Essays on Financial Econometrics

by

Yuan Xue

Department of Economics
Duke University

Date: ______________________

Approved:

__________________________
Tim Bollerslev, Co-supervisor

__________________________
Jia Li, Co-supervisor

__________________________
George Tauchen

__________________________
Federico Bugni

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2018
Abstract

Essays on Financial Econometrics

by

Yuan Xue

Department of Economics
Duke University

Date: ______________________
Approved: ______________________

____________________________
Tim Bollerslev, Co-supervisor

____________________________
Jia Li, Co-supervisor

____________________________
George Tauchen

____________________________
Federico Bugni

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University 2018
Abstract

This dissertation consists of three essays. The first essay, "Volume, volatility and Macroeconomic Announcements” studies the relationship between trading intensity and price volatility and how it is affected by investors’ disagreement on a common public signal around macroeconomic announcements. Inspired by a difference-of-opinion model in which investors agree to disagree, we use high frequency data and empirically show that the volume-volatility elasticity of S&P 500 ETF is uniformly below 1. Besides, the elasticity decreases with disagreement measures such as the forecast dispersion on unemployment rate and uncertainty measures, as well as a textual based tone measure constructed using FOMC statements. This paper provides new evidences on how information is processed in financial market.

The second essay, "Investor Sentiment and Volume Volatility Relationship” shows that investor sentiment plays a role on information processing in financial markets. We incorporate a one-factor asset pricing model into the difference-of-opinion model to derive the volume-volatility relationship for individual stocks. We separate the sample into high and low sentiment periods and use high frequency data to show that investors’ disagreement measures only significantly reduce volume-volatility elasticity around macroeconomic announcements during high-sentiment periods, for both the S&P 500 ETF and Dow Jones 30 components. This result is consistent with changes
in the confidence level of investors when sentiment regime shifts. Our estimates of
elasticity also decrease significantly with the ratio of idiosyncratic variance, which
indicates that higher idiosyncratic risks introduce larger dispersion among investors.

In the third essay, "Efficient Estimation of Integrated Functional of Variance with
Irregular Observation Time", we propose an efficient estimator of the integrated func-
tional of the variance with irregular observation time of prices. We propose the con-
sistency and central limit theorems, and then validate the theorems through proofs
and simulations.
To my husband, my daughter and my parents
Contents

Abstract iv
List of Tables xii
List of Figures xv
List of Abbreviations and Symbols xvii
Acknowledgements xviii

1 Introduction 1

2 Volume, Volatility and Public News Announcements 3
  2.1 Introduction .................................................. 4
  2.2 Theoretical motivation ........................................ 13
    2.2.1 New information and differences-of-opinion ............ 13
    2.2.2 Expected volume and volatility .......................... 15
  2.3 Data description and summary statistics .................... 19
    2.3.1 High-frequency market prices and trading volume .......... 19
    2.3.2 Macroeconomic news announcements ..................... 22
  2.4 A preliminary analysis of FOMC announcements .............. 24
  2.5 High-frequency econometric procedures ...................... 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6</td>
<td>Volume-volatility relationship around public announcements</td>
<td>38</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Jumps and announcements</td>
<td>38</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Volume-volatility elasticities around public news announcements</td>
<td>41</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Volume-volatility elasticities and disagreement measures</td>
<td>43</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Further analysis of FOMC announcements</td>
<td>46</td>
</tr>
<tr>
<td>2.6.5</td>
<td>Bond futures and FOMC announcements</td>
<td>51</td>
</tr>
<tr>
<td>2.7</td>
<td>Conclusion</td>
<td>52</td>
</tr>
</tbody>
</table>

3 Investor Sentiment and Volume-Volatility Relationship 57

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>58</td>
</tr>
<tr>
<td>3.2</td>
<td>Theoretical motivation</td>
<td>65</td>
</tr>
<tr>
<td>3.2.1</td>
<td>A one factor structure with heterogeneous beliefs</td>
<td>65</td>
</tr>
<tr>
<td>3.2.2</td>
<td>The volume-volatility elasticity</td>
<td>67</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Testable hypotheses of the model</td>
<td>70</td>
</tr>
<tr>
<td>3.3</td>
<td>Data description and summary statistics</td>
<td>71</td>
</tr>
<tr>
<td>3.3.1</td>
<td>High-frequency market prices and trading volume</td>
<td>71</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Macroeconomic news announcements</td>
<td>71</td>
</tr>
<tr>
<td>3.3.3</td>
<td>The Sentiment Measure</td>
<td>72</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Explanatory Variables: Disagreement Measures</td>
<td>74</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Explanatory Variables: Cross-sectional Variables</td>
<td>78</td>
</tr>
<tr>
<td>3.4</td>
<td>High-frequency econometric procedures</td>
<td>80</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Price and volume process</td>
<td>81</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Volume and volatility jump</td>
<td>83</td>
</tr>
</tbody>
</table>
3.4.3 Fixed effect panel regression ........................................ 86
3.5 Volume-volatility relationship around public announcements ....... 90
  3.5.1 Log volatility and volume jumps .................................. 90
  3.5.2 Volume-volatility elasticities around public news announcements ........................................ 91
  3.5.3 Dow Jones 30 Components ......................................... 97
  3.5.4 Treasury Bond Futures ........................................... 102
3.6 Robustness Test ...................................................... 105
  3.6.1 Different cutoff points for sentiment periods ................. 105
  3.6.2 Other disagreement measures ................................... 106
3.7 Conclusion .......................................................... 108

4 Efficient Estimation of Integrated Functional of Variance with Irreg-
ular Observation Time .................................................... 116
  4.1 Introduction ......................................................... 116
  4.2 Set up .............................................................. 119
    4.2.1 The stock price process ....................................... 119
    4.2.2 The observation times ....................................... 120
  4.3 Estimator .......................................................... 122
  4.4 Asymptotics ......................................................... 124
    4.4.1 Consistency .................................................... 124
    4.4.2 Central Limit Theorem ....................................... 125
    4.4.3 Efficiency bound .............................................. 126
  4.5 Simulation .......................................................... 127
### 4.5.1 A simple sampling framework .................................................. 127
### 4.5.2 Coverage rate ................................................................. 128
### 4.5.3 Jacod et al. (2013) under an irregular setting ......................... 129
### 4.6 Conclusion ............................................................................. 132

#### 5 Conclusion .............................................................................. 133

#### A Appendix for Chapter 2 .......................................................... 135

##### A.1 Technical background for econometric procedures ............... 135

- A.1.1 Continuous-time setup and definitions .................................. 135
- A.1.2 Estimation and inference .................................................... 137
- A.1.3 Proofs .............................................................................. 140

##### A.2 Additional data description .................................................. 148

- A.2.1 Macroeconomic news announcements ................................... 148
- A.2.2 A detailed description for the FOMC sentiment measure ...... 148
- A.2.3 Volume intensity sorted by five-minute normalized returns ... 151
- A.2.4 Raw estimates without using the DID method ...................... 151
- A.2.5 Results for DIA ETF .......................................................... 152
- A.2.6 Alternative dispersion and uncertainty measures ................. 153
- A.2.7 Results using dispersion in interest rate forecasts ................. 154
- A.2.8 Robustness check for removing the FOMC announcement on April 18, 2001 ................................................................. 155
- A.2.9 Robustness checks with respect to the local window size $k_n$ .... 156
- A.2.10 Estimation results with the 1-minute observations before announcements removed .............................................................. 157
List of Tables

2.1 Summary statistics of high-frequency price returns and volume data . 19
2.2 Macroeconomic news announcements . . . . . . . . . . . . . . . . . . 23
2.3 Volume and volatility jumps around public news announcements . . . 39
2.4 Volume-volatility elasticities around public news announcements . . 41
2.5 Volume-volatility elasticity estimates and disagreement measures . . 54
2.6 Volume-volatility elasticity estimates around FOMC announcements . 55
2.7 Volume-volatility elasticity estimates around FOMC announcements
   for Treasury bond futures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
3.1 Categories of macroeconomic news announcements . . . . . . . . . . 72
3.2 Summary statistics of the explanatory variables . . . . . . . . . . . . 78
3.3 Volume-volatility elasticity estimates for S&P 500 ETF . . . . . . . . 92
3.4 Volume-volatility elasticity estimates for S&P 500 ETF around FOMC
   announcements . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
3.5 Volume-volatility elasticity estimates for Dow Jones 30 components . 100
3.6 Volume-volatility elasticity estimates for T-bond futures . . . . . . . 103
3.7 Volume-volatility elasticity estimates for S&P 500 ETF (different cutoff
   points for sentiment periods) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 107
3.8 Volume-volatility elasticity estimates for S&P 500 ETF around FOMC announcements (different cutoff points for sentiment periods) 108
3.9 Volume-volatility elasticity estimates around for Dow Jones 30 components (different cutoff points for sentiment periods) 111
3.10 Volume-volatility elasticity estimates for S&P 500 ETF 112
3.11 Volume-volatility elasticity estimates for S&P 500 ETF around FOMC announcements 113
3.12 Volume-volatility elasticity estimates for Dow Jones 30 components 114
3.13 Volume-volatility elasticity estimates for T-bond futures 115
4.1 Confidence Interval Coverage Rate 130
4.2 Consistency: Average difference from Average theoretical value (in %) 131
4.3 Central limit theorem: 99%, 95% and 90% coverage rate 132
A.1 Other macroeconomic announcements 149
A.2 Word counts in FOMC Sentiment 150
A.3 Volume-volatility elasticities around public news announcements (no DID) 153
A.4 Volume-volatility elasticity estimates and disagreement measures (no DID) 154
A.5 Volume-volatility elasticity estimates around FOMC announcements (no DID) 155
A.6 Volume-volatility elasticity estimates and disagreement measures (DIA) 156
A.7 Volume-volatility elasticity estimates around FOMC announcements (DIA) 157
A.8 Volume-volatility elasticity estimates and disagreement measures based on one-year-ahead unemployment dispersion 158
A.9 Volume-volatility elasticity estimates around FOMC announcements based on one-year-ahead unemployment dispersion ........................................... 159
A.10 Volume-volatility elasticity estimates and disagreement measures based on one-year-ahead GDP dispersion ..................................................... 160
A.11 Volume-volatility elasticity estimates around FOMC announcements based on one-year-ahead GDP dispersion ........................................... 161
A.12 Volume-volatility elasticity estimates and disagreement measures based on Monthly Policy ................................................................. 162
A.13 Volume-volatility elasticity estimates around FOMC announcements based on Monthly Policy ................................................................. 163
A.14 Volume-volatility elasticity estimates around FOMC announcements ................................................................. 163
A.15 Volume-volatility elasticity estimates around FOMC announcements ................................................................. 164
A.16 Volume-volatility elasticity estimates and disagreement measures ................................................................. 165
A.17 Volume-volatility elasticity estimates around FOMC announcements ................................................................. 166
A.18 Volume-volatility elasticities around public news announcements ................................................................. 166
A.19 Volume-volatility elasticity estimates and disagreement measures ................................................................. 167
A.20 Volume-volatility elasticity estimates around FOMC announcements ................................................................. 168
# List of Figures

2.1 Equilibrium volume-volatility relations ........................................ 16  
2.2 Time series of volume and volatility ............................................. 20  
2.3 Intraday patterns of volatility and volume ..................................... 21  
2.4 Average volume intensity and volatility jumps around FOMC announce-  ments ................................................................. 26  
2.5 Volume and volatility around FOMC announcements .......................... 27  
2.6 Sorted volume around FOMC announcements ................................... 29  
2.7 Volume and volatility jumps around FOMC announcements ............... 30  
3.1 The Sentiment Measure .............................................................. 73  
3.2 Explanatory Variables ............................................................... 77  
3.3 Analysts’ dispersion in different sentiment periods .......................... 80  
3.4 the $\gamma$-$\beta$ ratio in different sentiment periods .......................... 81  
3.5 Scatter plot of log volume jumps on log volatility jumps ................... 85  
3.6 Average log volatility jumps in different sentiment periods ............... 91  
3.7 Average log volume jumps in different sentiment periods ................. 93  
3.8 The Sentiment Measure .............................................................. 106  
4.1 The empirical CDF of number of trades on Jan 31st, 2014 ............... 118
4.2 The empirical CDF of and smoothed empirical CDF of
the number of trades of USCR on Jan 31st, 2014 . . . . . . . . . . . . 123

4.3 Illustration of $T(t)$ when $\alpha = 0.4$ . . . . . . . . . . . . . . . . . . . . 129

A.1 Sorted volume around FOMC announcements . . . . . . . . . . . . . 152

A.2 T-bond volume and volatility around FOMC announcements . . . . . 164
List of Abbreviations and Symbols

Symbols

- $b$: the drift of the Itô semimartingale
- $X^c$: the continuous (local) martingale part of the semimartingale
- $E$: expectation operator
- $\mathcal{F}$: sigma-field
- $P$: probability function
- $W$: the standard Brownian motion
- $\Rightarrow$: stable convergence in law for the Skorokhod topology
- $\rightarrow^p$: convergence in probability

Abbreviations

- Idio-Sys Ratio: the ratio of idiosyncratic volatility to the systematic volatility
- TAQ: the NYSE Trade and Quote dataset
- VIX: the CBOE Volatility Index
- RV: realized volatility
Acknowledgements

I would like to thank my advisors, Prof. Tim Bollerslev and Prof. Jia Li for the guidance, encouragement and advice they provided throughout my time as their student. I have been extremely lucky to have advisors who care so much about my work and respond to all my queries promptly and patiently. I would also like to thank other committee members, Prof. George Tauchen and Prof. Federico Bugni for providing feedbacks and suggestions on my research progress. Besides, I am thankful to all participants in the financial econometrics lunchgroup for asking questions that inspire my research a lot.

I must express my gratitude to Yuqian Li, my husband and soul mate, for his continued support and encouragement. I would also like to thank my parents, who not only raised me to what I am, but also provided selfless help so that I could concentrate on writing the dissertation. I also wish to thank my classmates and friends, who never turned down the invitation to my mock presentations. I want to especially thank Mengke Wang and Chao Yang for their suggestion on presenting Chapter 3.

Completing this work would have been all the more difficult were it not for the many generous fellowships provided by Duke University, namely the Denis S. Karnosky/William Yohe Fellowship Endowment Fund and Duke Summer Research Fellowship. I am also grateful to the fund provided by NSF.
1

Introduction

My dissertation is mainly about information processing in financial markets. In the literature of market microstructure, it is of crucial importance to understand how prices are formed around information release. However, it is almost impossible to directly observe investors’ trading behavior. To have a clue about what happens around the arrival of information, researchers rely on the records traders leave behind: volume and price change.

Volume is the number of shares traded in a certain period of time. Volatility is the standard deviation of price change. These two variables, along with the price change, are very crucial variables to be studied in information processing. Some of the earliest paper that measure the effect of information on volatility are Engle and Ng (1993) and Ederington and Lee (1993a). In recent years, there are still growing literature on the reaction of prices to news announcements; see, e.g., Chan (2003), Vega (2006)
and Birz and Lott Jr (2011). Alongside are papers investigating the effect of news on volumes, including Tetlock (2007) and Yuan (2015).

My dissertation studies information processing and provides empirical evidences from the perspective of volume-volatility relationship. Chapter 2 investigates the volume-volatility elasticity of the market portfolio around macroeconomic announcements and show that it decreases with measures of investor’s disagreement. As a complement and extension, Chapter 3 studies the volume-volatility elasticity for the individual stock, and show that only during high sentiment periods will the measure of disagreement significantly decrease volume-volatility elasticity, which links the volume-volatility literature to the behavioral finance literature. Finally, Chapter 4 proposes an efficient estimator for the integrated functional of Variance with Irregular Observation Time. This estimator will give more efficient estimate of functional of variances for illiquid stocks, and can be used in future research on the effect of news on volatility.
Joint Work with Tim Bollerslev and Jia Li

**Summary:** We provide new empirical evidence for the way in which financial markets process information. Our results rely critically on high-frequency intraday price and volume data for the S&P 500 equity portfolio and U.S. Treasury bonds, along with new econometric techniques, for making inference on the relationship between trading intensity and spot volatility around public news announcements. Consistent with the predictions derived from a theoretical model in which investors agree to disagree, our estimates for the intraday volume-volatility elasticity around important news announcements are systematically below unity. Our elasticity estimates also decrease significantly with measures of disagreements in beliefs, economic uncertainty, and textual-based sentiment, further highlighting the key role played by differences-of-opinion.
Yuan Xue contributes to part of literature review, model and most of the coding and empirical analysis.

**Keywords:** Differences-of-opinion; high-frequency data; jumps; macroeconomic news announcements; trading volume; stochastic volatility; economic uncertainty; textual sentiment.

**JEL classification:** C51, C52, G12

2.1 Introduction

Trading volume and return volatility in financial markets typically, but not always, move in tandem. By studying the strength of this relationship around important public news announcements, we shed new light on the way in which financial markets function and process new information. Our empirical investigations rely critically on the use of high-frequency intraday price and volume data for the S&P 500 equity portfolio and U.S. Treasury bonds, together with new econometric inference procedures explicitly designed to deal with the unique complications that arise in the high-frequency data setting. Consistent with the implications derived from a stylized theoretical model in which investors agree to disagree, our estimates for the intraday volume-volatility elasticity around important news announcements are systematically below unity, the benchmark case that is obtained in the absence of any disagreement among investors. In line with the theoretical predictions from the same model, our estimates for the elasticity also decrease significantly with proxies for disagreements in beliefs, economic uncertainty, and textual-based sentiment, further corroborating the key role played by differences-of-opinion.
An extensive empirical literature has documented the existence of an on-average strong contemporaneous relation between trading volume and volatility; see Karpoff (1987) for a survey of some of the earliest empirical evidence. The mixture-of-distributions hypothesis (MDH) (see, e.g., Clark, 1973; Tauchen and Pitts, 1983; Andersen, 1996) provides a possible statistical explanation for the positive volume-volatility relationship based on the idea of a common news arrival process driving both the magnitude of returns and trading volume.\footnote{The MDH formally posits that price changes and trading volume are both subordinated to the same latent information, or news, arrival process. In its simplest form the hypothesis implies that conditional on the number of information arrivals observed over a non-trivial time interval, the corresponding price change and volume will follow a bivariate normal distribution.} The MDH, however, remains silent about the underlying economic mechanisms that link the actual trades and price adjustments to the news.

Meanwhile, a variety of equilibrium-based economic models have been developed to help better understand how prices and volume respond to new information. This includes the rational-expectations type models of Kyle (1985) and Kim and Verrecchia (1991, 1994) among many others, in which investors agree on the interpretation of the news, but their information sets differ. In this class of models the trading volume is mainly determined by liquidity trading and portfolio rebalancing needs. Although this is able to explain the on-average positive correlation between volume and volatility, the underlying trading motives would seem too small to account for the large trading volume observed empirically, especially when the price changes are close to zero (see, e.g., Hong and Stein, 2007, for additional arguments along these lines). Instead, models that feature differences-of-opinion, including those by Harrison and Kreps (1978), Harris and Raviv (1993), Kandel and Pearson (1995), Scheinkman and Xiong
(2003) and Banerjee and Kremer (2010) among others, in which investors agree to disagree, may help explain these oft observed empirical phenomena. In the differences-of-opinion class of models, the investors’ interpretation of the news and their updated valuations of the assets do not necessarily coincide, thereby providing an additional trading motive that is not directly tied to changes in the equilibrium price.

Much of empirical evidence presented in the literature in regards to the economic models discussed above, and the volume-volatility relationship in particular, have been based on daily or coarser frequency data (see, e.g., Tauchen and Pitts, 1983; Andersen, 1996).\(^2\) Meanwhile, another large and growing strand of literature has emphasized the advantages of the use of high-frequency intraday data for analyzing the way in which financial markets respond to new information and more accurately identifying price jumps. Harvey and Huang (1991) and Ederington and Lee (1993b), in particular, both find that the strong intraday volatility patterns observed in most financial markets may in part be attributed to regularly scheduled macroeconomic news announcements. Correspondingly large price jumps are often readily associated with specific news announcements; see, e.g., Andersen et al. (2007), Lee and Mykland (2008) and Lee (2012).\(^3\) This naturally suggests that by “zooming in” and analyzing how not only prices but also trading volume and volatility evolve around important

\(^2\) One notable exception is Chaboud et al. (2008), who document large trading volume in the foreign exchange market in the minutes immediately before macroeconomic announcements, even when the announcements are in line with market expectations and the actual price changes are small. In a recent work, Crego (2017) also uses high-frequency data to study the effect of the Weekly Petroleum Status Reports on bid-ask spreads, trading volume and returns for “oil firms” and “non-oil firms.”

\(^3\) Related to this, Savor and Wilson (2013) also document higher average excess market returns on days with important macroeconomic news releases compared to non-announcement days.
news announcements, a deeper understanding of the economic mechanisms at work and the functioning of markets may be forthcoming.

Set against this background, we provide new empirical evidence on the volume-volatility relationship around various macroeconomic news announcements. Our leading empirical investigations are based on high-frequency one-minute data for the aggregate S&P 500 equity market portfolio, but we also provide complimentary results for U.S. interest rates using one-minute Treasury bond futures data. We begin by documenting the occurrence of large increases in trading volume intensity around Federal Open Market Committee (FOMC) meetings without accompanying large price jumps. As noted above, this presents a challenge for models in which investors rationally update their beliefs based on the same interpretation of the news, and instead points to the importance of models allowing for disagreements or, differences-of-opinion, among investors.

To help further explore this thesis and guide our more in-depth empirical investigations, we derive an explicit expression for the elasticity of expected trading volume with respect to price volatility within the Kandel and Pearson (1995) differences-of-opinion model. We purposely focus our analysis on the elasticity as it may be conveniently estimated with high-frequency data and, importantly, has a clear economic interpretation in terms of model primitives. In particular, we show theoretically that the volume-volatility elasticity is monotonically decreasing in a well-defined measure of relative disagreement. Moreover, the elasticity is generally below one and reaches its upper bound of unity only in the benchmark case without disagreement.

The theoretical model underlying these predictions is inevitably stylized, focusing exclusively on the impact of public news announcements. As such, the theory mainly
speaks to the “abnormal” movements in volume and volatility observed around these news events. To identify the abnormal movements, and thus help mitigate the effects of other confounding forces, we rely on the “jumps” in the volume intensity and volatility around the news announcements. Our estimation of the jumps is based on the differences between the post- and pre-event levels of the instantaneous volume intensity and volatility, which we recover non-parametrically using high-frequency data. Even though the differencing step used in identifying the jumps effectively removes low-frequency dynamics in the volatility and volume series (including daily and lower frequency trending behavior) that might otherwise confound the estimates, the jump estimates are still affected by the well-documented strong intraday periodic patterns that exist in both volume and volatility (for some of the earliest empirical evidence, see Wood et al., 1985; Jain and Joh, 1988). In an effort to remove this additional confounding effect, we apply a second difference with respect to a control group of non-announcement days. The resulting “doubly-differenced” jump estimates in turn serve as our empirical analogues of the abnormal volume and volatility movements that we use in our regression-based analysis of the theoretical predictions.

Our empirical strategy for estimating the jumps may be viewed as a Difference-in-Difference (DID) type approach, as commonly used in empirical microeconomic studies (see, e.g., Ashenfelter and Card, 1985). The subsequent regression involving the jumps is similar in spirit to the “jump regressions” studied by Li et al. (2017), which in turn resembles the nonparametric estimation in (fuzzy) Regression Discontinuity Designs (RDD) (see, e.g., Lee and Lemieux, 2010).4 We hence refer to our

4 In parallel to Li et al. (2017), our “jump regressions” are designed to estimate equilibrium relationships between economic variables manifest in unusually large moves, or jumps, in those variables, rather than using the jumps (i.e, the discontinuities) for identifying causal treatment
new econometric method as a DID jump regression. However, our setup is distinctly different from conventional econometric settings, and the usual justification for the use of DID or RDD does not apply in the high-frequency data setting. Correspondingly, our new econometric procedures and the justification thereof entail two important distinctions. Firstly, to accommodate the strong dynamic dependencies in the volatility and volume intensity, we provide a rigorous theoretical justification based on a continuous-time infill asymptotic framework allowing for essentially unrestricted non-stationarity. Secondly, we provide an easy-to-implement local bootstrap method for conducting valid inference. By randomly resampling only locally in time (separately before and after each announcement), the method provides a simple solution to the issue of data heterogeneity, which otherwise presents a formidable challenge in the high-frequency data setting (see, e.g., Gonçalves and Meddahi, 2009).

Our actual empirical findings are closely in line with the theoretical predictions derived from the Kandel and Pearson (1995) model and the differences-of-opinion class of models more generally. In particular, we first document that the estimated volume-volatility elasticity around FOMC announcements is significantly below unity. This finding carries over to other important intraday public news announcements closely monitored by market participants. Interestingly, the volume-volatility elasticity estimates are lower for announcements that are released earlier in the monthly news cycle (see, e.g., Andersen et al., 2003), such as the ISM Manufacturing Index and the Consumer Confidence Index, reflecting the importance of the timing across the different announcements and the effect of learning.

effects. Hence, in spite of the similarity between the two procedures, our goal is distinctly different from that of RDD.
Going one step further, we show that the intraday volume-volatility elasticity around news announcements decreases significantly in response to increases in measures of dispersions-in-beliefs (based on the survey of professional forecasters as in, e.g., Van Nieuwerburgh and Veldkamp, 2006; Pasquariello and Vega, 2007) and economic uncertainty (based on the economic policy uncertainty index of Baker et al., 2015). This holds true for the S&P 500 aggregate equity portfolio as well as the U.S. Treasury bond market, and again corroborates our theoretical predictions and the key role played by differences-of-opinion. Our more detailed analysis of FOMC announcements, in which we employ an additional textual-based measure for the negative sentiment in the accompanying FOMC statements (based on the methodology of Loughran and McDonald, 2011), further underscores the time-varying nature of the high-frequency volume-volatility relationship and the way in which the market processes new information: when the textual sentiment in the FOMC statement is more negative, the relative disagreement among investors also tends to be higher, pushing down the volume-volatility elasticity.

In contrast to prior empirical studies related to the volume-volatility relationship (see, e.g., Tauchen and Pitts, 1983; Andersen, 1996), our analysis is much more closely guided by an economic model. In particular, based on the implications derived from the Kandel–Pearson model, we provide a new perspective on the volume-volatility relationship by directly linking the strength of the relationship to notions of investors’ disagreement. Since the Kandel–Pearson model explicitly concerns “abnormal” variations in volume and volatility around news announcements, we also derive the new DID jump regression framework in order to rigorously analyze the joint behavior of volume and volatility jumps. The use of high-frequency data are crucial in this re-
gard, as jumps are invariably short-lived in nature and would be difficult/impossible to accurately identify using data sampled at coarser, say daily, frequencies.

Putting the empirical results in the paper into a broader perspective, there is a large literature in market microstructure finance on the price impact of trades, and correspondingly the development of “optimal” trade execution strategies; empirical work along these lines include Hasbrouck (1991), Madhavan et al. (1997), and Chordia et al. (2002). In contrast to the volume-volatility relationship analyzed here, the price impact literature is explicitly concerned with directional price changes predicted by “signed” trading volume, or order flow. Green (2004), in particular, documents a significant increase in the impact of changes in order flow in the U.S. Treasury bond market on intraday bond prices immediately following macroeconomic news announcements, and goes on to suggest that this heightened price impact of trades may be attributed to increased informational asymmetry at the time of the announcements.5 Further corroborating this idea, Pasquariello and Vega (2007) find that the regression-based estimate for the impact of unanticipated daily order flow in the U.S. Treasury bond market is higher when the dispersion in beliefs among market participants, as measured by the standard deviation of the forecasts from professional forecasters, is high and when the public news announcement is more “noisy.” Our main empirical findings based on high-frequency intraday data for the S&P 500 aggregate market portfolio are generally in line with these existing results about the price impact of order flow in the U.S. Treasury bond market. At the same time, however, our focus on trading volume as opposed to order flow presents an important

5 This is consistent with the predictions from the theoretical model in Kim and Verrecchia (1994), in which earnings announcements may increase informational asymmetry due to the superior information processing skills of certain traders.
distinction from the aforementioned studies. In line with the insights of Hong and Stein (2007), by focusing on trading volume our empirical findings map more closely into the theoretical predictions from differences-of-opinion class of models in which the disagreement among traders provides an important trading motive beyond conventional rational-expectation type models (as, e.g., the classical model by Kyle, 1985). In addition, our empirical strategy relies critically on the use of high-frequency data coupled with new econometric techniques for non-parametrically uncovering both the instantaneous trading intensity and volatility under minimal statistical assumptions.

The rest of the paper is organized as follows. Section 2.2 presents the basic economic arguments and theoretical model that guide our empirical investigations. Section 2.3 describes the high-frequency intraday data and news announcements used in our empirical analysis. To help set the stage for our more detailed subsequent empirical investigations, Section 2.4 discusses some preliminary findings specifically related to the behavior of the aggregate stock market around the FOMC announcements. Section 2.5 describes the new inference procedures necessitated by our more in-depth high-frequency empirical analysis. Section 2.6 presents our main empirical findings based on the full set of news announcements, followed by our more detailed analysis of FOMC announcements. In addition to our main results based on data for the S&P 500 aggregate equity portfolio, we also present complementary results for the U.S. Treasury bond market. Section 2.7 concludes. Technical details concerning the new econometric inference procedure are provided in Appendix A.1.

As such, our empirical findings also do not speak directly to prior work that explicitly rely on net order flow data. Instead, we offer a new perspective on the widely studied volume-volatility relationship.
A.2 contains further data descriptions. Additional empirical results and robustness checks are relegated to a (not-for-publication) supplemental appendix.

2.2 Theoretical motivation

We rely on the theoretical volume-volatility relations derived from the differences-of-opinion model of Kandel and Pearson (1995) to help guide our empirical investigations. We purposely focus on a simplified version of the model designed to highlight the specific features that we are after, and the volume-volatility elasticity around news arrivals in particular. We begin by discussing the basic setup and assumptions.

2.2.1 New information and differences-of-opinion

Following Kandel and Pearson (1995), henceforth KP, we assume that a continuum of traders trade a risky asset and a risk-free asset in a competitive market. The random payoff of the risky asset, denoted $\tilde{u}$, is unknown to the traders. The risk-free rate is normalized to be zero. The traders’ utility functions have constant absolute risk aversion with risk tolerance $r$. There are only two types of traders, $i \in \{1, 2\}$, with the proportion of type 1 traders denoted $\alpha$. The traders have different prior beliefs about the payoff before the announcement, and they also disagree about the interpretation of the public signal at the time of the announcement. Type $i$ trader’s prior is given by a normal distribution with precision $s_i$.

After the announcement, the traders observe the same public signal $\tilde{u} + \tilde{\varepsilon}$, where the noise term $\tilde{\varepsilon}$ is normally distributed. The traders then update their beliefs about $\tilde{u}$ and optimally re-balance their positions. The key feature of the KP model is that the two types of traders agree to disagree on how to interpret the public signal when
updating their beliefs about the asset value: type \( i \) traders believe that \( \bar{\varepsilon} \) is drawn from the \( N(\mu_i, h^{-1}) \) distribution. Differences-of-opinion regarding the public signal among the traders thus corresponds to \( \mu_1 \neq \mu_2 \).

Following KP it is possible to show that in equilibrium\(^7\)

\[
\text{Volume} = |\beta_0 + \beta_1 \cdot \text{Price Change}|, \tag{2.1}
\]

where
\[
\beta_0 = r\alpha (1 - \alpha) h (\mu_1 - \mu_2), \quad \beta_1 = r\alpha (1 - \alpha) (s_1 - s_2). \tag{2.2}
\]

The coefficient \( \beta_0 \) is directly associated with the degree of differences-of-opinion concerning the interpretation of the public signal (i.e., \( \mu_1 - \mu_2 \)), while \( \beta_1 \) depends on the dispersion in the precisions of the prior beliefs about the payoff (i.e., \( s_1 - s_2 \)). Both of the coefficients are increasing in the degree of risk tolerance \( r \), and the degree of heterogeneity among the traders as measured by \( \alpha(1 - \alpha) \).

Looking at the equilibrium relationship in equation (2.1), the first \( \beta_0 \) term represents the “disagreement component.” This term becomes increasingly more important for higher levels of disagreement (i.e., \( |\mu_1 - \mu_2| \) is large) and/or when the traders are more confident about their interpretation of the public signal (i.e., \( h \) is large). Hence, other things being equal, the higher the level of disagreement, the weaker the relationship between trading activity and price changes. In the extreme case when the equilibrium price does not change as a result of the announcement and the second term on the right-hand side of (2.1) equals zero, there can still be large trading volume arising from disagreement among the traders because of the \( \beta_0 \) term. In this sense,

\(^7\) See equation (5) in Kandel and Pearson (1995).
disagreement generates an additional endogenous trading motive that is effectively “orthogonal” to any revision in the equilibrium price.

2.2.2 Expected volume and volatility

The implication of the KP model for the relationship between price adjustment and trading volume in response to new information is succinctly summarized by equations (2.1) and (2.2). These equations, however, depict an exact functional relationship between (observed) random quantities. A weaker, but empirically more realistic, implication can be obtained by thinking of this equilibrium relationship as only holding “on average.” Moment conditions corresponding to the stochastic version (2.1) formally capture this idea.

Specifically, let \( m(\sigma) \) denote the expected volume as a function of the volatility \( \sigma \) (i.e., the standard deviation of price change). We assume that the price changes are normally distributed with mean zero and standard deviation \( \sigma \). In particular, we note that the zero-mean assumption is empirically sound for high-frequency returns (see Table 2.1). It follows by direct integration of (2.1) that

\[
m(\sigma) = \sqrt{\frac{2}{\pi}} |\beta_1| \sigma \exp \left( -\frac{\beta_0^2}{2\beta_1^2 \sigma^2} \right) + |\beta_0| \left( 2\Phi \left( \frac{|\beta_0|}{|\beta_1| \sigma} \right) - 1 \right),
\]

where \( \Phi \) denotes the cumulative distribution function of the standard normal distribution. The expected volume \( m(\sigma) \) depends on \( \sigma \) and the \((\beta_0, \beta_1)\) coefficients in a somewhat complicated fashion. However, it is straightforward to show that \( m(\sigma) \) is increasing in \( \sigma \).

In order to gain further insight regarding this (expected) volume-volatility relationship, Figure 2.1 illustrates how the \( m(\sigma) \) function varies with the “disagreement...
Figure 2.1: Equilibrium volume-volatility relations

Notes: The figure shows the equilibrium relationship between expected trading volume $m$ and price volatility $\sigma$ in the Kandel–Pearson model for various levels of disagreement, ranging from $\beta_0 = 0$ (bottom) to 0.5, 1 and 1.5 (top). $\beta_1$ is fixed at one in all of the graphs.

Locally, when the volatility $\sigma$ is close to zero, the expected volume is positive if and only if the opinions of the traders differ (i.e., $\mu_1 \neq \mu_2$). Globally, as $\beta_0$ increases (from the bottom to the top curves in the figure), the equilibrium relationship between the expected volume and volatility “flattens out.” This pattern is consistent with the aforementioned intuition that disagreement among traders provide an additional trading motive which loosens the relationship between volume and volatility.

The exact non-linear expression in equation (2.3) inevitably depends on the specific assumptions and setup underlying the KP model. Although this expression
helps formalize the intuition about the way in which disagreement affects the volume-
volatility relationship, the KP model is obviously too stylized to allow for a direct
structural estimation. In our empirical analysis below, we therefore rely on a reduced-
form approach. In so doing, we aim to test the basic economic intuition and empirical
implications stemming from the differences-of-opinion class of models more generally. From an empirical perspective this is naturally accomplished by focusing on
the volume-volatility relationship expressed in terms of the elasticity of \( m(\sigma) \) with
respect to \( \sigma \).

Not only does this volume-volatility elasticity provide a convenient “reduced-form”
summary statistic, it also admits a clear economic interpretation within the KP model.
Let \( \mathcal{E} \) denote the volume-volatility elasticity. A straightforward calculation then yields

\[
\mathcal{E} \equiv \frac{\partial m(\sigma)/m(\sigma)}{\partial \sigma/\sigma} = \frac{1}{1 + \psi(\gamma/\sigma)},
\]

where

\[
\gamma \equiv \left| \frac{\beta_0}{\beta_1} \right| = \frac{h |\mu_1 - \mu_2|}{|s_1 - s_2|},
\]

and the function \( \psi \) is defined by \( \psi(x) \equiv x (\Phi(x) - 1/2)/\phi(x) \), with \( \phi \) being the density function of the standard normal distribution. The function \( \psi \) is strictly increasing
on \([0, \infty)\), with \( \psi(0) = 0 \) and \( \lim_{x \to \infty} \psi(x) = \infty \).

The expressions in (2.4) and (2.5) embody two key features in regards to the
volume-volatility elasticity. Firstly, \( \mathcal{E} \leq 1 \) with the equality and an elasticity of unity
obtaining if and only if \( \gamma = 0 \). Secondly, \( \mathcal{E} \) only depends on and is decreasing in \( \gamma/\sigma \).
This second feature provides a clear economic interpretation of the volume-volatility
elasticity \( \mathcal{E} \): it is low when differences-in-opinion is \emph{relatively} high, and vice versa,
with $\gamma/\sigma$ serving as the relative measure of the differences-of-opinion. This relative measure is higher when traders disagree more on how to interpret the public signal (i.e., larger $|\mu_1 - \mu_2|$) and with more confidence (i.e., larger $h$), relative to the degree of asymmetric private information (i.e., $|s_1 - s_2|$) and the overall price volatility (i.e., $\sigma$).

These two features in turn translate into directly testable implications that we use to guide our empirical analysis. In particular, taking the KP model and the expressions in (2.4) and (2.5) at face value, it is possible to test for the presence of differences-of-opinion among traders by testing whether the volume-volatility elasticity around important news arrivals is less than or equal to unity. This strict implication, however, hinges on a number of specific parametric and distributional modeling assumptions.\(^8\) We therefore also investigate the second more qualitative implication arising from the model that the volume-volatility elasticity is decreasing with the overall level of disagreement. This implication reflects the more general economic intuition that disagreement among traders provides an extra trading motive, and as such this implication should hold true more broadly. To allow for a focused estimation of the elasticities, we base our empirical investigations on intra-day high-frequency transaction data around well-defined public macroeconomic news announcements, along with various proxies for the heterogeneity in beliefs and economic uncertainty prevailing at the exact time of the announcements. We turn next to a discussion of the data that we use in doing so.

\(^8\) Among others, the assumption of CARA and normally distributed prior beliefs.
Table 2.1: Summary statistics of high-frequency price returns and volume data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Min</th>
<th>1%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>99%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return (percent)</td>
<td>0.000</td>
<td>-2.014</td>
<td>-0.149</td>
<td>-0.021</td>
<td>0.000</td>
<td>0.021</td>
<td>0.149</td>
<td>2.542</td>
</tr>
<tr>
<td>Volume (1,000 shares)</td>
<td>289.5</td>
<td>0.0</td>
<td>2.4</td>
<td>55.4</td>
<td>147.6</td>
<td>350.7</td>
<td>2005</td>
<td>31153</td>
</tr>
</tbody>
</table>

Notes: The table reports summary statistics of the one-minute returns and one-minute trading volumes for the SPY ETF during regular trading hours from April 10, 2001 to September 30, 2014.

2.3 Data description and summary statistics

Our empirical investigations are based on high-frequency intraday transaction prices and trading volume, together with precisely timed macroeconomic news announcements. We describe our data sets in turn.

2.3.1 High-frequency market prices and trading volume

Our primary data is comprised of intraday transaction prices and trading volume for the S&P 500 index ETF (ticker: SPY). All of the data are obtained from the TAQ database. The sample covers all regular trading days from April 10th, 2001 through September 30th, 2014. The raw data are cleaned following the procedures detailed in Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). Further, in order to mitigate the effect of market microstructure noise, we follow standard practice in the literature to sparsely sample the data at a one-minute sampling interval, resulting in a total of 1,315,470 one-minute return and volume observations.

Summary statistics for the SPY returns and trading volumes (number of shares) are reported in Table 2.1. Consistent with prior empirical evidence (see, e.g. Bollerslev and Todorov, 2011), the high-frequency one-minute returns appear close to be symmetrically distributed. The one-minute volume series, on the other hand, is highly skewed to the right, with occasionally very large values.
Figure 2.2: Time series of volume and volatility

Notes: The figure shows the daily logarithmic trading volume (top panel) and logarithmic realized volatility (bottom panel) for the SPY ETF. The daily volume (in millions) is constructed by accumulating the intraday volume. The daily realized volatility (annualized in percentage) is constructed as the sum of one-minute squared returns over the day.

To highlight the general dynamic dependencies inherent in the data, Figure 2.2 plots the daily logarithmic trading volume (constructed by summing the one-minute trading volumes over each of the different days) and the logarithmic daily realized volatilities (constructed as the sum of squared one-minute returns over each of the days in the sample). As the figure shows, both of the daily series vary in a highly predictable fashion. The volume series, in particular, seems to exhibit an upward trend over the first half of the sample, but then levels off over the second half. Meanwhile, consistent with the extensive prior empirical evidence discussed above, there are strong dynamic commonalities evident in the two series.
The volume and volatility series also both exhibit strong intraday patterns. To illustrate this, Figure 2.3 plots the square-root of the one-minute squared returns averaged across each minute-of-the-day (as an estimate for the volatility over that particular minute) and the average trading volume over each corresponding minute. In order to prevent abnormally large returns and volumes from distorting the picture, we only include non-announcement days that are discussed in Section 2.3.2 below. Consistent with the evidence in the extant literature, there is a clear U-shaped pattern in the average volatility and trading activity over the active part of the trading day.\footnote{See Wood et al. (1985), Harris (1986), Jain and Joh (1988), Baillie and Bollerslev (1990), and Andersen and Bollerslev (1997) for some of the earliest empirical evidence on the intraday patterns in volatility and volume.}

**Figure 2.3: Intraday patterns of volatility and volume**

*Notes:*

The figure shows the intraday volatility (annualized in percentage) for the SPY ETF (left panel) constructed as the square root of the one-minute squared returns averaged across all non-announcement days, along with the best quadratic fit. The intraday trading volume (in shares) for the SPY ETF (right panel) is similarly averaged across all non-announcement days.
In addition to our main empirical results based on the high-frequency intraday data for the S&P 500 aggregate equity portfolio discussed above, we also report complementary empirical evidence for the U.S. Treasury bond market. Our intraday price and volume data for the ten-year U.S. T-bond futures are obtained from TickData, and spans the slightly shorter sample period from July 1, 2003 to September 30, 2014. The intraday T-bond volume and volatility series naturally exhibit their own distinct dynamic dependencies and intraday patterns. However, as discussed further below, the key features pertaining to the volume and volatility “jumps” observed around important macroeconomic news announcements closely mirror those for the aggregate equity market.

2.3.2 Macroeconomic news announcements

The Economic Calendar Economic Release section in Bloomberg includes the date and exact within day release time for over one hundred regularly scheduled macroeconomic news announcements. Most of these announcements occur before the market opens or after it closes. We purposely focus on announcements that occur during regular trading hours only.\(^{10}\) While this leaves out some key announcements (most notably the monthly employment report), importantly it allows us to accurately estimate volume and volatility “jumps” by harnessing the rich information inherent in intraday high-frequency data about the way in which markets process new information. In addition to the intraday announcements pertaining to specific macroeconomic variables and indicators, we also consider FOMC rate decisions. FOMC announcements have

\(^{10}\) To ensure that there is a 30-minute pre-event (resp. post-event) window before (resp. after) each announcement, we also exclude announcements that are released during the first and the last 30 minutes of the trading day.
Table 2.2: Macroeconomic news announcements

<table>
<thead>
<tr>
<th></th>
<th>No.Obs</th>
<th>Time</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOMC</td>
<td>109</td>
<td>Varies</td>
<td>Federal Reserve Board</td>
</tr>
<tr>
<td>ISM Manufacturing</td>
<td>160</td>
<td>10:00</td>
<td>Institute of Supply Management</td>
</tr>
<tr>
<td>ISM Non-Manufacturing</td>
<td>158</td>
<td>10:00</td>
<td>Institute of Supply Management</td>
</tr>
<tr>
<td>Consumer Confidence</td>
<td>160</td>
<td>10:00</td>
<td>Conference Board</td>
</tr>
<tr>
<td>Other Indicators</td>
<td>1682</td>
<td>Varies</td>
<td>See Appendix A.2.1</td>
</tr>
</tbody>
</table>

Notes: The table reports the total number of observations, release time, and data source for each of the news announcements over the April 10, 2001 to September 30, 2014 sample.

been analyzed extensively in the existing literature, and we also pay special attention to those announcements in our empirical analysis discussed below.

Based on prior empirical evidence (see, e.g. Andersen et al., 2003; Boudt and Petitjean, 2014; Jiang et al., 2011; Lee, 2012), we identify four announcements as being the most important overall in the sense of having the on-average largest price impact among all of the regularly scheduled intraday news announcements. In addition to the aforementioned FOMC rate decisions, this includes announcements pertaining to the ISM Manufacturing Index (ISMM), the ISM Non-Manufacturing Index (ISMNW), and the Consumer Confidence (CC) Index. Table 3.1 provides the typical release times and the number of releases over the April 10, 2001 to September 30, 2014 sample period for each of these important announcements.\(^\text{11}\) The remaining announcements are categorized as “Others,” a full list of which is provided in Table A.1 in Appendix A.2. Some of the announcement times for these other indicators invariably coincide,\(^\text{11}\) The Economic Release section in Bloomberg reports 108 FOMC meetings during our sample period. However, there are also additional FOMC intermeeting announcements. All of these occurred outside the intraday trading hours, except the one at 10:55am on April 18, 2001. Including this one additional intermeeting announcement, thus leaves us with a total 109 FOMC announcements. Almost identical empirical results to the ones reported below are obtained by excluding this one FOMC intermeeting observation.
so that all-in-all our sample is comprised of a total of 2,130 unique intraday public news announcement times.

2.4 A preliminary analysis of FOMC announcements

To set the stage for our more in-depth subsequent empirical investigations, we begin by presenting a set of simple summary statistics and illustrative figures related to the volume-volatility relationship around FOMC announcements. We focus our preliminary analysis on FOMC announcements, because these are arguably among the most important public news announcements that occur during regular trading hours.\footnote{The reaction of market prices to FOMC announcements has been extensively studied in the recent literature; see, for example, Johnson and Paye (2015) and the many references therein.}

For each announcement, let $\tau$ denote the pre-scheduled announcement time. The time $\tau$ is naturally associated with the integer $i(\tau)$ such that $\tau = (i(\tau) - 1)\Delta_n$, where $\Delta_n = 1$ minute is the sampling interval of our intraday data. We define the event window as $((i(\tau) - 1)\Delta_n, i(\tau)\Delta_n]$. Further, we define the pre-event (resp. post-event) window to be the $k_n$-minute period immediately before (resp. after) the event. We denote the return and trading volume over the $j$th intraday time-interval $((j-1)\Delta_n, j\Delta_n]$ by $r_j$ and $V_{j\Delta_n}$, respectively. The volume intensity $m$ (i.e., the instantaneous mean volume) and the spot volatility $\sigma$ before and after the announcement, denoted by $m_{\tau-}$, $m_\tau$, $\sigma_{\tau-}$ and $\sigma_\tau$, respectively, can then be estimated by

\[
\begin{align*}
\hat{m}_{\tau-} & \equiv \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} V_{(i(\tau)-j)\Delta_n}, & \hat{m}_\tau & \equiv \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} V_{(i(\tau)+j)\Delta_n}, \\
\hat{\sigma}_{\tau-} & \equiv \sqrt{\frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} r^2_{(i(\tau)-j)}}, & \hat{\sigma}_\tau & \equiv \sqrt{\frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} r^2_{(i(\tau)+j)}}.
\end{align*}
\]
These estimators are entirely nonparametric, in the sense that they only rely on simple averages of the data in local windows around the event time. The window size $k_n$ plays the same role as the bandwidth parameter in conventional nonparametric analysis and regressions; the standard technical assumptions needed to formally justify the estimators are further detailed in Appendix A.1.\textsuperscript{13} In the empirical analysis reported below we set $k_n = 30$, corresponding to a 30-minute pre-event (resp. post-event) window.\textsuperscript{14}

As an initial illustration of the volume and volatility jumps observed around FOMC announcements, Figure 2.4 plots the average (across all FOMC announcements in the sample) estimated volume intensity (top panel) and spot volatility (bottom panel) processes for the 15 minutes before and after the announcements. As the figure clearly shows, both volume and volatility sharply increase at the time of the announcement. Consistent with the recent work of Bernile et al. (2016) and Kurov et al. (2016) and the finding that some price-adjustment seemingly occurs in anticipation of the actual news release, there is also a slight increase in both series leading up to the news announcement time. However, this increase is clearly small compared to the “jumps” that manifest at the time of the announcement, and as such will not materially affect any of our subsequent empirical analysis.\textsuperscript{15}

\textsuperscript{13} Technically, the consistent estimation of the volume intensity and spot volatility before and after the announcements only requires the processes to be right-continuous with left-limits. Correspondingly, the local window parameter $k_n$ is assumed to satisfy $k_n \to \infty$ and $k_n \Delta_n \to 0$, which directly mirrors the usual regularity conditions on the bandwidth parameter in nonparametric regressions.

\textsuperscript{14} Our main empirical findings are not sensitive to the choice of $k_n$; see Appendix ?? for some robustness checks.

\textsuperscript{15} As a simple robustness check, we also estimated the pre-event quantities excluding the one-minute observation immediately before the announcement, resulting in virtually the same estimates as the ones reported below; further details are available in Appendix A.1.
Figure 2.4: Average volume intensity and volatility jumps around FOMC announcements

Notes: The figure plots the nonparametric estimates of the volume intensity (in shares) and volatility (annualized in percentage) before and after an “average” FOMC announcement. We compute $\hat{m}_{\tau+s}$ (resp. $\hat{\sigma}_{\tau+s}$) for $s$ ranging from 15 minutes before to 15 minutes after each FOMC announcement, and plot the averages across all of the announcements in the top (resp. bottom) panel. The local window is set to $k_n = 30$.

To examine whether the “average jumps” evident in Figure 2.4 are representative, Figure 2.5 plots the time series of individually estimated logarithmic volume intensities (top panel) and logarithmic spot volatilities (bottom panel) before and after each of the FOMC announcements. Consistent with the on-average estimates depicted in Figure 2.4, the figure shows marked bursts in the trading volume following each of

---

16 The logarithmic transform is naturally motivated by our interest in the volume-volatility elasticity derived in Section 2.2.2. The log-transformation also helps reduce the heteroskedasticity in both series. The heteroskedasticity in the spot volatility estimates is directly attributable to estimation errors. Formally, the standard error of the $\hat{\sigma}_{\tau}$ estimate equals $\hat{\sigma}_{\tau} / \sqrt{2k_n}$, so that by the delta-method the standard error of $\log(\hat{\sigma}_{\tau})$ equals the constant $1 / \sqrt{2k_n}$. In addition, the logarithm effectively transforms the salient multiplicative trend in the volume series over the first half of the sample to an additive trend, which as seen in Figure 2.2 is close to linear in time.
Figure 2.5: Volume and volatility around FOMC announcements

Notes: The figure plots the log-volume intensity (top panel, in shares) and log-volatility (bottom panel, in percentage) around FOMC announcements. The volume intensity and volatility are calculated using equation (3.9) with $k_n = 30$.

the FOMC announcements, accompanied by positive jumps in the volatility. These jumps in the volume intensity and volatility are economically large, with average jump sizes (in log) of 1.41 and 1.09, respectively. As discussed in Section 2.6, they are also highly statistically significant. This therefore suggests that traders revise their beliefs about the stock market differently upon seeing the FOMC announcement. It is, of course, possible that traders have asymmetric private information about the overall market. However, in line with the reasoning of Hong and Stein (2007) regarding the large observed burst in trading volume, it seems much more likely that the differences are attributed to the traders’ disagreement regarding the news.\footnote{It also appears highly unlikely that there is any “insider information” pertaining to the actual FOMC announcements.}
In order to further buttress the importance of differences-of-opinion among investors, we consider two additional empirical approaches. Firstly, as discussed in Section 2.2.1, if all investors agreed on the interpretation of the FOMC announcements, the trading volumes observed around the news releases should be approximately proportional to the price changes. Consequently, if there were no disagreement we would expect to see large trading volume accompanied by large changes in prices and vice versa. To investigate this hypothesis, we sort all of the FOMC announcements by the normalized one-minute event returns \( r_i / \hat{\sigma}_{\tau} \sqrt{\Delta n} \), and plot the resulting time series of pre- and post-event volume intensity estimates in Figure 2.6.\(^\text{18}\) Consistent with the findings of KP (based on daily data), Figure 2.6 shows no systematic association between trading volumes and returns. Instead, we observe many sizable jumps in the volume intensity for events with absolute returns “close” to zero, that is, when they are less than one instantaneous standard deviation (highlighted by the shaded area).

The empirical approach underlying Figure 2.6 mainly focuses on events with price changes close to zero and, hence, is local in nature. Our second empirical approach seeks to exploit a more global feature of the differences-of-opinion type models, namely that the volume-volatility elasticity should be below unity. While this prediction was derived from the explicit solution for the KP model in equation (2.3), the underlying economic intuition holds more generally: differences-of-opinion provides an additional trading motive that is not tied to the traders’ average valuation of the asset and, hence, serves to “loosen” the relationship between trading volume and volatility.

\(^{18}\) The normalization with respect to the spot volatility serves as a scale adjustment to make the returns across announcements more comparable. A similar figure based on the five-minute returns is included in Appendix A.1.
Figure 2.6: Sorted volume around FOMC announcements

Notes: The figure shows the pre- and post-event log volume intensities (in shares) sorted on the basis of the 1-minute normalized returns $r_{i(\tau)}/\hat{\sigma}_{\tau} - \Delta n$ around FOMC announcements (dots). The normalized return increases from left to right. Announcements with normalized returns less than 1 are highlighted by the shaded area.

In order to robustly examine this prediction for the volume-volatility elasticity, without relying on the specific functional form in (2.4), we adopt a less restrictive reduced-form estimation strategy. Further along those lines, the theoretical models discussed in Section 2.2 are inevitably stylized in nature, abstracting from other factors that might affect actual trading activity (e.g., liquidity or life-cycle trading, reduction in trading costs, advances in trading technology, to name but a few). As such, the theoretical predictions are more appropriately thought of as predictions about “abnormal” variations in the volume intensity and volatility. In the high-frequency data setting, abnormal movements conceptually translate into “jumps.”
Notes: The figure shows the scatter of the jumps in the log-volume intensity (in shares) versus the jumps in the log-volatility (in percentage) around FOMC announcements. The line represents the least-square fit.

Below, we denote the corresponding log-volatility and log-volume intensity jumps by 
\( \Delta \log(\sigma) \equiv \log(\sigma) - \log(\sigma_{-\infty}) \) and \( \Delta \log(m) \equiv \log(m) - \log(m_{-\infty}) \), respectively.

Figure 2.7 shows a scatter plot of the estimated \( \Delta \log(\sigma) \) and \( \Delta \log(m) \) jumps around FOMC announcement times. As expected, there is a clear positive association between the two series, with a correlation coefficient of 0.57. Moreover, consistent with the theoretical predictions and the idea that traders interpret FOMC announcements differently, the estimate for the volume-volatility elasticity implied by the slope coefficient from a simple linear fit equals 0.66, which is much less than unity.

The summary statistics and figures discussed above all corroborate the conjecture that differences-of-opinion among investors play an important role in the way in which
the market responds to FOMC announcements. To proceed with a more formal empirical analysis involving other announcements and explanatory variables, we need econometric tools for conducting valid inference, to which we now turn.

2.5 High-frequency econometric procedures

The econometrics in the high-frequency setting is notably different from more conventional settings, necessitating the development of new econometric tools properly tailored to the data and the questions of interest. To streamline the discussion, we focus on the practical implementation and heuristics of the underlying econometric theory, deferring the technical details to Appendix A.1.

Following the discussion and the theoretical implications developed in Section 2.2.2, our primary interest centers on estimation and inference concerning the volume-volatility elasticity, and in particular, whether the elasticity decreases with the level of disagreement. To this end, we estimate the elasticity between announcement-induced “abnormal” volume and volatility variations by regressing jumps in the log-volume intensity on jumps in the log-volatility. Further, to help address how the elasticity is affected by disagreement, we parameterize the regression coefficient as a function of various measures of disagreement prevailing at the time of the announcement (e.g., the dispersion among professional forecasters). The resulting econometric model may be succinctly expressed as

$$\Delta \log (m_r) = (a_0 + b_0^T X_0,\tau) + (a_1 + b_1^T X_1,\tau) \cdot \Delta \log (\sigma_r), \quad (2.7)$$

$^{19}$ Aït-Sahalia and Jacod (2014) provide a comprehensive review of recent development on the econometrics of high-frequency data.
where $X_{\tau} \equiv (X_{0,\tau}, X_{1,\tau})$ is comprised of the different explanatory/control variables employed in the estimation. In particular, by including measures of disagreement in $X_{1,\tau}$ it is possible to directly assess the aforementioned theoretical predictions based on the statistical significance of the estimated $b_1$ coefficients.

From an econometric perspective, equation (3.11) is best understood as an instantaneous moment condition, in which the volume intensity process $m_t$ (resp. the spot volatility process $\sigma_t$) represents the latent local first (resp. second) moment of the volume (resp. price return) process. The sample analogue of (3.11) therefore takes the form

$$\Delta \log(m_{\tau}) = (a_0 + b_0^T X_{0,\tau}) + (a_1 + b_1^T X_{1,\tau}) \cdot \Delta \log(\sigma_{\tau}) + e_{\tau},$$

(2.8)

where the error term $e_{\tau}$ arises from the estimation errors associated with the local moments (i.e., $m$ and $\sigma$). Our goal is to conduct valid inference about the parameter vector $\theta \equiv (a_0, b_0, a_1, b_1)$, especially the components $a_1$ and $b_1$ that determine the volume-volatility elasticity.

---

20 This setting resembles the fuzzy RDD. To illustrate, consider the special case of (3.11) where $a_0$, $b_0$ and $b_1$ equal zero. In this situation, (3.11) becomes $\Delta \log(m_{\tau}) = a_1 \Delta \log(\sigma_{\tau})$, so $a_1$ is exactly equal to the ratio between the jump of $\log(m_{\tau})$ and that of $\log(\sigma_{\tau})$, where $m$ and $\sigma$ are local moments of the volume and return processes, respectively. Analogously, in the fuzzy RDD setting, the treatment effect is identified exactly as the ratio between the jump in the conditional mean of the outcome variable and that in the treatment propensity; see, e.g., Lee and Lemieux (2010). In the more general setting with multiple discontinuities (induced by announcements), we aggregate the information among the announcements assuming that heterogeneous effects can be controlled by the observed covariates in $X_{\tau}$. 
Consider the group $\mathcal{A}$ comprised of a total of $M$ announcement times.\textsuperscript{21} Further, define $S_\tau \equiv (m_{\tau-}, m_\tau, \sigma_{\tau-}, \sigma_\tau, X_\tau)$ and $S \equiv (S_\tau)_{\tau \in \mathcal{A}}$, where the latter collects the information on all announcements. Our estimator of $S$ may then be expressed as $\hat{S}_n \equiv (\hat{S}_\tau)_{\tau \in \mathcal{A}}$, where $\hat{S}_\tau \equiv (\hat{m}_{\tau-}, \hat{m}_\tau, \hat{\sigma}_{\tau-}, \hat{\sigma}_\tau, X_\tau)$ is formed using the nonparametric pre- and post-event volume intensity and volatility estimators previously defined in (3.9). Correspondingly, summary statistics pertaining to the jumps in the volume intensity and volatility for the group of announcement times $\mathcal{A}$ may be succinctly expressed as $f(\hat{S})$ for some smooth function $f(\cdot)$.\textsuperscript{22}

Moreover, we may estimate the parameter vector $\theta \equiv (a_0, b_0, a_1, b_1)$ in (3.11) for the group $\mathcal{A}$ using the following least-square estimator

$$
\hat{\theta} \equiv \arg\min_{\theta} \sum_{\tau \in \mathcal{A}} \left( \Delta \log(m_\tau) - (a_0 + b_0^T X_{0,\tau}) - (a_1 + b_1^T X_{1,\tau}) \cdot \Delta \log(\sigma_\tau) \right)^2. \tag{2.9}
$$

This estimator may similarly be expressed as $\hat{\theta} = f(\hat{S})$, albeit for a more complicated transform $f(\cdot)$. It can be shown that $\hat{S}$ is a consistent estimator of $S$, which in turn implies that $f(\hat{S})$ consistently estimates $f(S)$, provided $f(\cdot)$ is a smooth function of

\textsuperscript{21} Importantly, unlike conventional econometric settings, our asymptotic inference does not rely on an increasingly large number of announcements. Indeed, we assume that the sample span and the number of announcements within the span (i.e., $M$) are fixed. Our econometric theory exploits the fact that the high-frequency data are sampled at (asymptotically increasingly) short intervals. Our econometric setting allows for essentially arbitrary heterogeneity across the announcements and empirically realistic strong persistence in the volume intensity and volatility processes.

\textsuperscript{22} For example, the average jump sizes in the logarithmic volatility and volume intensity around the announcements are naturally measured by

$$
f_1(\hat{S}) = \frac{1}{M} \sum_{\tau \in \mathcal{A}} \Delta \log(\sigma_\tau) \quad \text{and} \quad f_2(\hat{S}) = \frac{1}{M} \sum_{\tau \in \mathcal{A}} \Delta \log(m_\tau),
$$

respectively.
the estimated quantities. The estimates that we reported in our preliminary analysis in Section 2.4 may be formally justified this way.\footnote{This econometric framework also readily accommodates more general functional forms than the log-linear specification in (2.9), corresponding to more complicated transforms $f(\cdot)$. That said, we focus our empirical investigations on the log-linear specification, as the scatter plot in Figure 2.7 does not suggest any obvious non-linear dependencies between the volume intensity and volatility jumps (in logs).}

The “raw” estimator defined above is asymptotically valid under general regularity conditions (again, we refer to Appendix A.1 for the specific details). However, the nonparametric estimators $\hat{\Delta \log(\sigma_\tau)}$ and $\hat{\Delta \log(m_\tau)}$ underlying the simple estimator in (2.9) do not take into account the strong intraday U-shaped patterns in trading volume and volatility documented in Figure 2.3. While the influence of the intraday patterns vanishes asymptotically, they invariably contaminate our estimates of the jumps in finite samples, and thus affect our use of the jump estimates as measures of “abnormal” volume and volatility movements, to which the economic theory speaks. A failure to adjust for this may therefore result in a mismatch between the empirical strategy and the economic theory.\footnote{For instance, the ISM indices and the Consumer Confidence Index are all released at 10:00 when volume and volatility both tend to be decreasing even on non-announcement days, while FOMC announcements mostly occur at 14:15 when volume and volatility are generally increasing (though to a much smaller extent).}

To remedy this, we correct for the influence of the intraday pattern by differencing it out with respect to a control group. Since this differencing step is applied to the jumps, which are themselves differences between the post- and pre-event quantities, our empirical strategy for jump estimation may be naturally thought of as a high-frequency DID type approach, in which we consider the event-control difference of the jump estimates as our measure for the abnormal movements in the volume intensity.
and volatility. We then regress these DID jump estimates to obtain the volume-volatility elasticity. Following the terminology of Li et al. (2017), who considered “jump regressions” involving discontinuous price increments, we refer to our new procedure as a DID jump regression estimator.\(^{25}\)

Formally, with each announcement time \(\tau\), we associate a control group \(C(\tau)\) of non-announcement times. Based on this control group, we then correct for the intraday patterns in the “raw” jump estimators by differencing out the corresponding estimates averaged within the control group, resulting in the adjusted jump estimators

\[
\begin{align*}
\overline{\Delta \log(m_\tau)} &\equiv \Delta \overline{\log(m_\tau)} - \frac{1}{N_C} \sum_{\tau' \in C(\tau)} \Delta \overline{\log(m_{\tau'})}, \\
\overline{\Delta \log(\sigma_\tau)} &\equiv \Delta \overline{\log(\sigma_\tau)} - \frac{1}{N_C} \sum_{\tau' \in C(\tau)} \Delta \overline{\log(\sigma_{\tau'})},
\end{align*}
\]

(2.10)

where \(N_C\) refers to the number of times in the control group.\(^{26}\) Analogously to (2.9), we then estimate the parameters of interest by regressing the DID jump estimates as

\[
\hat{\theta} \equiv \arg\min_{\theta} \sum_{\tau \in A} \left( \Delta \overline{\log(m_\tau)} - (a_0 + b_0^T X_{0,\tau}) - (a_1 + b_1^T X_{1,\tau}) \cdot \Delta \overline{\log(\sigma_\tau)} \right)^2.
\]

(2.11)

Note that \(\hat{\theta}\) depends not only on \((\hat{S}_\tau)_{\tau \in A}\) but also on \((\hat{S}_\tau')_{\tau \in C}\), where \(C \equiv \cup_{\tau \in A} C(\tau)\) contains the times of all control groups. This estimator can be expressed as \(\hat{\theta} = f(\tilde{S})\) where \(\tilde{S} \equiv (\hat{S}_\tau)_{\tau \in \mathcal{T}}\) for \(\mathcal{T} \equiv A \cup C\).

\(^{25}\) Unlike the price, the volume intensity and volatility are both latent processes and, as a result, our new estimator converges at a slower nonparametric rate, compared to the “jump regressions” studied by Li et al. (2017).

\(^{26}\) In our empirical analysis below, \(C(\tau)\) consists of the same time-of-day as \(\tau\) over the previous \(N_C = 22\) non-announcement days (roughly corresponding to the length of one trading month). We also experimented with the use of other control periods, including periods comprised of future non-event days, resulting in the same general conclusions as the DID results reported below.
In summary, our estimation procedure consists of two steps. The first step is to estimate the jumps in the volume intensity and volatility processes via DID. The second step consists in estimating the parameters that describe the relationship between the two via a least-square regression. Since jumps are discontinuities, the resulting DID jump regression estimator formally bears some resemblance to that in RDD commonly used in empirical microeconomics for the estimation of treatment effects (see, e.g., Lee and Lemieux, 2010). Importantly, however, our new econometric inference procedures (including the computation of standard errors) is very different and non-standard. Explicitly allowing for heterogeneity and dependence in the return and volume data, the sampling variability in \( \tilde{\theta} \) arises exclusively from the nonparametric estimation errors in the pre- and post-event high-frequency-based volume intensity and volatility estimators, \( \hat{m}_{\tau\pm} \) and \( \hat{\sigma}_{\pm} \), respectively. While in theory we can characterize the resulting asymptotic covariance matrix and it would be possible to use it in the design of “plug-in” type standard errors, the control groups \( C(\tau) \) used for the different announcement times often partially overlap, which would severely complicate the formal derivation and implementation of the requisite formulas.

In order to facilitate the practical implementation, we instead propose a novel easy-to-implement local i.i.d. bootstrap procedure for computing the standard errors, which is a localized version of the i.i.d. bootstrap of Gonçalves and Meddahi (2009). This procedure does not require the exact dependence of \( \tilde{\theta} \) on \( \tilde{S} \) to be fully specified. Instead, it merely requires repeated estimation over a large number of locally i.i.d. bootstrap samples for the pre-event and post-event windows around each of the announcement and control times. The “localization” is important, as it allows us to treat the conditional distributions as (nearly) constant, in turn permitting
the use of an \textit{i.i.d.} re-sampling scheme. The actual procedure is summarized by
the following algorithm, for which the formal theoretical justification is given in the
technical Appendix A.1.

\textbf{Bootstrap Algorithm}

Step 1: For each \(\tau \in \mathcal{T}\), generate \textit{i.i.d.} draws \((V^*_{i(\tau)-j}, r^*_{i(\tau)-j})_{1 \leq j \leq k_n}\) and
\((V^*_{i(\tau)+j}, r^*_{i(\tau)+j})_{1 \leq j \leq k_n}\) from \((V_{i(\tau)-j}, r_{i(\tau)-j})_{1 \leq j \leq k_n}\) and \((V_{i(\tau)+j}, r_{i(\tau)+j})_{1 \leq j \leq k_n}\), respectively.

Step 2: Compute \(\Delta \log (m_{\tau})^*\) and \(\Delta \log (\sigma_{\tau})^*\) the same way as \(\Delta \log (m_{\tau})\) and
\(\Delta \log (\sigma_{\tau})\), respectively, except that the original data \((V_{i(\tau)-j}, r_{i(\tau)-j})_{1 \leq |j| \leq k_n}\) is re-
placed with \((V^*_{i(\tau)-j}, r^*_{i(\tau)-j})_{1 \leq |j| \leq k_n}\). Similarly, compute \(\tilde{\theta}^*\) according to (3.14) using
re-sampled data.

Step 3: Repeat steps 1 and 2 a large number of times. Report the empirical
standard errors of (the components of) \(\tilde{\theta}^* - \tilde{\theta}\) as the standard errors of the original
estimator \(\tilde{\theta}\).

\(\square\)

It is important to stress that even though the re-sampling scheme is locally \textit{i.i.d.},
the validity of this bootstrap procedure does \textit{not} require the volume and return data to
be actually \textit{i.i.d.} Rather, the procedure allows for very general dynamic dependencies,
and potentially highly persistent and non-stationary volatility and volume intensity
processes.

Equipped with the new high-frequency DID jump regression estimator defined
in equation (3.14) and the accompanying bootstrap procedure outlined above for
calculating standard errors and conducting valid inference, we now turn to our main
empirical findings.
2.6 Volume-volatility relationship around public announcements

We begin our empirical investigations by verifying the occurrence of (on average) positive jumps in both trading volume intensity and return volatility around scheduled macroeconomic announcements. We document how these jumps, and the volume-volatility elasticity in particular, vary across different types of announcements. We then show how the variation in the elasticities observed across different announcements may be related to explanatory variables that serve as proxies for differences-of-opinion and, relatedly, notion of economic uncertainty. A more detailed analysis of FOMC announcements further highlights the important role played by the sentiment embedded in the FOMC statements accompanying each of the rate decisions. In addition to our main results based on the S&P 500 aggregate equity portfolio, we also report complementary results based on U.S. Treasury bond futures.

2.6.1 Jumps and announcements

Consistent with the basic tenet of information-based trading around public news announcements, the preliminary analysis underlying Figure 2.5 clearly suggests an increase in both volatility and trading intensity from thirty minutes before an FOMC announcement to the thirty minutes after the announcement. In order to more formally corroborate these empirical observations and extend them to a broader set of announcements, we report in Table 2.3 the average magnitudes of the logarithmic volatility and volume intensity jumps observed around our full combined set of news announcements, as well as the five specific news categories explicitly singled out in Table 3.1. In order to highlight the importance of properly correcting for the intra-
Table 2.3: Volume and volatility jumps around public news announcements

<table>
<thead>
<tr>
<th>Events</th>
<th>All</th>
<th>FOMC</th>
<th>ISMM</th>
<th>ISMNM</th>
<th>CC</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.Obs.</td>
<td>2130</td>
<td>109</td>
<td>160</td>
<td>158</td>
<td>160</td>
<td>1682</td>
</tr>
<tr>
<td>Log Volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No DID</td>
<td>0.090**</td>
<td>1.088**</td>
<td>0.162**</td>
<td>0.072**</td>
<td>0.114**</td>
<td>0.023**</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.027)</td>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.015)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log Volume</td>
<td>0.034**</td>
<td>1.410**</td>
<td>0.056**</td>
<td>-0.045**</td>
<td>0.049**</td>
<td>-0.045**</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.023)</td>
<td>(0.015)</td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log Volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DID</td>
<td>0.152**</td>
<td>1.037**</td>
<td>0.256**</td>
<td>0.165**</td>
<td>0.209**</td>
<td>0.087**</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.029)</td>
<td>(0.017)</td>
<td>(0.020)</td>
<td>(0.017)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log Volume</td>
<td>0.204**</td>
<td>1.329**</td>
<td>0.335**</td>
<td>0.233**</td>
<td>0.328**</td>
<td>0.118**</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.025)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

Notes: The table reports the average logarithmic volatility jumps and the average logarithmic volume intensity jumps around all announcements (All), which are further categorized into FOMC announcements (FOMC), ISM Manufacturing Index (ISMM), ISM Non-Manufacturing Index (ISMNM), Consumer Confidence Index (CC), and other pre-scheduled macroeconomic announcements (Others). The top panel reports the raw statistics. The bottom panel adjusts for the intraday pattern via the DID method using the past 22 non-announcement days as the control group. The sample period spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. ** indicates significance at the 1% level.

day patterns seen in Figure 2.3, we report and contrast both the “raw” \( \Delta \log (m_r) \) and \( \Delta \log (\sigma_r) \) jump estimates obtained without the DID correction and the DID corrected estimates formally based on equation (3.12), together with the new bootstrapped standard errors described in Section 2.5.

The top panel presents the “raw” jump statistics without the DID correction. The volatility jumps are always estimated to be positive and highly statistically significant. This is true for all of the announcements combined, as well as within each of the five separate categories. The volume jumps averaged across all news announcements are also significantly positive. However, the jumps in the volume intensities are estimated to be negative for two of the news categories: ISM Non-Manufacturing and Others. This is difficult to reconcile with any of the economic mechanisms and theoretical models discussed in Section 2.2. Instead, these negative estimates may be directly
attributed to the strong diurnal pattern evident in Figure 2.3. The ISM indices and most of the economic news included in the Others category are announced at 10:00am, when both volatility and trading volume tend to be falling, thus inducing a downward bias in the jump estimation.

To remedy this, the bottom panel of Table 2.3 reports the corresponding DID estimates based on equation (3.12) in which we rely on the previous 22 non-announcement days as the control group. As these estimates show, applying the DID correction results in significantly positive jumps for the spot volatility and trading intensity across all of the different news categories, ISMNM and Others included. This contrast directly underscores the importance of properly controlling for the intradaily features outside the stylized theoretical models when studying volume and volatility at the high-frequency intraday level. At the same time, the magnitude of the jump estimates associated with FOMC announcements, which mostly occur between 14:00 and 14:15 when volatility and trading volume both tend to be rising, is actually reduced by the DID correction. Nevertheless, FOMC clearly stands out among all of the different news categories, as having the largest (by a wide margin) average jump sizes in both volume and volatility.

Having documented the existence of highly significant positive jumps in both volume and volatility around public announcements, we next turn to the joint relationship between these jumps, focusing on the volume-volatility elasticity and the implications of the theoretical models discussed in Section 2.2.
2.6.2 Volume-volatility elasticities around public news announcements

The theoretical models that guide our empirical investigations are explicitly designed to highlight how trading volume and return volatility respond to well-defined public news announcements. As such, the models are inevitably stylized, with other influences (such as those underlying the intraday patterns and long-term trends evident in Figures 2.3 and 2.5, respectively) deliberately abstracted away. As discussed above, the DID method provides a way to guard against the non-trivial influence of the systematic intraday patterns. It also conveniently differences out other unmodeled nuisances, like trends, which would otherwise contaminate the estimates. Consequently, we rely on the DID jump regression approach throughout.

To begin, consider a basic specification of equation (3.14) without any explanatory variables (i.e., $X_{0,\tau}$ and $X_{1,\tau}$ are both absent). Table 2.4 reports the resulting estimates for each of the different news categories. All of the estimated intercepts (i.e., $a_0$) are positive and highly statistically significant, indicative of higher trading intensities following public news announcements, even in the absence of heightened return

<table>
<thead>
<tr>
<th></th>
<th>FOMC</th>
<th>ISMM</th>
<th>ISMNM</th>
<th>CC</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>0.586**</td>
<td>0.199**</td>
<td>0.119**</td>
<td>0.218**</td>
<td>0.050**</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.022)</td>
<td>(0.020)</td>
<td>(0.023)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.716**</td>
<td>0.529**</td>
<td>0.688**</td>
<td>0.521**</td>
<td>0.787**</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.064)</td>
<td>(0.069)</td>
<td>(0.075)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.330</td>
<td>0.155</td>
<td>0.220</td>
<td>0.109</td>
<td>0.287</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log \sigma - a_0 + a_1 \cdot \Delta \log \sigma$, using the past 22 non-announcement days as the control group. The sample spans the period from April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. ** indicates significance at the 1% level.
volatility.\textsuperscript{27} Put differently, abnormal bursts in trading volume around announcements are not always associated with abnormal price changes. This, of course, is directly in line with the key idea underlying the KP model that differences-of-opinion provides an additional trading motive over explicit shifts in investors’ average opinion. Further corroborating the rank of FOMC as the most important news category released during regular trading hours, the estimated intercept is the largest for FOMC announcements.\textsuperscript{28}

Turning to the volume-volatility elasticities (i.e., $a_1$), all of the estimates are below unity, and significantly so.\textsuperscript{29} The theoretical derivations in (2.4) and (2.5) based on the KP model also predict that in the presence of differences-of-opinion the volume-volatility elasticity should be below unity. Our empirical findings are therefore directly in line with this theoretical prediction, and further support the idea that disagreements among investors often provide an important motive for trading.

\textsuperscript{27} By contrast, the estimates obtained for $a_0$ without the DID correction, reported in the supplementary appendix, are significantly negative for ISMM, ISMN and the Others news categories, underscoring the importance of properly controlling for the strong intraday patterns in the volume intensity and volatility.

\textsuperscript{28} In our reduced-form specification (3.11), the intercept term captures the abnormal trading volume in the “limiting” case in which the price change is zero. In the KP model this corresponds to the $\beta_0$ term in equation (2.1), which from equation (2.2) is directly related to the level of disagreement, but is also confounded by the degree of risk tolerance. In addition to these influences, the large estimated intercepts for the FOMC announcements may simply reflect the fact that the FOMC announcements are more closely watched by a larger pool of investors than the other announcements. By contrast, the volume-volatility elasticity and the slope coefficient being a relative measure, as formalized within the context of the KP model in equations (2.4) and (2.5), is free from these confounding influences. We therefore focus our attention on the easier to interpret elasticity estimates.

\textsuperscript{29} The robust DID estimate for the elasticity around FOMC announcements reported in Table 2.4 is slightly larger than the preliminary raw non-DID estimate discussed in Section 2.4.
2.6.3 Volume-volatility elasticities and disagreement measures

In addition to the prediction that the volume-volatility elasticity should be less than unity, our theoretical derivations in Section 2.2 also predict that the elasticity should be decreasing with the overall level of disagreement among investors. In order to examine this more refined theoretical prediction, we include a set of additional explanatory variables (in the form of the $X_{1,\tau}$ variable in the specification in equation (3.14)) that serve as proxies for disagreement. To account for the category-specific heterogeneity in the volume-volatility elasticity estimates reported in Table 2.4, we also include a full set of category dummy variables (i.e., one for each of the FOMC, ISMM, ISMNM and CC news categories).

We consider two proxies for the overall level of investors’ disagreement that prevails at the time of the announcement. The first is the forecast dispersion of the one-quarter-ahead unemployment rate from the Survey of Professional Forecasters (SPF).\footnote{The SPF is a quarterly survey. It is released and collected in the second month of each quarter. To prevent any look-ahead bias (which may arise from using future information in the estimation of the current elasticity), we use the value from the previous quarter. Additional results for other forecast horizons and dispersion measures pertaining to other economic variables are reported in Appendix ??.
} For aggregated macroeconomic time series, like the unemployment rate, the dispersion among forecasters arises more likely from differences in their economic/statistical models, rather than their “private” information regarding the aggregate macroeconomy. Hence, the SPF dispersion measure speaks directly to the notion of disagreement in the Kandel–Pearson model, where investors’ disagreement arises from differences in their models used for interpreting the public signal. This same measure has also previously been used in other studies to gauge the degree of dis-
agreement; see, e.g., Van Nieuwerburgh and Veldkamp (2006) and Ilut and Schneider (2014) among others.

Secondly, as an indirect proxy for differences-of-opinion, we employ a weekly moving average of the economic policy uncertainty index developed by Baker et al. (2015).[^31] There is a voluminous literature that addresses the relation between disagreement and uncertainty, generally supporting the notion of a positive relation between the two; see, for example, Acemoglu et al. (2006) and Patton and Timmermann (2010). Below, we refer to these two proxies as Dispersion and Weekly Policy, respectively. To facilitate comparisons, we scale both measures with their own sample standard deviations.

The estimation results for different specifications including these additional explanatory variables in the volume-volatility elasticity are reported in Table 2.5.[^32] As a reference, the first column reports the results from a basic specification without any explanatory variables. The common elasticity is estimated to be 0.733 which, not surprisingly, is close to the average value of the category-specific estimates reported in Table 2.4. Underscoring the importance of disagreement more generally, the estimate is also significantly below one.

[^31]: The economic policy uncertainty index of Baker et al. (2015) is based on newspaper coverage frequency. We use the weekly moving average so as to reduce the noise in the daily index. The averaging also naturally addresses the weekly cycle in the media. Comparable results based on the monthly index are available in Appendix ??.

[^32]: We also include news-category dummies in the intercept $X_{0,t}$ in all of the different specifications, so as to control for the heterogeneity in the $a_0$ estimates in Table 2.4. Since our main focus centers on the volume-volatility elasticity, to conserve space we do not report these estimated $b_0$ dummy coefficients.
The specification in the second column includes the full set of news category dummies in the elasticity, with the baseline category being Others.\textsuperscript{33} The elasticity for the Others category, which includes by far the largest number of announcements, is estimated to be 0.776 and close to the value of 0.733 from the specification without any dummies. The estimates for FOMC and ISM Non-Manufacturing announcements are also both statistically indistinguishable from this value of 0.776. On the other hand, the volume-volatility elasticities estimated around ISM Manufacturing and Consumer Confidence announcements are both significantly lower, indicating that the levels of disagreement among investors are higher for these events. To help understand this latter finding, we note that ISM Manufacturing and Consumer Confidence announcements are both released early in the macroeconomic news cycle, as described in Andersen et al. (2003). As documented in Andersen et al. (2003, 2007), the first announcements in a given news cycle tend to have larger price impacts than later related news announcements, as much of the information contained in the later releases may have already been gleaned from the earlier news announcements. Our findings are in line with this logic. The estimated elasticities indicate relatively high levels of disagreement around the two early ISMM and CC announcements, and indirectly suggest that the release of these help resolve some of the economic uncertainty and decrease the overall level of disagreement, as manifest in the closer-to-one elasticity estimate for the later (in the news cycle) ISMNM releases.

\textsuperscript{33} Although the full set of news-category dummy variables are included in both the intercept and the elasticity specifications, the estimates in the second column in Table 2.5 are not exactly identical to those in Table 2.4, because some of the announcements across the different news categories occur concurrently.
The next two columns in the table report the results for specifications that include either Dispersion or Weekly Policy as an additional explanatory variable. Consistent with our theoretical prediction, both of these disagreement proxies significantly negatively impact the volume-volatility elasticity, and the estimated effects are also economically nontrivial. Moreover, the estimated elasticities obtained in the absence of any dispersion or economic policy uncertainty (i.e., the \( a_1 \) coefficients) are much closer to the no-disagreement benchmark of unity implied by the theoretical expressions in (2.4) and (2.5), than the corresponding estimates obtained without controlling for disagreement.

The last column shows that both of the disagreement measures remain statistically significant when included jointly, although less so for Dispersion. Interestingly, the baseline elasticity of 0.984 for Others (i.e., \( a_1 \)) is no longer statistically different from unity, nor are the estimated elasticities for FOMC and ISM Non-Manufacturing.

Taken as a whole, these results strongly corroborate the existence of a negative relationship between the volume-volatility elasticity and the level of disagreement among investors. The results also suggest that for a majority of the public news announcements, our two specific disagreement proxies, involving measures of forecast dispersion and economic policy uncertainty, are able to explain the deviation in the volume-volatility elasticity from the no-disagreement benchmark of unity.

2.6.4 Further analysis of FOMC announcements

The results discussed in the previous section were based on the joint estimation involving all of the macroeconomic news announcements that occur during regular trading hours. Meanwhile, as documented in Table 2.3, the FOMC rate decisions
rank supreme in inducing the on-average largest jumps in both trading activity and return volatility over our sample. These large responses occur in spite of the fact that the federal funds rate was fixed at the effective zero lower bound over much of the later half of the sample. Moreover, economists also routinely disagree about the interpretation of monetary policy. All of these unique features grant FOMC announcements of particular interest for our analysis pertaining to the role of disagreement in financial markets.

Before discussing the results from our more detailed analysis of the volume-volatility elasticity estimated exclusively around the times of FOMC announcements, it is important to stress some crucial differences between our analysis and prior empirical work related to FOMC announcements. In particular, there is already an extensive literature devoted to the study of the impact of FOMC announcements on equity returns. In addition to the earlier influential work by Bernanke and Kuttner (2005), this includes more recent studies specifically related to the behavior of monetary policy and market reactions when the rate is at or near the zero lower bound (e.g., Bernanke, 2012; Wright, 2012; van Dijk et al., 2014; Johnson and Paye, 2015). Other recent studies have also documented that most of the equity risk premium is earned in specific phases of the FOMC news release cycle (e.g., Savor and Wilson, 2014; Lucca and Moench, 2015; Cieslak et al., 2015). It is not our intent to add to this burgeoning literature on the determinant of the equity risk premium, and the functioning of monetary policy per se. Instead, we simply recognize the unique position of FOMC announcements as the most important news category in our sample. Motivated by this fact, we further investigate how the variation in the volume-volatility
elasticity more generally varies with measures of disagreement prevailing at the exact time of and directly extracted from the FOMC news releases.

For ease of reference, the first column in Table 2.6 reports the DID estimation results for the FOMC subsample and the benchmark specification that does not include any explanatory variables in the elasticity, as previously reported in Table 2.4. The second column includes the previously defined Dispersion measure as an explanatory variable. In parallel to the full-sample results in Table 2.5, the estimates show that higher levels of forecast dispersions are generally associated with lower volume-volatility elasticities. The estimate of -0.108 for the $b_1$ coefficient is also highly statistically significant. Moreover, after controlling for Dispersion, the baseline elasticity (i.e., $a_1$) is virtually one. This finding thus suggests that Dispersion alone, as a measure of differences-of-opinion, is able to successfully explain much of the deviation from unity in the volume-volatility elasticity at the times of FOMC announcements.

The $b_1$ estimate for the Weekly Policy variable reported in the third column of Table 2.6 is also negative. However, it is not significant at conventional levels. This lack of significance of the Weekly Policy variable in the FOMC subsample, stands in sharp contrast with its highly significant effect in Table 2.5 based on the full sample of all announcements. This therefore suggests that the Weekly Policy variable, which is constructed as a “catch-all” measure of economic uncertainty, is simply too diverse (or noisy) to satisfactorily explain the variation in the volume-volatility relationship observed exclusively around FOMC announcements.

To remedy this, we construct an alternative textual measure based on the actual FOMC press releases. The FOMC statements, in addition to announcing the new
target rates, also signal the future path of policy. In recent years, the statements also include brief summaries of the state of the economy, providing additional context underlying the rate decisions.\textsuperscript{34} We succinctly summarize this additional information by counting the number of negative words, in accordance with the financial-negative (Fin-Neg) word list compiled by Loughran and McDonald (2011). We refer to this textual measure as the FOMC Sentiment. A more detailed description of the construction is provided in Appendix A.2.2.\textsuperscript{35} Assuming that the use of more negative words provides additional room for investors to differ in their interpretation of the news, we consider this alternative FOMC Sentiment measure as a more direct proxy for the level of disagreement at the exact times of the FOMC announcements.

From the theoretical relations derived in Section 2.2, we would therefore expect to see lower volume-volatility elasticities in response to higher FOMC Sentiment measures. The estimation results reported in the fourth column of Table 2.6 supports this theoretical prediction. The estimated $b_1$ coefficient for our FOMC Sentiment measure equals -0.104. It is also highly statistically significant. Moreover, controlling for the

\textsuperscript{34} This is especially important over the later half of our sample period, when the target rate was consistently stuck at the zero lower bound and, hence, offered little new information by itself. Bernanke (2012) also explicitly emphasized the important role of “public communications” as a nontraditional policy tool of the Fed.

\textsuperscript{35} Loughran and McDonald (2011) originally constructed their Fin-Neg list for the purpose of analyzing corporate 10-K reports. Compared with negative words, positive words tend to be less informative due to their more frequent negation. Along those lines, some Fin-Neg words detected in the FOMC statement may not actually have a negative meaning. One example is the word “late” in the context “… are likely to warrant exceptionally low levels for the federal funds rate at least through late 2014.” Another example is “unemployment,” which routinely appears in the first two paragraphs of the statements over the later half of our sample. To examine the severity of this issue, we manually checked every word that was classified as negative using Loughran and McDonald’s list, and then refined this selection by only keeping words with an unambiguous negative meaning within their context. The regression results based on this refined measure are very similar to those based on the FOMC Sentiment measure reported here, and hence are omitted for brevity.
FOMC Sentiment, the baseline elasticity (i.e., $a_1$) is estimated to be 0.914, and this estimate is statistically indistinguishable from unity at conventional significance levels. Interestingly, the $b_1$ estimate of -0.104 for the FOMC Sentiment variable is also very close to the -0.108 estimate for the Dispersion measure reported in the second column.

In order to further gauge the relative merits of the Dispersion and FOMC Sentiment measures, the specification reported in the fifth column includes both as explanatory variables in the elasticity. Both of the estimated coefficients are negative and statistically significant. The coefficient estimates are also similar in magnitude, suggesting that the Dispersion and FOMC Sentiment measures are equally important in terms of capturing the disagreements-in-beliefs that motivate the abnormal trading at the times of FOMC announcements. Again, the estimate for the baseline elasticity $a_1$ is also not statistically different from the theoretical prediction of unity that should obtain in the absence of differences-of-opinion.

Further augmenting the specification to include the Weekly Policy measure as an additional explanatory variable in the elasticity does not change the key aspects of any of these findings, as shown by the results reported in the last column of the table. Counter to the previous empirical results and theoretical predictions, the estimated coefficient for the Weekly Policy variable in this expanded DID jump regression is actually positive, albeit not significant at conventional levels. This also indirectly supports our earlier conjecture that the textual-based FOMC Sentiment measure affords a much more pointed and accurate characterization of the economic uncertainty and differences-in-opinion at the exact times of the FOMC announcements, compared to the “catch-all” Weekly Policy measure.
2.6.5 *Bond futures and FOMC announcements*

Our empirical analysis so far has been focused exclusively on the aggregate stock market, as proxied by the S&P 500.\(^{36}\) The theoretical implications developed in Section 2.2, however, should hold true more broadly for other markets. Hence, to further buttress the theoretical implications, we briefly analyze the behavior of the volume-volatility elasticity in the fixed income market.\(^{37}\) We rely on high-frequency price and volume data for the ten-year U.S. Treasury bond futures sampled at the one-minute frequency. As discussed in Section 2.3, our sample for the T-bond futures spans the slightly shorter period from July 1, 2003 to September 30, 2014.

In parallel to the preliminary analysis for the stock market in Section 2.4 and the more detailed analysis described in the previous subsection, we focus our analysis on the volume-volatility elasticity estimated around the FOMC announcements.\(^{38}\) Table 2.7 reports the resulting estimates. The patterns are generally consistent with the comparable findings for the stock market reported in Table 2.6. Specifically, looking at the first column without any controls, the elasticity is estimated to be 0.531, and significantly lower than unity. Moreover, the coefficients associated with the various disagreement proxies included in columns two through five are all estimated to be

\(^{36}\) Additional robustness checks based on an ETF tracking the Dow Jones Industrial Average are provided in Appendix ??.

\(^{37}\) As previously noted, prior work, notably Green (2004) and Pasquariello and Vega (2007), have analyzed the price impact of order flow in the bond market around the time of macroeconomic news announcements. Our analysis specifically pertaining to the volume-volatility elasticity complements this work by allowing for a more direct interpretation of the estimates vis-a-vis the KP and differences-of-opinion class of models.

\(^{38}\) The graph for the log-volume intensity and log-spot volatility in the T-bond market immediately before and after all of the FOMC announcements included in Appendix ?? closely mirrors the corresponding Figure 2.5 for the aggregate stock market, clearly suggesting the presence of jumps in both series at the FOMC news announcement times.
negative (i.e., $b_1 < 0$), and controlling for these proxies pushes the baseline elasticity (i.e., $a_1$) closer towards one. In the most general specification reported in the last column, the Dispersion and FOMC Sentiment measures both exert a statistically significant influence on the elasticity, while the Weekly Policy measure is rendered insignificant by the inclusion of the former two measures. This compares directly to the results for the stock market in Table 2.6. Meanwhile, the baseline elasticity estimates for the bond market are slightly lower than for the stock market, suggesting that bond traders tend to disagree more (in a relative sense as formalized by equation (2.5) for the KP model) about the interpretation of the monetary policy compared to stock market traders.

All in all, however, our findings for the bond futures data are very similar to those for the aggregate stock market portfolio, supporting the theoretical implications of the KP model and the importance of differences-of-opinion more generally.

2.7 Conclusion

We provide new empirical evidence concerning the behavior of financial market volatility and trading activity in response to public news announcements. Our results are based on intraday prices and trading volume for the aggregate stock market portfolio and Treasury bond futures, along with new econometric procedures specifically designed to deal with the unique complications that arise in the high-frequency data setting. Explicitly zooming in on the volume-volatility changes right around the exact times of the announcements allows us to cast new light on the way in which financial markets process new information and function more generally.
Consistent with the implications from theoretical models involving economic agents who agree-to-disagree, we find that the sensitivity of abnormal volume changes with respect to those of volatility estimated around the times of the most important public news announcements, as embedded within the volume-volatility elasticity, is systematically below unity. Further corroborating the important role played by differences-of-opinion among market participants, the elasticity tends to be low during times of high economic policy uncertainty and high dispersion among professional economic forecasters. A direct textual-based measure of the negative sentiment in the FOMC statements accompanying the actual rate decisions also negatively impacts the elasticity estimated at FOMC announcement times, lending additional empirical support to our key theoretical predictions.
Table 2.5: Volume-volatility elasticity estimates and disagreement measures

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td>Elasticity ($a_1$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.044**</td>
<td>0.733**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041**</td>
<td>0.776**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.040**</td>
<td>0.906**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.043)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041**</td>
<td>0.921**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.036)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041**</td>
<td>0.984**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.045)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News-category dummy variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC</td>
<td>-0.060</td>
<td>-0.058</td>
<td>-0.048</td>
<td>-0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.072)</td>
<td>(0.072)</td>
<td>(0.072)</td>
<td></td>
</tr>
<tr>
<td>ISMM</td>
<td>-0.238**</td>
<td>-0.238**</td>
<td>-0.220**</td>
<td>-0.222**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.070)</td>
<td>(0.070)</td>
<td>(0.070)</td>
<td></td>
</tr>
<tr>
<td>ISMNMM</td>
<td>-0.090</td>
<td>-0.092</td>
<td>-0.082</td>
<td>-0.085</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.077)</td>
<td>(0.077)</td>
<td>(0.077)</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>-0.244**</td>
<td>-0.244**</td>
<td>-0.202**</td>
<td>-0.207**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.073)</td>
<td>(0.073)</td>
<td>(0.073)</td>
<td></td>
</tr>
<tr>
<td>Disagreement measures:</td>
<td>Dispersion</td>
<td></td>
<td></td>
<td>-0.031*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.051**</td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Weekly Policy</td>
<td>-0.079**</td>
<td></td>
<td></td>
<td>-0.070**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.481</td>
<td>0.482</td>
<td>0.483</td>
<td>0.486</td>
<td>0.486</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log(p_m) = a_0 + b_0 X_{0,t} + (a_1 + b_1 X_{1,t}) \cdot \Delta \log(\sigma_t)$ based on all of the public announcements, using the past 22 non-announcement days as the control group. In all specifications, $X_{0,t}$ includes category dummy variables for FOMC rate decision (FOMC), ISM Manufacturing Index (ISMM), ISM Non-Manufacturing Index (ISMNM) and Consumer Confidence Index (CC); the estimates of these dummies (i.e., $b_0$) are not reported for brevity. The Dispersion variable is constructed as the latest forecast dispersion of the one-quarter-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. The Weekly Policy variable is constructed as the weekly moving average before the announcement of the economic policy uncertainty index developed by Baker et al. (2015). Both variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
Table 2.6: Volume-volatility elasticity estimates around FOMC announcements

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>0.586**</td>
<td>0.579**</td>
<td>0.583**</td>
<td>0.545**</td>
<td>0.550**</td>
</tr>
<tr>
<td>(0.078)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.716**</td>
<td>0.996**</td>
<td>0.790**</td>
<td>0.913**</td>
<td>1.089**</td>
</tr>
<tr>
<td>(0.067)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
<td>-0.108**</td>
<td>-0.088**</td>
<td>-0.096**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Policy</td>
<td>-0.037</td>
<td></td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment</td>
<td>-0.104**</td>
<td>-0.077**</td>
<td>-0.097**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| $R^2$                                                   | 0.330  | 0.382  | 0.329  | 0.370  | 0.400  |

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log(m_{\tau}) = a_0 + (a_1 + b_1^I X_{1,\tau}) \cdot \Delta \log(\sigma_{\tau})$ based on FOMC announcements, using the past 22 non-announcement days as the control group. The news-category dummies, i.e., $X_{0,\tau}$, are absent from this specification for the FOMC subsample. Dispersion and Weekly Policy are constructed as in Table 2.5. FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. These variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
### Table 2.7: Volume-volatility elasticity estimates around FOMC announcements for Treasury bond futures

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td>Elasticity ($a_1$)</td>
</tr>
<tr>
<td></td>
<td>1.245** 1.197** 1.227** 1.241** 1.203** 1.218**</td>
<td>0.531** 0.813** 0.627** 0.649** 0.832** 0.786**</td>
</tr>
<tr>
<td></td>
<td>(0.076) (0.075) (0.078) (0.076) (0.076) (0.077)</td>
<td>(0.066) (0.086) (0.089) (0.074) (0.087) (0.094)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.314 0.364 0.315 0.337 0.364 0.364</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The table reports the results for the ten-year Treasury bond futures data from the DID jump regression in equation (3.14) for the specification $\Delta \log (m_r) = a_0 + (a_1 + b_1^T X_{1,r}) \cdot \Delta \log (\sigma_r)$ based on FOMC announcements, using the past 22 non-announcement days as the control group. The news-category dummies, i.e., $X_{0,r}$, are absent from this specification for the FOMC subsample. Dispersion and Weekly Policy are constructed as in Table 2.5. FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. These variables are scaled by their own sample standard deviations. The sample spans July 1, 2003 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
Summary: This paper shows the effect of investor sentiment on information processing in financial markets. We investigate how disagreement among investors affects the relationship between trading intensity and price volatility around macroeconomic announcements during high- and low-sentiment periods. By incorporating into the Kandel and Pearson (1995) model a one-factor structure with heterogeneous beliefs in the idiosyncratic component, we explicitly derive the volume-volatility elasticity for individual stocks around systematic information release. Our empirical results are based on intra-day transaction data for the S&P 500 ETF and Dow Jones 30 components, analyzed with high-frequency econometric tools in a multi-dimensional setting. Consistent with the model predictions, measures of investor disagreement only have a significantly negative effect on the volume-volatility elasticity in high-sentiment periods, and this is true both for the market portfolio and for individual
stocks. This result is in line with changes in the confidence level of investors when sentiment regime shifts. Our estimates of elasticity also decrease significantly with the ratio of idiosyncratic variance, which indicates that higher idiosyncratic risks introduce larger dispersion among investors.

**Keywords:** Differences-of-opinion; investor sentiment; high-frequency data; trading volume; stochastic volatility; economic uncertainty.

**JEL classification:** C52, G12, G41.

### 3.1 Introduction

How information is processed in financial markets is of high interest. Co-movement of volume and volatility upon information arrival has been widely documented in the literature. Empirically, researchers find large trading volumes and return volatility around the time when news is released\(^1\). One theoretical model often used to characterize this processes is the “differences-of-opinion” model, in which investors agree to disagree on common public signals. (see, e.g., Harris and Raviv (1993), Kandel and Pearson (1995), Scheinkman and Xiong (2003) and Banerjee and Kremer (2010)). In Bollerslev et al. (2017), the authors show the volume and volatility elasticity derived from Kandel and Pearson (1995) significantly decreases with the level of disagreement. We propose econometric tools to estimate the volatility-volume elasticity for the S&P 500 ETF. The empirical results are consistent with the model predictions.

In this “differences-of-opinion” model, investors interpret signals and change positions based on Bayesian learning, and this process can be subject to investor sen-

---

\(^1\) See Chaboud et al. (2008), Boudt and Petitjean (2014), Jiang et al. (2011), and Crego (2017)
timent. The literature of psychology proposes that emotions can distort how people process information and make decisions. Tiedens and Linton (2001) cites a large body of literature that associates certain emotions with a sense of uncertainty and claims that people engage in heuristic processing instead of systematic processing when they feel uncertain. This distortion of behavior in financial market is generally believed to contribute to price anomalies in financial markets, such as overreaction and underreaction to news (see, e.g., De Bondt and Thaler (1985), De Bondt and Thaler (1987), Bernard and Thomas (1990) Jegadeesh and Titman (1993)). Investor sentiment is one of the most influential emotions in financial markets, and so it is also one of the most thoroughly studied, as shown in Lee et al. (1991), Barberis et al. (1998) and Baker and Wurgler (2006). Using proxies for sentiment, researchers find disparity in investor behavior and asset prices during periods of high and low sentiment (see, e.g., Bergman and Roychowdhury (2008), Antoniou et al. (2013), Yu and Yuan (2011), Stambaugh et al. (2012)). Considering that the volume-volatility relationship is the result of investors’ interactions around information release, we hypothesize that it will become different when the sentiment level changes.

Accordingly, we separate our sample into high- and low-sentiment periods, based on the sample median of the sentiment index from Baker and Wurgler (2006). We estimate the effect of disagreement measures on volume-volatility elasticity in both high- and low-sentiment periods using high-frequency data, using the econometric methods in Bollerslev et al. (2017). The estimate is based on jumps in nonparametrically estimated volume intensity and spot volatility upon information arrival, which
follows the idea of regression discontinuity design as in Hahn et al. (2001)\textsuperscript{2}. We also implement difference-in-difference to control for intra-day seasonality. Measures of disagreement include forecast dispersion on real economy indicators, the level of economic uncertainty, and a text-based measure of negative tone. The subsample analysis reveals that the negative effect of the disagreement measures is only significant during high-sentiment periods for the S&P 500 ETF. In sharp contrast to this, none of the disagreement measures have significant explanatory power in low-sentiment periods.

This result is consistent with theoretical predictions. In the difference-of-opinions model, the effect of investor disagreement is magnified by investors’ confidence in interpreting the information from public news. In our sample, low-sentiment periods mainly correspond to recessions, so the diminished explanatory power of disagreement measures demonstrates a decrease in investors’ confidence during recessions. There can be several underlying reasons for this change. Psychologically, people become regretful and lose confidence when they suffer losses in a bearish market. Indeed, in finance literature, confidence and sentiment are closely related: in some papers researchers use the confidence index of confidence as a direct proxy for sentiment (see, e.g., Bergman and Roychowdhury (2008) and Antoniou et al. (2013)). In addition, according to Yu and Yuan (2011), more sentiment investors enter the market when sentiment is high, and they are more prone to be less rational and more naive. In other words, these investors are more likely to be overconfident, and their overconfidence can be further aggravated by self-attribution. The combined effect of these factors

\textsuperscript{2} Our aim is to focus only on the “abnormal” changes in the volume intensity and spot volatility around information release, which are independent of trading behaviors not triggered by news, such as noisy trading and low-frequency dynamics. Therefore, though the methodology is similar to regression discontinuity design, we do not intend to study the treatment effect, as most RDD literature does.
contributes to the co-movement of sentiment and confidence level. As a result, when sentiment is low, even though dispersion still exists in investors’ interpretation, it no longer affect the volume-volatility elasticity because investors themselves do not take these interpretations seriously.

We go one step further to investigate whether the same pattern holds for individual stocks. Individual stocks are both closely related to and different from the market ETF. Apart from their systematic components, they also have idiosyncratic components. Though in classic asset pricing models idiosyncratic risks are seldom priced, they can still introduce additional risks when a portfolio is not perfectly diversified. Idiosyncratic risks are priced and shown to have predictive power as shown in Merton (1987) and Hahn et al. (2001). Empirically, however, there is no consensus on how idiosyncratic risks affect prices; see, e.g., Ang et al. (2006), Fu (2009), and Huang et al. (2009) for some inconclusive results. In line with this strand of literature and the difference-of-opinion models, we assume that investors associate the idiosyncratic component of individual stocks with heterogeneous prices that can be different from zero.

We incorporate a one-factor model with disagreements on the idiosyncratic term to extend the Kandel and Pearson (1995) model to a multi-asset setting, which admits the explicit solution of the volume-volatility elasticity for individual stocks. We show theoretically that the elasticity decreases monotonically in the overall relative disagreement level, which can further be decomposed into investors’ relative disagreement on the interpretation of public signals and their disagreement on the idiosyncratic component. The model demonstrates that for both the market index and individual stocks, investor disagreement only matters when investors have high con-
fidence in interpreting the public signals. It demonstrates in addition that the higher
the ratio of idiosyncratic volatility is, the larger the dispersion that can be introduced
by the firm specific term.

To investigate the elasticities cross-sectionally, we implement a fixed effect panel
regression with difference-in-difference, and estimate the elasticity of individual stocks
around macroeconomic announcements using the Dow Jones 30 components. We
then propose a multi-dimensional bootstrap to facilitate feasible statistical inferences.
Since the relative measure of investors’ disagreement is dispersion in interpreting
common signals relative to private information asymmetry, in this case our relative
measure of disagreement is the difference between standardized disagreement levels
and analysts’ forecast dispersions. The idiosyncratic ratio is the ratio between the
idiosyncratic volatility and the systematic volatility, estimated using high-frequency
data. The elasticities are much lower than those for the S&P 500 ETF, indicating
increases in perceived complexity in pricing individual stocks. Consistent with the
model’s predictions, the idiosyncratic ratio decreases elasticity, with both statistical
and economic significance. This is quite intuitive. Indeed, due to heterogeneity in
idiosyncratic risks, it is harder for investors to reach a consensus on stocks that have
high idiosyncratic volatility. The relative disagreement measures also negatively affect
the elasticity. Importantly, this result is still mainly driven by the high-sentiment
periods, the same as for the S&P 500 ETF.

Finally, we implement the same analysis on 10-year Treasury bond futures as a
benchmark. We find that the negative effects of disagreement measures are significant
in both high- and low-sentiment periods, and that the differences of the coefficients
are much weaker than for the securities. This result is as expected, since all of the
sentiment proxies used in the PCA in Baker and Wurgler (2006) measure speculative intent in the stock market, so the sentiment index may not be able to capture the sentiment level in the Treasury bond market. Besides, the disparity in the effect of disagreement measures depends on a diminished confidence level in interpreting the assets’ payoff from public news while traders in the bond market may not perceive the same confidence loss during low-sentiment periods. Indeed, the “flight to quality” or “flight to liquidity” phenomena are widely documented in literature, and they indicate relatively high confidence for participants in the Treasury market during recessions (see, e.g., Bernanke et al. (1994), Scholes (2000), Longstaff (2002), Lang and Nakamura (1995) and Caballero and Kurlat (2008)).

For a robustness check, we conduct the same analysis for redefined high- and low-sentiment periods using different cutoff points. We also redo the analysis using different disagreement measures. This essentially leads to no changes in our main conclusions.

This paper contributes to the literature of sentiment. Daniel et al. (1998) and Barberis et al. (1998) propose models that include investor sentiment, confidence and self-attribution to address overreaction and underreaction. Our result is consistent with the change in confidence level during different sentiment periods, and thus adds evidence to these models. Researchers document differences in asset prices during high- and low-sentiment periods. Antoniou et al. (2013) finds that momentum strategy is only profitable when sentiment is optimistic, while Yu and Yuan (2011) shows that the the positive mean-variance trade-off and the negative correlation between returns and contemporaneous volatility innovations only exist in low-sentiment periods. The disparity of effects created by investor disagreement presented in this paper is consis-
tent with this literature. However, instead of studying prices, we focus on the short window surrounding information releases and determine investors’ trading behavior by studying the volume-volatility relationship. In this way, we are able to show that sentiment can influence the financial markets by affecting investors’ decision-making upon information arrival.

This paper also makes a contribution to the literature on volume and volatility. In terms of individual stocks, while there are papers on price change, trading intensity, and volatility (see Lamoureux and Lastrapes (1990), Darrat et al. (2007) and Giot et al. (2010), for example), our paper is the first to combine a theoretical model with empirical practices based on high-frequency data. We extend the difference-of-opinions model and provide a theoretical foundation for the volume-volatility relationship around the time of information release. To test our model’s predictions, we apply econometric methods adapted to a multi-dimensional setting and present the empirical results. This adds new evidences regarding information processing upon systematic news release for individual stocks.

A third contribution of this paper is to the literature of idiosyncratic risks. In line with many papers in the literature, we assume in our model that the idiosyncratic term is priced. However, empirically, instead of focusing on the relationship between price and idiosyncratic volatility, we show that a stock with a larger idiosyncratic term will have smaller volume-volatility elasticity. Our explanation for this empirical result is that among investors with heterogeneous beliefs, perceived idiosyncratic risks are associated with dispersions in stock payoffs. Thus, this paper demonstrates the role that the idiosyncratic term plays in financial markets from an additional perspective.
This paper precedes as follows. Section 3.2 presents the model that motivates our empirical investigations. Section 3.3 describes the data we use for this paper, and our econometric methods are explained in Section 3.4. The empirical results are shown in Section 3.5. Section 3.6 assesses the robustness of the findings. Section 3.7 concludes. The proofs of the validity of the econometric methods and details for the model are given separately in Appendix B.1 and Appendix B.2.

3.2 Theoretical motivation

Our model incorporates a one factor structure with disagreement on the idiosyncratic term into the Kandel and Pearson (1995) model so that we can explicitly derive the individual stock’s volume-volatility elasticity around macroeconomic announcements. Since all these macroeconomic announcements are systematic news, it is natural to decompose the payoff of individual stock to a systematic component and an idiosyncratic one. Following the setting of Kandel and Pearson (1995), we impose heterogeneous beliefs in both components. We present the main assumptions and model predictions in this section, and detailed derivation can be found in Appendix B.

3.2.1 A one factor structure with heterogeneous beliefs

Suppose in the market there are two categories of assets: a risk free asset with zero return rate and \( n \) risky securities with an uncertain payoff \( \mathbf{V} \) where \( \mathbf{V} = [V_1, V_2, \ldots, V_n] \). There are two groups of investors, and a proportion of \( \alpha \) are of type 1.

Assumption 1. Investor \( j \) believes the payoff of asset \( i \) is

\[
V_{ij} = \beta_i V_{mj} + \varepsilon_{ij} \quad (3.1)
\]
where $V_{mj} \sim N(x_{mj}, z_{mj}^2)$, and $\varepsilon_{ij} \sim N(\gamma_i \eta_j, \gamma_i^2 z_{mj}^2)$.

In this one factor structure, the market component follows a normal distribution with mean $x_{mj}$ and $z_{mj}^2$. Investors disagree on the mean and variance of the market payoff.

Notice that the idiosyncratic term of stock $i$ is $\gamma_i^2 z_{mj}^2$, where the idiosyncratic coefficient $\gamma_i$ measures the magnitude of idiosyncratic volatility relative to the market volatility. Therefore, $\gamma_i / \beta_i$ is the ratio of idiosyncratic volatility to systematic volatility for each stock. The larger the ratio is, the more distinct this stock is from the market index.

For investor $j$, the idiosyncratic term has mean $\gamma_i \eta_j$ instead of 0. This dispersion may be due to private information, expertise, or simply biased belief in a particular stock $i$. It is magnified by $\gamma_i$, with the intuition that stocks with larger idiosyncratic component tend to generate high heterogeneity in beliefs. This assumption is in line with Stambaugh et al. (2015), which documents a positive relationship between idiosyncratic volatility and mispricing. Similar to the setting in their paper, $\eta_j$ can be positive, negative or zero. Indeed, Our result does not depend on the sign of $\eta_j$ but only requires that $\eta_j$s be different: one group is relatively optimistic and another relatively pessimistic. However, in our assumption $\eta_j$s can aggregate to 0, and hence we do not rely on mispricing to generate our result.

In conclusion, the assumption states that these investors agree on $\beta_i$ and $\gamma_i$, which means they agree on the proportion of the systematic and idiosyncratic variance of each individual stock. However, they disagree on the mean and variance of the
systematic component, and thus by construction on the variance of the idiosyncratic component. They also disagree on the mean of unit idiosyncratic term, \( \eta_j \).

Admittedly, investors can also disagree on \( \beta_i \) and \( \gamma_i \), in which case investors disagree on the relative magnitude of the systematic component and idiosyncratic component. Consider the aim of the paper is to study how different interpretations of systematic news affect the volume-volatility elasticity, we do not find it necessary to introduce different beliefs on the two parameters. Besides, this will make the model intractable.

### 3.2.2 The volume-volatility elasticity

Following the same setting as Kandel and Pearson (1995), investors observe a common signal \( L \). Investor \( j \) believes that

\[
L = V_{mj} + \nu_j \sim N \left( \mu_j, \sigma^2_{mj} \right).
\]

In this equation, \( \nu_j \) is the difference between the observed signal and the market payoff and the two investors interpret it differently. The smaller \( \mu_j \) is, the more optimistic the investor is about the market payoff. With this setting, we can solve for the equilibrium prices in both the two stages. The equilibrium price change \( \Delta P^* \) is as follows:

\[
\Delta P^* = \beta F_1 (L) + \gamma \mu \eta
\]

Actually, \( F_1 (L) \) has exactly the same expression as the price change in Kandel and Pearson (1995), which can be interpreted as the price change of the market price around the release of the public signal. \( \mu \eta \propto \eta_1 - \eta_2 \) is a constant, the magnitude of which is small compared to that of \( F_1 (L) \) under certain constraints of the pa-
parameters. Therefore, $\Delta P^*$ almost follows a perfect one factor structure around news announcement.

We assume that $F_1 (L) \sim (0, \sigma_{pm}^2)$ where $\sigma_{pm}$ is the volatility of the market price change. On average the price jump is nearly zero for high frequency data, so it is reasonable to set the mean at 0. The volatility of each stock $i$ is then $\sigma_{pi} = \beta_i \sigma_{pm}$, and hence we can correspondingly derive the joint distribution of the price changes.

$$\Delta P^* \sim N (\gamma \mu_{\eta}, \beta^2 \sigma_{pm}^2)$$

With these assumptions, the equilibrium volume $\text{Vol}_i^*$ is a function of the price change $\Delta P^*$. Both of the volume and price change are random, and empirically we observe the volume and price change in an “on average” sense. Therefore, we apply the moment condition and further solve for the relationship between volume intensity $v_i$, the average volume, and return volatility $\sigma_{pi}$, a measure of the average magnitude of the price change for stock $i$. The elasticity of $v_i$ on $\sigma_{pi}$ is

$$\mathcal{E}_i \equiv \frac{\partial v_i(\sigma_{pi})/v_i(\sigma_{pi})}{\partial \sigma_{pi}/\sigma_{pi}} = \frac{1}{1 + \psi(\theta_i)} \quad (3.2)$$

where $\psi(x) = x [\Phi(x)/\phi(x) - \frac{1}{2}] > 0$ increases monotonously when $x > 0$. We define $\theta_i$ as the “total disagreement” measure. For simplicity, we further assume $\sigma_{m1}^2 = \sigma_{m2}^2 = h^{-1}$, then $h$ can be viewed as the confidence in interpreting the signals. Correspondingly, $s_j = z_{mj}^{-2}$ donotes the precision of the prior belief in the market...
payoff, then the total disagreement $\theta_i$ can be further decomposed as follows:

$$
\theta_i = \theta^\text{sys}_i + \theta^\text{idio}_i
$$

$$
\theta^\text{sys}_i = \frac{h}{|s_1 - s_2|} \frac{\Delta \mu_m}{\sigma_p m}
$$

$$
\theta^\text{idio}_i = \left( \frac{h}{|s_1 - s_2|} + \xi(h, s_1, s_2) \right) g\left( \frac{\gamma_i}{\beta_i} \right) \frac{\Delta \eta}{\sigma_p m}
$$

In the idiosyncratic disagreement component, $\xi(h, s_1, s_2)$ is a positive function, and remains positive even if $h = 0$. $g(\gamma_i/\beta_i)$ is an increasing function of $\gamma_i/\beta_i$. $\Delta \mu_m$ and $\Delta \eta$ characterize investors’ heterogeneous beliefs in the payoff of the interpreted market term and the (per unit) idiosyncratic term:

$$
\Delta \mu_m = \mu_{m2} - \mu_{m1}, \quad \Delta \eta = \eta_1 - \eta_2.
$$

The first term $\Delta \mu_m$ is positive if investor 1 is more optimistic in interpreting the public signal signal than investor 2, since $\mu_{mj}$ is the interpreted mean of the difference between the signal and the fundamental value of the market payoff. Correspondingly, $\Delta \eta$ is positive if investor 1 prices the idiosyncratic volatility term higher than investor 2. Without loss of generality, we assume $\Delta \mu_m > 0$, then $\Delta \eta$ can be both positive and negative. As it is reasonable to believe that people who are optimistic about the systematic component is also likely to be optimistic about the idiosyncratic component, we assume $\Delta \eta > 0$. Then the elasticity of individual stocks will be smaller than that of the market, as $\mathcal{E}_i$ decreases in $\theta_i$. 
3.2.3 Testable hypotheses of the model

When $\Delta \eta = 0$, the elasticity is exactly the market elasticity. $\sigma_{pm}$ normalizes $\Delta \mu_m$, and we define

$$\frac{\Delta \mu_m}{\sigma_{pm} |s_1 - s_2|}$$

as the relative disagreement. It measures the disagreement on the same signal relative to private information before the signal is released. Note that $\gamma_i/\beta_i$ is the ratio of idiosyncratic volatility to systematic volatility in our setting, and is thus defined as the idiosyncratic-systematic ratio (Idio-Sys-Ratio for short) in this paper. As the recall that $E_i$ is a decreasing function of $\theta_i$, we have the following prediction:

**Claim 1.** When $h$ is large, investors have high confidence in interpreting the public signals.

1. The relative disagreement decreases the elasticity for both the market ETF and individual stocks.

2. Idio-Sys-Ratio decreases elasticity for individual stocks.

**Claim 2.** When $h$ is small, investors have low confidence in interpreting the public signals.

1. The relative disagreement measure has no effect on the elasticity for both the market ETF and individual stocks.

2. Idio-Sys-Ratio still decreases elasticity for individual stocks.

We will test them in Section 3.4, but before turning to that we present the data used in this paper.
3.3 Data description and summary statistics

In this section, we describe the data we use, which include the one minute transaction data, the macroeconomic announcements, the sentiment index, as well as the explanatory variables in the econometric models.

3.3.1 High-frequency market prices and trading volume

Our individual stock sample is comprised of Dow Jones 30 Components, as updated in March 18th, 2015. Three stocks (CVX, TRV and V) are excluded because of large period of missing data during the entire sample, from April 10th, 2001 to December 30th, 2014. The S&P 500 index ETF (ticker: SPY) is used as market index, both in regressions for market index as well as when beta and idiosyncratic variances are calculated. These one minute prices and volume data are obtained from the TAQ database, cleaned following Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). We also drop all half trading days following previous literature practices, and this results in 1,340,430 one-minute return and volume observations spanning 3437 trading days. The transaction data of 19-year treasury bond futures are from TickData, and span from July 1, 2003 to September 30, 2014.

3.3.2 Macroeconomic news announcements

Following BLX, we use the macroeconomic news announcements from the Economic Calendar Economic Release section in Bloomberg. Since individual stocks do not react as strongly as the market index, we only keep the most important indexes used in literature, plus housing index. The important indexes include FOMC announcements, ISM manufacture index, ISM nonmanufacture index, as well as Consumer confidence
index. Housing index is composed of all news announcements related to housing, including Existing Home Index, Housing Price Index, NAHB’s HMI, New Home Sales Index and Pending Home Sale Index. Therefore, the housing index happen to include most important indicators in the housing market. In this way, our results are less prone to be contaminated by random noises. We report the sample size and the typical announcement time in Table 3.1. Since there are coincidences of events at the same time, our sample contains 1116 events in total, from April 10th, 2001 to December 31st, 2014.

Table 3.1: Categories of macroeconomic news announcements

<table>
<thead>
<tr>
<th>Number</th>
<th>Time</th>
<th>Release Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOMC Rate Decision</td>
<td>111</td>
<td>14:15\dagger</td>
</tr>
<tr>
<td>ISM Manufacturing</td>
<td>162</td>
<td>10:00</td>
</tr>
<tr>
<td>ISM Non-Manf. Composite</td>
<td>161</td>
<td>10:00</td>
</tr>
<tr>
<td>Consumer Confidence Index</td>
<td>163</td>
<td>10:00</td>
</tr>
<tr>
<td>Housing Index§</td>
<td>607</td>
<td>10:00 and 13:00</td>
</tr>
</tbody>
</table>

\dagger Exact time varies from 14:00 to 14:15.
§ Includes Existing Home Sales Index, House Price Index, NAHB’s HMI, New Home Sales Index and Pending Home Sale Index.

3.3.3 The Sentiment Measure

The sentiment measure is from Baker and Wurgler (2006). It is based on the first principal component of five sentiment proxies: the closed-end fund discount, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium\(^3\). This measure of sentiment fits better into our story than the classic consumer confidence type of indexes because it captures the comovement of

\(^3\) There are six proxies in the original paper, while NYSE share turnover is dropped since “turnover does not mean what it once did, given the explosion of institutional high-frequency trading and the migration of trading to a variety of venues”, as described in their data file.
several sentiment proxies in the financial market. In our paper, we zoom into the one hour window around each macroeconomic announcement and estimate the abnormal volume and volatility, which is the net increase and is more or less clean of noise trading already. Therefore, we care more about the sentiment of the professional financial market participants rather than that of household, though the two may be closely correlated. Following Stambaugh et al. (2012) and Yu and Yuan (2011), we separate our sample into two different subsamples: high sentiment periods and low sentiment periods depending on whether the sentiment measure is above (or equal to) or below the sample median.

**Figure 3.1: The Sentiment Measure**

*Notes:* The figure plots the monthly sentiment measure, the daily annualized realized volatility of S&P 500 ETF, as well as its daily volume. High sentiment periods are shaded.

Figure 3.1 shows the plot of the sentiment measure along with daily annualized realized volatility and daily volume. From April 10th, 2001 to December 31st, 2014,
there are two concentrated periods of high sentiment: the Internet bubble and the economy expansion before the subprime mortgage crisis. In the latter half of the sample there are some high-sentiment months scattering around the European debt crisis and near the end of our sample. On the other hand, the low-sentiment periods mainly correspond to economic recessions after market crashes.

The realized volatility and volume are highly correlated, especially during the latter half of the sample when high-frequency trading pervades. There are several peaks of realized volatility: around 2002, in the latter half of 2008 and around 2011. The sentiment index and volatility do not co-move throughout the full sample, though. During the Internet bubble burst, sentiment is high and volatility is also high. However, prior to the subprime mortgage crisis, the expansion in economy is characterized by low volatility and high sentiment. Consider that volatility is highly correlated to risk measures, the sentiment measure is distinctive from those traditional risk measures.

3.3.4 Explanatory Variables: Disagreement Measures

We use the dispersion in one year ahead unemployment rate and the real GDP growth (in the robustness test) to proxy for difference in opinion in the systematic component of the stock market, since they are of crucial importance in real economy activities. For the treasury bond futures, one year ahead and ten year ahead forecast dispersion in CPI (in the robustness test) are added to account for CPI’s significant impact in the bond market. To avoid peeping into the future, the previous quarter’s value is used as explanatory variable for the current quarter.
Consider the documented high correlation of uncertainty measures with disagreement, we also introduce some uncertainty measures. One measure is NVIX, the news implied volatility measure from Manela and Moreira (2017). This measure, though aiming to predict VIX, is based on textual analysis. It applies machine learning techniques on the title and abstract of all front-page articles of the Wall Street Journal and extracts features to predict VIX. In this way, it avoids direct measure of variation in VIX, which may be correlated to volatility jump, our regressor. This measure is positively correlated with the Unemployment (RGDP growth) rate dispersion with a correlation coefficient of 0.398 (0.396). It is a news based measure and thus captures some different aspects from the RGDP dispersion.

Another uncertainty measure is Monthly Index from Baker et al. (2016). Compared to the weekly policy measure in BLX which is the a weekly moving average of the Daily Policy, there are several reasons for this change. Firstly, Monthly Index has three components: newspaper coverage of policy-related economic uncertainty, the number of federal tax code provisions set to expire in future years, as well as the dispersion in Economic forecast dispersion. The third component includes the dispersions in three indexes: consumer price index (CPI), purchase of goods and services by state and local governments, and purchases of goods and services by the federal government. As stated on the data’s webpage, macroeconomic measures are directly influenced by monetary policy and fiscal policy actions, and thus can be a proxy for uncertainty directly from these policy changes. This measure also accounts for the uncertainty about the tax code, which is, as stated by the authors, a source of uncertainty for businesses and households. Therefore, this new hybrid measure is more

\[ \text{A full description of the data can be found on the data's website.} \]
comprehensive, which should be more suitable when we include individual stocks. A second reason for using this measure is that it is of lower frequency, and thus less noisy than Weekly Policy. Consider the increased noise introduced by estimation of the volume and volatility of individual stocks, we prefer an uncertainty measure with smaller noises.

We also add into elasticity the FOMC sentiment measure from BLX. It is a measure of how negative the tone of each FOMC statements is, constructed following Loughran and McDonald (2011). It is essentially similar to the textual measures mentioned above, but may have higher explanatory power for FOMC announcements since it is directly related to the contents of the announcements.

Figure 3.2 plots the explanatory series. We first look at the low disagreement periods: before the financial crisis, high sentiment is accompanied by low disagreement, while after the European debt crisis, both the sentiment and the dispersion are low. There are also two peaks of dispersions: the burst of Internet bubble is characterized by high dispersion and high sentiment, while the subprime mortgage crises is accompanied by high dispersion and low sentiment. Therefore, our sample includes various high-low combinations of dispersion and sentiment index.

Generally, the uncertainty measures are very similar in trend to the dispersion measures. The most obvious difference is that NVIX tends to stay higher than the dispersion measures after the subprime mortgage crisis though. That means, in the second economy expansion stage, the sentiment and dispersion is generally low, but uncertainty remains relatively high.

The summary statistics of these explanatory variables are presented in Table 3.2. Each time series has been normalized to have a standard deviation of 1. Consistent
with Kim et al. (2014), we find that the summary statistics of the disagreement measures are similar in high- and low-sentiment periods, but they seem to be slightly higher when sentiment is low. Similarly, the uncertainty measures are also higher with larger standard deviation when sentiment is low. Bloom (2014) finds that uncertainty rises sharply during recessions, so this table is consistent with his research, considering that low-sentiment periods coincide with recessions. The negative measure stays low during the first half of our sample and gradually climbs up after 2007 before dropping down at the beginning of 2014. As the sentiment is low in the latter half of our sample, it is natural to find FOMC sentiment higher on average during low sentiment periods.
Table 3.2: Summary statistics of the explanatory variables

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>Subsample</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Median</th>
<th>Min</th>
<th>Q0.25</th>
<th>Q0.75</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>2.530</td>
<td>1.000</td>
<td>2.334</td>
<td>1.167</td>
<td>1.541</td>
<td>3.500</td>
<td>4.667</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>2.909</td>
<td>0.943</td>
<td>2.684</td>
<td>1.167</td>
<td>2.334</td>
<td>3.500</td>
<td>4.667</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>2.968</td>
<td>1.000</td>
<td>2.692</td>
<td>1.457</td>
<td>2.200</td>
<td>3.704</td>
<td>6.243</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>3.534</td>
<td>0.895</td>
<td>3.535</td>
<td>2.163</td>
<td>2.656</td>
<td>4.213</td>
<td>6.243</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>3.409</td>
<td>1.000</td>
<td>3.497</td>
<td>1.796</td>
<td>2.616</td>
<td>3.961</td>
<td>7.634</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>3.888</td>
<td>0.928</td>
<td>3.826</td>
<td>2.220</td>
<td>3.373</td>
<td>4.229</td>
<td>7.634</td>
</tr>
<tr>
<td></td>
<td>Full</td>
<td>1.194</td>
<td>1.000</td>
<td>1.060</td>
<td>0.000</td>
<td>0.363</td>
<td>1.582</td>
<td>4.206</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>1.511</td>
<td>0.997</td>
<td>1.307</td>
<td>0.000</td>
<td>0.789</td>
<td>2.192</td>
<td>4.187</td>
</tr>
</tbody>
</table>

The table reports the summary statistics of the explanatory variables. The mean, standard deviation, median, minimum, 25% quantile, 75% quantile as well as the maximum of the explanatory variables are reported for the full sample, and the high(low)-sentiment subsample. Each time series has been normalized so that the standard deviation is 1 for the full sample.

3.3.5 Explanatory Variables: Cross-sectional Variables

Asymmetric information plays a larger role in individual stock’s valuation. For the market index and the treasury bonds, it is hard for investors to possess private information about the whole financial market. Even if investors have access to private information about individual stocks, these different sources of information will aggregate out so there will be no much variation in the amount of asymmetric information for the market. However, for individual stocks, there is extended literature in support of the existence of private information. Analysts’ earnings forecasts for firms composed of more intangibles are documented to contain higher proportions of private information in Barron et al. (2002). Mayew et al. (2013) show that analysts who ask questions in conference calls possess private information advantages. Actually, it has been recognized in literature that analysts make forecasts both on publicly available and privately acquired information (e.g., Arnold and Moizer (1984), Barker (1998)). Therefore, it is reasonable to assume the existence of variations in information asymmetry among investors, both cross-sectionally and over time.
One common proxy for private information is analysts’ dispersion. In this paper, analysts’ dispersion is defined as the monthly standard deviation of one year ahead forecasts of EPS, normalized by the average price of the same month. The standard deviation is from I/B/E/S while the prices are from Compustat. To avoid peeping into future, we use the previous month’s analysts’ dispersion for the current month.

The idiosyncratic-systematic-ratio defined in Section 3.3.5 can be approximated using high-frequency stock returns. Apply the methods detailed in Patton and Verardo (2012), we use 15 minutes’ return to obtain monthly beta, $\beta_{it}$ for each individual stock. The monthly idiosyncratic variances are calculated as follows:

$$\text{IdioRV}_{it} = \text{RV}_{it} - \beta_{it}^2 \text{RV}_{mt},$$

where $\text{RV}_{it}$ is the monthly realized variance of stock $i$, calculated using 15 minutes’ return again. $\text{RV}_{mt}$ is the concurrent market realized variance. To be consistent with the topic in this paper, we only include intraday returns. We define Idio-Sys Ratio as the ratio of idiosyncratic variance to systematic variance:

$$\text{Idio-Sys Ratio} = \sqrt{\frac{\text{IdioRV}_{it}}{\beta_{it}^2 \text{RV}_{mt}}}, \quad (3.3)$$

Figure 3.3 shows the 95% empirical confidence interval and the mean value of analysts’ dispersion in different sentiment periods. With a Wilcoxon Rank-Sum test, we find that analysts’ dispersion is significantly higher when sentiment is low except for IBM. This is consistent with the time series dispersion measures about the real
economy. Dispersions vary dramatically among stocks, which are high in financial and utility companies but low in retail and manufacure ones.

Figure 3.3: Analysts’ dispersion in different sentiment periods

Notes: The figure shows the distribution of analysts’ dispersion for each individual stock during high-sentiment subsample, the full sample and the low-sentiment subsample. When the text is red (blue) it means the mean of analysts’ dispersion is significantly higher using The Wilcoxon Rank-Sum Test.

Figure 3.4 is a similar plot of the Idio-Sys Ratio. All stocks except three have larger gamma-beta ratio in low-sentiment periods. This is mainly driven by a soaring idiosyncratic variance when sentiment is high.

3.4 High-frequency econometric procedures

We already have derived the elasticity as shown in Equation (3.2). However, the model is stylish and we need to estimate the elasticity from high-frequency data: one-
Figure 3.4: the $\gamma-\beta$ ratio in different sentiment periods

Notes: The figure shows the distribution of $\gamma-\beta$ ratio for each individual stock during high-sentiment subsample, the full sample and the low-sentiment subsample. When the text is red (blue) it means the mean of $\gamma-\beta$ ratio is significantly higher using The Wilcoxon Rank-Sum Test.

minute prices and volumes. Therefore, in this section we first present how volume and volatility can be estimated, and then continue to propose the econometric model from which we can estimate the elasticity and the effect of the explanatory variables. Finally, bootstrapping algorithms are shown for statistical inferences. Many of the ideas are essentially similar to those in Bollerslev et al. (2017), but we need to adapt to the panel data and subsample analysis in this section.

3.4.1 Price and volume process

We fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. Let $(P_{m,t})_{t \geq 0}$ denote the logarithmic price process of the market asset and $(P_{i,t})_{t \geq 0}$ be the price process of individual stock $i$. $P_{m,t}$ and $(P_{i,t})_{t \geq 0}$ follow from continuous-time finance literature (see, e.g.,
Merton (1992) and Duffie (2001)), and are assumed to be jump-diffusion processes of the form

\[ dP_{m,t} = b_{m,t}dt + \sigma_{m,t}dW_{m,t} + dJ_{m,t}, \tag{3.4} \]
\[ dP_{i,t} = b_{i,t}dt + \beta_{i,t}\sigma_{m,t}dW_{m,t} + \sigma_{i,t}dW_{i,t} + dJ_{i,t} \tag{3.5} \]

where \( b \) is an instantaneous drift process, \( \sigma \) is a stochastic spot volatility process, \( W \) is a Brownian motion, and \( J \) is a pure jump process. We assume that \( W_{m,t} \) is perpendicular to \( W_{i,t} \) and \( W_{i,t} \) is perpendicular to \( W_{j,t} \) if \( i \neq j \). In this way we decompose the price process of each individual stock as the sum of the market component and the idiosyncratic component. We do not specify the structure of the drift process or the jump process because they will not enter the spot volatility in our setting.

Let \( T \) denote the sample span and \( \Delta_n \) denote the sampling interval of the high-frequency data. Assume the prices are sampled at discrete times \( \{l\Delta_n : 0 \leq l \leq [T/\Delta_n]\} \). The corresponding high-frequency asset return is denoted by \( r_l \equiv P_{l\Delta_n} - P_{(l-1)\Delta_n} \).

Let \( V_{l\Delta_n} \) denote the trading volume within the high-frequency interval \( ((l-1)\Delta_n, l\Delta_n] \). Following Li and Xiu (2016), we consider a general state-space model for both the market index and individual stocks

\[ V_{m,l\Delta_n} = \mathcal{V}(\zeta_{m,l\Delta_n}, \epsilon_{m,l\Delta_n}), \tag{3.6} \]
\[ V_{i,l\Delta_n} = \mathcal{V}(\zeta_{i,l\Delta_n}, \epsilon_{i,l\Delta_n}), \tag{3.7} \]

where as stated in their paper, \( \zeta \) is a latent state process, \( (\epsilon_{\Delta_n}) \) are i.i.d. transitory shocks with distribution \( F_\epsilon \), and \( \mathcal{V}(\cdot) \) is a general transform function. The latent
state process \( \zeta \) captures time-varying conditioning information such as the intensity of order arrival and the shape of the order size distribution. Note that we are not imposing any restrictions on the structure of \( \zeta_{i,l,\Delta_n} \) (resp. \( \epsilon_{i,l,\Delta_n} \)) and it does not need to be perpendicular to \( \zeta_{m,l,\Delta_n} \) (resp. \( \epsilon_{m,l,\Delta_n} \)). We are not imposing restrictions on the form of \( F_\epsilon \) either, and the distributions can be different for the market index and individual stocks.

With this specification, the \textit{volume intensity} process is calculated by

\[
m_t = \int \mathcal{V}(\zeta_t, \epsilon) F_\epsilon \, (d\epsilon),
\]

which is the instantaneous conditional mean process of \( V \). This process is used as the instantaneous empirical analogue to the expected volume in the theoretical models discussed in Section 3.2, both for the market and the individual stocks. Correspondingly, the spot volatility processes \( \sigma \) is used as the instantaneous analogue of the return standard deviation.

### 3.4.2 Volume and volatility jump

With the setup above, we need to get the volume jump and volatility jump from the one minute return and one minute volume data. Let \( \tau \) denote the pre-scheduled announcement time. All announcements in our sample are pre-scheduled to be released on a certain minute, so we suppose \( \tau = (l(\tau) - 1) \Delta_n \) for a particular integer \( l(\tau) \),

where \( \Delta_n = 1 \) minute is the sampling interval of our intraday data. Then the event window is defined as \( (l(\tau) - 1) \Delta_n, l(\tau)\Delta_n) \).

\footnote{All announcements in our sample are indeed scheduled to be on the minute.}
Correspondingly, the $k_n$-minute periods immediate before (after) the event is defined as the pre-event (post-event) window. Denote stock $i$’s return and volume in the $((k - 1) \Delta_n, k \Delta_n)$ window by $r_{i,k}$ and $V_{i,k \Delta_n}$, respectively. Let $m_{i,\tau-}$ ($m_{i,\tau}$) denote the pre-event (post-event) volume intensity, and let $\sigma_{i,\tau-}$ ($\sigma_{i,\tau}$) denote the spot volatility, then they can be estimated as follows:

$$
\hat{m}_{i,\tau-} \equiv \frac{1}{k_n} \sum_{k=1}^{k_n} V_{i,(l(\tau)-k)\Delta_n}, \quad \hat{m}_{i,\tau} \equiv \frac{1}{k_n} \sum_{k=1}^{k_n} V_{i,(l(\tau)+k)\Delta_n},
$$

$$
\hat{\sigma}_{i,\tau-} \equiv \sqrt{\frac{1}{k_n\Delta_n} \sum_{k=1}^{k_n} r_{i,l(\tau)-k}^2}, \quad \hat{\sigma}_{i,\tau} \equiv \sqrt{\frac{1}{k_n\Delta_n} \sum_{k=1}^{k_n} r_{i,l(\tau)+k}^2}.
$$

(3.9)

We denote the corresponding log volatility and log volume jumps by $\Delta \log (\sigma_{i,\tau}) \equiv \log(\sigma_{i,\tau}) - \log(\sigma_{i,\tau-})$ and $\Delta \log (m_{i,\tau}) \equiv \log(m_{i,\tau}) - \log(m_{i,\tau-})$, respectively. The reason we use log difference is that both volume and volatility have time variation which is not of interest in this paper. For example, there is an obvious upward trend in volume for many stocks due to expanding high-frequency trading. Volumes and volatilities are also sensitive to liquidity and transaction costs, which are hardly affected by news announcements, and thus are not closely related to our topic either. By substracting pre-event volume (volatility) from post-event volume (volatility), we only take account of the “abnormal” volume and volatility in order to focus only on news-triggered trading behaviors. We nonparametrically estimate the log volatility and log volume jumps as follows:

$$
\Delta \log (m_{i,\tau}) = \log(\hat{m}_{i,\tau}) - \log(\hat{m}_{i,\tau-}), \quad \Delta \log (\sigma_{i,\tau}) = \log(\hat{\sigma}_{i,\tau}) - \log(\hat{\sigma}_{i,\tau-})
$$

(3.10)
Figure 3.5: Scatter plot of log volume jumps on log volatility jumps

Notes: The figure shows the scatter plot of log volume jumps on log volatility jumps for 27 Dow Jones 30 components and the market ETF.
Figure 3.5 demonstrates the positive correlation between log volume jumps and log volatility jumps for Dow Jones 30 components and the S&P 500 index.

3.4.3 Fixed effect panel regression

Instead of a univariate linear model, now we model the volume-volatility relationship of each stock as follows:

\[
\Delta \log (m_{i\tau}) = c_i + a_0^\top D_\tau + (b_0 + a_1^\top D_\tau + b_1^\top X_{i\tau}) \cdot \Delta \log (\sigma_{i\tau}), \tag{3.11}
\]

where \(\Delta \log (m_{i\tau})\) is the jump in stock \(i\)'s log volume intensity and \(\Delta \log (\sigma_{i\tau})\) is the jump in its log spot volatility. \(D_\tau\) are dummy variables that represent the five categories of announcements except FOMC announcements when we combine all announcements, so the estimates around FOMC announcements are the benchmark. \(X_{i\tau}\) is comprised of the different explanatory variables employed in the estimation. That means, though stocks differ in intercepts and slopes, they share the same specification in slopes: the coefficients of the explanatory variables are the same across stocks, though the values of these variables differ. One may also notice that different categories also share the same coefficients on the explanatory variable. The reason is, when we conducted the same analysis on each individual category, we found similar coefficients for these explanatory variable, and thus we combine these categories to increase sample size to reduce estimation error.

Consider the group \(\mathcal{A}\) which is a set of a total of \(M\) announcement times. We also implement DID to take account of intraday patterns, as shown in Equation 3.12, where \(\mathcal{C}(\tau)\) is a control group that occurs at the same time on different days and \(N_C\) is the size of the control groups. In our regressions, \(\mathcal{C}(\tau)\) is the set of the same
time as τ on the 22 days without announcements before the event day. That means, we use the average log volume (volatility) jump at the event time of the previous 22 no-announcement days as control.

\[
\Delta \log (m_{i\tau}) \equiv \Delta \log (m_{i\tau}) - \frac{1}{N_C} \sum_{\tau' \in C(\tau)} \Delta \log (m_{i\tau'}),
\]

\[
\Delta \log (\sigma_{i\tau}) \equiv \Delta \log (\sigma_{i\tau}) - \frac{1}{N_C} \sum_{\tau' \in C(\tau)} \Delta \log (\sigma_{i\tau'}),
\]

(3.12)

Therefore, our question to investigate what affects volume-volatility elasticity is resolved by a fixed-effect panel regression as follows:

\[
\Delta \log (m_{i\tau}) = c + a_0^\top D_{\tau} + (b_0 + a_1^\top D_{\tau} + b_1^\top X_{i\tau}) \cdot \Delta \log (\sigma_{i\tau}) + \varepsilon_{i\tau}.
\]

(3.13)

Note that this panel regression is different from the classic ones in that we do not assume the sample size goes to infinity. Instead, the size of the stock is relatively small, compared to the long time horizon. Therefore, fixed effect model does not decrease efficiency much, but ensures consistency. The asymptotic of the regression is not achieved by sample size going to infinity, but by the infinitely small sample interval and event window around announcements.

Besides the panel regression, we also run regressions for the market ETF and treasury bond futures. These are the univariate version of the above equation:

\[
\Delta \log (m_{\tau}) = c + a_0^\top D_{\tau} + (b_0 + a_1^\top D_{\tau} + b_1^\top X_{\tau}) \cdot \Delta \log (\sigma_{\tau}) + \varepsilon_{\tau}.
\]

(3.14)

For the market index, we also run regressions for FOMC announcements individually. In this case, \(D_{\tau} = 0\).
Our goal is therefore to conduct valid inference about the parameter vector \( \theta \equiv (a_0, b_0, a_1, b_1) \), especially the components \( b_0 \) and \( b_1 \) that shows the benchmark elasticity and the effect of explanatory variables.

Moreover, solving the fixed effect panel is equivalent to estimating the parameter vector \( \theta \equiv (a_0, b_0, a_1, b_1) \) in (3.13) for the group \( \mathcal{A} \) using the following least-square estimator

\[
\hat{\theta} \equiv \underset{\theta}{\text{argmin}} \sum_{i=1}^{n} \sum_{\tau \in \mathcal{A}} \left( -\phi \left( \Delta \log \left( m_{it} \right) \right) \right)^2 \tag{3.15}
\]

where

\[
\phi (Y_{it}) = Y_{it} - \frac{1}{M} \sum_{\tau \in \mathcal{A}} Y_{it}.
\]

The estimator \( \hat{\theta} \) requires the information set of both the treatment group and control group \( \mathcal{A} (\tau) \) and \( \mathcal{C} (\tau) \). Define \( S_\tau \equiv \cup_{i=1}^{n} (m_{i\tau-}, m_{i\tau}, \sigma_{i\tau-}, \sigma_{i\tau}, X_{i\tau})_{i \in \mathcal{N}} \), then \( S = (S_\tau)_{\tau \in \mathcal{A} \cup \mathcal{C}} \), contains all information needed for the estimation. Our estimator of \( S \) may then be expressed as \( \hat{S}_n = (\hat{S}_\tau)_{\tau \in \mathcal{A} \cup \mathcal{C}} \), where \( \hat{S}_\tau = \cup_{i=1}^{n} (\hat{m}_{i\tau-}, \hat{m}_{i\tau}, \hat{\sigma}_{i\tau-}, \hat{\sigma}_{i\tau}, X_{i\tau})_{i \in \mathcal{N}} \). \( X_{i\tau} \) are the explanatory variables as described in Section 3.3, and \( \hat{m}_{i\tau-}, \hat{m}_{i\tau}, \hat{\sigma}_{i\tau-}, \hat{\sigma}_{i\tau} \) are the pre and post event volume intensity or spot volatility as defined in (3.9). Then this estimator can be expressed as \( \hat{\theta} = f(\hat{S}) \).

In Appendix A we show that \( \hat{S} \) is a consistent estimator of \( S \), which in turn implies that \( f(\hat{S}) \) consistently estimates \( f(S) \), provided \( f(\cdot) \) is a smooth function of the estimated quantities.

The bootstrap methods is the multi-dimensional version of Bollerslev et al. (2017): we re-sample vectors of returns and volumes to make sure that we keep the co-
occurrences of trading across different stocks. The validity of the bootstrap algorithms depends still on the locally i.i.d assumption of the re-sampled vectors of volumes and returns, from the pre-announcement window and post-announcement window. The bootstrapped sample size is 1000 in all regressions.

**Bootstrap Algorithm**

Step 1: For each $\tau \in \mathcal{T}$, generate i.i.d. draws $(V_{i(\tau)-k}^*, r_{i(\tau)-j}^*)_{1 \leq k \leq n}$ and $(V_{i(\tau)+k}^*, r_{i(\tau)+k}^*)_{1 \leq k \leq n}$ from $(V_{i(\tau)-k}, r_{i(\tau)-k})_{1 \leq k \leq n}$ and $(V_{i(\tau)+k}, r_{i(\tau)+k})_{1 \leq k \leq n}$, respectively. $V_t$ is the vector of one minute volumes of the Dow Jones 30 components, while $r_t$ is the vector of one minute returns of the Dow Jones 30 components.

Step 2: Compute $\Delta \log (m_{i\tau})^*$ and $\Delta \log (\sigma_{i\tau})^*$ the same way as $\Delta \log (m_{i\tau})$ and $\Delta \log (\sigma_{i\tau})$, respectively, except that the original data $(V_{i(\tau)-k}, r_{i(\tau)-k})_{1 \leq k \leq n}$ is replaced with $(V_{i(\tau)-k}^*, r_{i(\tau)-k}^*)_{1 \leq k \leq n}$. Rerun the regression according to (3.13) and get $b^*$.

Step 3: Repeat steps 1 and 2 a large number of times. Report the empirical standard errors of each component of $b^* - b$ as the standard errors of the original estimator $b$.

**Differences of coefficients in high- and low-sentiment periods**

In this paper, we would like to investigate how regression results differ in high- and low-sentiment periods. Though the two samples do not overlap, they may share the same days as control, so the coefficients in the two subsamples are not independent. Therefore, we implement the following algorithm to generate statistical inference for the differences in coefficients in high- and low-sentiment periods:
Step 1: Separate the full sample into high-sentiment subsample and low-sentiment one. Run regressions for each subsample and get coefficients $b_H$ and $b_L$. Define the difference as $b_{\text{diff}} = b_H - b_L$.

Step 2: In each round of bootstrap, we repeat the above and get $b_{\text{diff}}^{(s)}$, where $s = 1, 2, \ldots, 1000$.

Step 3: Use the empirical standard errors of $b_{\text{diff}}^{(s)}$ as the standard errors of $b_{\text{diff}}$.

The validity of the bootstrap algorithms is proved in Appendix A.

3.5 Volume-volatility relationship around public announcements

In this section, we present the empirical results for the S&P 500 ETF (SPY) which represents the market index, the 10-year Treasury bond futures (TY) which represents the risk free assets, as well as the Dow Jones 30 components (DJ30) which are individual stocks. We first take a look at the log jumps of all these assets, and then study the regression results one by one.

3.5.1 Log volatility and volume jumps

Before we delve into the volume-volatility relationship in different sentiment periods, we check how the behavior of volume and volatility differs when sentiment is high and low. As can be seen from Figure 3.6, the differences between the two periods seem small for volatility. However, as shown in Figure 3.7, TY, SPY and more than one third of the stocks experience higher average volume jumps during high-sentiment periods. An interpretation for this is that investors’ dispersion is higher when sentiment is high. Garcia (2013) argues that in the literature of behavioral finance, volume jump should be zero when public information is released, if investors
are fully rational and possess symmetric information. Therefore, one reason of the higher volume jump may be the large inflow of less rational investors when sentiment is high. Besides, one can also notice much higher volume and volatility jumps around FOMC announcements, highlighting its importance for not only the S&P 500 ETF and Treasury bond futures, but also the individual stocks.

![Diagram](image)

**Figure 3.6: Average log volatility jumps in different sentiment periods**

*Notes:* The figure shows the mean and the 95% confidence interval of the daily volatility jumps of the Dow Jones 30 components, S&P 500 ETF (SPY) and ten year Treasury bond futures (TY) around FOMC announcements and all announcements. For each asset, the color of the ticker being red (blue) indicates that the mean is higher during high (low) sentiment period.

### 3.5.2 Volume-volatility elasticities around public news announcements

The results for regression (3.14) are reported in Table 3.3. To save space we do not report the coefficients of the dummy variables, so $b_0$ is the elasticity of volume-
Table 3.3: Volume-volatility elasticity estimates for S&P 500 ETF

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th>Full Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.696**</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.056**</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.065**</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.090**</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.547</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th>high-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.696**</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.115**</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.146**</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.160**</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th>Low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.757**</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>NVIX</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.558</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th>Difference between high and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.125**</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.191**</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.165**</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.558</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log(p_m) = c + a_0 T_a + (b_0 + a_1 T_a + b_1 T_b) X_t + \Delta \log(\sigma^2) + \varepsilon_t$. In all specifications, $D_a$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1$) are not reported for brevity. The UNEMP Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). NVIX is the news implied volatility measure from Manela and Moreira (2017). All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The four panels from top to bottom present regression results for the full sample, the high-sentiment subsample, the low-sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
Figure 3.7: Average log volume jumps in different sentiment periods

Notes: The figure shows the mean and the 95% confidence interval of the daily volume jumps of the Dow Jones 30 components, S&P 500 ETF (SPY) and ten year Treasury bond futures (TY) around FOMC announcements and all announcements. For each asset, the color of the ticker being red (blue) indicates that the mean is higher during high (low) sentiment period.

volatility elasticity around FOMC announcements. From the top to the bottom panel are results for the full sample, high-sentiment periods, low-sentiment periods and the differences of coefficients between high- and low-sentiment periods. The sentiment periods are separated according to the Baker and Wurgler (2006) sentiment index, and the differences and their statistics are generated by the algorithm in Section 3.4.3. Robustness check to other cutoffs of sentiment periods are presented in Section 3.6.1.

The results in the first panel are consistent with those in Bollerslev et al. (2017), where both the coefficients for unemployment dispersion and monthly index are significantly negative. When they are put together, the signs remain negative but the
unemployment dispersion variable no longer has significant explanatory power. We see that NVIX is the strongest among the three, directly driving the elasticity to around 1 and rendering the other two explanatory variables insignificant. One interpretation of this result is that the news implied VIX measure proxies difference of opinions best around all macroeconomic announcements. Considering that NVIX and Monthly Index are very similar inherently, in the following regressions we use only NVIX and put Monthly Index into the robustness tests, which are shown in Section 3.6.2.

A glance at the second and third panels will reveal the difference between the two sentiment periods: all the explanatory power of the disagreement measures only becomes significant in the high-sentiment periods. When we compare the results column by column, we notice that when no explanatory variables are added into the elasticity, the baseline elasticity $b_0$ is slightly smaller in high-sentiment period with insignificant difference. When the disagreement measures are added, the magnitudes of the coefficients in the high-sentiment periods are around twice as large as those in the full sample, and the coefficient of unemployment dispersion becomes significant even when MonthlyIndex is added in. NVIX still seems to dominate, driving the unemployment dispersion insignificant, but the latter’s negative sign remains. On the contrary, none of the coefficients are significant in the low-sentiment subsample individually, and the differences are significant for all these coefficients.

When we control for the disagreement measures, we see that during high sentiment periods the elasticity gradually increases to around 1, the theoretical value when there is no disagreement, which means that these explanatory variables captures the disagreement quite well in high-sentiment periods. In the low-sentiment periods, on
the contrary, the elasticity fluctuates but does not increases towards 1. The dramatically increased adjusted $R^2$ during high-sentiment regime adds further evidence to the effectiveness of the explanatory variables, contrasting the seldom changed $R^2$ in the low-sentiment period. We also conducted F tests and find the p values during high-sentiment periods always less than 0.002, while in low sentiment periods they never drop below than 0.1. With all these evidences we safely conclude that the explanatory variables only matter during high sentiment periods.

Next we focus our attention on FOMC announcements and report the results in Table 3.4. Consistent with Bollerslev et al. (2017), both the coefficients of unemployment dispersion and NVIX are significantly negative. NVIX still plays an important role and dominates the two. FOMC sentiment is also significantly negative, even when it is put together with NVIX and unemployment dispersion.

The patterns of the coefficients when the sample is split into two are similar as those for all announcements. All the explanatory variables individually are significantly negative, with much larger magnitude in high sentiment periods. Unemployment dispersion is now significantly negative even when other variables are added in, highlighting the importance of the dispersion about real economy around FOMC announcements. FOMC sentiment’s effect dominates that of NVIX. This reasonable considering that FOMC sentiment, a tone measurement of the FOMC statements themselves, is more closely related to FOMC announcements. Therefore, for FOMC announcements the FOMC sentiment measures seem to be a better measure of uncertainty compared with NVIX, as the latter is a more general uncertainty measures.

A sharp contrast to these highly significant variables are the ones in the low-sentiment period, none of which are significant anymore. Also contrasting to the
Table 3.4: Volume-volatility elasticity estimates for S&P 500 ETF around FOMC announcements

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>High-sentiment periods</th>
<th>Low-sentiment periods</th>
<th>Difference between high and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Elasticity ($b_0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.696**</td>
<td>0.142**</td>
<td>0.757**</td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.112)</td>
<td>(0.093)</td>
<td>(0.131)</td>
</tr>
<tr>
<td></td>
<td>0.884**</td>
<td>1.214**</td>
<td>0.736**</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td>(0.132)</td>
<td>(0.126)</td>
<td>(0.167)</td>
</tr>
<tr>
<td></td>
<td>1.040**</td>
<td>1.247**</td>
<td>0.824**</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.135)</td>
<td>(0.183)</td>
<td>(0.225)</td>
</tr>
<tr>
<td></td>
<td>1.057**</td>
<td>1.247**</td>
<td>0.807**</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(0.100)</td>
<td>(0.135)</td>
<td>(0.188)</td>
<td>(0.232)</td>
</tr>
<tr>
<td></td>
<td>0.828**</td>
<td>0.013</td>
<td>0.763**</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.025)</td>
<td>(0.114)</td>
<td>(0.030)</td>
</tr>
<tr>
<td></td>
<td>1.043**</td>
<td>0.013</td>
<td>0.808**</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.025)</td>
<td>(0.197)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UNEMP Dispersion</td>
<td>0.007</td>
<td>0.067</td>
<td>-0.209**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.051)</td>
</tr>
<tr>
<td></td>
<td>NVIX</td>
<td>-0.016</td>
<td>-0.021</td>
<td>-0.143**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.034)</td>
<td>(0.035)</td>
<td>(0.054)</td>
</tr>
<tr>
<td></td>
<td>FOMC Sentiment</td>
<td>-0.003</td>
<td>-0.001</td>
<td>-0.146**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.317</td>
<td>0.420</td>
<td>0.364</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>0.340</td>
<td>0.419</td>
<td>0.352</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>0.358</td>
<td>0.442</td>
<td>0.353</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>0.356</td>
<td>0.341</td>
<td>0.351</td>
<td>0.327</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log(p_{cm}) = \epsilon + (b_0 + b_1' X_t') \Delta \log(\sigma_\tau) + \xi_t$. In all specifications, the UNEMP Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. NVIX is the news implied volatility measure from Manela and Moreira (2017). FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The four panels from top to bottom present regression results for the full sample, the high-sentiment subsample, the low-sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
increasing adjusted $R^2$ in the high-sentiment periods are the continuously dropping adjusted $R^2$ during low-sentiment periods when more variables are added in. The F tests show that the p values are always less than 0.02 in the high-sentiment periods and are never smaller than 0.4 during the low-sentiment regime.

The last panel further confirms the significant differences between the coefficients in the two subsamples. This is consistent with our hypothesis that only when $h$ is high will disagreements significantly reduce elasticities. Intuitively, when sentiment is high, investors’ portfolios are performing well. Due to self-attribution they prone to be over-confident, which means $h$ is large and magnifies the relative disagreement term. On the contrary, low-sentiment period usually corresponds to recessions, during when investors suffer from losses. Under such circumstances they become frustrated and regretful, and gradually lose confidence in themselves. $h$ is so small at this time that relative disagreement no longer cast effect in the volume-volatility elasticity. Another possible explanation for this disparity is that, as argued in Yu and Yuan (2011), when sentiment is high there is higher proportion of sentiment traders, who tend to be inexperienced and naive, and thus are more likely to be victims of overconfidence and self–attribution. When market is filled more of investors with high $h$, their belief dispersion plays a larger role around information arrival.

3.5.3 Dow Jones 30 Components

In this section, we study what will affect individual stocks’ volume-volatility elasticity. Individual stocks are more complicated than the market index in that there can be variations in information asymmetry and idiosyncratic risks over time and across
stocks. In our model, the systematic disagreement term is

\[ h \frac{|\Delta \mu|}{|s_1 - s_2| \sigma_{pm}} \]

where \( |s_1 - s_2| \) is the difference in the prior precision, which can be viewed as asymmetry in private information. Note that \( \frac{|\Delta \mu|}{\sigma_{pm}} \) is the difference in normalized interpretations of the public information, and we define

\[ \frac{\Delta \mu}{\sigma_{pm} (s_1 - s_2)} \]

as the relative disagreement measure in Section 3.3, which is what really matters in this model.

For market index, asymmetry in priors may not play a large role over time, since it is hard to access private information about the market over time. Even if different investors possess asymmetric information in different stocks, they will aggregate out in the market portfolio. Therefore, it is reasonable to assume relatively stable information asymmetry along time. However, this is not the case for individual stocks. Investors may actively acquire private information, and hence there may be both time series variation and cross-sectional variation in information asymmetry for individual stocks, which we need to take account of.

To characterize this, we construct a “relative disagreement” measure, defined as follows:

\[ \text{Relative } X_{i\tau} = X_\tau - \widehat{AFD}_{i\tau} \]

(3.16)

\( \widehat{AFD}_{i\tau} \) is normalized \( AFD_{i\tau} \), which is the analysts’ dispersion for stock \( i \) at announcement time \( \tau \). As described in Section 3.3.4, it is the monthly standard deviation of
one-year-ahead forecasts for EPS, normalized by the average price of the same month. Therefore, \( AFD_{i\tau} \) is comparable across stocks and over time. \( X_{\tau} \) is the disagreement measure at time \( \tau \). To make \( AFD_{i\tau} \) comparable to \( X_{\tau} \) while keep its cross-sectional as well as time series variation, we divide each \( AFD_{i\tau} \) by a normalizing factor \( \sigma_{AFD} \), where

\[
\sigma_{AFD} \equiv \frac{1}{n} \sum_{i=1}^{n} \sqrt{\frac{1}{T-1} \sum_{\tau=1}^{T} \left( AFD_{i\tau} - \frac{1}{T} \sum_{\tau=1}^{T} AFD_{i\tau} \right)^2}
\]

\[
\widehat{AFD}_{i\tau} \equiv \frac{AFD_{i\tau}}{\sigma_{AFD}}
\]

The process is that, we first get the standard deviation of each time series, and then average across stocks to find the mean standard deviation \( \sigma_{AFD} \). After dividing each \( AFD_{i\tau} \) by \( \sigma_{AFD} \), \( \frac{1}{n} \sum_{i=1}^{n} \sigma_i = 1 \), where \( \sigma_i \) is the standard deviation of stock \( i \) over the full sample. Recall that the standard deviation of \( X_{\tau} \) is 1, then this normalization makes sure that \( X_{\tau} \) has similar magnitude of time variation as the average \( AFD_{i\tau} \). Then we get Relative \( X_{i\tau} \) and do the same normalization so that its time series standard deviation averaged across stocks is 1.

The reason for this subtraction is that analysts’ dispersion contain both disagreement and information asymmetry. The difference constructed above will be higher if asymmetric information is low, and lower when asymmetric information is high, and at the same time increases with the disagreement measure, which is quantitatively similar to the relative disagreement measure. Admittedly, another seemingly more natural normalization is to take ratios, but we do not employ it because analysts’ dispersion can be 0 or close to 0 for many stocks in a certain period, and the ratio will explode, driving the regression results invalid.
Table 3.5: Volume-volatility elasticity estimates for Dow Jones 30 components

<table>
<thead>
<tr>
<th>Full Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
</tr>
<tr>
<td>(0.018)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
</tr>
<tr>
<td>(0.003)</td>
</tr>
<tr>
<td>NVIX</td>
</tr>
<tr>
<td>(0.003)</td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
</tr>
<tr>
<td>(0.004)</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>high-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
</tr>
<tr>
<td>(0.022)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
</tr>
<tr>
<td>(0.009)</td>
</tr>
<tr>
<td>NVIX</td>
</tr>
<tr>
<td>(0.008)</td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
</tr>
<tr>
<td>(0.006)</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
</tr>
<tr>
<td>(0.027)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
</tr>
<tr>
<td>(0.003)</td>
</tr>
<tr>
<td>NVIX</td>
</tr>
<tr>
<td>(0.003)</td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
</tr>
<tr>
<td>(0.005)</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difference between high and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
</tr>
<tr>
<td>(0.034)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
</tr>
<tr>
<td>(0.009)</td>
</tr>
<tr>
<td>NVIX</td>
</tr>
<tr>
<td>(0.009)</td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
</tr>
<tr>
<td>(0.007)</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.13) for the specification $\Delta \log(m_{i\tau}) = c_i + a_i^0 D_{\tau} + (b_0 + a_i^1 D_{\tau} + b_1^X X_{i\tau}) \cdot \Delta \log(\sigma_{i\tau}) + \epsilon_{i\tau}$. In all specifications, $D_{\tau}$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_i^1$) are not reported for brevity. The UNEMP Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasts before the announcement. NVIX is the news implied volatility measure from Manela and Moreira (2017). Both these disagreement measure have been subtracted by normalized analysts’ dispersion, as in equation (3.16). Idio-Sys Ratio is the ratio of idiosyncratic variance to systematic variance constructed as in equation (3.3). All explanatory variables have been normalized to have average standard deviation 1. The sample spans April 10, 2001 to December 30, 2014. The four panels from top to bottom present regression results for the full sample, the high-sentiment subsample, the low-sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
Table 3.5 displays the results for regression (3.13). The elasticity of individual stocks is much lower than that of the market, indicating higher dispersions of investors regarding the valuation of individual stocks. NVIX is significantly negative, as predicted by the model. Its magnitude is much smaller than that of SPY, consistent with the fact that systematic term only composes a part of stocks’ pricing. Unemployment is negative but insignificant. Considering the multicollinearity by construction since analysts’ dispersion is subtracted from both of the disagreement measures, we do not put more than one disagreement measure in the model.

Next we turn to the Idio-Sys Ratio, and find it significantly decreases the elasticity. The magnitude is also much larger than that of the disagreement measures. The interpretation is that a similar unit of standard deviation in Idio-Sys Ratio will cast a much more negative effect than the disagreement measures. In our model, stocks with a high ratio of the idiosyncratic component relative to the systematic component will have lower elasticity. Intuitively, when the idiosyncratic part of a stock is high, it will be more difficult for investors to reach a consensus of the firms’ value due to this hard-to-estimate idiosyncratic component. We also notice that even when we put in both the relative disagreement proxy and Idio-Sys Ratio, the elasticity is still far below 1, indicating that further improvement on this model may be needed to account for disagreements on pricing individual stocks.

The sharp contrast between the coefficients of disagreement measures in panel 2 and panel 3 resembles that of SPY. When sentiment is high, all disagreement measures are significantly negative, with magnitudes several times of those for the full sample. On the contrary, the coefficient of the unemployment dispersion is not significant in the low-sentiment period, while the coefficient of NVIX is much smaller in terms
of magnitude. In high sentiment period both relative disagreement proxies increase adjusted $R^2$ while neither of them does so in low-sentiment periods. Also, elasticity increases by around 0.06 if we control for the effect of one relative disagreement proxy while the value is less than 0.02 in low-sentiment periods. This disparity in results is consistent with the model implication: during low-sentiment periods the increased uncertainty lowers investors’ confidence in interpreting public signals, and hence the effect of disagreement measures wanes. Interestingly, this decrease in explanatory power is made up for by Idio-Sys Ratio, and both elasticities increase by around 0.1 when both the disagreement measures and the Idio-Sys Ratio are included. The result that Idio-Sys Ratio plays a more important role is consistent with the model implication that when confidence in interpreting the public signal is low, the idiosyncratic disagreement component is the only term remained. All these differences are statistically significant, as manifest in panel 4.

### 3.5.4 Treasury Bond Futures

Finally we turn to the results for the 10-year Treasury bond futures as reported in Table 3.6. Consider the effect of CPI on Treasury bonds, we also include the one-year ahead forecast dispersion on CPI. Over the full sample, the regression results are very similar to those for SPY and individual stocks: all disagreement measures are significantly negative, and NVIX seems to play a dominating role again. The elasticity will be close to 1 when we have controlled for the disagreement measures. However, when divided into two subsamples, the results are different from those of SPY and Dow Jones 30 components. Most regressors are significant in both regimes,
Table 3.6: Volume-volatility elasticity estimates for T-bond futures

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th></th>
<th>High-sentiment periods</th>
<th>Low-sentiment periods</th>
<th>Difference between high and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.524**</td>
<td>0.760**</td>
<td>0.793**</td>
<td>0.998**</td>
<td>0.996**</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.095)</td>
<td>(0.091)</td>
<td>(0.085)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.049**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI Dispersion</td>
<td>-0.066**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.123**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.640</td>
<td>0.641</td>
<td>0.642</td>
<td>0.650</td>
<td>0.650</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline estimates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.330**</td>
<td>0.698**</td>
<td>0.559**</td>
<td>0.800**</td>
<td>0.822**</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.127)</td>
<td>(0.136)</td>
<td>(0.123)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.065**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI Dispersion</td>
<td>-0.049*</td>
<td></td>
<td>-0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.100**</td>
<td></td>
<td>-0.098**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.656</td>
<td>0.659</td>
<td>0.656</td>
<td>0.663</td>
<td>0.662</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline estimates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.841**</td>
<td>0.652**</td>
<td>1.092**</td>
<td>1.267**</td>
<td>1.319**</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.168)</td>
<td>(0.152)</td>
<td>(0.121)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI Dispersion</td>
<td>-0.077**</td>
<td></td>
<td>-0.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.028)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.165**</td>
<td></td>
<td>-0.156**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.033)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.631</td>
<td>0.631</td>
<td>0.634</td>
<td>0.643</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log (p_{mT}) = c + a_0 J + (b_0 + a_1 J) D_T + (b_1 + b_2 J) X_T \cdot \Delta \log (\sigma_T) + \epsilon_T$. In all specifications, $D_T$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1$) are not reported for brevity. The UNEMP Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. The CPI Dispersion is the forecast dispersion of the one-year-ahead CPI from the Survey of Professional Forecasters before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). NVIX is the news implied volatility measure from Manela and Moreira (2017). All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The four panels from top to bottom present regression results for the full sample, the high-sentiment subsample, the low-sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
with similar magnitudes. The only exception is unemployment dispersion, which turns insignificantly positive in the low-sentiment period. This may imply that dispersion in CPI acts as a better disagreement proxy for Treasury bond futures. Except for the unemployment dispersion, the differences are insignificant between the two regimes according to the last panel. In both periods, adjusted $R^2$ increases with add-ins of explanatory variables individually, and the p values of F test are smaller than 0.01 once NVIX is added into the regression, in both periods. Besides, we see large increases in elasticity when any of the proxies have been taken control of.

The most obvious difference between the two sentiment periods is the baseline estimate of elasticity. In high-sentiment periods, the log volume jump absent of volatility jump is significantly higher. However, the elasticity is lower during high-sentiment period, even after the disagreement proxies are already controlled. Therefore, though dispersion proxies play similar roles, there are still some differences that are not captured by the difference-of-opinion model around macroeconomic announcements when sentiment regime differs.

What we can conclude from Table 3.6 is that Treasury bond futures do not demonstrate a similar disparity in different sentiment regimes. As a “risk-free” asset, Treasury bill should be the least affected by sentiment, so it does not display the distinct pattern across subsamples as SPY does. Actually, with the “flight to quality” and “flight to liquidity” stories, investors may not really lose confidence when they hold Treasury bonds during low-sentiment periods. Besides, the sentiment index used in this paper is the first component of principal component analysis of five proxies in stock and mutual fund market, which may not accurately measure the sentiment in the Treasury bond market. Besides, there is even no well-defined sentiment measure
in the Treasury bond market considering the rarity of speculative activities in the Treasury bonds. Therefore, in our paper, the Treasury bond futures just act as a benchmark, which by comparison, highlights the sharp effect sentiment has on stock market.

3.6 Robustness Test

In this section, we test for robustness by redefining sentiment periods and using different explanatory variables.

3.6.1 Different cutoff points for sentiment periods

To make sure that the disparity between high and low sentiment period is not a coincidence from sample classification, we redivide the sample and redo the analysis for the S&P 500 ETF and Dow Jones 30 components. Instead of defining the high and low sentiment periods based on sample median, we use the 25% and 75% percent quantiles as the cutoff points. That means, a month is classified into high (low) sentiment period if the previous month’s sentiment is no less (no more) than the 75% (25%) sample quantile.

Figure 3.8 shows the plot of the sentiment measure along with daily annualized realized volatility and daily volume with high (low) sentiment periods shaded in red (blue). The results are shown in Table 3.7, Table 3.8 and Table 3.9.

The patterns are exactly the same for the S&P 500 ETF, though some of the differences between the two sentiment periods become insignificant around FOMC announcements, which can be explained by the reduced sample size and thus decreased power. For the individual stocks, when we use the 75% and 25% percent
Figure 3.8: The Sentiment Measure

Notes: The figure plots the monthly sentiment measure, the daily annualized realized volatility of S&P 500 ETF, as well as its daily volume. High sentiment periods are shaded in red and low sentiment periods are shaded in blue.

quantiles as thresholds, the disparity becomes even more obvious. All disagreement measures are significantly negative during high sentiment periods while lose significance when sentiment level is low. These results reinforce our argument that investors’ disagreement plays a much more important role during high sentiment period.

3.6.2 Other disagreement measures

We rerun the regressions using RGDP growth rate dispersion and Monthly Index measure instead of Unemployment rate dispersion and NVIX. Also, the one year ahead CPI dispersion is replaced by the ten year ahead CPI dispersion. To save space, we only keep the results of interest: rows that manifest the differences between the high and low sentiment periods.
Table 3.7: Volume-volatility elasticity estimates for S&P 500 ETF (different cutoff points for sentiment periods)

<table>
<thead>
<tr>
<th></th>
<th>high-sentiment periods</th>
<th>low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.634**</td>
<td>0.687**</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.148)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.129*</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.160**</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>NVIX</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.467</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td>0.473</td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>0.476</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.479</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.477</td>
<td>0.476</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>low-sentiment periods</th>
<th>high-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.966**</td>
<td>0.817**</td>
</tr>
<tr>
<td></td>
<td>(0.184)</td>
<td>(0.198)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.114*</td>
<td>-0.039</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>0.004</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>NVIX</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.477</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.478</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.477</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.477</td>
<td>0.476</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>low-sentiment periods</th>
<th>high-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.076</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.164*</td>
<td>-0.164*</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>NVIX</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.477</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.478</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.477</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>0.477</td>
<td>0.476</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log (r_t) = c + a_0 D_{0_t} + b_0 + a_1 D_{1_t} + b_1 X_t$, $\Delta \log (\sigma_t) + \epsilon_t$. In all specifications, $D_{0_t}$ and $D_{1_t}$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1$) are not reported for brevity. The UNEMP Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). NVIX is the news implied volatility measure from Manela and Moreira (2017). All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The three panels from top to bottom present regression results for the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. If the previous month’s sentiment level is above 75% (below 25%) of the sample sentiment level, this month is classified as high (low) sentiment period. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.

The results are shown in Table 3.10, Table 3.11, Table 3.12 and Table 3.13. For the individual stocks and the S&P 500 ETF, we can still see the disparity between high and low sentiment when RGDP growth rate dispersion and MonthlyIndex are used. The difference between the coefficients in high and low sentiment periods for RGDP growth rate dispersion becomes insignificant for the market ETF though. For the Treasury bond futures, we still do not see this disparity, though there are some
Table 3.8: Volume-volatility elasticity estimates for S&P 500 ETF around FOMC announcements (different cutoff points for sentiment periods)

<table>
<thead>
<tr>
<th></th>
<th>high-sentiment periods</th>
<th>low-sentiment periods</th>
<th>Difference between high- and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elasticity ($b_0$)</strong></td>
<td>0.634** 1.152** 1.551** 1.260** 0.874** 1.241**</td>
<td>0.687** 0.883** 0.865** 0.954** 0.764** 0.946**</td>
<td>0.053 0.217 0.395 0.186 0.190 0.294</td>
</tr>
<tr>
<td><strong>Estimates for explanatory variables in elasticity ($b_1$):</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.202** (0.063)</td>
<td>-0.047 (0.035)</td>
<td>-0.155* (0.073)</td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.178** (0.066)</td>
<td>-0.037 (0.037)</td>
<td>-0.141 (0.078)</td>
</tr>
<tr>
<td>Negative</td>
<td>-0.242* (0.103)</td>
<td>-0.064 (0.067)</td>
<td>-0.178 (0.124)</td>
</tr>
<tr>
<td><strong>$R^2$</strong></td>
<td>0.209 0.290 0.244 0.270 0.234 0.242</td>
<td>0.208 0.191 0.185 0.158 0.185 0.121</td>
<td>0.001 0.109 0.057 0.122 0.049 0.113</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log (\sigma) = c + (b_0 + b_1 X) \cdot \Delta \log (\sigma) + \varepsilon_t$. In all specifications, the UNEMP Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. NVIX is the news implied volatility measure from Manela and Moreira (2017). FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The three panels from top to bottom present regression results for the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. If the previous month’s sentiment level is above 75% (below 25%) of the sample sentiment level, this month is classified as high (low) sentiment period. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.

noticeable differences in the coefficients. These results show that, when we use other disagreement measures in our analysis, our conclusion remains essentially unchanged.

3.7 Conclusion

This paper provides evidences that sentiment can affect information processing in the financial market. During high-sentiment periods, investors are confident in their interpretations of the common public signal, while their confidence level is low during low
sentiment periods. As a result, dispersion in these interpretations plays a more important role in the volume-volatility relationship during high-sentiment periods, both for the market ETF and individual stocks. Empirically, we estimate the effect of disagreement measures on volume-volatility elasticity in high- and low-sentiment periods. We employ econometric tools including nonparametric estimation of volume intensity and spot volatility, regression discontinuity design and difference-in-difference. We conduct bootstraps for statistical inferences.

To derive the volume-volatility relationship of individual stocks around systematic news release, we incorporate a one factor structure with heterogeneous beliefs into the Kandel and Pearson (1995) differences-of-opinion model. The model predicts that individual stocks’ elasticity decreases with investor disagreement on the common public signals and the ratio of idiosyncratic volatility. Besides, during periods when investors’ confidence in interpreting the signals is low, the disagreement measures no longer have explanatory power.

Our empirical results are consistent with the model predictions. Forecast dispersion on real economy activities and uncertainty measures significantly decrease estimates of elasticity, so does the ratio of idiosyncratic volatility to systematic volatility. When we separate the sample into high- and low-sentiment periods based on the sentiment index, we find that the disagreement measures significantly reduces volume-volatility elasticity only in high-sentiment periods, both for the S&P 500 ETF and Dow Jones 30 components. The coefficient for idiosyncratic ratio is negative in both periods, as the theoretical model predicts. For Treasury bond futures, there is no disparity between high- and low-sentiment periods. The robustness tests use different
disagreement measures and new cutoff points for classifying sentiment periods, and
the results remain essentially the same.
Table 3.9: Volume-volatility elasticity estimates around for Dow Jones 30 components (different cutoff points for sentiment periods)

<table>
<thead>
<tr>
<th></th>
<th>high-sentiment periods</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.442**</td>
<td>0.521**</td>
<td>0.515**</td>
<td>0.481**</td>
<td>0.538**</td>
<td>0.540**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.029)</td>
<td>(0.035)</td>
<td>(0.037)</td>
<td>(0.036)</td>
<td>(0.040)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
<td></td>
<td>0.615**</td>
<td>0.615**</td>
<td>0.619**</td>
<td>0.734**</td>
<td>0.733**</td>
<td>0.739**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.303</td>
<td>0.305</td>
<td>0.304</td>
<td>0.304</td>
<td>0.305</td>
<td>0.304</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.048**</td>
<td>-0.045**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVIX</td>
<td>-0.040**</td>
<td></td>
<td>-0.038**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
<td></td>
<td>-0.013</td>
<td>-0.007</td>
<td>-0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>low-sentiment periods</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.615**</td>
<td>0.615**</td>
<td>0.619**</td>
<td>0.734**</td>
<td>0.733**</td>
<td>0.739**</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.308</td>
<td>0.308</td>
<td>0.308</td>
<td>0.312</td>
<td>0.311</td>
<td>0.312</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UNEMP Dispersion</td>
<td>-0.002</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVIX</td>
<td></td>
<td>-0.003</td>
<td>-0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
<td>-0.059**</td>
<td>-0.059**</td>
<td>-0.059**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Difference between high- and low-sentiment periods</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td></td>
<td>0.404</td>
<td>0.404</td>
<td>0.410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.318</td>
<td>0.318</td>
<td>0.322</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.13) for the specification

\[
\Delta \log(m_{p\tau}) = c_i + a_0 \cdot D_{\tau} + (b_0 + a_1 \cdot D_{\tau} + b_1 \cdot X_{\tau}) \cdot \Delta \log(\sigma_{p\tau}) + \epsilon_{p\tau}.
\]

In all specifications, $D_{\tau}$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1$) are not reported for brevity. The UNEMP Dispersion is constructed as the previous quarter's forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. NVIX is the news implied volatility measure from Manela and Moreira (2017). Both these disagreement measures have been subtracted by normalized analysts’ dispersion, as in equation (3.16). Idio-Sys Ratio is the ratio of idiosyncratic variance to systematic variance constructed as in equation (3.3). All explanatory variables have been normalized to have average standard deviation 1. The sample spans April 10, 2001 to December 30, 2014. The three panels from top to bottom present regression results for the full sample, the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. If the previous month’s sentiment level is above 75% (below 25%) of the sample sentiment level, this month is classified as high (low) sentiment period. Bootstrapped standard errors are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
### Table 3.10: Volume-volatility elasticity estimates for S&P 500 ETF

<table>
<thead>
<tr>
<th></th>
<th>High-sentiment periods</th>
<th>Low-sentiment periods</th>
<th>Difference between high- and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elasticity ($b_0$)</strong></td>
<td>0.696** (0.098)</td>
<td>0.757** (0.091)</td>
<td>0.061** (0.031)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.012 (0.034)</td>
<td>-0.023 (0.020)</td>
<td>0.011 (0.040)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.146** (0.033)</td>
<td>-0.191** (0.042)</td>
<td>-0.045** (0.039)</td>
</tr>
<tr>
<td>NVIX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.551</td>
<td>0.558</td>
<td>0.557</td>
</tr>
</tbody>
</table>

**Notes:** The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log (p_{t+1}) = c + a_0 D_T + (b_0 + a_1^T D_T + b_1^T X_T) \cdot \Delta \log (\sigma_T) + \epsilon_t$ In all specifications, $D_T$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1^T$) are not reported for brevity. The RGDP Growth Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead real GDP growth rate from the Survey of Professional Forecasters before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). NVIX is the news implied volatility measure from Manela and Moreira (2017). All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The three panels from top to bottom present regression results for the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
Table 3.11: Volume-volatility elasticity estimates for S&P 500 ETF around FOMC announcements

<table>
<thead>
<tr>
<th></th>
<th>high-sentiment periods</th>
<th>low-sentiment periods</th>
<th>Difference between high- and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline estimates:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.696**</td>
<td>1.048**</td>
<td>0.757**</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.171)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.101*</td>
<td>-0.069</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.044)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.234**</td>
<td>-0.213**</td>
<td>-0.276**</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.054)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Negative</td>
<td>-0.149**</td>
<td>-0.150**</td>
<td>-0.146**</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.041)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.312</td>
<td>0.330</td>
<td>0.364</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.118*</td>
<td>-0.179</td>
<td>-0.064</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.154)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.213**</td>
<td>-0.030</td>
<td>-0.268**</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.054)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>Negative</td>
<td>-0.112*</td>
<td>-0.038</td>
<td>-0.112*</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.033)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.378</td>
<td>0.378</td>
<td>0.359</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes:</td>
<td>The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log(m_{T}) = c + (b_0 + b_1' X_T') \Delta \log(\sigma_T) + \epsilon_t$. In all specifications, the RGDP Growth Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead real GDP growth rate from the Survey of Professional Forecasters before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. All variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to December 30, 2014. The three panels from top to bottom present regression results for the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.12: Volume-volatility elasticity estimates for Dow Jones 30 components

<table>
<thead>
<tr>
<th></th>
<th>high-sentiment periods</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Elasticity ($b_0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.458**</td>
<td>0.557**</td>
<td>0.509**</td>
<td>0.502**</td>
<td>0.583**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.024)</td>
<td>(0.029)</td>
<td>(0.027)</td>
<td>(0.029)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.045**</td>
<td></td>
<td></td>
<td>-0.042**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
<td></td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.041**</td>
<td></td>
<td></td>
<td>-0.040**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
<td>-0.015**</td>
<td>-0.011</td>
<td>-0.014*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.298</td>
<td>0.300</td>
<td>0.299</td>
<td>0.299</td>
<td>0.300</td>
<td>0.300</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>low-sentiment periods</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Elasticity ($b_0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.617**</td>
<td>0.637**</td>
<td>0.633**</td>
<td>0.684**</td>
<td>0.703**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.027)</td>
<td>(0.027)</td>
<td>(0.027)</td>
<td>(0.028)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.009**</td>
<td></td>
<td></td>
<td>-0.009**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.009**</td>
<td></td>
<td></td>
<td>-0.008**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
<td>-0.032**</td>
<td>-0.032**</td>
<td>-0.032**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.303</td>
<td>0.303</td>
<td>0.303</td>
<td>0.304</td>
<td>0.305</td>
<td>0.305</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Difference between high- and low-sentiment periods</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.036**</td>
<td></td>
<td></td>
<td>-0.033**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.033**</td>
<td></td>
<td></td>
<td>-0.032**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idio-Sys Ratio</td>
<td>0.017*</td>
<td>0.021**</td>
<td>0.018*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.13) for the specification $\Delta \log (p_{t, i}) = c_i + a_0' \mathbf{D}_r + (b_0 + a_1' \mathbf{D}_r + b_1' \mathbf{X}_{r,t}) \cdot \Delta \log (\sigma_{r,t}) + \epsilon_{r,t}$. In all specifications, $\mathbf{D}_r$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1$) are not reported for brevity. The RGDP Growth Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead real GDP growth rate from the Survey of Professional Forecasters before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). Both these disagreement measure have been subtracted by normalized analysts’ dispersion, as in equation (3.16). Idio-Sys Ratio is the ratio of idiosyncratic variance to systematic variance constructed as in equation (3.3). All explanatory variables have been normalized to have average standard deviation 1. The sample spans April 10, 2001 to December 30, 2014. The three panels from top to bottom present regression results for the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
Table 3.13: Volume-volatility elasticity estimates for T-bond futures

<table>
<thead>
<tr>
<th></th>
<th>high-sentiment periods</th>
<th>low-sentiment periods</th>
<th>Difference between high- and low-sentiment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity ($b_0$)</td>
<td>0.330** (0.072)</td>
<td>0.841** (0.107)</td>
<td>0.511* (0.041)</td>
</tr>
<tr>
<td></td>
<td>0.474** (0.100)</td>
<td>1.060** (0.118)</td>
<td>0.586* (0.044)</td>
</tr>
<tr>
<td></td>
<td>0.376** (0.136)</td>
<td>1.010** (0.116)</td>
<td>0.634* (0.045)</td>
</tr>
<tr>
<td></td>
<td>0.748** (0.141)</td>
<td>0.975** (0.117)</td>
<td>0.723* (0.045)</td>
</tr>
<tr>
<td></td>
<td>0.748** (0.140)</td>
<td>1.061** (0.120)</td>
<td>0.313* (0.045)</td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGDP Growth Dispersion</td>
<td>-0.041* (0.018)</td>
<td>-0.129** (0.036)</td>
<td>0.088* (0.041)</td>
</tr>
<tr>
<td>CPI Dispersion</td>
<td>-0.010 (0.023)</td>
<td>-0.072** (0.024)</td>
<td>0.061 (0.034)</td>
</tr>
<tr>
<td>MonthlyIndex</td>
<td>-0.106** (0.027)</td>
<td>-0.055** (0.020)</td>
<td>-0.050 (0.033)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.104* (0.043)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.656</td>
<td>0.656</td>
<td>0.655</td>
</tr>
<tr>
<td></td>
<td>0.656</td>
<td>0.655</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>0.655</td>
<td>0.655</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>0.660</td>
<td>0.655</td>
<td>0.659</td>
</tr>
<tr>
<td></td>
<td>0.659</td>
<td>0.659</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification $\Delta \log (m_{\tau}) = c + a_0 D_{\tau} + (b_0 + a_1 D_{\tau} + b_1 X_{\tau} ) \cdot \Delta \log (\sigma_{\tau}) + \epsilon_{\tau}$. In all specifications, $D_{\tau}$ are category dummy variables for ISM Manufacturing Index, ISM Non-Manufacturing Index, Consumer Confidence Index and the Housing Index. The coefficients of these dummies (i.e., $a_0$ and $a_1$) are not reported for brevity. The RGDP Growth Dispersion is constructed as the previous quarter’s forecast dispersion of the one-year-ahead real GDP growth rate from the Survey of Professional Forecasers before the announcement. The CPI Dispersion is the forecast dispersion of the ten-year-ahead CPI from the Survey of Professional Forecasers before the announcement. The MonthlyIndex variable is the Monthly Index from Baker et al. (2016). NVIX is the news implied volatility measure from Manela and Moreira (2017). All variables are scaled by their own sample standard deviations. The sample spans April 10, 2003 to December 30, 2014. The three panels from top to bottom present regression results for the high sentiment subsample, the low sentiment subsample, and the differences in the coefficients between the high and low subsamples. Bootstrapped standard errors are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.
4

Efficient Estimation of Integrated Functional of Variance with Irregular Observation Time

Summary: In this chapter we propose an efficient estimator for the functional of variance with irregular observation time. The estimator is an Riemann sum of the functional of the spot volatility, which achieves the efficiency bound defined in Renault et al. (2017). We validate the consistency and the central limit theorem and conduct simulations.

4.1 Introduction

In the literature of finance, most attention has been focused on liquid stocks. For example, when researchers conduct research on investment strategy and market microstructure, they usually use Dow Jones 30 components, S&P 100 components or S&P 500 components. However, in real life illiquid stocks are also an important com-
ponent in the financial market. For market makers, the bid-ask spreads for illiquid traded stocks are larger than the liquid stocks and thus constitute a large proportion of revenues in market making, even though the trading is much less frequent. Besides, hedge funds tend to pick less liquid stocks in their portfolios. It is therefore of importance to take account of illiquid stocks in finance research.

One important research in topic is the estimation of integrated volatility. In the literature of high frequency econometrics, regular sampling is widely employed. The practice is that for every n minutes \(^1\), the last price before time \(t\) is sampled as the price at time \(t\), denoted as \(p_t\). For frequently traded stocks, even though there can be slight lag between the sampled price and the exact price, the difference should be small considering that there are thousands of transactions within one minute. However, for less liquid stocks, the prices can be stale for a long time.

Figure 4.1 shows the empirical CDF of number of trades for four less liquid stocks. We extract all stocks in Compustat and rank the number of transactions on Jan 31st, 2014 from the largest to the smallest. The four stocks are ranked 500th (CTAS), 1000th (ONCY), 1500th (USCR), and 2000th (ULBI). The trading data are from NYSE TAQ. The slopes of the line can be seen as the intensity of trading. The plot shows that when the number of trades decreases, we see stage-like lines which means transactions are scattered within long periods of inactivity. Consistent with Andersen and Bollerslev (1998), the trading intensity is larger at the beginning and end of the day while smaller in the mid for most stocks.

This intraday pattern should be taken into consideration and controlled for in many settings (see Bollerslev et al. (2018) for example). Another topic that is fre-

\(^1\) Some widely seen examples of \(n\) are 1 and 5
Figure 4.1: The empirical CDF of number of trades on Jan 31st, 2014

Notes: The figure plots the empirical CDF of number of trades on Jan 31st, 2014 for CTAS, ONCY, USCR and ULBI.

Consequently studied in high frequency econometrics and will be affected is the smooth integral of volatility:

$$V(g)_t = \int_0^t g(c_s) \, ds.$$  

There have been growing research on this topic with irregular sampling times. A few example include Aït-Sahalia et al. (2010), Barndorff-Nielsen et al. (2011), Zhang (2011), Li et al. (2014) and Li and Xiu (2017). The sampling times are exogenous in some papers while endogenous in others. In Jacod et al. (2017), the authors propose estimator of market microstructure noise under irregular sampling. In Renault et al. (2017), though, they argue that many existing estimators are not even consistent if the sampling times are irregular and volatility are time varying. They propose the
nonparametric bounds for inference about functionals of volatility. However, they only propose a "nearly efficient" estimator.

In this chapter, inspired by Jacod et al. (2013), we propose an efficient estimator for the smooth functional of variance, under an irregular, random, but strongly predictive sampling. The stock price is an semimartingale process with jumps, while we do not include market microstructure noise in our setting. We first estimate the spot volatility which employs the interval length, and then estimate the integral by the means of Riemann sum. We propose the consistency and central limit theorem, and run simulations to support our theorems.

This paper precedes as follows. Section 4.2 presents the assumptions needed for the theorems. Section 4.3 proposes the estimator, and the asymptotics are discussed in Section 4.4. Section 4.5 shows the simulation results. Section 4.6 concludes. The proofs of are given in Appendix C.

4.2 Set up

We first propose several assumptions on the stock price process and the observation times.

4.2.1 The stock price process

Assume that the price of a stock follows the following semi-martingale process:

\[ X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_0^t \delta(s, z) 1_{[|\delta(s,z)| \leq 1]} (u - v) (ds, dz) \]

\[ + \int_0^t \int_E \delta (s,z) 1_{[|\delta(s,z)| > 1]} \mu (ds, dz) . \]
Assumption (H-r). There are a sequence \((J_n)\) of nonnegative bounded \(\lambda\)-integrable functions on \(E\) and a sequence \((\tau_n)\) of stopping times increasing to \(\infty\), such that

\[
\begin{align*}
t < \tau_n (\omega) & \implies \|b_t (\omega)\| \leq n, \quad \|\sigma_t (\omega)\| \leq n \\
t \leq \tau_n (\omega) & \implies \|\delta (\omega, t, z)\|^r \wedge 1 \leq J_n (z)
\end{align*}
\]

Suppose that \(c_t\) is again an Itô semimartingale, and we consider the following assumption:

Assumption (A-r). The process \(X\) satisfies Assumption (H-r), the associated volatility process \(c\) satisfies (H-2). The process \(b_t\) satisfies that when \(r \leq 1\),

\[
\begin{align*}
b_t' &= b_t - \int \delta (t, z) 1_{\{|\delta (t, z)| \leq 1\}} \lambda (dz) \text{ are càdlàg or càdlàg.}
\end{align*}
\]

These assumptions are very standard in the literature of high frequency econometrics.

4.2.2 The observation times

Follow the notation of Jacod and Protter (2011) (Page 430), we denote a restricted discretization scheme as \(T (n, i)\). A random discretization scheme is a double sequence \((T (n, i) : i \geq 0, n \geq 1)\) of finite stopping times, increasing in \(i\), with \(T (n, 0) = 0\) and

\[
\begin{align*}
\forall t > 0 : \pi^n_t := \sup_{i \geq 1} (T(n, i) \wedge t - T(n, i - 1) \wedge t) \quad &\xrightarrow{P} 0, \quad (4.1) \\
\Delta (n, i) &= T(n, i) - T(n, i - 1), \quad N_n (t) = \sum_{i \geq 1} 1_{\{T(n, i) \leq t\}}. \quad (4.2)
\end{align*}
\]

Denote \(\Delta_{n,i}\) as the interval between the two observation time, and \(N_n (t)\) the number of observations before \(t\). Then

\[
T_n (t) = T(n, N_n (t)), \quad I(n, i) = (T(n, i - 1), T(n, i)]. \quad (4.3)
\]

120
Now we assume there exists a stochastic process $T'(\cdot) : [0, t] \to \mathbb{R}$ s.t.

$$\frac{i}{n} = \int_0^{t_{n,i}} T'(s) ds.$$ 

$T'(t)$ can be seen as the observation intensity. It takes time $t_{n,i}$ to observe the proportion $\frac{i}{n}$ of all observations. The larger $T'(\cdot)$ is, the observations arrive at higher intensity. Define $\Delta_n = \frac{i}{n} = \frac{i}{n} - \frac{i-1}{n}$, and

$$\Delta_n = \int_{t_{n,i-1}}^{t_{n,i}} T'(s) ds.$$ 

$T(\cdot)$ is correspondingly defined as the CDF of $T'(\cdot)$.

**Assumption (T). Assumption on the sampling time**

**(T-1)** The sampling scheme $(t_{i,n})_{0 \leq i \leq N_n: n \geq 1}$ is $\{\mathcal{F}_t\}_{0 \leq t \leq 1}$-predictable; that is $(t_{j,n})_{i \leq j \leq k_n}$ is $\mathcal{F}_{t_{i-1,n}}$ measurable, for all $i, n$ and $k_n$ s.t. $k_n \Delta_n \to 0$.

**(T-2)** $T'(\cdot) \geq c$ for some $c > 0$.

**(T-3)** $T'(\cdot)$ is continuous.

**Remark:**

Assumption (T-1) is similar to the assumption 2 in Renault et al. (2017), and is a stronger assumption than the Assumption (C) in Hayashi et al. (2011), which only requires Conditional on $\mathcal{F}_{t_{n,i-1}}$, $t_{n,i}$ is independent of the price and volatility process. Their assumption ensures strong predictability of $(t_{j,n})_{i \leq j \leq k_n}$ conditional on $\mathcal{F}_{t_{i-1,n}}$. Assumption (T-2) states that the observation intensity cannot be 0, which is reasonable since all stocks are publicly traded during trading hours. Assumption (T-3)
amounts to saying that $T'(\cdot)$ is uniformly continuous, and so is $T'(\cdot)^{1-1}$. Admittedly all of them are strong assumptions but they simplify the proof a lot, and we may consider relax these restrictions in future research.

With assumptions (T-1) to (T-3) we have the following conclusions: When $T(\cdot)$ is defined above and satisfy the assumptions, we have the following conditions:

(C-1) $c\Delta_n \leq \Delta_{n,i} \leq \bar{c}\Delta_n$ where $c > 0$.

\[
(C-2) \left| \frac{\sum_{j=0}^{k_n-1} \Delta_{n,i-j}}{k_n\Delta_n} - \frac{\Delta_{n,i}}{\Delta_n} \right| \leq \delta^n_T \rightarrow 0 \text{ where } \delta^n_T = \sup_{|t_1-t_2| < c\Delta_n} \left| T(t_1)^{1-1} - T(t_2)^{1-1} \right|.
\]

Condition (C-1) states that the observed sampling interval is bounded (from both above and below) by a constant times $\Delta_n$. Condition (C-2) says that for the successive $k_n$ sample intervals, the average interval lengths are close to each other.

Figure 4.2 shows an illustration of the $T(\cdot)$ process, as the smoothed CDF of the trading intensity.

4.3 Estimator

In this section, we propose the estimator. Firstly we estimate the spot volatility, taking account of the sampling interval length, then we use Riemann sum to get the integral of the functional of the spot volatility.

The estimator of the spot volatility is

\[
\hat{c}_T^n(k_n, v) = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \frac{(\Delta_{n,i+m} X)^2}{\Delta_{n,i+m}} \frac{1(\|\Delta_{n,i+m} X\| \leq v_{n,i+m})}{(4.4)}.
\]
The empirical CDF of and smoothed empirical CDF of the number of trades of USCR on Jan 31st, 2014.

Notes: The figure plots the empirical CDF of and smoothed empirical CDF of the number of trades of USCR on Jan 31st, 2014. It is an illustration of the $T(\cdot)$ process.

where $v_{n,i} = \alpha \Delta_{n,i}^\varpi$ with $0 < \varpi < \frac{1}{2}$. This is similar to the spot volatility in Jacod et al. (2013). Approximately,

$$\Delta_n^\varpi X^c | \mathcal{F}_{n,i-1} \sim N(0, c_i \Delta_{n,i}).$$

Intuitively, returns that span a longer interval also have higher variance, so we scale the squared return using the interval length to get a “uniform” estimator of the magnitude of the local volatility. Choose an undersmoothed $k_n$ such that as $n \to \infty$,

$$k_n^3 \Delta_n \to \infty, k_n^2 \Delta_n \to 0.$$

Then the ”Statistical error” is of order $1/\sqrt{k_n}$. Since $k_n \Delta_n$ is asymptotically 0, the intuition is that the underlying variance over the interval $[t, t + k_n \Delta_n]$ stays almost constant. Then an estimator of the integrated smooth functional of the volatility is the
Riemann sum of the functional of the estimated spot volatility. With the estimator of spot volatility defined as in Equation (4.4). To estimate \( V_t = \int_0^t g(c_s) \, ds \) the following estimators is put forward:

\[
V^n(g, k_n, v_n, X)_t = \sum_{i=1}^{[\Delta n] - k_n + 1} \Delta(n, i) g(c^n_i(k_n, v_n)).
\] (4.5)

4.4 Asymptotics

In this section we claim the consistency and asymptotic normality of the estimator, and show that the estimator reaches the efficiency bound.

4.4.1 Consistency

**Theorem 4.4.1.** Assume that \( X \) satisfies (H) and \( T \) satisfies (T), and let \( g \) be a continuous function on \( \mathbb{R}^+ \), satisfying for some \( p \geq 0 \)

\[
|g(x)| \leq K(1 + \|x\|^p)
\]

1. If either \( X \) is continuous or \( p < 1 \), we have

\[
V^n(g, k_n, X)_t \xrightarrow{u.c.p.} \int_0^t g(c_s) \, ds
\] (4.6)

2. If either \( p \leq 1 \), or \( (H-r) \) holds for some \( r \in [0, 2) \) and

\[
p > 1, \quad \omega \geq \frac{p - 1}{2p - 4}
\] (4.7)

We have

\[
V^n(g, k_n, v_n, X)_t \xrightarrow{u.c.p.} \int_0^t g(c_s) \, ds
\] (4.8)
The consistency of \( V^n_t \) as defined in Equation (4.5) follows from Theorem 9.4.1 of Jacod and Protter (2011). The only extra assumption needed is Assumption (T-1):

\[
\zeta \Delta_n \leq \Delta_{n,i} \leq \epsilon \Delta_n,
\]

where \( \zeta > 0 \). Detailed proofs are provided in Section C.1.

4.4.2 Central Limit Theorem

To get rid of the bias term as in Theorem 3.1 in Jacod et al. (2013), we follow the steps below:

1. First estimate spot volatility

\[
\hat{c}_n = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \frac{(\Delta_{n+m} X)^2}{\Delta_{n,i+m}} 1(\|\Delta_n X\| \leq u_{n,i+m}),
\]

where \( u_{n,i+m} = O_p(\Delta_{n,i+m}^{1/2-k}) \).

2. Then

\[
V(g)_t^n = \frac{1}{k_n} \sum_{i=1}^{[t/\Delta_n]} \Delta_{n,i} g(\hat{c}_n) \xrightarrow{u.p.} V(g)_t := \int_0^t g(c_s) ds, \tag{4.9}
\]

\[
V'(g)_t^n = \frac{N_n(t) - k_n + 1}{\sum_{i=1}^{N_n(t) - k_n + 1}} \Delta_{n,i} \left( g \left( c_n \right) - k_n \sum_{j,k,l,m=1}^d \partial_j^2 g(\hat{c}_i) \times (\hat{c}_i^j) \right). \tag{4.10}
\]

**Theorem 4.4.2.** 1. Assume assumptions (A-r) for some \( r < 1 \) and Assumption (T). Let \( g \) be a \( C^3 \) function on \( \mathbb{R}^+ \) such that

\[
\|\partial_j g(x)\| \leq K \left( 1 + \|x\|^{p-j} \right), \quad j = 0, 1, 2, 3
\]

for some constants \( K > 0, p \geq 3 \).
2. Either suppose that \( X \) is continuous and \( u_n/\Delta_n^\varepsilon \to \infty \) for some \( \varepsilon < 1/2 \), or suppose that
\[
u_n \leq \Delta_n^\omega, \quad \frac{2p-1}{2(2p-1)} \leq \omega < \frac{1}{2}.
\]

3. Then we have the (functional in time) stable convergences in law
\[
\frac{1}{\sqrt{\Delta_n}} (V^n(g) - V(g)) \xrightarrow{\mathcal{L}^s} Z
\]
where \( Z \) is a process defined on an extension \( (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \geq 0}, \tilde{\mathbb{P}}) \) of \( (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}) \), which conditionally on \( \mathcal{F} \) is a continuous centered Gaussian martingale with variance
\[
\mathbb{E}((Z_t)^2 | \mathcal{F}) = 2 \int_0^t (\partial g(c_s))^2 \frac{1}{T'} (s) c_s^2 ds.
\]

Proofs can be found in Section C.2. We also propose another estimator that do not use overlapped moving window. We show in Section C.3 that this estimator follows the same central limit theorem.

\[
V^n(g)_i = \frac{[N_n(t)/k_n]^{-1}}{k_n} \sum_{i=0}^{[N_n(t)/k_n]-1} (T(n, (i+1) k_n) - T(n, ik_n))
\]
\[
\times \left( g(\tilde{c}_{ik_n+1}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \partial^2_{jk,lm} g(\tilde{c}_{ik_n+1}) \times \left( \tilde{c}_{ik_n+1}^{n,ji}, \tilde{c}_{ik_n+1}^{n,km}, \tilde{c}_{ik_n+1}^{n,ji}, \tilde{c}_{ik_n+1}^{n,km} \right) \right).
\]

4.4.3 Efficiency bound

In Renault et al. (2017), they define \( \psi_g(\sigma^2) = \int_0^1 g(u, \sigma^2(u)) w(u) du \) and state that under their Assumptions 1-7, any regular estimator for \( \psi_g(\sigma^2) \) as defined above based
on observations on the grid $T(n,i), i = 1, \cdots, n$, has, under $\mathbb{P}^{(n)}_{\sigma_0^2}$, a limiting variance of at least

$$V(g) = 2 \int_0^1 \left( \frac{\partial g}{\partial \sigma^2} (u, \sigma_0^2(u)) \right)^2 \sigma_0^2(u) \frac{w^2(u)}{T'(u)} du.$$ 

where $T'_n(u)^{-1} = \frac{\Delta n_i}{\Delta n}$ for $T(n,i-1) < u < T(n)$.

Besides, by their assumption $T'_n(u)$ converges to $T'(u)$ a.s.. Let $w(u) = 1$ and $g(u, \sigma^2(u)) = g(\sigma^2(u))$, We show that

$$V(g) = 2 \int_0^1 \left( \frac{\partial g}{\partial \sigma^2} (\sigma_0^2(u)) \right)^2 \sigma_0^2(u) \frac{1}{T'(u)} du,$$

which means our estimator achieves the efficiency bound.

4.5 Simulation

In this section, we set up simulations to show that our central limit theorem holds. Then we compare our estimator with other estimators to show its superiority under certain circumstances.

4.5.1 A simple sampling framework

We come up with a simple time function such that the transaction rate in the mid of one day is low while that at the beginning and end is high. Note that a beta distribution, $Beta(\alpha, \alpha)$ is suitable. As at point 0 and 1 the beta function goes to infinity, we need a truncated pdf.

$$g(x) = c \frac{1}{B(\alpha, \alpha)} \left( \frac{x}{1-x} \right)^{\alpha-1} \mathbb{1}_{[m_1 \leq x \leq m_2]} = cf(x) \mathbb{1}_{[m_1 \leq x \leq m_2]}.$$
If the CDF of the beta function is $F(x)$, then
\[ c = \frac{1}{F(m_2) - F(m_1)}. \]

Then we need to extend the window of $[m_1, m_2]$ to $[0, 1]$. Make a transformation as below:
\[ t = \frac{x - m_1}{m_2 - m_1}. \]

Then $t \in [0, 1]$. The "transaction density" function is:
\[ T'(t) = \frac{(m_2 - m_1)}{F(m_2) - F(m_1)} f ((m_2 - m_1) t + m_1) 1_{\{0 \leq t \leq 1\}}. \]

Integrate it we get
\[ T(t) = \frac{F((m_2 - m_1) t + m_1) - F(m_1)}{F(m_2) - F(m_1)}. \]

Figure shows an illustration of the time process.

4.5.2 Coverage rate

Simulate stock price as follows:
\[ dX_t = \left( r - \frac{1}{2} \nu_t \right) dt + \nu_t^{1/2} dW_t^Q, \]
\[ d\log(\nu_t) = \kappa (\log(\nu_0) - \log(\nu_t)) dt + \gamma dB_t^Q. \]

Specify the parameters as follows:
\[ \kappa = 5, \rho = 0, \gamma = 1, \nu_0 = 0.18, r = \frac{1}{2} v_0. \]
There are 23400 prices that can be sampled, which empirically states that prices change every second. The tuning parameters are as follows:

- $k_n$: Number of observations in the window to estimate spot volatility;
- $n$: number of the "observed" prices;
- $\alpha$: The variation in transaction intensity process.

The coverage rate of the 99%, 95% and 90% percentage are shown in Table 4.1. We see that generally the coverage rate is similar to the theoretical value, especially when the sample frequency is large.

4.5.3 Jacod et al. (2013) under an irregular setting

If we ignore the information about the observation time but use the Jacod et al. (2013) efficient estimator as if the prices are sampled regularly. Theoretically, the
Table 4.1: Confidence Interval Coverage Rate

<table>
<thead>
<tr>
<th></th>
<th>( \alpha = 1 )</th>
<th>( \alpha = 0.7 )</th>
<th>( \alpha = 0.4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 78 )</td>
<td>98.05 98.81 99.07</td>
<td>98.22 98.98 99.01</td>
<td>98.49 98.89 99.29</td>
</tr>
<tr>
<td>( n = 390 )</td>
<td>97.63 98.81 98.99</td>
<td>98.01 98.77 99.05</td>
<td>98.59 98.68 99.17</td>
</tr>
<tr>
<td>( n = 780 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: 99% Confidence interval

Panel B: 95% Confidence interval

Panel C: 90% Confidence interval

Table presents the results of for the function: \( g(x) = 2x^2 \) with 10000 simulations.

central limit theorem of Jacod et al. (2013) under infrequent trading and irregular sampling is as follows.

**Proposition 4.5.1.** 1. if either \( X \) is continuous or \( p < 1 \), we have

\[
C^n(g, k_n, X) \overset{u.c.p.}{\longrightarrow} \int_0^t T'(s) g\left(\frac{c_s}{T'(s)}\right)ds
\]

2. if either \( p \leq 1 \), or \((H-r)\) holds for some \( r \in [0, 2) \) and

\[
p > 1, \ \omega \geq \frac{p - 1}{2p - 4}
\]

we have

\[
C^n(g, k_n, v_n, X) \overset{u.c.p.}{\longrightarrow} \int_0^t T'(s) g\left(\frac{c_s}{T'(s)}\right)ds
\]

The estimator is not consistent unless \( g(\cdot) \) is linear. It turns out that when \( g(\cdot) \) is linear, the "regular estimator" is also efficient. Consider the following two functions:
Table 4.2: Consistency: Average difference from Average theoretical value (in %)

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Irregular Estimator</th>
<th>Panel B: Regularly Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0.8$</td>
<td>$\alpha = 0.4$</td>
</tr>
<tr>
<td></td>
<td>$n = 390$</td>
<td>$n = 780$</td>
</tr>
<tr>
<td>$g(x) = 2x^2$</td>
<td>$k_n = 30$ 0.74 0.79</td>
<td>$k_n = 60$ 1.43 0.89</td>
</tr>
<tr>
<td></td>
<td>$k_n = 60$ 1.71 1.25</td>
<td>$k_n = 60$ 1.69 1.14</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>$k_n = 30$ 0.25 -0.11 -0.25 -0.06</td>
<td>$k_n = 60$ 0.15 0.22 0.44 0.05</td>
</tr>
<tr>
<td></td>
<td>$k_n = 60$ -0.12 -0.09 -0.35 -0.37</td>
<td>$k_n = 60$ 1.35 0.44 1.66 0.61</td>
</tr>
</tbody>
</table>

Table presents the results with 2000 simulations.

- $f_1(x) = 2x^2$
- $f_2(x) = x$

In Panel A, we use the estimator we proposed in Section 4.3. In Panel B, we sample the prices regularly and then we use the efficient estimator in Jacod et al. (2013) and the central limit theorem in it. That is, if several observations in $((i-1)\Delta_n, i\Delta_n]$, $X_i = X_{t_{n,j}}$, where $t_{n,j}$ is the latest time in the interval. If no observation in $((i-1)\Delta_n, i\Delta_n]$, then $X_i = X_{i-1}$. Compare the bias and the coverage rate of the confidence intervals.

In Table 4.2 and Table 4.2, in Panel A we use the irregularly sampled estimator we propose in this chapter. In panel B, prices are generated irregularly, but the integrated function is estimated using Jacod et al. (2013) as if the prices are sampled regularly. Table 4.2 shows that when the function is $f_1(x) = 2x^2$ the bias from the theoretical value is large, and increases when there are more time variations. However, when the function is $f_2(x) = x$, the bias is small and similar to the irregular estimator. This is consistent with Proposition 4.5.1.

Table 4.2 shows that when the function is $f_1(x) = 2x^2$ the coverage rate is smaller than the theoretical value, and drops dramatically when there are more time variations. However, when the function is $f_2(x) = x$, the coverage rate is close to the
Table 4.3: Central limit theorem: 99%, 95% and 90% coverage rate

<table>
<thead>
<tr>
<th>Panel A: Irregular Estimator</th>
<th>Panel B: Regularly Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(x) = 2x^2$</td>
<td>$g(x) = x$</td>
</tr>
<tr>
<td>$\alpha = 0.8$</td>
<td>$\alpha = 0.4$</td>
</tr>
<tr>
<td>$\alpha = 0.4$</td>
<td>$\alpha = 0.4$</td>
</tr>
<tr>
<td>$n = 390$</td>
<td>$n = 390$</td>
</tr>
<tr>
<td>$n = 780$</td>
<td>$n = 780$</td>
</tr>
<tr>
<td>$k_n = 30$</td>
<td>$k_n = 30$</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>$k_n = 60$</td>
</tr>
<tr>
<td>$g(x) = 2x^2$</td>
<td>$g(x) = x$</td>
</tr>
<tr>
<td>$k_n = 30$</td>
<td>$k_n = 30$</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>$k_n = 60$</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>$g(x) = x$</td>
</tr>
<tr>
<td>$k_n = 30$</td>
<td>$k_n = 30$</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>$k_n = 60$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>99% Coverage Rate</th>
<th>99% Coverage Rate</th>
<th>95% Coverage Rate</th>
<th>95% Coverage Rate</th>
<th>90% Coverage Rate</th>
<th>90% Coverage Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_n = 30$</td>
<td>98.95</td>
<td>98.9</td>
<td>98.95</td>
<td>97.8</td>
<td>97.9</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>98.6</td>
<td>98.75</td>
<td>98.25</td>
<td>99</td>
<td>98.75</td>
</tr>
<tr>
<td>$99.3$</td>
<td>99.2</td>
<td>98.95</td>
<td>97.8</td>
<td>97.9</td>
<td>96.9</td>
</tr>
<tr>
<td>$k_n = 30$</td>
<td>98.6</td>
<td>98.75</td>
<td>98.25</td>
<td>99</td>
<td>98.75</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>98.95</td>
<td>99</td>
<td>98.95</td>
<td>97.8</td>
<td>97.9</td>
</tr>
<tr>
<td>$99.3$</td>
<td>99.2</td>
<td>98.95</td>
<td>97.8</td>
<td>97.9</td>
<td>96.9</td>
</tr>
<tr>
<td>$k_n = 30$</td>
<td>96.05</td>
<td>95.9</td>
<td>96.4</td>
<td>93.35</td>
<td>93.25</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>95.15</td>
<td>95.15</td>
<td>94.85</td>
<td>92.8</td>
<td>93.05</td>
</tr>
<tr>
<td>$95.9$</td>
<td>95.4</td>
<td>94.85</td>
<td>92.8</td>
<td>93.05</td>
<td>92.8</td>
</tr>
<tr>
<td>$k_n = 30$</td>
<td>95.6</td>
<td>95.1</td>
<td>94.7</td>
<td>95.35</td>
<td>94.95</td>
</tr>
<tr>
<td>$k_n = 60$</td>
<td>93.7</td>
<td>94.75</td>
<td>93.55</td>
<td>94.7</td>
<td>94.75</td>
</tr>
<tr>
<td>$95.1$</td>
<td>95.6</td>
<td>94.75</td>
<td>93.55</td>
<td>94.7</td>
<td>94.75</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>91.05</td>
<td>91.3</td>
<td>91.9</td>
<td>88.55</td>
<td>88.05</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>90.85</td>
<td>90.1</td>
<td>89.8</td>
<td>87.6</td>
<td>87.75</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>90.65</td>
<td>90.2</td>
<td>91.7</td>
<td>90.7</td>
<td>90</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>88.75</td>
<td>91.15</td>
<td>89.4</td>
<td>88.8</td>
<td>88.95</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>88.75</td>
<td>89.15</td>
<td>89.15</td>
<td>88.8</td>
<td>88.8</td>
</tr>
<tr>
<td>$g(x) = x$</td>
<td>88.75</td>
<td>89.15</td>
<td>89.15</td>
<td>88.8</td>
<td>88.8</td>
</tr>
</tbody>
</table>

Table presents the results with 2000 simulations.

theoretical value and similar to the irregular estimator. This is again consistent with Proposition 4.5.1.

4.6 Conclusion

In this chapter, we propose an estimator of functional of variance with irregular observation time that reaches the efficiency bound. We proved its consistency and central limit theorem. In the simulation, we show that the estimator achieves consistency under circumstances when the efficiency estimator in Jacod et al. (2013) does not due to infrequent trading and regular sampling.
Conclusion

My dissertation is mainly about the effect of macroeconomic release on the relationship between trading intensity and price volatility, which falls into the literature of market microstructure. Chapter 2 focuses on the market portfolio and finds that investors’ disagreement measures significantly reduce volume-volatility elasticity using high frequency data of S&P 500 ETF. Chapter 3 extends to individual stocks and finds that the same conclusion holds, and individual stocks with larger idiosyncratic volatility have lower elasticity. Besides, the negative effect of investors’ measures is only significant during high sentiment periods when investor’s speculative propensity is high, which is consistent with a model that includes investors’ confidence in interpreting public signals. Chapter 4 proposes an efficient estimator for the integrated functional of volatility for assets with irregularly observed prices. The consistency and central limit theorems are validated through proofs and simulations.
One further extension of my dissertation is the reaction of bid-ask spread around macroeconomic announcements and firm specific news announcements. Another extension is the volume-volatility relationship around firm specific news. With advances in textual analysis, it is possible to categorize news into more and less informative ones, and then one can study the intraday effect of different categories of news.
A.1 Technical background for econometric procedures

This appendix presents the formal econometric theory behind our high-frequency econometric estimation and inference procedures discussed in Section 2.5. Appendix A.1 describes the continuous-time setup for modeling the high-frequency price and volume data used in our empirical analysis. Appendix A.2 presents the main theoretical results, which we prove in Appendix A.3.

A.1.1 Continuous-time setup and definitions

Throughout, we fix a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\). Let \((P_t)_{t \geq 0}\) denote the logarithmic price process of an asset. As is standard in the continuous-time finance literature (see, e.g., Merton (1992) and Duffie (2001)), we assume that \(P\) is a
jump-diffusion process of the form

$$dP_t = b_t dt + \sigma_t dW_t + dJ_t,$$  \hfill (A.1)

where $b$ is an instantaneous drift process, $\sigma$ is a stochastic spot volatility process, $W$ is a Brownian motion, and $J$ is a pure jump process. The price is sampled at discrete times $\{i \Delta_n : 0 \leq i \leq [T/\Delta_n]\}$, where $T$ denotes the sample span and $\Delta_n$ denotes the sampling interval of the high-frequency data. We denote the corresponding high-frequency asset returns by $r_i \equiv P_{i\Delta_n} - P_{(i-1)\Delta_n}$.

Our empirical analysis is justified using an infill econometric theory with $\Delta_n \to 0$ and $T$ fixed. This setting is standard for analyzing high-frequency data (see, e.g., Aït-Sahalia and Jacod (2014) and Jacod and Protter (2012)) and it allows us to nonparametrically identify processes of interest in a general setting with essentially unrestricted nonstationarity and persistence.

We denote the trading volume within the high-frequency interval $((i-1)\Delta_n, i\Delta_n]$ by $V_{i\Delta_n}$. Unlike the price, the high-frequency volume data cannot be realistically modeled using the jump-diffusion model. Following Li and Xiu (2016), we consider a general state-space model

$$V_{i\Delta_n} = \mathcal{V}(\zeta_{i\Delta_n}, \epsilon_{i\Delta_n}),$$  \hfill (A.2)

where $\zeta$ is a latent state process, $(\epsilon_{i\Delta_n})$ are i.i.d. transitory shocks with distribution $F_{\epsilon}$, and $\mathcal{V}(\cdot)$ is a possibly unknown transform. The latent state process $\zeta$ captures time-varying conditioning information such as the intensity of order arrival and the shape of the order size distribution. Technically, this state-space model can be formally defined on an extension of the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$; see Li and Xiu (2016) for a formal construction.
This state-space model suits our empirical modeling of the trading volume well in two ways. Firstly, it conveniently allows the observed volume data to be discretely valued, while still being completely flexible on the state space of the latent process \( \zeta \). This allows us to model \( \zeta \) under minimal statistical restrictions without introducing unintended model inconsistency. Secondly, the setup does not restrict the dynamic persistence of the state process \( \zeta \). This is important as volume data often exhibit nonstationary behavior.

Lastly, to complete the notation, let

\[
m_t = \int \mathcal{V}(\zeta_t, \epsilon) F_\epsilon(d\epsilon)
\]

(A.3)

denote the instantaneous conditional mean process of \( V \), that is, the volume intensity process. We use this process as the instantaneous empirical analogue to the expected volume in the theoretical models discussed in Section 2.2. Correspondingly, we use the spot volatility processes \( \sigma \) as the instantaneous analogue of the return standard deviation. Empirically, we should and do allow for general stochastic behavior in these instantaneous moments so as to accommodate essentially arbitrary stochastic behaviors of time-varying conditioning information.

A.1.2 Estimation and inference

We now formally justify our inference procedures used in the main text. Below, we write \( \tilde{\theta}_n \) in place of \( \theta \) and similarly for other estimators, so as to emphasize the asymptotic stage. We focus on the DID jump regression estimator \( \tilde{\theta}_n \) defined by equation (3.14), which includes the “raw” estimator \( \hat{\theta}_n \) as a special case with the
control group \( \mathcal{C} \) set to be empty. We write \( \tilde{\theta}_n = f(\tilde{S}_n) \), where \( f(\cdot) \) is defined implicitly by the definitions (3.12) and (3.14).

We assume the following regularity conditions for the underlying processes.

**Assumption 2.**

(i) The price process \( P \) is given by (A.1) for \( J_t = \int_0^t \xi_s dN_s + \int_0^t \delta(s, z) \mu(ds, dz) \), where the processes \( b \) and \( \sigma \) are càdlàg (i.e., right continuous with left limit) and adapted; \( \sigma \) is positive for \( t \in [0, T] \) almost surely; the process \( \xi \) is predictable and locally bounded; \( N \) is a counting process that jumps at the scheduled announcement times which are specified by the set \( \mathcal{A} \); \( \delta \) is a predictable function; \( \mu \) is a Poisson random measure with compensator \( \nu(ds, dz) = ds \otimes \lambda(dz) \) for some finite measure \( \lambda \).

(ii) The volume process \( V \) satisfies (A.2). The process \( \zeta \) is càdlàg and adapted. The error terms \( \epsilon_i \) take values in some Polish space, are defined on an extension of \( (\Omega, \mathcal{F}) \), i.i.d. and independent of \( \mathcal{F} \).

(iii) For a sequence of stopping times \( (T_m)_{m \geq 1} \) increasing to infinity and constants \( (K_m)_{m \geq 1} \), we have \( \mathbb{E} \left| \sigma_{t \wedge T_m} - \sigma_{s \wedge T_m} \right|^2 + \mathbb{E} \left| \zeta_{t \wedge T_m} - \zeta_{s \wedge T_m} \right|^2 \leq K_m |t - s| \) for all \( t, s \) such that \( [s, t] \cap \mathcal{A} = \emptyset \).

Assumption 5 is fairly standard in the study of high-frequency data. Condition (i) allows the price process to contain jumps at both scheduled times and random times. Condition (ii) separates the conditional i.i.d. shocks \( \epsilon_i \) at observation times from the latent continuous-time state process \( \zeta_t \). This condition only mildly restricts the volume series, which can still exhibit essentially unrestricted conditional and unconditional heterogeneity through the (typically highly persistent) time-varying state process \( \zeta_t \). Condition (iii) imposes a mild smoothness condition on \( \sigma \) and \( \zeta \) only.
in expectation, while allowing for general forms of jumps in their sample paths. This condition is satisfied for any semimartingales with absolutely continuous predictable characteristics (possibly with discontinuity points in $A$) and for long-memory type processes driven by the fractional Brownian motion.

In addition, we need the following conditions for the nonparametric analysis, where we denote $M_p (\cdot) \equiv \int \mathcal{V} (\cdot, \epsilon)^p F_{\epsilon} (d\epsilon)$ for $p \geq 1$.

**Assumption 3.** $k_n \to \infty$ and $k_n^2 \Delta_n \to 0$.

**Assumption 4.** (i) The function $M_1 (\cdot)$ is Lipschitz on compact sets and the functions $M_2 (\cdot)$ and $M_4 (\cdot)$ are continuous.

(ii) Almost surely, the function $f$ is well-defined and continuously differentiable in a neighborhood of $\tilde{S}$.

Assumption 6 specifies the growth rate of the local window size $k_n$. As typical in nonparametric analysis, this condition features a type of undersmoothing, so as to permit feasible inference. Assumption 7 imposes some smoothness conditions that are very mild.

We need some notations for stating the asymptotic results. For notational simplicity, we denote $v_t = M_2 (\zeta_t) - M_2^1 (\zeta_t)$. Consider variables $(\eta_{r-}, \eta_r, \eta_{r-}, \eta_{r}^\prime)_{r \in T}$ which, conditionally on $\mathcal{F}$, are mutually independent, centered Gaussian with variances $(v_{r-}, v_r, \sigma_{r-}^2/2, \sigma_r^2/2)_{r \in T}$. We denote the first differential of $f$ at $\tilde{S}$ with increment $d\tilde{S}$ by $F (\tilde{S}; d\tilde{S})$. For a sequence $Y_n$ of random variables, we write $Y_n \xrightarrow{L^s} Y$ if $Y_n$ converges stably in law towards $Y$, meaning that $(Y_n, U)$ converges in distribution to $(Y, U)$ for any bounded $\mathcal{F}$-measurable random variable $U$. 

139
Theorem B.1.1 characterizes the asymptotic distribution of the estimator $\tilde{\theta}_n$, and further shows that the asymptotic distribution can be consistently approximated using the Bootstrap Algorithm described in Section 2.5.

**Theorem A.1.1.** (a) Under Assumptions 5, 6 and 7

$$
\sqrt{k_n(\bar{\theta}_n - \theta)} \overset{\mathcal{L}}{\to} F(\tilde{S}; (\eta_{r-}, \eta_r, \eta'_{r-}, \eta'_r, 0)_{r \in \mathcal{T}}).
$$

(A.4)

(b) Moreover, the conditional distribution function of $\sqrt{k_n(\bar{\theta}_n - \theta)}$ given the original data converges in probability to that of $F(\tilde{S}; (\eta_{r-}, \eta_r, \eta'_{r-}, \eta'_r, 0)_{r \in \mathcal{T}})$ under the uniform metric.

**Comment.** Since the variables $(\eta_{r-}, \eta_r, \eta'_{r-}, \eta'_r)_{r \in \mathcal{T}}$ are jointly $\mathcal{F}$-conditionally centered Gaussian, so is the limiting distribution given by (B.1).

### A.1.3 Proofs

Throughout the proofs, we use $K$ to denote a generic constant which may change from line to line. For a generic random sequence $Y_n$, we write $Y_n \overset{\mathcal{F}}{\to} Y$ if the $\mathcal{F}$-conditional distribution function of $Y_n$ converges in probability to that of $Y$ under the uniform metric. By a classical localization procedure (see, e.g., Section 4.4.1 in Jacod and Protter (2012)), we can assume that the processes $\sigma$ and $\zeta$ are bounded, and piecewise $(1/2)$-Hölder continuous under the $L^2$-norm (with possible discontinuity points given by $\mathcal{A}$) without loss of generality. We consider a sequence $\Omega_n$ of events given by

$$
\Omega_n = \{\text{the intervals } [\tau - 2k_n \Delta_n, \tau + 2k_n \Delta_n], \ \tau \in \mathcal{T}, \ \text{are mutually disjoint}\}.
$$
Since $\mathcal{T}$ is finite, $\mathbb{P}(\Omega_n) \to 1$. Therefore, we can focus attention on $\Omega_n$, again without loss of generality.

**Proof of Theorem B.1.1(a).** We first show that

$$
\sqrt{k_n}(\hat{m}_{n,\tau} - m_\tau, \hat{m}_{n,\tau} - m_\tau) \overset{\mathcal{L}_F}{\to} \mathbf{\eta}_\tau, \mathbf{\eta}_\tau \forall \tau \in \mathcal{T}.
$$

(A.5)

In restriction to $\Omega_n$, the estimators on the left-hand side of (A.5) are $\mathcal{F}$-conditionally independent. Therefore, it suffices to establish the convergence for the marginal distributions. We thus focus on $\sqrt{k_n}(\hat{m}_{n,\tau} - m_\tau)$, noting that the proof concerning $\sqrt{k_n}(\hat{m}_{n,\tau} - m_\tau)$ is similar.

We decompose $\sqrt{k_n}(\hat{m}_{n,\tau} - m_\tau) = A_n + R_n$, where

$$
A_n \equiv \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} \left( V_{i(\tau)+j} \Delta_n - M_1(\zeta_{i(\tau)+j} \Delta_n) \right),
$$

$$
R_n \equiv \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} \left( M_1(\zeta_{i(\tau)+j} \Delta_n) - M_1(\zeta_\tau) \right).
$$

By Assumption 7, $\mathbb{E}|R_n| \leq K k_n \Delta_n^{1/2}$, which goes to zero by Assumption 6.

It remains to consider the convergence of $A_n$. Since $(V_{i\Delta_n})_{i \geq 0}$ are $\mathcal{F}$-conditionally independent,

$$
\mathbb{E} \left[ A_n^2 | \mathcal{F} \right] = \frac{1}{k_n} \sum_{j=1}^{k_n} V_{i(\tau)+j} \Delta_n \to v_\tau,
$$

(A.6)

where the convergence holds because the process $v$ is càdlàg. Moreover, by the continuity of $M_4(\cdot)$ and the boundedness of $\zeta$, we can verify a Lyapunov-type condition:

$$
\mathbb{E} \left[ \left| V_{i(\tau)+j} \Delta_n - M_1(\zeta_{i(\tau)+j} \Delta_n) \right|^4 | \mathcal{F} \right] \leq K M_4 \left( \zeta_{i(\tau)+j} \Delta_n \right) \leq K.
$$

(A.7)
By (A.6) and (A.7), we can apply the Lindeberg central limit theorem under the $\mathcal{F}$-conditional probability, resulting in $\sqrt{k_n}(\hat{m}_\tau - m_\tau) \xrightarrow{\mathcal{F}} \eta_\tau$ as claimed.

Since the jumps of $P$ are of finite activity, the returns involved in $\hat{\sigma}_{n,\tau-}$ and $\hat{\sigma}_{n,\tau}$ do not contain jumps, with probability approaching one. By Theorem 13.3.3(c) of Jacod and Protter (2012), we have

$$\sqrt{k_n}(\hat{\sigma}_{n,\tau-} - \sigma_{n,\tau-}, \hat{\sigma}_{n,\tau} - \sigma_{n,\tau}) \xrightarrow{\mathcal{F}} (\eta_{\tau-}, \eta_{\tau}, \eta_{\tau-}, \eta_{\tau})_{\tau \in T}.$$  \hfill (A.8)

By Proposition 5 of Barndorff-Nielsen et al. (2008) and the property of stable convergence in law, we can combine (A.5) and (A.8), yielding

$$\sqrt{k_n}(\tilde{S}_n - \tilde{S}) \xrightarrow{\mathcal{F}} (\eta_{\tau-}, \eta_{\tau}, \eta_{\tau-}, \eta_{\tau})_{\tau \in T}.$$  \hfill (A.9)

The assertion (B.1) then follows from (B.4) and the delta method. \hfill Q.E.D.

**Proof of Theorem B.1.1(b).** Step 1. We divide the proof into several steps. Denote $\mathcal{G} \equiv \mathcal{F} \lor \sigma \{\epsilon_i : i \geq 0\}$. In this step, we show that, for each $\tau \in T$,

$$\sqrt{k_n}(\hat{\sigma}_{n,\tau-}^* - \hat{\sigma}_{n,\tau-} - \hat{\sigma}_{\tau-}^* - \hat{\sigma}_{\tau-}^*) \xrightarrow{\mathcal{G}} (\eta_{\tau-}, \eta_{\tau-}').$$  \hfill (A.10)

Observe that

$$\sqrt{k_n}\left(\hat{\sigma}_{\tau-}^2 - \hat{\sigma}_{\tau-}^2\right) = \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} \left(\frac{V_{(i(\tau)-j)}^* \Delta_n}{r_{i(\tau)-j}^2/\Delta_n} - \frac{1}{k_n} \sum_{j=1}^{k_n} \left(\frac{V_{(i(\tau)-j)} \Delta_n}{r_{i(\tau)-j}^2/\Delta_n}\right)\right).$$  \hfill (A.11)

By the construction of the bootstrap sample, the summands in the right-hand side of (B.6) are i.i.d. with zero mean conditional on $\mathcal{G}$. We denote the $\mathcal{G}$-conditional
covariance matrix of $\sqrt{k_n}(\hat{m}_{\tau-} - \bar{m}_{\tau-}, \hat{\sigma}_{\tau-}^2 - \bar{\sigma}_{\tau-}^2)$ by

$$
\Sigma_{n,\tau-} = \begin{pmatrix}
\Sigma_{n,\tau-}^{(11)} & \Sigma_{n,\tau-}^{(12)} \\
\Sigma_{n,\tau-}^{(12)} & \Sigma_{n,\tau-}^{(22)}
\end{pmatrix},
$$

where

$$
\left\{
\begin{array}{l}
\Sigma_{n,\tau-}^{(11)} = \frac{1}{k_n} \sum_{j=1}^{k_n} V^2_{(i(\tau)-j)\Delta_n} - \left( \frac{1}{k_n} \sum_{j=1}^{k_n} V_{(i(\tau)-j)\Delta_n} \right)^2, \\
\Sigma_{n,\tau-}^{(12)} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} V_{(i(\tau)-j)\Delta_n} V_{(i(\tau)-j)\Delta_n}^2 - \left( \frac{1}{k_n} \sum_{j=1}^{k_n} V_{(i(\tau)-j)\Delta_n} \right) \left( \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \tau_{i(\tau)-j} \right)^2, \\
\Sigma_{n,\tau-}^{(22)} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \tau_{i(\tau)-j}^4 - \left( \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \tau_{i(\tau)-j} \right)^2.
\end{array}\right.
$$

In step 2, we shall show that

$$
\left(\Sigma_{n,\tau-}^{(11)}, \Sigma_{n,\tau-}^{(12)}, \Sigma_{n,\tau-}^{(22)}\right) \xrightarrow{p} (\nu_{\tau-}, 0, 2\sigma_{\tau-}^4). \tag{A.12}
$$

For each subsequence $N_1 \subseteq \mathbb{N}$, we can find a further subsequence $N_2 \subseteq N_1$ such that the convergence in (B.7) holds almost surely along $N_2$. Moreover, on the paths for which this convergence holds, we apply a central limit theorem under the $G$-conditional probability to deduce the convergence of the conditional law of $\sqrt{k_n}(\hat{m}_{\tau-} - \bar{m}_{\tau-}, \hat{\sigma}_{\tau-}^2 - \bar{\sigma}_{\tau-}^2)$ towards that of $(\eta_{\tau-}, \eta'_{\tau-})$ along the subsequence $N_2$. From here, we deduce (B.5) for the original sequence by reversing the subsequence argument.

Step 2. We show (B.7) in this step. We start with $\Sigma_{n,\tau-}^{(11)}$. Consider the decomposition

$$
\frac{1}{k_n} \sum_{j=1}^{k_n} V^2_{(i(\tau)-j)\Delta_n} = A_{1,n} + A_{2,n},
$$

143
where

\[ A_{1,n} \equiv \frac{1}{k_n} \sum_{j=1}^{k_n} \left( V_{(i(j)-j)\Delta_n}^2 - M_2(\zeta_{(i(j)-j)\Delta_n}) \right), \]

\[ A_{2,n} \equiv \frac{1}{k_n} \sum_{j=1}^{k_n} M_2(\zeta_{(i(j)-j)\Delta_n}). \]

Note that the summands in \( A_{1,n} \) are \( \mathcal{F} \)-conditionally independent with zero mean. Therefore,

\[ \mathbb{E}[A_{1,n}^2 | \mathcal{F}] \leq K k_n^{-2} \sum_{j=1}^{k_n} M_4(\zeta_{(i(j)-j)\Delta_n}) \leq K k_n^{-1} \rightarrow 0, \]

which further implies that \( A_{1,n} = o_p(1) \). In addition, since \( M_2(\cdot) \) is continuous and \( \zeta \) is càdlàg, \( A_{2,n} \rightarrow M_2(\zeta_{-\nu}) \). From here and (B.4), it follows that

\[ \Sigma_{n,\tau_-}^{(11)} \overset{P}{\rightarrow} M_2(\zeta_{-\nu}) - M_1(\zeta_{-\nu})^2 = v_{\nu -}. \]

Next, we consider the limiting behavior of \( \Sigma_{n,\tau_-}^{(12)} \). We decompose

\[ \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} V_{(i(j)-j)\Delta_n}^2 r_{(i(j)-j)}^2 = A_{3,n} + A_{4,n} + A_{5,n} + A_{6,n}, \]
where

\[ A_{3,n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (V((i(\tau)-j)\Delta_n) - M_1(\zeta((i(\tau)-j)\Delta_n)) r_{i(\tau)-j}^2), \]

\[ A_{4,n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (M_1(\zeta((i(\tau)-j)\Delta_n)) - M_1(\zeta((i(\tau)-j-1)\Delta_n)) r_{i(\tau)-j}^2), \]

\[ A_{5,n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} M_1(\zeta((i(\tau)-j-1)\Delta_n)) (r_{i(\tau)-j}^2 - \mathbb{E}[r_{i(\tau)-j}^2 | F_{i(\tau)-j-1}])), \]

\[ A_{6,n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} M_1(\zeta((i(\tau)-j-1)\Delta_n)) \mathbb{E}[r_{i(\tau)-j}^2 | F_{i(\tau)-j-1}]. \]

We now show that

\[ A_{j,n} = o_p(1), \quad \text{for} \quad j = 3, 4, 5. \quad (A.13) \]

We note that \( A_{3,n} \) is an average of variables that are, conditionally on \( F \), independent with zero mean. Therefore, \( \mathbb{E}[A_{3,n}^2 | F] \leq k_n^{-2} \Delta_n^{-2} \sum_{j=1}^{k_n} M_2(\zeta((i(\tau)-j)\Delta_n) r_{i(\tau)-j}^4 \mathbb{E}[r_{i(\tau)-j}^4 | F_{i(\tau)-j-1}] \), which further implies \( \mathbb{E}[A_{3,n}^2] \leq K k_n^{-1} \rightarrow 0 \). Turning to \( A_{4,n} \), we observe

\[ \mathbb{E}[A_{4,n}] \leq \frac{K}{k_n \Delta_n} \sum_{j=1}^{k_n} \mathbb{E}\left[|\zeta((i(\tau)-j)\Delta_n) - \zeta((i(\tau)-j-1)\Delta_n)|^2 r_{i(\tau)-j}|^2\right] \]

\[ \leq K k_n \Delta_n \sum_{j=1}^{k_n} \left( \mathbb{E}\left[|\zeta((i(\tau)-j)\Delta_n) - \zeta((i(\tau)-j-1)\Delta_n)|^2\right] \mathbb{E}[r_{i(\tau)-j}^4] \right)^{1/2} \]

\[ \leq K \Delta_n^{1/2} \rightarrow 0, \]

where the first inequality holds because \( M_1(\cdot) \) is locally Lipschitz and \( \zeta \) is bounded; the second inequality follows from the Cauchy–Schwarz inequality; the third inequality is
due to the $(1/2)$-Hölder continuity of $\zeta$ and $\mathbb{E}[r_i\Delta n] \leq K\Delta_n^2$. Finally, we note that $A_{5,n}$ is an average of $k_n$ martingale difference terms with bounded second moments. Therefore, $\mathbb{E}[A_{5,n}^2] \leq K^{-1}k_n \to 0$. The proof for (A.13) is now finished.

We now consider $A_{6,n}$. By Itô’s formula and some standard estimates for continuous Itô semimartingales (noting that $r_i\tau_j$ does not contain jumps in restriction to $\Omega_n$),

$$\mathbb{E}\left[r_i\Delta_n^2 \mid \mathcal{F}_\tau\right] - \sigma_i^2(\tau) \Delta_n = \mathbb{E}\left[\sigma_i^2 - \sigma_i^2(\tau) \Delta_n \mid \mathcal{F}_\tau\right] \Delta_n ds + o_{pa}(\Delta_n),$$

where $o_{pa}(\Delta_n)$ denotes a term that is $o_p(\Delta_n)$ uniformly in $j$. Hence,

$$\left|A_{6,n} - \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{1}{\Delta_n} \left\{ M_1(\zeta_{(i\tau)\Delta_n}) \sigma_i^2 - \sigma_i^2(\tau) \Delta_n \right\} \right| \leq K \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{1}{\Delta_n} \mathbb{E}\left[\sigma_i^2 - \sigma_i^2(\tau) \Delta_n \mid \mathcal{F}_\tau\right] \Delta_n ds + o_p(1).$$

We further note that the expectation of the first term on the majorant side of the above display is bounded by $K\mathbb{E}\left[\sup_{s \epsilon [\tau-2k_n\Delta_n, \tau]} \sigma_s^2 - \sigma_\tau^2\right]$, which goes to zero since $\sigma$ is càdlàg and bounded. Therefore,

$$A_{6,n} = \frac{1}{k_n} \sum_{j=1}^{k_n} M_1(\zeta_{(i\tau)\Delta_n}) \sigma_i^2 - \sigma_i^2(\tau) \Delta_n + o_p(1) \overset{\mathbb{P}}{\to} M_1(\zeta_{\tau}) \sigma_\tau^2.$$

From here, it readily follows that

$$\Sigma_{n,\tau}^{(12)} \overset{\mathbb{P}}{\to} 0.$$
We now turn to the limiting behavior of $\Sigma_{n,\tau}^{(22)}$. For each $i \in \{i(\tau) - j : 1 \leq j \leq k\}$, we can decompose $r_i = r'_i + r''_i$ in restriction to $\Omega$ where

$$
\sigma_{(i-1)\Delta_n} (W_{i\Delta_n} - W_{(i-1)\Delta_n}),
$$

$$
r''_i = \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n} (\sigma_s - \sigma_{(i-1)\Delta_n}) \, dW_s.
$$

We can then decompose

$$
\frac{1}{k_n \Delta_n^2} \sum_{j=1}^{k_n} r^i_{(\tau)-j} = A_{7,n} + A_{8,n},
$$

where $A_{7,n} = \frac{1}{k_n \Delta_n^2} \sum_{j=1}^{k_n} r'^i_{(\tau)-j}$ and $A_{8,n}$ is defined implicitly by (B.9). Note that $E[r'^i_{(\tau)-j} | \mathcal{F}_{(i-1)\Delta_n}] = 3\sigma^4_{(i-1)\Delta_n}$ and the variance of the martingale difference sequence

$$
r'^i_{(\tau)-j} - E[r'^i_{(\tau)-j} | \mathcal{F}_{(i-1)\Delta_n}]$$

is bounded. It is then easy to see that $A_{7,n} \xrightarrow{P} 3\sigma^4_{\tau}$. Turning to the term $A_{8,n}$, we first observe

$$
|A_{8,n}| \leq \frac{K}{k_n \Delta_n^2} \sum_{j=1}^{k_n} |r'_i_{(\tau)-j}|^3 |r''_i_{(\tau)-j}| + \frac{K}{k_n \Delta_n^2} \sum_{j=1}^{k_n} |r''_i_{(\tau)-j}|^4.
$$

By the Burkholder–Davis–Gundy inequality and Hölder’s inequality,

$$
E \left[ |r''_i_{(\tau)-j}|^4 \right] \leq K \Delta_n^4 + K \Delta_n E \left[ \int_{(i(\tau)-j-1)\Delta_n}^{(i(\tau)-j)\Delta_n} (\sigma_s - \sigma_{(i(\tau)-j-1)\Delta_n})^4 \, ds \right].
$$

Hence,

$$
E \left[ \frac{1}{k_n \Delta_n^2} \sum_{j=1}^{k_n} |r''_i_{(\tau)-j}|^4 \right] \leq K \Delta_n^2 + K \frac{1}{k_n \Delta_n} \left[ \int_{(i(\tau)-k_n-1)\Delta_n}^{(i(\tau)-1)\Delta_n} (\sigma_s - \sigma_{(i(\tau)-k_n-1)\Delta_n})^4 \, ds \right]
$$

$$
\leq K \Delta_n^2 + K E \left[ \sup_{t \in [\tau-2k_n\Delta_n, \tau]} |\sigma_t - \sigma_{\tau-1}|^4 \right] \to 0.
$$
where the convergence follows from the bounded convergence theorem, because \( \sigma \) is càdlàg and bounded. Therefore, the second term on the majorant side of (A.15) is \( o_p(1) \). Since \( A_{\tau,n} = O_p(1) \), we can use Hölder’s inequality to show that the first term on majorant side of (A.15) is also \( o_p(1) \). Hence, \( A_{8,n} = o_p(1) \) and 
\[
k_n^{-1} \Delta_n^{-2} \sum_{j=1}^{k_n} r_{i(\tau) - j}^4 \xrightarrow{p} 3\sigma_\tau^4.
\]
Recall that 
\[
\frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} r_{i(\tau) - j}^2 \xrightarrow{p} \sigma_\tau^2.
\]
Hence, 
\[
\sum_{n,\tau}^{(22)} \xrightarrow{p} 2\sigma_\tau^4
\]
as asserted. The proof for (B.7) is now complete.

Step 3. We finish the proof of part (b) of Theorem B.1.1 in this step. By essentially the same argument for (B.5), we can show that
\[
\sqrt{k_n}(\hat{m}_{n,\tau} - \hat{m}_{n,\tau}, \hat{\sigma}_{n,\tau}^* - \hat{\sigma}_{n,\tau}) \xrightarrow{L_\theta} (\eta_\tau, \eta'_\tau).
\]
Moreover, since the variables in (B.5) and (B.11) are \( \mathcal{G} \)-conditionally independent between the pre- and post-event windows and across \( \tau \in \mathcal{T} \) (by the construction of the local i.i.d. resampling scheme), the joint convergence holds as well. Therefore,
\[
\sqrt{k_n}(\tilde{S}_n^* - \tilde{S}_n) \xrightarrow{L_\theta} (\eta_{\tau}, \eta_\tau, \eta'_{\tau}, \eta''_{\tau}, 0)_{\tau \in \mathcal{T}}.
\]
Note that, with probability approaching one, \( \tilde{S}_n \) falls in the neighborhood of \( \tilde{S} \) on which \( f \) is continuously differentiable. The assertion of the theorem then follows from the delta method.

Q.E.D.

A.2 Additional data description

A.2.1 Macroeconomic news announcements

A.2.2 A detailed description for the FOMC sentiment measure

Our detailed analysis of the volume-volatility relationship around FOMC announcements in Section 2.6.4 uses a textual measure of the negative sentiment in the actual
Table A.1: Other macroeconomic announcements

<table>
<thead>
<tr>
<th>Category by Bloomberg</th>
<th>Index</th>
<th>No.Obs.</th>
<th>Time†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor Market</td>
<td>JOLTS Job Openings</td>
<td>47</td>
<td>10:00</td>
</tr>
<tr>
<td>Retail and Wholesale</td>
<td>Wards Total Vehicle Sales</td>
<td>10</td>
<td>14:41</td>
</tr>
<tr>
<td>Retail and Wholesale</td>
<td>Wholesale Inventories</td>
<td>160</td>
<td>10:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>Chicago Purchasing Manager</td>
<td>69</td>
<td>10:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>Dallas Fed Manf Activity</td>
<td>68</td>
<td>10:30</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>ISM Milwaukee</td>
<td>58</td>
<td>10:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>Kansas City Fed Manf Activity</td>
<td>37</td>
<td>11:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>Leading Index</td>
<td>161</td>
<td>10:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>Philadelphia Fed Business Outlook</td>
<td>162</td>
<td>10:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>Richmond Fed Manf Index</td>
<td>108</td>
<td>10:00</td>
</tr>
<tr>
<td>Cyclical Indicators</td>
<td>U of Mich Sentiment</td>
<td>78</td>
<td>10:00</td>
</tr>
<tr>
<td>Personal Household Sector</td>
<td>Consumer Credit</td>
<td>159</td>
<td>15:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>Existing Home Sales</td>
<td>116</td>
<td>10:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>FHA House Price Index</td>
<td>56</td>
<td>10:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>House Price Purchase Index</td>
<td>16</td>
<td>10:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>NAHB Housing Market Index</td>
<td>138</td>
<td>13:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>New Home Sales</td>
<td>160</td>
<td>10:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>One Family Home Resales</td>
<td>47</td>
<td>10:00</td>
</tr>
<tr>
<td>Housing and Real Estate</td>
<td>Pending Home Sales</td>
<td>114</td>
<td>10:00</td>
</tr>
<tr>
<td>Government Finance And Debt</td>
<td>Monthly Budget Statement</td>
<td>155</td>
<td>14:00</td>
</tr>
</tbody>
</table>

Notes: The table lists all of the intraday news announcements included in the “Others” category.†Some of these release times vary over the sample. To ensure that the pre- and post-event windows both have 30 observations, we only keep announcements between 10:00 and 15:30. The times indicated in the table refer to the most common release times.

written statements accompanying each of the FOMC rate decisions. Our construction of this sentiment measure closely follows Loughran and McDonald (2011), henceforth LM.

Specifically, we begin by extracting the actual text for each of the FOMC statements in our sample, excluding the paragraph detailing the voting decisions and the paragraph stating the new target funds rate. The individual words in each of the statements are then compared with the Fin-Neg list of LM, resulting in a total of 79 different negative words. Table A.2 provides a count of ten such words that occur most frequently over the full sample. The count of words that are of different forms of the same word (i.e., inflections) are summed up. For example, the 65 count for the
Table A.2: Word counts in FOMC Sentiment

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
<th>unemployment</th>
<th>slow</th>
<th>weak</th>
<th>decline</th>
<th>against</th>
<th>depressed</th>
<th>downward</th>
<th>diminish</th>
<th>concern</th>
<th>persist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>73</td>
<td>65</td>
<td>53</td>
<td>39</td>
<td>23</td>
<td>23</td>
<td>21</td>
<td>18</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>

Notes: The table reports the counts of the ten most frequently occurring financial-negative words in the FOMC statements over the April, 2001 to September, 2014 sample period.

The word “slow” also includes occurrences of “slow,” “slowed,” “slower,” “slowing,” and “slowly.”

Following LM, we then construct a weighted measure for the $i$th negative word in the $j$th statement according to the following formula:

$$w_{i,j} = \begin{cases} 
\frac{1 + \log(tf_{i,j})}{1 + \log(a_j)} \log\left(\frac{N}{df_i}\right), & \text{for } t_{f_{i,j}} \geq 1, \\
0, & \text{otherwise}
\end{cases} \quad (A.19)$$

where $tf_{i,j}$ refers to the raw count of the $i$th word in the $j$th statement, $a_j$ denotes the total number of words in the $j$th statement, and $df_i$ is the number of statements out of the total number of statements $N$ containing at least one occurrence of the $i$th word. Importantly, this non-linear weighting scheme implies that negative words that seldom occur will receive a higher weight than more commonly used negative words. The overall negative sentiment for the $j$th statement is then simply obtained by summing these individual weights over all of the 79 negative words that occurred over the full sample, that is,

$$\text{FOMC Sentiment}_j = \sum_{i=1}^{79} w_{i,j}. \quad (A.20)$$

The words included in the Fin-Neg list are, of course, somewhat subjective. Also, certain words do not necessarily have a negative meaning in the context of FOMC announcements. For example, the most frequently used word “unemployment,” which
may sound negative in the 10-K filings analyzed by LM, may be a neutral word in the FOMC statements, simply used to summarize the economic conditions, whether good or bad. Similarly, the word “late” does not have a negative meaning when used in a sentence like: “... are likely to warrant exceptionally low levels for the federal funds rate at least through late 2014.” To guard against such ambiguities, we explicitly checked the usage of each of the words on the Fin-Neg list in the FOMC statements, and then select only those occurrences whose meaning were unambiguously negative. The resulting “selective” measure based on a total of 67 different words is highly correlated with FOMC Sentiment (correlation coefficient of 0.73). All of the results based on this more selective sentiment measure are also very close to those based on the original FOMC Sentiment measure reported in Section 2.6.4 of the main text. These additional results are available upon request.

A.2.3 Volume intensity sorted by five-minute normalized returns

Figure A.1 below shows that the patterns seen in Figure 2.6 in the main text is robust to changing the sampling frequency from one minute to five minutes.

A.2.4 Raw estimates without using the DID method

Tables A.3, A.4 and A.5, below, are the “no-DID” counterparts of Tables 2.4, 2.5 and 2.6 in the main text, respectively. Here, we report the “raw” estimates \( \hat{\theta} \) (see equation (2.9)), instead of \( \tilde{\theta} \) obtained using the DID method. The raw estimates of the intercept terms in Tables A.3 and A.4 are often (counterintuitive) negative, reaffirming the pattern reported in Table 2.3 of the main text, and further highlighting
Figure A.1: Sorted volume around FOMC announcements

Note: The figure shows the pre- and post-event log volume intensities (in shares) sorted on the basis of the 5-minute normalized returns $r_{i(\tau)}/\hat{\sigma}_{\tau-}\sqrt{\Delta_n}$, $\Delta_n = 5$ minutes, around FOMC announcements (dots). The normalized return increases from left to right. Announcements with normalized returns less than 1 are highlighted by the shaded area.

the importance of correcting the intraday U-shape pattern using the DID approach as done in the main text.

A.2.5 Results for DIA ETF

Tables A.6 and A.7, below, are counterparts of Tables 2.5 and 2.6 in the main text, with the SPY ETF replaced by the DIA ETF, designed to track the Dow Jones Industrial Average index. Our main findings using the SPY are re-affirmed using the DIA returns. Firstly, the volume-volatility elasticity is significantly below unity when we
Table A.3: Volume-volatility elasticities around public news announcements (no DID)

<table>
<thead>
<tr>
<th>Constant ($a_0$)</th>
<th>FOMC</th>
<th>ISMM</th>
<th>ISMN</th>
<th>CC</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.689**</td>
<td>0.043**</td>
<td>0.096**</td>
<td>0.029</td>
<td>0.065**</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.018)</td>
<td>(0.016)</td>
<td>(0.018)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.662**</td>
<td>0.614**</td>
<td>0.704**</td>
<td>0.681**</td>
<td>0.857**</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.062)</td>
<td>(0.069)</td>
<td>(0.074)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.317</td>
<td>0.216</td>
<td>0.256</td>
<td>0.185</td>
<td>0.313</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the “no-DID” regression in equation (2.9) for the specification $\Delta \log(p_m) = a_0 + a_1 \cdot \Delta \log(\sigma)$. The sample spans the period from April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. ** indicates significance at the 1% level.

Do not control for the level of disagreement. Secondly, the proxies of disagreement impact negatively the volume-volatility elasticity, although the Dispersion variable does not have a significant effect for the DIA data. Thirdly, after controlling for disagreement in the elasticity specification, the baseline elasticity is not statistically different from the no-disagreement benchmark of unity. Finally, for FOMC announcements, the FOMC sentiment measure negatively impacts the volume-volatility elasticity, and generally performs better than the “catch-all” Weekly Policy measure.

A.2.6 Alternative dispersion and uncertainty measures

Tables A.8-A.13 below present robustness checks in which we replace the disagreement proxies with similar alternatives as follows: (i) we use the one-year-ahead unemployment and GDP growth rates in place of the one-quarter-ahead unemployment rate when constructing the Dispersion measure; (ii) we use the Monthly Policy measure in place of the Weekly Policy measure. As the tables show, our findings in the main text are generally robustness to all of these changes.
Table A.4: Volume-volatility elasticity estimates and disagreement measures (no DID)

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>-0.067**</td>
<td>-0.068**</td>
<td>-0.069**</td>
<td>-0.070**</td>
<td>-0.070**</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.800**</td>
<td>0.858**</td>
<td>0.964**</td>
<td>0.952**</td>
<td>1.010**</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.022)</td>
<td>(0.041)</td>
<td>(0.034)</td>
<td>(0.044)</td>
</tr>
</tbody>
</table>

| Estimates for explanatory variables in elasticity ($b_1$): |   |   |   |   |   |
| News-category dummy variables: |   |   |   |   |   |
| FOMC                 | -0.196** | -0.199** | -0.191** | -0.194** |
|                     | (0.065)    | (0.065)    | (0.066)    | (0.065)    |
| ISMM                | -0.238** | -0.237** | -0.233** | -0.233** |
|                     | (0.067)    | (0.067)    | (0.068)    | (0.068)    |
| ISMNM               | -0.153*   | -0.153*   | -0.149*   | -0.150*   |
|                     | (0.073)    | (0.073)    | (0.073)    | (0.073)    |
| CC                  | -0.171*   | -0.177*   | -0.160*   | -0.166*   |
|                     | (0.073)    | (0.073)    | (0.073)    | (0.074)    |

| Disagreement measures: |   |   |   |   |   |
| Dispersion           | -0.041** | -0.030*  |
|                     | (0.013)    | (0.014)    |
| Weekly Policy        | -0.052** | -0.042*  |
|                     | (0.013)    | (0.014)    |

| $R^2$ | 0.539 | 0.54 | 0.541 | 0.541 | 0.542 |

Notes: The table reports the results from the “no-DID” regression in (2.9) for the specification $\Delta \log(p_m^\tau) - a_0 + b_0^J X_{0,\tau} + (a_1 + b_1^J X_{1,\tau}) \cdot \Delta \log(\sigma_\tau)$ based on all of the public announcements. In all specifications, $X_{0,\tau}$ includes category dummy variables for FOMC rate decision (FOMC), ISM Manufacturing Index (ISMM), ISM Non-Manufacturing Index (ISMNM) and Consumer Confidence Index (CC); the estimates of these dummies (i.e., $b_0$) are not reported for brevity. The Dispersion variable is constructed as the latest forecast dispersion of the one-quarter-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. The Weekly Policy variable is constructed as the weekly moving average of the economic policy uncertainty index developed by Baker et al. (2015) before the announcement. Both variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.

A.2.7 Results using dispersion in interest rate forecasts

After each FOMC meeting, the interest rate is announced in the FOMC release. Bloomberg reports a forecast dispersion for this interest rate announcement. In Table A.14 below, we include this dispersion variable (“Bloomberg Dispersion”) in the explanatory variable $X_{1,\tau}$ as a proxy of disagreement, and repeat the estimation in Table 6 in the main text. (Since the Bloomberg Dispersion variable is only available
Table A.5: Volume-volatility elasticity estimates around FOMC announcements (no DID)

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th>0.689**</th>
<th>0.692**</th>
<th>0.690**</th>
<th>0.641**</th>
<th>0.653**</th>
<th>0.642**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ( a_0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.076)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ( a_1 )</td>
<td>0.662**</td>
<td>0.897**</td>
<td>0.712**</td>
<td>0.865**</td>
<td>1.011**</td>
<td>0.976**</td>
</tr>
<tr>
<td>(0.063)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimates for explanatory variables in elasticity \( b_1 \):

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion -0.095**</td>
<td>-0.075**</td>
<td>-0.085**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Policy -0.027</td>
<td></td>
<td></td>
<td></td>
<td>0.055*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment -0.105**</td>
<td>-0.084**</td>
<td>-0.106**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( R^2 \)

<table>
<thead>
<tr>
<th></th>
<th>0.317</th>
<th>0.362</th>
<th>0.314</th>
<th>0.363</th>
<th>0.388</th>
<th>0.392</th>
</tr>
</thead>
</table>

Notes: The table reports the results from the “no-DID” regression in equation (2.9) for the specification \( \Delta \log(m_{\tau}) = a_0 + (a_1 + b_1 X_{1,\tau}) \cdot \Delta \log(\sigma_{\tau}) \) based on FOMC announcements. Dispersion and Weekly Policy are constructed as in Table 2.5. FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. These variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.

For the FOMC meetings, the announcement on April 18, 2001 is removed in this exercise.) For the ease of comparison, we include the specifications in Table 6 in Columns 1-6. The last two columns of Table A.14 shows the results with Bloomberg Dispersion included. We see that this variable is statistically insignificant; moreover, adding this variable results in little change in our overall estimation results.

A.2.8 Robustness check for removing the FOMC announcement on April 18, 2001

Among the 109 FOMC announcements in our sample, 108 are released after FOMC meetings, and one is released after a conference call on April 18, 2001. Here, we repeat our analysis for the FOMC sample (see Table 6 in the main text) with this
Table A.6: Volume-volatility elasticity estimates and disagreement measures (DIA)

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (a_0)</td>
<td>0.056**</td>
<td>0.057**</td>
<td>0.057**</td>
<td>0.057**</td>
<td>0.057**</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Elasticity (a_1)</td>
<td>0.786**</td>
<td>0.775**</td>
<td>0.853**</td>
<td>0.948**</td>
<td>0.956**</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.031)</td>
<td>(0.062)</td>
<td>(0.052)</td>
<td>(0.068)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity (b_1):</th>
</tr>
</thead>
<tbody>
<tr>
<td>News-category dummy variables:</td>
</tr>
<tr>
<td>FOMC</td>
</tr>
<tr>
<td>0.086</td>
</tr>
<tr>
<td>(0.096)</td>
</tr>
<tr>
<td>ISMM</td>
</tr>
<tr>
<td>0.265**</td>
</tr>
<tr>
<td>(0.101)</td>
</tr>
<tr>
<td>ISMNM</td>
</tr>
<tr>
<td>-0.128</td>
</tr>
<tr>
<td>(0.102)</td>
</tr>
<tr>
<td>CC</td>
</tr>
<tr>
<td>-0.114</td>
</tr>
<tr>
<td>(0.106)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disagreement measures:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
</tr>
<tr>
<td>-0.030</td>
</tr>
<tr>
<td>(0.019)</td>
</tr>
<tr>
<td>Weekly Policy</td>
</tr>
<tr>
<td>-0.096**</td>
</tr>
<tr>
<td>(0.019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.364</td>
</tr>
<tr>
<td>0.365</td>
</tr>
<tr>
<td>0.365</td>
</tr>
<tr>
<td>0.368</td>
</tr>
<tr>
<td>0.367</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification \(\Delta \log(p_{m,t}) - a_0 + b_0^J X_{0,t} + (a_1 + b_1^J X_{1,t}) \cdot \Delta \log(\sigma_t)\) based on all of the public announcements, using the past 22 non-announcement days as the control group. In all specifications, \(X_{0,t}\) includes category dummy variables for FOMC rate decision (FOMC), ISM Manufacturing Index (ISMM), ISM Non-Manufacturing Index (ISMNM) and Consumer Confidence Index (CC); the estimates of \(b_0\) are not reported for brevity. The Dispersion variable is constructed as the latest forecast dispersion of the one-quarter-ahead unemployment rate from the Survey of Professional Forecasters before the announcement. The Weekly Policy variable is constructed as the weekly moving average of the economic policy uncertainty index developed by Baker et al. (2015) before the announcement. Both variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.

A.2.9 Robustness checks with respect to the local window size \(k_n\)

In this subsection, we show that our main findings reported in Table 5 and Table 6 in the main text are robust to perturbations in the local window size \(k_n\). We repeat
Table A.7: Volume-volatility elasticity estimates around FOMC announcements (DIA)

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th>0.621**</th>
<th>0.617**</th>
<th>0.623**</th>
<th>0.597**</th>
<th>0.598**</th>
<th>0.592**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ((a_0))</td>
<td>(0.098)</td>
<td>(0.098)</td>
<td>(0.099)</td>
<td>(0.099)</td>
<td>(0.099)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>Elasticity ((a_1))</td>
<td>0.861**</td>
<td>1.004**</td>
<td>0.903**</td>
<td>1.013**</td>
<td>1.081**</td>
<td>1.060**</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.119)</td>
<td>(0.108)</td>
<td>(0.106)</td>
<td>(0.126)</td>
<td>(0.123)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ((b_1)):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
</tr>
<tr>
<td>-0.055</td>
</tr>
<tr>
<td>(0.030)</td>
</tr>
<tr>
<td>Weekly Policy</td>
</tr>
<tr>
<td>-0.023</td>
</tr>
<tr>
<td>(0.032)</td>
</tr>
<tr>
<td>FOMC Sentiment</td>
</tr>
<tr>
<td>-0.085**</td>
</tr>
<tr>
<td>(0.032)</td>
</tr>
<tr>
<td>(R^2)</td>
</tr>
<tr>
<td>0.298</td>
</tr>
<tr>
<td>0.300</td>
</tr>
<tr>
<td>0.293</td>
</tr>
<tr>
<td>0.310</td>
</tr>
<tr>
<td>0.306</td>
</tr>
<tr>
<td>0.301</td>
</tr>
</tbody>
</table>

Notes: The table reports the results from the DID jump regression in equation (3.14) for the specification \(\Delta \log (m_t) = a_0 + (a_1 + b_1^\top X_{1,t}) \cdot \Delta \log (\sigma_t)\) based on FOMC announcements, using the past 22 non-announcement days as the control group. Dispersion and Weekly Policy are constructed as in Table 2.5. FOMC Sentiment is a textual measure constructed using financial-negative words in the FOMC press release. These variables are scaled by their own sample standard deviations. The sample spans April 10, 2001 to September 30, 2014. Bootstrapped standard errors (1000 repetitions) are reported in parentheses. * and ** indicate significance at the 5% and 1% level, respectively.

the estimations in Tables 5 and 6, but replacing \(k_n = 30\) with \(k_n = 25\). Our main empirical findings remain the same. The details are given by Table A.16 and Table A.17 below.

A.2.10 Estimation results with the 1-minute observations before announcements removed

As a simple robustness check with respect to the pre-trend documented by Bernile et al. (2016) and Kurov et al. (2016), below, we estimate the pre-event quantities excluding the one-minute observations immediately before the announcements, and
Table A.8: Volume-volatility elasticity estimates and disagreement measures based on one-year-ahead unemployment dispersion

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td>0.044** 0.041** 0.041** 0.041** 0.041**</td>
<td>(0.007) (0.007) (0.007) (0.007) (0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elasticity ($a_1$)</td>
<td>0.733** 0.776** 0.921** 0.921** 0.990**</td>
<td>(0.020) (0.024) (0.043) (0.035) (0.046)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimates for explanatory variables in elasticity ($b_1$):</td>
<td>News-category dummy variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOMC</td>
<td>-0.060 -0.065 -0.048 -0.053</td>
<td>(0.073) (0.072) (0.073) (0.073)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISMM</td>
<td>-0.238** -0.244** -0.220** -0.226**</td>
<td>(0.067) (0.067) (0.067) (0.067)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISMNMM</td>
<td>-0.090 -0.086 -0.082 -0.081</td>
<td>(0.073) (0.073) (0.072) (0.073)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>-0.244** -0.248** -0.202** -0.210**</td>
<td>(0.077) (0.078) (0.078)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dispersion</td>
<td>-0.055** -0.033*</td>
<td>(0.013) (0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weekly Policy</td>
<td>-0.079** -0.069**</td>
<td>(0.013) (0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.481 0.482 0.484 0.486 0.486</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table is constructed in the same way as Table 5 in the main text, except that the Dispersion variable is constructed as the latest forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasters before the announcement.

then repeat the estimation in the main text. This modification results in very similar estimates as the ones in Section 2.6.

A.2.11 Volume intensity and volatility for U.S. Treasury bonds

Figure A.2 below plots the estimated logarithmic volume intensities (top panel) and logarithmic spot volatilities (bottom panel) for the U.S. ten-year T-bond futures contract before and after FOMC announcements.
Table A.9: Volume-volatility elasticity estimates around FOMC announcements based on one-year-ahead unemployment dispersion

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th>Estimates for explanatory variables in elasticity (b₁):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td>Elasticity ($a_1$)</td>
</tr>
<tr>
<td></td>
<td>$0.586^{<strong>}$ 0.592^{</strong>} 0.583^{<strong>} 0.545^{</strong>} 0.557^{<strong>} 0.555^{</strong>}</td>
<td>$0.716^{<strong>}$ 0.911^{</strong>} 0.790^{<strong>} 0.913^{</strong>} 0.986^{<strong>} 0.960^{</strong>}</td>
</tr>
<tr>
<td></td>
<td>$(0.076)$ $(0.076)$ $(0.076)$ $(0.075)$ $(0.076)$ $(0.076)$</td>
<td>$(0.065)$ $(0.083)$ $(0.081)$ $(0.081)$ $(0.088)$ $(0.089)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dispersion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$-0.077^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.019)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weekly Policy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$-0.037$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.021)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOMC Sentiment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$-0.104^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(0.022)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$R^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.330$ 0.354 0.329 0.370 0.373 0.37$</td>
</tr>
</tbody>
</table>

Notes: This table is constructed in the same way as Table 6 in the main text, except that the Dispersion variable is constructed as the latest forecast dispersion of the one-year-ahead unemployment rate from the Survey of Professional Forecasts before the announcement.
Table A.10: Volume-volatility elasticity estimates and disagreement measures based on one-year-ahead GDP dispersion

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
</tr>
</thead>
<tbody>
<tr>
<td>News-category dummy variables:</td>
</tr>
<tr>
<td>FOMC</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ISMM</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ISMNMM</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CC</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Disagreement measures:</td>
</tr>
<tr>
<td>Dispersion</td>
</tr>
<tr>
<td>Weekly Policy</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
</tbody>
</table>

Notes: This table is constructed in the same way as Table 5 in the main text, except that the Dispersion variable is constructed as the latest forecast dispersion of the one-year-ahead GDP growth rate from the Survey of Professional Forecasters before the announcement.
Table A.11: Volume-volatility elasticity estimates around FOMC announcements based on one-year-ahead GDP dispersion

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td>Dispersion</td>
</tr>
<tr>
<td></td>
<td>0.586** 0.582** 0.583** 0.545** 0.546** 0.539**</td>
<td>-0.073**</td>
</tr>
<tr>
<td></td>
<td>(0.078) (0.077) (0.078) (0.078) (0.078) (0.078)</td>
<td>(0.021)</td>
</tr>
<tr>
<td></td>
<td>Elasticity ($a_1$)</td>
<td>Weekly Policy</td>
</tr>
<tr>
<td></td>
<td>0.716** 1.016** 0.790** 0.913** 1.148** 1.154**</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>(0.067) (0.102) (0.080) (0.083) (0.108) (0.108)</td>
<td>(0.020)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOMC Sentiment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.104** -0.096** -0.121**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.022) (0.023) (0.027)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$R^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.33 0.349 0.329 0.37 0.382 0.385</td>
</tr>
</tbody>
</table>

Notes: This table is constructed in the same way as Table 6 in the main text, except that the Dispersion variable is constructed as the latest forecast dispersion of the one-year-ahead GDP growth rate from the Survey of Professional Forecasters before the announcement.
Table A.12: Volume-volatility elasticity estimates and disagreement measures based on Monthly Policy

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th>Disagreement measures:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.044**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.040**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.733**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.776**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.906**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.945**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.001**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>News-category dummy variables:</td>
<td>-0.060</td>
<td>-0.058</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.072)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>FOMC</td>
<td>-0.238**</td>
<td>-0.238**</td>
<td>-0.217**</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.070)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>ISMM</td>
<td>-0.090</td>
<td>-0.092</td>
<td>-0.086</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.076)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>ISMNM</td>
<td>-0.244**</td>
<td>-0.244**</td>
<td>-0.227**</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.072)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>CC</td>
<td>-0.051**</td>
<td>-0.029*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Disagreement measures:</td>
<td>Monthly Policy</td>
<td>-0.083**</td>
<td>-0.074**</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Monthly Policy</td>
<td>0.481</td>
<td>0.482</td>
<td>0.483</td>
</tr>
</tbody>
</table>

Notes: This table is constructed in the same way as Table 5 in the main text, except that we replace “Weekly Policy” with “Monthly Policy” that is constructed as the monthly moving average of the economic policy uncertainty index developed by Baker et al. (2015) before the announcement.
Table A.13: Volume-volatility elasticity estimates around FOMC announcements based on Monthly Policy

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>0.586**</td>
<td>0.579**</td>
<td>0.586**</td>
<td>0.545**</td>
<td>0.550**</td>
<td>0.539**</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.078)</td>
<td>(0.079)</td>
<td>(0.078)</td>
<td>(0.078)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.716**</td>
<td>0.996**</td>
<td>0.799**</td>
<td>0.913**</td>
<td>1.089**</td>
<td>1.059**</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.087)</td>
<td>(0.076)</td>
<td>(0.080)</td>
<td>(0.092)</td>
<td>(0.090)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
<td>-0.108**</td>
<td>-0.088**</td>
<td>-0.099**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly Policy</td>
<td>-0.040*</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment</td>
<td>-0.104**</td>
<td>-0.077**</td>
<td>-0.103**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ | 0.33 | 0.382 | 0.331 | 0.37 | 0.4 | 0.402

Notes: This table is constructed in the same way as Table 6 in the main text, except that we replace “Weekly Policy” with “Monthly Policy” that is constructed as the monthly moving average of the economic policy uncertainty index developed by Baker et al. (2015) before the announcement.

Table A.14: Volume-volatility elasticity estimates around FOMC announcements

<table>
<thead>
<tr>
<th>Baseline estimates:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>0.568**</td>
<td>0.570**</td>
<td>0.562**</td>
<td>0.536**</td>
<td>0.545**</td>
<td>0.544**</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.076)</td>
<td>(0.077)</td>
<td>(0.075)</td>
<td>(0.075)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.737**</td>
<td>0.999**</td>
<td>0.830**</td>
<td>0.920**</td>
<td>1.089**</td>
<td>1.056**</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.082)</td>
<td>(0.079)</td>
<td>(0.077)</td>
<td>(0.087)</td>
<td>(0.088)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
<td>-0.105**</td>
<td>-0.086**</td>
<td>-0.095**</td>
<td>-0.085**</td>
<td>-0.094**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.023)</td>
<td>(0.021)</td>
<td>(0.023)</td>
<td></td>
</tr>
<tr>
<td>Weekly Policy</td>
<td>-0.045*</td>
<td>0.045</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.030)</td>
<td>(0.030)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment</td>
<td>-0.101**</td>
<td>-0.076**</td>
<td>-0.096**</td>
<td>-0.072**</td>
<td>-0.091**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.027)</td>
<td>(0.023)</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td>Bloomberg Dispersion</td>
<td>0.030</td>
<td>0.021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ | 0.336 | 0.383 | 0.337 | 0.372 | 0.399 | 0.399 | 0.397 | 0.395

Notes: This table is constructed in the same way as Table 6 in that main text, except that we include “Bloomberg Dispersion” as an additional explanatory variable in $X_{1,r}$. Since this dispersion variable is only available for FOMC meetings, we remove the conference-call observation on April 18, 2001 from our sample.
Table A.15: Volume-volatility elasticity estimates around FOMC announcements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ($a_0$)</td>
<td>0.568**</td>
<td>0.570**</td>
<td>0.562**</td>
<td>0.536**</td>
<td>0.545**</td>
<td>0.544**</td>
<td>0.079</td>
<td>0.077</td>
<td>0.077</td>
<td>0.077</td>
<td>0.077</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.737**</td>
<td>0.999**</td>
<td>0.830**</td>
<td>0.920**</td>
<td>1.089**</td>
<td>1.056**</td>
<td>0.068</td>
<td>0.082</td>
<td>0.081</td>
<td>0.079</td>
<td>0.086</td>
</tr>
<tr>
<td><strong>Estimates for explanatory variables in elasticity ($b_1$):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispersion</td>
<td>-0.105**</td>
<td>-0.086**</td>
<td>-0.095**</td>
<td>0.019</td>
<td>0.020</td>
<td>0.021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Policy</td>
<td>-0.045*</td>
<td>0.45</td>
<td></td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment</td>
<td>-0.101**</td>
<td>-0.076**</td>
<td>-0.096**</td>
<td>0.022</td>
<td>0.022</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.336</td>
<td>0.383</td>
<td>0.337</td>
<td>0.372</td>
<td>0.399</td>
<td>0.399</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table is constructed in the same way as Table 6 in the main text, except that we remove the announcement on April 18, 2001.

Figure A.2: T-bond volume and volatility around FOMC announcements

Notes: The figure plots the log-volume intensity (top panel, in shares) and log-volatility (bottom panel, in percentage) around FOMC announcements. The volume intensity and volatility are calculated using equation (3.9) with $k_n = 30$. 

164
Table A.16: Volume-volatility elasticity estimates and disagreement measures

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ($a_0$)</td>
<td>0.050**</td>
<td>0.046**</td>
<td>0.046**</td>
<td>0.047**</td>
<td>0.047**</td>
</tr>
<tr>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.698**</td>
<td>0.736**</td>
<td>0.834**</td>
<td>0.865**</td>
<td>0.908**</td>
</tr>
<tr>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Estimates for explanatory variables in elasticity ($b_1$): |
| **News-category dummy variables:** |
| FOMC                   | -0.074 | -0.069 | -0.071 | -0.069 |
| (0.071)                |       |       | (0.071) |       |
| ISMM                   | -0.186** | -0.189** | -0.173* | -0.176* |
| (0.072)                |       |       | (0.072) |       |
| ISMNMM                 | -0.094 | -0.099 | -0.089 | -0.093 |
| (0.072)                |       |       | (0.072) |       |
| CC                     | -0.214** | -0.217** | -0.177* | -0.182* |
| (0.073)                |       |       | (0.074) |       |

| **Disagreement measures:**   |       |       |       |       |       |
| Dispersion                  | -0.038** |       | -0.021 |       |
| (0.014)                     |       |       | (0.014) |       |
| Weekly Policy               |       | -0.070** | -0.064** |       |
| (0.013)                     |       |       | (0.014) |       |

| $R^2$                       | 0.473 | 0.474 | 0.474 | 0.477 | 0.477 |

Notes: This table is constructed in the same way as Table 5 in the main text, except that we change the local window $k_n$ from 30 to 25.
Table A.17: Volume-volatility elasticity estimates around FOMC announcements

<table>
<thead>
<tr>
<th></th>
<th>FOMC</th>
<th>ISMM</th>
<th>ISMNM</th>
<th>CC</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>0.665**</td>
<td>0.649**</td>
<td>0.665**</td>
<td>0.634**</td>
<td>0.630**</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.079)</td>
<td>(0.080)</td>
<td>(0.079)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.662**</td>
<td>0.941**</td>
<td>0.756**</td>
<td>0.830**</td>
<td>1.014**</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.088)</td>
<td>(0.082)</td>
<td>(0.081)</td>
<td>(0.094)</td>
</tr>
</tbody>
</table>

Estimates for explanatory variables in elasticity ($b_1$):

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
<td>-0.104**</td>
<td>-0.087**</td>
<td>-0.091**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Policy</td>
<td>-0.049*</td>
<td>0.021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment</td>
<td>-0.092**</td>
<td>-0.064**</td>
<td>-0.072**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.026)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ 0.323 0.375 0.327 0.354 0.386 0.382

Notes: This table is constructed in the same way as Table 6 in the main text, except that we change the local window $k_n$ from 30 to 25.

Table A.18: Volume-volatility elasticities around public news announcements

<table>
<thead>
<tr>
<th></th>
<th>FOMC</th>
<th>ISMM</th>
<th>ISMNM</th>
<th>CC</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($a_0$)</td>
<td>0.505**</td>
<td>0.202**</td>
<td>0.113**</td>
<td>0.214**</td>
<td>0.048**</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.023)</td>
<td>(0.019)</td>
<td>(0.023)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Elasticity ($a_1$)</td>
<td>0.785**</td>
<td>0.528**</td>
<td>0.717**</td>
<td>0.540**</td>
<td>0.789**</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.064)</td>
<td>(0.071)</td>
<td>(0.074)</td>
<td>(0.022)</td>
</tr>
</tbody>
</table>

$R^2$ 0.357 0.153 0.237 0.113 0.290

Notes: The table is constructed as Table 2.4 in the main text, except that we remove the 1-minute observations immediately before the announcements in the estimation of pre-event volume intensity and volatility.
Table A.19: Volume-volatility elasticity estimates and disagreement measures

<table>
<thead>
<tr>
<th></th>
<th>Estimate 1</th>
<th>Estimate 2</th>
<th>Estimate 3</th>
<th>Estimate 4</th>
<th>Estimate 5</th>
<th>( \text{SE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline estimates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ((a_0))</td>
<td>0.042**</td>
<td>0.039**</td>
<td>0.039**</td>
<td>0.040**</td>
<td>0.040**</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Elasticity ((a_1))</td>
<td>0.742**</td>
<td>0.776**</td>
<td>0.891**</td>
<td>0.926**</td>
<td>0.979**</td>
<td>(0.020)</td>
</tr>
<tr>
<td><strong>Estimates for explanatory variables in elasticity ((b_1)):</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News-category dummy variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOMC</td>
<td>0.010</td>
<td>0.004</td>
<td>0.013</td>
<td>0.010</td>
<td></td>
<td>(0.070)</td>
</tr>
<tr>
<td>ISMM</td>
<td>-0.239**</td>
<td>-0.239**</td>
<td>-0.221**</td>
<td>-0.222**</td>
<td></td>
<td>(0.070)</td>
</tr>
<tr>
<td>ISMNM</td>
<td>-0.060</td>
<td>-0.062</td>
<td>-0.052</td>
<td>-0.054</td>
<td></td>
<td>(0.073)</td>
</tr>
<tr>
<td>CC</td>
<td>-0.224**</td>
<td>-0.222**</td>
<td>-0.178*</td>
<td>-0.181*</td>
<td></td>
<td>(0.075)</td>
</tr>
<tr>
<td><strong>Disagreement measures:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispersion</td>
<td>-0.045**</td>
<td></td>
<td>-0.026</td>
<td></td>
<td></td>
<td>(0.013)</td>
</tr>
<tr>
<td>Weekly Policy</td>
<td></td>
<td>-0.082**</td>
<td>-0.075**</td>
<td></td>
<td></td>
<td>(0.013)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.500</td>
<td>0.501</td>
<td>0.502</td>
<td>0.505</td>
<td>0.505</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The table is constructed as Table 2.5 in the main text, except that we remove the 1-minute observations immediately before the announcements in the estimation of pre-event volume intensity and volatility.
Table A.20: Volume-volatility elasticity estimates around FOMC announcements

<table>
<thead>
<tr>
<th></th>
<th>Baseline estimates:</th>
<th>Estimates for explanatory variables in elasticity ($b_1$):</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant ($a_0$)</td>
<td>Elasticity ($a_1$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.505**</td>
<td>0.785**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.068)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.512**</td>
<td>1.017**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.083)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.506**</td>
<td>0.876**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.077)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.482**</td>
<td>0.946**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.077)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.493**</td>
<td>1.089**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.492**</td>
<td>1.081**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td>Dispersion</td>
<td>-0.096**</td>
<td>-0.076**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Weekly Policy</td>
<td>-0.049**</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.023)</td>
<td></td>
</tr>
<tr>
<td>FOMC Sentiment</td>
<td>-0.095**</td>
<td>-0.071**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.357</td>
<td>0.403</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.362</td>
<td>0.396</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.421</td>
<td>0.416</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table is constructed as Table 2.6 in the main text, except that we remove the 1-minute observations immediately before the announcements in the estimation of pre-event volume intensity and volatility.
Appendix B

Appendix for Chapter 3

B.1 Proofs for the econometric methods

This appendix section proves the validity of the econometric methods and bootstraps proposed in Section 2.5.

B.1.1 Estimation and inference

We now begin to validate the estimation and inference methods. We focus on the most general estimation, $\hat{\theta}_n$ defined by equation (3.13). We write $\hat{\theta}_n = f(\tilde{S}_n)$, where $f(\cdot)$ is defined implicitly by the definitions (3.12) and (3.13). Besides, $S \equiv (S_{\tau})_{\tau \in \{A,U,C\}}$ contains all information needed for the estimation.

The following assumptions and notations are from Bollerslev et al. (2017) except the last one, which imposes finite bound on $\beta_{it}$. In these definitions, we suppress the subscript if we refer to the asset in regardless of whether it is the market ETF or an individual stock.
Assumption 5. (i) The price process $P_t$ is given by (A.1) or (3.5) for $J_t = \int_0^t \xi_s dN_s + \int_0^t \int_\mathbb{R} \delta(s,z) \mu(ds,dz)$, for both the market and individual stocks. The processes $b_m, \sigma_m, b_i, \sigma_i$ are càdlàg (i.e., right continuous with left limit) and adapted; $\sigma_{mt}$ and $\sigma_{it}$ are positive for $t \in [0,T]$ almost surely; the process $\xi$ is predictable and locally bounded; $N$ is a counting process that jumps at the scheduled announcement times which are specified by the set $A$; $\delta$ is a predictable function; $\mu$ is a Poisson random measure with compensator $\nu(ds,dz) = ds \otimes \lambda(dz)$ for some finite measure $\lambda$.

(ii) The volume process $V_t$ satisfies (A.2) or (3.7). The process $\zeta$ is càdlàg and adapted. The error terms $(\epsilon_i)$ take values in some Polish space, are defined on an extension of $(\Omega, \mathcal{F})$, i.i.d. and independent of $\mathcal{F}$.

(iii) For a sequence of stopping times $(T_m)_{m \geq 1}$ increasing to infinity and constants $(K_m)_{m \geq 1}$, we have $\mathbb{E} |\sigma_{t \wedge T_m} - \sigma_{s \wedge T_m}|^2 + \mathbb{E} |\zeta_{t \wedge T_m} - \zeta_{s \wedge T_m}|^2 \leq K_m |t-s|$ for all $t, s$ such that $[s,t] \cap A = \emptyset$.

(iv) $\beta_{it}$ given by (3.5) satisfies that $\sup_t |\beta_{it}| \leq K$.

Assumption 5 is fairly standard in the study of high-frequency data. See Bollerslev et al. (2017) for detailed explanation. We add the assumption that the $\beta$ process is bounded, which is not a binding consumption as empirically $\beta$ is believed to be around 1. We also denote $M_p(\cdot) \equiv \int \mathcal{V}(\cdot, \epsilon)^p F_{\epsilon} (d\epsilon)$ for $p \geq 1$ and denote $v_{it} = M_2(\zeta_{it}) - M_1^2(\zeta_{it})$. To accommodate to the multidimensional case, we also denote $M_{2p}^{ij}(\cdot, \cdot) = \int \mathcal{V}(\zeta_{it}, \epsilon_i)^p \mathcal{V}(\zeta_{jt}, \epsilon_j)^p F_{\epsilon_i, \epsilon_j} (d\epsilon_i, d\epsilon_j)$ and denote $v_{ij}^{ij} = M_2(\zeta_{it}, \zeta_{jt}) - M_1(\zeta_{it}) M_1(\zeta_{jt})$. $F_{\epsilon_i, \epsilon_j}$ is the joint distribution of $\epsilon_i$ and $\epsilon_j$.

Assumption 6. $k_n \to \infty$ and $k_n^2 \Delta_n \to 0$. 

170
Assumption 7. (i) The function $M_1(\cdot)$ is Lipschitz on compact sets and the functions $M_2(\cdot)$, $M_2^{ij}(\cdot, \cdot)$ and $M_4(\cdot)$ are continuous.

(ii) Almost surely, the function $f$ is well-defined and continuously differentiable in a neighborhood of $\bar{S}$.

Consider variables $(\eta_{i\tau}, \eta_{i',\tau}, \eta_{i''\tau}, \eta_{i'''\tau})_{\forall \tau} \in T$ which, conditionally on $F$, are mutually independent, centered Gaussian with variances $(\nu_{i\tau}, \nu_{i',\tau}, \sigma_{i\tau}^2/2, \sigma_{i'\tau}^2/2)_{\forall \tau \in N, \forall \tau \in T}$, where $N$ is the set of stock indexes. We denote the first differential of $f$ at $\bar{S}$ with increment $d\bar{S}$ by $F(\bar{S}; d\bar{S})$. For a sequence $Y_n$ of random variables, we write $Y_n \xrightarrow{L} Y$ if $Y_n$ converges stably in law towards $Y$, meaning that $(Y_n, U)$ converges in distribution to $(Y, U)$ for any bounded $F$-measurable random variable $U$.

Assumption 6 characterizes the growth rate of $k_n$. The undersmoothing condition eliminates the bias terms asymptotically, and thus facilitates the statistical inference. With these notations, we get a multi-dimensional version of the Theorem 1 in Bollerslev et al. (2017) as follows:

**Theorem B.1.1.** (a) Under Assumptions 5, 6 and 7

\[
\sqrt{k_n}(\hat{\theta}_n - \theta) \xrightarrow{L} F(\bar{S}; (\eta_{i\tau}, \eta_{i',\tau}, \eta_{i''\tau}, 0)_{\forall \tau \in N, \forall \tau \in T}).
\]  

(B.1)

(b) Moreover, the conditional distribution function of $\sqrt{k_n}(\hat{\theta}_n^2 - \hat{\theta}_n)$ given the original data converges in probability to that of $F(\bar{S}; (\eta_{i\tau}, \eta_{i',\tau}, \eta_{i''\tau}, 0)_{\forall \tau \in N, \forall \tau \in T})$ under the uniform metric.

B.1.2 Proofs

**Proof of Theorem B.1.1(a)**
According to their Equation (A.5) and (A.8) in Bollerslev et al. (2017), for each stock \( i \),

\[
\sqrt{k_n}(\hat{m}_{n,i\tau} - m_{i\tau}, \hat{m}_{n,i\tau} - m_{i\tau}) \xrightarrow{CLT} (\eta_{i\tau}, \eta_{i\tau})_{\tau \in \mathcal{T}} \tag{B.2}
\]

\[
\sqrt{k_n}(\hat{\sigma}_{n,i\tau} - \sigma_{n,i\tau}, \hat{\sigma}_{n,i\tau} - \sigma_{n,i\tau}) \xrightarrow{CLT} (\eta'_{i\tau}, \eta'_{i\tau})_{\tau \in \mathcal{T}} \tag{B.3}
\]

By Proposition 5 of Barndorff-Nielsen et al. (2008) and the property of stable convergence in law, we can combine (B.2) and (B.3), yielding

\[
\sqrt{k_n}(\hat{S}_n - \hat{S}) \xrightarrow{CLT} (\eta_{\tau}, \eta'_{\tau}, \eta'_{\tau}, 0)_{\tau \in \mathcal{T}} \tag{B.4}
\]

The assertion (B.1) then follows from (B.4) and the delta method, since \( f \) is still a smooth function.

Q.E.D.

**Proof of Theorem B.1.1(b).** Step 1. We divide the proof into several steps. Denote \( \mathcal{G} \equiv \mathcal{F} \vee \sigma \{ \epsilon_i : i \geq 0 \} \). In this step, we show that, for each \( \tau \in \mathcal{T} \),

\[
\sqrt{k_n} \begin{pmatrix}
\hat{m}_{n,1\tau} - m_{1\tau} \\
\vdots \\
\hat{m}_{n,1\tau} - m_{1\tau} \\
\hat{\sigma}_{n,1\tau} - \sigma_{1\tau} \\
\vdots \\
\hat{\sigma}_{n,1\tau} - \sigma_{1\tau} \\
\hat{\sigma}_{n,1\tau} - \sigma_{1\tau} \\
\hat{\sigma}_{n,1\tau} - \sigma_{1\tau}
\end{pmatrix} \xrightarrow{CLT} \begin{pmatrix}
\eta_{1\tau} \\
\vdots \\
\eta_{1\tau} \\
\eta'_{1\tau} \\
\vdots \\
\eta'_{1\tau} \\
\eta'_{1\tau}
\end{pmatrix} \tag{B.5}
\]
Notice that in this section we use $I$ instead of $n$ to denote the total number of stocks, since $n$ is used for asymptotic proofs. Observe that

\[
\sqrt{k_n} \left( \begin{array}{c}
\hat{m}_{n,1\tau} - \hat{m}_{1\tau} \\
\vdots \\
\hat{m}_{n,I\tau} - \hat{m}_{I\tau} \\
\hat{\sigma}_{n,1\tau} - \hat{\sigma}_{1\tau} \\
\vdots \\
\hat{\sigma}_{n,I\tau} - \hat{\sigma}_{I\tau}
\end{array} \right) = \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} \left( \begin{array}{c}
V_{1,1(l(\tau)-j)} \Delta_n \\
\vdots \\
V_{1,1(l(\tau)-j)} \Delta_n \\
V_{1,I(l(\tau)-j)} \Delta_n \\
\vdots \\
V_{1,I(l(\tau)-j)} \Delta_n
\end{array} \right) - \frac{1}{k_n} \sum_{j=1}^{k_n} \left( \begin{array}{c}
r_{1,1(l(\tau)-j)} \Delta_n \\
r_{1,1(l(\tau)-j)} \Delta_n \\
r_{1,l(l(\tau)-j)} \Delta_n \\
r_{1,l(l(\tau)-j)} \Delta_n \\
r_{1,l(l(\tau)-j)} \Delta_n
\end{array} \right).
\]

(B.6)

We bootstrap with replacement within a small window, so the summands in the right-hand side of (B.6) are i.i.d. with zero mean conditional on $\mathcal{G}$ when the window size goes to infinitely small. We denote the $\mathcal{G}$-conditional covariance matrix of the left-hand side by

\[
\Sigma_{n,\tau} = \left( \begin{array}{cc}
H_{n,\tau} & Q_{n,\tau}
\end{array} \right)
\]

For matrix $X$, denote its $(i, j)$ element as $X^{(ij)}$, then

\[
\left\{ \begin{array}{l}
H_{n,\tau}^{(ij)} = \frac{1}{k_n} \sum_{k=1}^{k_n} V_{i,1(l(\tau)-k)} \Delta_n V_{j,1(l(\tau)-k)} \Delta_n \\
- \left( \frac{1}{k_n} \sum_{k=1}^{k_n} V_{i,1(l(\tau)-k)} \Delta_n \right) \left( \frac{1}{k_n} \sum_{k=1}^{k_n} V_{j,1(l(\tau)-k)} \Delta_n \right),
\end{array} \right.
\]

\[
Q_{n,\tau}^{(ij)} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} V_{i,1(l(\tau)-k)} \Delta_n \sum_{k=1}^{k_n} V_{j,1(l(\tau)-k)} \Delta_n \\
- \left( \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} V_{i,1(l(\tau)-k)} \Delta_n \right) \left( \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} V_{j,1(l(\tau)-k)} \Delta_n \right),
\]

\[
Z_{n,\tau}^{(ij)} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \sum_{k=1}^{k_n} r_{1,1(l(\tau)-k)}^2 \sum_{k=1}^{k_n} \sum_{k=1}^{k_n} \tau_{1,1(l(\tau)-k)}^2 \\
- \left( \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \sum_{k=1}^{k_n} \tau_{1,1(l(\tau)-k)}^2 \right) \left( \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \sum_{k=1}^{k_n} \tau_{1,1(l(\tau)-k)}^2 \right).
\]

In step 2, we shall show that

\[
\left( H_{n,\tau}^{(ij)}, Q_{n,\tau}^{(ij)}, Z_{n,\tau}^{(ij)} \right) \xrightarrow{p} \left( v_{\gamma,\tau}^{ij}, 0, 2\beta_{1,\tau}^2 \beta_{1,\tau}^2 \sigma_{m,\tau}^4 \right).
\]

(B.7)

Then we get (B.5) following exactly the same arguments as in Bollerslev et al. (2017).
Step 2. We aim to prove (B.7). Starting with \( \mathbf{H}^{(ij)}_{n,\tau} \), we do the following decomposition

\[
\frac{1}{k_n} \sum_{k=1}^{k_n} V_i(l(\tau)-k) \Delta_n V_j(l(\tau)-k) \Delta_n = A_{1,n} + A_{2,n},
\]

where

\[
A_{1,n} \equiv \frac{1}{k_n} \sum_{k=1}^{k_n} \left( V_i(l(\tau)-k) \Delta_n V_j(l(\tau)-k) \Delta_n - M_2(\zeta_i(l(\tau)-k) \Delta_n, \zeta_j(l(\tau)-k) \Delta_n) \right),
\]

\[
A_{2,n} \equiv \frac{1}{k_n} \sum_{k=1}^{k_n} M_2(\zeta_i(l(\tau)-k) \Delta_n, \zeta_j(l(\tau)-k) \Delta_n)
\]

Since each summand in \( A_{1,n} \) is \( \mathcal{F} \)-conditionally independent with zero mean, by Hölder’s inequality we get

\[
\mathbb{E}[A_{1,n}^2|\mathcal{F}] \leq K k_n^{-2} \sum_{k=1}^{k_n} \sqrt{M_2(\zeta_i(l(\tau)-k) \Delta_n) M_2(\zeta_j(l(\tau)-k) \Delta_n)} \leq K k_n^{-1} \rightarrow 0,
\]

That means, \( A_{1,n} = o_p(1) \). In addition, since \( M_2(\cdot, \cdot) \) is continuous and \( \zeta \) is càdlàg, \( A_{2,n} \rightarrow M_2(\zeta_{i\tau-}, \zeta_{j\tau-}) \). This together with (B.2) lead to

\[
\mathbf{H}^{(ij)}_{n,\tau-} \xrightarrow{p} M_2(\zeta_{i\tau-}, \zeta_{j\tau-}) - M_1(\zeta_{i\tau-}) M_1(\zeta_{j\tau-}) = v^{ij}_{\tau-}.
\]

Next we turn to the limiting behavior of \( Q^{(ij)}_{n,\tau-} \). We decompose

\[
\frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} V_i(l(\tau)-k) \Delta_n V_j(l(\tau)-k) = A_{3,n} + A_{4,n} + A_{5,n} + A_{6,n},
\]

\[
\Rightarrow
\]
where

\[ A_{3,n} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} (V_{i_1(\ell(\tau) - k) \Delta_n} - M_1(\zeta_{i_1(\ell(\tau) - k) \Delta_n})) r_{j_1(\tau) - k}^2, \]

\[ A_{4,n} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} (M_1(\zeta_{i_1(\ell(\tau) - k) \Delta_n}) - M_1(\zeta_{i_1(\ell(\tau) - (k-1) \Delta_n})) r_{j_1(\tau) - k}^2, \]

\[ A_{5,n} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} M_1(\zeta_{i_1(\ell(\tau) - (k-1) \Delta_n}) (r_{j_1(\tau) - k}^2 - \mathbb{E}[r_{j_1(\tau) - k}^2|F_{i_1(\tau) - (k-1)}]), \]

\[ A_{6,n} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} M_1(\zeta_{i_1(\ell(\tau) - (k-1) \Delta_n}) \mathbb{E}[r_{j_1(\tau) - k}^2|F_{i_1(\tau) - (k-1)}]. \]

Then by replacing all \( r_{i_1(\tau) - k}^2 \) with \( r_{j_1(\tau) - k}^2 \) and all the proofs can easily follow from the proof of limiting behavior of \( \Sigma_{n,\tau}^{(12)} \) in Bollerslev et al. (2017). Then we safely conclude that

\[ Q_{n,\tau}^{(ij)} \overset{p}{\longrightarrow} 0. \]

Finally there comes the limiting behavior of \( Z_{n,\tau}^{(ij)} \). For each \( \ell \in \{l(\tau) - k : 1 \leq k \leq k_n\} \), we can decompose \( r_{i_1,\tau} = r'_{i_1,\tau} + r''_{i_1,\tau} \) in restriction to \( \Omega_n \) where

\[
dP_{i_1,\tau} = b_{i_1,\tau} dt + \beta_{i_1,\tau} \sigma_{m_1,\tau} dW_{m_1,\tau} + \sigma_{i_1,\tau} dW_{i_1,\tau} + dJ_{i_1,\tau} \tag{B.8} \]

\[
r'_{i_1,\tau} = \beta_{i_1,\ell-1} \Delta_n \sigma_{m_1,\ell-1} \Delta_n (W_{m_1,\tau} - W_{m_1,\ell-1} \Delta_n) + \sigma_{i_1,\ell-1} \Delta_n (W_{i_1,\tau} - W_{i_1,\ell-1} \Delta_n) \]

\[
r''_{i_1,\tau} = \int_{(i-1) \Delta_n}^{\ell \Delta_n} b_s ds + \int_{(i-1) \Delta_n}^{\ell \Delta_n} (\beta_{i_1,\ell} \sigma_{m_1,\ell} - \beta_{i_1,\ell-1} \Delta_n \sigma_{m_1,\ell-1} \Delta_n) dW_{m_1,\tau}, \]

\[ + \int_{(i-1) \Delta_n}^{\ell \Delta_n} (\sigma_{i_1,\tau} - \sigma_{i_1,\ell-1} \Delta_n) dW_{i_1,\tau}. \]
We can then decompose

\[
\frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \Delta_n \left( \frac{r_{i,l}(\tau) - k}{\Delta_n} \right)^2 = A_{7,n} + A_{8,n}, \tag{B.9}
\]

where \( A_{7,n} \equiv \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \Delta_n \left( \frac{r_{i,l}(\tau) - k}{\Delta_n} \right)^2 \) and \( A_{8,n} \) is defined implicitly by (B.9). Here we assume \( i \neq j \) as the case \( i = j \) has already been proved in Bollerslev et al. (2017).

Note that

\[
\mathbb{E}[r_{i,i}^2 r_{j,j}^2 / \Delta_n^2 | \mathcal{F}(i-1) \Delta_n] = 3 \beta_i^2 (\delta(\mathcal{H} \Delta_n)) \left[ \beta_j^2 (\delta(\mathcal{H} \Delta_n)) \sigma_{m,\tau}^2 \Delta_n + \beta_j^2 (\delta(\mathcal{H} \Delta_n)) \sigma_{m,\tau}^2 \Delta_n + \sigma_i^2 (\delta(\mathcal{H} \Delta_n)) \sigma_{m,\tau}^2 \Delta_n + \sigma_i^2 (\delta(\mathcal{H} \Delta_n)) \sigma_{m,\tau}^2 \Delta_n \right].
\]

The variance of the martingale difference sequence \( r_{i,i}^2 r_{j,j}^2 / \Delta_n^2 - \mathbb{E}[r_{i,i}^2 r_{j,j}^2 / \Delta_n^2 | \mathcal{F}(i-1) \Delta_n] \) is bounded. \( \beta_{it} \) is bounded, so it is then easy to show that

\[
A_{7,n} \overset{p}{\rightarrow} 3 \beta_i^2 \beta_j^2 \sigma_{m,\tau}^2 + \beta_i^2 \beta_j^2 \sigma_{m,\tau}^2 + \beta_j^2 \beta_i^2 \sigma_{m,\tau}^2 + \sigma_i^2 \sigma_{m,\tau}^2.
\]

Similarly, we have the following decomposition:

\[
\frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \Delta_n \left( \frac{r_{i,l}(\tau) - k}{\Delta_n} \right)^2 = A_{9,n} + A_{10,n},
\]

where \( A_{9,n} = \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} \Delta_n \left( \frac{r_{i,l}(\tau) - k}{\Delta_n} \right)^2 \). Then we can similarly prove that

\[
A_{9,n} \overset{p}{\rightarrow} \beta_i^2 \sigma_{m,\tau}^2 + \sigma_i^2.
\]

Indeed, \( \mathbb{E}[r_{i,i}^2 / \Delta_n | \mathcal{F}(i-1) \Delta_n] = \beta_i^2 (\delta(\mathcal{H} \Delta_n)) \sigma_{m,\tau}^2 \Delta_n + \sigma_i^2 (\delta(\mathcal{H} \Delta_n)) \sigma_{m,\tau}^2 \Delta_n \).

\[
|A_{10,n}| \leq \frac{K}{k_n \Delta_n} \sum_{k=1}^{k_n} \left| \frac{r_{i,l}(\tau) - k}{\Delta_n} \right|^2 + \frac{K}{k_n \Delta_n} \sum_{k=1}^{k_n} \left| \frac{r_{i,l}(\tau) - k}{\Delta_n} \right|^2.
\]

176
By Hölder’s inequality,

$$
\mathbb{E} \left[ |r'_{i,l}(\tau) - k| \cdot |r''_{i,l}(\tau) - k| \right] \leq \sqrt{\mathbb{E} \left[ |r'_{i,l}(\tau) - k|^2 \right] \mathbb{E} \left[ |r''_{i,l}(\tau) - k|^2 \right]}.
$$

As \( \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} r''_{i,l(\tau) - k} = O_p(1) \), we only need to prove that \( \sum_{k=1}^{k_n} |r''_{i,l(\tau) - k}|^2 = o_p(1) \).

This can be concluded from \( \frac{1}{k_n \Delta_n} \sum_{k=1}^{k_n} r''_{i,l(\tau) - k} = O_p(1) \), as shown in the appendix A of Bollerslev et al. (2017).

With symmetry in \( i \) and \( j \),

$$
|A_{8,n}| \leq \frac{K}{k_n \Delta_n^2} \sum_{k=1}^{k_n} \left| r'_{i,l(\tau) - k} \right|^2 \left| r'_{j,l(\tau) - k} \right| \left| r''_{j,l(\tau) - k} \right| + \frac{K}{k_n \Delta_n^2} \sum_{k=1}^{k_n} \left| r'_{i,l(\tau) - k} \right| \left| r''_{i,l(\tau) - k} \right| \left| r''_{j,l(\tau) - k} \right|^2 + \frac{K}{k_n \Delta_n^2} \sum_{k=1}^{k_n} \left| r''_{i,l(\tau) - k} \right| \left| r''_{j,l(\tau) - k} \right|^2.
$$

Apply Hölder’s inequality to each of the three components we get

$$
\mathbb{E} \left[ |r'_{i,l(\tau) - k}| \cdot |r'_{j,l(\tau) - k}| \cdot |r''_{j,l(\tau) - k}| \right] \leq \mathbb{E} \left[ |r'_{i,l(\tau) - k}|^4 \right]^{1/4} \mathbb{E} \left[ |r'_{j,l(\tau) - k}|^4 \right]^{1/4} \mathbb{E} \left[ |r''_{j,l(\tau) - k}|^4 \right]^{1/4},
$$

$$
\mathbb{E} \left[ |r'_{i,l(\tau) - k}| \cdot |r''_{i,l(\tau) - k}| \cdot |r''_{j,l(\tau) - k}|^2 \right] \leq \mathbb{E} \left[ |r'_{i,l(\tau) - k}|^4 \right]^{1/4} \mathbb{E} \left[ |r''_{i,l(\tau) - k}|^4 \right] \mathbb{E} \left[ |r''_{j,l(\tau) - k}|^4 \right]^{1/4},
$$

$$
\mathbb{E} \left[ |r''_{i,l(\tau) - k}| \cdot |r''_{j,l(\tau) - k}| \cdot |r''_{j,l(\tau) - k}|^2 \right] \leq \sqrt{\mathbb{E} \left[ |r''_{i,l(\tau) - k}|^4 \right] \mathbb{E} \left[ |r''_{j,l(\tau) - k}|^4 \right]},
$$

177
Therefore, we only need to prove that $E \left[ \frac{1}{k_n \Delta_n^2} \sum_{j=1}^{k_n} \left| \eta_{i(l(\tau)-j)} \right|^4 \right] \to 0$. By the Burkholder–Davis–Gundy inequality and Hölder’s inequality,

$$E \left[ \left| \eta_{i(l(\tau)-j)} \right|^4 \right] \leq K \Delta_n^4$$

$$+ \sup_t |\beta_t|^4 K \Delta_n \mathbb{E} \left[ \int_{(l(\tau)-k-1) \Delta_n}^{(l(\tau)-k) \Delta_n} \left( \sigma_{m,s} - \sigma_{m,((l(\tau)-k-1) \Delta_n)} \right)^4 ds \right]$$

$$+ K \Delta_n \mathbb{E} \left[ \int_{(l(\tau)-k-1) \Delta_n}^{(l(\tau)-k) \Delta_n} \left( \sigma_{i,s} - \sigma_{i,((l(\tau)-k-1) \Delta_n)} \right)^4 ds \right],$$

$$E \left[ \frac{1}{k_n \Delta_n^2} \sum_{j=1}^{k_n} \left| \eta_{i(l(\tau)-j)} \right|^4 \right] \leq K \Delta_n^2 + K E \left[ \sup_{t \in [\tau-2k_n \Delta_n, \tau]} |\sigma_t - \sigma_\tau|^4 \right] \to 0.$$

Therefore, the third term of (B.10) is $o_p(1)$. As $\frac{1}{k_n \Delta_n^2} \sum_{k=1}^{k_n} \eta_{i,l(\tau)-k} = O_p(1)$, we can conclude that the first two terms of (B.10) are also $o_p(1)$. Therefore, $A_{8,n} = o_p(1)$. Hence,

$$Z_{n,\tau}^{(ij)} \xrightarrow{p} 3\beta_{i,\tau}^2 - \beta_{j,\tau}^2 - \sigma_{m,\tau}^4 \sigma_{i,\tau} - \beta_{i,\tau}^2 - \sigma_{i,\tau}^2 - \sigma_{m,\tau}^2 \sigma_{j,\tau} - \sigma_{i,\tau}^2 \sigma_{j,\tau} + \sigma_{i,\tau}^2 \sigma_{j,\tau} -$$

$$- \left( \beta_{i,\tau}^2 - \sigma_{m,\tau}^2 \right) \left( \beta_{j,\tau}^2 - \sigma_{m,\tau}^2 \right),$$

$$Z_{n,\tau}^{(ij)} \xrightarrow{p} 2\beta_{i,\tau}^2 - \beta_{j,\tau}^2 - \sigma_{m,\tau}^4.$$

This marks the end of proof for (B.7).

Step 3. For this step, we need to prove the multi-dimensional version of Equation (A.17) in Bollerslev et al. (2017). Actually, by essentially the same argument for
(B.5), we can show that

\[
\sqrt{k_n} \begin{pmatrix}
\hat{m}_{n,1\tau} - \hat{m}_{1\tau} \\
\vdots \\
\hat{m}_{n,I\tau} - \hat{m}_{I\tau} \\
\hat{\sigma}_{n,1\tau} - \hat{\sigma}_{1\tau} \\
\vdots \\
\hat{\sigma}_{n,I\tau} - \hat{\sigma}_{I\tau}
\end{pmatrix} \overset{\mathcal{L}}{\rightarrow} \begin{pmatrix}
\eta_{1\tau} \\
\vdots \\
\eta_{I\tau} \\
\eta_{1\tau}^I \\
\vdots \\
\eta_{I\tau}^I
\end{pmatrix}
\]  

(B.11)

The proof is finished. \( Q.E.D. \)

B.2 Derivation of the Model

B.2.1 The general case when there are \( n \) assets

We first derive the relationship between trading intensity and return volatility when there are \( n \) assets with a general variance distribution.

Prior to the release of the signal

Suppose that prior to the signal is released, investor \( j \) believes that the pricing of all stocks has the prior distribution as below:

\[
V_j \sim N(X_j, \Sigma_j),
\]

where \( V_j = [V_{j1} \ V_{j2} \ \cdots \ V_{jn}] \) is the vector of fundamental values. \( X_j = [X_{j1} \ X_{j2} \ \cdots \ X_{jn}] \) is a vector of means. \( \Sigma_j \) is the \( n \times n \) covariance matrix.

Each investor maximizes his expected utility

\[
\max_{m_i(1)} \mathbb{E}_i^{(1)} \left[ -\exp \left\{ -\lambda m_i^{(1)\gamma} (X_i - P^{(1)}) \right\} \right].
\]

The superscript (1) means the value corresponds to stage 1. \( P^{(1)} \) is the vector of prices before the investors see the signal. Then the optimal portfolio for investor \( j \)
can be solved:

\[ m_j^{(1)} = \lambda^{-1} \Sigma_j^{-1} (X_j - P^{(1)}) \]

When market clears,

\[ \alpha m_1^{(1)} + (1 - \alpha) m_2^{(1)} = 0 \]

Then the equilibrium price \( P^{(1)*} \) and equilibrium positions \( m_j^{(1)*} \) can be solved.

The signal is released

Upon receiving the signal, for investor \( j \), the distribution of \( V \) interpreted from the signal \( L \) is \( N(\mu_j(L), \Omega_j) \). Then under a Bayesian framework, the posterior distribution is

\[ V_j | L \sim N \left(Y_j, (\Sigma_j^{-1} + \Omega_j^{-1})^{-1}\right) \]

\( Y_j = \rho_j X_j + (1 - \rho_j) \mu_j(L) \) where \( \rho_j = (\Sigma_j^{-1} + \Omega_j^{-1})^{-1} \Sigma_j^{-1} \). We can similarly solve for the equilibrium prices and positions. If we use the following notations (\( L \) is suppressed for simplicity):

\[
\hat{\Sigma}^{-1} \equiv \alpha \Sigma_1^{-1} + (1 - \alpha) \Sigma_2^{-1} \\
\hat{\Omega}^{-1} \equiv \alpha \Omega_1^{-1} + (1 - \alpha) \Omega_2^{-1} \\
\hat{X} \equiv \alpha \Sigma_1^{-1} X_1 + (1 - \alpha) \Sigma_2^{-1} X_2 \\
\hat{\mu} \equiv \alpha \Omega_1^{-1} \mu_1 + (1 - \alpha) \Omega_2^{-1} \mu_2
\]

The equilibrium prices and positions are

\[ P^{(1)*} = \hat{\Sigma} \hat{X}, \]

\[ P^{(2)*} = \left(\hat{\Sigma}^{-1} + \hat{\Omega}^{-1}\right)^{-1} \hat{X} + \hat{\mu}. \]
The equilibrium holdings are
\[ m_2^{(1)*} = \Sigma_1^{-1} (X_1 - P^{(1)*}) , \]
\[ m_1^{(2)*} = \lambda^{-1} (\Sigma_1^{-1} + \Omega_1^{-1}) (Y_1 - P^{(2)*}) . \]

The price change is
\[ \Delta P^* = P^{(2)*} - P^{(1)*} \]
\[ = \left( \hat{\Sigma}^{-1} + \hat{\Omega}^{-1} \right)^{-1} \left( \hat{\mu} - \hat{\Omega}^{-1} P^{(1)*} \right). \]  

(B.12)

Volume is the change in position:
\[ \text{Vol}^* = \lambda^{-1} \alpha \left[ m_1^{(2)*} - m_1^{(1)*} \right] \]
\[ = \lambda^{-1} \alpha (1 - \alpha) \left| \Omega_1^{-1} \hat{\Omega} \Omega_2^{-1} (\mu_1 - \mu_2) + (\Omega_2 \Sigma_2^{-1} - \Omega_1 \Sigma_1^{-1}) \Delta P^* \right| . \]

(B.13)

(B.14)

Equation (B.14) is obtained by plunging \( P^{(1)*} \) in as a function of \( \Delta P^* \) from equation (B.12). Therefore, the equilibrium volume and price change relationship is as follows:
\[ \text{Vol}^* = \lambda^{-1} \alpha (1 - \alpha) \left| \Omega_1^{-1} \hat{\Omega} \Omega_2^{-1} (\mu_1 - \mu_2) + (\Omega_2 \Sigma_2^{-1} - \Omega_1 \Sigma_1^{-1}) \Delta P^* \right| . \]

B.2.2 Implement the one factor model with heterogeneous beliefs

Suppose the pricing of all stocks for investor \( j \) follows the distribution below:
\[ V \sim N (X_j, \Sigma_j) . \]
Hence after we incorporate the one factor structure with heterogeneous beliefs, the expressions of the parameters are:

\[
X_j = \beta x_{mj} + \gamma \eta_j,
\]

\[
\Sigma_j = (\beta \beta' + S) z_{mj}^2,
\]

where \( \beta = [\beta_1 \beta_2 \cdots \beta_n] \), \( \gamma = [\gamma_1 \gamma_2 \cdots \gamma_n] \) and \( S = \text{diag}(\gamma_1^2 \gamma_2^2 \cdots \gamma_n^2) \). As the structure of the covariance matrix persists, we can simplify the proof if we introduce the following matrix:

\[
M = \begin{bmatrix}
\beta_1^2 + \gamma_1^2 & \beta_1 \beta_2 & \cdots & \beta_1 \beta_n \\
\beta_2 \beta_1 & \beta_2^2 + \gamma_2^2 & \cdots & \beta_2 \beta_n \\
\vdots & \vdots & \ddots & \vdots \\
\beta_n \beta_1 & \beta_n \beta_2 & \cdots & \beta_n^2 + \gamma_n^2
\end{bmatrix}.
\]

It can be regarded as the covariance coefficients. The joint distribution of the stock prices can be written as functions of \( \beta, \gamma, M \) as well as the parameters of the distribution of the market payoff.

\[
X_j = \beta x_{Mj} + \gamma \eta_j,
\]

\[
\Sigma_j = z_{mj}^2 M,
\]

\[
\mu_j (L) = \beta (L - \mu_{mj}) + \gamma \eta_j,
\]

\[
\Omega_j = \sigma_{mj}^2 M.
\]

Plug them in equation (B.14) we get

\[
\text{Vol}^* = c \left| M^{-1} (\mu_1 - \mu_2) + \left( \frac{\sigma_{m1}^2}{\sigma_{m1}^2} - \frac{\sigma_{m2}^2}{\sigma_{m2}^2} \right) M^{-1} \Delta P^* \right|, \quad \text{(B.15)}
\]

182
where $c = \lambda^{-1} \alpha (1 - \alpha) \left( \alpha \sigma_{m1}^{-2} + (1 - \alpha) \sigma_{m2}^{-2} \right)^{-1} \sigma_{m1}^{2} \sigma_{m2}^{2}$. With the notation $\Delta \mu_m = \mu_{m2} - \mu_{m1}$ and $\Delta \eta = \eta_1 - \eta_2$,

$$\mu_1 (L) - \mu_2 (L) = \beta \Delta \mu_m + \gamma \Delta \eta.$$  

For any variable $x$, we define

$$\tilde{x} = \frac{\alpha \sigma_{m1}^{-2} x_1 + (1 - \alpha) \sigma_{m2}^{-2} x_2}{\alpha \sigma_{m1}^{-2} + (1 - \alpha) \sigma_{m2}^{-2}},$$

$$\tilde{x} = \frac{\alpha z_{m1}^{-2} x_1 + (1 - \alpha) z_{m2}^{-2} x_2}{\alpha z_{m1}^{-2} + (1 - \alpha) z_{m2}^{-2}},$$

$$\kappa_m = \frac{\alpha \sigma_{m1}^{-2} + (1 - \alpha) \sigma_{m2}^{-2}}{\alpha \sigma_{m1}^{-2} + (1 - \alpha) \sigma_{m2}^{-2} + \alpha z_{m1}^{-2} + (1 - \alpha) z_{m2}^{-2}},$$

then the price changes of the stocks are

$$\Delta P^* = P^{(2)*} - P^{(1)*}$$

$$= \left( \tilde{\Sigma}^{-1} + \tilde{\Omega}^{-1} \right)^{-1} \left( \tilde{\mu} (L) - \tilde{\Omega}^{-1} \tilde{\Sigma} \tilde{x} \right)$$

$$= \beta F_1 + \gamma F_2,$$

where

$$F_1 = \kappa_m (L - \tilde{\mu} - \tilde{x}),$$

$$F_2 = \kappa_m (\tilde{\eta} - \tilde{\eta}).$$

We notice that $F_1 = \Delta P^*_m$ has exact the same expression as in Kandel and Pearson (1995), so it can be treated as the equilibrium price change for the market. $F_2$ is a constant, the value of which is

$$\mu_\eta = \alpha (1 - \alpha) \left( \alpha + (1 - \alpha) \frac{\sigma_{m1}^{2}}{\sigma_{m2}^{2}} \right)^{-1} \left( \frac{z_{m2}^{2}}{z_{m1}^{2}} + (1 - \alpha) \right)^{-1} \left( 1 - \frac{\sigma_{m1}^{2} z_{m2}^{2}}{\sigma_{m2}^{2} z_{m1}^{2}} \right) \Delta \eta.$$
When the prices of stocks have such a factor structure, the distribution of their equilibrium prices is

\[ \Delta P^* \sim N(\gamma \mu, \Sigma_p), \]

where

\[ \Sigma_p = \begin{bmatrix} \beta_1^2 & \beta_1 \beta_2 & \cdots & \beta_1 \beta_n \\ \beta_2 \beta_1 & \beta_2^2 & \cdots & \beta_2 \beta_n \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n \beta_1 & \beta_n \beta_2 & \cdots & \beta_n^2 \end{bmatrix}. \]

**B.2.3 Volume-volatility elasticity**

We aim to find the individual stock’s volume-volatility elasticity from equation (B.15):

\[ \text{Vol}^* = c \left| M^{-1} (\mu_1 - \mu_2) + \left( \frac{\sigma_{m1}^2}{\sigma_{m1}^2 - \sigma_{m2}^2} \right) M^{-1} \Delta P^* \right|. \]

According to Kan and Robotti (2017), if \( X \sim N(\mu, \Sigma) \) and \( Y = c|X| \),

\[ E(Vol_i^*) = \mu_i \text{erf} \left( \frac{\mu_i}{\sqrt{2} \sigma_i} \right) + 2 \sigma_i \phi \left( \frac{\mu_i}{\sigma_i} \right) + \mu_i \left[ 2 \Phi \left( \frac{\mu_i}{\sigma_i} \right) - 1 \right] + \sqrt{\frac{2}{\pi}} \sigma_i \exp \left\{ -\frac{\mu_i^2}{2 \sigma_i^2} \right\}. \]

where \( \mu_i \) is the \( i \)th element of \( \mu \) and \( \sigma_i^2 \) is the \( i \)th element on the diagonal of \( \Sigma \), so the volume-volatility elasticity is exactly the same as the univariate case in Bollerslev et al. (2017). Since we are not interested in a particular equilibrium volume and price change, but instead in the “average” level, we continue to find out the relationship between expected volume and price change volatility. Denote \( v_i = E[Vol_i^*] \), and

184
\( \sigma_{p_i} \) be the volatility of \( p_i \), the price change of stock \( i \), then the relationship between \( v_i \) and \( \sigma_{p_i} \) is
\[
\mathcal{E}_i \equiv \frac{\partial v_i(\sigma_{p_i})}{\partial \sigma_{p_i}} \frac{v_i(\sigma_{p_i})}{\sigma_{p_i}} = \frac{1}{1 + \psi(\theta_i)},
\]
where \( \theta_i \) is the ratio of the mean to the volatility of the \( i \)'th stock. Therefore, we need to find the distribution of \( \text{Vol}^* \) and \( \Delta \text{P}^* \). Let
\[
A \equiv \begin{pmatrix} \frac{\sigma_{m_1}^2}{z_{m_1}^2} & -\frac{\sigma_{m_2}^2}{z_{m_2}^2} \end{pmatrix} M^{-1}, \quad B \equiv M^{-1} (\mu_1 - \mu_2) .
\]
then \( \text{Vol}^* = c|B + A\Delta \text{P}^*| \), and the distribution of \( B + A\Delta \text{P}^* \) is \( N(A\gamma\mu_\eta + B, AS\Sigma_p A^T) \). Without loss of generality, we look at the first stock. For the first stock, \( \theta_1 \) is the ratio of the first element of \( A\gamma\mu_\eta + B \) to the square root of the \((1, 1)\) element of \( A\Sigma_p A^T \). For at \( A\Sigma_p A^T \), it is easy to prove that
\[
M^{-1} = \left( \frac{\sum_{i=1}^n \beta_i \prod_{j \neq i} \gamma_j^2 \gamma_i^2}{\prod_{i=1}^n \gamma_i^2} \right)^{-1} H,
\]
where the \((k, k)\) element of \( H \) is \( \sum_{i \neq k} \prod_{j \neq k} \beta_j \gamma_j^2 + \prod_{i \neq k} \gamma_i^2 \) and the \((k, l) \quad (k \neq j)\) element of \( H \) is \( -\beta_1 \beta_2 \prod_{j \neq k, l} \gamma_j^2 \). Hence the \((1, 1)\) element of \( A\Sigma_p A^T \) is
\[
\sqrt{f_{11}(\Sigma_p)} = \frac{\left| \frac{\sigma_{q_1}^2}{z_{q_1}^2} - \frac{\sigma_{q_2}^2}{z_{q_2}^2} \right| \beta_1 \prod_{i \neq 1} \gamma_i^2 \sigma_{p_1}}{\sum_{i=1}^n \beta_i \prod_{j \neq i} \gamma_j^2 + \prod_{i=1}^n \gamma_i^2}.
\]
Next we turn to the mean vector
\[
B + A\gamma\mu_\eta = M^{-1} (\beta \Delta \mu + \gamma \Delta \eta) + \frac{\sum_{i=1}^n \beta_i \prod_{j \neq i} \gamma_j^2 + \prod_{i=1}^n \gamma_i^2}{\beta_1 \prod_{i \neq 1} \gamma_i^2} M^{-1} \gamma \mu_\eta,
\]
185
the first component of which is

\[
\frac{\prod_{i \neq 1} \gamma_{i}^{2} \beta_{1}}{\sum_{i=1}^{n} \beta_{i} \prod_{j \neq i} \gamma_{j}^{2} + \prod_{i=1}^{n} \gamma_{i}^{2}} \Delta \mu + \frac{\left(\sum_{i \neq 1, j \neq i, 1} \beta_{i}^{2} \gamma_{j}^{2} + \prod_{i \neq 1} \gamma_{i}^{2}\right) \gamma_{1} - \beta_{1} \sum_{i \neq 1, j \neq i, 1} \beta_{i} \gamma_{i} \gamma_{j}^{2}}{\sum_{i=1}^{n} \beta_{i} \prod_{j \neq i} \gamma_{j}^{2} + \prod_{i=1}^{n} \gamma_{i}^{2}} \Delta \eta
\]

\[
\left(\frac{\sigma_{m_{1}}^{2}}{\sigma_{m_{1}}^{2}} - \frac{\sigma_{m_{2}}^{2}}{\sigma_{m_{2}}^{2}}\right) \left(\left(\sum_{i \neq 1, j \neq i, 1} \beta_{i}^{2} \gamma_{j}^{2} + \prod_{i \neq 1} \gamma_{i}^{2}\right) \gamma_{1} - \beta_{1} \sum_{i \neq 1, j \neq i, 1} \beta_{i} \gamma_{i} \gamma_{j}^{2}\right) \frac{\Delta \eta}{\sum_{i=1}^{n} \beta_{i} \prod_{j \neq i} \gamma_{j}^{2} + \prod_{i=1}^{n} \gamma_{i}^{2}} \mu_{\eta}.
\]

Divide it by \( \sqrt{f_{11}(\Sigma_{p})} \) we get

\[
\theta_{1} = \frac{1}{\sigma_{pm}} \left| \frac{\sigma_{m_{1}}^{2}}{z_{m_{1}}^{2}} - \frac{\sigma_{m_{2}}^{2}}{z_{m_{2}}^{2}} \right|^{-1} \frac{\Delta \mu}{\sigma_{pm}} + \frac{1}{\sigma_{pm}} \left| \frac{\sigma_{m_{1}}^{2}}{z_{m_{1}}^{2}} - \frac{\sigma_{m_{2}}^{2}}{z_{m_{2}}^{2}} \right|^{-1} \left( \frac{\gamma_{1}}{\beta_{1}} \left( \sum_{i=2}^{n} \beta_{i}^{2} \gamma_{i}^{2} + 1 \right) - \sum_{i=2}^{n} \frac{\beta_{i}}{\gamma_{i}} \right) \frac{\Delta \eta}{\sigma_{pm}}
\]

\[
+ \text{sgn} \left( \frac{\sigma_{m_{1}}^{2}}{z_{m_{1}}^{2}} - \frac{\sigma_{m_{2}}^{2}}{z_{m_{2}}^{2}} \right) \left( \frac{\gamma_{1}}{\beta_{1}} \left( \sum_{i=2}^{n} \beta_{i}^{2} \gamma_{i}^{2} + 1 \right) - \sum_{i=2}^{n} \frac{\beta_{i}}{\gamma_{i}} \right) \frac{\mu_{\eta}}{\sigma_{pm}}.
\]

Recall that

\[
\mu_{\eta} = \alpha (1 - \alpha) \left( \alpha \sigma_{m_{2}}^{2} + (1 - \alpha) \sigma_{m_{1}}^{2} \right)^{-1} \left( \alpha z_{m_{1}}^{2} + (1 - \alpha) z_{m_{2}}^{2} \right)^{-1} \left( \frac{\sigma_{m_{2}}^{2}}{z_{m_{2}}^{2}} - \frac{\sigma_{m_{1}}^{2}}{z_{m_{1}}^{2}} \right) \Delta \eta.
\]

If we denote

\[
\Delta w = \left| \frac{\sigma_{m_{1}}^{2}}{z_{m_{1}}^{2}} - \frac{\sigma_{m_{2}}^{2}}{z_{m_{2}}^{2}} \right|,
\]

\[
g \left( \frac{\gamma_{i}}{\beta_{i}} \right) = \frac{\gamma_{i}}{\beta_{i}} \left( \sum_{k=1}^{n} \frac{\beta_{i}^{2}}{\gamma_{k}^{2}} + 1 \right) - \sum_{j=1}^{n} \frac{\beta_{k}}{\gamma_{k}},
\]

\[
c_{m} = \alpha (1 - \alpha) \left( \alpha \sigma_{m_{2}}^{2} + (1 - \alpha) \sigma_{m_{1}}^{2} \right)^{-1} \left( \alpha z_{m_{1}}^{2} + (1 - \alpha) z_{m_{2}}^{2} \right)^{-1},
\]

186
then the elasticity of stock \( i \) can be written as

\[
\mathcal{E}_i = \frac{1}{1 + \psi(\theta_i)},
\]

\[
\theta_i = \frac{\Delta w^{-1} \Delta \mu}{\sigma_{pm}} + \left(\Delta w^{-1} + c_m \Delta w\right) g \left(\frac{\gamma_i}{\beta_i}\right) \frac{\Delta \eta}{\sigma_{pm}}.
\]

To further simplify the analysis, consider a special case, when \( \sigma_{m1}^2 = \sigma_{m2}^2 = h^{-1} \), and \( s_j = z_{mj}^{-2} \) denote the precisions,

\[
\theta_i = \frac{h}{|s_1 - s_2|} \frac{\Delta \mu_m}{\sigma_{pm}} + \left(\frac{h}{|s_1 - s_2|} + \alpha (1 - \alpha) \frac{|s_1 - s_2|}{\alpha s_1 + (1 - \alpha) s_2}\right) g \left(\frac{\gamma_i}{\beta_i}\right) \frac{\Delta \eta}{\sigma_{pm}}.
\]

Further, let

\[
\xi(h, s_1, s_2) \equiv \frac{h}{|s_1 - s_2|} + \alpha (1 - \alpha) \frac{|s_1 - s_2|}{\alpha s_1 + (1 - \alpha) s_2},
\]

We have completed the model derivation.
Appendix C

Appendix for Chapter 4

C.1 Proof of Theorem 4.4.1

Jacod and Protter (2011) have argued that the consistency of irregular estimator hold in their Chapter 3, for the unnormalized version of the estimator. The consistency of the normalized ones can be proved following Theorem 9.4.1. Using their words, the only challenge is to do the approximation for \( c_t \) itself, since \( g \) is continuous and thus \( g(\hat{c}_t^n) \) approximates \( g(c_t) \) immediately. Besides, \( g(\hat{c}_t^n) \) has the same good performance as a direct approximation of \( g(c_t) \). Hence, we start with a local approximation of \( c_t \).

Define

\[
\hat{c}_t^n(k_n)^j = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \frac{\Delta_{i+m}^n X_j \Delta_{i+m}^n X_t}{\Delta_{i+m}^n},
\]

where \( \Delta_{n,i} = T(n, i) - T(n, i - 1) \) and \( t \in [T(n, i - 1), T(n)] \).
Then
\[ \hat{c}_i^n(k_n, t-)^{jl} = \frac{1}{k_n} \sum_{m=-k_n}^{-1} \frac{\Delta_{i+m} X^j \Delta_{i+m} X^l}{\Delta_{i+m}^n}, \]
which uses \(k_n\) windows prior to time \(t\).

\[ \hat{c}_i^n(k_n, t)^{jl} = \frac{1}{k_n} \sum_{m=1}^{k_n+1} \frac{\Delta_{i+m} X^j \Delta_{i+m} X^l}{\Delta_{i+m}^n}, \]
which uses \(k_n\) windows after time \(t\). Assume that \(X\) is one dimensional and \(T\) is bounded, which is not restrictive in our setting. Introduce the following processes:

\[ Y^n_t = \sigma_{T_n}(W_t - W_{T_n})1_{\{T_n \leq t\}}, \quad Y'_n = \int_{T_n \wedge t}^t (\sigma_s - \sigma_{T_n})dW_s. \]

Simplify the notation such that \(T_n = T(n, N_n(T) + 1)\) and \(i_n = N_n + 2\), then again \(T \in [T_{n-1}, T_n)\).

\[ \hat{c}_i^n(k_n, Y^n) = c_{T_n} \hat{c}_n \quad \text{and} \quad \hat{c}_n = \frac{1}{k_n} \sum_{i=0}^{k_n-1} \frac{(\Delta_{i+n+i} W)^2}{\Delta_{i+n+i}^n}, \tag{C.1} \]

where \(T_n\) is the earliest observed time after time \(T\) and \(i_n\) is the immediately next observation after \(T_n\). Similar as in Jacod and Protter (2011), since \(T_n\) is a stopping time, the variables \(\Delta_{i+n+i}^n W\) for \(i \geq 0\) are i.i.d. with the law \(N(0, \sqrt{\Delta_n})\). It is straightforward that \(\hat{c}_n \overset{p}{\rightarrow} 1\) as it has mean 1 and variance going to 0. They arrive at the conclusion \(\hat{c}_{i,n}^n(k_n, Y^n) - c_{T_n} \overset{p}{\rightarrow} 0\) if \(\Delta_{i+n+i}^n W \sim N(0, \sqrt{\Delta_{i+n+i}^n})\) is strongly predictable w.t. \(\mathcal{F}_T\) then it is straightforward that \(\hat{c}_n \overset{p}{\rightarrow} 1\) as it has mean 1 and variance going to 0. Indeed,
This leads to the conclusion that

\[ \tilde{c}_{i,n}^n(k_n, Y^n) - c_{T_n} \xrightarrow{p} 0. \]

It should hold when the series \( T_n \) is strongly predictive. If this holds, the variables \( \Delta_{n,i} W \) for \( i \geq 0 \) are conditionally independent with the law \( N(0, \sqrt{\Delta_{n,i}}) \). We know that \( \tilde{c}_n \xrightarrow{p} 1 \) as it has mean 1 and variance going to 0. Next, we prove

\[ \tilde{c}_n(k_n, v_n, X) \xrightarrow{p} \tilde{c}_n(k_n, Y^n), \]

where

\[ \tilde{c}_n(k_n, v_n)^{jl} = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \frac{\Delta_{n,i+m} X_j \Delta_{n,i+m} X^l}{\Delta_{n,i+m}} 1(\|\Delta_{n,i+m} X\| \leq v_{n,i+m}) \quad \text{and} \quad v_{n,i} = \alpha \Delta_{n,i}. \]

Define \( x = \Delta_{n,i} Y^n / \sqrt{\Delta_{n,i} \Delta_{n,i} + 1}, y = \Delta_{n,i} X'' / \sqrt{\Delta_{n,i} \Delta_{n,i} + 1}, z = \Delta_{n,i} Y'' / \sqrt{\Delta_{n,i} \Delta_{n,i} + 1} \) and

\[ w = \Delta_{n,i} B'' / \sqrt{\Delta_{n,i} \Delta_{n,i} + 1}. \]

So \( x + y + z + w = \Delta_{n,i} X / \sqrt{\Delta_{n,i}} \) and \( v = v_{n,i} + i / \sqrt{\Delta_{n,i} \Delta_{n,i} + 1} \).
\[ \alpha^{1/2} \Delta_n^{\infty-1/2} \] then the conclusion will be easily reached using step 3 on page 258 of Jacod and Protter (2011). The only additional assumption needed is \( \max |\Delta(n, i)| = O(\Delta_n) \), which is a conclusion of the assumption on the process \( T \). subsectionFrom Local Approximation to Global Approximation The theorem is an analogue of Theorem 9.4.1 in Jacod and Protter (2011). Recall that

\[
\hat{c}_i^n(k_n, t)^{jl} = \frac{1}{k_n} \sum_{m=1}^{k_n} \frac{\Delta_n^m X^j \Delta_n^m X^l}{\Delta_n^{i+m}}.
\]

which uses \( k_n \) windows after time \( t \).

\[
\hat{c}_i^n(k_n, v_n)^{jl} = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \frac{\Delta_n^m X^j \Delta_n^m X^l}{\Delta_n^{i+m}} \mathbf{1}_{\{\|\Delta_n^{i+m} X\| \leq v_n,i+m\}}.
\]

where \( v_{n,i} = \alpha \Delta_n^{\infty} \) First prove the results when \( g \) is bounded. For any \( t > 0 \) we have \( \hat{c}_i^n(k_n, t) = \hat{c}_i^n(k_n) \) when \( T(n, i - 1) \leq t < T(n, i) \), hence

\[
C^n(g, k_n, X)_t = \Delta_n 1 g(\hat{c}_i^n(k_n)) + \int_0^{T(n,N_n(t)-k_n)} g(\hat{c}_i^n(k_n, s)) ds. \tag{C.2}
\]

If we let

\[
E\left[ \left| (C^n(g, k_n, X)_t - \int_0^t g(c_s) ds) \right| \right] \leq K |T(n, N_n(t)) - T(n, N_n(t) - k_n)| + \int_0^{T(n,N_n(t)-k_n)} a_n(s) ds,
\]

where \( a_n(s) = E\left[ |g(\hat{c}_i^n(k_n, s) - g(c_s)| \right] \), then \( \hat{c}_i^n(k_n, T) \xrightarrow{p} c_T \) implies that \( a_n(s) \to 0 \) for each \( s \) and stays bounded uniformly in \( (n, s) \) because \( g \) is bounded. Hence 1 follows from the dominated convergence theorem, and 2 is obtained in the same way.
The twist from Chapter 9.4 is that it remains to prove that
\[
\lim_{m \to \infty} \lim_{n \to \infty} \sup \mathbb{E} \left( \frac{1}{[t/\Delta_n]} \sum_{i=1}^{[t/\Delta_n]} \Delta_{n,i} \left\| \hat{c}_i^n (k_n, v_n) \right\|^p \mathbb{1}\{\|\hat{c}_i^n (k_n, v_n)\| > m\} \right) = 0. \tag{C.3}
\]

We start with (C.3). Similar we have
\[
\mathbb{E} \left( \|\Delta_n^p X\|^q \right) \leq K_q \left( \Delta_n^{q/2} + \kappa \Delta_n^{(q/2) \wedge 1} \right).
\]
Assume \(\sup \Delta_{n,i} = O_p (\Delta_n)\), then with some algebra we conclude
\[
\mathbb{E} \left( \|\hat{c}_i^n (k_n)\|^q \right) \leq K_q \left( 1 + \kappa \Delta_n^{0 \wedge 1 - q} \right).
\]
This leads to the conclusion that
\[
\left\| \hat{c}_i^n (k_n, v_n) \right\|^p \mathbb{1}\{\|\hat{c}_i^n (k_n, v_n)\| > m\} \leq \frac{K}{mp^q} + \sup \Delta_{n,i+j}^{w} \phi_{n,i+j}.
\]
where \(\phi_{n,i+j} \to 0\). We have finished the proof.

C.2 Proof of Theorem 4.4.2

The logic of this proof is to decompose \(V'(g)^n_t\) as defined in Equation 4.10 and prove that one of the component generates the central limit theorem while the others converges to zero.

C.2.1 Decompose into components

First we decompose \(\frac{1}{\sqrt{\Delta_n}} \left( V'(g)^n_t - V(g)_t \right)\). Use the notation from last section,

\[
\hat{c}_i^n (k_n, v_n) = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \Delta_{n,i+m} \Delta_{n,i+m} X' \Delta_{n,i+m} X' \Delta_{n,i+m} \mathbb{1}\{\|\Delta_{n,i+m} X' \| \leq u_{n,i+m}\}.
\]
where \( u_{n,i+m} = O_p(\Delta_n i^{1/2-\omega}) \) and \( c_i^n = c_T(n,i-1) \). From the last section we know

\[
V(g)_t^n := \sum_{i=1}^{[\gamma/\Delta_n] - k_n + 1} \Delta_{n,i} g(\tilde{c}_i^n) \overset{u.c.p.}{\longrightarrow} V(g)_t := \int_0^t g(c_s) \, ds.
\]

Define

\[
V'(g)_t^n = \sum_{i=1}^{N_n(t)-k_n + 1} \Delta_{n,i} \left( g \left( \tilde{c}_i^n \right) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \partial^2_{jk,l,m} g(\tilde{c}_i^n) \times \left( \tilde{c}_i^{n,jl} \tilde{c}_i^{n,km} + \tilde{c}_i^{n,jm} \tilde{c}_i^{n,kl} \right) \right),
\]

then \( \frac{1}{\sqrt{k_n}} (V'(g)_t^n - V(g)_t^n) = \sum_{j=1}^5 V^{(n,j)}_t \), as soon as \( N_n(t) > k_n \).

\[
V^{(1)}_t^n = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{N_n(t)-k_n + 1} \Delta_{n,i}
\]

\[
\times \left( \left( g \left( \tilde{c}_i^n \right) - g \left( \tilde{c}_i^{n,j} \right) \right) \right)
\]

\[
- \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \left( \partial^2_{jk,l,m} g(\tilde{c}_i^n) \times \left( \tilde{c}_i^{n,jl} \tilde{c}_i^{n,km} + \tilde{c}_i^{n,jm} \tilde{c}_i^{n,kl} \right) \right)
\]

\[
V^{(2)}_t^n = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{N_n(t)-k_n + 1} \int_{T(n,i-1)}^T g(c_i^n - g(c_s)) \, ds - \frac{1}{\sqrt{k_n}} \int_{T(n,N_n(t)-k_n+1)}^T g(c_s) \, ds
\]

\[
V^{(3)}_t^n = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{N_n(t)-k_n + 1} \Delta_{n,i} \sum_{l,m=1}^d \partial_{lm} g(\tilde{c}_i^n) \frac{1}{k_n} \sum_{j=0}^{k_n-1} (c_i^{n,lm})
\]

\[
V^{(4)}_t^n = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{N_n(t)-k_n + 1} \Delta_{n,i}
\]

\[
\times \left( g \left( c_i^n + \beta_i^n \right) - g \left( c_i^n \right) + \frac{1}{k_n} \sum_{l,m=1}^d \partial_{lm} g(\tilde{c}_i^n) \beta_i^{n,lm} \right)
\]

\[
- \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \partial^2_{jk,l,m} g(\tilde{c}_i^n + \beta_i^n)
\]

\[
\times \left( (c_i^{n,j} + \beta_i^{n,j}) (c_i^{n,km} + \beta_i^{n,km}) + (c_i^{n,jm} + \beta_i^{n,jm}) (c_i^{n,kl} + \beta_i^{n,kl}) \right)
\]

\[
V^{(5)}_t^n = \frac{1}{k_n \sqrt{k_n}} \sum_{i=1}^{N_n(t)-k_n + 1} \Delta_{n,i} \sum_{l,m=1}^d \partial_{lm} g(\tilde{c}_i^n) \sum_{j=0}^{k_n-1} (c_i^{n,lm}) \frac{1}{\Delta_{n,i+j}}
\]

193
C.2.2 Asymptotics for each component

Lemma C.2.1. For $v = 1, 2, 3, 4$ we have $V^{n,v} \overset{a.c.p.}{\longrightarrow} 0$.

$v = 1$

With some simple algebra, $\mathbb{E} \left( \sup \left| V_{s}^{n,1} \right| \right) \leq K t \sum_{i=1}^{N_{n}(t)} a_{n} \Delta_{n,i}^{(2p - r) \omega + 2 - p} \Delta_{n}^{-1/2}$ for some sequence $a_{n} \to 0$. Let $w = (2p - r) \omega + 1 - p - 1/2 > 0$,

$$\mathbb{E} \left( \sup \left| V_{s}^{n,1} \right| \right) \leq K t \sum_{i=1}^{N_{n}(t)} a_{n} \Delta_{n,i}^{w+1} \left( \frac{\Delta_{n,i}}{\Delta_{n}} \right)^{1/2} \to 0.$$ 

$v = 2$

$$V_{t}^{n,2} = \frac{1}{\sqrt{\Delta_{n}}} \sum_{i=1}^{N_{n}(t) - k_{n} + 1} \int_{T(n,i-1)}^{T(n,i)} (g(c_{i}^{n}) - g(c_{s})) \, ds - \frac{1}{\sqrt{\Delta_{n}}} \int_{T(n,N_{n}(t)-k_{n}+1)}^{t} g(c_{s}) \, ds.$$

Since $\sum_{i=1}^{j} \Delta_{n,i} \leq K \sum_{i=1}^{j} \Delta_{n} \leq K k_{n} \Delta_{n}$ for $j \leq k_{n}$,

$$\left| t - T(n, N_{n}(t) - k_{n} + 1) \right| \leq K k_{n} \sqrt{\Delta_{n}} \to 0,$$

and thus the second term goes to zero. For the first term, we turn to the proof of (5.3.24) in Jacod and Protter (2011) with modification. We want to prove that $\mathbb{E} \left( \sup \left| V_{s}^{n,2} \right| \right) \to 0$. Define

$$A^{n}(2)_{t} = \frac{1}{\sqrt{\Delta_{n}}} \left( \sum_{i=1}^{[t/\Delta_{n}]} \Delta_{n,i} \rho_{CT(n,i-1)}(f) - \int_{0}^{t} \rho_{c_{s}}(f) \, ds \right).$$

According to the notation of (3.4.4) $\mathcal{M}^{+}_{d \times d}$ is the set of all $d \times d$ symmetric nonnegative matrices: if $\rho \in \mathcal{M}^{+}_{d \times d}$, then $\rho_{a}$ denotes the centered Gaussian law with covariance
matrix $\alpha$, and $\rho_{\alpha} (f) = \int f(x) \rho_{\alpha} (dx)$ And our aim is to prove

$$A^n (2)_t \overset{u.c.p.} \Rightarrow 0. \quad (C.6)$$

Their (5.3.27) should still hold, that is

$$\psi (\alpha) = \rho_{\alpha \alpha^*} (f) = \mathbb{E} (f (\alpha U)).$$

A Taylor expansion gives $A^n (2)_t = -\eta^n_t - \sum_{i=1}^{N_n(t)} (\eta^n_i + \eta^n_{i+1} + \eta^n_{i-1})$, where

$$\eta^n_t = \frac{1}{\sqrt{\Delta_n}} \int_{T(n,N_n(t))}^t \psi (\sigma_s) ds,$$

$$\eta^n_i = \frac{1}{\sqrt{\Delta_n}} \nabla \psi (\sigma_{T(n,i-1)}) \int_{T(n,i-1)}^{T(n,i)} (\sigma_u - \sigma_{T(n,i-1)}) du,$$

$$\eta^n_{i+1} = \frac{1}{\sqrt{\Delta_n}} \int_{T(n,i-1)}^{T(n,i)} (\psi (\sigma_u) - \psi (\sigma_{T(n,i-1)})) - \nabla \psi (\sigma_{T(n,i-1)}) (\sigma_u - \sigma_{T(n,i-1)}) du.$$

By Doob’s inequality,

$$\mathbb{E} \left( \sup_{s \leq t} |\eta^n_s| \right) \leq K \left( t + \sqrt{t} \right) \sqrt{\Delta_n}.$$

Besides,

$$\mathbb{E} \int_{T(n,i-1)}^{T(n,i)} \| \sigma_u - \sigma_{T(n,i-1)} \|^2 du = \mathbb{E} \left( \mathbb{E} \left( \int_{T(n,i-1)}^{T(n,i)} \| \sigma_u - \sigma_{T(n,i-1)} \|^2 du | \mathcal{F}_{T(n,i-1)} \right) \right)$$

$$= \mathbb{E} \left( \int_{T(n,i-1)}^{T(n,i)} \mathbb{E} \left( \| \sigma_u - \sigma_{T(n,i-1)} \|^2 | \mathcal{F}_{T(n,i-1)} \right) du \right)$$

$$\leq K \Delta^2_{n,i}.$$
\[
\mathbb{E} \left( \int_{T(n,i)}^{T(n,i+1)} \left\| \sigma_u - \sigma_{T(n,i)} \right\| \, du \right) \\
\leq \mathbb{E} \left( \left( \int_{T(n,i)}^{T(n,i+1)} 1^2 \, du \right)^{1/2} \right) \left( \int_{T(n,i)}^{T(n,i+1)} \mathbb{E} \left[ \left\| \sigma_u - \sigma_{T(n,i)} \right\|^2 \right] \mathcal{F}_{T(n,i-1)} \right)^{1/2} \, du \\
\leq K \Delta_n^{3/2},
\]

Then

\[
\mathbb{E} \left( |\eta^n_t| \right) \leq K \psi(\varepsilon) \Delta_n^{3/2} \Delta_n^{-1/2} + \frac{K}{\varepsilon} \Delta_n^{2} \Delta_n^{-1/2} \leq K \psi(\varepsilon) \Delta_n + \frac{K}{\varepsilon} \Delta_n^{3/2}.
\]

So we have

\[
\limsup_n \mathbb{E} \left( \sup_{s \leq t} |A^n(t)_s| \right) \leq K \left( t + \sqrt{t} \right) \left( \psi(\varepsilon) + \frac{\sqrt{\Delta_n}}{\varepsilon} \right).
\]

The property that \( \lim_{\varepsilon \to 0} \psi(\varepsilon) = 0 \) gives Equation (C.6).

\( v = 3 \)

Define

\[
\zeta^n_i = \sum_{l,m=1}^d \Delta_{n,i} \hat{c}_{lm} g(c^n_i) \frac{1}{k_n} \sum_{u=0}^{k_n-1} \left( c_{i+u}^{n,lm} - c_{i}^{n,lm} \right).
\]

We let \( N(n,j,t) \) be the integer part of \( (N_j(t) - k_n j + 1)/k_n \), then

\[
V^{n,3}_t = \frac{1}{\sqrt{\Delta_n}} \sum_{j=1}^{k_n} H(j)_t^n, \text{ where } H(j)_t^n = \sum_{i=0}^{N(n,j,t)} \Delta_{n,j+k_n} \zeta^n_{j+k_n,i}.
\]
As \( \mathbb{E}( \sup_{w \in [0,s]} \| c_{t+w} - c_t \| | F_t ) \leq K s^{1\wedge (q/2)} \) and \( \| \mathbb{E}(c_{t+s} - c_t | F_s) \| \leq K s \),

\[
\| \mathbb{E}(z^n_i | F^n_i) \| \leq \sup_{i \in \{0, \ldots, k_n\}} \left| \sum_{l, m=1}^{d} \Delta_{lm} g(c^n_i) \left( \frac{1}{k_n} \sum_{u=0}^{k_n-1} \mathbb{E}(c^n_{i+u} - c^n_i | F^n_i) \right) \right| \leq K |T(n, i + k_n) - T(n, i)|.
\]

\[
\mathbb{E}(k^n_i^2 | F^n_i) \leq K \mathbb{E} \left( \left( \frac{1}{k_n} \sum_{u=0}^{k_n-1} (c^n_{i+u} - c^n_i)^2 \right) | F^n_i \right) \leq K |T(n, i + k_n) - T(n, i)|.
\]

By Doob’s inequality,

\[
\mathbb{E} \left( \sup_{s \leq t} |H(j)^n_i| \right) \leq \sum_{i=0}^{N(n, j, t)} \mathbb{E} \left( \left( \Delta_{n, j+k_n, i} \zeta^n_{j+k_n} | F^n_{j+k_n} \right) \right) + \left( 4 \sum_{i=0}^{N(n, j, t)} \mathbb{E} \left( \left( \Delta_{n, j+k_n, i} \zeta^n_{j+k_n} \right)^2 \right) \right)^{1/2} \leq K(t + \sqrt{t}) \sup |\Delta_{n,i}|
\]

\( V_t^{n,3} \leq \frac{1}{\sqrt{\Delta_n}} \sum_{j=1}^{k_n} |H(j)^n_t| \) and \( k_n \sqrt{\Delta_n} \to 0 \), so we get the result for \( v = 3 \).

\( v = 4 \)

Define the \( i \)th summand in the definition of \( V_t^{n,A} \) as \( v^n_i + w^n_i \), where

\[
\hat{c}_{i+n} = \frac{1}{k_n} \sum_{m=0}^{k_n-1} \frac{\Delta_{i+m} X' \Delta_{i+m} X' s}{\Delta_{n,i+m}}
\]

\[
\alpha^n_i = \Delta^n_i X' \Delta^n_i X' s - c^n_i \Delta_{n,i}, \quad \beta^n_i = \hat{c}_{i+n} - c^n_i = \frac{1}{k_n} \sum_{j=0}^{k_n-1} \frac{( \alpha^n_{i+j} + (c^n_{i+j} - c^n_i) \Delta_{n,i+j} )}{\Delta_{n,i+j}}.
\]
\[ v^n_i = \frac{1}{2} \sum_{j,k,l,m=1}^d \partial_{jk,lm}^2 g(c^n_i) \left( \beta^{n,jk}_i \beta^{n,lm}_i - \frac{1}{k_n} \left( c^{n,kl}_i c^{n,km}_i + c^{n,jm}_i c^{n,km}_i \right) \right), \]

then \( V_t^{n,4} = G^n_t + \sum_{j=1}^{k_n} H(j)_t^n \) where \( G^n_t = \sum_{i=0}^{N_n(t)-k_n+1} \frac{\Delta_n}{\sqrt{\Delta_n}} (w^n_i + E(v^n_i|F)) \).

We need to prove several lemmas.

**Lemma C.2.2.**

\[
\left| \mathbb{E} \left( \Delta X^i_t \Delta X^m_t | F^n_i \right) - c^{i,jm}_i \Delta_{n,i} \right| \leq K \Delta_{n,i}^{3/2} \left( \sqrt{\Delta_{n,i}} + \eta_{0,t} \right)
\]

\[
\left| \mathbb{E} \left( \Delta X^i_t \Delta X^{rk}_t \Delta X^l_t | F^n_i \right) - \left( c^{n,jk}_i c^{n,lm}_i + c^{n,jm}_i c^{n,kl}_i + c^{n,jm}_i c^{n,km}_i \right) \Delta_{n,i}^2 \right| \leq K \Delta_{n,i}^{5/2}
\]

where \( \eta_{t,s} = \sup \left( \| u_{t+u} - u_t \|^2 : u \in [0,s] \right) \)

\[
\eta^n_{i,j} = \sqrt{E \sup \left( \| u_u - u_{u_{t,n,i+j}} \|^2 : u \in [T_{t,n,i+j-1}, T_{t,n,i+j-1+k_n}] \right)}
\]

The following can easily be shown: \( \| E(\alpha^n_i | F^n_i) \| \leq K \Delta_{n,i}^{3/2} \left( \sqrt{\Delta_{n,i}} + \eta^n_{1,i} \right), \)

\[ \left| E(\alpha^n_i \alpha^n_l | F^n_i) - \left( c^{n,ji}_i c^{n,kl}_i + c^{n,jm}_i c^{n,km}_i \right) \Delta_{n,i}^2 \right| \leq K \Delta_{n,i}^{5/2}. \]

Indeed, \( \left| \mathbb{E} \left( \Delta X^i_t \Delta X^m_t | F_0 \right) - t c^{i,lm}_0 \right| \leq K t^{3/2} \left( \sqrt{t} + \eta_{0,t} \right), \) \( t = \Delta_{n,i} \) gives the first claim. The proof of the second claim is straightforward.

**Lemma C.2.3.** For all \( t > 0 \) we have \( \Delta_{n,i} E \left( \sum_{i=1}^{N_n(t)} \eta^n_i \right) \rightarrow 0, \) and for all \( j, k \) such that \( j + k \leq k_n \) we have \( E \left( \eta^n_{i+j,k} | F^n_i \right) \leq \eta^n_i \)

For the first claim, as
\begin{align*}
\eta_i^n &= \sqrt{\mathbb{E} \sup \left( \left\| u - u_{T_n,i-1}^n \right\|^2 : u \in [T_{n,i-1}, T_{n,i+k_n-1}]|F_{T_n,i-1}^{n} \right)}, \\
\mathbb{E} \left( \left( \eta_i^n \right)^2 \right) &= \mathbb{E} \left( \mathbb{E} \sup \left( \left\| u - u_{T_n,i}^n \right\|^2 : u \in [T_{n,i-1}, T_{n,i+k_n-1}]|F_{T_n,i-1}^{n} \right) \right) \leq K.
\end{align*}

When \( i \geq 2 \), if \( s \in (T_{n,i-2}, T_{n,i-1}) \), since

\[ (T_{n,i-1}, T_{n,i+k_n-1}) \subseteq (T_{n,i-1}, T_{n,i-2+k_n+1}) \subseteq (s, 2k_n + 1), \]

one immediately gets

\[ \mathbb{E} \left( \left( \eta_{s,2k_n+1} \right)^2 \right) \geq \mathbb{E} \left( \left( \eta_i^n \right)^2 \right) \text{ for } \forall s \in (T_{n,i-2}, T_{n,i-1}). \]

Therefore,

\[ \mathbb{E} \left( \left( \eta_i^n \right)^2 \right) \leq \frac{1}{\Delta_{n,i}} \int_{T_{n,i-2}}^{T_{n,i-1}} \mathbb{E} \left( \left( \eta_{s,2k_n+1} \right)^2 \right) \ ds. \]

Then one can get that

\[ \mathbb{E} \left( \sum_{i=1}^{N_n(t)} \Delta_n \eta_i^n \right) \leq \left( K t \Delta_n + \mathbb{E} \left( t \int_0^t \left( \eta_{s,2k_n+1} \right)^2 \ ds \right) \right)^{1/2}, \]

and the first claim is proved. For the second claim,

\[ \eta_{T_{n,i-1}, T_{n,i-1+k_n}, T_{n,i-1}} = \sup \left( \left\| u - u_{T_n,i-1} \right\|^2 : u \in [T_{n,i-1}, T_{n,i-1+k_n}] \right), \]
\[ \eta_{n,j,kn}^n = \sqrt{E \sup \left( \left\| u - u_{T_n,i+j-1}^\beta \right\|^2 : u \in [T_{n,i+j-1}, T_{n,i+j-1+\kappa_n}^\beta] \right) \}, \]

\[ E \left( \eta_{n,i+k,i}^n \mid F_{i}^n \right) \leq \sqrt{E \left( E \sup \left( \left\| u - u_{T_n,i-1}^\beta \right\|^2 : u \in [T_{n,i-1}, T_{n,i-1+\kappa_n}^\beta] \right) \mid F_{i}^n \right) \mid F_{i}^n } \]

\[ = \sqrt{E \sup \left( \left\| u - u_{T_n,i-1}^\beta \right\|^2 : u \in [T_{n,i-1}, T_{n,i-1+\kappa_n}^\beta] \right) \mid F_{i}^n } \]

\[ = \eta_{i,k}^n. \]

**Lemma C.2.4.**

\[ \| E (\beta_i^n \mid F_{i}^n) \| \leq K \sqrt{\Delta_n} \left( k_n \sqrt{\Delta_n + \eta_i^n} \right) \]

\[ \left\| E \left( \beta_i^n, \beta_i^n \mid F_{i}^n \right) - \frac{1}{k_n} \left( c_i^{n,jk}, c_i^{n,jm} c_i^{n,kl} \right) \right\| \]

\[ \leq K \sqrt{\Delta_n} \left( k_n^{-1/2} + k_n \sqrt{\Delta_n + \eta_i^n} \right) \]

From Lemma C.2.2 we know that \( \| E (\alpha_i^n \mid F_{i}^n) \| \leq K \Delta_{n,i}^{3/2} (\sqrt{\Delta_{n,i}} + \eta_{0,t}) \).

\[ \left\| E \left( \alpha_i^{n,jk} \alpha_i^{n,jm} \right) \mid F_{i}^n \right\| \leq \left\| E \left( \alpha_i^{n,jk} \mid F_{i}^n \right) \right\| + \left\| E \left( (c_i^n - c_i^n) \Delta_{n,j} \mid F_{i}^n \right) \right\| \]

\[ \leq K \left\| E \left( \Delta_{n,i}^{3/2} (\sqrt{\Delta_{n,i}} + \eta_{n,i+1}^n) \mid F_{i}^n \right) \right\| \leq K \Delta_{n,i}^{3/2} \left( \Delta_{n,i} + \eta_{n,i+1}^n \right) \]

With the assumption that \( \Delta_{n,i} = O_p (\Delta_n) \),

\[ \mathbb{E} \left( \left\| \sum_{j=0}^{k_n-1} \alpha_i^{n,j} \mid F_{i}^n \right\|^q \right) \leq K q \Delta_{n,i}^{q/2} k_n^{q/2}. \]

\[ \beta_i^n = \hat{\beta}_i^n - c_i^n = \frac{1}{k_n} \sum_{j=0}^{k_n-1} \frac{(c_i^n - c_i^n) \Delta_{n,i+j}}{\Delta_{n,i+j}} \]
\[ \mathbb{E}(\|\beta^n\|^q | F^n_i) \leq K_q \left( \mathbb{E}\left( \left\| \frac{1}{k_n} \sum_{j=0}^{k_n-1} \frac{\alpha_{i+j}^n}{\Delta_{n,i+j}} \right\|^q | F_i^n \right) + \mathbb{E}\left( \left\| \frac{1}{k_n} \sum_{j=0}^{k_n-1} (c_{i+j}^n - c_i^n) \right\|^q | F_i^n \right) \right). \]

\[ \mathbb{E}\left( \left\| \frac{1}{k_n} \sum_{j=0}^{k_n-1} \frac{\alpha_{i+j}^n}{\Delta_{n,i+j}} \right\|^q | F_i^n \right) \leq K_q k^{-q} \Delta_n^{-q} \mathbb{E}\left( \left\| \frac{1}{k_n} \sum_{j=0}^{k_n-1} \alpha_{i+j}^n \right\|^q | F_i^n \right) \]

\[ \leq K_q k^{-q} \Delta_n^{-q} \Delta_n k_n^{q/2} = K_q k_n^{-q/2}. \]

\[ \mathbb{E}\left( \left\| \frac{1}{k_n} \sum_{j=0}^{k_n-1} (c_{i+j}^n - c_i^n) \right\|^q | F_i^n \right) \leq \mathbb{E}\left( \left( \frac{1}{k_n} \sum_{j=0}^{k_n-1} (c_{i+j}^n - c_i^n)^2 \right)^{q/2} | F_i^n \right) \leq \mathbb{E}(\sup_{j \leq k_n} (c_{i+j}^n - c_i^n)^q | F_i^n) \]

\[ \leq K_q |T_{n,i+k_n-1} - T_{n,i-1}| = O_p(k_n \Delta_n). \]

\[ \|\mathbb{E}(\alpha_i^n | F_i^n)\| \leq K \Delta_{n,i}^{3/2} \left( \sqrt{\Delta_{n,i} + \eta_{i,1}^n} \right), \]

\[ \|\mathbb{E}(\alpha_{i+j}^n | F_i^n)\| \leq K \Delta_{n,i+j}^{3/2} \left( \sqrt{\Delta_{n,i+j} + \eta_{i,1}^n} \right). \]

Use the strong predictability condition and that \( \Delta_{n,i} = O_p(\Delta_n) \),

\[ \|\mathbb{E}(\beta_i^n | F_i^n)\| \leq \frac{1}{k_n} \sum_{j=0}^{k_n-1} \mathbb{E}\left( \frac{\alpha_{i+j}^n}{\Delta_{n,i+j}} | F_i^n \right) \leq + \frac{1}{k_n} \sum_{j=0}^{k_n-1} \mathbb{E}(c_{i+j}^n - c_i^n | F_i^n) \]

\[ \leq K_q |T_{n,i+k_n-1} - T_{n,i-1}| + K \frac{1}{k_n} |T_{n,i+k_n-1} - T_{n,i-1}| \]

+ \[ + K \frac{1}{k_n} \sum_{j=0}^{k_n-1} \sqrt{\Delta_{n,i+j} \eta_{i,1}^n} \]

\[ \leq K_q k_n \Delta_n + K \sqrt{\Delta_n \eta_i^n}. \]
Next we turn to \( \left| \mathbb{E} \left( \frac{\beta_i^{n,j}}{k_n} \beta_i^{n,l,m} \big| F_i^n \right) \right| - \frac{1}{k_n} \left( c_i^{n,j,l} c_i^{n,k,m} + c_i^{n,j,m} c_i^{n,k,l} \right) \). Use the following notations:

\[
\xi_i^n = c_i^{n,j,l} c_i^{n,k,m} + c_i^{n,j,m} c_i^{n,k,l}, \quad \varsigma_i^n = \alpha_i^n + (c_i^{n,j} - c_i^n) \Delta_{n,i+j}, \quad (C.7)
\]

\[
\beta_i^{n,j,k} = \beta_i^{n,l,m} = \frac{1}{k_n} \sum_{u=0}^{k_n-1} \frac{\epsilon_i^{n,j,k} \epsilon_i^{n,l,m}}{\Delta_{n,i+u}} + \frac{1}{k_n} \sum_{u=0}^{k_n-1} \sum_{v=u+1}^{k_n-1} \frac{\epsilon_i^{n,j,k} \epsilon_i^{n,l,m}}{\Delta_{n,i+u} \Delta_{n,i+u+v}} + \frac{1}{k_n} \sum_{u=0}^{k_n-1} \sum_{v=u+1}^{k_n-1} \frac{\epsilon_i^{n,l,m} \epsilon_i^{n,j,k}}{\Delta_{n,i+u} \Delta_{n,i+u+v}}.
\]

We work on the three terms one by one.

\[
\left| \frac{\epsilon_i^{n,j,k} \epsilon_i^{n,l,m} - \alpha_i^{n,j,k} \alpha_i^{n,l,m}}{\Delta_{n,i+u}} \right| \leq 2 \Delta_{n,i+u} \left( \| \alpha_i^n \|^2 + \Delta_{n,i+u}^2 \| \alpha_i^n \|^2 \right).
\]

Let \( \delta_{n,i} = |T_{n,i+kn-1} - T_{n,i-1}| \), then

\[
\left\| \mathbb{E} \left( \Delta_{n,i+u} \left\| c_i^n - c_i^n \right\| \big| F_i^n \right) \right\| \leq \Delta_{n,i+u} \sqrt{\mathbb{E} \left( \left\| c_i^n - c_i^n \right\|^2 \big| F_i^n \right) \right) \left\| \mathbb{E} \left( \left\| \alpha_i^n \right\|^2 \big| F_i^n \right) \right\| \leq K \delta_{n,i}^{1/2} \Delta_{n,i+u}^2.
\]

\[
\left\| \mathbb{E} \left( \frac{\epsilon_i^{n,j,k} \epsilon_i^{n,l,m}}{\Delta_{n,i+u}} - \alpha_i^{n,j,k} \alpha_i^{n,l,m} \big| F_i^n \right) \right\| \leq K \left( \Delta_{n,i+u}^{5/2} + \delta_{n,i}^{1/2} \Delta_{n,i+u}^2 \right) \leq K \Delta_{n}^{5/2} k_n^{1/2}.
\]

By boundness of \( c_i \) and (4.3) in Jacod et al. (2013), \( \left| \mathbb{E}(\xi_i^n | F_i) - \xi_i^n \right| \leq K \delta_{n}, \)

\[
\mathbb{E} \left( \frac{1}{k_n} \sum_{u=0}^{k_n-1} \frac{\epsilon_i^{n,j,k} \epsilon_i^{n,l,m}}{\Delta_{n,i+u}} \big| F_i \right) \right) - \frac{1}{k_n} \left( c_i^{n,j,l} c_i^{n,k,m} + c_i^{n,j,m} c_i^{n,k,l} \right) \]

\[
\leq \mathbb{E} \left( \frac{1}{k_n} \sum_{u=0}^{k_n-1} \frac{\epsilon_i^{n,j,k} \epsilon_i^{n,l,m}}{\Delta_{n,i+u}^2} - \alpha_i^{n,j,k} \alpha_i^{n,l,m} \big| F_i \right) \right) \Delta_{n,i+u}^2 + \mathbb{E} \left( \frac{1}{k_n} \sum_{u=0}^{k_n-1} \alpha_i^{n,j,k} \alpha_i^{n,l,m} - \left( c_i^{n,j,l} c_i^{n,k,m} + c_i^{n,j,m} c_i^{n,k,l} \right) \big| F_i \right) \right) \Delta_{n,i+u}^2.
\]

202
We find the upper bound for each term.

\[
\mathbb{E} \left( \frac{1}{k_n^2} \sum_{u=0}^{k_n-2} \sum_{v=u+1}^{k_n-2} \frac{\alpha_{n,jk} \alpha_{n,jk}}{\Delta_{n,i+v} \Delta_{n,i+v}} | F_i^n \right) \leq K \Delta_n^{1/2} \left( \sqrt{\Delta_n^{1/2} + \eta_n^{1/2}} \right).
\]

Since

\[
\mathbb{E} \left( \left( c_{n,jk}^{i,i} - c_{n,jk}^{i,i+1} \right) (c_{n,jk}^{n,jk} - c_{n,jk}^{i,i}) | F_i^n \right) \leq K \delta_n,
\]

we have

\[
\mathbb{E} \left( \frac{1}{k_n^2} \sum_{u=0}^{k_n-2} \sum_{v=u+1}^{k_n-2} \frac{\alpha_{n,jk} \alpha_{n,jk}}{\Delta_{n,i+v} \Delta_{n,i+v}} | F_i^n \right) \leq K \Delta_n^{1/2} \left( \sqrt{\Delta_n^{1/2} + \eta_n^{1/2}} \right).
\]
\[ \leq K\delta_n \leq Kk_n\Delta_n. \]

As \( c_i \) is bounded,
\[
\left\| E \left( \frac{1}{k_n^2} \sum_{u=0}^{k_n-2} \sum_{v=u+1}^{k_n-1} \left( \frac{\alpha_i^{n,jk}}{\Delta_n,i+u} |F^n_{i+u}| \right) \right\| \leq \frac{1}{k_n^2} \sum_{u=0}^{k_n-2} \sum_{v=u+1}^{k_n-1} E \left( \left\| E \left( \frac{\alpha_i^{n,jk}}{\Delta_n,i+u} |F^n_{i+u}| \right) \right\| \right) \]
\[
\leq K \frac{1}{k_n} \sum_{u=0}^{k_n-2} \Delta_{n,i+u}^{1/2} \left( \sqrt{|\Delta_n,i+u| + \eta_{i+u,1}^n} |F^n_{i+u}| \right) \]
\[
\leq K \frac{1}{k_n} \sum_{u=0}^{k_n-2} \Delta_{n,i+u}^{1/2} \left( \sqrt{|\Delta_n,i+u| + \eta_{i,1}^n} \right) \]
\[
\leq K\Delta_n^{1/2} \left( \sqrt{|\Delta_n| + \eta_{i,1}^n} \right) \]

Similarly,
\[
\left\| E \left( \frac{1}{k_n^2} \sum_{u=0}^{k_n-2} \sum_{v=u+1}^{k_n-1} \left( \frac{\alpha_i^{n,jk}}{\Delta_n,i+v} |F^n_i| \right) \right\| \leq \frac{1}{k_n^2} \sum_{u=0}^{k_n-2} \sum_{v=u+1}^{k_n-1} E \left( \left\| E \left( \frac{\alpha_i^{n,jk}}{\Delta_n,i+v} |F^n_{i+u}| \right) \right\| \right) \]
\[
\leq K\Delta_n^{1/2} \left( \sqrt{|\Delta_n| + \eta_{i,1}^n} \right) \]

Add the four components up we get the last two terms of
\[
|\beta_i^{n,jk} \beta_i^{n,lm} - \frac{1}{k_n} \left( c_i^{n,jl} c_i^{n,km} + c_i^{n,jm} c_i^{n,kl} \right) | \]

is bounded by \( Kk_n\Delta_n + K\sqrt{\Delta_n} (\sqrt{|\Delta_n| + \eta_{i,1}^n}) \).

So we find the upper bound: \( K\sqrt{|\Delta_n| \left( k_n^{-1/2} + k_n \sqrt{|\Delta_n| + \eta_{i,1}^n} \right) \}

The proof of Lemma C.2.4 is finished.

Next we continue the proof of Lemma 4.
\[
|w_i^n| \leq K \left( 1 + \| \beta_i^n \|^{p-3} \| \beta_i^n \|^{3} + \frac{k_n}{k_n} \left( 1 + \| \beta_i^n \|^{p-1} \right) \| \beta_i^n \|. \right) V_i^n = G_i^n + \sum_{j=1}^{k_n} H(j)_i^n \text{ as defined below:}
\]

\[
G_i^n = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{N_n(t)-k_n+1} \Delta_{n,i} \left( w_i^n + E \left( v_i^n | F_i^n \right) \right), \quad H(j)_i^n = \sum_{i=0}^{N(n,j,i)} \zeta(j)_i^n,
\]

where \( \zeta(j)_i^n = \frac{\Delta_{n,i}}{\sqrt{\Delta_n}} \left( v_{j+k_{n,i}} - E \left( v_{j+k_{n,i}} | F_{j+k_{n,i}} \right) \right). \)

From 4.10 in Jacod et al. (2013) we know that for all \( q \geq 0, \)

\[
E \left( \| \alpha_i^n \|^q | F_i^n \right) \leq K q \Delta_{n,i}^q, \quad \| E(\alpha_i^n | F_i^n) \| \leq K \Delta_{n,i}^{3/2}.
\]

Besides, one can show that \( E \left( \left\| \sum_{j=0}^{k_n-1} \alpha_i^{n,j} \right\|^q | F_i^n \right) \leq K k_n^{3/2} \Delta_{n,i}^q, \) and these lead to

\[
E \left( |w_i^n| \right) \leq K \left( \frac{1}{k_n^{3/2}} + k_n \Delta_n + \frac{\sqrt{k_n \Delta_n}}{k_n} \right). \quad (C.8)
\]

Then

\[
E \left( \sum_{i=1}^{N_n(t)-k_n+1} \frac{\Delta_{n,i}}{\sqrt{\Delta_n}} |w_i^n| \right) \leq K \sum_{i=1}^{N_n(t)-k_n+1} \frac{\Delta_{n,i}}{\sqrt{\Delta_n}} \left( \frac{1}{k_n^{3/2}} + k_n \Delta_n + \frac{\sqrt{k_n \Delta_n}}{k_n} \right) \to 0.
\]

With the assumption that \( \Delta_{n,i} = O_p(\Delta_n) \) and \( \Delta_n = O_p(\Delta_{n,i}), \) we get

\[
E \left( \sum_{i=1}^{N_n(t)-k_n+1} \sqrt{\Delta_{n,i}} |E \left( v_i^n | F_i^n \right)| \right) \leq K \sum_{i=1}^{N_n(t)-k_n+1} \sqrt{\Delta_{n,i}} K \sqrt{\Delta_n} \left( k_n \sqrt{\Delta_n} + \eta_i^n \right)
\]

\[
= K \sum_{i=1}^{N_n(t)-k_n+1} \Delta_{n,i}^n \eta_i^n \sqrt{\Delta_n} + K \sum_{i=1}^{N_n(t)-k_n+1} k_n \Delta_n^{3/2} \frac{\sqrt{\Delta_{n,i}}}{\sqrt{\Delta_n}}
\]

\[
\to 0,
\]

205
\[ E \left( \sup_{s \leq t} |G^n_s| \right) = E \left( \sup_{s \leq t} \left| \frac{N_n(t) - k_n + 1}{\sqrt{\Delta_n}} \left( w^n_i \right) E \left( v^n_i | F^n_i \right) \right| \right) \]

\[ \leq E \left( \sum_{i=1}^{N_n(t) - k_n + 1} \frac{\Delta_{n,i}}{\sqrt{\Delta_n}} \left( |w^n_i| + \left| E \left( v^n_i | F^n_i \right) \right| \right) \right) \rightarrow 0. \]

\[ H(j)_i^n = \sum_{j=0}^{N(n,j,t)} \zeta(j)_i^n \text{ where } \zeta(j)_i^n = \Delta_{n,j,k_n} \left( v^n_{j+k_n+i} - E \left( v^n_{j+k_n+i} | F^n_{j+k_n+i} \right) \right). \]

\[ E \left( \sup_{s \leq t} |H(j)_i^n| \right) \leq \left( \sum_{j=0}^{N(n,j,t)} E \left( |\zeta(j)_i^n|^2 \right) \right)^{1/2} \]

\[ \leq K \left( \sum_{j=0}^{N(n,j,t)} k_n \Delta^2_{n,j+k_n} \right)^{1/2} = K t \sup \frac{\Delta_{n,i}}{\sqrt{\Delta_n}} \rightarrow 0. \]

Since \( |V^n_{t,n}| \leq |G^n_t| + \sum_{j=1}^{k_n} |H(j)_i^n| \), we deduce the result for \( v = 4 \).

C.2.3 The component that gives the central limit theorem

Lemma C.2.5. With \( Z \) as in Theorem 4.4.2, we have the functional stable convergence in law

\[ V^n_{t,5} \xrightarrow{\mathcal{L}} Z. \]

Proof: \( V^n_{t,5} = \sqrt{\Delta_n} \sum_{i=1}^{N_n(t) - k_n + 1} w^n_i \Delta_{n,i} \) where

\[ w^n_i = \frac{(i-1) \wedge (k_n-1)}{k_n} \sum_{j=(i-N_n(t)+k_n-1)^+}^{(i-1) \wedge (k_n-1)} \partial_{lm} (c^n_{i-j}) \Delta_{n,i-j} \rightarrow \partial_{lm} (c^n_i) \frac{\Delta_{n,i}}{\Delta_n}. \]
We need the convergence of both the two parts: \( \frac{\partial_{im} g (c_i^n)_{r \to} \Delta_n, i \to j} {\Delta_n} \), for \( j = 0, 1, \cdots, k_n \), then

\[
V_i^{n,5} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{N_n(t)} \frac{1}{\Delta_n} \sum_{l,m=1}^{d} \frac{(i-1) \Delta_n + (k_n-1)}{\sum_{l,m=1}^{d} \Delta_{n,j} \alpha_{i}^{n,lm}} \Delta_{n,j} \frac{\Delta_{n, i} \alpha_{i}^{n,lm}}{\Delta_{n,i}}.
\]

\[
\Delta_n \sum_{i=1}^{N_n(t)} \Delta_{n,i} \frac{w_i^{n,jk} w_i^{n,lm}}{E} \left( \frac{\alpha_i^{n,lm} \alpha_i^{n,jk}}{\Delta_{n,i}} \right) \frac{E_i^{n}}{E_i^{n}}.
\]

Let \( \tilde{w}_i^{n,jk} = \partial_{lm} g (c_i^n) \frac{\Delta_{n,i}}{\Delta_n} \) and \( w_i^{n,lm} = \frac{1}{\Delta_n} \sum_{j=(i-N_n(t)+k_n-1)}^{(i-1) \Delta_n + (k_n-1)} \partial_{lm} g (c_i^n) \frac{\Delta_{n,i}}{\Delta_n} \), we need to prove that \( \Delta_n \sum_{i=1}^{N_n(t)} \Delta_{n, i} \frac{w_i^{n,jk} w_i^{n,lm} - \tilde{w}_i^{n,jk} \tilde{w}_i^{n,lm}}{E} \left( c_i^{n,lm} c_i^{n,jk} + c_i^{n,km} c_i^{n,kl} + O_p(\Delta_{n,i}) \right) \to 0 \).

As both \( c_i^n \) and \( w_i^n \) are bounded, this is satisfied if \( w_i^{n,jk} - \tilde{w}_i^{n,jk} \to 0 \).

\[
w_i^{n,lm} - \tilde{w}_i^{n,jk} = \frac{1}{\Delta_n} \sum_{j=0}^{k_n-1} \partial_{lm} g (c_i^n) \frac{\Delta_{n,i} - j} {\Delta_n} - \partial_{lm} g (c_i^n) \frac{\Delta_{n,i} - j} {\Delta_n}.
\]

\[
\partial_{lm} g (c_i^n) \left( \frac{\sum_{j=0}^{k_n-1} \Delta_{n,i} - j} {k_n \Delta_n} - \frac{\Delta_{n,i} - j} {\Delta_n} \right)
\]

\[
+ \frac{1}{k_n} \sum_{j=0}^{k_n-1} \left( \partial_{lm} g (c_i^n) - \partial_{lm} g (c_i^n) \right) \frac{\Delta_{n,i} - j} {\Delta_n}.
\]

As \( \lim_{k_n \Delta_n \to 0} \sum_{j=0}^{k_n-1} \frac{\Delta_{n,i} - j} {k_n \Delta_n} \), the result holds with the assumption that \( T (\cdot) \) is continuously differentiable. Then by the assumption in The-
orem 3.1 that \( g \) is a \( C^3 \) function on \( \mathcal{M}_a^+ \) such that \( \| \partial_j g(x) \| \leq K \left( 1 + \| x \|^p \right) \) and \( c_s \) is bounded we conclude that \( \partial_{lm} g(c^n_i) \) is bounded. Therefore,

\[
\lim_{k_n \Delta_n \to 0} \partial_{lm} g(c^n_i) \left( \frac{\sum_{j=0}^{k_n-1} \Delta_{n,i-j}}{k_n \Delta_n} - \frac{\Delta_{n,i}}{\Delta_n} \right) = 0.
\]

Since \( \partial_{lm} g(c^n_{i-j}) \) is continuous and \( c^n_{i-j} \to c^n_i \) as \( n \to \infty \), as well as that \( \frac{\Delta_{n,i-j}}{\Delta_n} \leq K \), we have

\[
\frac{1}{k_n} \sum_{j=0}^{k_n-1} \left( \partial_{lm} g(c^n_{i-j}) - \partial_{lm} g(c^n_i) \right) \frac{\Delta_{n,i-j}}{\Delta_n} \xrightarrow{n \to \infty} 0.
\]

Hence \( w^{n,lm}_i - \tilde{w}^{n,lm}_i \to 0 \). As \( c \) is bounded, this leads to that

\[
\left( \tilde{w}^{n,lm}_i w^{n,jk}_i - w^{n,lm}_i w^{n,jk}_i \right) \left( c^n_i c^n_{n,jk} + c^n_{i,jm} c^n_{i,kl} \right) \xrightarrow{n \to \infty} 0.
\]

With the above conclusion,

\[
\Delta_n \sum_{i=1}^{N_n(t) - k_n + 1} w^{n,lm}_i \partial_{lm} g(c^n_i) \left( c^n_i c^n_{n,jk} + c^n_{i,jm} c^n_{i,kl} \right)
\]

\[
= \Delta_n \sum_{i=1}^{N_n(t) - k_n + 1} \left( \tilde{w}^{n,lm}_i \tilde{w}^{n,jk}_i - w^{n,lm}_i w^{n,jk}_i \right) \left( c^n_i c^n_{n,jk} + c^n_{i,jm} c^n_{i,kl} \right)
\]

\[
+ \Delta_n \sum_{i=1}^{N_n(t) - k_n + 1} \tilde{w}^{n,lm}_i \tilde{w}^{n,jk}_i \left( c^n_i c^n_{n,jk} + c^n_{i,jm} c^n_{i,kl} \right)
\]

\[
\to \sum_{i=1}^{N_n(t) - k_n + 1} \partial_{jk} g(c^n_i) \partial_{lm} g(c^n_i) \left( \frac{\Delta_{n,i}}{\Delta_n} \right)^2 \left( c^n_i c^n_{n,jk} + c^n_{i,jm} c^n_{i,kl} \right) \Delta_n
\]

\[
= \sum_{i=1}^{N_n(t) - k_n + 1} \partial_{jk} g(c^n_i) \partial_{lm} g(c^n_i) \frac{\Delta_{n,i}}{\Delta_n} \left( c^n_i c^n_{n,jk} + c^n_{i,jm} c^n_{i,kl} \right) \Delta_{n,i}
\]

208
Define \( T'_n(u) = \frac{\Delta_n u}{\Delta_n} \) for \( T(n, i - 1) < u < T(n, i) \). Then

\[
\Delta_n \sum_{i=1}^{N_n(t)-k_n+1} w_i^{n,jk} w_i^{n,lm} \left( c_i^{n,lm} c_i^{n,jk} + c_i^{n,jm} c_i^{n,kl} \right)
\]

\[
= \sum_{i=1}^{N_n(t)-k_n+1} \partial_{jk} g \left( c_i^{n} \right) \partial_{lm} g \left( c_i^{n} \right) T'_n \left( T(n, i) \right) \left( c_i^{n,lm} c_i^{n,jk} + c_i^{n,jm} c_i^{n,kl} \right) \Delta_{n,i}
\]

For each \( t \) we have \( T'_n \left( T(n, i(t), t) \right) \to T'(t) \), and as \( k_n \Delta_n \to 0 \),

\[
\partial_{jk} g \left( c_i^{n(t)} \right) \partial_{lm} g \left( c_i^{n(t)} \right) \to \partial_{jk} g \left( c_s^{n} \right) \partial_{lm} g \left( c_s^{n} \right) \text{ a.s.},
\]

\[
c_i^{n,lm} c_i^{n,jk} + c_i^{n,jm} c_i^{n,kl} \to c_t^{n,lm} c_t^{n,jk} + c_t^{n,jm} c_t^{n,kl} \text{ a.s.}
\]

By the dominated convergence theorem,

\[
\Delta_n \sum_{i=1}^{N_n(t)-k_n+1} w_i^{n,jk} w_i^{n,lm} \left( c_i^{n,lm} c_i^{n,jk} + c_i^{n,jm} c_i^{n,kl} \right) \geq 0
\]

\[
P \int_0^t \partial_{jk} g \left( c_s^{n} \right) \partial_{lm} g \left( c_s^{n} \right) T'(s) \left( c_s^{n,lm} c_s^{n,jk} + c_s^{n,jm} c_s^{n,kl} \right) ds.
\]

This is consistent with Renault et al. (2017)'s conclusion, though the definition of \( T'(s) \) is different. Now let the other terms go to zero.

\[
\sqrt{\Delta_n} \sum_{i=1}^{N_n(t)-k_n+1} \sum_{l,m=1}^{d} w_i^{n,lm} \left( \frac{\alpha_i^{n,lm}}{\Delta_{n,i}} \right) \leq K \sqrt{\Delta_{n,i}} \left( \sqrt{\Delta_{n,i}} + \eta_{i,1} \right) \to 0.
\]
Since in lemma C.2.3 we get $\Delta_{n,i}E \left( \sum_{i=1}^{\lceil t/\Delta_n \rceil} \eta_i^n \right) \rightarrow 0$.

$$\Delta_n^2 \sum_{i=1}^{N_n(t)-k_n+1} \|w_i^n\|^4 E \left( \left\| \frac{\eta_i^n}{\Delta_{n,i}} \right\|^4 |F_i^n \right) \leq \Delta_n^2 \sum_{i=1}^{N_n(t)-k_n+1} K \rightarrow 0.$$ 

$$\mathbb{E} \left( \alpha_i^{n,lm} \Delta_i^n N |F_i^n \right) = 0$$ for all $N$ defined as $N$ where $N = W^k$ for some $k$, or is an arbitrary bounded martingale, orthogonal to $W$, so

$$\sqrt{\Delta_n} \sum_{i=1}^{N_n(t)-k_n+1} \Delta^{n,lm}_i N \mathbb{E} \left( \frac{\alpha_i^{n,lm}}{\Delta_{n,i}} \Delta_i^n N |F_i^n \right) \rightarrow 0.$$ This finishes the proof.

C.3 Proof for the Alternative Estimator (without Overlap)

$$V^n(g)_i^n = \sum_{i=0}^{[N_n(t)/k_n]-1} (T(n, (i+1) k_n) - T(n, ik_n))$$

$$\times \left( g \left( \tilde{c}_{ikn+1}^n \right) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \partial_{l,j,k,m}^2 g(\tilde{c}_{ikn+1}^n) \left( \tilde{c}_{j,km}^n + \tilde{c}_{j,lm}^n + \tilde{c}_{j,kl}^n \right) \right).$$

We first look at the fifth component:

$$V^n_{i,5} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=0}^{[N_n(t)/k_n]-1} \left( T(n, (i+1) k_n) - T(n, ik_n) \right) \frac{\partial_{lm} g(\tilde{c}_{ikn+1}^n)}{k_n} \sum_{j=0}^{\frac{k_n-1}{k_n}} \frac{\alpha_{ikn+1+j}^{n,lm}}{\Delta_{i,n,ikn+1+j}}$$

$$= \sqrt{\Delta_n} \sum_{i=1}^{[N_n(t)/k_n]} \left( T(n, \left( \left\lfloor \frac{i-1}{k_n} \right\rfloor + 1 \right) k_n) - T(n, \left\lfloor \frac{i-1}{k_n} \right\rfloor k_n) \right) \frac{\partial_{lm} g(\tilde{c}_{ikn+1}^n)}{k_n}$$

$$\times \sum_{l,m=1}^d \frac{\alpha_{ikn+1+j}^{n,lm}}{\Delta_{i,n,ikn+1+j}}$$

$$= \sqrt{\Delta_n} \sum_{i=1}^{[N_n(t)/k_n]} \Delta_{i,n} \sum_{l,m=1}^d w_i^n \alpha_{i,n,lm}^{n,lm} \frac{\alpha_{i}^{n,lm}}{\Delta_{i,n}}$$

210
\[ w_{i,n}^{n,lm} = \left( T\left(n, \left[ \frac{i-1}{k_n} \right] + 1 \right) k_n \right) - T\left(n, \left[ \frac{i-1}{k_n} \right] k_n \right) \sum_{l,m=1}^{d} \partial_{lm} \left(c_{k_n}^{n,\left[ \frac{i-1}{k_n} \right]+1} \right). \]

With the strong predictability assumption,

\[ \lim_{k_n \Delta_n \to 0} \frac{T\left(n, \left[ \frac{i-1}{k_n} \right] + 1 \right) k_n}{k_n \Delta_n} - T\left(n, \left[ \frac{i-1}{k_n} \right] k_n \right) = \lim_{k_n \Delta_n \to 0} \frac{\Delta_{n,i}}{\Delta_n}, \]

if the derivative changes slowly over time.

Then \( w_{i,(n,t)}^{n,lm} \to \partial_{lm} \left( c_{t}^{n,\left[ \frac{i(n,t)-1}{k_n} \right]} + 1 \right) \) \( c_{t}^{n} \) a.s.. \)

\[ \Delta_n \sum_{i=1}^{[N(n,t)/k_n]k_n} w_{i,n}^{n,ik} w_{i,n}^{n,lm} \left( c_{i}^{n,ik} c_{i}^{n,lm} + c_{i}^{n,jm} \right) \]

\[ k_n \Delta_n \to 0 \Delta_n \sum_{i=1}^{[N(n,t)/k_n]k_n} \left( \frac{\Delta_{n,i}(n,t)}{\Delta_n} \right)^2 \partial_{jk} \left( c_{i(n,t)} \right) \partial_{jk} \left( c_{i(n,t)} \right) \left( c_{i(n,t)}^{n,ik} c_{i(n,t)}^{n,lm} + c_{i(n,t)}^{n,jm} \right) \]

\[ \mathbb{P} \int_{0}^{t} \partial_{jk} \left( c_{s} \right) \partial_{lm} \left( c_{s} \right) T \left( s \right) \left( c_{s}^{n,ik} c_{s}^{n,lm} + c_{s}^{n,jk} c_{s}^{n,lm} \right) ds. \]

by the dominated convergence theorem.

For the other components: \( \nabla_{n,1}^{n,1} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=0}^{[N(n,t)/k_n]-1} \left( T(n, (i+1) k_n) - T(n, ik_n) \right) \)

\[ \times \left( \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \left( \partial_{jk,lm}^{2} \left( c_{ik_n+1}^{n} \right) \partial_{lm} \left( c_{ik_n+1}^{n} \right) \left( c_{ik_n+1}^{n} + c_{ik_n+1}^{n} \right) \right) \right) \]

\[ \nabla_{n,2}^{n,2} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=0}^{[N(n,t)/k_n]-1} \left( T(n, (i+1) k_n) \right) \left( g(c_{ik_n+1}^{n}) - g(c_{ik_n}^{n}) \right) ds - \frac{1}{\sqrt{\Delta_n}} \int_{T(n, (i+1) k_n)}^{T(n, ik_n)} g(c_{s}) ds \]
\[
\bar{V}_t^{n,3} = \frac{1}{\sqrt{2n}} \sum_{i=0}^{[N_n(t)/k_n]-1} \left( \frac{1}{k_n} \sum_{j=0}^{k_n-1} \frac{1}{c_{ik_n+1}^{n,lm}} \left( c_{ik_n+1+j}^{n,lm} - c_{ik_n+1}^{n,lm} \right) \right)
\]

\[
\bar{V}_t^{n,4} = \frac{1}{\sqrt{2n}} \sum_{i=0}^{[N_n(t)/k_n]-1} \left( T(n, (i+1) k_n) - T(n, ik_n) \right)
\]

\[
\times \left( g \left( c_{ik_n+1}^{n} + \beta_{ik_n+1}^{n} \right) - \left( g \left( c_{ik_n+1}^{n} \right) + \sum_{l,m=1}^{d} \partial_{lm} g \left( c_{ik_n+1}^{n} \right) \beta_{ik_n+1}^{n,lm} \right) \right)
\]

\[
\times \left( -\frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial_{j,k,lm}^2 g \left( c_{ik_n+1}^{n} + \beta_{ik_n+1}^{n} \right) \right)
\]

\[
\times \left( (c_{ik_n+1}^{n,ji} + \beta_{ik_n+1}^{n,ji}) (c_{ik_n+1}^{n,km} + \beta_{ik_n+1}^{n,km}) + (c_{ik_n+1}^{n,jm} + \beta_{ik_n+1}^{n,jm}) (c_{ik_n+1}^{n,kl} + \beta_{ik_n+1}^{n,kl}) \right)
\]

Again we can prove similarly that for \( v = 1,2,3,4 \), \( V_{n,v} \longrightarrow 0 \). This essentially states that the alternative estimator has the same asymptotics.
Bibliography


217


Biography

Yuan Xue was born in Chengdu, China in 1989. She attended Peking University between 2007 and 2011 majoring in finance in Guanghua School of Management. She also got a double degree in statistics from the Department of Mathematics. After that, she got a master degree in economics from Duke University in 2012. She will graduate from Duke University with PhD in economics. She will be joining Quora as data scientist shortly after graduation in 2018.