

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/117255>

Copyright and reuse:

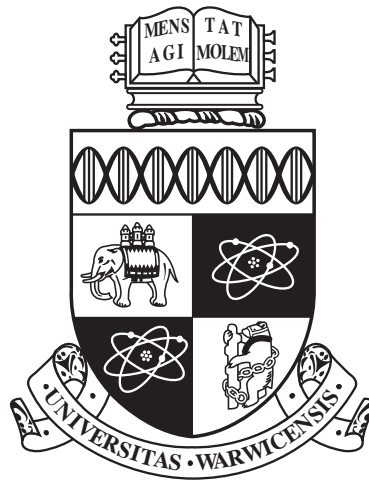
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



Three Essays on Empirical Finance

by

Chao Wang

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Business School

September 2017

THE UNIVERSITY OF
WARWICK

Contents

List of Tables	iii
List of Figures	v
Acknowledgments	vii
Declarations	viii
Abstract	ix
Chapter 1 Introduction	1
Chapter 2 Central Bank Intervention: Its Effects and the Endogeneity Problem	10
2.1 Introduction	10
2.2 Instrumental Variable	19
2.2.1 Wavelet Analysis	20
2.2.2 Wavelet Based Instrumental Variable	23
2.3 Data	26
2.4 Models and Empirical Results	27
2.4.1 Effect on the Price Level	28
2.4.2 Effect on the Volatility	30
2.4.3 Coordination Channel	32
2.4.4 Damping Channel	34
2.4.5 Duration of the Price Impact	35

2.4.6	Volatility Regimes and Price Impact	37
2.5	Conclusions	41
Chapter 3 Calibrating the Predictability of Exchange Rates: A Model-Independent Approach		54
3.1	Introduction	54
3.2	Literature Review	59
3.3	Methodology	60
3.4	Information Theory and the Predictability of Time Series	65
3.5	Data	70
3.6	Estimation and Empirical Results	72
3.6.1	Contemporaneous Transfer Entropy and Explanatory Power	72
3.6.2	Transfer Entropy, Auto-Information-Transfer and Predictive Power	76
3.7	Conclusions and Discussion	81
3.8	Appendix: Mechanism of Noisy Channel Communication	83
Chapter 4 Transient and Permanent Jumps in the High Frequency Data		91
4.1	Introduction	91
4.2	Physical Clock versus Volume Clock	97
4.3	Literature Review	99
4.4	Jump Test	102
4.5	Data	107
4.6	Jump Statistics	110
4.7	Panel Regression Results	115
4.8	conclusion	120
Chapter 5 Concluding Remarks		151

List of Tables

2.1	Effect on the price level, without the lagged variables	45
2.2	Effect on the price level, with the lagged variables	46
2.3	Effect on the volatility	47
2.4	Effect on the market order flow	48
2.5	Effect on the number of trades	49
2.6	Damping Channel	50
2.7	Price impact at different time scales	51
2.8	Price impact and volatility regimes	52
2.9	Price impact of the order flow and the volatility regimes	53
3.1	Explanatory power at the hourly frequency: Contemporaneous Transfer Entropy and R^2	85
3.2	Explanatory power at the daily frequency: Contemporaneous Transfer Entropy and R^2	86
3.3	Predictive power at the hourly frequency: Transfer Entropy and AIT	87
3.4	Predictive power at the daily frequency: Transfer Entropy and AIT	88
3.5	Predictive power at the monthly frequency: Transfer Entropy and AIT	89
4.1	Fixed effect panel regressions with market capitalization buckets	121
4.2	Fixed effect panel regressions with volume buckets	122
4.3	Fixed effect panel regressions with price buckets	123
4.4	Fixed effect panel regressions with effective spread buckets	124
4.5	Fixed effect panel regressions with quoted spread buckets	125
4.6	Fixed effect panel regressions with quote-to-trade ratio buckets	126

4.7	Fixed effect panel regressions with market depth buckets	127
4.8	Fixed effect panel regressions with bipower variation buckets	128
4.9	The effects of NYSE auto-quote, based on the regression results	129

List of Figures

2.1	Morlet wavelet	42
2.2	Haar wavelet	42
2.3	Japanese central bank intervention and its wavelet decomposition	43
2.4	Correspondence between the states of the Markov switching model and the volatility regimes	44
3.1	The equivalence between R^2 and Gaussian estimate of CTE at the hourly frequency	90
3.2	The equivalence between R^2 and Gaussian estimate of CTE at the daily frequency	90
4.1	Average numbers of jumps in a month for market capitalization quartiles	130
4.2	Average numbers of jump days in a month for market capitalization quartiles . .	131
4.3	Average numbers of jumps per jump day in a month for market capitalization quartiles	132
4.4	The percentage of jump stocks for market capitalization quartiles	133
4.5	Average numbers of jumps in a month for volume quartiles	134
4.6	Average numbers of jump days in a month for volume quartiles	135
4.7	Average numbers of jumps per jump day in a month for volume quartiles	136
4.8	The percentage of jump stocks for volume quartile	137
4.9	Average numbers of jumps in a month for price buckets	138
4.10	Average numbers of jump days in a month for price buckets	139
4.11	Average numbers of jumps per jump day in a month for price buckets	140
4.12	Average numbers of jumps in a month for quoted spread quartiles	141
4.13	Average numbers of jump days in a month for quoted spread quartiles	142
4.14	Average numbers of jumps per jump day in a month for quoted spread quartiles	143

4.15	Average numbers of jumps in a month for effective spread quartiles	144
4.16	Average numbers of jump days within a month for effective spread quartiles . . .	145
4.17	Average numbers of jumps per jump day within a month for effective spread quartiles	146
4.18	Average numbers of jumps in a month for exchange buckets	147
4.19	Average numbers of jump days in a month for exchange buckets	148
4.20	Average numbers of jumps per jump day in a month for exchange buckets	149
4.21	The percentage of jump stocks for different exchanges buckets	150

Acknowledgments

Four years time of my Ph.D. in Warwick Business School is a challenging but rewarding journey. I have learned a lot from my supervisors, faculty members of the finance group, and my colleagues. Not only the finance and economics knowledge has been acquired and internalized, more importantly, the capability of learning new knowledge, thinking and analyzing from different perspectives, and proposing new research ideas, is developed during my Ph.D. study. I would like to express my gratitude towards my supervisors, Prof. Roman Kozhan and Prof. Michael Moore, who have been guiding and supporting me along the whole Ph.D. study. This Ph.D. would not have been done without their help. Other faculty members of the WBS finance group also gave me lots of useful advice as well, especially Dr. Arie Gozluklu and Prof. Alex Stremme. I would also want to express my gratitude to my two examiners, Prof. Richard Payne and Dr. Arie Gozluklu, for their valuable inputs and advice.

Six years ago, I was a physics student who had never learned any economics or finance. As the coordinator of the research master programme, Prof. Marco Haan (University of Groningen) had confidence in me and helped me in securing the prestigious Huygens Scholarship for my two-year research master at University of Groningen. My supervisor at Groningen, prof. Paul Bekker, who is a remarkable econometrician, taught me most of my knowledge in econometrics. The good time in the Netherlands will be missed.

I would like to take the opportunity to thank my family, my parents and my uncles. They are always there for me, and care about me. They made me who I am. Last but not least, I want to say thank you to my primary school teachers Ms. Yang and Ms. Bao, for cultivating my confidence when I was young and changing the trajectory of my life. They surely would be very pleased to know that I am getting my Ph.D.

Declarations

I declare that any material contained in this thesis has not been submitted for a degree to any other university. I further declare that Chapter 4 of the this thesis is joint work with Dr. Roman Kozhan and Dr. Wing Wah Tham.

Abstract

This thesis comprises three essays on different topics of empirical finance. Particularly, the first paper (chapter 2) studies the effects of central bank intervention; The second paper (chapter 3) proposes a novel method to calibrate the predictability of exchange rates; The third paper (chapter 4) investigates the price stability in the era of high frequency trading: whether the U.S. stocks suffer more from transient jumps than they used to.

Chapter 2 studies the effects of Japanese central bank intervention from the market microstructure perspective. We measure its impact on the price and volatility level, and other market participants' behavior. The endogeneity problem is solved by adopting a novel instrumental variable: the long run part of the intervention time series. The empirical results indicate that adopting this IV does make qualitative differences. We show that the intervention successfully moves the price level of JPY/USD, and increases the volatility. The empirical evidence supports the damping channel, but not the coordination channel. The intervention has relatively persistent price impact, which lasts for roughly 16 days. Moreover, the Markov switching model shows that the intervention has greater price impact in the high volatility regime, and *vice versa*.

Chapter 3 introduces a nonparametric model-independent methodology to calibrate the predictability of exchange rates. In order to predict the exchange rates, the predictors should contain enough information about the future return, regardless of the specification of the model. The information transfer from the predictors to future return can be measured by their mutual information. According to Shannon's channel coding theorem, it is *the* measure of the statistical dependence, linear and nonlinear, and not just the linear dependence as the correlation coefficient measures. The information transfer also acts as the upper bound of the predictive power of any model based on these predictors. Empirically, we evaluate the predictability at the hourly, daily and monthly frequency, and find that exchange rates are systematically predictable intraday, but not on other frequencies. Moreover, the linear model is suboptimal and fails to capture most of the information, which explains why it is so hard for the traditional linear models to outperform the random walk benchmark.

Chapter 4 conducts a comprehensive empirical exercise to identify all the permanent and transient jumps in the 20 years Trade and Quote (TAQ) data. We aim to answer the following questions: whether the markets have become more unstable, and if HFTs are responsible for causing the increased instability. Empirically, we document the cross-sectional variations: the small market cap, low volume and market liquidity, and low-priced stocks become more unstable, and most of the increased jumps are transient, which cannot be attributed to the increase in information events. On the other hand, the large market cap, high volume and liquidity, and high-priced stocks suffer from fewer price jumps. The structural change of the jumps happens around 2003, which coincides with the auto-quote implementation. We also find further supporting evidence that more HFT participation is associated with more jumps.

Chapter 1

Introduction

This doctoral thesis comprises of three essays on three topics of empirical finance. They are in the fields of market microstructure, foreign exchange, and their intersection. The first essay (chapter 2) is on the effects of central bank intervention, from a market microstructure point of view. In particular, we first resolve the endogeneity problem by introducing an instrumental variable. Then the effects of the intervention on different variables can be evaluated accurately with the help of the new IV. The second essay (chapter 3) introduces a model-independent method to calibrate the predictability of exchange rates. The ideas in information theory are applied to time series predictability problem. Mutual information calibrates the upper bound for the predictive power of any model based on a given set of predictors. The third essay (chapter 4) conducts a comprehensive empirical exercise to identify all the transient and permanent jumps in the intraday stock prices. We document the cross-sectional variations in the price jumps for different stock buckets, especially around 2003, when auto-quote greatly reduces the trading latency. We also try to answer if HFTs are responsible for causing the increased transient jumps in the small market cap, low volume and market liquidity, and low-priced stocks. In the following paragraphs, we will discuss the main points of these three essays one by one.

The first essay (chapter 2) aims at evaluating the effects of the intervention on various quantities, such as the price and volatility level of the exchange rate, based on Japanese official foreign exchange interventions from Japan's Ministry of Finance and the Reuters interdealer order flow data. To be more specific, we focus on the following questions. Can the interventions move

the price level of the exchange rate in the desired way? Is the impact of the intervention on the exchange rate permanent or transitory? In other words, does it decay over time, and how long does it take for the price impact to dissipate completely? Another question of interest is that whether central bank intervention will calm the market or it would raise the volatility. Moreover, do the interventions coordinate the private order flow into the direction set by the central bank (coordination channel)? Would the price impact of private order flow be reduced or even eliminated by the interventions, as suggested in the theoretical argument of the damping channel? Last but not least, whether the effectiveness of the intervention depends on the market condition, for instance, does the price impact of the intervention depends on the prevailing volatility level of the exchange rate?

However, the intervention is very likely to be an endogenous variable, which means that the results could be biased or even misleading if one fails to separate the causality from the reverse causality. The importance of correctly assessing the effects of the interventions cannot be overlooked. The central bankers cannot make the right decision without the accurate feedback. For instance, considering the case that the central bank intends to support the exchange rate ‘leaning against the wind’, but if they ignore the endogeneity problem and wrongly conclude that the intervention is not effective, they would decide not to intervene in the future. But actually, the intervention may have strong price impact on the price level, once the endogeneity issue is resolved.

The research on the effects of central bank intervention is suffered from severe endogeneity problem (Fatum and Hutchison 2003, Neely 2005). It is the main difficulty of evaluating the effects of the intervention on various financial variables, such as price level, volatility, order flow and the number of trades. For instance, previous studies report that intervention has a much stronger impact when the central bank intervenes ‘with-the-market’, in contrast to the intervention ‘leaning against the wind’ (for instance, Payne and Vitale 2003). This clearly reveals the possibility that the regressor, intervention, is endogenous with respect to the dependent variable, the change in price level. With regard to the volatility, Dominguez (2003) documents that, based on USD/DEM returns, the volatility on Fed intervention days is always higher than that in the control sample. The problem is that whether it is due to the fact that intervention is usually conducted on high volatility days, or the intervention raises the volatility level on the

intervention day. Simple OLS regressions would produce biased estimates, or may even wrongly draw completely opposite conclusions.

Previous studies use several methods to alleviate the endogeneity issue. First of all, intraday data provide opportunities for more precise analyses, compared with daily data. Menkhoff (2010) points out that ‘Intraday data focus on a narrow time window which contains less noise, i.e. fewer other influences on exchange rates occurring during the day, and which tentatively overcomes the endogeneity problem of interventions’. Based on the intraday intervention data made available by the Swiss National Bank, the high-frequency studies such as Fischer and Zurlinden (1999) and Payne and Vitale (2003), which tend to be less afflicted by the endogeneity problems, have also found intervention to be effective at moving the exchange rate. The intraday studies rely on press reports, such as Reuters newswire. But there are serious mistakes in the press reports on interventions (e.g., Frenkel et al., 2004, Fischer 2006). Furthermore, the central bank would be more interested in the long term effects, instead of the intraday price impact. Resolving the endogeneity problem by going to high frequency has a disadvantage: the effectiveness at the high frequency (intraday) does not guarantee the effectiveness at the daily frequency or longer horizon. For instance, the price would move very dramatically at the very minute/hour of the intervention, but it may revert back to the original level by the end of the day.

Another possible way to evaluate the ‘real’ effect of the intervention on the exchange rate is to construct the “counterfactual”, which is the hypothetical exchange rate movement in the absence of intervention. Fatum and Hutchison (2010) introduce the method of propensity-score matching to estimate the “average treatment effect” of the intervention. They match the intervention days (treatment) and non-intervention days (control) based on the estimated probability of intervention (a propensity score) for each day in the given sample. Chamon, Garcia and Souza (2015) use a synthetic control approach to estimate the effect of Brazilian intervention: constructing a synthetic control group provides a counterfactual exchange rate. But the methodology is used for the one-time event, and not appropriate for studying the effect of frequent interventions.

One of the contributions of chapter 2 is adopting a novel instrumental variable to study central bank intervention. The instrumental variable introduced in this chapter makes sure that the

researchers are able to obtain an accurate assessment of the effects of central bank intervention. To obtain the instrumental variable, we decompose the time series of intervention into two orthogonal components using wavelet analysis: the short scale part, which has all the variations within the one-day horizon, and the long term part, which contains all the variations with the horizon longer than (including) two days. Intuitively speaking, since the long term component does not contain the contemporaneous information within the day, we expect it not to be directly correlated with the dependent variable, the contemporaneous price movement. On the other hand, the long term component is highly correlated with the endogenous explanatory variable by design: the long term component preserves most of the variations of the original time series.

Empirical, we document that the intervention has positive and significant price impact, which means that interventions can move the price level of the exchange rate in the desired way. The price impact of intervention gradually decays as the horizon becomes longer, and we show that it would last for approximately 16 trading days. With regard to the volatility, Japanese intervention would increase the volatility of JPY/USD. Furthermore, the empirical results do not support the coordination channel: the intervention does not change market order flow. But it would increase the number of trades contemporaneously, and decrease the number of trades in the following day. In other words, the intervention cannot align the market order flow to the direction of the intervention, but there is weak evidence that market participants do trade on the intervention events. However, we find empirical evidence that supports the damping channel: the private trade's price impact is reduced considerably in the presence of the intervention. Last but not least, we also extend the linear model studying price impact into a two-state Markov switching model, and show that intervention would have greater price impact in the high volatility regime. Therefore, the price impact of the intervention depends on the market condition, and if the objective of the central bank is influencing the price level, the intervention should be conducted when the volatility level is high.

Chapter 3 proposes a new methodology to assess the out-of-sample predictability of financial time series. Since Meese and Rogoff (1983a,b, 1988), it is well known that exchange rates are very difficult to predict. The random walk model generally outperforms the economic models in the out-of-sample forecasting exercises, which is called “the Meese and Rogoff puzzle”. The traditional model-dependent method cannot differentiate whether the failure is due to misspec-

ification of the forecasting model or lacking information in the predictors. If it is the former, different economic or econometric models might improve the forecasting performance. On the other hand, if the predictors contain no information about the future price movement, it is impossible to predict based on these predictors, regardless of the specification of the forecasting model.

We use an information-theoretical quantity called mutual information to calibrate the information transfer from the predictors to the future return. Mutual information measures all statistical dependence, linear and nonlinear, and not just the linear dependence as the correlation coefficient measures. Measuring the information transfer can help us to distinguish lack of information from model misspecification. Pinpointing the underlying cause of failure is very important for time series predictability problem, since the lack of information should be solved by searching for better predictors, rather than trying different economic and econometric models.

Our goal is to provide a model-independent criterion for calibrating the predictability of time series by drawing an analogy between predictability and Shannon's channel coding theorem (Shannon 1948, Shannon and Weaver 1949). The logic is straightforward: in order to be able to predict, the predictors have to carry information about the future return of the exchange rate. In the language of statistics, the predictors and the future price movement must be statistical dependent. This necessity is independent of the specific economic model or econometric methodology adopted by the researcher. On the other hand, for a given set of predictors, the predictive power of any model must be bounded by an upper limit, which is the information transfer from the predictors to the future return.

Why is mutual information the correct measure of information transfer or statistical dependence? In information theory, mutual information specifies a noisy channel's reliable information transmission capacity, beyond which error-free communication is impossible. In other words, mutual information specifies the upper bound of the information transmission rate of the noise communication channel. This is the great insight of Shannon's channel coding theorem. Essentially, a noisy communication channel is two statistical dependent random variables. Mutual information is the capacity of information transfer between these two variables. Therefore, mutual information is the right measure that can calibrate the information transfer from the predictors to the future return.

We investigate the exchange rate predictability at various horizons: hourly, daily and monthly. The exchange rates in our dataset are the most frequently traded ones in the FX markets. Empirically, we find that the exchange rates are systematically predictable at the hourly frequency. The intraday predictive power is mainly from the interdependence of the exchange rate returns at the hourly frequency, whereas order flow has very small predictive power for the future return. One important point is that the linear model that are frequently used in the literature fails to capture most of the information transfer, which explains why it is so hard for the linear model to outperform the random walk benchmark, i.e. the Meese and Rogoff puzzle. It also suggests that the optimal forecasting model based on historical returns of the exchange rate must be non-linear. At daily and monthly frequency, the exchange rates are not systematically predictable based on historical returns, order flow, and macroeconomic fundamentals. Our finding suggests that for more than half of the currency pairs, the factors have small but significant information transfer, but for all the exchange rates, the linear model does not have any significant predictive power. Furthermore, unlike the hourly frequency, the historical movements of exchange rates do not have any predictive power for the future return at the daily and monthly frequency.

In chapter 4, we are motivated to conduct a thorough empirical study on the price jumps in the high frequency data, and further answer the question whether the markets have become more unstable in the era of high frequency trading. We are also very interested in finding out whether the increased instability is due to the changes in permanent or transient jumps. While the increase in permanent jumps can be explained as more information events in recent years, the increase in extreme transient jumps can only be attributed to the deterioration of market quality.

Based on the Trade and Quote (TAQ) database, we identify all the permanent/transient jumps of the CRSP common shares in the period from January 1995 to December 2014. With the Flash Crash type of transient jump in mind, we detect transient jumps that last 2.5 minutes to one hour. Considering that jump properties of the stocks may be dependent on the stock characteristics, especially the market microstructure and liquidity related quantities, in each month, we divide the stocks with jumps into four buckets (quartiles) according to the value of a chosen characteristic, and calculate the mean values of the number of jumps (permanent and transient), the number of jump days and the percentage of jump stocks. We observe a structural

change in the jump statistics around 2003. It coincides with the introduction of auto-quote, which reduce the trading latency considerably and made high frequency trading much easier.

Meanwhile, there is cross-sectional variation in the jump properties across different buckets of stocks: in terms of the price jumps, some of the stocks become more stable, while others suffer from much more transient jumps. General speaking, the stocks with low market cap and volume, large bid-ask spread and relative tick size experience more jump after 2003, while they are the relatively more stable ones before 2003. These are the thinly-traded penny stocks become more unstable. Notably, most of the increase in jumps is the changes in transient jumps, which are not information-driven. It means that the worsening price stability is not due to more frequent market-relevant news after 2003. On contrary, the considerable increase in transient jumps actually reveals the deterioration of market quality for these stocks in recent years. On the other hand, the stocks on the other end of the characteristic spectrum (high market cap and volume, small bid-ask spread and relative tick size) actually become more stable after 2003. These are the large firms and blue-chip stocks, which are traded very actively by market participants. These stocks become more stable in the era of high frequency trading. This observation is in line with the empirical literature that high frequency trading improves the market liquidity (Hendershott, Jones, and Menkveld 2011). In conclusion, when trying to answer the question whether the stock price has become unstable in the era of HFT, one needs to be specific on the characteristics of the stock.

We are also interested in finding out whether the high frequency traders should be held accountable for the increase or decrease in the jumps. HFTs could play a role in these transient price jumps for several reasons (Biais and Foucault 2014). First, because of the similar trading strategies, HFTs may all react at the same time to erroneous signals by sending buy or sell market orders consuming market liquidity, triggering sharp price movements. Alternatively, after the arrival of a large sell (buy) market order, the limit orders may all got canceled by HFTs for safety reasons, and if the large market order does not appear to be informationally motivated, HFTs would resubmit new limit orders very quickly. Moreover, as the endogenous liquidity providers, HFTs may withdraw their quotes and stop marking the market in the adverse market conditions. In any case, waves of cancellations or market orders submissions by HFTs reacting to the same event may exacerbate the transient price jumps.

To study this issue, we run panel regressions around the event of auto-quote introduction on the NYSE around early 2003. By fitting a linear model with dummy variables of auto-quote and quartiles as well as their interaction terms to the data of NYSE shares, we can estimate the number of jumps (permanent/transient) and jump days for each quartile before and after the auto-quote introduction. The panel data regressions using characteristic quartile dummy variables confirm the results obtained previously: the structural shift and heterogeneous changes in the jump properties across different characteristic buckets of stocks.

More importantly, when the stocks are grouped into quartiles according to the quote-to-trade ratio, the panel data regression results indicate that the stocks in the quartile with the highest quote-to-trade ratio become much more unstable after auto-quote, and most of the incremental jumps are transient. As we all know that one of the distinguishing features of HFTs is the high quote-to-trade ratio, because HFTs would submit and quickly cancel their quotes at milliseconds or higher frequency, the actual trades that go through are merely a very small fraction of the total number of quote posted by the HFTs. Similarly, as the message-to-trade ratio used in Hasbrouck and Saar (2013), the quote-to-trade ratio can be used as a proxy for HFT. The stock with high quote-to-trade ratio has a higher percentage of HFT participation. Therefore, the dramatic increase in transient jump for the high quote-to-trade ratio stocks after auto-quote suggests that HFT does cause instability in the stock price.

Another supporting evidence is that we also empirically document that the quartile with the greatest relative tick size become much more unstable: the number of jumps, especially the transient ones, increase considerably after auto-quote. Considering that the large relative tick size is associated with a high percentage of HFT participation (Yao, Ye 2014), this empirical evidence also implies that HFTs cause more transient jumps in these low-priced stocks.

Overall speaking, the empirical evidence points at the fact that HFTs increase jumps in the small market cap, low volume and market liquidity, and large relative tick size ('penny') stocks, while for the large market cap stocks with high volume and market liquidity, their price stability is actually improved considerably. Due to the lack of high frequency trading data, our indirect evidence cannot completely pin down this claim. Nonetheless, the empirical fact of the deterioration in market quality and price stability for these small illiquid low-priced stocks are genuine, regardless whether it is caused by HFT or not.

Finally, Chapter 5 concludes with a discussion of the findings and contributions of this thesis, and discusses future potential research questions.

Chapter 2

Central Bank Intervention: Its Effects and the Endogeneity Problem

2.1 Introduction

This paper introduces a novel instrumental variable to study some of the most important research questions of central bank intervention, from the market microstructure point of view. Based on Japanese official foreign exchange interventions from Japan's Ministry of Finance and the Reuters interdealer order flow data, we want to find out the effects of the interventions on various quantities, such as the price and volatility level of the exchange rate. To be more specific, this paper focuses on the following questions. Can the interventions move the price level of the exchange rate in the desired way? Is the impact of the intervention on the exchange rate permanent or transitory? In other words, does it decay over time, and how long does it take for the price impact to dissipate completely? Another question of interest is that whether central bank intervention will calm the market or it would raise the volatility. Moreover, do the interventions coordinate the private order flow into the direction set by the central bank (coordination channel)? Would the price impact of private order flow be reduced or even eliminated by the interventions, as suggested in the theoretical argument of the damping channel? Last

but not least, whether the effectiveness of the intervention depends on the market condition, for instance, does the price impact of the intervention depends on the prevailing volatility level of the exchange rate?

This paper underlines the importance of using the instrumental variable when evaluating central bank interventions. First of all, the intervention is very likely to be an endogenous variable, which means that the results could be biased or even misleading if one fails to separate the causality from the reverse causality. Therefore, the importance of correctly assessing the effects of the interventions cannot be overlooked. The central bankers decide to intervene with certain intention: for instance, supporting or suppressing the value of the currency, or reducing the volatility and calming the market. Regardless the underlying intention of intervention, the key is evaluating the effects of the intervention accurately, without which the central bankers cannot get the right feedback and adjust their action accordingly. For instance, considering the case that the central bank intends to support the exchange rate ‘leaning against the wind’, but if they ignore the endogeneity problem and wrongly conclude that the intervention is not effective, they may decide not to intervene in the future. But actually, the intervention may have strong price impact on the price level, once the endogeneity issue is taken care of.

We are going to evaluate the effects of the Japanese central bank interventions on JPY/USD. JPY/USD is one of the most heavily traded exchange rates, and Japan is one of the advanced economies that still conduct interventions in recent years. More importantly, the data on Japanese central bank intervention is publicly available. Another desirable fact is that Japanese foreign exchange intervention was automatically sterilized (Ito 2002, 2007). The intervention is formally considered sterilized as the dollar purchases are financed by the sale of yen assets issued by the Ministry of Finance. From the researcher’s point of view, sterilized intervention is more interesting, because any changes in monetary base due to foreign exchange intervention are absorbed by open market operations in the opposite direction so that monetary basis remains unchanged (Ito 2007). The sterilized foreign exchange interventions tend to be less effective at moving exchange rates compared with unsterilized interventions: the unsterilized intervention results in the expansion of domestic money base so that the interest rate declines, which encourages capital outflows and cause the home currency to depreciate.

Central bank intervention is defined as “purchases and sales of foreign currencies by the mone-

tary authorities with an intention to influence the foreign exchange rate” (Ito 2007). There are several points need to be clarified here. First one is that the monetary authorities often buy and sell foreign exchanges for its ordinary activities, which are inevitable, but these transactions without the intentions of influencing the exchange rate should not be considered as intervention. Second, as Ito (2007) points out, intervention is fiscal operation rather than monetary operation, and the balance is shown in the budget rather than the balance sheet of the Central Bank. There are other important issues that influence the exchange rate, such as monetary policy, quantitative easing (QE), and currency swaps line among central banks, but these issues are beyond our scope of the central bank intervention studied in this paper.

As for the intention of Japanese central bank intervention, Nanto (2007) states that “*In Japans case, the Bank of Japan (in consultation with the Ministry of Finance) has bought U.S. Treasury securities and other liquid dollar assets at times when the value of the dollar relative to the yen was declining. The intended result was to keep the value of the yen from appreciating too quickly in order to keep the price of Japanese exports from rising in markets such as those in the United States and to maintain the profitability of those exports.*” In other extreme scenarios, such as the unfortunate earthquake and tsunami happened in 2011, G7 central banks coordinately intervened to stabilize the yen, tamping its value down after Japans devastating earthquake triggered a yen surge and raised fears about the global economy.

The Japanese central bank interventions were carried out in the spot market, and they are automatically sterilized. No derivative, including forward, was used. The interventions are carried out directly by the Bank of Japan or carried out by other Central Banks on behalf of the Bank of Japan (Ito 2007). However, the disclosed intervention data do not specify the exact time of day and which market intervention was conducted.

Since the end of last century, many advanced economies no longer intervene the exchange rate, such as the United States, the Euro area, and the United Kingdom. They have moved away from actual interventions to communication or oral intervention, which would guide and influence foreign exchange markets (Fratzscher, 2008). Although central bank intervention has lost its role among certain industrialized countries, it is still a very important research topic. First of all, most of the currencies in the world are not fully flexible, especially those of the emerging markets, such as Chinese Renminbi. Moreover, some developed economies

still intervene occasionally, for instance, Reserve Bank of Australia, which has not intervened since September 2001, conducted two brief phases of intervention in August 2007 and October - November 2008, the global financial crisis, in order to provide market liquidity and address market dysfunction (Newman, Potter, and Wright, 2011). Therefore, a better and more precise understanding of the effect of central bank intervention is still of great use.

The early literature on foreign exchange is surveyed by Sarno and Taylor (2001), which typically focuses on advanced economies and generally concludes that sterilized intervention is not very effective at influencing the price level of the exchange rates. More recently, many papers using daily intervention and exchange rate data to investigate the effect of central bank intervention: Fatum and Hutchison (2003, 2006), Edison, Cashin, and Liang (2003), Aguilar and Nydalh (2000), Kim, Kortian, and Sheen (2000), Ito (2002), and Chaboud and Humpage (2005). These studies obtain mixed evidence for the hypothesis that intervention influences exchange rates in the desired direction, but coordinated interventions are found to be much more successful (e.g., Ito 2002). On the other hand, papers based on intraday data reach the consensus that interventions successfully move the price, at least in the very short term, within one-day horizon (Fischer and Zurlinden 1999, Payne and Vitale 2003, Pasquariello 2007, Fatum and King 2005, Menkhoff 2010).

Another strand of literature on central bank intervention focuses on the effect on exchange rate volatility. Theoretically, the forward-looking rational expectations exchange rate model implies that a credible central bank intervention should either dampen exchange rate volatility or should not affect volatility at all. If interventions are not credible, or the signals sent by monetary authorities are ambiguous, interventions should amplify exchange rate volatility (Dominguez 1998). Empirically, the results are mixed. There are studies find that intervention operations increase exchange rate volatility, for instance, Frankel, Pierdzioch and Stadtmann (2005), and, Beine et al. (2007), whereas Dominguez (1998) examines the effects of US, German and Japanese intervention policies and finds that interventions generally increase exchange rate volatility for the 1977-1994 period, except in the mid-1980s, interventions appear to have reduced exchange rate volatility. Some studies based on intraday data find that interventions increase volatility intraday, but at the daily frequency, they may reduce volatility overall (Menkhoff 2010). Kim (2007) studies the Japanese interventions from 1991 to 2004, and finds that interventions reduce

volatility significantly overnight, but not contemporaneously. Dominguez (2006) reports that interventions increase volatility in the short run (intraday and daily frequency), but no evidence indicates that interventions influence longer-term volatility.

The research on the effects of central bank intervention is suffered from severe endogeneity problem (Fatum and Hutchison 2003, Neely 2005). It is the main difficulty of evaluating the effects of the intervention on various financial variables, such as price level, volatility, order flow and the number of trades. For instance, previous studies report that intervention has a much stronger impact when the central bank intervenes ‘with-the-market’, in contrast to the intervention ‘leaning against the wind’ (for instance, Payne and Vitale 2003). This clearly reveals the possibility that the regressor, intervention, is endogenous with respect to the dependent variable, the change in price level. With regard to the volatility, Dominguez (2003) documents that, based on USD/DEM returns, the volatility on Fed intervention days is always higher than in the control sample. The problem is that whether it is due to the fact that intervention is usually conducted on high volatility days, or the intervention raises the volatility level on the intervention day. Simple OLS regressions would produce biased estimates, or may even wrongly draw completely opposite conclusions because of the reverse causality. Ideally, in order to gauge the effects correctly, one has to compare the case when the central bank intervenes with the ‘counterfactual’: what it would have been without the interventions. To separate the causality from central bank interventions to other economic variables, we use the instrumental variables, which is a widely used econometric method to address the endogeneity.

Previous studies use several methods to alleviate the endogeneity issue. First of all, intraday data provide opportunities for more precise analyses, compared with daily data. Menkhoff (2010) points out that ‘Intraday data focus on a narrow time window which contains less noise, i.e. fewer other influences on exchange rates occurring during the day, and which tentatively overcomes the endogeneity problem of interventions’. Based on the intraday intervention data made available by the Swiss National Bank, the high-frequency studies such as Fischer and Zurlinden (1999) and Payne and Vitale (2003), which tend to be less afflicted by the endogeneity problems, have also found intervention to be effective at moving the exchange rate. The intraday studies rely on press reports, such as Reuters newswire. But there are serious mistakes in the press reports on interventions (e.g., Frenkel et al., 2004, Fischer 2006). Furthermore, the

central bank would be more interested in the long term effects, instead of the intraday price impact. Resolving the endogeneity problem by going to high frequency has a disadvantage: the effectiveness at the high frequency (intraday) does not guarantee the effectiveness at the daily or lower frequency. For instance, the price would move very dramatically at the very minute/hour of the intervention, but it may revert back to the original level by the end of the day.

Another possible way to evaluate the ‘real’ effect of the intervention on the exchange rate is to construct the “counterfactual”, which is the hypothetical exchange rate movement in the absence of intervention. Fatum and Hutchison (2010) introduce the method of propensity-score matching to estimate the “average treatment effect” of the intervention. They match the intervention days (treatment) and non-intervention days (control) based on the estimated probability of intervention (a propensity score) for each day in the given sample. Chamon, Garcia and Souza (2015) use a synthetic control approach to estimate the effect of Brazilian intervention: constructing a synthetic control group provides a counterfactual exchange rate. But the methodology is used for the one-time event, and not appropriate for studying the effect of frequent interventions.

One of the contributions of this paper is adopting a novel instrumental variable to study central bank intervention. The instrumental variable introduced in this study makes sure that the researchers are able to obtain an accurate assessment of the effects of central bank intervention. To obtain the instrumental variable, we decompose the time series of intervention into two orthogonal components using wavelet analysis: the short scale component, which has all the variations within the one-day horizon, and the long term component, which contains all the variations with the horizon longer than two days. We are going to argue that since the long term component of the intervention does not contain the contemporaneous information within the day, we expect it not to be directly caused by the dependent variable, the contemporaneous price movement.

To evaluate the effect of the intervention on various variables, we regress the dependent variable on the intervention and other control variables. The dependent variable, such as the return, can also be treated as the sum of two orthogonal components as well. The short term component of the dependent variable that contains variations of one-day scale is statistically uncorrelated with the instrumental variable by design (the wavelet decomposition at different scales are orthogonal

to each other since the wavelet basis functions at different scales are orthogonal to each other). Meanwhile, the long term component of the return contains variations longer than two days, in other words, the long term component contains the long term smoothed movement or trend from current to the future, not the variation within current day. The intervention decision of the central bank generally is based on past and current price movements. It is not possible that the intervention decision is based on the long term movement from now to the future, which is essentially the future information that is not available to the central banker at real time. To sum up, the long term component of the return does not directly cause the IV, and remembering the fact that the short term component of the return is uncorrelated with the IV by design, as discussed earlier, therefore, The instrumental variable, the long term component of the intervention time series, is expected to be a valid instrumental.

Another concern is that the central bank may forecast the future exchange rate, and made the decision of the intervention directly based on the future exchange rate, hence the long term component of the dependent variable would directly cause the intervention. This concern is raised quite naturally. But there are two things to point out, first of all is that the future exchange rate is very hard to forecast, as we will discuss in the next chapter of this dissertation. The likelihood that central bank can forecast the exchange rate very accurately and made intervention decision directly based on that is very small, given the fact that we are considering the exchange rate in the floating regime. Second of all, even if the central bank can forecast the exchange rate to a certain degree, it does not necessarily mean that the long term component of the exchange rate return will directly cause the long term component of the intervention, since the current value of the intervention long term component is the smoothed/averaged of current and future intervention. Moreover, whether the instrumental variable is indeed exogenous is always subject to econometric test. Applying the instrumental variable merely based on economic or logic reasoning is a useful but not reliable way to conduct empirical research. In the next subsection, we will carry out the overidentification test to confirm that the instrumental variable is valid with respect to various dependent variables.

The second validity check for the instrumental variable is that it should contain information of the original endogenous explanatory variable. It suggests that finding an exogenous but uncorrelated variable would not help. In our case, the long term component is highly correlated

with the endogenous explanatory variable by design, since the long term component is part of the intervention time series, it preserves most of the variations of the original time series.

Empirically, for each regression, we verify that the instrumental variables are truly exogenous using the overidentification test, then apply the IV and two-stage least square (2SLS) to estimate the model. Comparing the results based on 2SLS, we can find out whether the conclusions based on OLS are qualitatively different. Furthermore, one can also test the exogeneity of the original regressors using the Hausman test, which essentially compares the OLS and 2SLS estimates. For instance, if the Hausman tests fail to reject the null hypothesis that intervention is exogenous, then one can treat the intervention as an exogenous variable with respect to that specific dependent variable.

A novel method is used in this paper, namely the wavelet analysis. Wavelet analysis is originated in the applied math and signal processing studies, but lately have been applied to economics and financial problems. Just to name a few, Vacha and Barunik (2012) study the co-movement of energy commodities using wavelet coherence analysis; In and Kim (2007) examine how well the Fama-French factor model works on different time scales; In and Kim (2006c) study the relationship between stock and futures markets with wavelet cross-correlation; Gencay et al. (2005) use wavelet method to estimate the systematic risk of an asset; Gencay et al. (2004) show that the leverage effect is weak at high frequencies but becomes prominent at lower frequencies.

Wavelet analysis decomposes the time series into orthogonal components at different scales or horizons, while preserving the information/variations localized in time. The motivation for using this novel method is mainly twofold. First of all, this method produces the instrumental variable for evaluating the effects of the intervention. As discussed earlier, the decomposed long-term component is adopted as the instrumental variable. Second of all, decomposing time series into components at different scales would allow us to study whether the economic relationship varies at different time scale. To be more specific, we want to evaluate the persistence of the price impact. The orthogonality of the variable components at different time scales would guarantee that, at each time scale, we can regress the component of return on the corresponding component of the intervention, and estimate the scale-specific coefficient of the intervention, without worrying about the potential bias when different scales are coupled together. In such a way, we can examine the decay of price impact by carrying out the regressions scale by scale.

Wavelet analysis can be used to decompose the time series into the long-term component and short-term component. In this aspect, it is similar in nature to other signal processing/econometric methods. Broadly speaking, when wavelet method is used to decompose time series into long/short term components, it can be view as a type of high/low band-pass filter. There are other methods, such as band-pass filter based on Fourier analysis, and Hodrick-Prescott filter, which aims to fit a smooth line as close as possible to the original time series while penalizing the curvature of the fitted curve. However, these do not preserve the temporal information localized in time. Moreover, Hodrick-Prescott filter uses an ad-hoc parameter for the regularization term, and tuning the parameter can be difficult and unreliable. More importantly, to study the persistence of price impact, we will use wavelet analysis to do multi-resolution analysis, i.e., decomposing the time series into orthogonal components at different time horizons. In my opinion, wavelet analysis can handle these research problems quite efficiently.

We document that one unit of intervention has a positive and significant price impact, which means that interventions can move the price level of the exchange rate in the desired way; but for the first subsample, the estimate is not strongly significant. The price impact of intervention gradually decays as the horizon becomes longer, and we show that it would last for approximately 16 trading days. With regard to the volatility, Japanese intervention would increase the volatility of JPY/USD. Furthermore, the empirical results do not support the coordination channel: the intervention does not change market order flow. But it would increase the number of trades contemporaneously, and decrease the number of trades in the following day. In other words, the intervention cannot align the market order flow to the direction of the intervention, but market participants do trade on the intervention events. However, we find empirical evidence that supports the damping channel: the private trade's price impact is reduced considerably in the presence of the intervention. Last but not least, this paper also extends the linear model studying price impact to a two-state Markov switching model, and show that intervention would have a greater price impact in the high volatility regime. Therefore, the price impact of the intervention depends on the market condition, and if the objective of the central bank is influencing the price level, the intervention should be conducted when the volatility level is high.

2.2 Instrumental Variable

To assess the effects of central bank intervention, various quantities of interest are treated as the dependent variable, such as the return, volatility, market order flow, the number of trades, while the explanatory variables are the intervention and other relevant variables, depending on the model specification. As mentioned before, in order to correctly estimate the effects of the central bank intervention, potential endogeneity problem in these regressions needs to be resolved. The results of simple OLS regression could be biased or even completely misleading due to the fact that intervention could be endogenous. Furthermore, without the valid instrumental variable, there is no way to find out whether the potential endogenous regressors are truly endogenous. Therefore, the instrumental variable is of the essence for evaluating the effects of the intervention.

Let's consider the case where central bank aims at influencing the price level of the exchange rate. On one hand, the interventions may arise from the fact that exchange rate deviates from the desired price level, in other words, undesirable movements of the exchange rate cause the central bank to intervene; on the other hand, interventions cause the exchange rate price level to move as well. Researchers are aiming at assessing the causality from the interventions to the price level movements, but the reverse causality could be much stronger. For instance, if central bank intervenes leaning against the wind to support the price of its currency, the intervention may not be able to reverse the trend, but had there not been an intervention, the exchange rate would have depreciated much more. In this case, the OLS estimate of the intervention's coefficient would be negative, which is qualitatively different from the positive estimate using the 2SLS regression.

The same logic applies to other variables, such as the volatility of the exchange rate. If the central bank aims at calming down the market by intervening, the interventions would occur when volatility is unusually high. In this case, the OLS regression will wrongly draw the conclusion that central bank intervention increases the volatility. However, it might be the case that the interventions successfully bring the volatility down.

2.2.1 Wavelet Analysis

As mentioned before, wavelets analysis is used to obtain the orthogonal decomposition of the relevant economic variables at different time scales. Due to limited space, we will discuss the wavelet analysis very briefly here. We refer the interested readers to Ramsey and Lampart (1998), and Ramsey (2002) for more details.

Wavelet analysis can be seen as an extension of the Fourier analysis. In Fourier analysis, the sine and cosine functions at various frequencies are used to expand the given function or time series. Suppose the original function or time series is $X(t)$, where $t \in [0, T]$. Based on Fourier analysis, $X(t)$ can be decomposed into components at different frequencies:

$$X(t) = a_0 + \sum_{\omega=1}^{\infty} \{a_{\omega} \cos(2\pi\omega t/T) + b_{\omega} \sin(2\pi\omega t/T)\}, \quad t \in [0, T] \quad (2.1)$$

The orthogonal basis functions $\cos(2\pi\omega t/T)$ and $\sin(2\pi\omega t/T)$ capture the variations at different frequency $\omega = 1, 2, 3, 4, \dots$, and a_{ω} and b_{ω} are the amplitudes of corresponding frequencies. Intuitively, the component at frequency $\omega = 1$, $a_1 \cos(2\pi t/T) + b_1 \sin(2\pi t/T)$, represents the long term component or trend in the original time series that fluctuates slowly. While the high frequency components, for instance the component at the frequency $\omega = 100$, $a_{100} \cos(200\pi t/T) + b_{100} \sin(200\pi t/T)$, would captures the high frequency variations that oscillates 100 time faster. The superposition of components at different frequencies recovers the original time series $X(t)$.

However, the drawback of Fourier analysis is that it assumes that the frequency content of the time series is invariant across time. After transforming the original time series from time domain into frequency domain, only the frequency information is preserved, all the temporal information is lost. Moreover, the sine and cosine basis functions do not die out, hence they are not adaptive to changes that localized in time. For example, if the time series have high frequency oscillation in the first ten percent of the time series, and continue with low frequency fluctuations for the rest, Fourier analysis would correctly identify that there are two frequency components for the whole time series, but it cannot differentiate the change in frequency content localized in time, i.e., the temporal information is completely lost. However, the temporal order of the time series is of the essence for many economic/financial studies, if we want to study the

relationship and causality among different variables based on their temporal information.

Compared with Fourier analysis, wavelet analysis are more suitable for handling financial/economic time series. The examples of the Morlet and Haar wavelets are shown in figures 2.1 and 2.2. As oppose to sine and cosine functions in Fourier analysis, the basis function of the wavelet analysis are localized in both time and frequency, hence the temporal information would be preserved as we study the time series variations at different scale/horizon. The decomposed components based on wavelet analysis at different scale can adaptively capture the local behavior of the time series in different time periods.

Insert figure 2.1 about here

Insert figure 2.2 about here

To study the variations at different scale/horizon, the scaling or dilation property of wavelets function is particularly important. Given $\psi(t)$ is the wavelet function at scale 1 and centered at 0. The Haar wavelet used in this study is in the following form:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

which is shown graphically in figure 2.2. The basis function at scale s and centered at time u is defined as

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right), \quad (2.3)$$

Intuitively, the wavelet functions at large scales (lower frequency) are dilated or stretched in the time dimension by s times, and in order to capture the variation around time u , the whole wavelet function is shifted to the new location of time u . Therefore, the wavelet $\psi_{u,s}(t)$ is concentrated in a neighborhood of size proportional to s and centered around time u . The factor $\frac{1}{\sqrt{s}}$ ensure that the energy of wavelet is normalized to one, which means the integration of the squared wavelet function $\psi_{u,s}^2(t)$ equals to one. Function $\psi_{u,s}(t)$ is also referred as the mother wavelet function, as opposed to the father wavelet or scaling function $\psi_{u,s}(t)$ that we will discuss next.

Notice that the integration of the wavelet function $\psi_{u,s}(t)$ (eq. 2.3) with respect to t equals zero, which suggest that at scale s , wavelet function $\psi_{u,s}(t)$ captures the fluctuations around a local average level, but it cannot represent the local average level of the original time series. Therefore, we need the scaling functions to span the average level of the time series. The scaling function of Haar wavelet is as follows

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

The scaling function at scale s and centered at time u is defined similarly as equation 2.3:

$$\phi_{u,s}(t) = \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right), \quad (2.5)$$

Therefore, it is easy to see that the scale function $\phi_{u,s}(t)$ captures the local average of the original function at scale s and location u .

When wavelet functions are used for multi-resolution analysis, time series X_t is decomposed into orthogonal components by scale:

$$X_t = \sum_{j=1}^5 X_t[d_j] + X_t[a_5] = X_t[d_1] + X_t[d_2] + X_t[d_3] + X_t[d_4] + X_t[d_5] + X_t[a_5], \quad (2.6)$$

where X_t could be one of the variables, such as Δs_t , mo_t , and I_t , and $X_t[d_j]$ is the component of X_t at the j th scale. Intuitively speaking, the component $X_t[d_i]$ captures the variations at 2^{j-1} days horizon, and none of the other scales:

$$X_t[d_i] = \sum_u c_u \psi_{u,d_i}(t), \quad (2.7)$$

where $\psi_{u,d_i}(t)$ is the wavelet function at scale d_i and location u , and c_u is the amplitude of $\psi_{u,d_i}(t)$ at location u .

The multi-resolution analysis (equation 2.6) can be seen as the process of peeling the onion. First of all, we extract the highest frequency variations at the 1-day horizon, which comprise the component $X_t[d_1]$; after $X_t[d_1]$ has been taken from the original time series, $X_t[d_2]$ captures the

variations between 1 day and 2 day horizon; similarly, $X_t[d_3]$ captures the variations between 2-day and 4-day horizon, and so on. At last, all the details from scale 1 to 5 have been extracted from the time series, what is left is the long-term variations that are not captured by $X_t[d_i]$, $i = 1, 2, 3, 4, 5$, and they all encapsulated in component $X_t[a_5]$. $X_t[a_i]$ represents the long term variations at the horizon of 2^5 days or greater:

$$X_t[a_i] = \sum_u f_u \phi_{u,a_i}(t), \quad (2.8)$$

where f_u is the coefficient of the scaling function $\phi_{u,a_i}(t)$ at time u .

Moreover, the components at different scales are completely uncorrelated from each other by design, since the wavelet basis functions at any two different scales are orthogonal to each other:

$$\int_{\mathbf{R}} \psi_{u_1,s_1}(t) \psi_{u_2,s_2}(t) dt = \delta_{u_1,u_2} \delta_{s_1,s_2}, \quad (2.9)$$

where function δ_{u_1,u_2} represents the Kronecker delta, which equals to 1 if $u_1 = u_2$, and zero otherwise. Equation 2.9 suggests that the basis functions at different scales or different locations are orthogonal to each other. Moreover, we also have the detail function $\psi_{u,s}(t)$ (such as equation 2.2) and scaling function $\phi_{u,s}(t)$ (such as equation 2.4) are orthogonal to each other at all scales. This is why we use wavelet analysis to decompose the relationship between the exchange rate movement and the intervention by time scale, without worrying about the relationship at different scales being coupled together.

2.2.2 Wavelet Based Instrumental Variable

Due to the potential severe endogeneity issue, one needs a valid instrumental variable (IV) for evaluating the effects of the interventions correctly. To construct the IV, we decompose the time series of intervention into two orthogonal components using wavelet analysis (Haar wavelet). The original time series of intervention denoted as I_t , is decomposed into two orthogonal components: $I_t = I_t[d_1] + I_t[a_1]$. The first part $I_t[d_1]$ contains all of the short-term variations or small-scale ‘details’ within one day horizon; the second component $I_t[a_1]$ is the long-term part of intervention, which is an approximation of the original time series: it contains all the variations

at two-day horizon or longer, which are not captured by $I_t[d_1]$. The long-term component $I_t[a_1]$ can be viewed as the result when we apply the low pass filter on the original time series of intervention: the one-day variations are averaged out using current and next day's variation. The component $I_t[a_1]$ is used as an instrumental variable to resolve the possible endogeneity issue. The time series of the original central bank intervention and its long-term and short-term components are shown in figure 2.3.

Insert figure 2.3

To estimate the effect of the intervention, the variable of interest X_t is regressed on the intervention:

$$X_t = \alpha + \beta I_t + \epsilon_t, \quad (2.10)$$

The intervention I_t is decomposed into two parts using Haar wavelet:

$$I_t = I_t[a_1] + I_t[d_1], \quad (2.11)$$

where $I_t[d_1]$ contains the “details” at one-day horizon, and $I_t[a_1]$ captures the variation at two-day horizon or longer. We are going to use $I_t[a_1]$ as the instrumental variable. On the other hand, dependent variable X_t can be decomposed into two orthogonal components as well:

$$X_t = X_t[a_1] + X_t[d_1]. \quad (2.12)$$

Now the regression is written in the following form

$$X_t[a_1] + X_t[d_1] = \alpha + \beta I_t + \epsilon_t. \quad (2.13)$$

We apply the instrumental variable $I_t[a_1]$ on both sides of the equation. $X_t[d_1]$ is orthogonal to $I_t[a_1]$ by design. Now we need to argue that $X_t[a_1]$ does not directly cause $I_t[a_1]$. Let's consider the case that X_t is the return time series. Then $X_t[a_1]$, as the long term component of the return, contains the return from now to more than two days in the future, meanwhile, $X_t[a_1]$ does not have the variations within current day. The intervention decision of the central bank generally is based on past and current price movements. It is not possible that the intervention decision is based on the long term movement from now to the future, which is essentially

the future information that is not available to the central banker in real time. Hence, the long term component of the return $X_t[a_1]$ does not directly cause the IV, as well as $I_t[a_1]$. Therefore, instrumental variable $I_t[a_1]$ should be exogenous with respect to the dependent variable. Moreover, instrumental variable $I_t[a_1]$ still preserves the long term variations of the original time series I_t , so it is expected to be highly correlated with the original explanatory variable, the intervention I_t .

However, eventually, the validity of instrumental variable is subjected to the empirical test. The overidentification tests the joint null hypothesis that the excluded instruments are valid instruments, i.e., uncorrelated with the error term and correctly excluded from the estimated equation. A rejection of the test suggests the invalidity of the instruments. The p value of the overidentification test reported in each regression (tables 2.1 to 2.5) ensures the validity of the IV. Moreover, because the long term component contains most of the information of the original time series, the strength of the IV is very good. Meanwhile, the R^2 of the 2SLS regression is also an indicator of the strength of the instrumental variables: If the R^2 of the 2SLS regression is zero, the instrumental variables have very little information about the original regressors, in which case all coefficients would be not significantly different from zero, due to the weakness or irrelevance of the instrumental variables. In other words, one can easily find some truly exogenous variables, but they still are not valid instrumental variables because they do not have explanatory power for the endogenous regressors.

We are going to observe that the two-stage least square (2SLS) regressions do produce some qualitative different results, in comparison to those of the OLS regressions. For instance, using 2SLS, we conclude that intervention would not change the market order flow, but OLS results indicate otherwise. Estimates from OLS regressions wrongly suggest that intervention would increase the market order flow at the event day, and decrease it the next day. This study underlines the necessity and importance of the instrumental variable. The central bankers can correctly assess the effects of the intervention, and check whether the realized effects fit their original intention for the interventions. In contrast, without the help of the instrumental variable, the feedback obtained by the central bank could be biased or even very misleading, which leads to suboptimal decision making.

2.3 Data

Our sample includes data from the Reuters trading system Dealing 3000. The sample periods are January 1996 to December 2002, and November 2003 to February 2013. The original data consists of a continuous record of transactions and orders. For the purpose of our study, the data is aggregated into daily frequency. Market order flow mo is measured as the difference between the trades initiated by buyers and the trades initiated by sellers from period $t - 1$ to t . The unit of the market order flow is thousand of trades. Since the minimum size of the trade is 1 million USD, approximately the unit of mo is 1 billion USD. The number of trades N_t measure the total number of the trades from period $t - 1$ to t : it is the summation of the trade quantities initiated by buyers and the trade quantities initiated by sellers. Since we don't know the dollar amount of the trade, the number of trades N_t is a proxy measure for the trading volume.

The concern with the Reuters data is that since EBS (Electronic Broking Services) is the primary trading venue for JPY/USD, while Reuters dealing system is only a minor trading venue, the order flow data from Reuters has weak explanatory power. Regressing the concurrent return on the order flow produces R^2 equals to 10%, approximately. This is a disadvantage of our order flow data. However, it is still beneficial to include the interdealer order flow as a control variable, when we regress the variable of interest on the order flow of central bank intervention.

With respect to the price level, the log of the closing price of the exchange rate at 21:00 GMT on day t is denoted by s_t . The exchange rate return or difference of the logged price, Δs_t , is calculated as the difference between the log midpoint exchange rate at date t and $t - 1$. The unit of Δs_t is basis points. The volatility of the exchange rate σ_t is the realized volatility calculated based on the intraday exchange rate movements from 21:00 GMT on day $t - 1$ to the same time on day t .

The Japanese official exchange rate intervention data is publicly available on the website of Ministry of Finance. The data contains the exact date of the intervention, the currency pairs, the amount of the intervention, as well as its direction (for example, USD is bought and JPY is sold) from January 1996 to February 2013. The intervention I_t is the time series contains the signed amount of the intervention. The unit of the intervention is 100 billion JPY, which is equivalent to 1 billion USD, roughly speaking. The sign of the intervention is defined as positive

if USD is sold and JPY is bought. Moreover, if there is no intervention at day t , the value of I_t is simply set as zero.

2.4 Models and Empirical Results

In this section, we empirically study the impacts of Japanese central bank interventions on various variables, such as the price and volatility levels of the exchange rate, the order flow and the total number of trades. Naturally, the central bank intervenes the exchange rate with certain goals in mind. To determine whether the intervention is successful or not, ideally one needs to know what is the purpose behind the interventions. For instance, simply determining the success of the interventions based on the price impact would not be appropriate if the central bank wants to calm the market and smooth the adjustments of the exchange rate. However, there is no way for us to find out the true incentive of the central bank. The next best thing is to carry out an exhaustive empirical study on the variables of interest.

For simplicity, the linear model is adopted for most of the econometric exercises in this paper. The variable of interest would be put on the left hand side as the dependent variable, while the intervention, other control variables, and the lagged terms are included as the independent variables on the right hand side. The linear model will be estimated using the OLS and 2SLS methods. In addition, there are two extra research questions need to be answered with more complicated models. First of all, we would like to find out whether the price impact of the intervention is transient or permanent. If the central bank aims at influencing the price level, but the price impact of the intervention is transient and decays to zero the very next day, for example, then it is not worth the effort to intervene in the first place. Therefore, it would be interesting to measure the duration of the price impact. Moreover, we also want to check if the price impact of the intervention depends on the market condition, i.e., the volatility level of the exchange rate. Because the price impact of the order flow is dependent on the volatility level: the price impact of the order flow is high when the volatility is high, and vice versa. Therefore, it is natural to expect that the price impact of central bank interventions would be high during the high volatility regime. Using the Markov-switching model, we can capture the asymmetry of the price impact in different regimes, as well as identify the state probability at each point

of time.

2.4.1 Effect on the Price Level

First of all, does the Japanese central bank intervention successfully move the price of the exchange rate? To answer this question, we study the price impact of the interventions on the price level of the exchange rate (JPY/USD). Note that the Japanese central bank interventions are universally sterilized. Sterilization of intervention means that any changes in monetary base due to foreign exchange intervention are absorbed by open market operations in the opposite direction, so that monetary basis remains unchanged (Ito 2007). In theory, sterilized foreign exchange interventions tend to be less effective at moving exchange rates than unsterilized interventions. This is due to the fact that unsterilized intervention to buy the foreign currency results in the expansion of domestic money base so that the interest rate declines. The lower interest rate will encourage capital outflows and cause the home currency to depreciate. However, if the central bank absorbs the expanded money due to intervention by selling domestic bonds in open market operations, then the monetary base and the interest rate remain unchanged, so there is no policy effect on capital flows. Therefore, theoretically, the unsterilized interventions are effective in moving the price level of the exchange rate, while the sterilized interventions would not be effective. Although the sterilized interventions have no impact on the domestic money supply, they would still alter the public relative supplies of available yen and dollar assets.

The specification of the linear model is as follows:

$$\Delta s_t = \beta_0 + \beta_s \Delta s_{t-1} + \beta_I I_t + \beta_{mo} mo_t + \text{lagged terms} + \epsilon_t, \quad (2.14)$$

where Δs_t is the difference of logged price of the exchange rate, i.e., the daily return, mo_t and I_t are the size of market order flow and central bank intervention, respectively. In addition, lagged terms will also be added to this equation. The model is a natural extension of the Evans-Lyons model. If intervention successfully moves the price level, the loading of intervention β_I should be significant and positive.

Model (2.14) also controls for the market order flow mo_t , which is the most important determinant of the changes in exchange rate. Evans and Lyons (2002) argue that order flow contains

private information on the exchange rate. Their empirical finding is that for two major exchange rates (mark/dollar and yen/dollar), including order flow as regressor would increase the R^2 dramatically, from 1-5% (the case when the only regressor is the interest rate differential) to 40-60%. Further research shows that the information content of order flow is not just private information. Order flow plays an intermediary role between exchange rates and macroeconomic fundamentals (Rime et al. 2010). Since placing an order is a willingness to back one's beliefs with real money, order flow is also a vehicle to aggregate macroeconomic information among all the market participants. Other studies (Love and Payne 2008, Evans and Lyons 2008; Carlson and Lo 2006) also document the empirical evidence to show that the impact of macro news operates primarily through order flow.

The data on central bank intervention and order flow is sampled at the daily frequency. The interest rate is also available at the daily frequency, but as a determinant of the exchange rate movement, interest rate differentials can only explain a few percent of the total variation, if any at all. This is the “exchange rate disconnect” puzzle (Meese and Rogoff, 1983; Cheung et al., 2005), which points out that fundamental variables, such as the underlying fundamentals: interest rates, inflation rates, and output, cannot explain the changes in exchange rates, especially in the short run (daily or monthly frequency). Therefore, we only control for the order flow that influences the exchange rate return beyond interventions. Omitting the order flow, the most important determinant, may cause some biases when evaluating the price impact of the intervention.

Insert table 2.1 about here

Insert table 2.2 about here

The empirical results of model (2.14) are reported in tables 2.1 and 2.2. As a robustness check, table 2.2 contains the regressions with the additional lagged terms of the independent variables. The upper, middle and lower parts of the tables report the results based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the sample (Nov. 2003 to Feb. 2013), respectively. Moreover, the estimates on the left and right hand sides are estimated using two-stage least square (2SLS) and ordinary least square (OLS) regressions, respectively. In table 2.1, the price impacts of the market order flow and intervention are both positive.

Every unit of market order flow (one thousand number of trades) has price impact of around 400 basis points, and every unit of intervention (100 billion JPY) can move the price level for about 3 or 4 basis points. The price impact of each unit of the intervention is much smaller than that of the market order flow. Based on the 2SLS estimates, the impact of the intervention is statistically significant for the whole sample and the second subsample, but not for the first subsample. The 2SLS and OLS estimates are qualitatively similar, but the coefficient of the intervention is slightly greater for OLS estimates. The similarity of the intervention coefficients based on 2SLS and OLS suggests that the intervention is unlikely to be endogenous with respect to the daily exchange rate movement. For the whole sample, first and second subsample, the Hausman tests fail to reject the null hypothesis that intervention is exogenous. Hence we treat the intervention as an exogenous variable when the dependent variable is the exchange rate movement. As reported in table 2.2, the intervention has similar price impact when the lagged terms are included.

2.4.2 Effect on the Volatility

To study the effect of the intervention on the volatility, the endogeneity issue also needs to be addressed. As we have mentioned, if the central bank tends to intervene when the market is extremely volatile, the intervention is an endogenous variable with respect to the dependent variable, the volatility of exchange rate. If it is indeed the case, OLS estimates are biased or even mistaken. Similarly, the absolute value of the market order flow $|mo_t|$ and the number of trades N_t can also be endogenous with respect to the volatility. For example, the order flow imbalance may cause the volatility to increase, simultaneously, the volatility could also drive the order flow imbalance. The similar argument also applies to the number of trades N_t . To address the endogeneity problem, we use the long term components of the time series $|I_t|$, $|mo_t|$ and N_t as the instrumental variables to evaluate the effect of Japanese central bank intervention on the volatility of JPY/USD.

One thing to note is that if the volatility σ_t is the dependent variable, R^2 of the regression would be very close to 1, which suggests the existence of the non-stationarity. The augmented Dickey-Fuller (ADF) test fails to reject the null hypothesis that there is a unit root in the time series of the volatility σ_t . To avoid the results of the spurious regression, we take the first difference

of the volatility σ_t and use it ($\Delta\sigma_t$) as the dependent variable. The model specification is as follows:

$$\Delta\sigma_t = \beta_0 + \beta_\sigma\Delta\sigma_{t-1} + \beta_I|I_t| + \beta_{mo}|mo_t| + \beta_N N_t + \text{lagged terms} + \epsilon_t, \quad (2.15)$$

where N_t is the number of trade at day t , $|I_t|$ and $|mo_t|$ are the absolute values of the intervention I_t and market order flow mo_t , respectively. Due to the symmetry of the exchange rate, the positive and negative I_t and mo_t are expected to have the same impact on the volatility: the order flow of selling JPY for USD should have the same effect on the volatility of JPY/USD as the order flow of selling USD for JPY. Therefore, in model 2.15, the absolute values of I_t and mo_t are used as the regressors for the dependent variable, $\Delta\sigma_t$. If the central bank intervention would raise the volatility of the exchange rate, the estimates of β_I in model 2.15 should be positive and significant. In our empirical exercise, additional lagged terms of the regressors are also included in the regression 2.15.

Insert table 2.3 about here

To avoid the potential endogeneity problem, we estimate model (2.15) using OLS as well as 2SLS regressions. The estimates based on 2SLS and OLS are reported on the left and right hand sides of the table 2.3, respectively. The estimates from simple OLS regressions indicate that central bank intervention would raise the volatility level, especially the day after the intervention, because the coefficient of $|I_{t-1}|$, the lagged absolute value of the intervention, is positive and statistically significant. For the two-stage least square (2SLS) regression, first of all, we use the overidentification tests to make sure that the instrumental variables are exogenous. The overidentification tests reported in table 2.3 confirm the exogeneity of the instrumental variables: the joint null hypothesis that the excluded instruments are valid instruments cannot be rejected, because the p values are greater than 0.05. Moreover, the R^2 of the 2SLS regression is very close to that of the OLS regression. It indicates that the instrumental variables have very good explanatory power for the original explanatory variables as well. The estimates of 2SLS regressions reported in table 2.3 are qualitatively identical to the OLS estimates, which support the conclusion that the Japanese central bank intervention increases the volatility level of JPY/USD.

2.4.3 Coordination Channel

Another interesting topic is to see how the intervention changes the behavior of other market participants. The intervention can work indirectly by inducing changes in the behavior of the market participants. This indirect channel is referred as the coordination channel. Taylor (2005) and Hung (1997) provide theoretical models for the coordination channel. In the Taylor (2005), intervention serves as a coordinating signal that induces fundamentalists to trade jointly, and this shift in private behavior accounts for intervention's success. Hung (1997) is in the same spirit, the difference is that the intervention induces a shift in the behavior of non-fundamental noise traders.

Empirically, we investigate whether the intervention causes the order flow and the number of trades to change. For the effect on the market order flow, we have the following model:

$$mo_t = \beta_0 + \sum_{i=1}^5 \beta_{mo_i} mo_{t-i} + \sum_{i=0}^2 \beta_{I_i} \hat{I}_{t-i} + \sum_{i=0}^2 \beta_{s_i} \Delta s_{t-i} + \epsilon_t, \quad (2.16)$$

where the market order flow mo_t is the dependent variable, and the explanatory variables are the sign function of the central bank intervention \hat{I}_t , difference of the logged price (return) Δs_t , and the lagged terms of these variables. The sign function of the intervention is defined as $\hat{I}_t = \text{sign}(I_t)$. If the central bank intervenes by selling JPY for USD at day t , $\hat{I}_t = 1$; if it sells USD for JPY, $\hat{I}_t = -1$; if there is no intervention, \hat{I}_t is zero. In theory, the coordination channel of the intervention suggests that the market participants would trade in line with the direction of the central bank intervention. Hence we expect that the coefficient β_I to be positive and significant. In model 2.17, the intervention \hat{I}_t and price change Δs_t could be endogenous with respect to the dependent variable mo_t . Similar as before, the long term component of the intervention time series and other exogenous variables are introduced as the IV.

The results based on 2SLS and OLS are reported on the left and right hand sides of table 2.4, respectively. In this case, OLS and 2SLS produce qualitatively different results, which substantiates the importance of the instrumental variables. The OLS estimates indicate that the intervention would increase the contemporaneous market order flow and decrease the market order of the next day. Aggregately, the intervention slightly increases the market order flow. However, with the help of the instrumental variables, two-stage least square regressions draw

a completely different conclusion. The intervention does not have any significant effect on the market order flow, since the parameters of the intervention and the lagged terms are not significantly different from zero. Moreover, the overidentification tests reported in table 2.4 guarantee us that the instrumental variables are valid, since the p values are much greater than 0.05. The R^2 of 2SLS regressions confirm the explanatory power of the instrumental variables for the original dependent variables. In conclusion, the empirical results based on 2SLS do not support the coordination channel. Since the intervention does not induce the market orders to become more aligned with the direction set by the central bank.

Insert table 2.4 about here

Following the same procedure, we also study the effect of the intervention on the number of trades. The model specification is as follows:

$$N_t = \beta_0 + \sum_{i=1}^6 \beta_{N_i} N_{t-i} + \sum_{i=0}^2 \beta_{I_i} |\hat{I}_{t-i}| + \sum_{i=0}^1 \beta_{mo_i} |mo_{t-i}| + \sum_{i=0}^1 \beta_{s_i} |\Delta s_{t-i}| + \epsilon_t, \quad (2.17)$$

where the regressors are the indicator of the intervention $|\hat{I}_t|$, whose value would be 1 if central bank intervenes, otherwise its value is zero; $|mo_t|$ and $|\Delta s_{t-i}|$ are the absolute values of the market order flow and price movement, respectively. The empirical results are reported in table 2.5. The 2SLS and OLS regressions produce qualitatively similar estimates, although the statistical significance of the 2SLS estimates is weaker. For the whole sample and the first subsample of the data, the results imply that Japanese central bank intervention would increase the number of trade on the contemporaneous day, and decrease the number of trades the next day. For the second subsample, the 2SLS regression suggests that the intervention does not change the number of trades at all. Moreover, similarly as before, the validity of the instrumental variables is assured by the results of the overidentification tests and the R^2 of the 2SLS regressions.

Insert table 2.5 about here

To sum up, the Japanese central bank intervention does not cause the order flow imbalance to change at all. It does not support the coordination channel, which suggests that the market order flow would be coordinated into the direction set by the central bank. However, for the

first subsample of the data, the intervention does change the number of trades: the number of trades would increase significantly at the intervention day, then decrease the day after the intervention. It means that the market participant would actively trade at the event of the intervention. So the intervention does cause the trading volume to increase, but the direction of the order imbalance is not aligned by the central bank intervention.

2.4.4 Damping Channel

Another indirect channel of the intervention is the damping channel, if the intervention is credible, it would damp the price impact of the private trades (Vitale 1999, Killeen, Lyons, and Moore 2006). To be more specific, intervention represents a signal about the exchange rate target of the central bank. The stronger the commitment, the more effective the intervention. At the extreme case, where the intervention is most effective, the informativeness of order flow vanishes, and private trades have no price impact. In this case, the elasticity of private demand becomes infinite and the volatility shrinks to zero. This is the fixed exchange rate regime. In contrast, when the intervention is perfectly ineffective at influencing the price level of the exchange rate, the price impact of the order flow is not changed by the intervention at all. Therefore, coordination channel suggests that the intervention is more effective when it dampens the price impact of the private order flow.

Insert table 2.6 about here

The model for evaluating the damping channel is as follows:

$$\Delta s_t = \beta_0 + \beta_{s,1}\Delta s_{t-1} + \beta_{s,2}\Delta s_{t-2} + \beta_I I_t + \beta_{mo} mo_t + \beta_* mo_t * \hat{I}_t + \epsilon_t, \quad (2.18)$$

where $mo_t * \hat{I}_t$ is the interaction term between the order flow mo_t and the indicator function of intervention. The model is the same one evaluating the price impact of the intervention and order flow, with the extra interaction term. When there is an intervention, the price impact of the order flow mo_t would be $\beta_{mo} + \beta_*$; if there is no intervention, the price impact is β_{mo} . Therefore, if the damping channel is effective, the coefficient of the interaction term should be negative and significantly different from zero, which means that the private orders' price impact would be smaller if there is an intervention. The estimates of model 2.18 are reported in table 2.6.

It is easy to observe that for the whole sample and the first subsample, the empirical evidence supports the damping channel: the loading of the interaction term is negative and significant. It suggests that the price impact of market order flow would decrease in the presence of central bank intervention. However, the coefficient of the interaction term is not significant from zero for the second subsample. It is due to the fact that there are fewer interventions in the second subsample.

2.4.5 Duration of the Price Impact

After confirming the statistical significance of the price impact of the intervention, we would also like to find out whether the price impact is permanent or transitory. To be more specific, what is the price impact of the intervention at different horizons? Does it last for a day or a month? From the point of view of the central bank, the persistence of the price impact matters, because if the effect of interventions is transient, then it is not worth the effort to intervene in the first place.

We will measure the duration of the price impact by studying the price impact at different scales or horizons. In other words, we want to decompose the economic relationship between the exchange rate movement and the intervention by timescale. To accomplish this goal, we decompose the original time series into orthogonal components at different horizons or time scales. Each component captures all the variations at that particular time scale, and ‘orthogonal’ means that the components at different scales are completely uncorrelated from each other. To estimate the price impact at a particular scale, we regress the component of the price movements on the corresponding components of the regressors. Furthermore, since the intervention is exogenous with respect to the price movement, as established before, we simply use OLS to estimate the parameters scale by scale. Based on the coefficients, one can check if the price impact of the intervention decays as the time scale grows.

The methodology we adopt is similar as Ramsey and Lampart (1998), which studies the economic relationship between expenditure and income at different time scales. Ramsey and Lampart (1998) decompose the time series of the expenditure and income into orthogonal components at various scales. They document the statistically significant relationship between the long term

expenditure and long term income, but not between the short term ones. It means that in the long term, higher income would increase the expenditure, but there is no clear relationship between them in the short term. It sounds reasonable and is consistent with the economic theory.

In this paper, wavelets analysis is used to obtain the orthogonal decomposition of the relevant economic variables at five different time scales. Due to limited space, we will not discuss wavelet analysis here. We refer the interested readers to Ramsey and Lampart (1998), and Ramsey (2002) for more details. Briefly speaking, the time series of the return (Δs_t), market order flow (mo_t) and intervention (I_t) are decomposed into orthogonal components by scale:

$$X_t = \sum_{j=1}^5 X_t[d_j] + X_t[a_5] = X_t[d_1] + X_t[d_2] + X_t[d_3] + X_t[d_4] + X_t[d_5] + X_t[a_5], \quad (2.19)$$

where X represents one of the variables, such as Δs_t , mo_t , and I_t , and $X_t[d_j]$ is the component of X_t at the j th scale. Intuitively speaking, the component $X_t[d_i]$ captures the variations at 2^{j-1} days horizon, and none of the other scales. To be more specific, $X_t[d_1]$ is the highest frequency variations at the 1 day horizon; $X_t[d_2]$ captures the variations between 1 day and 2 day horizon; $X_t[d_3]$ captures the variations between 2 day and 4 day horizon, and so on ... At last, the component $X_t[a_5]$ contains all the variations with the horizon greater than (including) 2^5 days. Essentially, $\{X_t[d_i], i = 1, 2, \dots, 5\}$ capture the variations or ‘details’ at various scales, while $X_t[a_5]$ contains the long term variations that are not captured by $\{X_t[d_i], i = 1, 2, \dots, 5\}$, so $X_t[a_5]$ is the long term ‘average’ or ‘approximation’ of the original time series X_t . Moreover, the components at different scales are completely uncorrelated from each other by design: the wavelet bases at different scale are orthogonal to each other. This is why we can decompose the relationship between the exchange rate movement and the intervention by time scale, without worrying about the relationship at different scales being coupled together.

Insert table 2.7 about here

At scale j , we regress the component $\Delta s_t[d_j]$ on the corresponding components of mo_t and I_t :

$$\Delta s_t[d_j] = \beta_{0[d_j]} + \beta_{s[d_j]} \Delta s_{t-1}[d_j] + \beta_{I[d_j]} I_t[d_j] + \beta_{mo[d_j]} mo_t[d_j] + \epsilon_t, \quad j = 1, 2, \dots, 5, \quad (2.20)$$

where $\Delta s_t[d_j]$, $I_t[d_j]$, and $mo_t[d_j]$ are the components of Δs_t , mo_t , and I_t at scale j , respectively. The coefficient $\beta_{s[d_j]}$ measures the price impact at scale j . If the central bank interventions have long lasting effect on the price level, we expect that the coefficients on the intervention will be significant at small as well as large scales. The model 2.20 is estimated using OLS method, and the regression results at scales 1 to 5 are reported in table 2.7. The estimates of $\beta_{I[d_j]}$ suggest that the price impact of intervention gradually decays to zero as the scale becomes greater: the price impact starts to decay at scale 4; and at scale $j = 5$, or the horizon of 16 (2^4) trading days, the impact of intervention would dissipate completely. Notice that the price impact of the market order flow has much more persistent price impact. It gradually decays, but is still strongly significant from zero after 2^5 trading days. Therefore, we conclude that Japanese central bank interventions have a relatively long-lasting price impact.

2.4.6 Volatility Regimes and Price Impact

Last but not least, we want to investigate whether the price impact of the intervention depends on the market condition. The previous researchers find that the effect of the intervention is higher if trading volumes are higher, or if the intervention occurs shortly after important macro announcements (e.g., Dominguez 2003). The conjecture we are going to test is that the price impact of the intervention should be greater in the high volatility regime. It is because that the Evans-Lyons model (Killeen, Lyons, and Moore 2006) suggests that price impact of order flow would increase with the volatility level. Due to the fact that the elasticity of public's speculative demand is finite and inversely related to the volatility of the exchange rate, if the volatility is high, the elasticity of public's speculative demand would be low, which means the public would trade less aggressively. Moreover, since the dealers don't want to hold overnight positions, at the end of the trading day, if the market is volatile, they have to make more adjustment to the quote, in order to induce the speculators to re-absorb all the positions. Therefore, the order flow will have a greater impact on the exchange rate in the high volatility regime.

To measure the asymmetry of the price impact in high and low volatility regimes, we use the Markov switching model. It is a very parsimonious extension of the single regime model. There is a strand of research uses Markov switching model to study the secular changes in the economic system. Hamilton's (1989) seminal work applies the regime switching model to study business

cycle, recessions and expansions. The regimes in his paper are closely tied to the recession indicators as identified *ex post* by the NBER business cycle dating committee. Sim and Zha (2006) use a multivariate regime-switching model to investigate U.S. monetary policy between the 1970s and the 1980s, during which the monetary policy changed a great deal. They find three estimated regimes corresponding to periods when most observers believe that monetary policy actually differed. Perez-Quiros and Timmermann (2000) adopt the regime switching framework to model the variation of small and large firms' risk over the economic cycle: small firms with little collateral is more strongly affected by tighter credit market conditions in a recession state than large ones. For further applications of Markov switching model in finance, see Henkel, Martin and Nardari (2011), Engel and Hamilton (1990), and Ichiue and Koyama (2011).

As the returns of other financial assets, time series of the changes in exchange rates display two features: recurring alternation of stable and volatile regimes and volatility clustering, which can be observed in the upper part in figure (2.4). The Markov switching model can accommodate these two features, while has the flexibility of allowing the linear relationship between economic variables to be different in different regimes of the world. Our Markov-switching model is a two-regime extension of the linear model (2.14):

$$\Delta S_t = \beta_{0,s} + \beta_{S,s}\Delta S_{t-1} + \beta_{mo,s}mo_t + \beta_{I,s}I_t + \epsilon_{t,s}, \quad \epsilon_{t,s} \sim N(0, \sigma_s^2) \quad \text{for } s=1 \text{ or } 2. \quad (2.21)$$

where the subscript s in equation (2.21) denotes the particular state the system is in. To be more specific, the model is as follows

$$\Delta S_t = \beta_{0,1} + \beta_{S,1}\Delta S_{t-1} + \beta_{mo,1}mo_t + \beta_{I,1}I_t + \epsilon_{t,1}, \quad \epsilon_{t,1} \sim N(0, \sigma_1^2) \quad \text{for state 1}, \quad (2.22)$$

$$\Delta S_t = \beta_{0,2} + \beta_{S,2}\Delta S_{t-1} + \beta_{mo,2}mo_t + \beta_{I,2}I_t + \epsilon_{t,2}, \quad \epsilon_{t,2} \sim N(0, \sigma_2^2) \quad \text{for state 2}. \quad (2.23)$$

At time t , the system is either in state 1 or state 2. Each pair of the corresponding parameters in different states can be different, for example, the price impact of the market order flow could be different: $\beta_{mo,1} \neq \beta_{mo,2}$. The transition of states, or regimes, is stochastic: from time t to $t + 1$, the system has some probability to stay in the original state, or it can jump to the other

state with a certain probability. The evolution of the regime or state is governed by a first-order Markov chain, and the transition probability from state i to state j is

$$Pr(s_{t+1} = j | s_t = i) = p_{ij} , \quad (2.24)$$

where we assume p_{ij} is a constant, which needs to be estimated empirically. There are four transition probabilities, $p_{ij}, i = 1, 2$, but only two of them are independent, because of the following relationships:

$$Pr(s_t = 1 | s_{t-1} = 1) = p_{11} = 1 - Pr(s_t = 2 | s_{t-1} = 1) = 1 - p_{12} , \quad (2.25)$$

and

$$Pr(s_t = 1 | s_{t-1} = 2) = p_{21} = 1 - Pr(s_t = 2 | s_{t-1} = 2) = 1 - p_{22} . \quad (2.26)$$

Another thing worth mentioning is that in the specification of our model, equation (2.21), we don't impose any assumption related to the high/low volatility regimes *ex ante*. We simply allow the possibility of the existence of two states at each point of time, and each pair of parameters across two states could be different. The identification of the state probability at each point of time and the values of the parameters are estimated *ex post* by the econometric method, based on the data.

Insert table 2.8 here

We estimate the probabilities of the system in state 1 and 2 at every point of time t , as well as the coefficients in equation (2.21). Figure (2.4) shows the daily price changes in JPY/USD, conditional standard deviation of equation (2.21), and the smoothed state probability of state 2 based on the second subsample of the data. It is easy to observe that the smoothed state probability of state 2 exactly coincides with the turbulent periods of the exchange rate: at each point of time, if the volatility level is high, the state probability of state 2 would be one or close to one, and *vice versa*. Hence, state 1 and 2 correspond to the low and high volatility regimes, respectively. Note that in the original assumption of the model 2.21, we do not specify the correspondence between state 2 and the high volatility regime: states 1 and 2 are completely symmetry *ex ante*. Based on the data, the smoothed state probability of state 2 at each point

of time is inferred. By comparing the time series of the volatility and smoothed probability of state 2, we can identify that the state 2 actually corresponds to the high volatility regime.

The estimates of the Markov switching model based on the whole sample, first and second subsamples are reported in table 2.8. It is easy to observe the asymmetry of the coefficients across two states: $\beta_{mo,1} < \beta_{mo,2}$, and $\beta_{I,1} < \beta_{I,2}$, which indicate that both market order flow and central bank intervention have greater price impact in the high volatility regime, compared with those in the low volatility regime. For instance, the estimated coefficients of intervention in state 1 and 2 are 3.34 and 12.81, respectively; and the loadings of the market order flow in state 1 and 2 are 370.9 and 604.1, respectively. The results are consistent with the theoretical argument discussed before.

With regards to the other properties of the Markov switching model, one can observe that p_{11} and p_{22} are much greater than p_{12} and p_{21} , respectively. It implies that the system is much more likely to stay in the same regime than transfer to the other regime: periods of high volatility are followed by periods of high volatility, and periods of low volatility are followed by periods low volatility. This is the feature of volatility clustering. The low volatility regime is much more persistent, as p_{11} is very close to one. We can also observe that the low volatility regime lasts longer, compared with the high volatility one: the expected state durations of the low volatility regime (state 1) are much greater than those of the high volatility regime (state 2).

Furthermore, we should point out that it is a universal fact that the price impact of the order flow is greater in the high volatility regime. It is not just specific to JPY/USD or happens by accident. Based on the order flow data of thirteen exchange rates actively traded on the Reuters dealing system, we study the price impact of market order flow using Markov switching model. The sample period is 2003 to 2013. Table (2.9) summarizes the estimates of the Markov switching model with market order flow mo_t as explanatory variable. State 2 corresponds the high volatility regime. The results verify that, for all the thirteen different exchange rates, the market order flow has much greater impact on the exchange rates when the volatility is high.

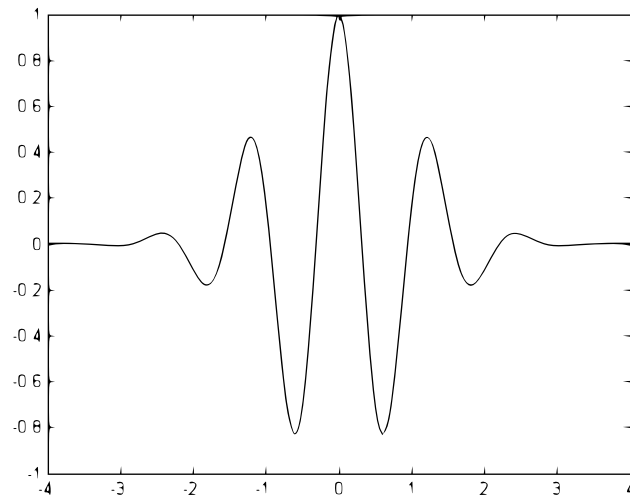
The fact that the price impact is volatility dependent has policy implications as well. First of all, with the intention to move the price level of the exchange rate, central bank should intervene when it is volatile, because the intervention would have much greater price impact;

on the other hand, if the central bank intends to adjust its portfolio holdings (foreign reserve) without disturbing the exchange rate, the trades should be carried out when the volatility is low. If central bank executes transactions representing its government, it would be better off to trade when the volatility level is low, because the greater price impact during volatile periods would incur greater transaction cost.

2.5 Conclusions

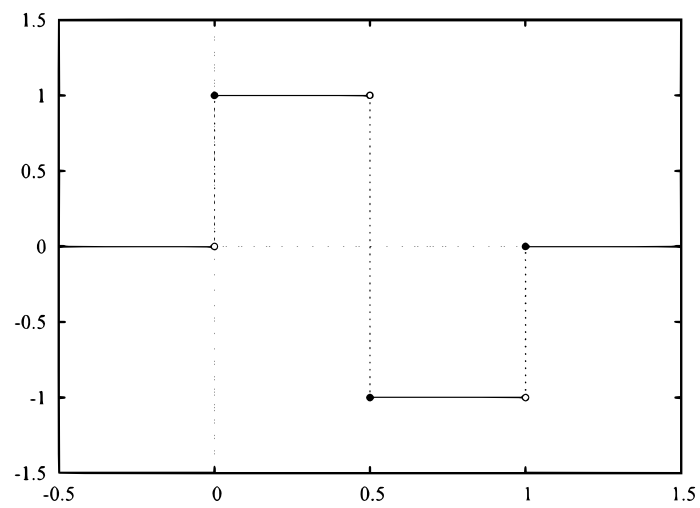
This paper conducts a thorough empirical investigation on the effects of the central bank intervention. To evaluate the effects of the intervention accurately, we propose a novel instrumental variable to resolve the endogeneity issue. Since the central bank intervention is highly likely to be endogenous variable with respect to various dependent variables, the instrumental variable method in this paper would be indispensable for the central bank to obtain accurate feedbacks on its interventions. Empirically, we find that the Japanese central bank intervention successfully moves the price level of the exchange rate, and it also increases the volatility level. With respect to the effects on the market participants behavior, on one hand, the market order is not changed by the intervention, which does not support the coordination channel; on the other hand, the intervention increases the total number of trades on the intervention day but decreases it in the following day. With respect to the damping channel, we do find the evidence that the private trades' price impact is damped by the central bank intervention. Moreover, the price impact of the intervention lasts for 16 trading days, while the price impact of the market order flow is much more persistent. We also document that the price impact of the intervention depends on the volatility of the exchange rate: the price impact would be high when the volatility is high, and *vice versa*. Last but not least, we discuss the policy implications of our empirical results.

Figure 2.1: Morlet wavelet



This figure shows the basis function of Morlet wavelet.

Figure 2.2: Haar wavelet



This figure shows the basis function of Haar wavelet.

Figure 2.3: Japanese central bank intervention and its wavelet decomposition

This figure shows the time series of Japanese central bank intervention and its wavelet decomposition using Haar wavelet. The sample period is 1996 to 2013. The first panel shows the original central bank intervention. The second and third panels shows the long-term component ‘a1’ ($I_t[a_1]$) and the short term component ‘d1’ ($I_t[d_1]$). The unit is 100 million JPY.

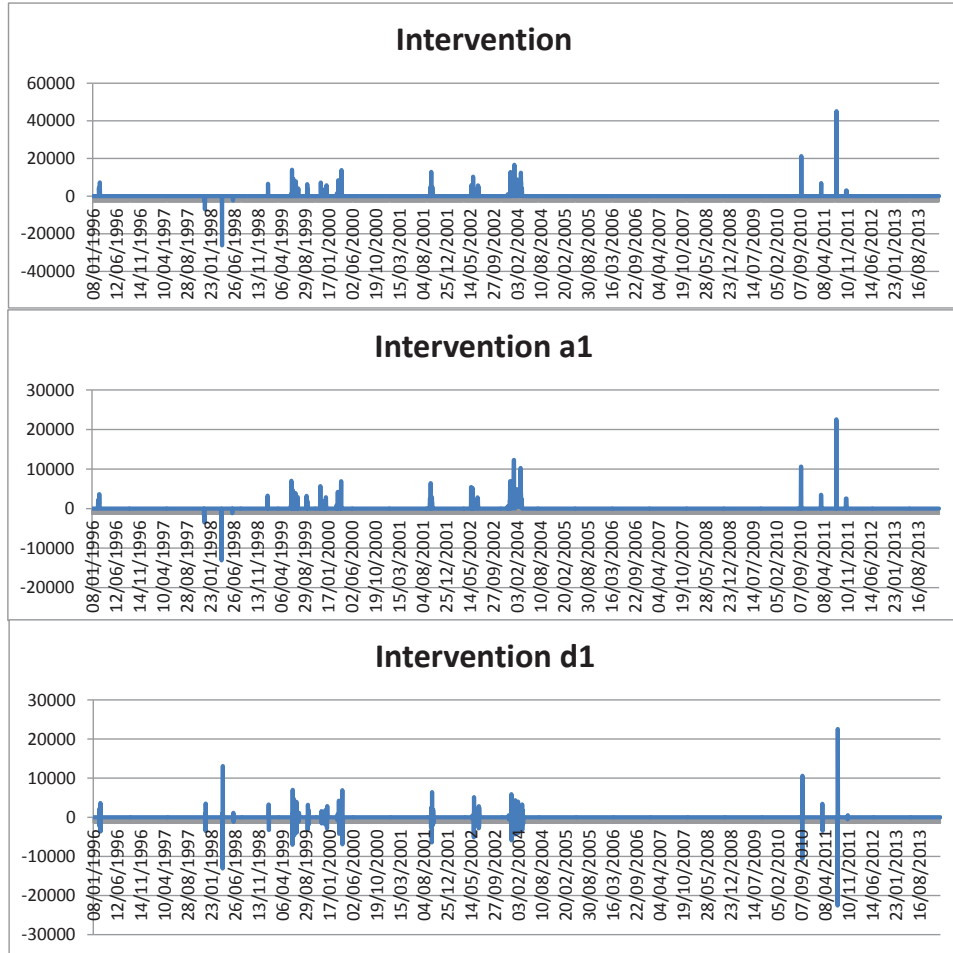
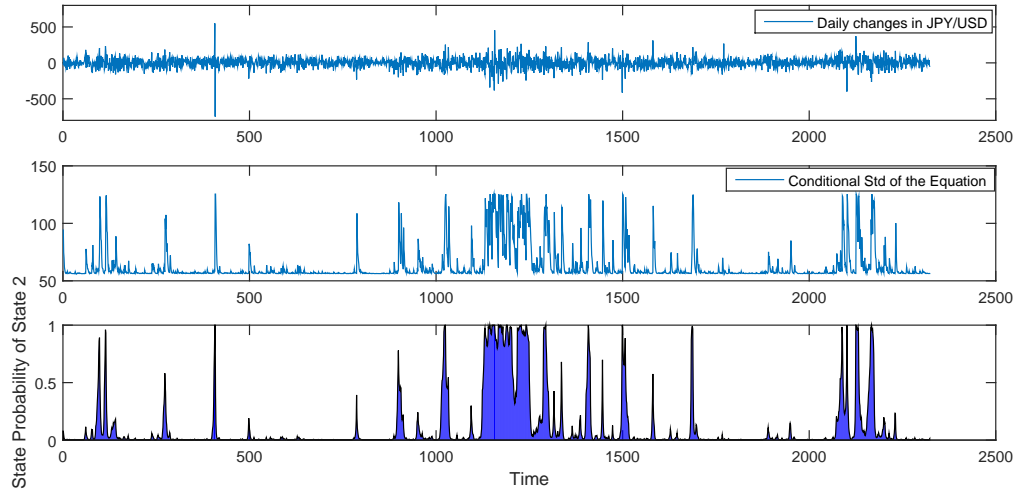


Figure 2.4: Correspondence between the states of the Markov switching model and the volatility regimes



This figure shows the correspondence between the states of the Markov switching model and the high/low volatility regimes of JPY/USD. The two-state Markov switching model is as follows: $\Delta S_t = \beta_{0,s} + \beta_{S,s}\Delta S_{t-1} + \beta_{mo,s}mo_t + \beta_{I,s}I_t + \epsilon_{t,s}$, $\epsilon_{t,s} \sim N(0, \sigma_s^2)$, for $s=1$ or 2 , where subscript s labels the state of the system, mo_t is the market order flow, I_t is the intervention, ΔS_t is the daily changes in JPY/USD (difference of logged price). The upper diagram shows the daily changes in JPY/USD, the middle diagram reports the conditional standard deviation of the equation, and the lower diagram is the smoothed state probability of state 2. The results are based on the second part of the data (Nov. 2003 to Feb. 2013)

Table 2.1: Effect on the price level, without the lagged variables

2SLS, the whole sample				OLS, the whole sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-3.5920	1.0794	-3.33	Intercept	-3.7488	1.0773	-3.48
ΔS_{t-1}	-0.1139	0.0138	-8.26	ΔS_{t-1}	-0.1130	0.0138	-8.21
I_t	3.3101	1.2325	2.69	I_t	4.8442	0.9346	5.18
mo_t	393.0025	15.3793	25.55	mo_t	390.8935	15.3399	25.48
R-Square	0.1372			R-Square	0.1403		
Adj R-Sq	0.1366			Adj R-Sq	0.1398		
Test for Overidentifying Restrictions				p value = 0.2266			
2SLS, first part of the sample				OLS, first part of the sample, OLS			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-8.0189	1.6328	-4.91	Intercept	-8.1542	1.6308	-5
ΔS_{t-1}	-0.1250	0.0191	-6.56	ΔS_{t-1}	-0.1241	0.0190	-6.52
I_t	3.3370	2.2048	1.51	I_t	4.6234	1.6091	2.87
mo_t	393.7743	17.7514	22.18	mo_t	392.2864	17.6883	22.18
R-Square	0.1937			R-Square	0.1953		
Adj R-Sq	0.1926			Adj R-Sq	0.1943		
Test for Overidentifying Restrictions				p value = 0.6016			
2SLS, Second part of the sample				OLS, second part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	0.8369	1.4398	0.58	Intercept	0.6509	1.4342	0.45
ΔS_{t-1}	-0.1017	0.0200	-5.08	ΔS_{t-1}	-0.1011	0.0200	-5.05
I_t	2.9016	1.4554	1.99	I_t	4.6720	1.1251	4.15
mo_t	463.7536	37.5765	12.34	mo_t	458.5447	37.4481	12.24
R-Square	0.0728			R-Square	0.0777		
Adj R-Sq	0.0716			Adj R-Sq	0.0766		
Test for Overidentifying Restrictions				p value = 0.1949			

The table reports the results of the model $\Delta s_t = \beta_0 + \beta_s \Delta s_{t-1} + \beta_I I_t + \beta_{mo} mo_t + \epsilon_t$, where Δs_t is the return (difference of the logged price), I_t is the central bank intervention, and mo_t is the market order flow. The coefficients of I_t and mo_t measure the price impact of the intervention and market order flow, respectively. The 2SLS and OLS estimates are shown on the left and right hand side of the table. The results are based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the data (Nov. 2003 to Feb. 2013), respectively. The p values of the overidentification tests indicate the exogeneity of the instrumental variables cannot be rejected.

Table 2.2: Effect on the price level, with the lagged variables

2SLS, the whole sample				OLS, the whole sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-3.1177	1.0914	-2.86	Intercept	-3.1913	1.0903	-2.93
ΔS_{t-1}	-0.1061	0.0147	-7.2	ΔS_{t-1}	-0.1055	0.0147	-7.17
I_t	3.9267	1.2477	3.15	I_t	5.2543	0.9515	5.52
I_{t-1}	0.1552	0.9647	0.16	I_{t-1}	-0.0108	0.9593	-0.01
I_{t-2}	-2.1685	0.9571	-2.27	I_{t-2}	-2.2996	0.9536	-2.41
I_{t-3}	-1.4429	0.9509	-1.52	I_{t-3}	-1.5298	0.9493	-1.61
mo_t	398.2107	15.6464	25.45	mo_t	396.7122	15.6166	25.4
mo_{t-1}	-27.5054	16.6077	-1.66	mo_{t-1}	-26.9673	16.6010	-1.62
R-Square	0.1394			R-Square	0.1428		
Adj R-Sq	0.1381			Adj R-Sq	0.1415		
Test for Overidentifying Restrictions				p value = 0.2837			
2SLS, first part of the sample				OLS, first part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-7.7522	1.6822	-4.61	Intercept	-7.8543	1.6807	-4.67
ΔS_{t-1}	-0.1257	0.0210	-5.99	ΔS_{t-1}	-0.1248	0.0210	-5.95
I_t	2.4608	1.9840	1.24	I_t	4.7483	1.6181	2.93
I_{t-1}	1.7945	1.6247	1.1	I_{t-1}	1.6466	1.6223	1.01
I_{t-2}	-2.3623	1.6173	-1.46	I_{t-2}	-2.5371	1.6142	-1.57
I_{t-3}	-2.0716	1.6133	-1.28	I_{t-3}	-2.1025	1.6125	-1.30
mo_t	396.2563	17.9413	22.09	mo_t	394.2039	17.9038	22.02
mo_{t-1}	-4.0073	19.6591	-0.20	mo_{t-1}	-3.0782	19.6448	-0.16
R-Square	0.1951			R-Square	0.1973		
Adj R-Sq	0.1926			Adj R-Sq	0.1948		
Test for Overidentifying Restrictions				p value = 0.6293			
2SLS, Second part of the sample				OLS, second part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	1.0428	1.4458	0.72	Intercept	0.9407	1.4434	0.65
ΔS_{t-1}	-0.0918	0.0207	-4.44	ΔS_{t-1}	-0.0910	0.0207	-4.4
I_t	4.0910	1.5260	2.68	I_t	5.4277	1.1573	4.69
I_{t-1}	-0.9492	1.1816	-0.80	I_{t-1}	-1.1571	1.1711	-0.99
I_{t-2}	-2.0300	1.1670	-1.74	I_{t-2}	-2.1677	1.1621	-1.87
I_{t-3}	-1.1427	1.1574	-0.99	I_{t-3}	-1.2659	1.1534	-1.10
mo_t	478.5274	38.4071	12.46	mo_t	474.8952	38.3009	12.4
mo_{t-1}	-71.9244	39.3667	-1.83	mo_{t-1}	-71.9911	39.3553	-1.83
R-Square	0.0769			R-Square	0.0823		
Adj R-Sq	0.0741			Adj R-Sq	0.0795		
Test for Overidentifying Restrictions				p value = 0.2255			

The table reports the results of the model $\Delta s_t = \beta_0 + \beta_s \Delta s_{t-1} + \beta_I I_t + \beta_{mo} mo_t + \text{lagged terms} + \epsilon_t$, where Δs_t is the return, I_t is the central bank intervention, and mo_t is the market order flow. The coefficients of I_t and mo_t measure the price impact of the intervention and market order flow, respectively. The 2SLS and OLS estimates are shown on the left and right hand side of the table. The results are based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the data (Nov. 2003 to Feb. 2013), respectively. The p values of the overidentification tests indicate the exogeneity of the instrumental variables.

Table 2.3: Effect on the volatility

2SLS, the whole sample				OLS, the whole sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.1526	0.0758	-2.01	Intercept	-0.1719	0.0716	-2.4
$\Delta\sigma_{t-1}$	0.1143	0.0148	7.74	$\Delta\sigma_{t-1}$	0.1140	0.0148	7.73
$\Delta\sigma_{t-2}$	-0.0131	0.0146	-0.89	$\Delta\sigma_{t-2}$	-0.0130	0.0146	-0.89
$ I_t $	-0.0484	0.0560	-0.86	$ I_t $	0.0191	0.0427	0.45
$ I_{t-1} $	0.2773	0.0432	6.41	$ I_{t-1} $	0.2683	0.0430	6.24
$ I_{t-2} $	-0.0869	0.0433	-2.01	$ I_{t-2} $	-0.0934	0.0431	-2.16
$ I_{t-3} $	-0.0326	0.0430	-0.76	$ I_{t-3} $	-0.0365	0.0429	-0.85
$ mo_t $	-3.1807	1.7268	-1.84	$ mo_t $	-1.7182	1.0980	-1.56
$ mo_{t-1} $	4.7882	1.1134	4.3	$ mo_{t-1} $	4.5863	1.1010	4.17
N_t	-0.2298	0.1230	-1.87	N_t	-0.3363	0.0956	-3.52
N_{t-1}	0.7941	0.1235	6.43	N_{t-1}	0.8401	0.1154	7.28
N_{t-2}	-0.4703	0.0825	-5.7	N_{t-2}	-0.4679	0.0824	-5.68
R-Square	0.0484	Adj R-Sq	0.0461	R-Square	0.0502	Adj R-Sq	0.0479
Test for Overidentifying Restrictions				p value = 0.1631			
2SLS, first part of the sample				OLS, first part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.4854	0.1844	-2.63	Intercept	-0.4340	0.1733	-2.5
$\Delta\sigma_{t-1}$	0.0438	0.0212	2.07	$\Delta\sigma_{t-1}$	0.0434	0.0212	2.05
$\Delta\sigma_{t-2}$	-0.0015	0.0209	-0.07	$\Delta\sigma_{t-2}$	-0.0011	0.0209	-0.05
$ I_t $	0.1062	0.0994	1.07	$ I_t $	0.1440	0.0811	1.77
$ I_{t-1} $	0.1492	0.0817	1.83	$ I_{t-1} $	0.1423	0.0815	1.75
$ I_{t-2} $	-0.0571	0.0816	-0.7	$ I_{t-2} $	-0.0602	0.0814	-0.74
$ I_{t-3} $	-0.0424	0.0812	-0.52	$ I_{t-3} $	-0.0433	0.0811	-0.53
$ mo_t $	-3.4131	2.2964	-1.49	$ mo_t $	-2.1320	1.4178	-1.5
$ mo_{t-1} $	3.6490	1.4301	2.55	$ mo_{t-1} $	3.4620	1.4215	2.44
N_t	-0.1281	0.1626	-0.79	N_t	-0.2770	0.1217	-2.28
N_{t-1}	0.7920	0.1469	5.39	N_{t-1}	0.8595	0.1379	6.23
N_{t-2}	-0.3103	0.1059	-2.93	N_{t-2}	-0.3250	0.1049	-3.1
R-Square	0.0399	Adj R-Sq	0.0351	R-Square	0.0438	Adj R-Sq	0.0391
Test for Overidentifying Restrictions				p value = 0.1446			
2SLS, Second part of the sample				OLS, second part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.3583	0.1109	-3.23	Intercept	-0.4089	0.1092	-3.74
$\Delta\sigma_{t-1}$	0.2588	0.0206	12.57	$\Delta\sigma_{t-1}$	0.2596	0.0206	12.63
$\Delta\sigma_{t-2}$	-0.0721	0.0203	-3.54	$\Delta\sigma_{t-2}$	-0.0706	0.0203	-3.48
$ I_t $	-0.0857	0.0566	-1.51	$ I_t $	-0.0555	0.0430	-1.29
$ I_{t-1} $	0.3581	0.0438	8.19	$ I_{t-1} $	0.3546	0.0433	8.18
$ I_{t-2} $	-0.1572	0.0442	-3.56	$ I_{t-2} $	-0.1579	0.0440	-3.59
$ I_{t-3} $	0.0133	0.0438	0.3	$ I_{t-3} $	0.0098	0.0436	0.22
$ mo_t $	-5.8072	3.2106	-1.81	$ mo_t $	-4.2816	2.1814	-1.96
$ mo_{t-1} $	4.5211	2.2207	2.04	$ mo_{t-1} $	4.1623	2.1809	1.91
N_t	-0.2561	0.9979	-0.26	N_t	0.7131	0.7975	0.89
N_{t-1}	7.0790	0.8749	8.09	N_{t-1}	6.5847	0.8439	7.8
N_{t-2}	-4.4021	0.7500	-5.87	N_{t-2}	-4.7518	0.7373	-6.44
R-Square	0.12835	Adj R-Sq	0.12419	R-Square	0.1279	Adj R-Sq	0.1237
Test for Overidentifying Restrictions				p value = 0.6024			

The table reports the results of the model $\Delta\sigma_t = \beta_0 + \beta_\sigma\Delta\sigma_{t-1} + \beta_I|I_t| + \beta_{mo}|mo_t| + \beta_N N_t + \text{lagged terms} + \epsilon_t$, where $\Delta\sigma_t$ is the difference of the volatility; $|I_t|$ and $|mo_t|$ are the absolute value of the intervention and market order flow, respectively; N_t is the number of trades. The 2SLS and OLS estimates are shown on the left and right hand side of the table. The results are based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the data (Nov. 2003 to Feb. 2013), respectively.

Table 2.4: Effect on the market order flow

2SLS, the whole sample				OLS, the whole sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	0.0060	0.0010	6.04	Intercept	0.0059	0.0010	6.07
mo_{t-1}	0.1502	0.0149	10.07	mo_{t-1}	0.1508	0.0149	10.16
mo_{t-2}	0.0829	0.0150	5.53	mo_{t-2}	0.0831	0.0150	5.55
mo_{t-3}	0.0530	0.0141	3.77	mo_{t-3}	0.0529	0.0140	3.76
mo_{t-4}	0.0621	0.0140	4.44	mo_{t-4}	0.0624	0.0140	4.46
mo_{t-5}	0.0134	0.0139	0.96	mo_{t-5}	0.0137	0.0138	0.99
\hat{I}_t	0.0147	0.0283	0.52	\hat{I}_t	0.0260	0.0082	3.16
\hat{I}_{t-1}	-0.0124	0.0142	-0.87	\hat{I}_{t-1}	-0.0171	0.0086	-1.99
\hat{I}_{t-2}	0.0117	0.0115	1.02	\hat{I}_{t-2}	0.0084	0.0082	1.02
ΔS_t	0.0003	0.0000	24.79	ΔS_t	0.0003	0.0000	25.48
ΔS_{t-1}	0.0000	0.0000	2.79	ΔS_{t-1}	0.0000	0.0000	2.88
ΔS_{t-2}	0.0000	0.0000	-0.56	ΔS_{t-2}	0.0000	0.0000	-0.53
R-Square	0.1793	Adj R-Sq	0.1773	R-Square	0.1808	Adj R-Sq	0.1788
Test for Overidentifying Restrictions				p value = 0.8194			
2SLS, first part of the sample				OLS, first part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	0.0162	0.0019	8.41	Intercept	0.0161	0.0019	8.45
mo_{t-1}	0.1174	0.0213	5.51	mo_{t-1}	0.1181	0.0213	5.56
mo_{t-2}	0.0506	0.0214	2.37	mo_{t-2}	0.0509	0.0214	2.38
mo_{t-3}	0.0365	0.0194	1.88	mo_{t-3}	0.0363	0.0194	1.87
mo_{t-4}	0.0420	0.0194	2.17	mo_{t-4}	0.0426	0.0194	2.2
mo_{t-5}	-0.0014	0.0193	-0.07	mo_{t-5}	-0.0007	0.0192	-0.04
\hat{I}_t	0.0169	0.0336	0.5	\hat{I}_t	0.0326	0.0134	2.44
\hat{I}_{t-1}	-0.0049	0.0149	-0.33	\hat{I}_{t-1}	-0.0082	0.0134	-0.61
\hat{I}_{t-2}	0.0240	0.0149	1.62	\hat{I}_{t-2}	0.0207	0.0134	1.55
ΔS_t	0.0005	0.0000	21.36	ΔS_t	0.0004	0.0000	21.83
ΔS_{t-1}	0.0001	0.0000	2.61	ΔS_{t-1}	0.0001	0.0000	2.66
ΔS_{t-2}	0.0000	0.0000	0.41	ΔS_{t-2}	0.0000	0.0000	0.43
R-Square	0.2116	Adj R-Sq	0.2077	R-Square	0.2133	Adj R-Sq	0.2094
Test for Overidentifying Restrictions				p value = 0.6990			
2SLS, second part of the sample				OLS, second part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.0015	0.0008	-1.93	Intercept	-0.0014	0.0008	-1.88
mo_{t-1}	0.1694	0.0209	8.12	mo_{t-1}	0.1700	0.0208	8.16
mo_{t-2}	0.1106	0.0211	5.25	mo_{t-2}	0.1103	0.0210	5.24
mo_{t-3}	0.0672	0.0205	3.27	mo_{t-3}	0.0669	0.0205	3.26
mo_{t-4}	0.0638	0.0204	3.12	mo_{t-4}	0.0633	0.0204	3.1
mo_{t-5}	0.0046	0.0202	0.23	mo_{t-5}	0.0052	0.0201	0.26
\hat{I}_t	0.0276	0.0321	0.86	\hat{I}_t	0.0124	0.0086	1.45
\hat{I}_{t-1}	-0.0163	0.0220	-0.74	\hat{I}_{t-1}	-0.0067	0.0100	-0.67
\hat{I}_{t-2}	-0.0010	0.0113	-0.09	\hat{I}_{t-2}	0.0025	0.0086	0.29
ΔS_t	0.0001	0.0000	12.55	ΔS_t	0.0001	0.0000	12.83
ΔS_{t-1}	0.0000	0.0000	2.69	ΔS_{t-1}	0.0000	0.0000	2.64
ΔS_{t-2}	0.0000	0.0000	-0.56	ΔS_{t-2}	0.0000	0.0000	-0.6
R-Square	0.1421	Adj R-Sq	0.1380	R-Square	0.1427	Adj R-Sq	0.1386
Test for Overidentifying Restrictions				p value = 0.4611			

The table reports the results of the model $mo_t = \beta_0 + \sum_{i=1}^5 \beta_{mo_i} mo_{t-i} + \sum_{i=0}^2 \beta_{I_i} \hat{I}_{t-i} + \sum_{i=0}^2 \beta_{s_i} \Delta s_{t-i} + \epsilon_t$, where mo_t is the market order flow, $\hat{I}_{t-i} = \text{sign}(I_{t-i})$, the sign function of the intervention, and Δs_t is the return. The 2SLS and OLS estimates are shown on the left and right hand side of the table. The results are based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the data (Nov. 2003 to Feb. 2013), respectively.

Table 2.5: Effect on the number of trades

2SLS, the whole sample				OLS, the whole sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.0394	0.0324	-1.21	Intercept	-0.0902	0.0127	-7.12
N_{t-1}	0.5685	0.0355	16.03	N_{t-1}	0.5121	0.0138	37.11
N_{t-2}	-0.1520	0.0199	-7.66	N_{t-2}	-0.1283	0.0137	-9.34
N_{t-3}	0.0838	0.0167	5.01	N_{t-3}	0.0699	0.0138	5.08
N_{t-4}	0.0160	0.0146	1.1	N_{t-4}	0.0162	0.0138	1.18
N_{t-5}	-0.0134	0.0151	-0.88	N_{t-5}	-0.0056	0.0136	-0.41
N_{t-6}	0.3688	0.0248	14.84	N_{t-6}	0.3309	0.0117	28.39
$ \hat{I}_t $	0.2391	0.2143	1.12	$ \hat{I}_t $	0.2174	0.0557	3.9
$ \hat{I}_{t-1} $	-0.2664	0.1083	-2.46	$ \hat{I}_{t-1} $	-0.2354	0.0582	-4.05
$ \hat{I}_{t-2} $	0.1181	0.0836	1.41	$ \hat{I}_{t-2} $	0.1008	0.0559	1.8
$ mo_t $	1.7591	1.7497	1.01	$ mo_t $	4.7887	0.1292	37.07
$ mo_{t-1} $	-0.3789	0.2434	-1.56	$ mo_{t-1} $	-0.6991	0.1466	-4.77
$ \Delta S_t $	0.0018	0.0004	4.71	$ \Delta S_t $	0.0013	0.0001	11.04
$ \Delta S_{t-1} $	-0.0006	0.0002	-2.9	$ \Delta S_{t-1} $	-0.0003	0.0001	-2.73
R-Square	0.7037	Adj R-Sq	0.7028	R-Square	0.7515	Adj R-Sq	0.7508
Test for Overidentifying Restrictions				p value = 0.6840			
2SLS, first part of the sample				OLS, first part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	0.0989	0.0951	1.04	Intercept	0.0450	0.0369	1.22
N_{t-1}	0.4929	0.0563	8.75	N_{t-1}	0.4637	0.0201	23.08
N_{t-2}	-0.1539	0.0339	-4.53	N_{t-2}	-0.1386	0.0192	-7.24
N_{t-3}	0.0562	0.0220	2.55	N_{t-3}	0.0509	0.0191	2.66
N_{t-4}	-0.0085	0.0203	-0.42	N_{t-4}	-0.0055	0.0191	-0.29
N_{t-5}	-0.0198	0.0213	-0.93	N_{t-5}	-0.0141	0.0189	-0.74
N_{t-6}	0.3143	0.0320	9.83	N_{t-6}	0.2991	0.0166	18.05
$ \hat{I}_t $	0.3957	0.3749	1.06	$ \hat{I}_t $	0.3354	0.0981	3.42
$ \hat{I}_{t-1} $	-0.2642	0.1266	-2.09	$ \hat{I}_{t-1} $	-0.2407	0.0979	-2.46
$ \hat{I}_{t-2} $	0.1406	0.1240	1.13	$ \hat{I}_{t-2} $	0.1317	0.0980	1.34
$ mo_t $	3.6901	2.8680	1.29	$ mo_t $	5.2363	0.1949	26.87
$ mo_{t-1} $	-0.3725	0.3102	-1.2	$ mo_{t-1} $	-0.4745	0.2234	-2.12
$ \Delta S_t $	0.0024	0.0008	2.83	$ \Delta S_t $	0.0021	0.0002	9.63
$ \Delta S_{t-1} $	-0.0007	0.0005	-1.48	$ \Delta S_{t-1} $	-0.0005	0.0002	-2.23
R-Square	0.5001	Adj R-Sq	0.4971	R-Square	0.5908	Adj R-Sq	0.5884
Test for Overidentifying Restrictions				p value = 0.9957			
2SLS, second part of the sample				OLS, second part of the sample			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	0.0075	0.0038	1.96	Intercept	0.0082	0.0032	2.54
N_{t-1}	0.3514	0.0241	14.56	N_{t-1}	0.3531	0.0208	16.97
N_{t-2}	0.1125	0.0244	4.6	N_{t-2}	0.1153	0.0199	5.8
N_{t-3}	0.0870	0.0200	4.34	N_{t-3}	0.0880	0.0199	4.42
N_{t-4}	0.0757	0.0213	3.56	N_{t-4}	0.0776	0.0199	3.89
N_{t-5}	0.0239	0.0210	1.14	N_{t-5}	0.0257	0.0198	1.3
N_{t-6}	0.0509	0.0205	2.49	N_{t-6}	0.0529	0.0186	2.85
$ \hat{I}_t $	-0.0097	0.0600	-0.16	$ \hat{I}_t $	0.0226	0.0153	1.48
$ \hat{I}_{t-1} $	0.0078	0.0403	0.19	$ \hat{I}_{t-1} $	-0.0124	0.0178	-0.7
$ \hat{I}_{t-2} $	-0.0019	0.0208	-0.09	$ \hat{I}_{t-2} $	-0.0097	0.0153	-0.63
$ mo_t $	1.0769	0.4329	2.49	$ mo_t $	0.9907	0.0500	19.82
$ mo_{t-1} $	-0.1349	0.0946	-1.43	$ mo_{t-1} $	-0.1205	0.0539	-2.23
$ \Delta S_t $	0.0003	0.0001	6.18	$ \Delta S_t $	0.0003	0.0000	11.91
$ \Delta S_{t-1} $	-0.0001	0.0000	-1.62	$ \Delta S_{t-1} $	-0.0001	0.0000	-1.8
R-Square	0.55459	Adj R-Sq	0.55207	R-Square	0.5973	Adj R-Sq	0.595
Test for Overidentifying Restrictions				p value = 0.6101			

The model is $N_t = \beta_0 + \sum_{i=1}^6 \beta_{N_i} N_{t-i} + \sum_{i=0}^2 \beta_{I_i} |\hat{I}_{t-i}| + \sum_{i=0}^1 \beta_{mo_i} |mo_{t-i}| + \sum_{i=0}^1 \beta_{s_i} |\Delta s_{t-i}| + \epsilon_t$, where N_t is the number of trades, $|\hat{I}_{t-i}|$ is the indicator function of the intervention, and $|mo_t|$ is the absolute value of the market order flow. The 2SLS and OLS estimates are shown on the left and right hand side of the table. The results are based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the data (Nov. 2003 to Feb. 2013), respectively.

Table 2.6: Damping Channel

	OLS, the whole sample			OLS, first part of sample			OLS, second part of sample		
Variable	Parameter	S.E.	t Value	Parameter	S.E.	t Value	Parameter	S.E.	t Value
Intercept	-3.67	1.08	-3.4	-7.98	1.63	-4.9	0.64	1.44	0.45
ΔS_{t-1}	-0.11	0.01	-8.29	-0.13	0.02	-6.69	-0.10	0.02	-5
ΔS_{t-2}	-0.01	0.01	-0.87	-0.03	0.02	-1.53	0.01	0.02	0.44
I_t	5.41	0.99	5.49	5.57	1.68	3.31	4.19	1.43	2.92
mo_t	394.18	15.45	25.52	396.32	17.79	22.28	455.82	37.70	12.09
$mo_t * \hat{I}_t$	-167.46	92.20	-1.82	-185.03	99.77	-1.85	196.44	358.84	0.55
R-Square	0.1408			0.1969			0.078		
Adj R-Sq	0.1399			0.1951			0.076		

The model is $\Delta s_t = \beta_0 + \beta_{s,1}\Delta s_{t-1} + \beta_{s,2}\Delta s_{t-2} + \beta_I I_t + \beta_{mo} mo_t + \beta_* mo_t * \hat{I}_t + \epsilon_t$, where Δs_t is the change in exchange rate, I_t is the intervention, and mo_t is the market order flow, and $mo_t * \hat{I}_t$ is the interaction term between the market order flow and indicator function of the intervention. The value of \hat{I}_t is 1 if the central bank intervenes, otherwise it is zero. The OLS estimates are based on the whole sample, the first part (Jan. 1996 to Dec. 2002), and the second part of the data (Nov. 2003 to Feb. 2013), respectively.

Table 2.7: Price impact at different time scales

Original data				scale 3			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-3.7488	1.0773	-3.48	Intercept	0.0042	0.2961	0.01
Δs_{t-1}	-0.1130	0.0138	-8.21	$\Delta s_{t-1}[d_3]$	0.5695	0.0115	49.49
I_t	4.8442	0.9346	5.18	$I_t[d_3]$	5.6790	0.7596	7.48
mo_t	390.8935	15.3399	25.48	$mo_t[d_3]$	203.9709	11.8670	17.19
R-Square	0.1403			R-Square	0.4285		
Adj R-Sq	0.1398			Adj R-Sq	0.4281		
scale 1				scale 4			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.0213	0.6839	-0.03	Intercept	0.0011	0.1470	0.01
$\Delta s_{t-1}[d_1]$	-0.4778	0.0121	-39.5	$\Delta s_{t-1}[d_4]$	0.7627	0.0088	86.4
$I_t[d_1]$	5.2145	0.9243	5.64	$I_t[d_4]$	0.6100	0.4157	1.47
$mo_t[d_1]$	337.0337	15.3557	21.95	$mo_t[d_4]$	149.7648	8.6533	17.31
R-Square	0.3537			R-Square	0.6845		
Adj R-Sq	0.3533			Adj R-Sq	0.6843		
scale 2				scale 5			
Variable	Parameter	S.E.	t Value	Variable	Parameter	S.E.	t Value
Intercept	-0.0054	0.4955	-0.01	Intercept	-0.0024	0.0833	-0.03
$\Delta s_{t-1}[d_2]$	0.2042	0.0136	15.04	$\Delta s_{t-1}[d_5]$	0.8635	0.0067	128.24
$I_t[d_2]$	4.7500	0.9058	5.24	$I_t[d_5]$	-0.3644	0.3512	-1.04
$mo_t[d_2]$	367.8855	15.4285	23.84	$mo_t[d_5]$	83.7317	5.5334	15.13
R-Square	0.1675			R-Square	0.8328		
Adj R-Sq	0.1670			Adj R-Sq	0.8327		

The table reports the results of the model $\Delta s_t[d_j] = \beta_{0[d_j]} + \beta_{s[d_j]}\Delta s_{t-1}[d_j] + \beta_{I[d_j]}I_t[d_j] + \beta_{mo[d_j]}mo_t[d_j] + \epsilon_t$, $j = 1, 2, \dots, 5$, where $[d_j]$ labels the component of the variable at the scale j . For instance, the daily return Δs_t is decomposed into independent components at different scales, and $\Delta s_t[d_j]$ denotes the component of Δs_t at the time scale j . The loading of the component of the intervention at scale j , $\beta_{I[d_j]}$, measures the price impact of the intervention at scale j . Due to the fact that the intervention is exogenous with respect to Δs_t , which is verified empirically, the model is estimated using OLS.

Table 2.8: Price impact and volatility regimes

The whole sample			First part of the sample			Second part of the sample		
	state 1	state 2		state 1	state 2		state 1	state 2
Intercept	-2.1398	-3.0127	Intercept	-5.7689	-25.9446	Intercept	1.4614	-2.9925
p value	0.02	0.36	p value	0.00	0.08	p value	0.22	0.61
ΔS_{t-1}	-0.1449	-0.0658	ΔS_{t-1}	-0.2002	0.0622	ΔS_{t-1}	-0.0744	-0.1486
p value	0.00	0.19	p value	0.00	0.43	p value	0.00	0.02
mo_t	370.9045	604.1050	mo_t	377.2795	574.8658	mo_t	442.0392	559.5118
p value	0.00	0.00	p value	0.00	0.00	p value	0.00	0.04
I_t	3.3407	12.8133	I_t	2.1598	19.5289	I_t	3.7885	14.2441
p value	0.00	0.03	p value	0.21	0.10	p value	0.00	0.08
State duration	39.46	4.01	State duration	42.59	3.60	State duration	73.80	9.56
P_{j1}	0.97	0.25	P_{j1}	0.98	0.28	P_{j1}	0.99	0.10
P_{j2}	0.03	0.75	P_{j2}	0.02	0.72	P_{j2}	0.01	0.90
p value	0.00	0.00	p value	0.00	0.03	p value	0.00	0.00

The table reports the results of the two-state Markov switching model $\Delta S_t = \beta_{0,s} + \beta_{S,s}\Delta S_{t-1} + \beta_{mo,s}mo_t + \beta_{I,s}I_t + \epsilon_{t,s}$, $\epsilon_{t,s} \sim N(0, \sigma_s^2)$, for $s=1$ or 2 , where subscript s labels the state of the system. The coefficients $\beta_{I,1}$ and $\beta_{I,2}$ measure the price impact of the intervention at state 1 and 2, respectively. The transition probability P_{ij} is the probability of transitioning from state i to state j in a single step. The state duration is the expected number of periods (days) the system stays in the particular state before it transits to the other state. Based on the smoothed probability reported in figure 2.4, state 2 correspond to the high volatility regime, thus state 1 corresponds to the low volatility regime.

Table 2.9: Price impact of the order flow and the volatility regimes

		$\beta_{0,s}$	$SE_{\beta_{0,s}}$	$t_{\beta_{0,s}}$	$\beta_{mo,s}$	$SE_{\beta_{mo,s}}$	$t_{\beta_{mo,s}}$	StateDuration	TrP p_{i1}	TrP p_{i2}	P_{TrP}
AUD/USD	state 1	-1.6660	1.0531	-1.5820	122.1298	3.1676	38.5565	72.5030	0.9862	0.0138	0
AUD/USD	state 2	-46.7278	8.7481	-5.3415	229.7329	22.1173	10.3870	13.0939	0.0764	0.9236	0
CAD/USD	state 1	-10.5622	0.9518	-11.0973	115.9651	3.7159	31.2078	87.0696	0.9885	0.0115	0
CAD/USD	state 2	-13.1039	2.8739	-4.5597	174.6527	11.3517	15.3856	44.5347	0.0225	0.9775	0
EUR/USD	state 1	-7.9222	1.0954	-7.2319	296.9106	9.1175	32.5650	447.7359	0.9978	0.0022	0
EUR/USD	state 2	1.8583	2.3201	0.8010	428.9224	29.3408	14.6186	152.9111	0.0065	0.9935	0
EUR/GBP	state 1	-5.7320	0.8500	-6.7436	91.3033	4.4829	20.3672	53.7676	0.9814	0.0186	0
EUR/GBP	state 2	-0.7237	1.6853	-0.4294	118.4898	14.0854	8.4123	18.8641	0.0530	0.9470	0
EUR/NOK	state 1	-2.6175	0.7251	-3.6098	243.8860	8.7705	27.8076	89.2448	0.9888	0.0112	0
EUR/NOK	state 2	-3.0136	3.4423	-0.8755	474.7056	51.0166	9.3049	16.8228	0.0594	0.9406	0
EUR/SEK	state 1	-4.0759	0.7025	-5.8019	201.4736	9.1024	22.1341	57.9345	0.9827	0.0173	0
EUR/SEK	state 2	-9.7351	3.1857	-3.0559	365.0681	34.5563	10.5644	18.1854	0.0550	0.9450	0
GBP/USD	state 1	-5.1792	0.8517	-6.0808	118.1860	3.1581	37.4229	563.1606	0.9982	0.0018	0
GBP/USD	state 2	-22.4445	5.5630	-4.0346	189.9040	19.3740	9.8020	84.5461	0.0118	0.9882	0
HKD/USD	state 1	-0.0065	0.0276	-0.2367	5.3025	0.6737	7.8708	13.9152	0.0719	0.9281	0
HKD/USD	state 2	-0.5115	0.1128	-4.5339	35.6890	1.7851	19.9926	16.8474	0.9406	0.0594	0
JPY/USD	state 1	1.8369	1.2109	1.5169	428.1934	36.5310	11.7214	61.7060	0.9838	0.0162	0
JPY/USD	state 2	-4.4538	6.3296	-0.7036	679.6838	257.8614	2.6358	8.2129	0.1218	0.8782	0
MXN/USD	state 1	-11.5018	1.1229	-10.2426	151.3133	7.2685	20.8178	56.4272	0.9823	0.0177	0
MXN/USD	state 2	11.4457	5.9978	1.9083	181.8231	36.7518	4.9473	16.1884	0.0618	0.9382	0
NZD/USD	state 1	-1.9448	1.3146	-1.4794	284.5357	11.0595	25.7278	104.1812	0.9904	0.0096	0
NZD/USD	state 2	-22.4289	5.0719	-4.4222	520.4838	41.0740	12.6718	36.0628	0.0277	0.9723	0
SGD/USD	state 1	-6.0628	0.5744	-10.5544	159.2413	7.6867	20.7164	47.7526	0.9791	0.0209	0
SGD/USD	state 2	-7.6523	2.1566	-3.5483	234.9747	19.6622	11.9506	17.0801	0.0585	0.9415	0
ZAR/USD	state 1	-26.7563	2.0305	-13.1769	430.4965	17.1726	25.0688	64.1344	0.9844	0.0156	0
ZAR/USD	state 2	-35.1039	9.3424	-3.7575	775.5638	71.2427	10.8862	17.8321	0.0561	0.9439	0

The table shows that the price impact of the order flow is high in the high volatility regime and vice versa, based on the results of 13 different currency pairs. The two-state Markov switching model is as follows: $\Delta S_t = \beta_{0,s} + \beta_{mo,s} mo_t + \epsilon_{t,s}$, $\epsilon_{t,s} \sim N(0, \sigma_s^2)$, for $s = 1, 2$, where mo_t is the market order flow, ΔS_t is the difference of logged price. The loadings of the market order flow in different states, $\beta_{mo,1}$ and $\beta_{mo,2}$, capture the asymmetry of the price impact in different states. The smoothed state probability identifies that State 2 corresponds to the high volatility state, while state 1 is the low volatility state. The last three columns contain the transition probabilities p_{i1} , p_{i1} , and the p value of the transition probability, respectively.

Chapter 3

Calibrating the Predictability of Exchange Rates: A Model-Independent Approach

3.1 Introduction

The main purpose of this paper is to propose a new methodology to assess the out-of-sample predictability of financial time series. Traditionally, the predictability is studied based on model-dependent method: the researcher selects the forecasting model, carries out the econometric exercise, then compares the forecasting performance with that of the benchmark, such as random walk, according to the chosen evaluation method, for instance, mean squared forecast error. The disadvantage of the model-dependent approach is that one cannot differentiate whether the failure is due to misspecification of the forecasting model or lacking information in the predictors. If it is the former, different economic or econometric models might improve the forecasting performance. On the other hand, if the predictors contain no information about the future price movement, it is impossible to predict based on these predictors, regardless of the specification of the forecasting model. In this paper, we use an information-theoretical quantity called mutual information to calibrate the information transfer from the predictors to the future return. Mutual information measures all statistical dependence, linear and nonlinear, and not

just the linear dependence as the correlation coefficient measures.

Measuring the information transfer (in other words, the true statistical dependence) will help us to distinguish lack of information from model misspecification. Pinpointing the underlying cause is very important for time series predictability problem, since the lack of information can only be solved by searching for better predictors, rather than trying different economic and econometric models. Failure to find any predictor with statistically significant information transfer suggests that it is impossible to predict at all. This is precisely the issue with the model-dependent exchange rate forecasting literature: researchers keep trying different models and methods based on a limited set of predictors, without asking whether the failure in forecasting exchange rates is due to model misspecification or lack of information.

Generally speaking, the linear measures of statistical dependence such as auto-correlation and cross-correlation are not unbiased measures of the statistical dependence. They are unbiased under the circumstance where the variables are multivariate Gaussian distributed and their interaction is linear. So correlation is not very useful when it comes to quantifying the predictability of time series. However, based on Shannon's channel coding theorem, we can establish the fact that the correct measure of the statistical dependence should be an information theoretical quantity called mutual information. It can be used to calibrate the information transfer from the predictors to the future return, without imposing any *ex ante* assumption on the distribution of the variables and their linear or nonlinear relationship. Moreover, due to the fact that the information transfer is the upper bound of the predictive power of any model based on given predictors, the statistical significance of mutual information is the necessary condition for the time series predictability, as we will elaborate later.

In this paper we apply the nonparametric model-independent method to answer one important question: are exchange rates predictable? This question has puzzled the researchers for several decades. There has been a huge number of model-dependent studies devoted to predicting the exchange rates, using different economic models and econometric methods. However, since Meese and Rogoff (1983a,b, 1988), it is well known that exchange rates are very difficult to predict. The benchmark random walk model generally outperforms the economic models in the out-of-sample forecasting exercises, which is called "the Meese and Rogoff puzzle". Rossi (2013) provides an excellent review of the related literature. She finds that the performance

usually depends on the choice of predictor, sample period, the specification of the model, and the forecast evaluation method. Overall speaking, there is no forecasting model can consistently predict the exchange rates.

As we have mentioned, the shortcoming of the traditional model-dependent methodology is that it cannot differentiate lack of information from model misspecification. Failure of the previous models and predictors does not rule out the possibility of future success. In principle, there are a countless number of model specifications based on various predictors. Testing the performance of these model specifications one by one is time-consuming. There is still no conclusive answer after more than three decades of model-dependent researches. Therefore model-dependent approach is not very effective in determining the predictability of exchange rates.

In this paper, we approach the predictability problem from a different perspective. The goal is to provide a model-independent criterion for calibrating the predictability of time series by drawing an analogy between predictability and Shannon's channel coding theorem (Shannon 1948, Shannon and Weaver 1949). The logic is straightforward: in order to be able to predict, the predictors have to carry information about the future return of the exchange rate. In the language of statistics, the predictors and the future price movement must be statistical dependent. This necessity is independent of the specific economic model or econometric methodology adopted by the researcher. On the other hand, for a given set of predictors, the predictive power of any model must be bounded by an upper limit, which is the information transfer from the predictors to the future return.

Why is mutual information the correct measure of information transfer or statistical dependence? In information theory, mutual information specifies a noisy channel's reliable information transmission capacity, beyond which error-free communication is impossible. In other words, mutual information specifies the upper bound of the information transmission rate of the noise communication channel. This is the great insight of Shannon's channel coding theorem. Essentially, a noisy communication channel is two statistical dependent random variables. Mutual information is the capacity of information transfer between these two variables. Therefore, mutual information is the right measure that can calibrate the information transfer from the predictors to the future return.

The last thing needed to clarify is what this study is not about, and how our approach is different from the traditional model-dependent methods in the literature. Very different from the model-dependent forecasting method, this study aims to provide a measure for the upper bound of the exchange rate predictability via measure the true statistical dependence between the predictors and future return. However, we cannot stress enough, this study does not actually carry out the out-of-sample forecasting exercise in the traditional model-dependent sense, which is to find the model, estimate the model parameters using a sample of rolling window, then make one step forward forecasts. This study we don't carry out any model-dependent forecasting exercise as such, since the whole purpose of this study is to propose a non-parametric model-independent method to calibrate the predictability. Our method calibrates the time series predictability, but it does not specify the analytical form of the optimal forecasting model. Therefore, we do not have traditional measures of forecasting performance, such as RMSE or OOS- R^2 .

We investigate the exchange rate predictability at various horizons: hourly, daily and monthly. The exchange rates in our dataset are the most frequently traded ones in the FX markets. At the hourly, daily and monthly frequency, there are 13, 15 and 10 currency pairs, respectively. With both economic and market microstructure literature in mind, the variables used to predict are limit and market order flow and the available monetary fundamentals: differentials of interest rate, money supply, inflation, and industrial production. From the market microstructure point of view, order flow is the most important determinant of exchange rate movements, since it aggregates the information among the market participants (Evans and Lyons 2002, Rime, Sarno, and Sojli 2010, Kozhan, Moore and Payne 2014). From the macroeconomic perspective, differentials of interest rate, money supply, inflation, and industrial production are the fundamentals usually used in economic models (Rossi 2013). The availability of the data is frequency-dependent: order flow is available at the hourly and daily frequency; interest rate is the only fundamental documented at the daily frequency; at the monthly frequency, the following macroeconomic fundamentals are available: interest rate, money supply, inflation, and industrial production (output).

Empirically, we find that the exchange rates are systematically predictable at the hourly frequency. The intraday predictive power is mainly from the interdependence of the exchange rate returns at the hourly frequency, whereas order flow has very small predictive power for

the future return. One important point is that the linear model that are frequently used in the literature fails to capture most of the information transfer, which explains why it is so hard for the linear model to outperform the random walk benchmark, i.e. the Meese and Rogoff puzzle. It also suggests that the optimal forecasting model based on historical returns of the exchange rate must be nonlinear ¹. At daily and monthly frequency, the exchange rates are not systematically predictable based on historical returns, order flow, and macroeconomic fundamentals. Our study finds that for more than half of the currency pairs, the factors have small but significant information transfer, but for all the exchange rates, the linear model does not have any significant predictive power. Furthermore, unlike the hourly frequency, the historical movements of exchange rates do not have any predictive power for the future return at the daily and monthly frequency.

This study has several contributions to the literature. First of all, as far as we know, this is the first study to use the model-independent methodology to calibrate the time series predictability. we apply information theory to study predictability, and shed some light on the Meese and Rogoff puzzle: why the exchange rates are so difficult to predict by traditional linear models. Second, given the set of predictors, the model-independent non-parametric method proposed in this paper can be used to assess the predictability of time series, and estimate the upper bound of the predictive power of any forecasting model. Moreover, it can be used for predictor selection: one should adopt the predictors with significant information transfer from the predictors to the future return. Third, it also provides some guidance for the model specification. For instance, whether the model is optimal or suboptimal, and if it is suboptimal, how good is its performance compared to the optimal forecasting model. Our method can be useful for model selection as well: one can rank the performance of different models based on the information transfer from the model to the future return. Furthermore, once the functional form of the model is chosen, it can be useful for parameter estimation: the parameters should be chosen such that the information transfer from the model to the future return is maximized. Last but not least, the method is applicable for other predictability questions as well. For instance, the predictability of stock returns or any other economic variable of interest.

¹The nonlinear forecasting model $f(x)$ predicts ΔS out-of-sample.

3.2 Literature Review

There are a large number of studies on the model-dependent predictability of the exchange rates. Please refer to Rossi (2013), Frankel and Rose (1995), Mark (1995), Engel, Mark and West (2007) for the literature reviews on this topic. In the early 1980s, the seminal papers by Meese and Rogoff (1983a,b) first examine the out-of-sample forecasting performance of three exchange rate models. They find that economic models do not outperform the random walk benchmark, which essentially suggests that the exchange rates are not predictable. Cheung, Chinn and Pascual (2005) provide a comprehensive examination on the out-of-sample forecasting performance of various economic models. They confirm that these models cannot consistently outperform the random walk model. Other researchers also conduct the empirical exercises based on traditional predictors, such as interest rate differentials (Alquist and Chinn 2008, Clark and West 2006, and Molodtsova and Papell 2009), price and inflation differentials (Rogoff 1996), money and output differentials (Chinn and Meese 1995, Taylor and Sarno 1998), productivity differentials (Wright 2008), portfolio balance (Cheung, Chinn and Pascual 2005). In some studies, certain economic models occasionally outperform the random walk benchmark for some countries and certain time periods, but generally speaking, the exchange rates are not predictable systematically. More recently, studies such as Molodtsova and Papell (2009), Giacomini and Rossi (2010), Inoue and Rossi (2012), find empirical evidence that Taylor rule fundamentals have better forecasting performance than the traditional predictors. Rossi (2013) concludes that none of the predictors, models, or tests systematically find the empirical support of superior exchange rate forecasting ability across all countries and time periods.

As regards to the market microstructure approach to the exchange rates, Evans and Lyons (2002) introduce a portfolio shifts model, which argues that changes in exchange rates are determined by a combination of innovations in public and private information. Private information is made public through the process of trading. In other words, private information is revealed by order flow, which is measured as the net of buyer- over seller-initiated trades in the foreign exchange market. Evans and Lyons (2002) show that order flow is a critical determinant of two major exchange rates (mark/dollar and yen/dollar). The empirical finding is that R^2 of the linear regression increases from 1-5% to 40-60% after adding market order flow as a regressor.

Further research shows that order flow plays an intermediary role between exchange rates and macroeconomic fundamentals (Rime, Sarno and Sojli, 2010). Order flow does not only contain private information, but it is also a vehicle for aggregating macroeconomic information among all the market participants. The reason that order flow has more explanatory power than macroeconomic fundamentals is due to the fact that the conventional measures of expected future fundamentals are so imprecise that an order flow “proxy” performs much better than macroeconomic fundamentals in explaining the exchange rates. Rime et al. (2010) also show that order flow is a predictor of daily exchange rates in an out-of-sample exercise. The forecast analysis is based on economic criteria, where an investor can earn profits from an asset allocation strategy that exploits this predictability. Evans and Lyons (2005, 2006) argue that gradual learning in the foreign exchange market can generate not only explanatory, but also forecasting power in order flow. More recently, Kozhan, Moore and Payne (2014) show that besides the market order flow, the limit order flow is also an important determinant of the exchange rates.

With respect to the literature on the applications of information theory. Schreiber (2000) uses Transfer Entropy to study the information transfer between two time series. Barnett, Barnett and Seth (2009) show that Granger causality and Transfer Entropy are equivalent for Gaussian variables. Information theoretical measures such as transfer entropy have been used to study complex system and structure of the network, especially in the field of neuroscience (for instance, Vicente, Wibral, Lindner, and Pipa, 2011, Wibral, Vicente, and Lizier, 2014). However, these inter-discipline studies just use the mutual information as a measure for information transfer. None of them use information theory to discuss the issue of time series predictability, nor any of them gives a solid rationale for the fact that mutual information is the right measure of the statistical dependence. Answering these questions will give us the dominating reasons that information theory is relevant for time series predictability.

3.3 Methodology

The setting of out-of-sample exchange rate forecasting is as follows. There are time series of the exchange rate returns, $\Delta \mathbf{S}$, and the factors \mathbf{X} , which comprises of the variables adopted by the forecasting model, such as order flow and monetary fundamentals. At each point of

time t , to forecast the return in the next period, ΔS_{t+1} , we can use the available information of the factors \mathbf{X} and the exchange rate movements $\Delta \mathbf{S}$ up to and including time t : $\mathbf{X}_t^{(k)} = \{\mathbf{X}_{t-k+1}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t\}$ and $\Delta \mathbf{S}_t^{(k)} = \{\Delta S_{t-k+1}, \dots, \Delta S_{t-1}, \Delta S_t\}$. Here only a finite number of lagged observations are included, because it is reasonable to assume that the observations prior to time $t - k + 1$ are irrelevant to the exchange rate return ΔS_{t+1} . It is not a restrictive assumption, since the number of the lagged periods k can be adjusted with ease.

If the future return, ΔS_{t+1} , is predictable, there should be some statistical dependence between the predictors $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ and the future return ΔS_{t+1} . More precisely, $\mathbf{X}_t^{(k)}$ and $\Delta \mathbf{S}_t^{(k)}$ should contain some information about the future return ΔS_{t+1} . Empirically, the estimated information transfer from $\mathbf{X}_t^{(k)}$ and $\Delta \mathbf{S}_t^{(k)}$ to ΔS_{t+1} should be positive and significant from zero. On the other hand, if $\mathbf{X}_t^{(k)}$ and $\Delta \mathbf{S}_t^{(k)}$ contain no information about the future return ΔS_{t+1} , forecasting the future movements of the exchange rate using $\mathbf{X}_t^{(k)}$ and $\Delta \mathbf{S}_t^{(k)}$ would be an impossible mission, because $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ and ΔS_{t+1} are complete statistically independent. As we are going to argue in the next section, the correct measure of information transfer from $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ to ΔS_{t+1} is their mutual information, which captures the linear and nonlinear statistical dependence between the predictors and the future return. It is denoted as $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$, where the semicolon separates the predictors and the future price movement.

Moreover, mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is the upper bound on the predictive power of any model based on $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$:

$$I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1}) \geq I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1}), \quad (3.1)$$

where $f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$ is the forecasting model. This is the data processing inequality of information theory. Intuitively speaking, it is impossible to increase the information carried by a given set of predictors about the future return via any operations or processing. In other words, the information content of a signal cannot be increased via any local operation. Therefore, the necessary condition for the predictability is that mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is positive and statistically different from zero. If $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is positive and significant, there should exist an optimal model $f^*(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$ that can fully take advantage of the pre-

dictive power of $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ for the future return ΔS_{t+1} . On the other hand, if the mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is zero, any forecasting model $f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$ should satisfy $I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1}) = 0$, due to the inequality (3.1). Therefore, any endeavor based on predictors $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ would be fruitless. Given the set of predictors, this necessary condition for predictability is independent of the specification of the forecasting model. However, in the case that the predictability has been established, our methodology does not directly show us how to predict. In other words, in this paper we focus on the “existence” of the optimal model $f^*(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$, but not its specific form.

Based on the chain-rule for mutual information, the information transfer from $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ to ΔS_{t+1} can be further decomposed into two parts:

$$I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1}) = I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1}) + I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)}). \quad (3.2)$$

The first part measures the auto-predictive power of the exchange rate. To be more specific, it is the predictive power of the historical movements $\Delta \mathbf{S}_t^{(k)}$ for the future return ΔS_{t+1} . For convenience, we refer it as Auto-Information-Transfer (AIT). The other part is the predictive power of the factors $\mathbf{X}_t^{(k)}$ for the future movement ΔS_{t+1} , which is called Transfer Entropy (TE).

Equation 3.2 can be viewed as follows: the total information about future return contained in two predictors $\Delta \mathbf{S}_t^{(k)}$ and $\mathbf{X}_t^{(k)}$ are consisted of two parts: the first part is the information contained in the historical price movement $\Delta \mathbf{S}_t^{(k)}$; and the second part is the information contained in predictor $\mathbf{X}_t^{(k)}$, given $\Delta \mathbf{S}_t^{(k)}$ is known. This decomposition is related to the chain rule of the probability $P(X_2, X_1 | Z) = P(X_2 | X_1, Z) \cdot P(X_1 | Z)$. We will postpone the details to next section.

This decomposition provides us with the opportunity to investigate these two sources of predictive power separately. We will estimate AIT and TE empirically, and test their statistical significance. The statistical significance of TE (or AIT) means that there is positive information transfer from $\mathbf{X}_t^{(k)}$ (or $\Delta \mathbf{S}_t^{(k)}$) to the future return ΔS_{t+1} . If it is the case, it is possible to predict the exchange rate using certain models based on $\mathbf{X}_t^{(k)}$ (or $\Delta \mathbf{S}_t^{(k)}$). On the other hand, if there is no statistically significant information transfer (AIT and TE), then it is absolutely

impossible to predict future exchange rate based on the given predictors.

Before investigating the predictability empirically, we are going to study the contemporaneous information transfer, or equivalently, the explanatory power of factors \mathbf{X} for the contemporaneous exchange rate movement. The purpose of studying the contemporaneous case before the predictive case is two folds. First, as a sanity check for this new methodology, we empirically establish the equivalence between Contemporaneous Transfer Entropy (CTE) $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ under the linear-Gaussian assumption, and R^2 from the linear regression $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$. The equivalence is proved analytically by Barnett, Barnett and Seth (2009). For all the exchange rates, we empirically establish the equivalence between these two measures: R^2 and the Gaussian estimate of CTE. Moreover, the nonparametric estimates of CTE are consistent with the known empirical facts as well, for instance, market order flow has good explanatory power for the contemporaneous return. Second, more importantly, the numerical values of Contemporaneous Transfer Entropy (CTE) provide us with a frame of reference to interpret the magnitude of information transfer. The issue is that after estimating the number of nats or bits of the information transfer, we still don't have any sense of the scale of large or small information transfer. For instance, how large is 0.01 nats, and what does it mean? One straightforward method is to compare the information transfer across time with its contemporaneous counterparts: the value of CTE. To be more specific, the results from the contemporaneous case indicate that market order flow (*mo*) has very good explanatory power, while interest rate differential (*IR*) has very small explanatory power for the contemporaneous exchange rate movements. Therefore the values of CTE can be used as benchmarks to assess the magnitudes of information transfer as well as their economic significance. Please note that we are not aiming to compare the information transfer across time with the contemporaneous information transfer. It will not be a fair competition because the contemporaneous information transfer would be much greater, since there are variables has good contemporaneous explanatory power, such as the order flow (Evans and Lyons 2002). We merely use the values of the contemporaneous information transfer as a frame of reference, which helps us to make sense out of the units in nats.

In this paper, the measures of information transfer are empirically estimated with two different methods: multivariate Gaussian model and the Kraskov, Stogbauer, and Grassberger (2004) (KSG) method. The first method imposes the linear-Gaussian model *ex ante*, which assumes

that the variables are multivariate Gaussian distributed, and their relationship is linear. Estimating information measure boils down to estimating the sample covariance matrix. Naturally, this method is equivalent to the frequent-used linear regression, which is also under the same linear-Gaussian assumption. The shortcoming of this method is that if the true underlying model (or data generating process) governing the time series are not linear-Gaussian, the estimates subjected to linear-Gaussian assumption would be biased. KSG method improves the method of Kozachenko and Leonenko (1987) based on kernel estimation. It uses a dynamically altered kernel width to adjust to the density of samples in the neighborhood of any given observation, in order to correct the bias in the probability density function estimation, especially when the sample is finite. KSG method does not impose the Gaussian or any other parametric model *ex ante*. In comparison to the Gaussian estimator, it can accommodate non-Gaussian distribution and the non-linear relationships among the variables. In other words, this method is nonparametric and model-independent. Vicente and Wibral (2014) provide an updated survey of various types of estimators.

Obviously, once we estimate the information transfer, their statistical significance also needs to be tested. If the null hypothesis that TE $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ (or AIT $I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$) equals to zero cannot be rejected, factors \mathbf{X} (or the historical movements of exchange rate) do not have any predictive power for the future exchange rate return. To test the statistical significance of the estimate, the simulated samples under the null hypothesis is computed. The null is that there is no information transfer. By comparing the original estimate and the estimates based on simulated samples under the null, we can find if the original estimate is statistically significant. We refer to it as resampling method. It is in nature the same procedure as the bootstrap method.

Here we briefly explain the resampling method for testing the statistical significance of conditional mutual information $I(Y; Z|W)$ or mutual information $I(Y; Z)$. Let's assume we want to test the significance of conditional mutual information $I(Y; Z|W)$. First, we generate random permutations of the original sample $\{y_t, t = 1, 2, \dots, T\}$ by randomly shuffling the sequential order of the original sample while keeping all the observations. These simulated samples are denoted as $\{y_t^s, t = 1, 2, \dots, T\}$, where $s = 1, 2, \dots, S$, and the total number of the simulated samples can be set as $S = 1000$, for example. Next, for each simulated sample $\{y_t^s, t = 1, 2, \dots, T\}$,

conditional mutual information $I(Y^s; Z|W)$ can be estimated. In such way, we can generate a population of $I(Y^s; Z|W)$, $s = 1, 2, \dots, S$, under the null hypothesis that the conditional mutual information is zero. This is due to the fact that by doing random permutation of the original sample, the distribution of Y is kept: $p(y^s) = p(y)$, while any potential conditional dependence between Y and Z is destroyed: $p(z|y^s, w) = p(z|w)$, because the sequential correspondence between Y and Z has been destroyed by random shuffling. Equivalently, we have $I(Y^s; Z|W) = 0$. Hence $I(Y^s; Z|W)$ satisfies the null hypothesis that the conditional mutual information $I(Y; Z|W)$ is zero. Once a population of $I(Y^s; Z|W)$ has been estimated, we can obtain the p -value of $I(Y; Z|W)$ by comparing its value with the empirical distribution of $I(Y^s; Z|W)$. Alternatively, one can also calculate the standard deviation of $I(Y^s; Z|W)$, and determine whether conditional mutual information $I(Y; Z|W)$ is statistically significant based on the usual two standard deviation rule of thumb. For simplicity, we only report the p -values in the later section. Similarly, the statistical significance of mutual information $I(Y; Z)$ can also be tested.

3.4 Information Theory and the Predictability of Time Series

In this section, we are going to lay the theoretical foundation for our methodology. It can be skipped for the readers without the interest in going into the details of information theory. The measure of information transfer in this study is based on information theory. It is a vast area, due to the limited space, we only discuss the most important points relevant to our study. Instead of being mathematically rigorous, the discussion here just tries to be straightforward and intuitive. For a full-fledged treatment, please refer to the monographs on information theory, such as Gray (2011), Cover and Thomas (2006).

The most fundamental concept of information theory is Shannon's Entropy, which is a measure of information content. Assume there is a random variable Y , with probability density function $p(y)$. The measure of the information content of a particular event $Y = y$ is defined as

$$h(y) = \log \frac{1}{p(y)}. \quad (3.3)$$

The logic behinds this is that if the probability of event y is very large, for example, close to 1, which means that it is very likely to happen, the knowledge that event $Y = y$ occurred provides very little information; in contrast, if something very improbable happened, it conveys a lot of information. It is a measure of information gained, or the level of “surprise”, in finding out this event.

The measure of information, such as $h(y)$ in equation 3.3, is in units of “bits” if logs in the equations are the logarithm to the base 2, or in units of “nats” if logs are base- e logarithm. In other words, if we have chosen base 2 for the information formula, one unit of information is called a bit; and if we have chosen base e for the information formula, one unit of information is called a nat.

The entropy of a random variable Y , is defined as the expected information content of all events with positive probability:

$$H(Y) = \int p(y)h(y) dy = \int p(y) \log \frac{1}{p(y)} dy = - \int p(y) \log p(y) dy. \quad (3.4)$$

Entropy $H(Y)$ measures the information contained in Y . It is the expected degree of uncertainty of the probability distribution $p(y)$. A related concept is conditional entropy. It quantifies the amount of information in one random variable (Z) given we already know the other (Y). The conditional entropy of Z conditional on Y is

$$H(Z|Y) = \int_Y p(y) H(Z|Y = y) dy = - \int_Y \int_Z p(z, y) \log p(z|y) dz dy. \quad (3.5)$$

Conditional entropy $H(Z|Y)$ is a measure of what Y does not say about Z , or the amount of uncertainty remaining about Z after Y is known. $H(Z|Y) = 0$ if and only if the value of Z is completely determined by the value of Y , in this case, knowing Y completely reduces the uncertainty or information content of Z to zero. Conversely, $H(Z|Y) = H(Z)$ if and only if Z and Y are independent random variables: knowing Y tells us nothing about Z .

The most important quantity for out-of-sample predictability is called mutual information (MI). The mutual information of two random variables is a measure of the mutual dependence between these two variables. More specifically, it quantifies the amount of information obtained about

one random variable (Z), through the other random variable (Y). MI is defined as

$$I(Y; Z) = \int_Y \int_Z p(z, y) \log \left(\frac{p(z, y)}{p(z)p(y)} \right) dz dy \quad (3.6)$$

$$= \int_Y \int_Z p(z, y) \log \left(\frac{p(z|y)}{p(z)} \right) dz dy \quad (3.7)$$

$$= H(Z) - H(Z|Y). \quad (3.8)$$

It can be shown that mutual information $I(Y; Z)$ is non-negative. If Y and Z are independent: $P(Y, Z) = P(Y)P(Z)$, we have $I(Y; Z) = 0$, which means that Y does not have any information about Z , or the information transfer from Y to Z should be zero. On the other hand, if the value of Z is completely determined by the value of Y , we have $H(Z|Y) = 0$, which means that $I(Y; Z) = H(Z)$. In this case, knowing variable Y provides all the information about the variable Z . Overall speaking, mutual information is a general measure to quantify the dependence between two variables. Moreover, mutual information precisely measures the information transfer from one variable to the other.

A related concept is conditional mutual information, $I(Y; Z|W)$, which is the expected value of the mutual information of two random variables, Y and Z , given the value of a third variable, W . It is defined as

$$I(Y; Z|W) = \mathbb{E}_W(I(Y; Z)|W) \quad (3.9)$$

$$= \int_W p(w) \int_Y \int_Z p(z, y|w) \log \frac{p(z, y|w)}{p(z|w)p(y|w)} dy dz dw \quad (3.10)$$

$$= \int_W \int_Y \int_Z p(w)p(z, y|w) \log \frac{p(z|y, w)}{p(z|w)} dy dz dw \quad (3.11)$$

$$= H(Z|W) - H(Z|Y, W). \quad (3.12)$$

It measures the reduction in the uncertainty of Z due to knowledge of Y when W is given. The last useful equality is the chain-rule for mutual information: the mutual information between

$\{Y, W\}$ and Z can be decomposed into two parts:

$$I(Y, W; Z) = H(Z) - H(Z|Y, W) \quad (3.13)$$

$$= H(Z) - H(Z|W) + H(Z|W) - H(Z|Y, W) \quad (3.14)$$

$$= I(W; Z) + I(Y; Z|W), \quad (3.15)$$

where $I(W; Z)$ is the mutual information between W and Z , and $I(Y; Z|W)$ is the conditional mutual information between Y and Z given W .

Information theory mainly comprises of two parts, source encoding and channel communication. The first one is about data compression while the later one is related to the topic of information transfer via a noisy communication channel. In this paper, we want to link the theory of channel communication and information transfer to the predictability of time series. For the details of the mechanism of information transfer through a noisy channel, please refer to the appendix. Here we just present the channel coding theorem, which is based on the Theorem 7.7.1 in Cover and Thomas (2006):

Channel Coding Theorem. For information channel Y to Z , we define the channel capacity as $C = \max_{p(y)} I(Y; Z)$. All the information transmission rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of codes of length n and transmission rate R with the maximum probability of error goes to zero as $n \rightarrow \infty$. Conversely, if any sequence of codes of length n and transmission rate R with the probability of error goes to zero as $n \rightarrow \infty$, the transmission rate R must satisfy $R \leq C$.

With respect to the predictability of time series, the relevant point in Shannon's channel coding theorem is that the information transfer from Y to Z is measured by mutual information $I(Y; Z)$. Statistically speaking, the channel coding theorem indicates that $I(Y; Z)$ measures the statistical dependence between Y and Z . In contrast to other frequently-used measures of linear statistical dependence, such as covariance, mutual information is *the* right measure of the 'dependence' or information transfer from random variable Y to Z . This is exactly due to the fact that Shannon has proved in the channel coding theorem: reliable information transfer is achievable for any information transmission rate lower than $I(Y; Z)$, while information transfer is impossible for any transmission rate greater than $I(Y; Z)$.

The previous literature usually gives a different argument based on the source encoding theorem to validate that mutual information measures the information transfer (for example, Schreiber 2000). However, the reasoning is not very convincing. The source encoding theorem is another important component of Shannon's information theory, but in our opinion, source encoding is about data compression and does not directly related to information transfer. On the other hand, noisy channel communication is all about information transfer from one variable (Y) to the other (Z). Hence we believe the channel coding theorem is the right part of information theory relevant to information transfer and time series predictability.

For the exchange rate to be predictable, there has to be positive information transfer from $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ to ΔS_{t+1} , where $\mathbf{X}_t^{(k)} = \{\mathbf{X}_{t-k+1}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t\}$ and $\Delta \mathbf{S}_t^{(k)} = \{\Delta S_{t-k+1}, \dots, \Delta S_{t-1}, \Delta S_t\}$. Furthermore, mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is actually the upper bound of the predictive power of any model, $f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$. It is based on another important result of information theory, the data processing inequality. Given the set of predictors $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$, the researcher can process these predictors with any localized operation, and use model f of any functional form, linear or nonlinear, the data processing inequality (Gray, 2011, Cover and Thomas, 2006) guarantees that $f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$ can only carry less or equal amount of information about the future exchange rate ΔS_{t+1} :

$$I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1}) \leq I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1}). \quad (3.16)$$

In other words, one can not create more information based on the information he already has. The mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is the upper bound of the predictive power of any model $f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$. Therefore, if the exchange rate is predictable by predictors $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$, mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ has to be positive and significantly different from zero. As to the forecasting performance of the model $f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)})$, we should expect that the model with mutual information $I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1})$ closer to the upper bound to have greater predictive power. In other words, mutual information $I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1})$ can be used as a metric to evaluate the out-of-sample forecasting performance of the model. The performance of different economic models can be ranked based on mutual information $I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1})$. The total information transfer from $\mathbf{X}_t^{(k)}$ and $\Delta \mathbf{S}_t^{(k)}$ to ΔS_{t+1} can be further decomposed into

two parts:

$$I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1}) = I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1}) + I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)}). \quad (3.17)$$

Here we have used the chain-rule for mutual information, equation (3.15). Equation (3.17) simply points out that the information transfer from $\mathbf{X}_t^{(k)}$ and $\mathbf{S}_t^{(k)}$ to ΔS_{t+1} can be decomposed into two parts: the information transfer from current and past movements of the exchange rate to its future return, and the information transfer from factors $\mathbf{X}_t^{(k)}$ to the future return ΔS_{t+1} conditional on the historical returns of the exchange rate. We refer the first term on the right hand side of equation (3.17) as the Auto-Information-Transfer (AIT). AIT quantifies the auto-predictive power of the exchange rates, as well as the “memory” of the time series across time. The second term is known as Transfer Entropy (TE):

$$TE = I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)}). \quad (3.18)$$

It calibrates the predictive power of factors \mathbf{X} for the future return ΔS_{t+1} . In the following sections, we will estimate AIT and TE using the time series of exchange rates and various factors, and discuss exchange rate predictability at the hourly, daily and monthly frequency based on the empirical results.

3.5 Data

Three types of data are used in our study: exchange rates, order flow, and macroeconomic fundamentals. At the daily and hourly frequency, the sample period is November 19th, 2003 to February 28th, 2014. The exchange rates and order flows data are obtained from Thomson Reuters Dealing system. At the daily frequency, there are fifteen exchange rates in the sample: Australian Dollar/U.S. Dollar (AUD), Canadian Dollar/U.S. Dollar (CAD), Swiss Franc/U.S. Dollar (CHF), Euro/U.S. Dollar (EUR), Euro/British Pound (EURGBP), Euro/Swiss Franc (EURCHF), Euro/Norwegian Krone (EURNOK), Euro/Swedish Krona (EURSEK), British Pound/U.S. Dollar (GBP), Hong Kong Dollar/U.S. Dollar (HKD), Japanese yen/U.S. Dollar (JPY), Mexican Peso/U.S. Dollar (MXN), New Zealand Dollar/U.S. Dollar (NZD), Singapore

Dollar/U.S. Dollar (SGD), South African Rand/U.S. Dollar (ZAR). At the hourly frequency, there are thirteen exchange rates, including the exchange rates listed above, except Swiss Franc/U.S. Dollar (CHF) and Euro/Swiss Franc (EURCHF). They are excluded due to the fact that these two exchange rates are mainly traded on the EBS system instead of the Reuters system, as a result, a small percentage of the hourly order flow observations are missing.

Market order flow mo is measured as the difference between the trade quantities initiated by buyers and the trade quantities initiated by sellers in each period. The limit order flow lo is calculated as the sum of all bid order quantities arriving at the best price (either price improving or best price matching) minus the sum of all offer order quantities at the best over the corresponding period. Moreover, the data also contains 3-month interest rates of the corresponding countries, obtained from Datastream. The interest rate is the only macroeconomic fundamental that is available at the daily frequency.

The data at the monthly frequency is collected separately. Because based on the original ten-year exchange rates data, the monthly sampled time series would be too short for accurate nonparametric estimation. Aiming at better information transfer estimation, we should enlarge the sample period as long as possible. Meanwhile, some difficult tradeoffs have to be taken into account during the data collection. The first one is the availability of the data: the exchange rates and macroeconomic fundamentals of different countries are of very different length. The consideration is that time series in our sample should have the greatest length possible. Moreover, we also try to collect the fundamentals for as many currency pairs as possible, since our aim is to assess whether the exchange rates can be predicted systemically. Taking these factors into account, we set the sample period as January 1980 to December 2013, which means every time series contains 34 years of data and 408 monthly observations. The macroeconomic fundamentals at the monthly frequency are the interest rate, money supply, inflation, and industrial production. Ten currency pairs are selected: Australian Dollar/U.S. Dollar (AUD), Canadian Dollar/U.S. Dollar (CAD), Swiss Franc/U.S. Dollar (CHF), Danish Krone/U.S. Dollar (DKK), British Pound/U.S. Dollar (GBP), Japanese yen/U.S. Dollar (JPY), Norwegian Krone/U.S. Dollar (NOK), New Zealand Dollar/U.S. Dollar (NZD), Swedish Krona/U.S. Dollar (SEK), South African Rand/U.S. Dollar (ZAR).

Another caveat is that for some macroeconomic fundamentals, the stationarity of the time series

could be questionable. For instance, for some currency pairs, the log difference of the monetary supplies of the two corresponding countries is not stationary: an obvious trend can be spotted by eyeballing the time series plot.

The non-stationarity of these time series may be due to various economic reasons, a recommended further exercise is to carry out sub-sample analysis. However, for monthly data, there are merely 12 data points each year, while our non-parametric estimation method definably requires more than several hundred data points for accurate and sensible estimates. To avoid any spurious results, we will use the growth rate of the monetary supply, instead of the monetary supply itself. For similar reasons, we use the difference of inflations as a factor, rather than the difference of the price levels. Another point is that even for most of the econometric methods such as regression, to avoid spurious results, stationarity of the time series are also required.

3.6 Estimation and Empirical Results

In the first and second parts of this section, we will discuss the contemporaneous and predictive information transfer, respectively. The former quantifies the explanatory power of the factors for the concurrent exchange rate movements, and the later gauges the predictive power for the future returns. In the first subsections, we will examine the contemporaneous information transfer (or the explanatory power) of the factors, and compare the results with the known empirical results based on linear regression. The contemporaneous results are supportive for the validity of the new measure of information transfer, which contains the traditional method as a special case under the linear-Gaussian assumption. At the same time, they serve as the frame of reference for the out-of-sample case. Then, in the second subsection, we are going to investigate the predictability of the exchange rates by estimating Transfer Entropy (TE) and Auto-Information-Transfer (AIT), using data at the hourly, daily and monthly frequency.

3.6.1 Contemporaneous Transfer Entropy and Explanatory Power

The contemporaneous explanatory power of factors \mathbf{X} are measured by Contemporaneous Transfer Entropy (CTE) $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$. This measure is similar as Transfer Entropy

$I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ (equation 3.18), but with $k = 1$, and the delay of factors \mathbf{X} with respect to ΔS is zero. CTE quantifies the contemporaneous information transfer from \mathbf{X}_t to ΔS_t . It is a model-independent measure for the explanatory power of factors \mathbf{X} . At hourly frequency, three different sets of factors \mathbf{X} are used: limit order flow lo , market order flow mo , and their combination $\{lo, mo\}$. Because from the market microstructure perspective, limit order flow (lo) and market order flow (mo) are crucial explanatory variables, which have been documented repeatedly in the literature. Table 3.1 reports the estimates and p -values of CTE $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$, and the values of R^2 from the linear regressions $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$, at the hourly frequency.

Insert table 3.1 about here

The first three subsections of table 3.1 report Gaussian estimates of the Contemporaneous Transfer Entropy $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ and the corresponding R^2 from the linear regressions. The values of R^2 suggest that mo has very good explanatory power for the contemporaneous exchange rate movements, and lo has small explanatory power. These results are consistent with the established results in previous studies, such as Evans and Lyons (2002), Chinn and Moore (2011), and Kozhan, Moore and Payne (2014). For several currency pairs mainly traded on the EBS dealing system, our order flow data from the Reuters system have weaker explanatory power. In other words, the occasional low explanatory power is not due to the fact that order flow is a poor explanatory variable. In table 3.1, the Gaussian estimates of CTE are labeled with subscript G . The p -values indicate that the Gaussian estimates of CTE are all significant from zero, for both market order flow mo and limit order lo .

The results in the last three subsections are estimated by the KSG method. Compared with Gaussian estimates, KSG estimates are the model-independent nonparametric measurements of the information transfer, without any *ex ante* assumption about the distribution of the time series or the relationship among them. In table 3.1, all KSG estimates are positive and statistically significant from zero. The CTE of mo and $\{mo, lo\}$ are at the order of magnitude of 0.1 nats, and the CTE of lo is roughly 10 times smaller, at the order of magnitude of 0.01 nats. One thing to notice is that the Gaussian estimates are smaller than the corresponding

KSG estimates. It is logical because Gaussian estimates are constrained by the linear-Gaussian model, while it is not the case for KSG estimates. Comparing the Gaussian and KSG estimates in table 3.1, we find that the linear-Gaussian model is able to capture a large fraction of the total contemporaneous information transfer.

Insert figure 3.1 about here

As we have discussed, mutual information (equation 3.6) is *the* correct measure to quantify the statistical dependence between two variables. Compared with its linear counterpart, mutual information is more general, since it can capture the linear as well as nonlinear dependence. If we impose the assumption that the time series have linear interaction, and the distribution is multivariate Gaussian, we expect that there is a simple equivalence between CTE $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ under linear-Gaussian assumption, and R^2 from the linear regression $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$ (Barnett, Barnett and Seth, 2009). Remind that R^2 captures the explanatory power of regressors \mathbf{X}_t for the dependent variable ΔS_t . In the language of information theory, R^2 is the information obtained about ΔS_t through the regressors \mathbf{X}_t in the framework of linear model, which is exactly the same as CTE $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ under linear-Gaussian assumption. Therefore their equivalence is very natural. To illustrate the equivalence empirically, we plot the values of R^2 against the corresponding Gaussian estimates of CTE in figure 3.1. The Gaussian estimates of CTE are on the X axis, and the corresponding R^2 s are on the Y axis. The coordinates are the values reported in the first three subsections of table 3.1. One can easily spot the linear correspondence between these two measures.

One of the purposes of studying contemporaneous information transfer is to provide a frame of reference for interpreting the level of the information transfer. The equivalence between Gaussian estimates of CTE and R^2 of the linear regression establishes the connection between the unit of the new measure and that of the traditional one. At least qualitatively, it helps us to make sense of the scale or order of magnitude of the information transfer. In the next subsection, when evaluating the economic significance of information transfer from $\{\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}\}$ to ΔS_{t+1} , the contemporaneous counterpart can be used to evaluate how many nats (or bits) of information transfer is economically big enough. In table 3.1, the Gaussian estimates $CTE_{G,mo}$

and $CTE_{G,lo}$, which are at the order of magnitude of 0.1 nats and 0.01 nats, respectively. Considering the well-known empirical fact that the explanatory power of mo is good, and explanatory power of lo is quite small, the estimates of CTE can be used as yardsticks to assess the magnitude of predictive power. To be more specific, we are going to compare the estimates of the Auto-Information-Transfer and Transfer Entropy with the estimates of CTE. For instance, if the estimate of TE $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ is at the order of magnitude of 0.001 nats, the information transfer from $\mathbf{X}_t^{(k)}$ to $\Delta \mathbf{S}_t^{(k)}$ is approximately one hundred (or ten) times smaller than the CTE of mo (or lo), we would say that the predictive power of factors \mathbf{X} is negligible, and TE is not economic significant. In this case, even though TE itself might be statistically significant, but the magnitude of information transfer is so small such that there is no model based on \mathbf{X} is able to predict the exchange rate in any economic meaningful way.

Table 3.2 show the contemporaneous results at daily frequency. There are fifteen exchange rates, and three different sets of factors are used: interest rate differential IR , market order flow mo , and their combination $\{mo, IR\}$. The estimates of Contemporaneous Transfer Entropy (CTE) $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ and R^2 from the linear regression are reported in table 3.2. Similarly as before, the estimates of CTE in the first three subsections of table 3.2 are based on Gaussian method, while those in the last three subsections are estimated using KSG method. The results are mainly twofold. First, the values of R^2 suggest that IR has very small explanatory power. CTE of IR are at the order of magnitude of 0.01 nats, and some of them are not statistically significant. Second, based on the values of R^2 , market order flow mo has good explanatory power in general, except for several exchange rates that are mainly traded on EBS dealing system, as we explained before. The CTE of mo is generally at the order of magnitude of 0.1 nats, and significantly different from zero. Furthermore, the equivalence of R^2 and Gaussian estimates of $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ is shown in figure 3.2, similarly as the case at hourly frequency. Moreover, the majority of the total information transfer, which is measured by KSG estimates of the CTE, can be captured by the linear-Gaussian model.

Insert table 3.2 about here

Insert figure 3.2 about here

3.6.2 Transfer Entropy, Auto-Information-Transfer and Predictive Power

Having studied the contemporaneous explanatory power of various factors, now we are going to investigate the out-of-sample predictability of exchange rates. As we have discussed, time series predictability depends on the information transfer from factors $\mathbf{X}_t^{(k)}$ and historical movements $\Delta \mathbf{S}_t^{(k)}$ to the future return in the next period, ΔS_{t+1} . We have argued that based on Shannon's channel coding theorem, the correct information transfer is measured by mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$. For any model based on $\mathbf{X}_t^{(k)}$ and $\Delta \mathbf{S}_t^{(k)}$, $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is the upper bound of the predictive power. Therefore, as a measure of the predictability, mutual information $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ is independent from the economic model or econometric method adopted for forecasting. Empirically, we will estimate Transfer Entropy (TE) $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ and Auto-Information-Transfer (AIT) $I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$. They calibrate the predictive power of two different sources: the chosen factors \mathbf{X} and historical movements of the exchange rate. The other advantage of our method is that information transfer can be estimated based on the nonparametric method as well as the linear-Gaussian method. In such way, the performance of the linear-Gaussian model can be compared with that of the optimal scenario. For instance, what fraction of the total information can be harnessed by the frequently-used linear model?

For the predictability at the hourly frequency, the estimates of TE and AIT with $k = 1$ are reported in table 3.3. In the first two subsections of this table, we have the Gaussian and KSG estimates of TE, and the last two subsections show the Gaussian and KSG estimates of AIT. The subscript G in table 3.3 indicates the Gaussian estimate of TE (or AIT), which measures the predictive power of the linear-Gaussian model based on factors \mathbf{X} (or the historical returns $\Delta \mathbf{S}_t^{(k)}$). The KSG estimate of TE (or AIT) is the non-parametric model-independent estimation of information transfer, it quantifies the maximum predictive power of factors \mathbf{X} (or $\Delta \mathbf{S}_t^{(k)}$) for the future return.

The first subsection of table 3.3 shows that the Gaussian estimates $TE_{G,lo}$ are not significant from zero in eight out of thirteen cases. For market order flow, mo , and their combination, $molo$, roughly half of the exchange rates have statistically significant Gaussian estimates $TE_{G,mo}$ and $TE_{G,molo}$. However, their values are very small, at the order of magnitude of 0.0001 nats or even

smaller. As we have discussed, the contemporaneous information transfer (CTE) from *mo* and *lo* to the exchange rate are at the order of magnitude of 0.1 nats and 0.01 nats, respectively, which concrete the levels of “large” and “small” information transfer. Compared to these yardsticks, 0.0001 nats is 1000 and 100 times smaller. Hence we think it is safe to say that the Gaussian estimates of TE are not economically significant. In other words, at the hourly frequency, the linear-Gaussian model based on limit and market order flow has negligible predictive power.

The model-independent KSG estimates of TE $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ with $k = 1$ are reported in the second subsection of table 3.3. It can be easily observed that KSG estimates are greater than their Gaussian counterparts, which is consistent with our expectation. For *mo* and *molo*, more than half of the KSG estimates of TE are statistically significant, and roughly at the order of magnitude of 0.01 nats. Remember that in the previous subsection, the CTE of *lo* is also approximately 0.01 nats. Therefore, for these exchange rates with significant TE_{mo}/TE_{molo} , the optimal model based on *mo* or *molo* has small predictive power. The optimal model can fully exploit the information contained in *mo* or *molo* about the future exchange rate movements. However, the specification of the optimal model is beyond the scope of this paper. Our method evaluate the predictability but does not show the specification of the optimal model. However, we do know that the linear model has already been excluded. By comparing the KSG estimates with their linear-Gaussian counterparts, we see that the linear-Gaussian model merely captures a very small fraction of the total information transfer.

Now let’s look into the predictability based on the historical returns. The Gaussian and KSG estimates of AIT $I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ with $k = 1$ are presented in the last two subsections of table 3.3. The p -values suggest that they are all significantly different from zero. It means that at the hourly frequency, the time series of the exchange rate returns have strong interdependence across time. Another thing to notice is that there is an enormous difference between the magnitudes of Gaussian and KSG estimates of AIT. The Gaussian estimates of AIT are at the order of magnitude of 0.001 nats, which suggests that the predictive power of the linear model based on historical exchange rate movements is very small. In comparison, the KSG estimates are at the order of magnitude of 0.1 or even 1 nats, which are 100 times greater than the Gaussian estimates. Compared to the CTE benchmarks, the information transfer from $\Delta \mathbf{S}_t^{(k)}$ to ΔS_{t+1} is very large, generally speaking. Therefore, at the hourly frequency, the exchange rates are

predictable using the optimal model based on its historical returns. On the other hand, the Gaussian estimates of AIT indicate that the predictive power of the linear model is 100 times smaller than the optimal model. In other words, the frequently-used linear-Gaussian model fails to capture most of the predictive power of the historical returns. The optimal forecasting model must be nonlinear. However, once again, our approach just examines the predictability and the existence of the optimal model, but not its specification.

Insert table 3.3 about here

Before we move to discuss the results at the daily and monthly frequency, we should point out that since the estimates of these information measures are carried out using numerical method, there would be small numerical errors incurred in the numerical computation process. Therefore it could be the case that the true value is zero, and its numeric estimate is zero plus some small numerical errors, positive or negative. In this case, the estimate could be a small negative value. Since we believe all numerics should be faithfully reported, we will not automatically correct these small negative estimates and set them as zero. The way we cope with this issue is based on the statistical significance test: if the small negative value is negative but not significantly different from zero at all, there is nothing to worry about. The non-significant negative value is not due to something fundamentally incorrect. As with any numerical method, statistical/econometric estimation incurs small numerical error as well. There is no need to feel alerted here. The small numerical error is inevitable for any empirical studies.

With respect to the predictability at the daily frequency, the forecasting factors are interest rate differential IR , market order flow mo , and their combination, denoted as $moIR$. The estimates are summarized in table 3.4. In the first subsection of table 3.4, the Gaussian estimates of TE are very small, roughly at the order of magnitude of 0.0001 nats, and not statistically significant, which suggests that it is impossible to predict exchange rate using the linear-Gaussian model with mo and IR as predictors. On the other hand, the KSG estimates in the second subsection of table 3.4 indicate that TE_{IR} is at the order of magnitude of 0.01 nats and statistically significant for most of the exchange rates, while TE_{mo} is generally not significant. Hence for these exchange rates with significant TE_{IR} , there exists an optimal model based on IR with

small predictive power at the daily frequency. However, none of the model based on *mo* has any predictive power.

For the auto-predictive power at the daily frequency, some of the Gaussian estimates of AIT are significantly different from zero, but their values are too small to be economically significant. Moreover, the majority of the KSG estimates of AIT are not significant, which suggests that at the daily frequency, the historical observations of the exchange rate have no predictive power for its future movement. This is consistent with the stylized empirical fact that daily returns are not serially correlated across time. However, there is an astronomical difference between the time series predictability at the hourly and daily frequency. The time series of exchange rates are systematically predictable at the hourly frequency, but not at the daily frequency. It is mainly due to the disappearance of the statistical dependence across time when the sampling frequency moves from hourly to daily. In a word, the time series predictability is not time scale invariant.

Insert table 3.4 about here

At the monthly frequency, the macroeconomic fundamentals for forecasting are the differentials of interest rate (*IR*), inflation (*IF*), money growth *M*, and industrial production *prod*. As we have discussed, the log difference of monetary supply is non-stationary for some exchange rates. Hence we will use the difference in the growth rate of monetary supply instead. For the similar reason, the difference of inflation is adopted as a forecasting factor, rather than the difference of price level. Another relevant issue is the length of the time series. In our dataset, every time series sampled at the monthly frequency has 408 observations in the thirty-four-year sample period. The number of the observations is not ideal for nonparametric estimation, but the empirical estimates seem reasonable, as we will see soon. The caveat for working with quarterly or even annually data is that the time series with merely 100 or 40 observations may be too short for reliable nonparametric estimation of the mutual information.

Table 3.5 reports the predictability of ten different exchange rates at the monthly frequency. The estimates of Transfer Entropy (TE) are labeled with subscripts 1 to 4, which correspond to four different sets of factors: $\{IR\}$, $\{IR \text{ and } IF\}$, $\{IR, IF \text{ and } M\}$, $\{IR, IF, M \text{ and } prod\}$.

$prod\}$, respectively. For the estimation of information transfer, similar as before, two different methods are used: the linear-Gaussian model and model-independent KSG method. In table 3.5, the results with subscript G are the Gaussian estimates, and the rest of them are the KSG estimates.

Insert table 3.5 about here

The following conclusion can be drawn with respect to the predictability at the monthly frequency. First of all, based on the Gaussian estimates of the Transfer Entropy (TE), the predictive power of the linear-Gaussian model is not statistically significant for every single exchange rate. It explains why it is very hard for the linear model to outperform the random walk benchmark, i.e., the Meese and Rogoff puzzle. Second, most of the model-independent KSG estimates are actually significant, especially when all the fundamentals are included. The KSG estimates of TE are at the order of magnitude of 0.01 nats, which suggests that the optimal forecasting model would have small predictive power for the future exchange rate return. But the p -values suggest that the exchange rates are not predictable systematically. Statistical significance of the predictive power depends on the specific currency pairs. Third, comparing the Gaussian and KSG estimates, we can tell that the linear model actually fails to capture most of the information transfer, as the cases at the hourly and daily frequency. Furthermore, regarding the predictive power of the historical returns, most of the AIT estimates are not significant. The only exception is the exchange rate of Swedish Krona/U.S. Dollar (SEK). Therefore, there is no statistical dependence cross time the at the daily and monthly frequency, but at the hourly frequency, the time series are highly interdependent.

To summarize, our empirical results suggest that time series predictability of the exchange rates are frequency dependent: they are not systematically predictable at the daily and monthly frequency, but at a higher frequency, such as the hourly frequency, they are predictable. The strong intraday predictive power mainly comes from the historical observations of the exchange rate. Meanwhile, we also document that the intraday market order flow has small but significant predictive power. The intraday results indicate that there exists an optimal model that can predict exchange rates systematically. Its specification is beyond the scope of this study, al-

though we do know that it should be nonlinear, because the Gaussian estimates of AIT and TE suggest that the forecasting performance of the linear-Gaussian model is generally quite poor. The frequently-used linear-Gaussian model fails to exploit most of the information transferred across time. At daily and monthly frequency, for more than half of the exchange rates, the macroeconomic fundamentals have small predictive power for some exchange rates. However, the linear model is suboptimal and does not have significant predictive power.

Once we have established there is non-linear stability, the next step is to carry out forecasting exercise using various non-linear models. The investigation of the best non-linear model for forecasting exchange rate is beyond the scope of this study. However, as we have mentioned, the upper bound of predictability estimated with our method should be used as the benchmark to evaluate forecasting models.

3.7 Conclusions and Discussion

We propose a model-independent nonparametric method to examine the predictability of any time series. This paper studies the time series predictability in the context of exchange rates, because their out-of-sample predictability has puzzled the researchers for several decades. Intuitively, in order to predict exchange rate systematically, the predictors should contain enough information about the future movement of the exchange rate. The information transfer from predictors to the future exchange rate movement is measured by their mutual information, which also acts as the upper bound of the predictive power of any model based on the same set of predictors. Empirically, we show that the exchange rates are predictable at the hourly frequency, but not at the daily and monthly frequency. The intraday predictability mainly comes from the historical observations of the exchange rate itself. For the predictability at daily and monthly frequency, the information transfer from historical returns to the future return is not statistically different from zero, which means the exchange rate movements are statistically independent across time. However, for some currency pairs, there is small information transfer from macroeconomic fundamentals to the future exchange rate movement. It depends on the specific exchange rates, and they are not predictable systematically. The frequently used linear-Gaussian model has very poor forecasting performance, because it fails to capture most

of the information transfer across time. It explains why it is so hard for the linear forecasting models to outperform the random walk benchmark. The optimal forecasting model should be nonlinear.

Last but not least, let's discuss the contributions and potential applications of our nonparametric method based on information theory. First, it introduces a model-independent methodology to examine the predictability of time series, once the set of predictors is given. The predictability can be evaluated using this approach, before one starts to try various forecasting models, because statistically significant information transfer is the necessary condition for the out-of-sample predictability. If there is no significant information transfer, the fruitless effort of testing various economic or econometric models can be avoided. Second, our method is very useful for predictor selection. One should choose the predictors with significant information transfer and discard those with no information transfer. Third, the methodology proposed in this paper is not limited to the predictability of exchange rates. It is also applicable to other financial or economic predictability questions, for example, the predictability of stock returns, or any other economic/financial time series. The only caveat is that there should be enough data points in the sample, in order for the nonparametric estimation to work properly. Fourth, although the method does not directly show us how to predict, it can provide some useful guidance: for instance, does the linear model have good forecasting performance? What is the upper bound of the predictive power? Has the model reached the upper limit, or is there any room for improvement? Moreover, our method is useful for comparing the forecasting performance of the models (model selection): the model f with information transfer $I(f(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}); \Delta S_{t+1})$ closer to the upper limit $I(\mathbf{X}_t^{(k)}, \Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$ must have better forecasting performance. Therefore the pecking order of the models can be established by comparing the rates of information transfer. Moreover, another potential application of our method is parameter estimation: one can estimate the parameters of the forecasting model based on the "principle of maximum information transfer". To be more specific, once the model and the predictors are selected, one can search for the 'optimal' values of the parameters by maximizing the information transfer from model f to the future return.

3.8 Appendix: Mechanism of Noisy Channel Communication

In this paper, we want to link the theory of channel communication and information transfer to the predictability of time series. To see the relevance, let's we briefly discuss the mechanism of information transfer through a noisy channel.

A noisy channel has two ends, one is input Y , and the other is output Z . The output Z and the input Y must be statistically dependent ($P(Y, Z) \neq P(Y)P(Z)$), otherwise it is not possible to communicate through this channel at all. Assume the encoder at the input end needs to send a message to the receiver. He encodes the message into a length- n string of source symbols \mathbf{Y}^n , and send these symbols one by one through the noisy channel. The receiver at the other end will receive a length- n string of symbols \mathbf{Z}^n , then try to decode \mathbf{Z}^n in order to recover the original message \mathbf{Y}^n . To be more concrete, in the digital age, the source symbols \mathbf{Y}^n and the received symbols \mathbf{Z}^n can be long strings of 1s and 0s, such as 010100111001001... The noisy channel can be a copper wire, or a wireless network connecting two devices; and the message can be the content of a book, a movie, or a piece of music.

Communication via a perfect channel is easy, the receiver would obtain the exact symbol/signal sent by the encoder: for instance, if encoder sends 1 via the perfect channel, the receiver will get 1 for sure. There will be no distortion. Hence the receiver on the end of the perfect channel can recover the original message effortlessly. However, the practical problem is that in the real world, all the communication channels are noisy. The received symbols \mathbf{Z}^n are subjected to some distortion: $\mathbf{Z}^n \neq \mathbf{Y}^n$. For example, when the encoder sends a particular symbol from one end, say $y = 1$, instead of receiving the exact same signal with certainty, the output end may receive something different, $z = 0$, with a certain probability. This is clearly an error. Due to the noise and distortion in the communication channel, the received symbols \mathbf{Z}^n will never be exactly the same as the source symbol \mathbf{Y}^n , hence the decoder may never recover the original message sent by the encoder. Considering the unfortunate fact that all communication channels in the real world are inherently noisy, reliable error-free communication was once thought as impossible, until Shannon's channel coding theorem pointed out the right direction.

Shannon's great insight is that even though a channel is noisy, through which error-free communication is still achievable, as long as the information transmission rate, which is defined as

the amount of information per source symbol (bits per source symbol), is lower than a certain bound, also known as the channel capacity. Shannon showed that the channel capacity of the noisy channel Y to Z is measured by the mutual information between Y and Z : $I(Y; Z)$ ². Mutual information $I(Y; Z)$ acts as the upper bound of the transmission rate is due to the fact that for a given channel Y to Z , the encoder of the message always has the flexibility of lowering the information rate in his source symbols \mathbf{Y}^n , just by adding more redundancy. However, if one tries to transfer information at a rate greater than the channel capacity, the probability of error at the receiver end would go to one exponentially fast, as the length of the symbols goes to infinity. Reliable information transmission is infeasible at any rate greater than the channel capacity $I(Y; Z)$. This is the channel coding theorem. To understand this point from more fundamental principles, one needs to know the concepts of the asymptotic equipartition property (AEP) and the Joint Typicality. Due to the limited space of our study, we refer the reader to the monographs for more detailed discussion, such as Gray (2011), Cover and Thomas (2006).

²More precisely, in information theory, the channel capacity is $\max_{p(y)} I(Y; Z)$. It is because when communicating through channel Y to Z , the sender has the flexibility of choosing the most favorable $p(Y)$ to maximize $I(Y; Z)$, in order to maximize the information transmission rate of this channel. On the other hand, the conditional probability $p(Z|Y)$ is the intrinsic property of the noisy channel Y to Z . Therefore, $p(Z|Y)$ is predetermined and not adjustable.

Table 3.1: Explanatory power at the hourly frequency: Contemporaneous Transfer Entropy and R^2

$k = 1$	AUD	CAD	EUR	EUR GBP	EUR NOK	EUR SEK	GBP	HKD	JPY	MXN	NZD	SGD	ZAR
CTE $_{G,lo}$	0.0096	0.0143	0.0256	0.0132	0.0021	0.0016	0.0200	0.0015	0.0032	0.0021	0.0053	0.0021	0.0006
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2_{lo}	0.0191	0.0283	0.0505	0.0252	0.0041	0.0031	0.0395	0.0029	0.0066	0.0042	0.0107	0.0043	0.0012
adj R^2_{lo}	0.0191	0.0283	0.0505	0.0252	0.0041	0.0031	0.0395	0.0028	0.0066	0.0042	0.0107	0.0042	0.0011
CTE $_{G,mo}$	0.1797	0.1615	0.1417	0.0499	0.1368	0.1390	0.1496	0.1360	0.0261	0.0772	0.1058	0.1584	0.1423
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2_{mo}	0.2995	0.2751	0.2430	0.0898	0.2365	0.2415	0.2560	0.2311	0.0497	0.1438	0.1878	0.2669	0.2432
adj R^2_{mo}	0.2995	0.2751	0.2430	0.0898	0.2365	0.2415	0.2560	0.2311	0.0496	0.1437	0.1878	0.2669	0.2432
CTE $_{G,molo}$	0.1973	0.1838	0.1543	0.0543	0.1378	0.1405	0.1624	0.1408	0.0291	0.0781	0.1143	0.1656	0.1435
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2_{molo}	0.3240	0.3068	0.2627	0.0976	0.2379	0.2437	0.2752	0.2383	0.0556	0.1453	0.2015	0.2771	0.2451
adj R^2_{molo}	0.3239	0.3068	0.2626	0.0976	0.2379	0.2437	0.2752	0.2383	0.0556	0.1453	0.2015	0.2771	0.2451
CTE $_{lo}$	0.0335	0.0375	0.0573	0.0484	0.0234	0.0196	0.0477	0.0317	0.0403	0.0341	0.0219	0.0247	0.0194
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CTE $_{mo}$	0.2579	0.2085	0.1683	0.1379	0.1924	0.1929	0.1996	0.1741	0.0576	0.1698	0.1404	0.1960	0.2095
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CTE $_{molo}$	0.2981	0.2717	0.1995	0.1586	0.2131	0.2252	0.2342	0.2190	0.0852	0.1934	0.1673	0.2398	0.2457
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The contemporaneous explanatory power at the hourly frequency. We estimate the Contemporaneous Transfer Entropy (CTE) of the following factors \mathbf{X} : limit order flow (lo) and market order flow (mo), and their combination ($molo$). There are thirteen exchange rates in the sample. The first three subsections report the Contemporaneous Transfer Entropy $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ estimated using the linear-Gaussian model, and R^2 of the linear regression $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$. The Gaussian estimates are labeled with subscript G . In the last three subsection of the table, Contemporaneous Transfer Entropy $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ is estimated using KSG method. The p -value is calculated using the empirical distribution of CTE estimates based on the simulated samples under the null hypothesis that there is no information transfer. The number of the simulated samples is $S = 1000$.

Table 3.2: Explanatory power at the daily frequency: Contemporaneous Transfer Entropy and R^2

$k = 1$	AUD	CAD	CHF	EUR	EUR CHF	EUR GBP	EUR NOK	EUR SEK	GBP	HKD	JPY	MXN	NZD	SGD	ZAR
CTE _{G,IR}	0.0003	0.0011	0.0000	0.0003	0.0000	0.0000	0.0001	0.0000	0.0012	0.0009	0.0000	0.0000	0.0012	0.0001	0.0001
p -value	0.1830	0.0180	0.8480	0.2400	0.7720	0.6270	0.4240	0.6260	0.0090	0.0510	0.8430	0.9410	0.0070	0.6000	0.5420
R^2_{IR}	0.0006	0.0020	0.0000	0.0006	0.0000	0.0001	0.0003	0.0001	0.0025	0.0016	0.0000	0.0000	0.0024	0.0001	0.0001
adj R^2_{IR}	0.0002	0.0016	-0.0004	0.0001	-0.0004	-0.0003	-0.0001	-0.0004	0.0021	0.0011	-0.0004	-0.0004	0.0020	-0.0003	-0.0003
CTE _{G,mo}	0.1725	0.1611	0.0073	0.1925	0.0274	0.0850	0.1513	0.1214	0.1864	0.1164	0.0332	0.0470	0.1291	0.1546	0.1488
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2_{mo}	0.2875	0.2684	0.0145	0.3165	0.0532	0.1590	0.2595	0.2148	0.3127	0.2049	0.0604	0.0915	0.2261	0.2638	0.2563
adj R^2_{mo}	0.2872	0.2681	0.0141	0.3162	0.0528	0.1586	0.2592	0.2145	0.3125	0.2046	0.0600	0.0911	0.2258	0.2635	0.2560
CTE _{G,moIR}	0.1727	0.1632	0.0075	0.1933	0.0276	0.0867	0.1519	0.1226	0.1931	0.1191	0.0333	0.0475	0.1329	0.1552	0.1511
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2_{moIR}	0.2878	0.2709	0.0149	0.3177	0.0535	0.1618	0.2604	0.2167	0.3220	0.2089	0.0604	0.0923	0.2316	0.2647	0.2599
adj R^2_{moIR}	0.2873	0.2703	0.0140	0.3171	0.0527	0.1611	0.2598	0.2160	0.3214	0.2082	0.0596	0.0915	0.2310	0.2641	0.2593
CTE _{IR}	0.0485	0.0131	0.0077	0.0383	0.0374	0.0337	0.0292	0.0461	0.0111	0.0390	0.0206	0.0406	0.0201	0.0150	0.0170
p -value	0.0000	0.0910	0.2460	0.0020	0.0000	0.0020	0.0070	0.0000	0.1360	0.0000	0.0240	0.0000	0.0330	0.0810	0.0800
CTE _{mo}	0.2803	0.2062	0.0072	0.2247	0.0784	0.0959	0.1861	0.1539	0.2349	0.1110	0.0318	0.1173	0.1718	0.2069	0.1883
p -value	0.0000	0.0000	0.2170	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0040	0.0000	0.0000	0.0000	0.0000
CTE _{moIR}	0.3426	0.2490	0.0470	0.2806	0.1228	0.1428	0.2059	0.1780	0.2607	0.1282	0.0843	0.1457	0.1935	0.2584	0.2470
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The contemporaneous explanatory power at the daily frequency. We estimate the Contemporaneous Transfer Entropy (CTE) of the following factors \mathbf{X} : market order flow (mo), interest rate differential (IR), and their combination ($moIR$). There are fifteen exchange rates in the sample. The first three subsections report the Contemporaneous Transfer Entropy $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ estimated using the linear-Gaussian model, and R^2 of the linear regression $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$. The Gaussian estimates are labeled with subscript G . In the last three subsection of the table, Contemporaneous Transfer Entropy $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ is estimated using KSG method. The p -value is calculated using the empirical distribution of CTE estimates based on the simulated samples under the null hypothesis that there is no information transfer. The number of the simulated samples is $S = 5000$.

Table 3.3: Predictive power at the hourly frequency: Transfer Entropy and AIT

$k = 1$	AUD	CAD	EUR	EUR GBP	EUR NOK	EUR SEK	GBP	HKD	JPY	MXN	NZD	SGD	ZAR
$TE_{G,lo}$	0.0000	0.0000	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
p -value	0.0760	0.0250	0.0010	0.0000	0.0160	0.1250	0.4890	0.0700	0.0000	0.0680	0.5470	0.2000	0.1040
$TE_{G,mo}$	0.0000	0.0001	0.0002	0.0006	0.0000	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000
p -value	0.8680	0.0010	0.0000	0.0000	0.6450	0.2990	0.0000	0.0010	0.4070	0.0090	0.0020	0.4180	0.3650
$TE_{G,molo}$	0.0000	0.0001	0.0003	0.0006	0.0001	0.0000	0.0001	0.0002	0.0002	0.0001	0.0001	0.0000	0.0000
p -value	0.2080	0.0010	0.0000	0.0000	0.0400	0.1800	0.0000	0.0040	0.0030	0.0140	0.0020	0.3810	0.1840
TE_{lo}	0.0111	0.0115	0.0011	0.0113	0.0056	0.0083	0.0022	0.0038	-0.0015	0.0101	0.0084	-0.0029	0.0066
p -value	0.0000	0.0000	0.3210	0.0000	0.0220	0.0030	0.1570	0.0060	0.6720	0.0000	0.0000	0.7730	0.0080
TE_{mo}	0.0104	0.0111	0.0061	0.0097	0.0098	0.0110	0.0077	-0.0015	0.0022	0.0092	0.0021	-0.0018	0.0099
p -value	0.0000	0.0000	0.0100	0.0000	0.0000	0.0000	0.0000	0.2750	0.2340	0.0000	0.1700	0.6430	0.0000
TE_{molo}	0.0071	0.0164	0.0069	0.0108	0.0084	0.0111	0.0116	0.0027	0.0065	0.0146	0.0024	0.0021	0.0096
p -value	0.0040	0.0000	0.0100	0.0000	0.0010	0.0000	0.0000	0.0770	0.0240	0.0000	0.1820	0.2010	0.0010
AIT_G	0.0005	0.0007	0.0009	0.0107	0.0034	0.0018	0.0004	0.0044	0.0011	0.0051	0.0015	0.0019	0.0014
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AIT	0.8595	1.0769	0.7142	1.0352	0.3866	0.4656	0.6442	2.2566	0.2856	0.1510	0.9598	1.2745	0.1260
p -value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The predictive power at the hourly frequency. The factors \mathbf{X} are limit order flow (*lo*), market order flow (*mo*), and their combination (*molo*). In total, there are thirteen exchange rates in the sample. This table reports the estimates of Transfer Entropy (TE) $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ and Auto-Information-Transfer (AIT) $I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$. The first two subsections of this table contain the Gaussian and KSG estimates of TE, and the last two subsections show the Gaussian and KSG estimates of AIT. The Gaussian estimates are labeled with subscript G . The p -value is calculated using the empirical distribution of CTE/AIT estimates based on the simulated samples under the null hypothesis that there is no information transfer. The number of the simulated samples is $S = 1000$.

Table 3.4: Predictive power at the daily frequency: Transfer Entropy and AIT

k=1	AUD	CAD	CHF	EUR	EUR CHF	EUR GBP	EUR NOK	EUR SEK	GBP	HKD	JPY	MXN	NZD	SGD	ZAR
$TE_{G,IR}$	0.0003	0.0008	0.0000	0.0003	0.0000	0.0001	0.0004	0.0000	0.0012	0.0006	0.0000	0.0000	0.0011	0.0001	0.0001
p -value	0.1867	0.0337	0.9433	0.2497	0.8900	0.5860	0.1800	0.6827	0.0077	0.0970	0.8730	0.9370	0.0160	0.5233	0.5463
$TE_{G,mo}$	0.0000	0.0000	0.0003	0.0001	0.0000	0.0002	0.0003	0.0000	0.0000	0.0001	0.0002	0.0013	0.0002	0.0007	0.0001
p -value	0.9753	0.9503	0.2450	0.4693	0.9173	0.2840	0.2177	0.7447	0.9450	0.5863	0.3930	0.0133	0.2640	0.0690	0.5043
$TE_{G,moIR}$	0.0003	0.0008	0.0003	0.0005	0.0000	0.0003	0.0008	0.0001	0.0012	0.0007	0.0002	0.0014	0.0014	0.0008	0.0002
p -value	0.4430	0.1117	0.4880	0.3547	0.9837	0.4270	0.1540	0.8630	0.0353	0.2077	0.6690	0.0533	0.0253	0.1680	0.6007
TE_{IR}	0.0468	0.0128	0.0137	0.0362	0.0312	0.0341	0.0368	0.0464	0.0134	0.0252	0.0228	0.0339	0.0304	0.0224	0.0119
p -value	0.0000	0.1283	0.1140	0.0013	0.0040	0.0007	0.0017	0.0000	0.1023	0.0117	0.0210	0.0033	0.0027	0.0187	0.1633
TE_{mo}	0.0050	-0.0113	-0.0007	-0.0135	-0.0077	0.0179	0.0165	0.0036	0.0229	0.0108	0.0084	-0.0045	-0.0086	0.0117	-0.0069
p -value	0.3257	0.8373	0.5143	0.8740	0.7550	0.0577	0.0903	0.3773	0.0220	0.1783	0.2417	0.6310	0.7687	0.1453	0.7240
TE_{moIR}	0.0239	0.0366	0.0235	0.0245	0.0362	0.0576	0.0198	0.0463	0.0140	0.0238	0.0032	0.0137	0.0280	0.0165	0.0191
p -value	0.0117	0.0000	0.0183	0.0210	0.0003	0.0000	0.0487	0.0003	0.1033	0.0193	0.4063	0.1247	0.0077	0.0797	0.0527
AIT_G	0.0009	0.0006	0.0002	0.0000	0.0003	0.0018	0.0001	0.0008	0.0012	0.0011	0.0035	0.0019	0.0001	0.0004	0.0003
p -value	0.0263	0.0697	0.3527	0.7257	0.2297	0.0017	0.4827	0.0457	0.0097	0.0343	0.0000	0.0067	0.5997	0.2010	0.2150
AIT	-0.0087	0.0221	-0.0196	0.0033	0.1030	0.0144	0.0119	0.0143	0.0240	0.1155	0.0046	0.0305	0.0219	0.0484	0.0188
p -value	0.7290	0.0523	0.9157	0.4027	0.0000	0.1527	0.2157	0.1640	0.0453	0.0000	0.3777	0.0233	0.0613	0.0017	0.1220

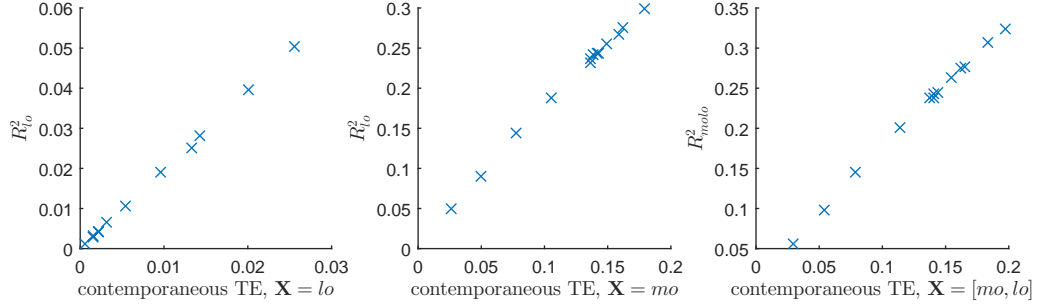
The predictive power at the daily frequency. The factors \mathbf{X} are market order flow (mo), interest rate differential (IR) and their combination ($moIR$). In total, there are fifteen exchange rates in the sample. This table reports the estimates of Transfer Entropy (TE) $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ and Auto-Information-Transfer (AIT) $I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$. The first two subsections of this table contain the Gaussian and KSG estimates of TE, and the last two subsections show the Gaussian and KSG estimates of AIT. The Gaussian estimates are labeled with subscript G . The p -value is calculated using the empirical distribution of CTE/AIT estimates based on the simulated samples under the null hypothesis that there is no information transfer. The number of the simulated samples is $S = 5000$.

Table 3.5: Predictive power at the monthly frequency: Transfer Entropy and AIT

k=1	AUD	CAD	CHF	DKK	GBP	JPY	NOK	NZD	SEK	ZAR
$TE_{1,G}$	0.00000	0.00013	0.00551	0.00256	0.00562	0.00650	0.00007	0.00191	0.00015	0.00032
p -value	0.99760	0.74020	0.03600	0.15020	0.03620	0.01940	0.81340	0.20580	0.72380	0.60920
$TE_{2,G}$	0.00420	0.00380	0.00760	0.00552	0.00603	0.00654	0.00088	0.00219	0.00148	0.00167
p -value	0.18720	0.21420	0.05160	0.10300	0.09180	0.06800	0.66940	0.41180	0.54040	0.51033
$TE_{3,G}$	0.00427	0.00384	0.00849	0.00820	0.00616	0.00659	0.00107	0.00221	0.00148	0.00179
p -value	0.32840	0.36840	0.07920	0.07480	0.16140	0.13900	0.82540	0.60900	0.75260	0.68340
$TE_{4,G}$	0.00572	0.00450	0.00851	0.00976	0.01119	0.00913	0.00139	0.00829	0.00732	0.00594
p -value	0.31640	0.46240	0.15240	0.08400	0.05840	0.09560	0.88000	0.13140	0.19560	0.28880
TE_1	0.01474	-0.02085	0.01891	0.05923	0.02053	0.01388	0.02939	0.03278	0.09616	0.05617
p -value	0.26640	0.79960	0.22100	0.00660	0.20080	0.27280	0.11320	0.07780	0.00000	0.00940
TE_2	0.06852	0.02299	0.05210	0.02593	0.06184	-0.03029	0.02317	0.03203	0.06405	0.05184
p -value	0.00400	0.16000	0.01800	0.14000	0.00920	0.90120	0.16860	0.09220	0.00720	0.01620
TE_3	0.06966	0.02396	0.04308	0.02870	0.05503	-0.01664	0.06946	0.07774	0.05004	0.03586
p -value	0.00180	0.14560	0.03900	0.11860	0.01180	0.75140	0.00220	0.00080	0.01920	0.05220
TE_4	0.04426	0.02560	0.05851	0.05260	0.06348	0.03241	0.05298	0.10281	0.07616	0.03135
p -value	0.03040	0.13280	0.00620	0.01280	0.00420	0.07380	0.00900	0.00000	0.00040	0.07660
AIT_G	0.00227	0.00190	0.00040	0.00137	0.00364	0.00132	0.00063	0.00001	0.00600	0.00095
p -value	0.17620	0.20680	0.57880	0.28560	0.08560	0.30700	0.47440	0.94040	0.02280	0.36220
AIT	-0.03457	-0.03031	0.06653	0.02035	-0.05336	0.01516	0.00656	-0.03054	0.05755	-0.02215
p -value	0.85380	0.81620	0.03060	0.25680	0.95140	0.31620	0.41460	0.82700	0.04920	0.74780

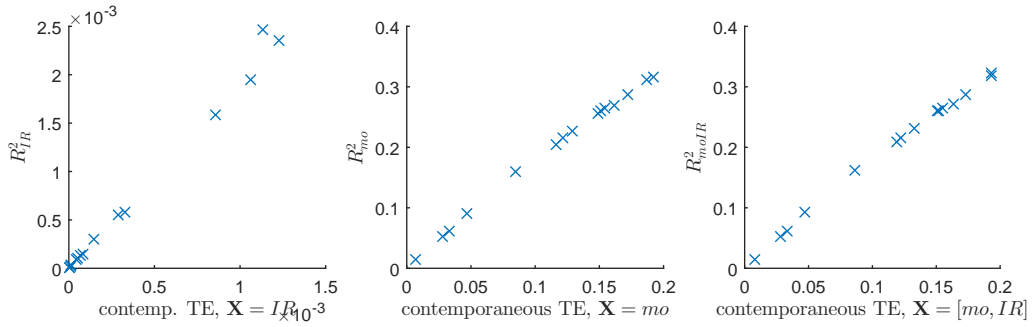
The predictive power at the monthly frequency. The macroeconomic fundamentals are the interest rate (IR), inflation (IF), money growth M , and industrial production differentials $prod$. The estimates of the Transfer Entropy (TE) and Auto-Information-Transfer (AIT) are labeled with subscripts 1 to 4, which corresponds to four different sets of factors: $\{IR\}$, $\{IR \text{ and } IF\}$, $\{IR, IF \text{ and } M\}$, $\{IR, IF, M \text{ and } prod\}$, respectively. There are ten exchange rates in the sample. This table reports the estimates of Transfer Entropy (TE) $I(\mathbf{X}_t^{(k)}; \Delta S_{t+1} | \Delta \mathbf{S}_t^{(k)})$ and Auto-Information-Transfer (AIT) $I(\Delta \mathbf{S}_t^{(k)}; \Delta S_{t+1})$. The first two subsections of the table contain the Gaussian and KSG estimates of TE, and the last two subsections show the Gaussian and KSG estimates of AIT. The Gaussian estimates are labeled with subscript G . The p -value is calculated using the empirical distribution of CTE/AIT estimates based on the simulated samples under the null hypothesis that there is no information transfer. The number of the simulated samples is $S = 5000$.

Figure 3.1: The equivalence between R^2 and Gaussian estimate of CTE at the hourly frequency



The three figures show the linear equivalence between the Gaussian estimates of Contemporaneous Transfer Entropy (CTE) $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ and R^2 of the linear regression $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$ at the hourly frequency. The factors \mathbf{X} are limit order flow (*lo*), market order flow (*mo*), and their combination (*molo*). The coordinates are the Gaussian estimates of CTE and linear regression R^2 reported in table 3.1.

Figure 3.2: The equivalence between R^2 and Gaussian estimate of CTE at the daily frequency



The three figures show the linear equivalence between the Gaussian estimates of Contemporaneous Transfer Entropy (CTE) $I(\mathbf{X}_t; \Delta S_t | \Delta S_{t-1})$ and R^2 of the linear regression $\Delta S_t = \alpha + \beta \mathbf{X}_t + \epsilon_t$ at the daily frequency. The factors \mathbf{X} are market order flow (*mo*), interest rate differential (*IR*) and their combination (*moIR*). The coordinates are the Gaussian estimates of CTE and linear regression R^2 reported in table 3.2.

Chapter 4

Transient and Permanent Jumps in the High Frequency Data

4.1 Introduction

In recent years, market participants and media often complain that the financial markets have become more and more unstable, and accuse the high frequency traders of causing it, especially after the Flash Crash, which the U.S. financial markets experienced on May 6, 2010. The heated debate is related to the issue of market quality and stability. There are two pressing questions need to be investigated in a scientific and objective way: whether the financial markets have become more stable or unstable, and if the high frequency trading is responsible for it? The extreme transient price jumps reflect the unnecessary price and market instability. Their prevalence is a manifestation the low market quality. In this paper, we will answer these questions by empirically identifying all the transient and permanent jumps in the U.S. common shares of recent twenty years, then aggregate the data into jump-related quantities and study the evolution of these properties over time, in the presence of several separate cross-sections of firms.

The stock price can be viewed as a process consisting of two components: continuous price movement in addition to a discontinuous jump component. The jump component of the stock price can be either permanent or transient. The permanent ones are caused by arrivals of new

information: such as the press releases from the Federal Reserve, earning announcements of the companies, and critical political events, etc. Theoretically speaking, when the relevant news hit the markets, the asset price will incorporate the news immediately, such that the price reflects all the available information (Fama 1970, Malkiel 2003). At the moment of news arrival, the asset price would adjust accordingly and jump to the new fair value of the asset. In contrast, the transient price jump does not result from the arrival of any relevant information. The dramatic oscillation of stock price does not reflect the changes in the fundamental value of underlying asset price. Therefore the disturbance in the price should not be permanent. The transient jump is actually the extreme case of price instability. A notable example of the transient jumps is the 2010 Flash Crash, during which S&P 500, Dow Jones Industrial Average and Nasdaq Composite, collapsed and rebounded very rapidly. The whole process lasted for approximately 36 minutes, after which the market indexes roughly recovered to their original values.

While permanent jump is resulted from information arrival, the transient jump without any fundamental reason is the kind of risk that bothers market participants, especially when they happen systematically across the markets, such as the Flash Crash. The existence of frequent or systematic large transient jumps indicates that the markets are fragile and dysfunctional. The price would deviate significantly from the market consensus all of a sudden for no apparent reason, so the price cannot reflect the fair value of the fundamentals. At the same time, the transient jumps can be seen as a source of unnecessary risk and price instability that the market participants abhor. Transient extreme price movements could harm the investors if the price that their trades locks in happens to be much worse than the fair value. The prevalence of transient jumps reflects low market quality, which would harm the investors' confidence.

The market participants are worried about an increase in so-called “mini flash crashes”, the dramatic transient price jumps in individual stocks. For instance, an article of from the media USA today writes (“Mini flash crashes worry traders”, USA today, May 17, 2011): *“Mini flash crashes still occur routinely with individual stocks ... Despite efforts to prevent another flash crash, the infamous day on May 6, 2010, when the Dow Jones industrials fell roughly 900 points, only to quickly recover, regulators and markets have moved to implement safeguards. Yet, traders and market observers are still seeing individual stocks and ETFs suffer flash-crash-like events, when stocks fall suddenly for no reason and quickly rebound, suggesting many of the*

underlying problems haven't been solved. ... For the first month and three days of 2011, stocks showed perplexing moves in 139 cases, rising or falling about 1% or more in less than a second, only to recover, says Nanex. There were 1,818 such occurrences in 2010 and 2,715 in 2009, Nanex says."

Therefore, we are motivated to conduct a thorough empirical study on the price jumps in the high frequency data, and further answer the question whether the markets have become more unstable in the era of high frequency trading. We are also very interested in finding out whether the increased instability is due to permanent or transient jumps. Because the increase in permanent jumps can be explained as more information events in recent years, however, the increase in extreme transient jumps can only be attributed to the deterioration of market stability.

Based on the Trade and Quote (TAQ) database, we identify all the permanent/transient jumps of the CRSP common shares in the period from January 1995 to December 2014. With the Flash Crash type of transient jump in mind, we detect transient jumps that last 5 minutes to one hour. The econometric procedures of the jump test and permanent/transient classification will be discussed in detail later. For empirical investigation, the intraday detected jumps are aggregated to monthly frequency. In each month, the stocks are grouped into four buckets (quartiles) according to the value of the chosen characteristic. For each bucket (quartile), we calculate the mean values of the number of jumps (permanent and transient), the number of jump days, and the percentage of jump stocks. As a result, we find that the properties of detected jumps depend on stock characteristics, especially the market microstructure related quantities. More importantly, we document a structural change in the jump statistics around 2003. It coincides with the introduction of auto-quote, which greatly reduces the trading latency and made high frequency trading much easier.

Why does the implementation of auto-quote matter? First of all, we should have some understanding of what auto-quote is and its meaning to the high frequency trading. Prior to 2003, the specialists on the NYSE were responsible for manually disseminating the inside quote. The manual procedure would slow down the speed of algorithmic traders: the latency was quite high. But started from early 2003, the traditional manual quote was replaced by a newly automated quote system. From then on NYSE electronically updates any change to the NYSE limit order

book. This market structure change provides much quicker feedback to trading algorithms, thus made the high frequency trading easier. Hendershott, Jones, and Menkveld (2011) document that auto-quote is associated with a significant increase in their proxy for algorithmic trading activity. In early 2003, NYSE gradually phases in the auto-quote among all the stocks, 515 of them adopted auto-quote in the first three months of 2003, whereas for the rest of the stocks, the auto-quote started on May 27, 2003. The gradual implementation of auto-quote on the NYSE can be used to study the effects of high frequency trading on the stability of the markets.

We document that the jump properties are heterogeneous across different characteristic-based buckets of stocks. Meanwhile, there is an evident structural break in jump statistics around the start of 2003, which coincides with the auto-quote introduction. General speaking, the stocks with low market cap and volume, large bid-ask spread and relative tick size experience more jump after 2003, in contrast, they are the relatively more stable stocks before auto-quote. Notably, most of the increased jumps are transient ones, which are not information-driven. It means that the worsening price stability is not due to more frequent market-relevant news after 2003. The sizable increase in transient jumps actually reveals the deterioration of market quality for these small-cap thinly-traded low-priced stocks.

On the other hand, it is a completely different picture for the stocks on the other end of the spectrum: stocks with high market cap and volume, small bid-ask spread and relative tick size, have actually become more stable after 2003. In the era of high frequency trading, the large-cap stocks with active trading are less prone to jumps. This observation is in line with the empirical literature that high frequency trading improves the market liquidity for those stocks priced higher than five dollars (Hendershott, Jones, and Menkveld 2011). In conclusion, whether the stock price is prone to transient jump in the new era of electronic trading depends on the stock characteristics.

Furthermore, as we have documented that the structural change in jump properties coincides with the auto-quote introduction, it is to natural ask whether the high frequency traders should be held accountable for the increase or decrease in the jumps. HFTs could play a role in these transient price jumps for several reasons (Biais and Foucault 2014). First, because of the similar trading strategies, HFTs may all react at the same time to erroneous signals by sending buy or sell market orders consuming market liquidity, triggering sharp price movements. Alternatively,

after the arrival of a large sell (buy) market order, the limit orders may all got canceled by HFTs for safety reasons, and if the large market order does not appear to be informationally motivated, HFTs would resubmit new limit orders very quickly. Moreover, as the endogenous liquidity providers, HFTs may withdraw their quotes and stop marking the market in the adverse market conditions. In any case, waves of cancellations or market orders submissions by HFTs reacting to the same event may exacerbate the transient price jumps. For the thinly-traded small-cap stocks, traditional market makers are crowded out by HFTs after the great latency reduction in 2003, and as endogenous liquidity providers, HFTs have the less long-term risk-bearing capacity and they might withdraw liquidity or stop making the market under adverse market conditions. Thus the thinly-traded small-cap stocks is more prone to frequent transient jumps or mini flash crashes. We will go into this discussion in detail later.

To study these research questions more rigorously, we run panel regressions around the event of auto-quote implementation on the NYSE around early 2003. By fitting a linear model with dummy variables of auto-quote and quartiles as well as their interaction terms to the data, we can estimate the number of jumps (permanent/transient) and jump days for each quartile before and after the auto-quote introduction. The panel data regressions confirm the results obtained previously: the structural shift in early 2003 and the heterogeneous jump properties across different characteristic-based buckets of stocks.

More importantly, when the stocks are grouped into quartiles according to the quote-to-trade ratio, the regression results indicate that the stocks in the highest quote-to-trade ratio quartile have become much more unstable after auto-quote, and most of the increased jumps are transient. As we all know that one of the distinguishing features of HFTs is the high quote-to-trade ratio, because HFTs would submit and quickly cancel their quotes at a very higher frequency, the actual trades are merely a very small fraction of the total number of quote posted by the HFTs. The quote-to-trade ratio can be used as a proxy for HFT (Hagstromer and Norden, 2013, Friederich and Payne, 2015). Similarly, the message-to-trade ratio proxy is also used (Hasbrouck and Saar, 2013). The stock with high quote-to-trade ratio has a higher percentage of HFT participation. Therefore, the dramatic increase in transient jumps for the high quote-to-trade ratio stocks after auto-quote suggests that HFTs cause more extreme price movements.

Another supporting evidence is that we also document that the quartile with the greatest

relative tick size become much more unstable: the number of jumps in these stocks, especially the transient ones, have increased considerably after auto-quote. Considering that the large relative tick size is associated with a high percentage of HFT participation (Yao and Ye 2014), this empirical evidence also implies that HFTs cause more transient jumps in these low-priced stocks.

Overall speaking, the empirical evidence points to the fact that HFTs increase jumps in the small market cap, low volume and market liquidity, and large relative tick size (low-priced) stocks. The stocks with high HFT participation (proxied by high quote-to-trade ratio, and large relative tick size) are more prone to transient jumps. As for the large market cap stocks with high volume and market liquidity, their price stability is actually improved considerably. However, due to lack of high frequency trading data, our indirect evidence cannot completely pin down the association between the increased price instability and HFT activities. Nonetheless, the empirical fact of market stability deterioration for small illiquid low-priced stocks is iron-clad, regardless whether it is caused by HFT or not.

One may wonder why HFT could have different effects on large-cap, high volume and liquidity stocks versus small-cap, low volume and liquidity stocks. First of all, the daily dollar trading volume of large-cap, high volume stock is thousands of times greater than that of the small-cap illiquid stock, which is mainly traded by retail investors through online brokerage accounts. For the heavily-traded large-cap liquid stocks, when the trading latency decreases due to auto-quote, the trade size decreases and the number of trades increase, meanwhile, the non-stop trading could take place at a finer and finer time scale (milliseconds or higher) as the trading technology advances. Thus the price at the macroscopic time scale would become more stable compared with the pre-auto-quote time when the lumpiness of supply and demand for liquidity happens at the scale of minutes. On the other hand, for thinly traded small-cap stocks, the liquidity traders arrive at a much lower rate and the daily trading volume is low. As the latency greatly reduced by auto-quote, slow market makers would be driven out of the market due to much higher exposure to the pick-off risk, because the liquidity traders still arrive at the same rate, but now slow market makers would be picked off more frequently by HFTs as the trading latency reduced. Another contributing factor is that in the pre-auto-quote time, there are market specialists with long-term risk bearing capability maintain the market liquidity and

continuous trading. In the new era of electronic trading, they are replaced by endogenous liquidity providers (HFTs).

There are the other factors that change over time, e.g., Regulation NMS (order protection) rules, fragmentation, etc. Can these changes be another explanation for increased jumps? The answer embeds in the empirical evidence, as we will observe later, the structural change in jump statistics happens around the beginning of 2003, which coincides with the auto-quote implementation. Regulation NMS Trading Phase Date occurs on February 5, 2007, and is fully implemented by October 15, 2007. The market fragmentation mainly happened after 2007 Regulation NMS. SEC (2013) says: “prior to the implementation of Regulation NMS in 2007, the market for NYSE-listed stocks was highly centralized, with the NYSE executing 79% of volume in its listings. The remaining 21% was executed primarily off-exchange by broker-dealer internalizers... After Regulation NMS, the NYSE’s market share in its listings declined from 79% in 2005 to 25% in 2009, while the total volume in NYSE-listed stocks during this period increased by 181%”. As for Nasdaq, the time line of market fragmentation is similar. For those interested in the details, please refer to SEC (2013). While these market changes could have some influences on the jumps as well, none of them coincides with the fundamental structural shifts in jump dynamic around the start of 2003, which is the focus of this study.

This paper is the first one that documents the cross-sectional variation in the effects of algorithmic trading on transient price jumps. In particular, while the current market making system is good for majority of the stocks, the thinly-traded small market cap, low-priced stocks would benefit from the traditional specialist market making system, where the liquidity providers are the exchange-regulated market makers subjected to affirmative obligations, such as requirements for continuous liquidity provision on both sides of the market.

4.2 Physical Clock versus Volume Clock

We recognize that the volume clock is useful in analyzing certain problems, for instance, VPIN (Easley, Lopez de Prado, and O’Hara 2012), and trade execution (Easley, Lopez de Prado, and O’Hara 2015). However, testing jumps using price time series under volume clock may not be a sensible choice, for the following reasons:

First of all, one difficulty of using volume clock comes from the cross-sectional heterogeneity of the volumes. In our data sample, there are liquid large market cap stocks such as Coca-Cola (NYSE: KO), as well as low-priced small-cap stocks, such as JMP group (NYSE: JMP). The volume of Coca-Cola is more than 500 times greater than JMP. We are not sure if it makes sense to sample the prices of these two stocks based on same volume bucket, and compare the volume-clock-based jumps of Coca-Cola and JMP. To be more specific, if we set the volume clock in such a way that there are 360 volume minutes per calendar trading day for Coca-Cola, its price is sampled at the minute frequency. Using the same volume bucket as the clock, one calendar trading day of JMP is only considered as less than one volume minutes since the volume of JMP is 500 times smaller. As a consequence, the price of JMP would be sampled at the daily frequency. It completely contradicts our purpose of investigating the “flash crash” or “mini flash crash” type of jumps. In this case, we essentially focus on testing the jumps in high volume stocks while overlooking the jumps in thinly-traded stocks.

One may consider using different volume buckets to time Coca-Cola and JMP separately. But now the concern becomes what is the sensible choice? It could be a quite arbitrary decision. Moreover, if one needs to choose different volume buckets for each unique stock in the U.S. markets, making economic sensible and convincing choice would be a challenge.

In our empirical analysis, the detected jumps in different stocks are compared cross-sectionally based on stock characteristics. Because we are aware that the trading protocols, rules, and technology are quite different now and ten years ago. Cross-sectional comparison mitigates the issue of varying market conditions over the years. At any point in time, the cross-sectional comparison among different characteristic buckets should be reliable.

Second, volume clock shifts out focus on the jumps in these high volume periods while overlooking the jumps in normal market condition and the liquidity dry-up periods. Let’s consider the market condition when the market liquidity dries up. At the moment, trading volume is low, while the asset prices become very unstable. As the market liquidity dries up, the price instability can be caused by the unusually large price impact of the trade, or by the canceled quotes as market makers are withdrawing their liquidity provision. In these scenarios, price jumps are in companion with thin trading volume. If the clock based on volume bucket is adopted, we would have slowed the time, and as consequence, the extreme price movements during liquidity

dry-up periods would look more benign. From a different perspective, during liquidity dry-up periods, since volume is low, we would downsample the price data points comparing to the normal times (we would sample the data points at a much lower frequency because the volume clock runs slower than the physical clock). As a result, there will be fewer jumps identified using volume clock because there are fewer data points sampled during these extreme periods. On the other hand, there are also very volatile times with high trading volume. If the volume clock is adopted, we essentially focus on the jumps during high volume periods while overlooking the jumps in normal market condition and the liquidity dry-up periods. But this deviates from our aim to investigate all the transient and permanent jumps in the U.S. stocks.

What's more concerned is that by dilating or shrinking the local time frame may completely change the nature of the identified jumps. For instance, the econometric test identifies a jump if there is a large sharp change in price relative to the local volatility. However, if the local time frame is dilated (stretched), i.e., the volume clock runs much slower than the physical clock, the sharp price change would look much milder and not so different from random walk. Then the jump based on physical clock would not be classified as a jump using volume clock.

Third, testing the jump based on volume clock would run into some econometric difficulties as well. As the current jump tests are all based on the price movements with the physical clock. We are not sure if there is an econometrically sound way to test price jumps when the price is timed with the volume-based clock.

Therefore, we think it makes more sense to simply using the physical clock, because this paper aims to investigate the market stability from an average investor point of view. The time experienced by the average investors is the physical time, not volume-bucket time artificially introduce by researchers. Let alone to say, the actual meaning of volume clock is still a very debatable issue (Andersen and Bondarenko 2014a, 2014b).

4.3 Literature Review

The research that is closest to ours is Gao and Mizrach (2016), which is recently referred to us by Prof. Richard Payne. They study the breakups and breakdowns of the limit order book.

To be more specific, they examine stocks whose bids (asks) move more than 10% at the NBBO between 09:35 and 15:55 but recover within the trading day. There is a breakdown (or breakup) of the limit order book happen if the national best bids (asks) fall (rise) 10% below (above) the 09:35 price, and rebound within 2.5% of the 09:35 price at 15:55. They find that market quality (in terms of the limit order book breakups and breakdowns) has improved since the implementation of Reg.NMS, since mid-October 2007.

Although Gao and Mizrach (2016) share some similar starting points as our study, i.e., the market stability issue, their research is completely different from ours in the following ways. First of all, Gao and Mizrach (2016) interest in the breakdowns/breakups of the limit order book while we are interested in the transient jumps that last from 5 minutes to 1 hour. To be more specific, Gao and Mizrach (2016) study bids and asks separately for identifying breakdowns in bids and breakups in asks. They focus on the dysfunction of the limit order book and scrutinize every change in the NBBO. In contrast, our research aims to differentiate transient price jumps from instantaneous volatility and market microstructure noise: we sample mid prices at a lower frequency (2.5 minutes) and mitigate the market microstructure noise with pre-averaging procedure. Second, their definition of breakdown (breakups) does not differentiate transient extreme price movements from high volatility: when simulating the whole day's price path using random walks, one would generate many realizations of price paths that satisfy the breakdown/breakup conditions, especially when the volatility is high. In other words, the bids "breakdown" to a level lower than 10% of 09:35 price and come back within 2.5% of the 09:35 price do not necessarily mean dysfunction or instability of the markets, although it does contain the events of market instability as its subset. While our transient jump test aims at identifying the jumps using econometrically rigorous method, then test if the subsequent price movements after the jump come back inside the volatility cone within 1 hour. Last but not least, although the sample period of their data (1993-2013) is similar as ours, the conclusions of their study are quite different from ours: Gao and Mizrach (2016) reports the breakdown frequency in all U.S. stocks aggregately, while we document that the jump properties depend on stock characteristics. Moreover, the event in their study is the implementation of Reg.NMS in October 2007, while we focus on the implementation of auto-quote in early 2003, around which the major structural changes in jump properties happened.

With respect to other relevant studies, Hendershott, Jones, and Menkveld (2011) use the implementation of auto-quote on NYSE in 2003 to study the effect of algorithmic trading on market liquidity. They find that the market liquidity, the quoted and effective bid-ask spread, improves after the introduction of auto-quote for the large market cap stocks, while there is no significant effect on market liquidity for small market cap stocks. Additionally, they show the reduction in trading cost is driven by a reduction in adverse selection component of the bid-ask spread. Note that Hendershott, Jones, and Menkveld (2011) exclude the stocks with price smaller than 5 dollars, for which we find that the transient jumps increase dramatically after 2003. Considering that the low-priced stock is usually the stocks with small market capitalization and low market liquidity, the dataset of Hendershott, Jones, and Menkveld (2011) excludes these stocks. Boehmer, Fong and Wu (2012) obtain similar findings based on a wide range of countries. Interestingly, they find cross-sectional variations in the effect of algorithmic trading (AT): while AT improves liquidity and informational efficiency for the large market cap and high-priced stocks, greater AT reduces liquidity and worsens the volatility for the smallest capitalization stocks. In line with Boehmer, Fong and Wu (2012), we also document completely different effects of algorithmic trading on the large cap high-priced stocks and small cap low-priced stocks. The difference is that Boehmer, Fong and Wu (2012) focus on market liquidity and volatility, while our paper is about market stability, especially the transient price jumps that are not information driven.

A related topic is the effect of algorithmic trading (AT) on volatility. However, the empirical findings are rather mixed. Hasbrouck and Saar (2013) use the number of linked messages as a proxy for algorithmic trading and find a negative effect of AT on volatility. Using the ban on short-sales in the U.S. markets for about three weeks in September and October 2008, Brogaard (2011) finds a negative effect of HFT on volatility. However, Boehmer, Fong and Wu (2012) document a positive association between their measure of algorithmic trading and volatility, based on their international sample of stocks.

Brogaard et al. (2016) study the extreme price movements and the behavior of HFTs. The extreme price movements (EPMs) are defined as the 10-second absolute midquote returns that belong to the 99.9th percentile of the return distribution. Brogaard et al. (2016) find that HFTs provide liquidity during extreme price movements (EPMs) by absorbing imbalances created by

non-high frequency traders (nHFTs). But this observation is limited to EPMs in single stocks. When several stocks experience simultaneous EPMs, HFT liquidity demand dominates their supply.

Menkveld and Zoican (2017) investigate the case when the trading platforms reduce their latency, the market liquidity could be hurt because the market makers are more likely to meet the high frequency bandits and less likely to meet the liquidity traders. The argument is relevant to our study on market stability. The implementation of auto-quote on NYSE reduces latency, thus the manual market makers, who have greater long-term risk-bearing capability are more likely to meet the high frequency arbitragers (bandits). This creates an adverse selection problem for the slow market makers who would be crowded out of the market by HFTs. In the extreme market conditions, the absence of slow market makers could impair the resilience of the market (Biais and Foucault 2014, and the references therein).

4.4 Jump Test

In this section, we briefly introduce the related literature on jump tests first, then discuss the econometrics of the jump test used in this paper. Most of the previous studies consider testing jumps with the low frequency data, for which the market microstructure is irrelevant. Aït-Sahalia (2002) proposes a test based on the transition function of the stochastic process. Carr and Wu (2003) introduce a method based on short dated options. Barndorff-Nielsen and Shephard (2006) use the realized variance and the realized bi-power variation to identify days with jumps. Jiang and Oomen (2008) propose a method based on the variance swap contract and the higher order return moments to identify days with jumps. Lee and Mykland (2008) test the large increment relative to the local volatility. Aït-Sahalia and Jacod (2009) sample the power variations at different frequencies in order to test the jump. More recently, Christensen, Oomen and Podolskij (2014), and Lee Mykland (2012) propose nonparametric jump tests which allow us to asymptotically remove the market microstructure noise and discover jumps in the high frequency data. However, these jump tests do not differentiate whether the jump is permanent and transient. In this paper, we further extend the pre-averaged Lee Mykland test in Christensen, Oomen and Podolskij (2014) to classify the jump as either transient or

permanent.

The nonparametric jump test of Lee and Mykland (2008) detects the existence of a jump at time t_i , without making assumptions about whether there are jumps before or after t_i . The intuition of the LM jump test is as follows. At the time of the jump, we would expect the stock price has a very large realized return, much greater than usual continuous innovations. But since the data is sampled at the discrete time, we need to differentiate whether the large realized return is due to high spot volatility or it is actually a jump. To distinguish those two cases, the return at time t_i is standardized by *instantaneous volatility*, which measures the local variation from the continuous part of the process. To be more specific, the statistic of the jump test is the ratio of realized return to instantaneous volatility. If the test statistic is greater than the selected threshold (test statistic), the price change in the current period is large enough relative to the instantaneous volatility, the price movement should be identified as a jump.

In the presence of the jumps, the traditional variance estimator, realized variance (or quadratic variation), is not good for estimating the integrated variance of the underlying continuous part of price process. Realized variance includes the integrated variance as well as the jump variation. A consistent estimator for the integrated variance is called the realized bipower variation, which is robust to the large or small jumps added to the diffusive part of the process. (Barndorff-Nielsen and Shephard, 2004; and Aït-Sahalia, 2004). The realized variance is equal to the sum of squared returns, while the realized bipower variation is calculated by taking the sum of products of consecutive absolute returns, as in equation (4.2). Since jumps are unlikely to occur in two consecutive intraday periods, when intervals are small enough, the realized bipower variation will converge to the integrated variance asymptotically. The difference between realized variance and realized bipower variation estimates the quadratic variation of the jump component. In other words, the quadratic variation of the process can be separated into its continuous and jump components.

The instantaneous volatility at time t_i is estimated similarly as the realized bipower variation, using a local movement of the process within a window size K prior to time t_i . The window size K is chosen in such a way that the effect of jumps on the volatility estimation disappears, on the other hand, the window size K should not be too large, since the estimates of the instantaneous volatility should reflect the local behavior prior to t_i . Lee and Mykland (2008) suggest using

a K value of $\sqrt{252 \times N}$, where N is the number of the observations within a day. The test statistic of LM jump test is

$$L(t_i) = \frac{\log S(t_i) - \log S(t_{i-1})}{\hat{\sigma}(t_i)}, \quad (4.1)$$

where $\hat{\sigma}(t_i)$ is the instantaneous volatility at time t_i , which is the square root of the instantaneous variance estimated with the realized bi-power variation over a historical window of K observations prior to time t_i :

$$\hat{\sigma}^2(t_i) = \frac{\pi}{2} \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} |\log S(t_j) - \log S(t_{j-1})| |\log S(t_{j-1}) - \log S(t_{j-2})|. \quad (4.2)$$

Following Lee and Mykland (2008), the rejection threshold of $\xi \equiv \frac{|L(t_i)| - C_n}{S_n}$ is β^* , such that $P(\xi \leq \beta^*) = \exp(-\exp(-\beta^*))$, where $C_n = \frac{(2 \log n)^2}{c} - \frac{\log \pi + \log(\log n)}{2c(2 \log n)^{0.5}}$ and $S_n = 1/(c(2 \log n)^{0.5})$, and $c = \sqrt{2}/\sqrt{\pi}$. For instance, for 1% significant level, $\beta^* = -\log(-\log(0.99)) = 4.6001$. If $\frac{|L(t_i)| - C_n}{S_n} > 4.6001$, the null hypothesis of no jump at t_i is rejected.

When testing the jumps using the high frequency financial data, one needs to distinguish jumps in efficient prices from microstructure noise. The microstructure noise is induced by the frictions such as the discreteness of the prices, bid and ask bounce, the discreteness of the observations, and other trading mechanics (O'Hara 1995, and Hasbrouck 2007). The presence of microstructure noise has a negative impact on the previously described jump tests (Andersen and Benzoni, 2009). It is important to distinguish the efficient price jumps from noise. Recently, Christensen, Oomen and Podolskij (2014) propose a test to detect jumps in high frequency data by combining pre-averaging with the jump test of Lee and Mykland (2008). Pre-averaging is useful to smoothen the impact of microstructure noise in the high frequency data. Jacod et al. (2009) and Podolskij and Vetter (2009) propose the local averaging procedure, which asymptotically removes the noise and approximate the true underlying prices.

In this paper, the pre-averaged return is averaged in a local neighborhood with K^* observations, and $K^* \geq 2$ and even. Christensen, Oomen and Podolskij (2014) suggest to use $K^* = \theta\sqrt{N}$ and the value of θ should be in the range of 0.1 to 2. In our empirical jump test, we set $\theta = 0.4$

and $K^* = 4$. The pre-averaged return is calculated using the following equation:

$$r_{t,i,K^*}^* = \frac{1}{K^*} \left(\sum_{j=K^*/2}^{K^*-1} \log S_{t-i-j} - \sum_{l=0}^{K^*/2-1} \log S_{t-i-l} \right). \quad (4.3)$$

The LM pre-average statistic $L^*(t_i)$ is calculated similarly as before:

$$L^*(t_i) = \frac{r_{t,i,K^*}^*}{\hat{\sigma}_{t,i,K^*}^*}, \quad (4.4)$$

where $\hat{\sigma}_{t,i,K^*}^*$ is the square root of the instantaneous variance:

$$\hat{\sigma}_{t,i,K^*}^{*2} = \frac{\pi}{2} \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} |r_{t,j-K^*,K^*}^*| |r_{t,j,K^*}^*|. \quad (4.5)$$

Note that the instantaneous variance $\hat{\sigma}_{t,i,K^*}^{*2}$ is also calculated using the pre-averaged return. The reject threshold of the pre-averaged statistic $L^*(t_i)$ is the same as the previous case. The test statistics follow the same distribution. The only difference is the additional pre-average procedure before constructing the test statistic.

We identify the jumps using the pre-average LM test, then further classify the detected jumps into permanent or transient ones. Differentiating transient jumps from the permanent ones is important because our focus is the market instability in the form of transient extreme price movements, which are not information driven. For stock i , at each time point t , firstly we apply the pre-average LM test to examine the return time series and label all the jumps. For each detected jump, we check if its subsequent price path reverts back inside the diffusion region (volatility cone) of the initial price $S_{i,t}$ within a certain period of time. The diffusion region is the price range that is reachable by the continuous price process following the geometric Brownian motion. We set the diffusion region of the price as $(S_{i,t} - \hat{\sigma}_{t,i,K^*}^* \sqrt{T-t}, S_{i,t} + \hat{\sigma}_{t,i,K^*}^* \sqrt{T-t})$, where $S_{i,t}$ is the log price right before the jump, $T-t$ is the interval from time T to the reference point of the jump, t . At the reference point of the jump ($T=t$), the size of the diffusion region is zero. As time passes, the diffusion region grows in proportional to the instantaneous volatility and square root of the time interval.

In this paper, we define the maximum duration of the transient jump as one hour. It means

that after the initial jump, if the price would come back and enter the diffusion region within one hour, then the jump is labeled as transient, otherwise it is permanent. Intuitively, if the stock price has jumped and keep staying at the new price level without coming back to the pre-jump level in a certain period of time, the jump should be labeled as a permanent jump. On the other hand, it should be taken as a transient jump if the price reverts back to the small neighborhood of the initial price level.

Of course, the one-hour duration cutoff is a relatively subjective choice. However, the maximum duration cannot be set too long, since the diffusion region of the random walk grows as the square root of the interval. If the interval is set as a very large value, the diffusion region would be so large that every jump's subsequent path would enter the diffusion region, merely because of the randomness. In this case, all permanent jumps would be falsely identified as transient. From a different point of view, one-hour interval cutoff is appropriate since this paper is interested in the flash crash type of transient jumps, which have the duration of 5 minutes to 1 hour. With respect to the jumps at much higher frequency, as we will elaborate in the next section, for various reasons, these high frequency price oscillations are not the subject of this study.

Another issue with the one-hour cutoff of transient jump duration is that some jumps that are more transient in nature may be classified as permanent jumps. For instance, if the price reverts back to diffusion region around the pre-jump price level in 62 minutes, because it is longer than one hour, the jump would be labeled as permanent. Clearly, this would be a misclassification. This misclassification is inevitable. What's need to be pointed out is that misclassifying transient jumps as permanent would only weaken our claim that the majority of the increased jump are transient. In other words, the potential misclassification implies that the actual results could be slightly stronger than the results reported here. This being said, we document the supporting evidence for the successful classification of permanent and transient jumps: the number of permanent jumps per jump day is invariant across time and stock characteristics. We will discuss this in the section on jump statistics.

4.5 Data

Using the pre-average LM test specified in the previous section, we detect all the transient and permanent price jumps of CSRP stocks based on WRDS NBBO data from January 1995 to December 2014. The WRDS NBBO (National Best Bid and Offer) stores all the output datasets that contain the NBBOs at each second for all stocks with quotes data in TAQ. The NBBO is updated throughout the day to show the highest bids and lowest offers for a security among all exchanges and market makers. We calculate the mid-price based on the NBBO. The mid-price does not have the bid-ask bounce of the trade price, hence it can mitigate the market microstructure noise.

This paper is not about the jumps at the frequency of seconds or tick by tick. Because the movements of quoted prices at very high frequency are more relevant for high frequency traders. The transient jumps at the frequency of minutes or longer are more important or ‘observable’ to the average investors and non-HFT market participants. Moreover, the estimator of the realized variance is biased when sampling the price at a high frequency due to the existence of market microstructure noise. Therefore, prices are often sampled at a lower frequency. Andersen and Benzoni (2009) suggest that the volatility signature plot is useful for determining an optimum sampling frequency. The volatility signature plot of SPDR S&P 500 ETF (SPY) tells us that the sampling frequency should be around 30 seconds at least. Meanwhile, we also need to consider the feasibility of the computational task when carrying out the empirical exercise. As we need to process 20 years worth of TAQ data on U.S. equities, it would be computationally expensive if the data is sampled at a very high frequency. Considering all these factors and we are interested in the transient jumps with the duration of 5 minutes to 1 hour, the NBBO is sub-sampled at the sampling frequency of 2.5 minutes. In such a way, the minimum duration of the detected transient jump would be 5 minutes. Because for the transient jump that jumps away from the pre-jump price level then immediately reverts back to the neighborhood of the pre-jump price, the duration of the whole process (jumps away and reverts back) is 5 minutes. As for the choice of 1 hour maximum duration, please refer to the discussion in the previous section.

Following Lee and Mykland (2008), the window size for the instantaneous volatility estimation

is chosen as $K = \sqrt{252 \times N} = 199$, where N is the number of observations in a trading day. It means that to test the jump at time t_i , we will use the 199 observations prior to t_i to estimate the instantaneous volatility. The pre-average window is set as $K^* = 4$. Considering that we are interested in identifying the extreme price jumps, so the confidence level is set as $\alpha = 0.999$, equivalently, the rejection threshold for the statistic $L^*(t_i)$ (equation 4.4) is 4.9166. Intuitively, if the pre-averaged return at time t_i , r_{t,i,K^*}^* , is 4.9166 times of the pre-averaged instantaneous volatility, then the price movement is too drastic relative to the local volatility such that it should be classified as a jump.

We only consider the jumps in the intraday price movements: the gap between the opening price and the closing price of two consecutive days is not considered as a jump. Moreover, the jumps in the first half hour of the trading day are not taken into account, because the price movements tend to be very volatile in the first half hour. For the detected jumps, we document the stock symbol, date, test statistic, the type of jump (transient or permanent), the beginning time of the jump, and come-back time if it is transient. We further aggregate the jump data into monthly frequency and calculate the number of jumps, transient jumps and permanent jumps, which are denoted as $n_{i,t}^J$, $n_{i,t}^{tran}$ and $n_{i,t}^{perm}$, respectively, as well as the number of jump days $D_{i,t}^J$. The subscripts i, t represent that the variable is for stock i and month t .

To study the cross-sectional properties of the jumps, we also need the relevant stocks characteristics. Hence, we merge the dataset of detected jumps based on TAQ with other datasets such as CRSP, Compustat, I/B/E/S, etc. First, we link every symbol in the TAQ jump dataset with its corresponding CUSIP via the TAQ Master file. Then, based on the matched CUSIP and date, we merge the jump dataset with the CRSP and compustat, etc. Furthermore, we only keep the common shares with the share code equals to 10 or 11 (shrcd=10, 11). The following stock characteristics are collected: price (relative tick size), volume, market capitalization, bid-ask spread percentage, market depth, quote-to-trade ratio, book to market ratio, market beta, institutional holding, short interest, analyst coverage, bi-power variation, and the exchange code.

The relative tick size is relevant because the percentage of high frequency trading volume increases with the relative tick size: the larger the relative tick size, the more participation of the high frequency trading (Yao, Ye 2014). Intuitively speaking, the HFTs prefer to compete

in speed rather than in price (by undercutting the price), because of their speed advantage against the traditional market makers. The large relative tick size gives the HFTs an advantage to compete in speed: the traditional market makers cannot compete by improving the quoted price by a fraction of the tick size, because tick size is the minimum increment in price; on the other hand, the larger the relative tick size, the greater the rent HFT can extract from its speed advantage. The stock price is inversely related to the relative tick size. Therefore, the percentage of HFT volume is positively related to the relative tick size, and negatively related to the stock price. The relative tick size is an important characteristic from the market microstructure point of view.

With respect to other market microstructure related characteristics, the volume is the total number of shares of a stock traded during that month. The market capitalization is calculated as the shares outstanding times the stock price. Moreover, as for the liquidity measure, we calculate the quoted and effective bid-ask spread percentage using the TAQ data, using the method in Holden and Jacobsen (2014). The market depth is defined as the average of the dollar ask and bid depths. The dollar ask (bid) depth is the dollar amount available at the best ask (bid) quote from the exchange or market maker with the largest size quoted at that price. The quote-to-trade ratio is the number of quotes divided by the number of trades. Because HFT strategies involve frequent cancel-and-replace quote traffic, the number of quotes that leads to a trade is typically much greater for HFTs than for non-algorithmic traders. The quote-to-trade ratio is used as a proxy for the HFT in the literature (Hagstromer and Norden, 2013, Friederich and Payne, 2015). Moreover, the data set also contains the exchange codes, which use 1, 2, and 3 to represent NYSE, AMEX, and Nasdaq, respectively.

The second category is the traditional asset pricing characteristics, such as book-to-market ratio, market beta, institutional holding, short interest, analyst coverage, volatility (bi-power variation). The book-to-market ratio is a financial ratio used to compare a company's current market value to its book value. The market beta of the stock is estimated using a rolling window based regression. To estimate $\beta_{t,i}$ of stock i at month t , we regress the return of stock i on the market's return, using the twenty-four-month data prior to current month t . For institutional holding, we calculate the percentage of institutional holding based on the Thomson (Thomson Reuters) institutional holdings data. The short interest data is the SHORTINT variable of

compustat database. It reflects the short positions resulting from short sales. Short selling is the selling of a security that the seller does not own, or any sale that is completed by the delivery of a security borrowed by the seller. Furthermore, we use two variables to describe the analyst coverage: the number of analysts following a stock, reported in the NUMEST variable of I/B/E/S, and analyst forecast dispersion, which is the standard deviation calculated on the forecast issued during the month, labeled as *stdev*. Bi-power variation is the square root of the pre-averaged instantaneous variance (equation 4.5), which captures the variation of the continuous price component.

To show the cross-sectional variation in the jump properties, we group the stocks into 4 buckets based on the characteristics discussed below. The market capitalization quartiles are constructed according to the Fama-French monthly market equity (ME) breakpoints, which are available from Kenneth French's data library. The stock price is inversely related to the relative tick size. For the price buckets, the stocks are grouped into four categories based on the following price ranges: smaller than 5 dollars, 5 to 25 dollars, 25 to 50 dollars, and greater than 50 dollars. With respect to the following characteristics: volume, effective/quoted bid-ask spread percentage, market depth, quote-to-trade ratio, book-to-market ratio, market beta, institutional holding, short interest, bi-power variation, we categorize the stocks into four equal-sized groups based on the respective characteristic in each month. Last but not least, for the analyst coverage, the stocks are categorized into four buckets based on the following criteria: if the number of analyst forecasts is smaller than 3, greater than or equal to 3 and smaller than 6, greater than or equal to 6 and smaller than 11, greater than or equal to 11. For each characteristic, the buckets of stocks are labeled from 0 to 3, among which quartile 0 has the lowest value of the corresponding characteristic while quartile 3 has the greatest. For example, stocks in the quartile 0 of the market capitalization have the smallest market cap, and those in the quartile 3 have the greatest market cap.

4.6 Jump Statistics

Before carrying out the econometric analysis, it would be useful to visualize the statistics of the detected jumps and observe how they vary across different quartiles and evolve over time. Based

on the stocks with jumps, the following statistics will be studied: the total number of jumps per stock in a month (n^J), the number of jump days per stock in a month (D^J), the average number of jumps per jump day (\bar{n}_D^J), and the percentage of jump stocks in a month (p^J), which is defined as the number of the stocks with jump(s) divided by the total number of the stocks in a month. Moreover, it would also be very interesting to differentiate the transient/permanent jumps or positive/negative ones, and investigate their properties separately. It turns out that the statistics of positive and negative jumps are completely symmetric, so it would not be interesting to report them separately. On the other hand, the differentiation between transient and permanent jump is economic meaningful, as we have argued before, the transient ones are the flash crash type of price instability that worries the market participants, in contrast, the permanent jumps are caused by arrivals of the new information. The results of these jump statistics are aggregated into monthly frequency and shown in the figures 4.1 to 4.21.

Empirically, we document a clearly structural change in the jump properties across time. The structural change happened around 2003, which coincided with the introduction of auto-quote. Moreover, another interesting observation is that the properties of the jumps are heterogeneous across different characteristic buckets. General speaking, the stocks with large market capitalization and volume, small relative tick size, low quote-to-trade ratio, and high market liquidity (in terms of the bid-ask spread and market depth), become more stable after 2003. On the other hand, the stocks with small market capitalization and volume, large relative tick size, high quote-to-trade ratio, and low market liquidity, become much more unstable and experience more jumps after 2003. Moreover, prior to 2003, before the introduction of auto-quote, these small cap, low volume and market liquidity stocks are actually more stable compared with those on the other end of the spectrum. Note that for those stocks experienced a dramatic structural change in the jump properties, the majority of the changes comes from the transient jumps, which are the instability of the stock price, not driven by new information. In the following paragraphs, we are going to discuss these empirical results in detail.

For those small-cap, relative illiquid, low-priced and high quote-to-trade ratio stocks, the number of permanent jumps rise as well, as the number of transient jumps increases after 2003. We would like to argue that the increased permanent jumps are due to the fact that there are more information-driven events: because the empirical results on the number of jumps per

jump day (\bar{n}_D^J) show that the numbers of permanent jumps per jump day are roughly the same for different characteristic quartiles/groups, before and after 2003. It implies that one news event incurs the same number of permanent jumps, and the number of permanent jumps per information event is approximately invariant across time and stock characteristics. This is the supporting evidence for the successful classification of transient and permanent jumps. If the misclassification of transient jump as permanent is a severe issue, we would definitely see the number of permanent jumps per jump day rise with the number of transient jumps per jump day. Based on this conclusion, we can further infer that the number of information events/days increase after 2003 for these small-cap illiquid stocks (since the number of permanent jumps increases but the number of permanent jump per jump day does not change).

Now let's discuss the jump properties based on various stock characteristics. First of all, for the stocks in the lowest market cap quartile (*capt0*), the total number of jumps per stock (n^J) started to increase dramatically from 2003 and onwards, but prior to 2003, it had the smallest number of jumps compared to other large market cap quartiles. On the other hand, for the stocks in other three market cap quartiles, the number of jumps occurred in a month slightly decreased since 2003. When the jumps are separated into transient and permanent ones and studied separately, it is easy to tell that the majority of the increase in the total number of jumps comes from the increase in transient jumps, whereas the increase in permanent jumps is much smaller. Note that since we classify any jump with the duration longer than one hour as permanent, so it is possible that some jumps are more transient in nature are labeled as permanent. Hence there is an even greater increase in the number of transient jumps than it appears here. The claim that the increase in the number of jumps is mainly due to transient jumps is stronger than it seems.

Similar structural change is observed when we study the number of jump days per stock in a month (D^J). The ranking of D^J across different market cap quartiles reverses before and after 2003: at the beginning of our sample period, the large market cap stocks have the largest number of jump days (D^J) while the small market cap stocks have the smallest. However, from 2003 and onwards, the number of jump days increases considerably for the smallest market cap quartile. In contrast, the number of jump days decreases for the other three market cap quartiles, while the largest market cap quartile has the greatest reduction in the jump days.

Regarding the different types of the jumps, we can observe that, for the smallest market cap quartile, the number of transient jump days has a greater increase than the number of permanent jump days.

For the number of jumps per jump day in a month (\bar{n}_D^J), one can see that there is a steady increase in \bar{n}_D^J for the small market cap quartile after 2003, and the increase mainly comes from transient jumps. In contrast, for permanent jumps, the number of jumps per jump day is roughly stable over time, and does not have any structural change after 2003. Another thing to notice is that after 2003, \bar{n}_D^J of the large market cap quartiles has some large spikes in certain months, but we don't have a good explanation so far.

With respect to the percentage of the jump stocks (p^J), we see that for the small market cap stocks, p^J increase significantly and steadily, from several percent in 1995 to roughly 80 percent after 2003. It suggests the small market cap stocks used to be very stable now become very unstable: most of the stocks would experience sudden jumps, transient or permanent. For large market cap stocks, the percentage of jump stocks is relatively large in 1995, more than 60 percent, and it becomes oscillatory since 2003: in certain months, most of the stocks experience jumps, while in other months, it is not the case.

We have discussed the results of market cap quartiles in detail. Similar results are obtained for the quartiles of other market microstructure related characteristics. For example, the relevant properties of the jump data also experienced a structural shift around 2003 for the volume quartiles (figures 4.5 to 4.8). The ranking of the number of jumps n^J and jump days D^J across different volume quartiles has completely reversed after 2003. For the quartile of the largest volume, *vol3*, the number of jumps and jump days in a month remain stable and even decrease slightly. For other quartiles, the smaller the volume, the greater the increase in n^J and D^J after 2003. For the number of jumps per jump day, the small volume stocks have great increases in \bar{n}_D^J , while those of the large volume stocks does not increase on average, but in certain months from now and then, the number of jumps per jump day becomes very large. Another thing to notice is that the changes in \bar{n}_D^J after 2003 mainly come the transient jumps, because \bar{n}_D^J roughly remain stable across all the volume quartiles. For the percentage of jump stocks (p^J), the smaller the volume of the stock, the greater the increase in p^J . Aggregately, most of the stocks didn't experience any jump before 2003, but the story completely changed later: most

of the stocks do experience some price jump every month after 2003.

When the jump dataset is grouped into quartiles based on the price level, which is inversely related to relative tick size, one can easily observe similar structural change around 2003 as well (figures 4.9 to 4.11). As discussed before, four price ranges are chosen: smaller than 5 dollars, between 5 and 25 dollars, between 25 and 50 dollars, and greater than 50 dollars, according to which the stocks are categorized into four groups. Briefly speaking, the results are similar as before: the stocks with low stock price (thus large relative tick size) become significantly unstable, in terms of the total number of jumps, and the number of jump days, etc; whereas those with high-priced stocks (with small relative tick size) do not experience more price jumps after 2003. Therefore, one can conclude that large relative tick size is associated with more price jumps after auto-quote. Due to the fact that the percentage of the high frequency trading increases with the relative tick size, in addition to the structural change happened in 2003, it suggests that the high frequency trading is associated with more jumps of the low-priced stocks.

We also study the properties of the jumps with the market liquidity as the characteristic, for instance, the quoted and effective spread percentage (figures 4.12 to 4.17). The monthly percent quoted spread is aggregated by taking the time-weighted average of percent quoted spreads over the whole month, and monthly percent effective spread is the dollar-volume-weighted average of percent effective spread computed over all trades in the month. When the stocks with jumps are grouped into quartiles of bid-ask spread percentage, as expected, we obtain similar results: the structural change of the jump properties over time and the heterogeneity across different levels of market liquidity. For the most liquid stocks, the total number of jumps in a month did not increase, however illiquid stocks experienced more jumps as well as more jump days after 2003. Of course, most of the changes come from the transient jumps. Furthermore, other market liquidity measures have the similar results as well. Due to the limit space, we will not discuss the rest of the characteristics one by one.

It would be interesting to see the jump properties across different exchanges as well (figures 4.18 to 4.21). The stocks are categorized based on the exchanges they list. In the CRSP data, *EXCHCD* is a code indicating the exchange. Most of the CRSP stocks have normal exchange codes 1, 2, and 3, which correspond to NYSE, AMEX and the Nasdaq Stock Markets. The number of jumps n^J and the number of jumps days D^J of the NYSE stocks do not increase

over time, whereas those of the AMEX and Nasdaq stocks do increase significantly after 2003. Between the two, AMEX stocks have the greater surge in price jumps. Hence NYSE stocks are the most stable ones in terms of the jumps and jump days. Moreover, we also observe that for NYSE stocks, the number of jumps per jump day \bar{n}_D^J roughly stays stable on average, except for the spikes every few months, while the stocks listed on the other two exchanges have a considerable increase in \bar{n}_D^J . Comparing the transient and permanent jumps, we see that the increase since 2003 is mainly due to the increase in transient jumps. For permanent jumps, the number of jumps per jump day (\bar{n}_D^J) stays flat, it does not have any structural change after 2003, regardless which exchange the stock is listed.

4.7 Panel Regression Results

Having visualized the jump statistics of different characteristic buckets evolving over time, it is beneficial to study the research question in a more rigorous way. Based on the auto-quote implementation dates of NYSE stocks, we study the structural change around 2003 and the heterogeneous jump properties across different characteristic quartiles. The fixed effect panel data regression with auto-quote and characteristic quartiles/buckets dummy variables, as well as their interaction terms is used to capture these properties of the jumps. The dummy variable of auto-quote equals 0 before the introduction of auto-quote, which happened around early 2003, and 1 after auto-quote. The dummy variable of the characteristic quartile/buckets equals 1 for the stocks belong to the quartile/buckets, and 0 otherwise. The interaction term of auto-quote and quartile/group dummies is 1 if and only if the stock belongs to the quartile/group and the auto-quote has been implemented for the stock as well, otherwise the interaction term is 0.

The monthly aggregated dataset which contains 1379 NYSE stocks from January 1999 to December 2006. Note that as pointed out before, the NYSE listed stocks are the most stable ones after 2003, compared with the stocks listed on other exchanges. Hence it is natural to expect the before-after effects of auto-quote estimated based on NYSE stocks would be weaker than those based on the whole CRSP universe.

The fixed effect panel data regressions are used to study the number of jumps and jump days

around the event of NYSE's auto-quote implementation:

$$Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^j + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*j} + controls, \quad (4.6)$$

where the independent variable $Y_{i,t}$ is one of the variables that quantify the price instability: the number of jumps ($n_{i,t}^J$), the number of transient/permanent jumps ($n_{i,t}^{trans}/n_{i,t}^{perm}$) and the number of jump days ($D_{i,t}^J$). The subscripts i and t denote that it is the observation of stock i at month t . On the right hand side of equation (4.6), $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t . The value of $I_{i,t}^{auto}$ is 0 for the periods prior to auto-quote of stock i , and 1 after auto-quote. Moreover, $I_{i,t}^j, j = 0, 1, 2, 3$, is the quartile indicator: for stock i in quartile k , $I_{i,t}^j = 0$ if $j \neq k$, and $I_{i,t}^j = 1$ if $j = k$. The quartiles are grouped based on one of the characteristics. In equation (4.6), as the quartile 0 is the base case, $I_{i,t}^0$ is not included in the regression. Regressors $I_{i,t}^{auto*j}$ is the interaction terms of $I_{i,t}^{auto}$ and $I_{i,t}^j$. Last but not least, to test the robustness of the results, certain control variables are added in the regressions, such as market cap, volume, inverse price, bid-ask spread percentage, bipower variation, etc., which will be reported in tables 4.1 to 4.9.

Table 4.1 reports the fixed effect panel data regressions when the stocks are grouped into market capitalization quartiles. The model specification is the same as equation (4.6), in which the dependent variables are the number of jumps $n_{i,t}^J$, the number of permanent jumps $n_{i,t}^{perm}$, the number of transient jumps $n_{i,t}^{tran}$, and the number of jump days $D_{i,t}^J$. The results are presented by column: for example, first two columns show the results with $n_{i,t}^J$ as the dependent variable, the next two show the results for $n_{i,t}^{tran}$, and so forth.

In table 4.1, the coefficients for relevant variables are all significantly different from zero, and the results are robust to control variables. The before-and-after effects of the NYSE auto-quote implementation across different market capitalization quartiles can be evaluated based on these coefficient estimates. To be more specific, the base case is the number of jumps of market cap quartile 0 (*capt0*) before auto-quote, its value is set as zero for convenience (Please refer to the jump statistics in the previous section for the scale of the baseline, which should be smaller than the counterpart in the previous section, since we only consider the stocks with jumps previously). After auto-quote, $n_{i,t}^{J,capt0}$ of quartile *capt0* becomes 5.077 per month (as the

coefficient of $I_{i,t}^{auto}$ is 5.077), in other words, $n_{i,t}^{J,capt0}$ increases because of the auto-quote. For market quartile 1 ($capt1$), before auto-quote, $n_{i,t}^{J,capt1} = \beta_{i,t}^{capt1} I_{i,t}^{capt1} = 4.102$, however, after auto-quote, $n_{i,t}^{J,capt1} = \beta_{i,t}^{auto} I_{i,t}^{auto} + \beta_{i,t}^{capt1} I_{i,t}^{capt1} + \beta_{i,t}^{capt1*auto} I_{i,t}^{capt1*auto} = 5.077 + 4.102 - 9.965 = -0.786$. Similarly, for market cap quartile 2 ($capt2$), before auto-quote, $n_{i,t}^{J,capt2}$ equals 3.051, after the auto-quote, it becomes $5.077 + 3.051 - 9.693 = -1.565$. For market quartile 3 ($capt3$), before auto-quote, $n_{i,t}^{J,capt3}$ is 1.010 relative to the base case, after the auto-quote, it becomes $5.077 + 1.010 - 7.420 = -1.333$. Similarly, we can obtain the following results for the number of transient jumps $n_{i,t}^{trans}$: before auto-quote, the base case $capt0$ is set as zero, the number of jump days for $capt1$, $capt2$, and $capt3$ are 2.964, 2.075, 0.496, respectively; after auto-quote, the number of transient jumps become 4.05, -0.47 , -1.08 , -1.00 for $capt0$ to $capt3$, respectively. For the convenience of comparison, the before-and-after effects are computed based on the coefficients estimates, and compiled in table 4.9. Qualitatively speaking, the results are completely aligned with our previous research: quartile $capt0$ has the smallest market capitalization, and it is the most stable bucket before auto-quote. However, while the other stocks become more stable after auto-quote, the number of jumps for stocks in quartile $capt0$ increase considerably after auto-quote.

Furthermore, by observing the results in the subsequent columns, we find that the majority of the changes in $n_{i,t}^J$ is because of transient jumps. To be more specific, for quartile $capt0$, $n_{i,t}^J$ increases by 5.077, out of which 4.046 is the transient jumps, only 1.031 is the permanent jump. Considering that we may misclassify some of the transient-in-nature jumps as permanent, the conclusion that the price instability deterioration is mainly due to the increased transient jumps after auto-quote could be stronger than it appears there. Moreover, for quartiles $capt1$ to $capt3$, the transient jumps decrease considerably, which reflects the fact that most of those mid to large market cap stocks actually become more stable.

Tables 4.2 to 4.8 report the fixed effect panel regressions with the stock buckets of volume, price, effective spread, quoted spread, quote-to-trade ratio, market depth, bi-power variation, respectively. For each stock characteristic, quartile/bucket 0 has the lowest value of the corresponding characteristic, while quartile/bucket 3 has the greatest value of the characteristic. Similarly, as the results of market cap quartiles, the estimates of relevant coefficients are significantly different from zero, and robust when additional control variables are added. The net effects on $n_{i,t}^J$,

$n_{i,t}^{perm}$, $n_{i,t}^{tran}$, and $D_{i,t}^J$ before and after auto-quote implementation are calculated and compiled in table 4.9. The calculations are similar as shown previously for the market cap quartiles. To avoid the repetitive discussion, we will not discuss tables 4.2 to 4.8 one by one, as we did for the market cap quartiles. Let's just point out that these results based on NYSE stocks provide us with very similar empirical results as the previous results based on all CRSP common shares: the low-priced stocks and the ones with the smallest market cap, trading volume, and market liquidity, measured by effective/quoted spread and market depth, become more unstable after 2003, especially the drastic increase in the transient jumps. However, these stocks are the most stable bucket cross-sectionally before 2003. In contrast, the stocks on the other end of the characteristic spectra: the high-priced, large market cap and volume, high market liquidity stocks, become stable after auto-quote.

The relative tick size (inverse stock price) and quote-to-trade ratio are the two characteristics that particularly interesting to us, because they are related to the high frequency trading activities, as discussed before. Table 4.3 shows the panel data regressions of price quartiles. The four stocks buckets have the following price ranges: smaller than 5 dollars (*prc0*), between 5 and 25 dollars (*prc1*), between 25 and 50 dollars (*prc2*), and greater than 50 dollars (*prc3*). The stock price is inversely related to relative tick size. The low-priced stocks with large relative tick size have more high frequency activities (Yao and Ye, 2014). The results compiled in table 4.9 (third panel) indicate that the low-priced stocks (with large relative tick size) become significantly unstable, especially in terms of the transient jumps; whereas the high-priced stocks (with small relative tick size) experience less price jumps after 2003. Due to the fact that high frequency trading increases with the relative tick size, one can conclude that for these low-priced stocks, high frequency trading causes the price instability.

Furthermore, when the stocks are divided into buckets based on the quote-to-trade ratio, which is a proxy for high frequency trading, we observe that the stocks in quartile *qt_ratio3*, which has the greatest quote-to-trade ratio, becomes very unstable after auto-quote. Before auto-quote implementation, quartile *qt_ratio3* is the most unstable one ($n_{i,t}^J = 3.624$) compared with the other quartiles; after auto-quote implementation, $n_{i,t}^J$ increases to 12.598, out of which more than 10 increased jumps are transient. At the same time, the other quartiles with less high frequency trading become slightly more stable, as in table 4.9. It indicates that the quartile

of stocks with the most HFT activities are more prone to extreme price movements, especially the transient jumps, while other stocks with less HFT activities become more stable after the auto-quote implementation.

Theoretically, we can interpret the results as follows: the auto-quote reduces the latency considerably thus algorithmic trading becomes much easier. After the latency reduction, the slow market makers are more likely to meet high frequency bandits, relative to the chance of them trading with liquidity traders, as discussed in Menkveld and Zoican (2017). Thus slow market makers are more likely to be picked off than they are used to be. The slow market makers, who have greater long-term risk-bearing capability, would be crowded out of the market by HFTs, especially for the thinly-traded stocks. Compared with stock high volume and market liquidity, the slow market makers are facing more severe adverse selection problem in the thinly-traded stocks after the latency deduction. On top of this, the traditional designated market makers are replaced by endogenous liquidity providers, who may synchronously withdraw from the market when it is too risky to make the market, such as what has happened during the flash crash. (Biais and Foucault 2014, Kirilenko, Kyle, Samadi and Tuzun 2017).

Based on the empirical evidence, high frequency trading can be a double-edged sword. While the stocks with large market cap, volume, and market liquidity, and small relative stock size have become more stable after the reduction of latency, which is consistent with the previous empirical studies that suggest HFT is beneficial, the low-priced stocks with small market cap, volume and liquidity are suffered from large surges of transient jumps that cannot be attributed to information arrivals. The effects of high frequency trading can be destabilizing, which is in line with the argument of Menkveld and Zoican (2017).

Interestingly, the bi-power variation, which is the continuous part of the total variation, is not a relevant characteristic with respect to the jump properties: when stocks are grouped into bi-power variation buckets, the variables $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$, and $D_{i,t}^J$ all uniformly decrease across different quartiles. We also document that the quartile with smallest short interest and analyst coverage become more unstable after auto-quote, while the other quartiles have the fewer number of jumps and jump days, and the majority of the increase is because of transient jumps. The regression results based on other characteristics, such as book-to-market ratio, beta, institutional holding, short interest, analyst coverage, are not reported here in tables.

Another thing worth mentioning is that some results for the middle quartiles seem slightly different from the diagrams before. The reasons are mainly two-fold: first of all, the diagrams in the previous section are based on all of the stocks in the CRSP universe, which contains the stocks listed on all the exchanges. However, the panel data regressions are merely based on the NYSE stocks, due to the limitation that only the dates of NYSE auto-quote implementation are available to us. Moreover, the diagrams of the previous section are only based on the jump stocks, while the non-jump stocks are not included. The variables such as the number of jumps are averaged among all the stocks with jumps, not among all the stocks.

4.8 conclusion

This paper studies the price stability of U.S. common stocks in the recent 20 years. Based on the high frequency TAQ data, we first identify all the transient and permanent jumps in the stock price. We find that there is a structural change in the jump properties after the implementation of auto-quote, which greatly reduce the latency of the trading platform and make high frequency trading much easier. The empirical findings indicate that the effects of HFT (or algorithmic trading) and auto-quote implementation are very different cross-sectionally. It depends on the characteristics of the stocks: the small market cap, low volume and market liquidity, low-priced stocks become unstable. The number of jumps, especially the transient jumps, increase dramatically. It indicates that the market stability deteriorates for these stocks, since the increase in transient jumps cannot be explained by more news events and information arrivals. On the other hand, most of the stocks, in particular, the stocks with large market cap, high volume and liquidity, high-priced stocks are more stable after auto-quote. The empirical results suggest that HFT does not only have positive effects on the market liquidity and stability, which consistent with the theoretical literature. The policy implication is that for these small cap, low-priced and thinly-traded stocks, the existence of the designated specialists and slow market makers is good for the market stability. But these long-term risk-bearing market makers are likely to be crowded out by the high frequency traders when the trading latency is very low. The policy makers may take this into account for the market design.

Table 4.1: Fixed effect panel regressions with market capitalization buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	5.077	5.603	1.031	1.027	4.046	4.576	1.789	2.088
	(6.326)	(6.603)	(4.749)	(4.463)	(6.801)	(7.277)	(5.397)	(6.061)
$I_{i,t}^{capt1}$	4.102	4.301	1.137	1.077	2.964	3.224	1.634	1.662
	(6.166)	(6.300)	(6.567)	(5.942)	(5.952)	(6.358)	(6.288)	(6.209)
$I_{i,t}^{capt2}$	3.051	3.661	0.976	0.915	2.075	2.746	1.336	1.523
	(3.946)	(4.316)	(4.510)	(3.823)	(3.648)	(4.431)	(4.055)	(4.279)
$I_{i,t}^{capt3}$	1.010	1.983	0.513	0.449	0.496	1.533	0.471	0.774
	(1.422)	(2.359)	(2.425)	(1.830)	(0.955)	(2.498)	(1.540)	(2.211)
$I_{i,t}^{auto*capt1}$	-9.965	-9.391	-2.481	-2.375	-7.484	-7.017	-4.117	-3.910
	(-13.95)	(-13.54)	(-12.92)	(-12.46)	(-14.14)	(-13.77)	(-15.32)	(-14.77)
$I_{i,t}^{auto*capt2}$	-9.693	-9.448	-2.493	-2.418	-7.200	-7.030	-4.276	-4.226
	(-10.04)	(-9.864)	(-9.021)	(-8.668)	(-10.33)	(-10.25)	(-10.47)	(-10.41)
$I_{i,t}^{auto*capt3}$	-7.420	-7.510	-1.883	-1.838	-5.538	-5.672	-3.213	-3.300
	(-7.517)	(-7.589)	(-6.438)	(-6.168)	(-7.858)	(-8.088)	(-7.343)	(-7.567)
log_vol		-0.442		0.0354		-0.477		-0.279
		(-2.946)		(0.896)		(-4.169)		(-4.741)
inv_prc		4.768		0.306		4.462		0.980
		(4.124)		(1.161)		(4.849)		(2.448)
effective_sprd		-29.82		-8.530		-21.29		-14.11
		(-1.927)		(-1.809)		(-1.975)		(-1.726)
btm		4.70e-05		4.82e-06		4.22e-05		8.86e-06
		(6.260)		(2.374)		(7.651)		(5.795)
beta		-0.345		-0.113		-0.232		-0.183
		(-4.409)		(-4.516)		(-4.099)		(-5.570)
inst_hld		-3.730		-0.613		-3.117		-1.170
		(-6.445)		(-3.689)		(-7.134)		(-4.600)
shortint		-3.399		-1.210		-2.189		-2.020
		(-1.890)		(-2.335)		(-1.629)		(-3.190)
numest		-0.0500		-0.00893		-0.0411		-0.0127
		(-3.515)		(-1.890)		(-3.996)		(-1.950)
stdev		0.0355		0.00713		0.0284		0.00749
		(1.613)		(1.761)		(1.566)		(1.322)
depth		0.000300		5.48e-05		0.000246		7.79e-05
		(2.225)		(2.258)		(2.202)		(2.218)
qt_ratio		0.0374		0.00750		0.0299		0.00957
		(2.897)		(2.947)		(2.871)		(2.534)
bipower		5.01e-05		1.73e-05		3.28e-05		-4.36e-05
		(0.486)		(0.453)		(0.372)		(-1.220)
Constant	1.98e-09	1.18e-09	1.37e-09	1.34e-09	9.47e-10	1.69e-10	1.59e-09	1.21e-09
	(1.87e-08)	(1.08e-08)	(4.43e-08)	(4.42e-08)	(1.18e-08)	(2.01e-09)	(3.01e-08)	(2.35e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.110	0.145	0.071	0.081	0.103	0.146	0.155	0.179

The stocks are grouped into four market capitalization buckets: quartile 0 (*capt0*) has the smallest market cap, while quartile 3 (*capt3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{captj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*captj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{captj}$ is the indicator of market cap quartile j , and $I_{i,t}^{auto*captj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^j$. The base case is quartile 0 (*capt0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_vol* is the logged volume; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *effective_sprd* is the effective spread percentage; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *qt_ratio* is the quote-to-trade ratio; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.2: Fixed effect panel regressions with volume buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	19.95	19.84	4.553	4.559	15.40	15.28	6.686	6.824
	(22.09)	(20.83)	(18.50)	(17.96)	(22.67)	(21.14)	(21.46)	(20.72)
$I_{i,t}^{vol1}$	6.522	6.545	1.907	1.905	4.614	4.640	2.495	2.526
	(12.18)	(12.54)	(11.96)	(12.14)	(11.93)	(12.36)	(13.68)	(13.77)
$I_{i,t}^{vol2}$	9.782	10.39	2.964	3.068	6.817	7.318	4.104	4.379
	(14.52)	(15.33)	(15.27)	(15.64)	(13.86)	(14.85)	(16.40)	(16.63)
$I_{i,t}^{vol3}$	8.494	9.879	2.769	3.022	5.724	6.857	3.390	4.007
	(12.12)	(13.39)	(14.02)	(14.84)	(11.08)	(12.53)	(11.72)	(12.58)
$I_{i,t}^{auto*vol1}$	-18.08	-17.03	-4.321	-4.149	-13.76	-12.88	-5.766	-5.482
	(-20.98)	(-19.35)	(-17.67)	(-16.99)	(-21.57)	(-19.57)	(-20.53)	(-19.07)
$I_{i,t}^{auto*vol2}$	-24.72	-23.44	-6.007	-5.788	-18.72	-17.65	-9.117	-8.744
	(-24.97)	(-22.88)	(-21.33)	(-20.25)	(-25.78)	(-23.32)	(-26.38)	(-24.22)
$I_{i,t}^{auto*vol3}$	-23.22	-22.31	-5.631	-5.473	-17.59	-16.83	-8.497	-8.279
	(-21.59)	(-19.98)	(-18.49)	(-17.62)	(-22.28)	(-20.40)	(-20.43)	(-19.16)
log_capt		-1.685		-0.392		-1.294		-0.814
		(-10.08)		(-8.058)		(-10.26)		(-10.29)
inv_prc		-0.146		-0.825		0.679		-1.412
		(-0.128)		(-3.004)		(0.742)		(-3.471)
effective_sprd		-30.58		-8.865		-21.72		-14.69
		(-1.679)		(-1.608)		(-1.707)		(-1.575)
btm		4.78e-05		5.01e-06		4.28e-05		9.18e-06
		(5.530)		(2.145)		(6.746)		(4.681)
beta		-0.337		-0.101		-0.237		-0.177
		(-4.526)		(-4.218)		(-4.361)		(-5.594)
inst_hld		-1.739		-0.0910		-1.648		-0.443
		(-3.320)		(-0.578)		(-4.203)		(-1.932)
shortint		-4.247		-0.990		-3.257		-2.504
		(-2.872)		(-2.246)		(-2.924)		(-4.363)
numest		-0.0366		-0.00569		-0.0310		-0.00664
		(-2.680)		(-1.241)		(-3.142)		(-1.074)
stdev		0.0358		0.00734		0.0285		0.00765
		(1.611)		(1.770)		(1.561)		(1.323)
depth		0.000307		5.66e-05		0.000250		8.14e-05
		(2.237)		(2.272)		(2.214)		(2.237)
qt_ratio		0.0208		0.00327		0.0175		0.00464
		(2.509)		(2.277)		(2.520)		(1.996)
bipower		3.40e-05		2.22e-05		1.18e-05		-5.53e-05
		(0.324)		(0.545)		(0.134)		(-1.557)
Constant	-1.31e-08	-1.30e-08	-2.15e-09	-2.14e-09	-1.06e-08	-1.05e-08	-3.66e-09	-3.75e-09
	(-1.27e-07)	(-1.28e-07)	(-7.13e-08)	(-7.21e-08)	(-1.36e-07)	(-1.37e-07)	(-7.19e-08)	(-7.97e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.195	0.221	0.109	0.117	0.192	0.223	0.215	0.237

The stocks are grouped into four trading volume buckets: quartile 0 (*capt0*) has the smallest volume, while quartile 3 (*capt3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{volj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*volj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{volj}$ is the indicator of volume quartile j , and $I_{i,t}^{auto*volj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{volj}$. The base case is quartile 0 (*vol0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *effective_sprd* is the effective spread percentage; *btm* is the book-to-market ratio; *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *qt_ratio* is the quote-to-trade ratio; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.3: Fixed effect panel regressions with price buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	7.933 (9.835)	7.719 (9.740)	1.719 (9.433)	1.619 (8.869)	6.214 (9.622)	6.100 (9.673)	2.767 (9.658)	2.799 (9.841)
$I_{i,t}^{prc1}$	1.240 (1.783)	2.357 (3.490)	0.681 (4.423)	0.898 (5.747)	0.559 (1.019)	1.460 (2.765)	0.820 (3.139)	1.180 (4.396)
$I_{i,t}^{prc2}$	2.268 (2.940)	4.046 (5.223)	1.106 (6.114)	1.456 (7.590)	1.162 (1.933)	2.590 (4.348)	1.253 (4.076)	1.852 (5.658)
$I_{i,t}^{prc3}$	1.402 (2.034)	3.576 (4.978)	1.007 (6.184)	1.459 (7.866)	0.395 (0.726)	2.117 (3.814)	0.865 (3.181)	1.603 (5.305)
$I_{i,t}^{auto*prc1}$	-7.876 (-9.788)	-7.109 (-9.274)	-1.938 (-10.78)	-1.785 (-10.21)	-5.938 (-9.191)	-5.324 (-8.681)	-3.058 (-10.80)	-2.814 (-10.29)
$I_{i,t}^{auto*prc2}$	-11.68 (-12.52)	-10.66 (-11.86)	-2.952 (-13.05)	-2.744 (-12.33)	-8.727 (-11.96)	-7.912 (-11.32)	-4.777 (-13.34)	-4.443 (-12.62)
$I_{i,t}^{auto*prc3}$	-11.07 (-12.35)	-9.947 (-11.39)	-2.951 (-12.71)	-2.721 (-11.86)	-8.116 (-11.67)	-7.226 (-10.69)	-4.569 (-13.01)	-4.205 (-12.06)
log_capt		-1.178 (-6.204)		-0.316 (-5.730)		-0.863 (-6.105)		-0.433 (-5.409)
log_vol		-0.233 (-1.651)		0.0856 (2.352)		-0.319 (-2.929)		-0.200 (-3.729)
effective_sprd		-35.64 (-1.919)		-10.21 (-1.809)		-25.43 (-1.963)		-17.09 (-1.732)
btm		4.79e-05 (5.895)		5.08e-06 (2.296)		4.28e-05 (7.208)		9.19e-06 (5.172)
beta		-0.268 (-3.207)		-0.0929 (-3.545)		-0.175 (-2.881)		-0.151 (-4.429)
inst_hld		-3.153 (-5.413)		-0.484 (-2.943)		-2.669 (-6.003)		-0.842 (-3.608)
shortint		-3.504 (-1.747)		-1.312 (-2.359)		-2.192 (-1.455)		-1.944 (-2.848)
numest		-0.0320 (-2.116)		-0.00377 (-0.775)		-0.0283 (-2.564)		-0.00549 (-0.796)
stdev		0.0348 (1.619)		0.00678 (1.754)		0.0280 (1.575)		0.00737 (1.325)
depth		0.000296 (2.257)		5.21e-05 (2.258)		0.000244 (2.239)		7.75e-05 (2.260)
qt_ratio		0.0446 (3.000)		0.00927 (3.080)		0.0353 (2.969)		0.0126 (2.725)
bipower		7.99e-05 (0.773)		1.99e-05 (0.555)		5.99e-05 (0.653)		-2.82e-05 (-0.765)
Constant	-8.33e-09 (-8.52e-08)	-7.83e-09 (-7.52e-08)	-1.17e-09 (-4.02e-08)	-9.84e-10 (-3.39e-08)	-6.83e-09 (-9.27e-08)	-6.51e-09 (-8.11e-08)	-2.54e-09 (-5.29e-08)	-2.48e-09 (-5.20e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.064	0.102	0.048	0.061	0.058	0.102	0.101	0.129

The stocks are grouped into four price buckets based on the following price ranges: smaller than 5 dollars (*prc0*), 5 to 25 dollars (*prc1*), 25 to 50 dollars (*prc2*), and greater than 50 dollars (*prc3*). The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{prcj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*prcj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{prcj}$ is the indicator of price bucket j , and $I_{i,t}^{auto*prcj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{prcj}$. The base case is bucket 0 (*prc0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *log_vol* is the logged volume; *effective_sprd* is the effective spread percentage; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *qt_ratio* is the quote-to-trade ratio; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.4: Fixed effect panel regressions with effective spread buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	-3.822	-2.973	-1.255	-1.124	-2.567	-1.849	-2.080	-1.640
	(-15.05)	(-10.35)	(-14.77)	(-11.69)	(-14.39)	(-8.995)	(-16.64)	(-11.79)
$I_{i,t}^{effsprd1}$	-0.355	-0.918	-0.273	-0.350	-0.0829	-0.569	-0.348	-0.634
	(-1.125)	(-2.640)	(-2.883)	(-3.444)	(-0.361)	(-2.236)	(-2.338)	(-3.797)
$I_{i,t}^{effsprd2}$	-2.843	-3.881	-1.269	-1.348	-1.575	-2.533	-1.610	-2.132
	(-4.118)	(-5.465)	(-7.177)	(-7.118)	(-3.004)	(-4.760)	(-5.795)	(-7.073)
$I_{i,t}^{effsprd3}$	-2.474	-3.650	-1.311	-1.236	-1.163	-2.414	-1.885	-2.465
	(-2.515)	(-4.055)	(-5.326)	(-4.964)	(-1.534)	(-3.575)	(-5.512)	(-7.093)
$I_{i,t}^{auto*effsprd1}$	1.816	2.135	0.612	0.642	1.204	1.494	1.252	1.433
	(2.880)	(3.484)	(3.182)	(3.398)	(2.709)	(3.469)	(4.244)	(5.046)
$I_{i,t}^{auto*effsprd2}$	14.31	14.13	3.782	3.714	10.53	10.42	5.988	6.011
	(14.51)	(15.01)	(14.09)	(13.97)	(14.37)	(15.08)	(14.79)	(15.43)
$I_{i,t}^{auto*effsprd3}$	21.06	19.62	5.139	4.860	15.92	14.76	7.666	7.280
	(15.26)	(13.99)	(14.24)	(13.08)	(15.19)	(13.92)	(15.94)	(14.87)
log_capt		-1.130		-0.333		-0.797		-0.558
		(-6.035)		(-5.939)		(-5.708)		(-6.437)
log_vol		-0.169		0.103		-0.272		-0.173
		(-1.295)		(2.894)		(-2.718)		(-3.267)
inv_prc		-2.582		-1.398		-1.184		-1.279
		(-2.110)		(-4.890)		(-1.203)		(-2.945)
btm		4.85e-05		5.21e-06		4.32e-05		9.45e-06
		(5.610)		(2.240)		(6.820)		(4.788)
beta		-0.447		-0.137		-0.311		-0.221
		(-5.936)		(-5.537)		(-5.756)		(-6.997)
inst_hld		-1.490		-0.112		-1.379		-0.399
		(-2.882)		(-0.750)		(-3.488)		(-1.844)
shortint		-4.480		-1.531		-2.948		-2.480
		(-2.853)		(-3.242)		(-2.530)		(-4.422)
numest		-0.0270		-0.00322		-0.0237		-0.00267
		(-1.935)		(-0.693)		(-2.362)		(-0.423)
stdev		0.0328		0.00645		0.0264		0.00656
		(1.609)		(1.740)		(1.563)		(1.283)
depth		0.000279		4.90e-05		0.000230		7.02e-05
		(2.260)		(2.260)		(2.239)		(2.263)
qt_ratio		0.0268		0.00501		0.0218		0.00678
		(2.597)		(2.580)		(2.578)		(2.244)
bipower		-6.33e-05		-1.05e-05		-5.28e-05		-8.26e-05
		(-0.508)		(-0.331)		(-0.464)		(-1.795)
Constant	7.84e-09	6.92e-09	2.93e-09	2.81e-09	5.24e-09	4.45e-09	4.10e-09	3.65e-09
	(7.06e-08)	(6.28e-08)	(9.33e-08)	(9.24e-08)	(6.22e-08)	(5.25e-08)	(7.54e-08)	(7.15e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.169	0.190	0.097	0.104	0.167	0.192	0.187	0.207

The stocks are grouped into four effective spread buckets: quartile 0 (*effsprd0*) has the smallest effective spread, while quartile 3 (*effsprd3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{effsprdj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*effsprdj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{effsprdj}$ is the indicator of effective spread quartile j , and $I_{i,t}^{auto*effsprdj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{effsprdj}$. The base case is quartile 0 (*effsprd0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *log_vol* is the logged volume; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *qt_ratio* is the quote-to-trade ratio; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.5: Fixed effect panel regressions with quoted spread buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	-3.767 (-14.59)	-2.913 (-10.03)	-1.240 (-14.11)	-1.093 (-11.05)	-2.526 (-14.10)	-1.820 (-8.881)	-2.066 (-16.20)	-1.631 (-11.60)
$I_{i,t}^{quotsprd1}$	-0.123 (-0.412)	-0.853 (-2.604)	-0.282 (-2.994)	-0.400 (-4.032)	0.159 (0.750)	-0.452 (-1.920)	-0.161 (-1.138)	-0.509 (-3.251)
$I_{i,t}^{quotsprd2}$	-3.178 (-4.689)	-4.566 (-6.424)	-1.489 (-8.725)	-1.675 (-9.144)	-1.689 (-3.275)	-2.891 (-5.385)	-1.634 (-5.919)	-2.268 (-7.550)
$I_{i,t}^{quotsprd3}$	-3.202 (-3.124)	-5.083 (-5.134)	-1.625 (-6.527)	-1.760 (-7.014)	-1.577 (-1.991)	-3.323 (-4.398)	-2.043 (-5.774)	-2.822 (-7.790)
$I_{i,t}^{auto*quotsprd1}$	1.345 (2.094)	1.719 (2.745)	0.504 (2.495)	0.554 (2.785)	0.841 (1.882)	1.164 (2.685)	1.003 (3.297)	1.202 (4.083)
$I_{i,t}^{auto*quotsprd2}$	13.81 (13.78)	13.76 (14.11)	3.635 (13.17)	3.608 (13.07)	10.17 (13.80)	10.16 (14.30)	5.868 (13.88)	5.930 (14.42)
$I_{i,t}^{auto*quotsprd3}$	22.09 (15.54)	20.80 (14.26)	5.291 (14.11)	5.024 (13.08)	16.80 (15.73)	15.78 (14.38)	8.107 (16.29)	7.762 (15.21)
log_capt		-1.319 (-7.154)		-0.394 (-7.398)		-0.925 (-6.629)		-0.621 (-7.361)
log_vol		-0.121 (-0.926)		0.105 (2.927)		-0.226 (-2.263)		-0.158 (-2.934)
inv_prc		-1.964 (-1.487)		-1.161 (-4.043)		-0.803 (-0.749)		-1.254 (-2.796)
btm		4.84e-05 (5.573)		5.19e-06 (2.229)		4.32e-05 (6.768)		9.44e-06 (4.714)
beta		-0.421 (-5.496)		-0.127 (-5.154)		-0.294 (-5.314)		-0.212 (-6.669)
inst_hld		-1.504 (-2.890)		-0.144 (-0.972)		-1.359 (-3.405)		-0.379 (-1.722)
shortint		-3.437 (-2.126)		-1.294 (-2.624)		-2.143 (-1.807)		-2.078 (-3.537)
numest		-0.0273 (-1.961)		-0.00345 (-0.742)		-0.0238 (-2.375)		-0.00200 (-0.318)
stdev		0.0337 (1.607)		0.00669 (1.744)		0.0270 (1.560)		0.00685 (1.295)
depth		0.000284 (2.253)		5.07e-05 (2.268)		0.000234 (2.231)		7.22e-05 (2.262)
qt_ratio		0.0264 (2.602)		0.00511 (2.611)		0.0213 (2.579)		0.00657 (2.215)
bipower		-9.44e-05 (-0.747)		-1.55e-05 (-0.481)		-7.89e-05 (-0.682)		-9.70e-05 (-2.032)
Constant	7.30e-09 (6.64e-08)	6.72e-09 (6.15e-08)	2.91e-09 (9.38e-08)	2.82e-09 (9.31e-08)	4.72e-09 (5.65e-08)	4.23e-09 (5.04e-08)	4.02e-09 (7.49e-08)	3.71e-09 (7.36e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.168	0.190	0.094	0.101	0.168	0.193	0.188	0.209

The stocks are grouped into four quoted spread buckets: quartile 0 (*quotsprd0*) has the smallest quoted spread, while quartile 3 (*quotsprd3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{quotsprdj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*quotsprdj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{quotsprdj}$ is the indicator of quoted spread quartile j , and $I_{i,t}^{auto*quotsprdj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{quotsprdj}$. The base case is quartile 0 (*quotsprd0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *log_vol* is the logged volume; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *qt_ratio* is the quote-to-trade ratio; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.6: Fixed effect panel regressions with quote-to-trade ratio buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	-0.452 (-2.384)	0.0881 (0.410)	-0.175 (-2.767)	-0.0870 (-1.268)	-0.277 (-1.884)	0.175 (1.083)	-0.418 (-5.099)	-0.169 (-1.782)
$I_{i,t}^{qt_ratio1}$	2.118 (11.79)	2.230 (12.13)	0.541 (10.32)	0.634 (11.74)	1.577 (11.68)	1.595 (11.60)	0.974 (14.23)	1.018 (14.32)
$I_{i,t}^{qt_ratio2}$	2.970 (16.83)	3.209 (13.89)	0.961 (14.04)	1.154 (15.54)	2.009 (16.06)	2.055 (11.92)	1.465 (17.78)	1.555 (15.32)
$I_{i,t}^{qt_ratio3}$	3.624 (11.50)	4.022 (9.902)	1.078 (11.97)	1.393 (12.98)	2.545 (10.60)	2.629 (8.415)	1.755 (12.52)	1.870 (10.45)
$I_{i,t}^{auto*qt_ratio1}$	-2.063 (-10.99)	-1.980 (-11.16)	-0.554 (-10.12)	-0.556 (-10.27)	-1.509 (-10.39)	-1.424 (-10.44)	-0.885 (-11.04)	-0.873 (-11.57)
$I_{i,t}^{auto*qt_ratio2}$	-0.208 (-0.371)	-0.155 (-0.294)	-0.400 (-2.656)	-0.438 (-3.021)	0.191 (0.454)	0.282 (0.719)	-0.131 (-0.474)	-0.127 (-0.499)
$I_{i,t}^{auto*qt_ratio3}$	13.05 (10.67)	12.88 (10.38)	2.628 (8.351)	2.593 (8.101)	10.43 (11.30)	10.28 (11.02)	4.263 (10.32)	4.183 (9.980)
log_capt		-1.568 (-8.343)		-0.424 (-7.349)		-1.144 (-8.324)		-0.750 (-9.152)
log_vol		0.703 (4.610)		0.371 (8.819)		0.331 (2.900)		0.231 (3.640)
inv_prc		0.654 (0.519)		-0.710 (-2.171)		1.365 (1.394)		-0.942 (-1.951)
effective_sprd		-42.68 (-1.673)		-12.33 (-1.619)		-30.35 (-1.695)		-20.50 (-1.583)
btm		4.85e-05 (5.752)		5.25e-06 (2.318)		4.32e-05 (6.982)		9.48e-06 (5.058)
beta		-0.355 (-4.269)		-0.119 (-4.600)		-0.236 (-3.905)		-0.184 (-5.248)
inst_hld		-1.520 (-2.701)		-0.0310 (-0.194)		-1.489 (-3.511)		-0.253 (-1.019)
shortint		-2.380 (-1.330)		-0.843 (-1.651)		-1.537 (-1.148)		-1.476 (-2.265)
numest		-0.0667 (-5.048)		-0.0169 (-4.019)		-0.0498 (-5.091)		-0.0214 (-3.633)
stdev		0.0301 (1.557)		0.00514 (1.716)		0.0250 (1.514)		0.00501 (1.123)
depth		0.000270 (2.197)		4.25e-05 (2.241)		0.000227 (2.174)		6.39e-05 (2.189)
bipower		0.000114 (1.156)		3.19e-05 (0.754)		8.19e-05 (1.037)		-1.24e-05 (-0.427)
Constant	3.20e-09 (2.59e-08)	2.86e-09 (2.38e-08)	1.44e-09 (4.30e-08)	1.44e-09 (4.44e-08)	2.10e-09 (2.23e-08)	1.76e-09 (1.92e-08)	1.89e-09 (3.12e-08)	1.72e-09 (3.05e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.151	0.172	0.082	0.094	0.151	0.175	0.164	0.183

The stocks are grouped into four quote-to-trade ratio buckets: quartile 0 (*quotsprd0*) has the smallest quote-to-trade ratio, while quartile 3 (*quotsprd3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{qt_ratioj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*qt_ratioj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{qt_ratioj}$ is the indicator of quote-to-trade ratio quartile j , and $I_{i,t}^{auto*qt_ratioj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{qt_ratioj}$. The base case is quartile 0 (*qt.ratio0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *log_vol* is the logged volume; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *effective_sprd* is the effective spread percentage; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.7: Fixed effect panel regressions with market depth buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	4.182 (5.220)	5.391 (6.230)	0.821 (3.839)	0.992 (4.270)	3.360 (5.634)	4.399 (6.833)	1.432 (4.242)	2.003 (5.626)
$I_{i,t}^{depth1}$	4.520 (8.176)	5.122 (9.098)	1.236 (8.818)	1.299 (8.938)	3.285 (7.882)	3.822 (9.058)	1.859 (8.365)	2.109 (9.322)
$I_{i,t}^{depth2}$	6.127 (8.142)	7.707 (9.249)	1.782 (8.881)	1.999 (9.014)	4.345 (7.796)	5.708 (9.239)	2.507 (7.673)	3.170 (8.915)
$I_{i,t}^{depth3}$	5.878 (7.545)	8.427 (9.056)	1.663 (7.717)	2.045 (8.111)	4.215 (7.364)	6.382 (9.268)	2.148 (6.461)	3.220 (8.309)
$I_{i,t}^{auto*depth1}$	-7.536 (-12.39)	-6.834 (-11.85)	-1.880 (-11.26)	-1.728 (-10.62)	-5.656 (-12.63)	-5.107 (-12.16)	-3.221 (-13.61)	-2.968 (-13.08)
$I_{i,t}^{auto*depth2}$	-8.737 (-9.371)	-8.442 (-9.175)	-2.287 (-8.688)	-2.189 (-8.301)	-6.450 (-9.539)	-6.253 (-9.430)	-3.869 (-9.589)	-3.782 (-9.513)
$I_{i,t}^{auto*depth3}$	-6.376 (-6.309)	-6.860 (-6.658)	-1.640 (-5.654)	-1.682 (-5.627)	-4.737 (-6.484)	-5.178 (-6.991)	-2.746 (-6.233)	-2.971 (-6.688)
log_capt		-1.689 (-9.274)		-0.442 (-8.075)		-1.247 (-9.148)		-0.745 (-8.802)
log_vol		-0.718 (-4.895)		-0.0305 (-0.785)		-0.687 (-6.089)		-0.359 (-6.588)
inv_prc		0.763 (0.622)		-0.735 (-2.592)		1.498 (1.518)		-0.806 (-1.843)
effective_sprd		-34.81 (-1.851)		-9.779 (-1.756)		-25.03 (-1.888)		-16.36 (-1.686)
btm		5.01e-05 (6.057)		5.42e-06 (2.461)		4.47e-05 (7.333)		9.74e-06 (5.538)
beta		-0.227 (-2.859)		-0.0851 (-3.378)		-0.142 (-2.460)		-0.137 (-4.206)
inst_hld		-3.001 (-5.749)		-0.423 (-2.758)		-2.579 (-6.520)		-0.819 (-3.790)
shortint		-4.459 (-2.383)		-1.457 (-2.761)		-3.002 (-2.135)		-2.215 (-3.420)
numest		-0.0436 (-3.046)		-0.00756 (-1.604)		-0.0361 (-3.484)		-0.0102 (-1.576)
stdev		-0.00338 (-1.802)		4.98e-05 (0.0808)		-0.00343 (-1.773)		-0.00260 (-2.206)
qt_ratio		0.0391 (2.927)		0.00794 (2.989)		0.0312 (2.899)		0.0104 (2.597)
bipower		8.52e-05 (0.787)		2.42e-05 (0.646)		6.10e-05 (0.650)		-3.16e-05 (-0.870)
Constant	-5.31e-09 (-5.06e-08)	-6.54e-09 (-6.11e-08)	-4.80e-10 (-1.58e-08)	-6.16e-10 (-2.04e-08)	-4.50e-09 (-5.65e-08)	-5.58e-09 (-6.83e-08)	-1.58e-09 (-3.02e-08)	-2.16e-09 (-4.38e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.087	0.133	0.063	0.077	0.078	0.133	0.127	0.167

The stocks are grouped into four market depth buckets: quartile 0 (*depth0*) has the smallest market depth, while quartile 3 (*depth3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{depthj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*depthj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{depthj}$ is the indicator of market depth quartile j , and $I_{i,t}^{auto*depthj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{depthj}$. The base case is quartile 0 (*depth0*) and auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *log_vol* is the logged volume; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *effective_sprd* is the effective spread percentage; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *qt_ratio* is the quote-to-trade ratio; *bipower* is the bi-power variation. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.8: Fixed effect panel regressions with bipower variation buckets

depend variable	$n_{i,t}^J$	$n_{i,t}^J$	$n_{i,t}^{perm}$	$n_{i,t}^{perm}$	$n_{i,t}^{tran}$	$n_{i,t}^{tran}$	$D_{i,t}^J$	$D_{i,t}^J$
$I_{i,t}^{auto}$	-0.996 (-3.134)	-0.379 (-1.064)	-0.543 (-5.830)	-0.476 (-4.596)	-0.453 (-1.908)	0.0971 (0.362)	-1.037 (-7.508)	-0.703 (-4.538)
$I_{i,t}^{bipower1}$	0.602 (3.298)	0.636 (3.339)	0.159 (2.632)	0.144 (2.288)	0.443 (3.436)	0.492 (3.662)	0.154 (1.899)	0.192 (2.264)
$I_{i,t}^{bipower2}$	0.138 (0.634)	0.198 (0.859)	0.0343 (0.487)	0.00767 (0.102)	0.103 (0.674)	0.190 (1.170)	-0.0934 (-0.967)	-0.0343 (-0.337)
$I_{i,t}^{bipower3}$	-0.847 (-2.951)	-0.841 (-2.750)	-0.180 (-2.124)	-0.219 (-2.323)	-0.667 (-3.180)	-0.622 (-2.844)	-0.719 (-6.161)	-0.639 (-5.028)
$I_{i,t}^{auto*bipower1}$	-1.665 (-7.448)	-1.438 (-6.645)	-0.387 (-5.995)	-0.348 (-5.437)	-1.278 (-7.631)	-1.090 (-6.802)	-0.501 (-5.835)	-0.433 (-5.070)
$I_{i,t}^{auto*bipower2}$	-1.373 (-4.784)	-1.056 (-3.885)	-0.287 (-3.644)	-0.238 (-3.080)	-1.086 (-5.018)	-0.819 (-4.041)	-0.308 (-2.721)	-0.200 (-1.846)
$I_{i,t}^{auto*bipower3}$	0.194 (0.518)	0.475 (1.381)	0.0622 (0.631)	0.0885 (0.922)	0.132 (0.458)	0.386 (1.492)	0.502 (3.291)	0.591 (4.182)
log_capt		-1.547 (-7.803)		-0.411 (-6.843)		-1.136 (-7.754)		-0.732 (-8.410)
log_vol		-0.113 (-0.735)		0.124 (3.042)		-0.238 (-2.001)		-0.111 (-1.887)
inv_prc		-0.451 (-0.332)		-1.066 (-3.171)		0.615 (0.578)		-1.336 (-2.735)
effective_sprd		-45.52 (-1.805)		-12.52 (-1.739)		-33.00 (-1.830)		-20.52 (-1.690)
btm		4.66e-05 (6.732)		4.73e-06 (2.494)		4.19e-05 (8.274)		8.73e-06 (6.553)
beta		-0.210 (-2.381)		-0.0777 (-2.890)		-0.132 (-2.050)		-0.125 (-3.458)
inst_hld		-3.134 (-5.282)		-0.470 (-2.741)		-2.664 (-5.998)		-0.911 (-3.677)
shortint		-3.257 (-1.596)		-1.217 (-2.137)		-2.040 (-1.335)		-1.919 (-2.671)
numest		-0.0414 (-2.653)		-0.00665 (-1.349)		-0.0348 (-3.047)		-0.00975 (-1.366)
stdev		0.0368 (1.611)		0.00740 (1.746)		0.0294 (1.567)		0.00819 (1.341)
depth		0.000319 (2.244)		5.92e-05 (2.251)		0.000260 (2.227)		8.71e-05 (2.244)
qt_ratio		0.0475 (3.037)		0.0101 (3.128)		0.0374 (3.004)		0.0141 (2.802)
Constant	2.94e-09 (2.86e-08)	2.41e-09 (2.28e-08)	1.68e-09 (5.56e-08)	1.64e-09 (5.44e-08)	1.59e-09 (2.05e-08)	1.11e-09 (1.37e-08)	2.27e-09 (4.37e-08)	1.98e-09 (4.01e-08)
Observations	91,522	91,522	91,522	91,522	91,522	91,522	91,522	91,522
R-squared	0.024	0.073	0.027	0.043	0.017	0.075	0.057	0.096

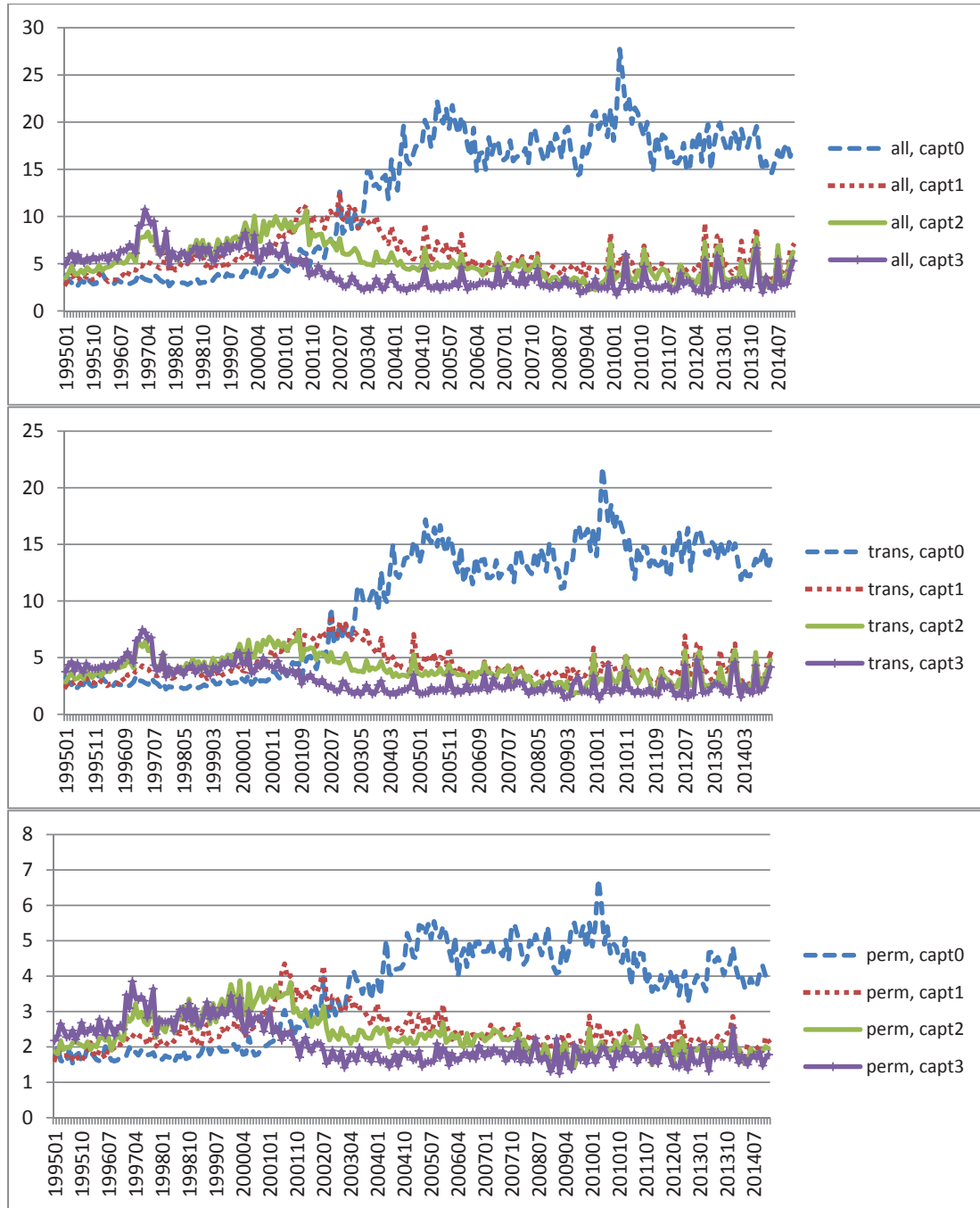
The stocks are grouped into four bi-power variation buckets: quartile 0 (*bipower0*) has the smallest market depth, while quartile 3 (*bipower3*) has the greatest. The model specification is $Y_{i,t} = \alpha_i + \lambda_t + \beta_a I_{i,t}^{auto} + \sum_{j=1}^3 \beta_j I_{i,t}^{bipowerj} + \sum_{j=1}^3 \gamma_j I_{i,t}^{auto*bipowerj} + controls$, where the independent variable $Y_{i,t}$ is one of the variables: $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$ and $D_{i,t}^J$. $I_{i,t}^{auto}$ is the indicator of auto-quote for stock i at month t , $I_{i,t}^{bipowerj}$ is the indicator of bi-power variation quartile j , and $I_{i,t}^{auto*bipowerj}$ is the interaction term between $I_{i,t}^{auto}$ and $I_{i,t}^{bipowerj}$. The base case is quartile 0 (*bipower0*) before auto-quotation in 2003. The regression results are accounted for the two-dimension clustered standard errors. With respect to the control variables, *log_capt* is the logged market cap; *log_vol* is the logged volume; *inv_prc* is the inverse of price, which is proportional to the relative tick size; *effective_sprd* is the effective spread percentage; *btm* is the book-to-market ratio, *beta* is the beta of the stock estimated using the two-year rolling window prior to the month t ; *inst_hld* is the institution holding percentage; *shortint* is the short interest percentage; *numest* is the number of analyst forecasts; *stdev* is the standard deviation of analyst forecasts; *depth* is the market depth; *qt_ratio* is the quote-to-trade ratio. The sample is 1379 NYSE stocks from January 1999 to December 2006.

Table 4.9: The effects of NYSE auto-quote, based on the regression results

no control	$n_{i,t}^J$			$n_{i,t}^{perm}$			$n_{i,t}^{trans}$			$D_{i,t}^J$		
	before	after	change	before	after	change	before	after	change	before	after	change
capt0	0.00	5.077	5.08	0.00	1.031	1.03	0.00	4.046	4.05	0.00	1.789	1.79
capt1	4.102	-0.79	-4.89	1.137	-0.31	-1.45	2.964	-0.47	-3.44	1.634	-0.69	-2.33
capt2	3.051	-1.57	-4.62	0.976	-0.49	-1.46	2.075	-1.08	-3.15	1.336	-1.15	-2.49
capt3	1.010	-1.33	-2.34	0.513	-0.34	-0.85	0.496	-1.00	-1.49	0.471	-0.95	-1.42
vol0	0.00	19.95	19.95	0.00	4.553	4.55	0.00	15.40	15.40	0.00	6.686	6.69
vol1	6.522	8.39	1.87	1.907	2.14	0.23	4.614	6.26	1.65	2.495	3.42	0.92
vol2	9.782	5.01	-4.77	2.964	1.51	-1.45	6.817	3.52	-3.30	4.104	1.67	-2.43
vol3	8.494	5.22	-3.27	2.769	1.69	-1.08	5.724	3.56	-2.16	3.390	1.58	-1.81
prc0	0.00	7.933	7.93	0.00	1.719	1.72	0.00	6.214	6.21	0.00	2.767	2.77
prc1	1.240	1.30	0.06	0.681	0.46	-0.22	0.559	0.84	0.28	0.820	0.53	-0.29
prc2	2.268	-1.48	-3.75	1.106	-0.13	-1.23	1.162	-1.35	-2.51	1.253	-0.76	-2.01
prc3	1.402	-1.74	-3.14	1.007	-0.23	-1.23	0.395	-1.51	-1.90	0.865	-0.94	-1.80
effsprd0	0.00	-3.822	-3.82	0.00	-1.255	-1.26	0.00	-2.567	-2.57	0.00	-2.080	-2.08
effsprd1	-0.355	-2.36	-2.01	-0.273	-0.92	-0.64	-0.0829	-1.45	-1.36	-0.348	-1.18	-0.83
effsprd2	-2.843	7.65	10.49	-1.269	1.26	2.53	-1.575	6.39	7.96	-1.610	2.30	3.91
effsprd3	-2.474	14.76	17.24	-1.311	2.57	3.88	-1.163	12.19	13.35	-1.885	3.70	5.59
quotsprd0	0.00	-3.767	-3.77	0.00	-1.240	-1.24	0.00	-2.526	-2.53	0.00	-2.066	-2.07
quotsprd1	-0.123	-2.55	-2.42	-0.282	-1.02	-0.74	0.159	-1.53	-1.69	-0.161	-1.22	-1.06
quotsprd2	-3.178	6.87	10.04	-1.489	0.91	2.40	-1.689	5.96	7.64	-1.634	2.17	3.80
quotsprd3	-3.202	15.12	18.32	-1.625	2.43	4.05	-1.577	12.70	14.27	-2.043	4.00	6.04
qt_ratio0	0.00	-0.452	-0.452	0.00	-0.175	-0.175	0.00	-0.277	-0.277	0.00	-0.418	-0.418
qt_ratio1	2.118	-0.397	-2.515	0.541	-0.188	-0.729	1.577	-0.209	-1.786	0.974	-0.329	-1.303
qt_ratio2	2.970	2.31	-0.66	0.961	0.386	-0.575	2.009	1.923	-0.086	1.465	0.916	-0.549
qt_ratio3	3.624	16.222	12.598	1.078	3.531	2.453	2.545	12.698	10.153	1.755	5.6	3.845
depth0	0.00	4.182	4.182	0.00	0.821	0.821	0.00	3.360	3.36	0.00	1.432	1.432
depth1	4.520	1.166	-3.354	1.236	0.177	-1.059	3.285	0.989	-2.296	1.859	0.07	-1.789
depth2	6.127	1.572	-4.555	1.782	0.316	-1.466	4.345	1.255	-3.09	2.507	0.07	-2.437
depth3	5.878	3.684	-2.194	1.663	0.844	-0.819	4.215	2.838	-1.377	2.148	0.834	-1.314
bipower0	0.00	-0.996	-0.996	0.00	-0.543	-0.543	0.00	-0.453	-0.453	0.00	-1.037	-1.037
bipower1	0.602	-2.059	-2.661	0.159	-0.771	-0.93	0.443	-1.288	-1.731	0.154	-1.384	-1.538
bipower2	0.138	-2.231	-2.369	0.0343	-0.7957	-0.83	0.103	-1.436	-1.539	-0.0934	-1.4384	-1.345
bipower3	-0.847	-1.649	-0.802	-0.180	-0.6608	-0.4808	-0.667	-0.988	-0.321	-0.719	-1.254	-0.535

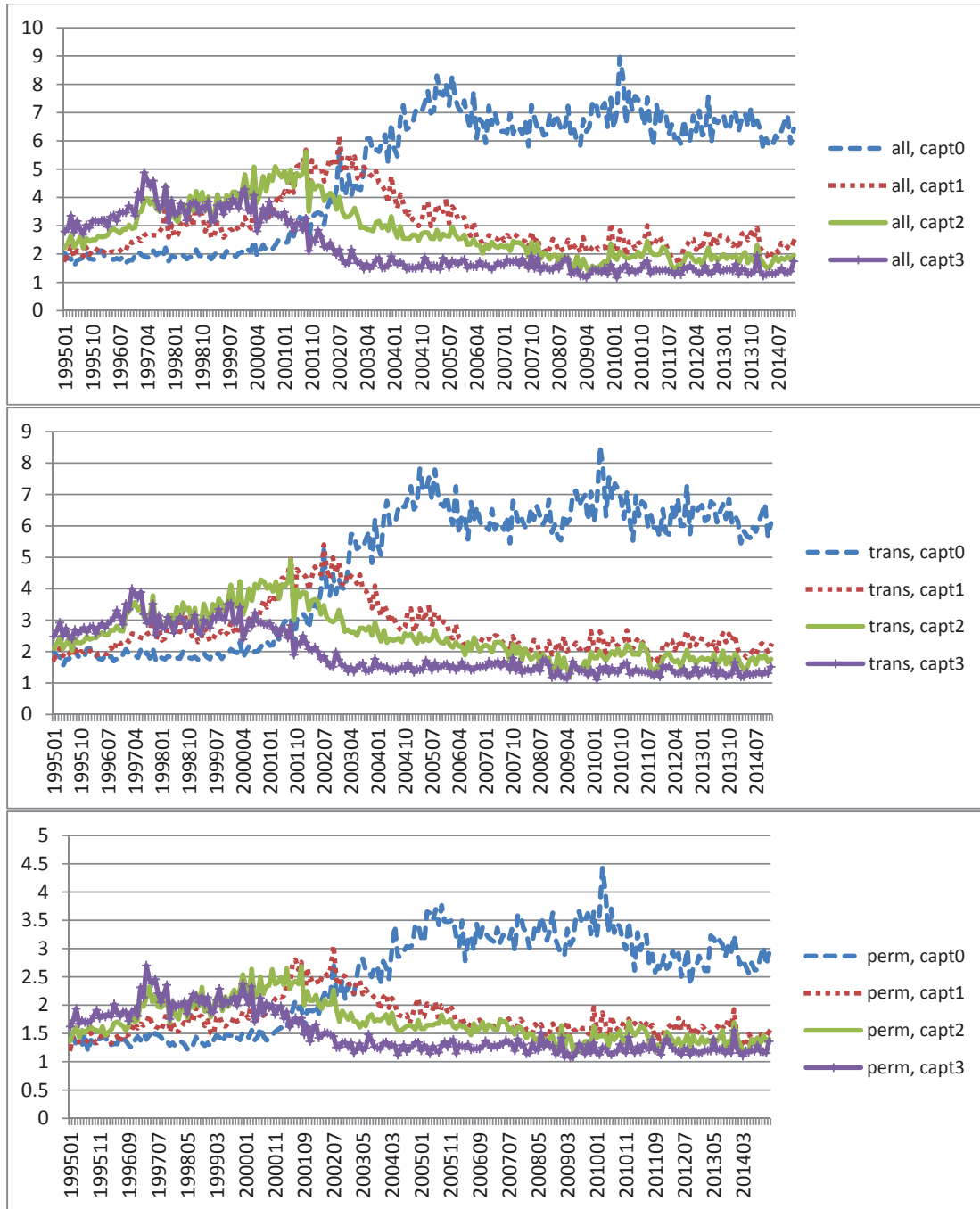
This table shows the effects of NYSE auto-quote implementation on the following variables across different characteristic buckets: the number of jumps $n_{i,t}^J$, the number of permanent jumps $n_{i,t}^{perm}$, the number of transient jumps $n_{i,t}^{trans}$, and jump days $D_{i,t}^J$ before and after the introduction of auto-quote. The change is the difference of the values before and after auto-quote. The results are calculated based on the panel regression estimates without the control variables, which are reported in tables 4.1 to 4.8. For each characteristic, the base case is quartile/bucket 0, which has the smallest value of the corresponding characteristic. For the base case before auto-quote, the values of $n_{i,t}^J$, $n_{i,t}^{perm}$, $n_{i,t}^{trans}$, and $D_{i,t}^J$ are set as zero.

Figure 4.1: Average numbers of jumps in a month for market capitalization quartiles



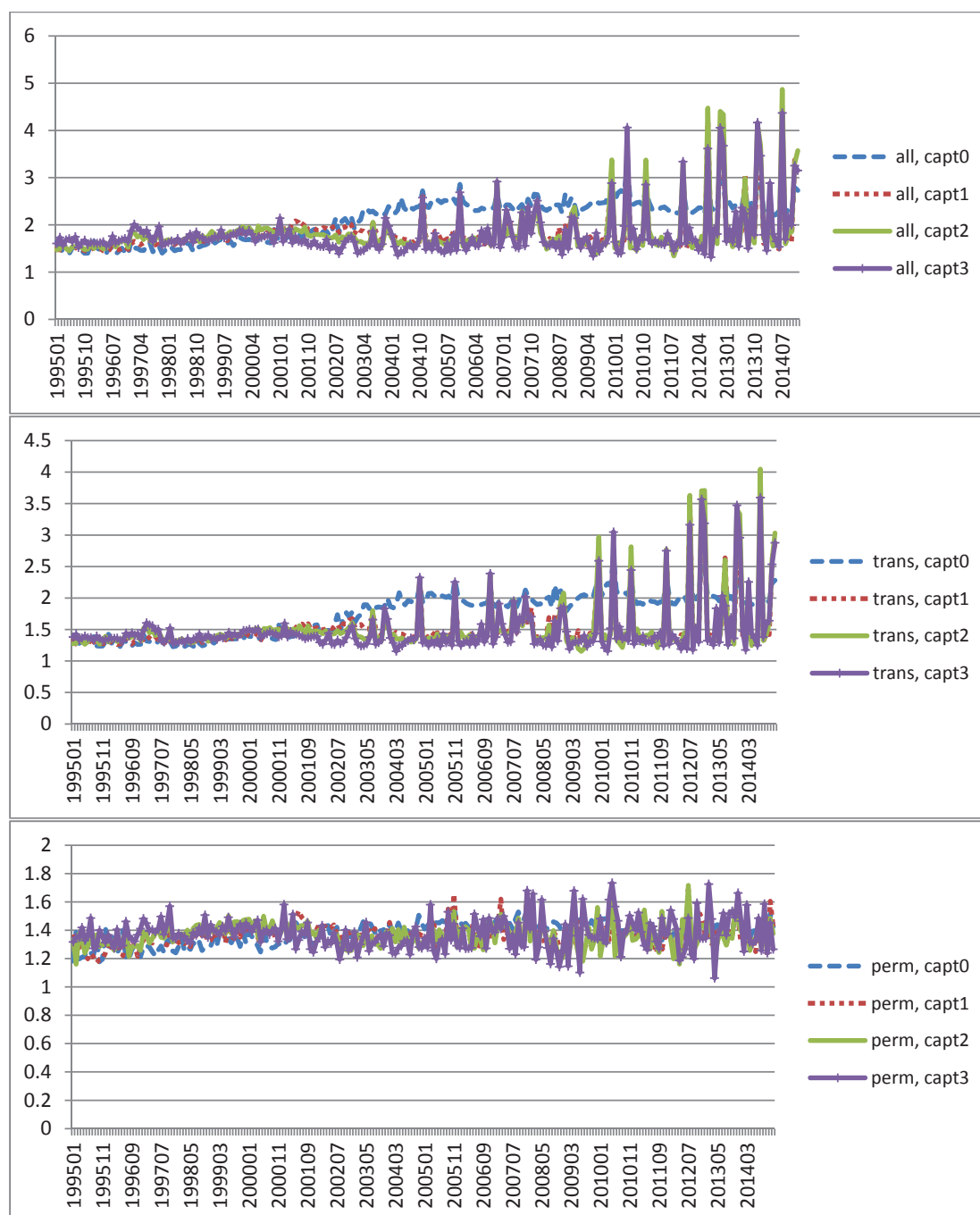
This table reports the number of jumps in a month. In each month, the jump stocks are grouped into quartiles based on the market cap. Quartile 0 has the smallest market cap while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.2: Average numbers of jump days in a month for market capitalization quartiles



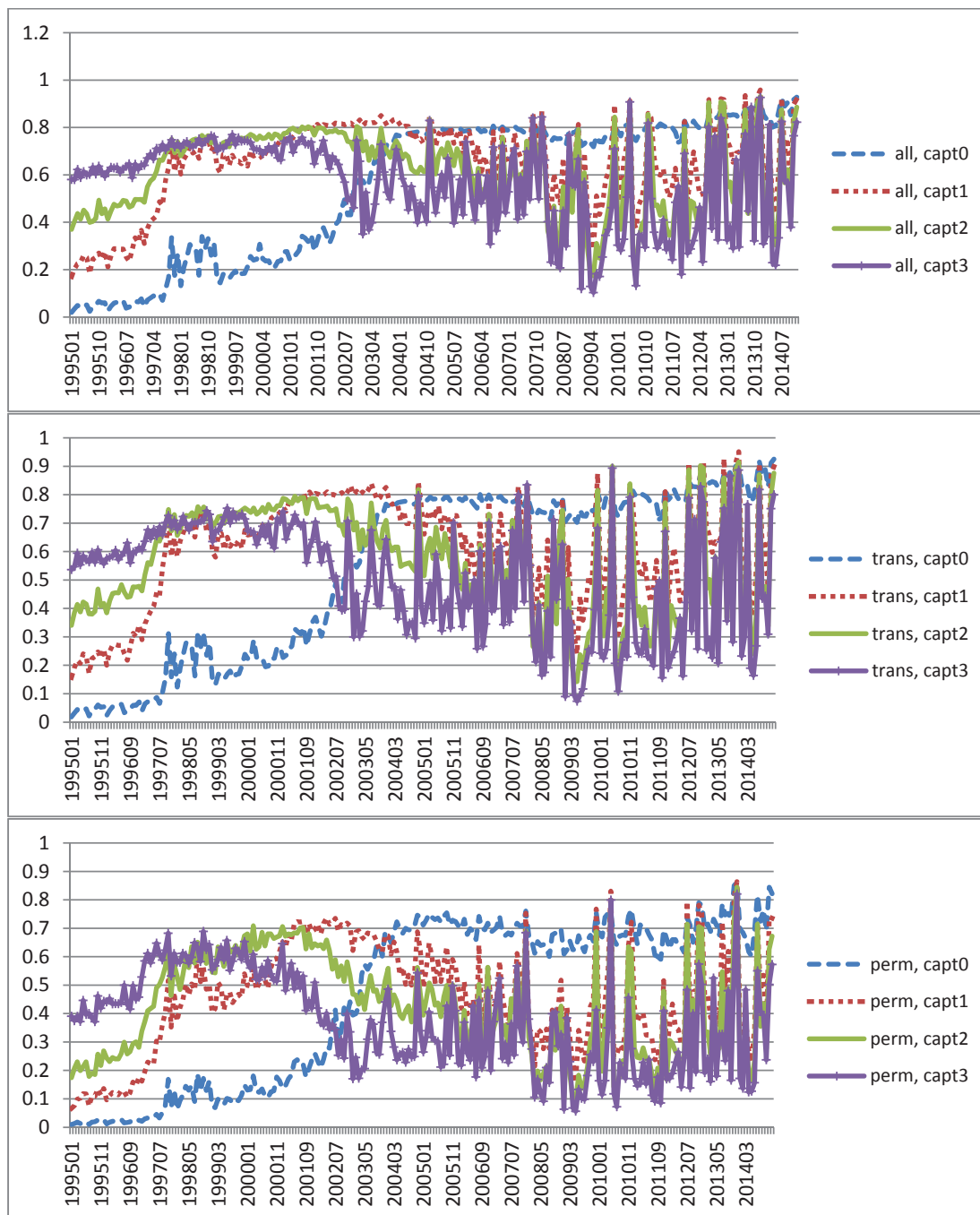
This table reports the number of jump days in a month. In each month, the jump stocks are grouped into quartiles based on the market cap. Quartile 0 has the smallest market cap while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.3: Average numbers of jumps per jump day in a month for market capitalization quartiles



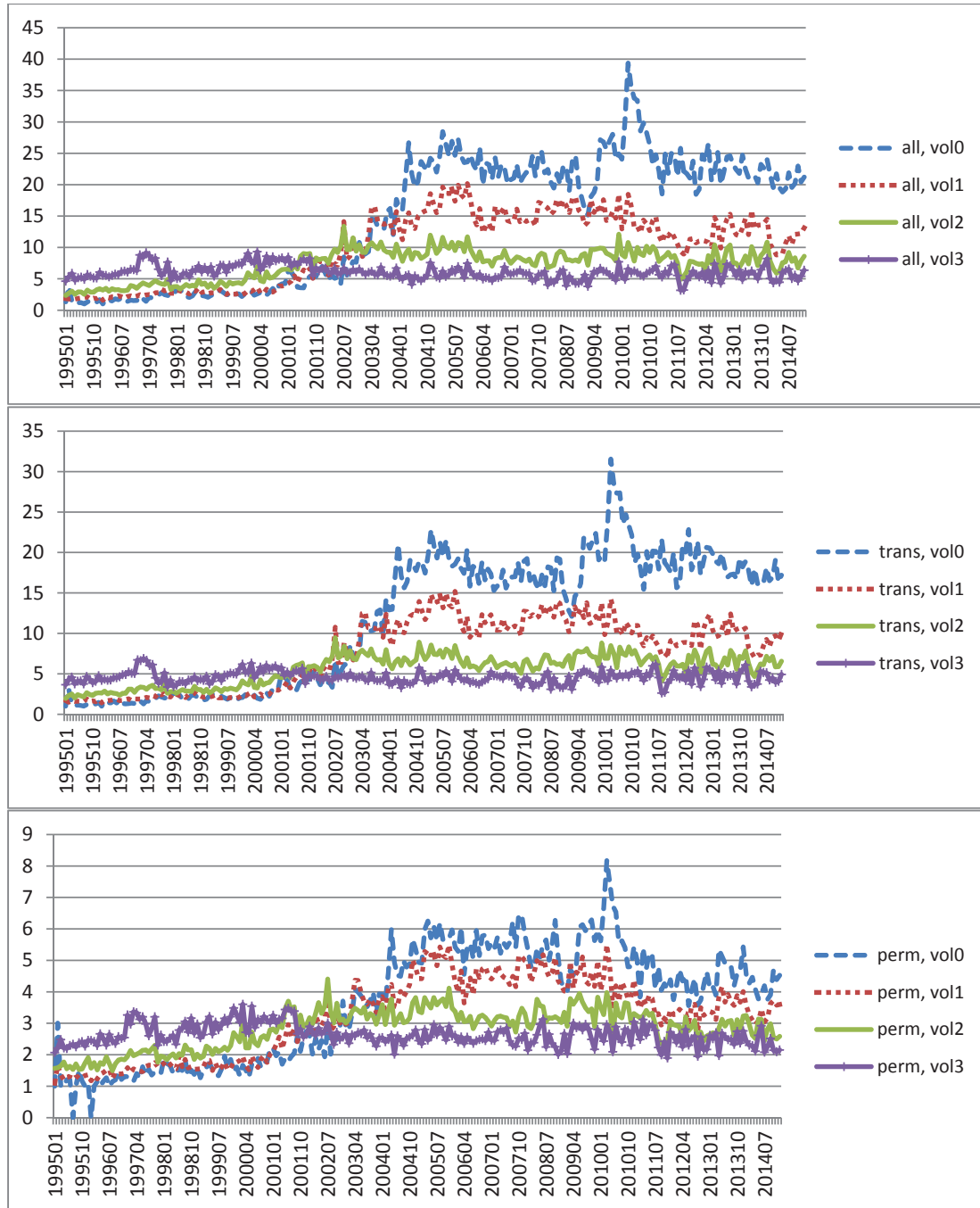
This table reports the number of jumps per jump day. In each month, the jump stocks are grouped into quartiles based on the market cap. Quartile 0 has the smallest market cap while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.4: The percentage of jump stocks for market capitalization quartiles



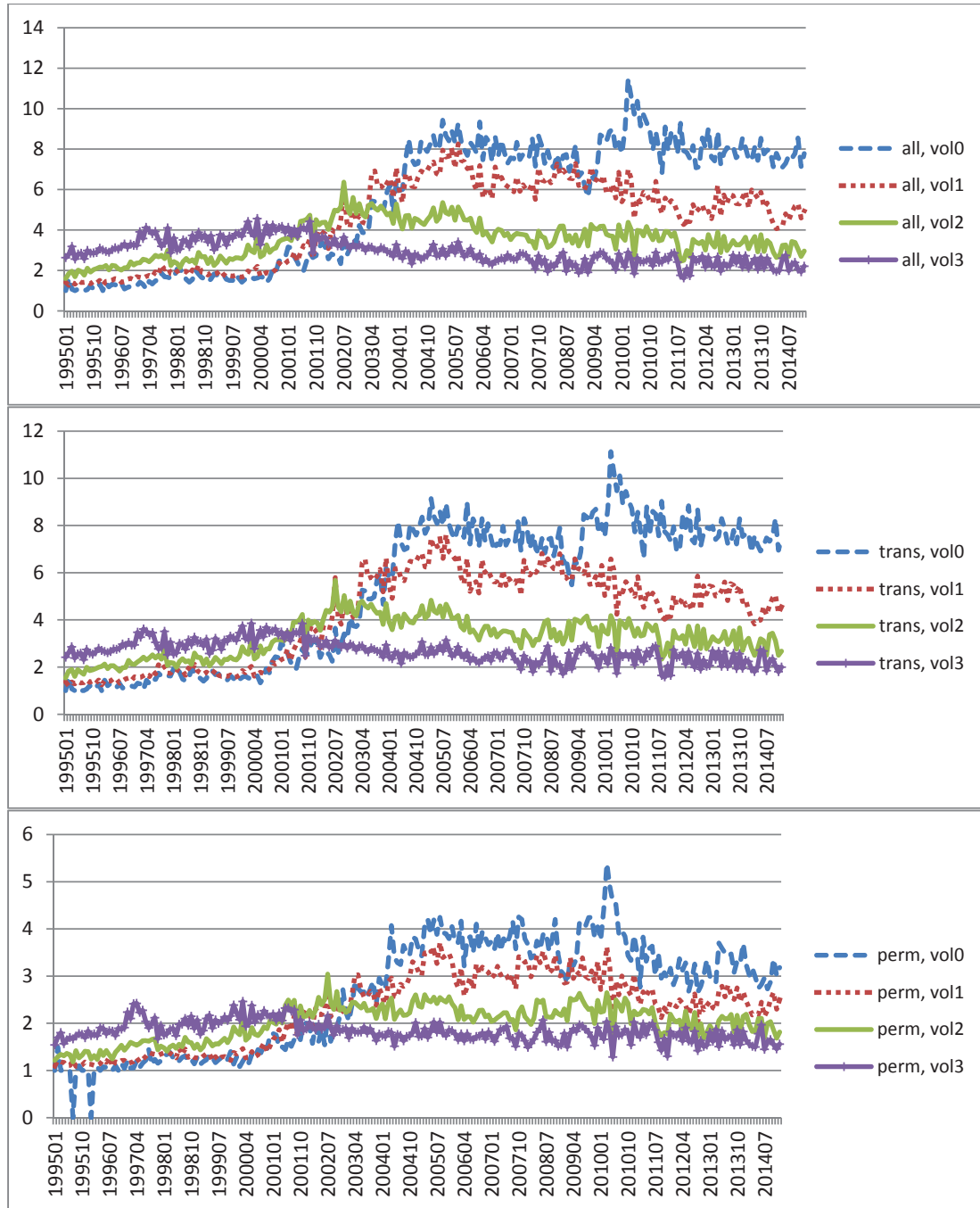
This table reports the percentage of jump stocks. In each month, the jump stocks are grouped into quartiles based on the market cap. The percentage of jump stocks is the number of the jump stocks divided by the total number of stocks in each quartile. Quartile 0 has the smallest market cap while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.5: Average numbers of jumps in a month for volume quartiles



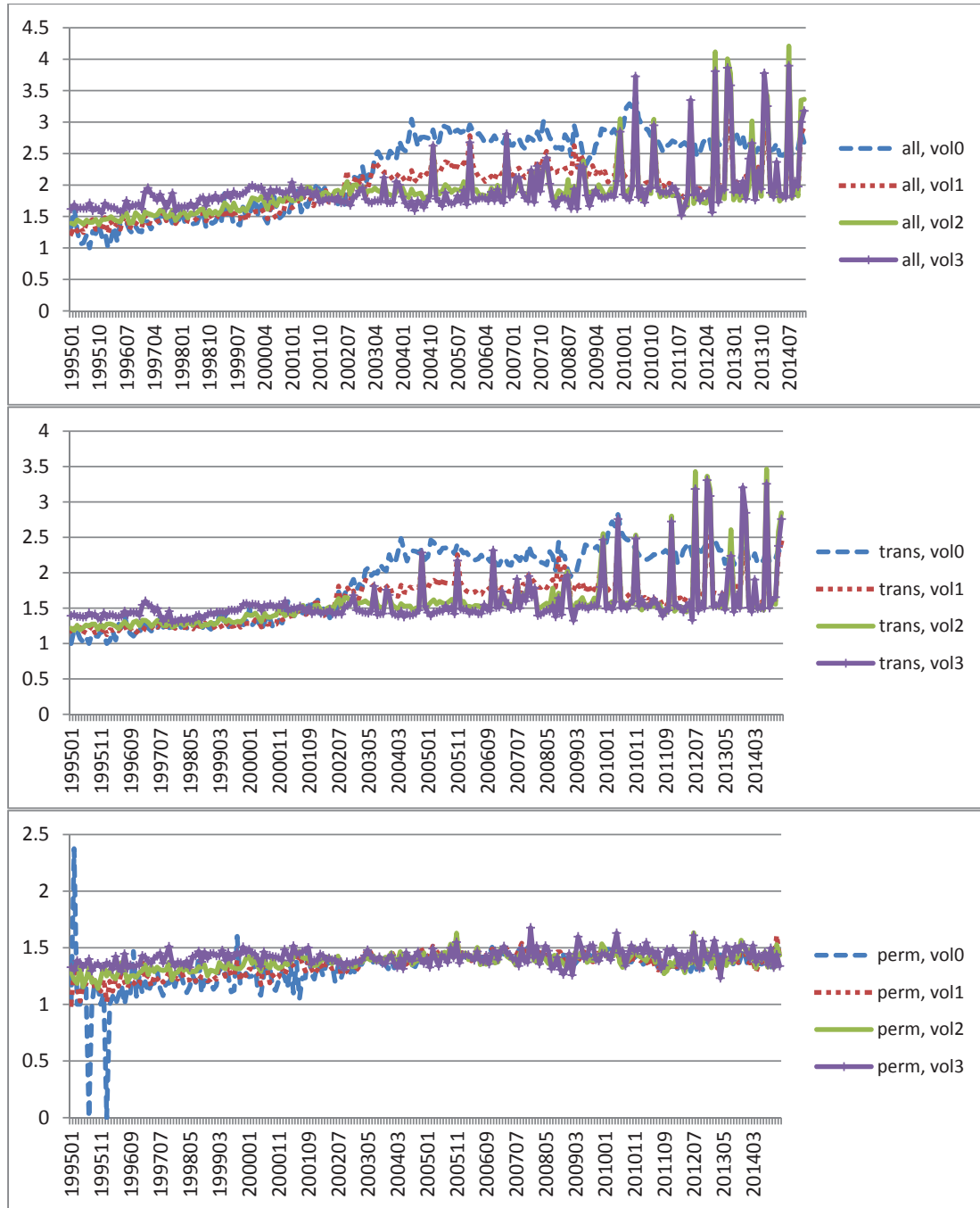
This table reports the number of jumps in a month. In each month, the jump stocks are grouped into quartiles based on the trading volume. Quartile 0 has the smallest volume while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.6: Average numbers of jump days in a month for volume quartiles



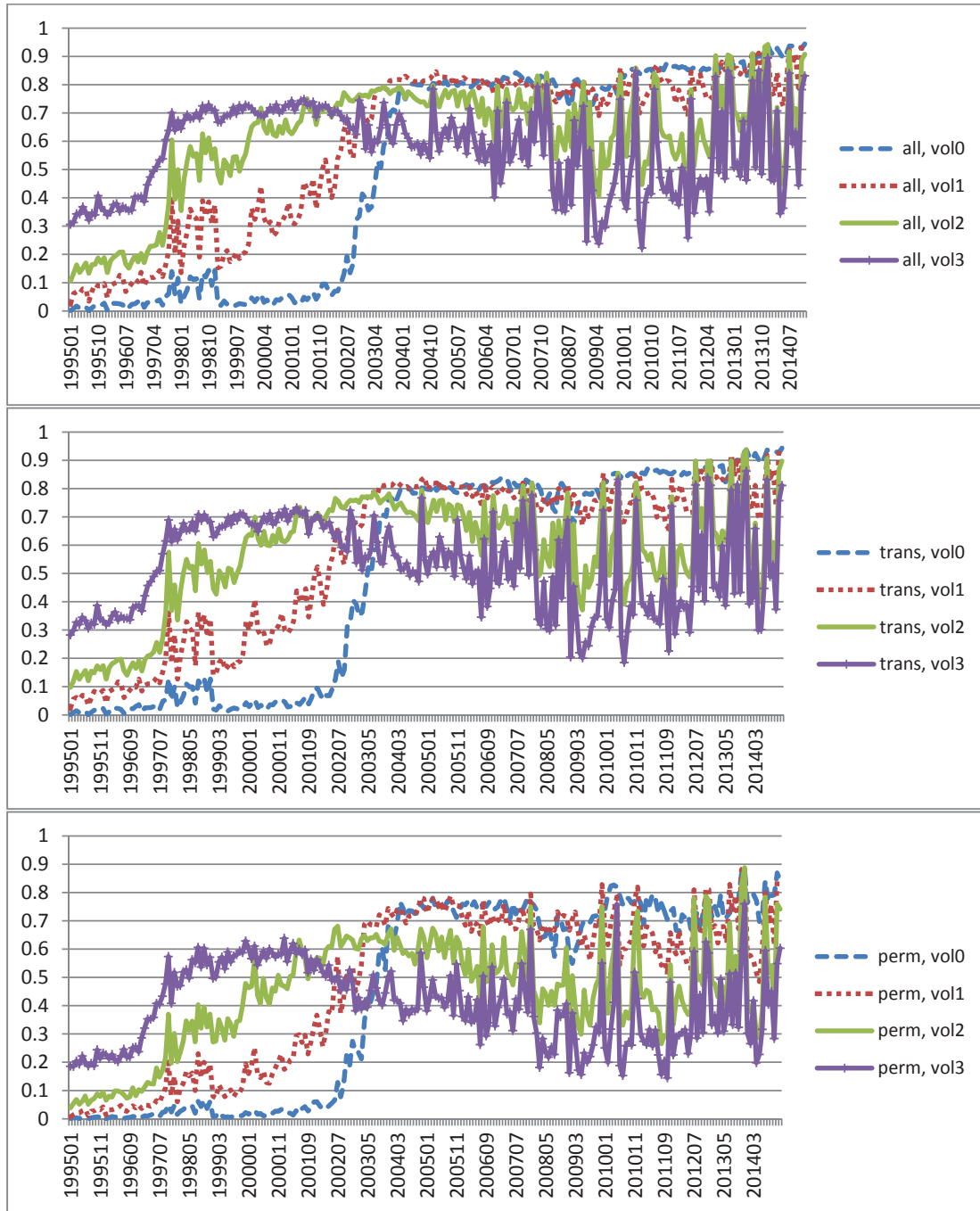
This table reports the number of jump days in a month. In each month, the jump stocks are grouped into quartiles based on the trading volume. Quartile 0 has the smallest volume while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.7: Average numbers of jumps per jump day in a month for volume quartiles



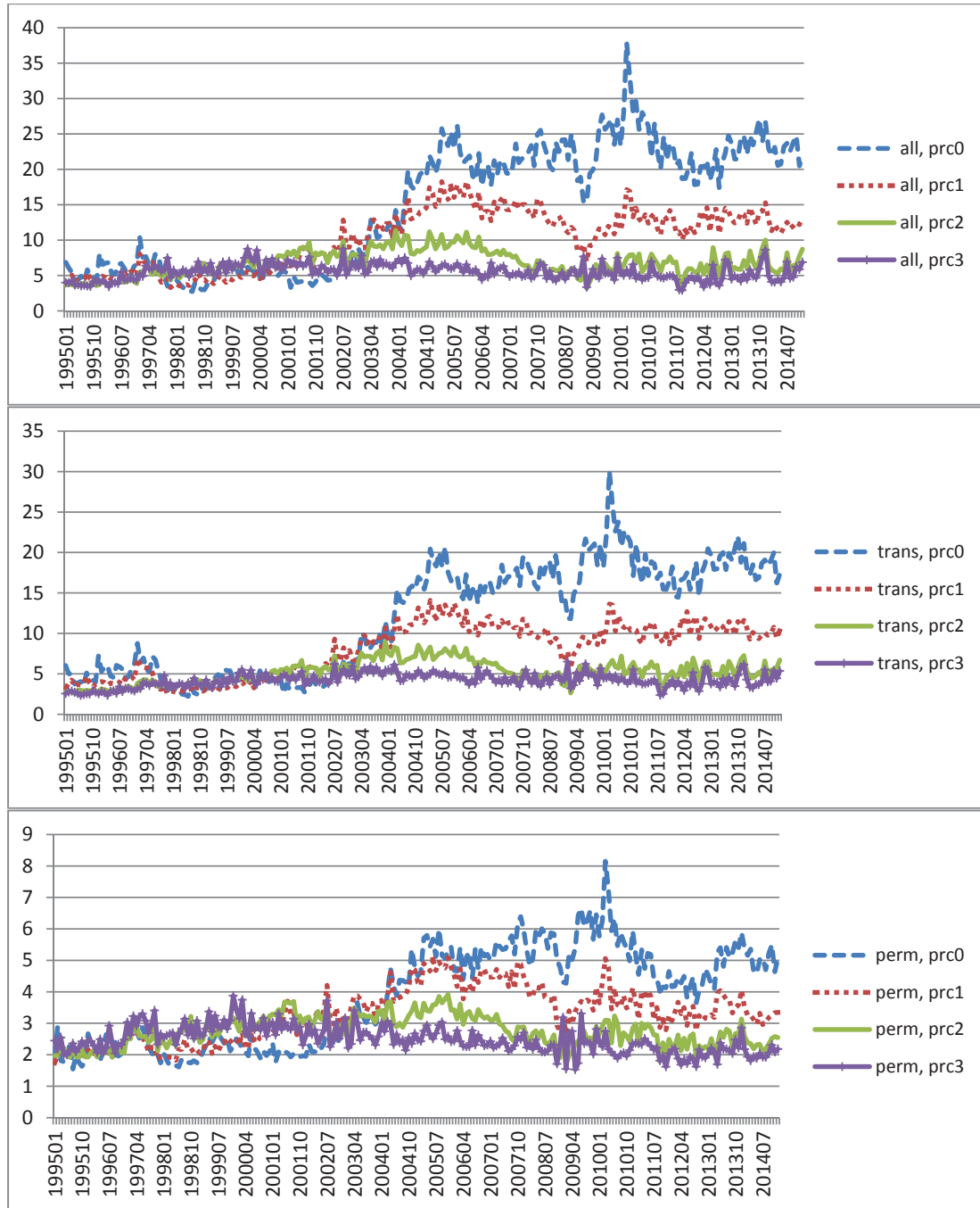
This table reports the number of jumps per jump day in a month. In each month, the jump stocks are grouped into quartiles based on the trading volume. Quartile 0 has the smallest volume while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.8: The percentage of jump stocks for volume quartile



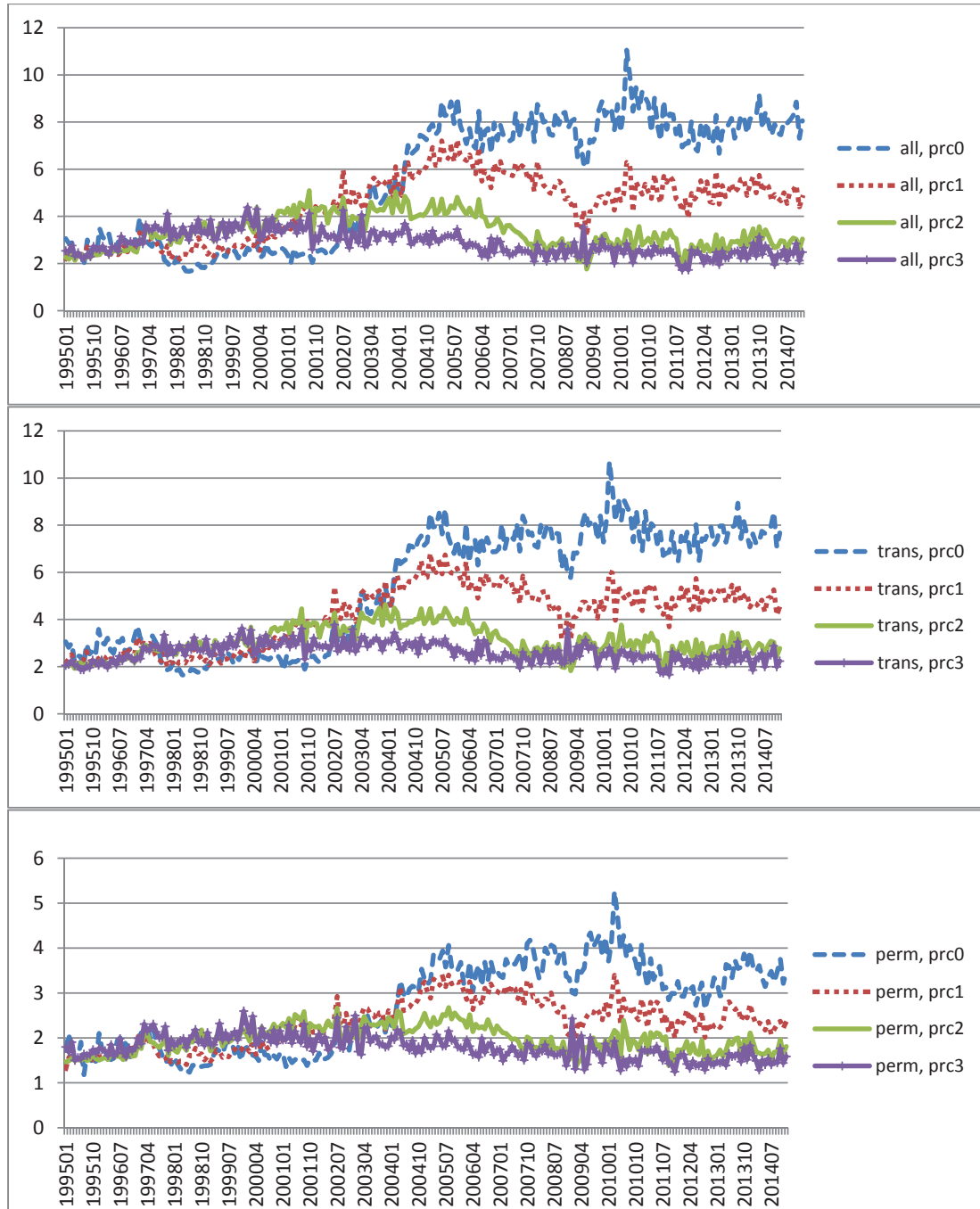
This table reports the percentage of jump stocks. In each month, the jump stocks are grouped into quartiles based on trading volume. The percentage of jump stocks is the number of the jump stocks divided by the total number of stocks in each quartile. Quartile 0 has the smallest volume while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.9: Average numbers of jumps in a month for price buckets



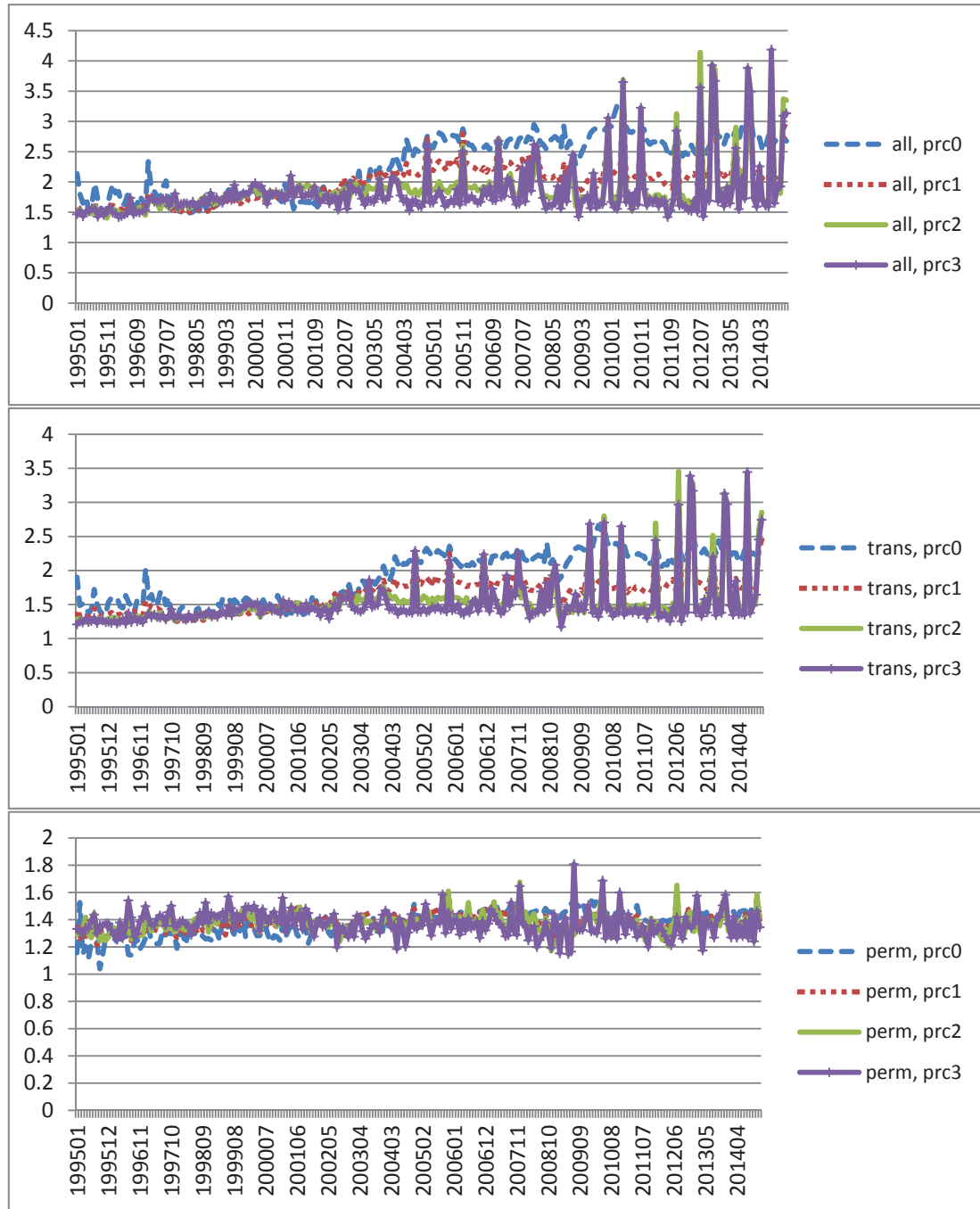
This table reports the number of jumps in a month. In each month, the jump stocks are grouped into buckets based on the stock price: smaller than 5 dollars (prc0), 5 to 25 dollars (prc1), 25 to 50 dollars (prc2), and greater than 50 dollars (prc3). First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.10: Average numbers of jump days in a month for price buckets



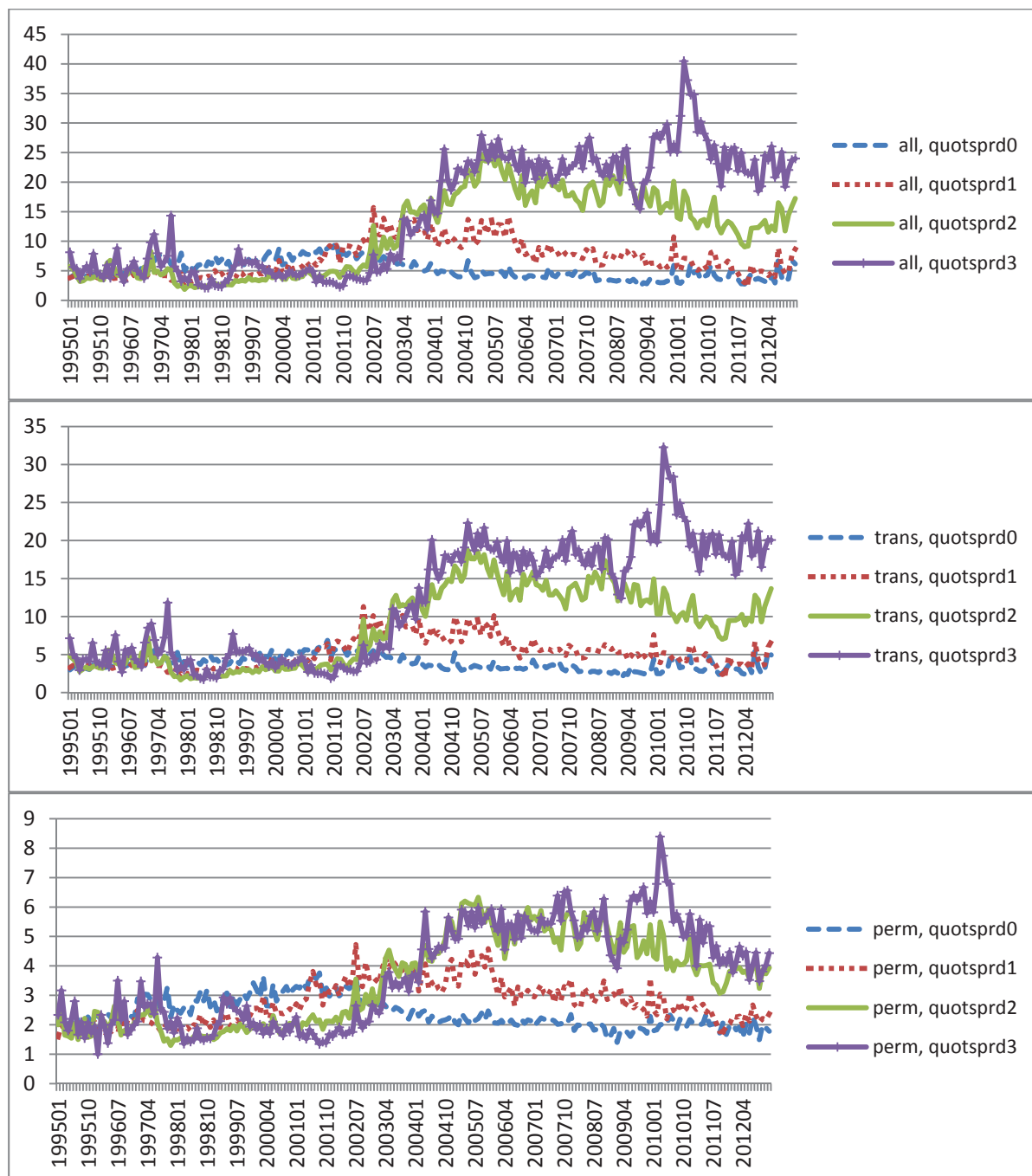
This table reports the number of jump days in a month. In each month, the jump stocks are grouped into buckets based on the stock price: smaller than 5 dollars (prc0), 5 to 25 dollars (prc1), 25 to 50 dollars (prc2), and greater than 50 dollars (prc3). First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.11: Average numbers of jumps per jump day in a month for price buckets



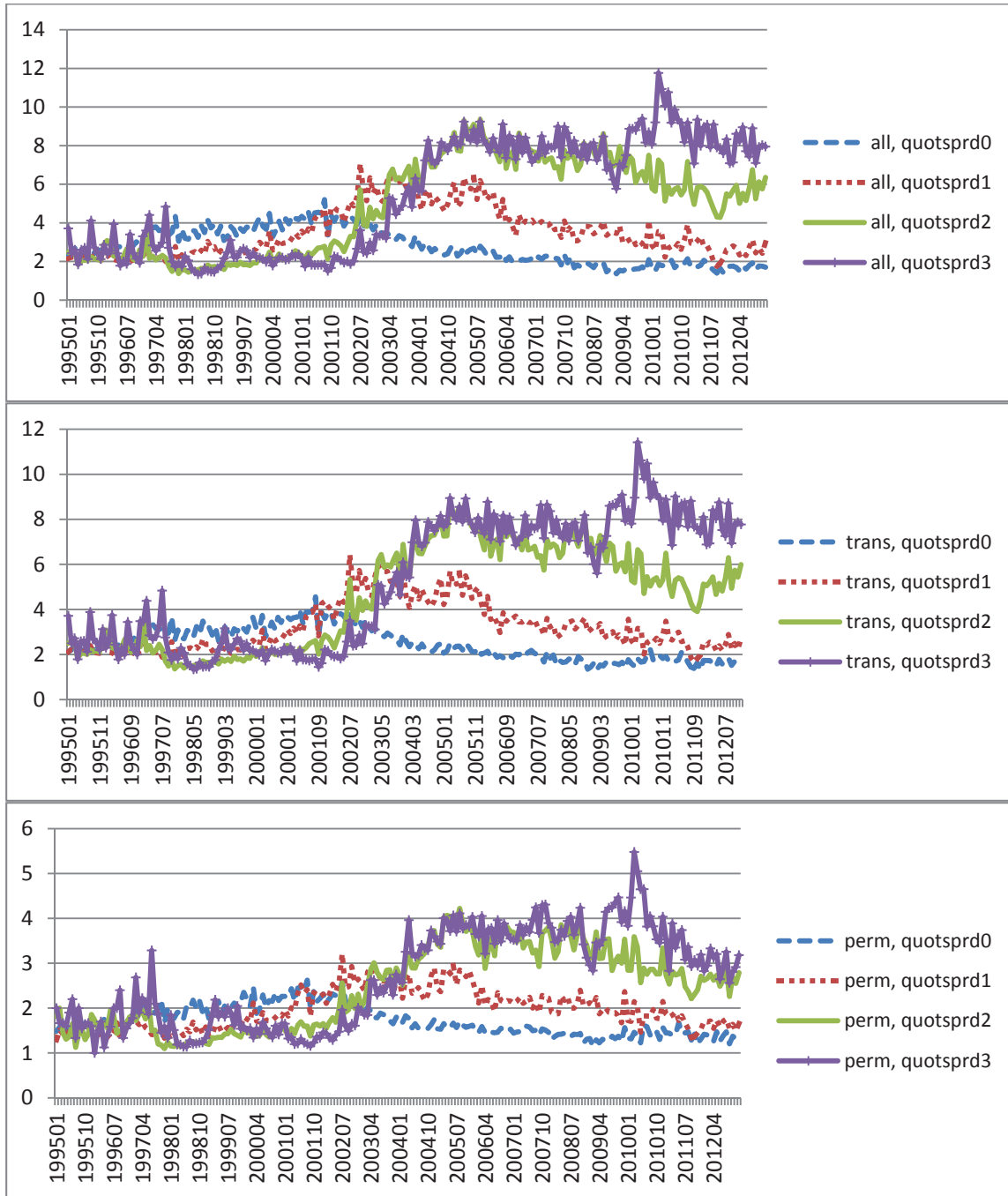
This table reports the number of jumps per jump day in a month. In each month, the jump stocks are grouped into buckets based on the stock price: smaller than 5 dollars (prc0), 5 to 25 dollars (prc1), 25 to 50 dollars (prc2), and greater than 50 dollars (prc3). First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.12: Average numbers of jumps in a month for quoted spread quartiles



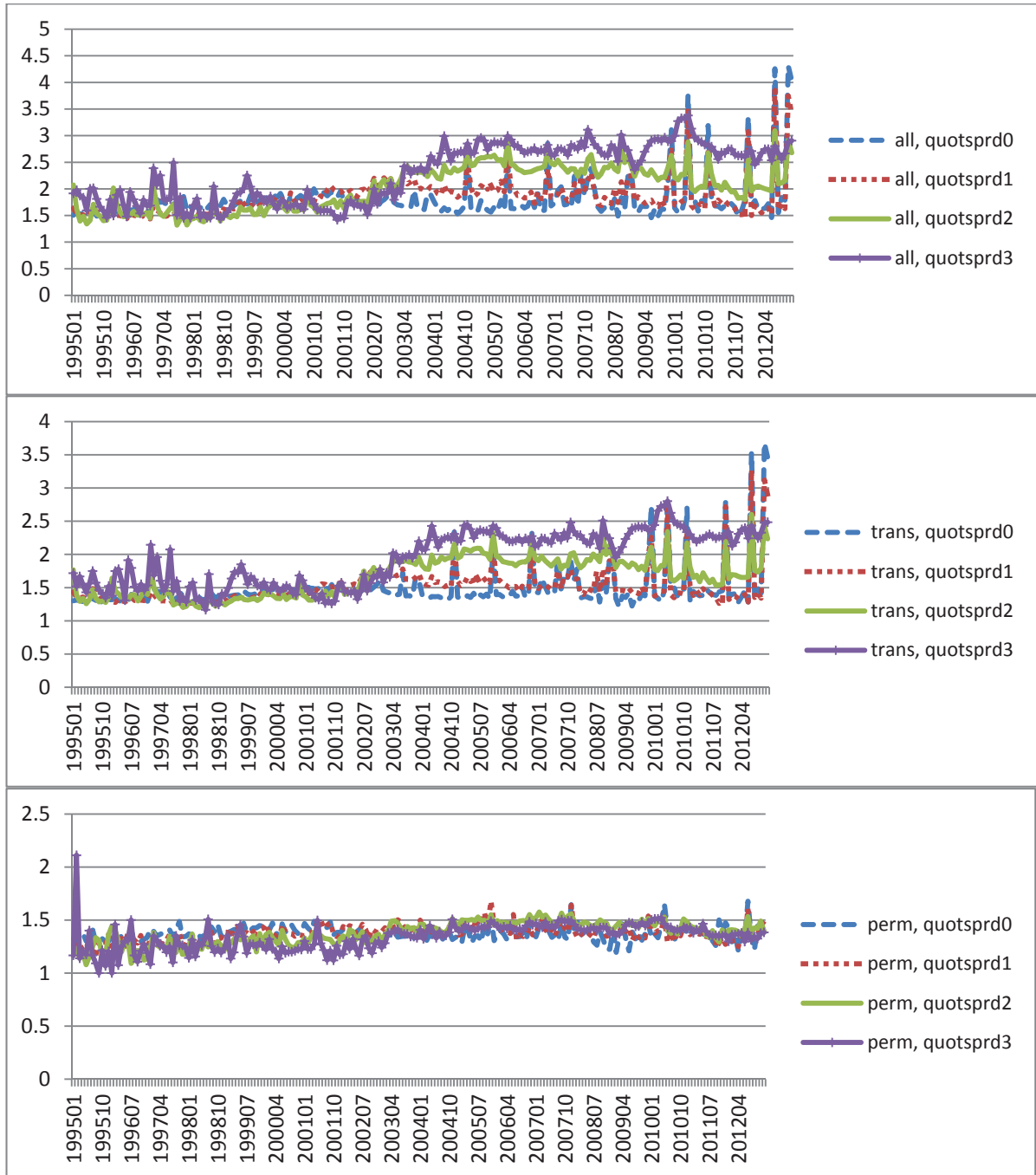
This table reports the number of jumps in a month. In each month, the jump stocks are grouped into quartiles based on the quoted bid-ask spread. Quartile 0 has the smallest quoted spread while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2012.

Figure 4.13: Average numbers of jump days in a month for quoted spread quartiles



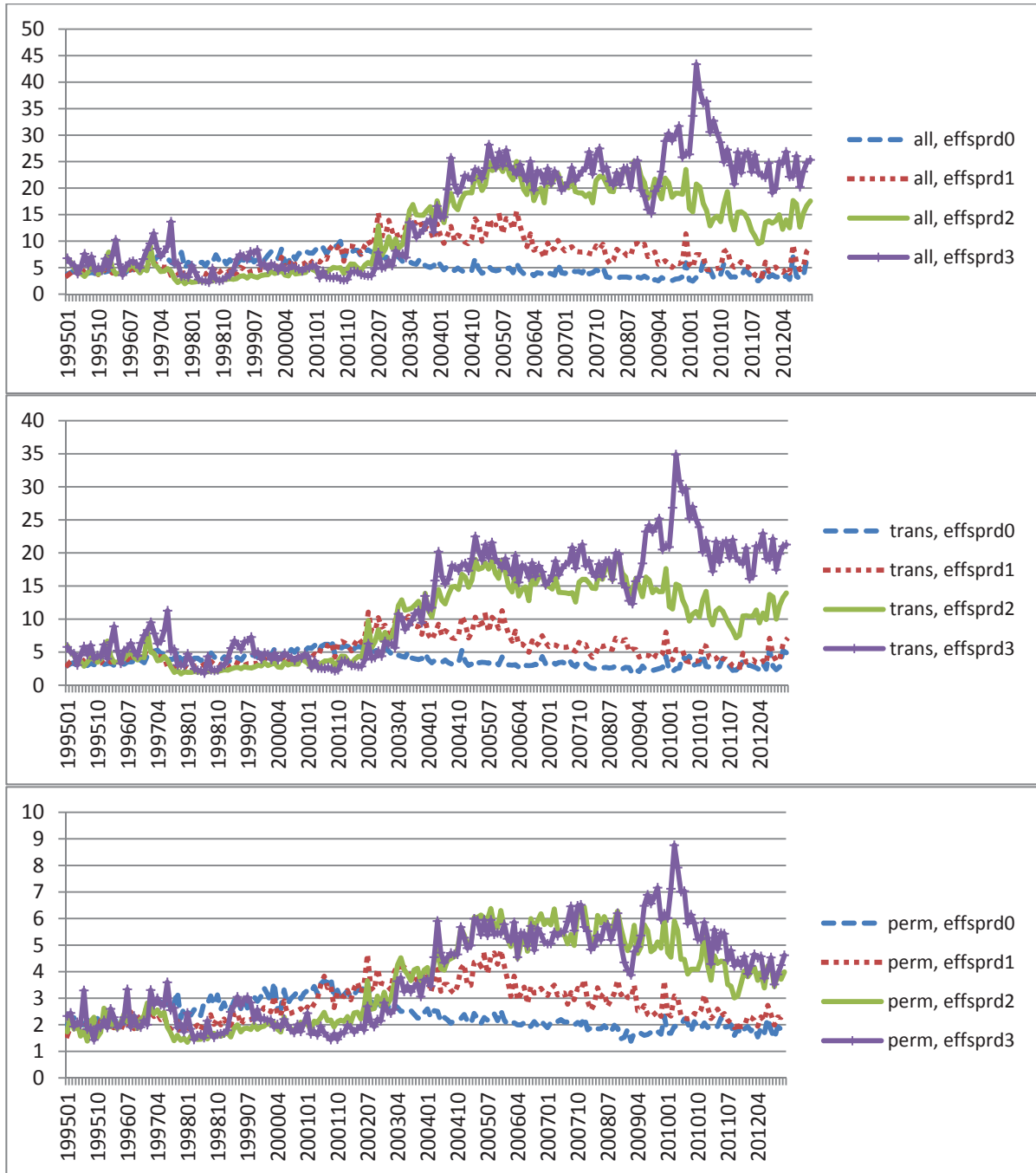
This table reports the number of jump days in a month. In each month, the jump stocks are grouped into quartiles based on the quoted bid-ask spread. Quartile 0 has the smallest quoted spread while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2012.

Figure 4.14: Average numbers of jumps per jump day in a month for quoted spread quartiles



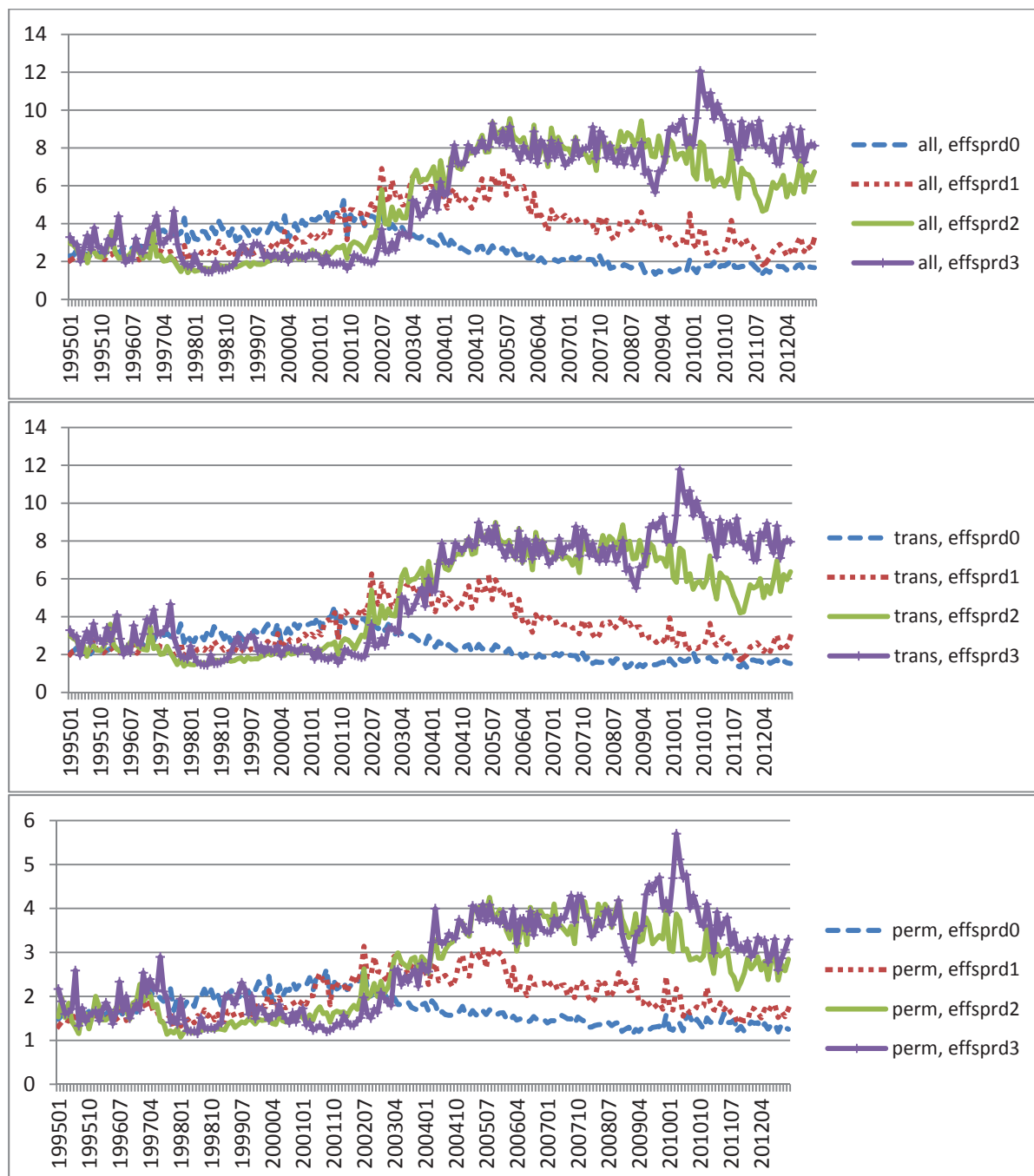
This table reports the number of jumps per jump day in a month. In each month, the jump stocks are grouped into quartiles based on the quoted bid-ask spread. Quartile 0 has the smallest quoted spread while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2012.

Figure 4.15: Average numbers of jumps in a month for effective spread quartiles



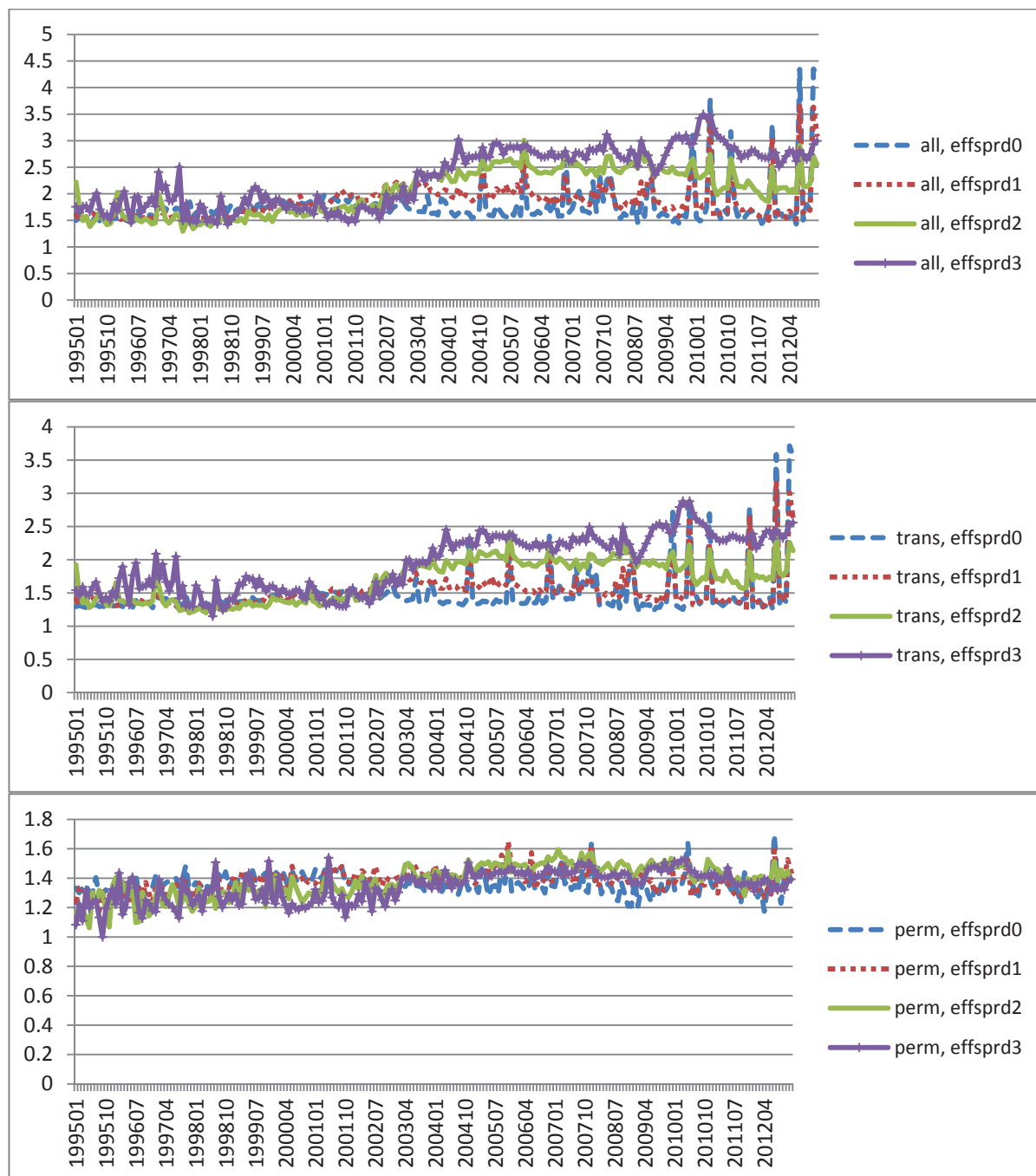
This table reports the number of jumps in a month. In each month, the jump stocks are grouped into quartiles based on the effective bid-ask spread. Quartile 0 has the smallest effective spread while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2012.

Figure 4.16: Average numbers of jump days within a month for effective spread quartiles



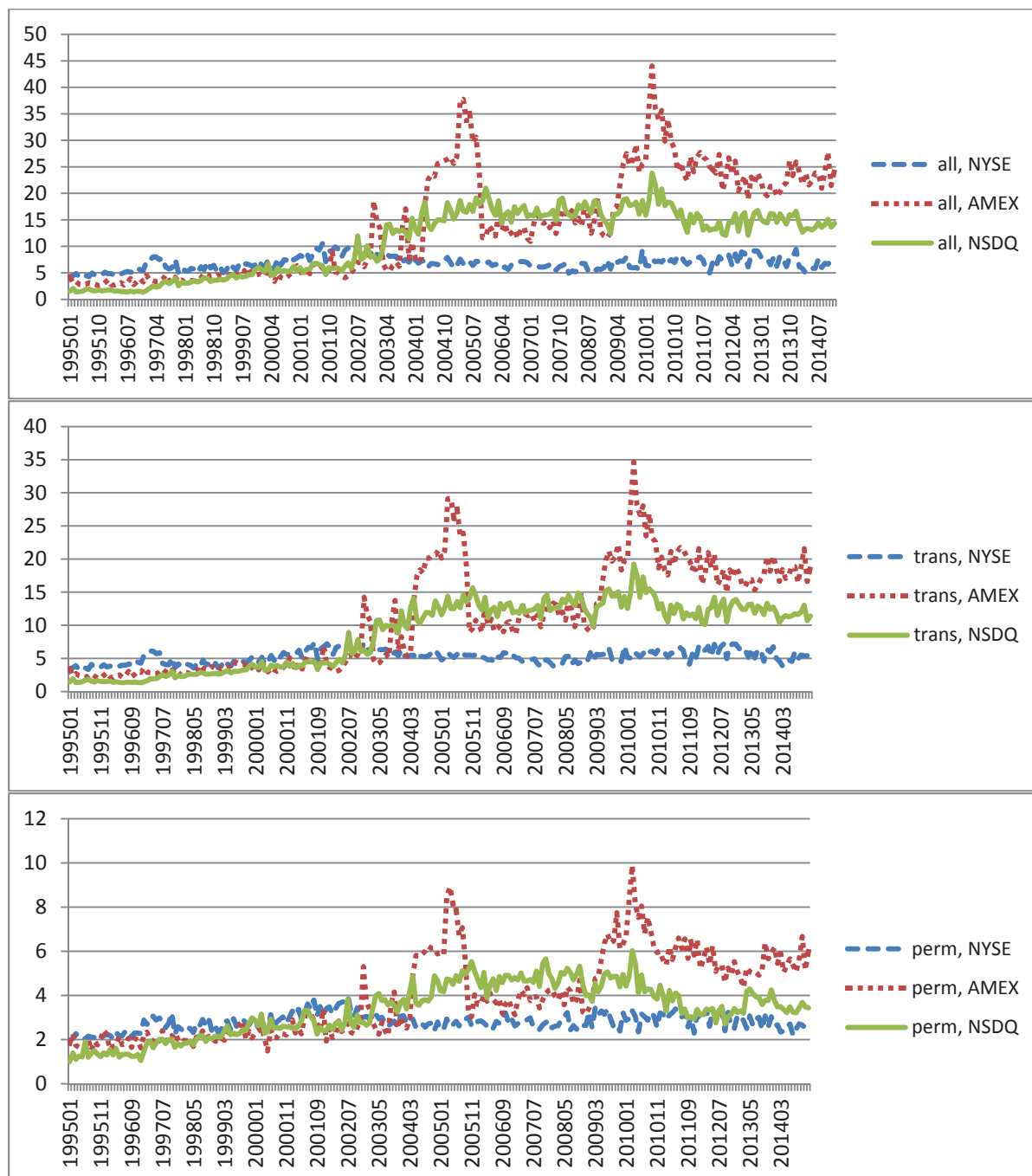
This table reports the number of jump days in a month. In each month, the jump stocks are grouped into quartiles based on the effective bid-ask spread. Quartile 0 has the smallest effective spread while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2012.

Figure 4.17: Average numbers of jumps per jump day within a month for effective spread quartiles



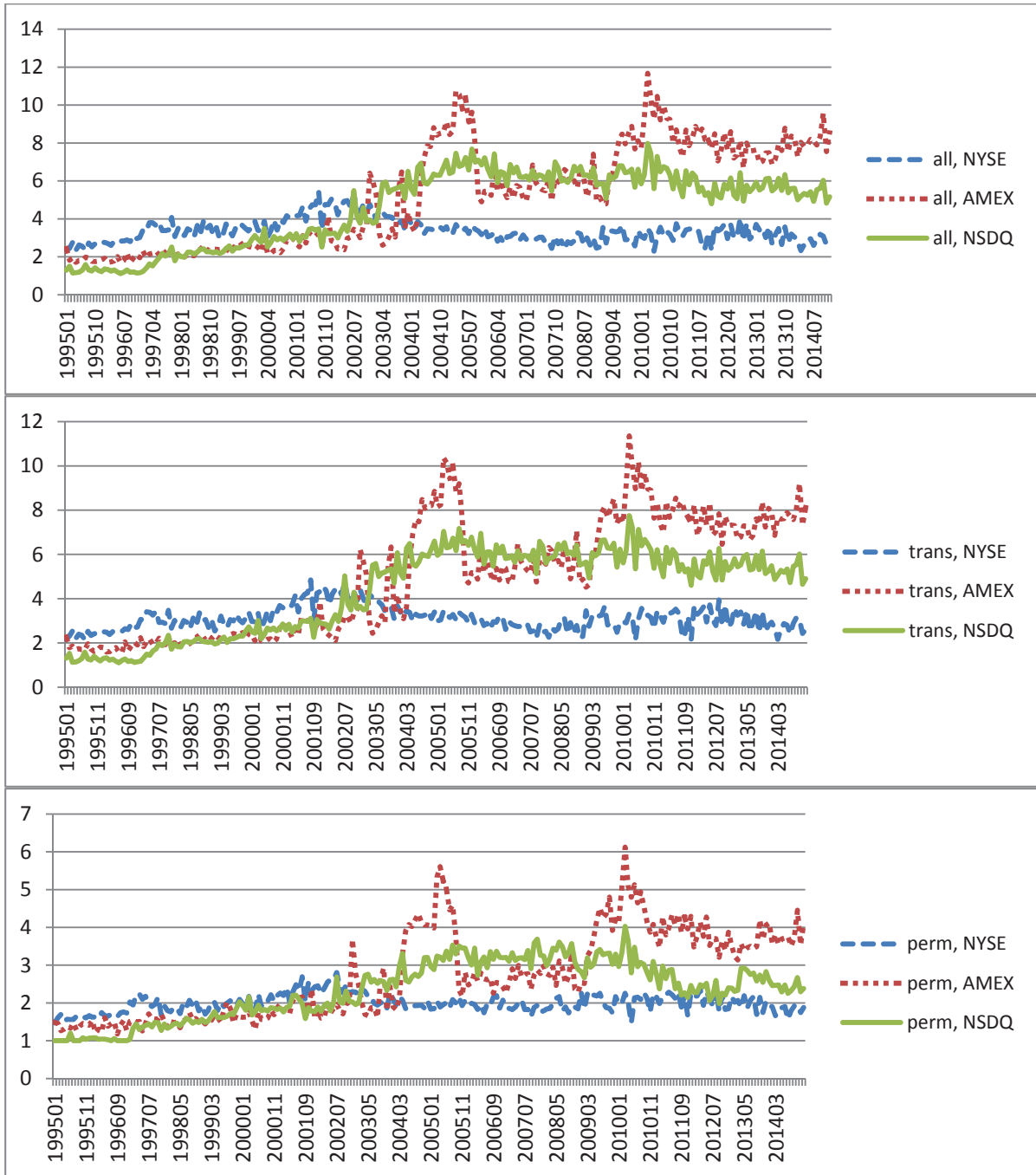
This table reports the number of jumps per jump day in a month. In each month, the jump stocks are grouped into quartiles based on the effective bid-ask spread. Quartile 0 has the smallest effective spread while quartile 3 has the greatest. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2012.

Figure 4.18: Average numbers of jumps in a month for exchange buckets



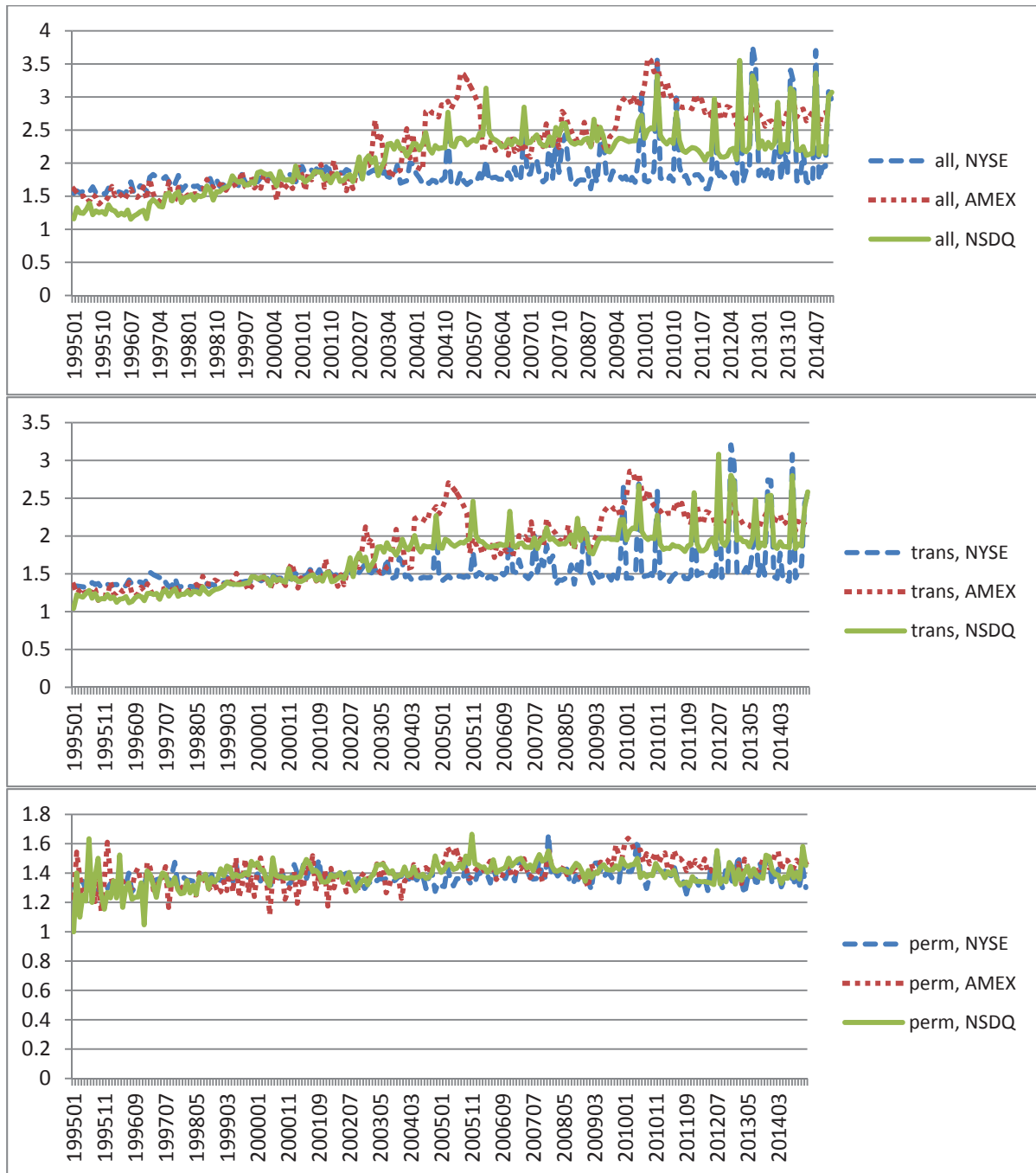
This table reports the number of jumps in a month. The jump stocks are grouped into buckets based on the exchanges on which the stock are listed: the exchange codes, which are respectively 1, 2, and 3 for NYSE, AMEX, and Nasdaq. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.19: Average numbers of jump days in a month for exchange buckets



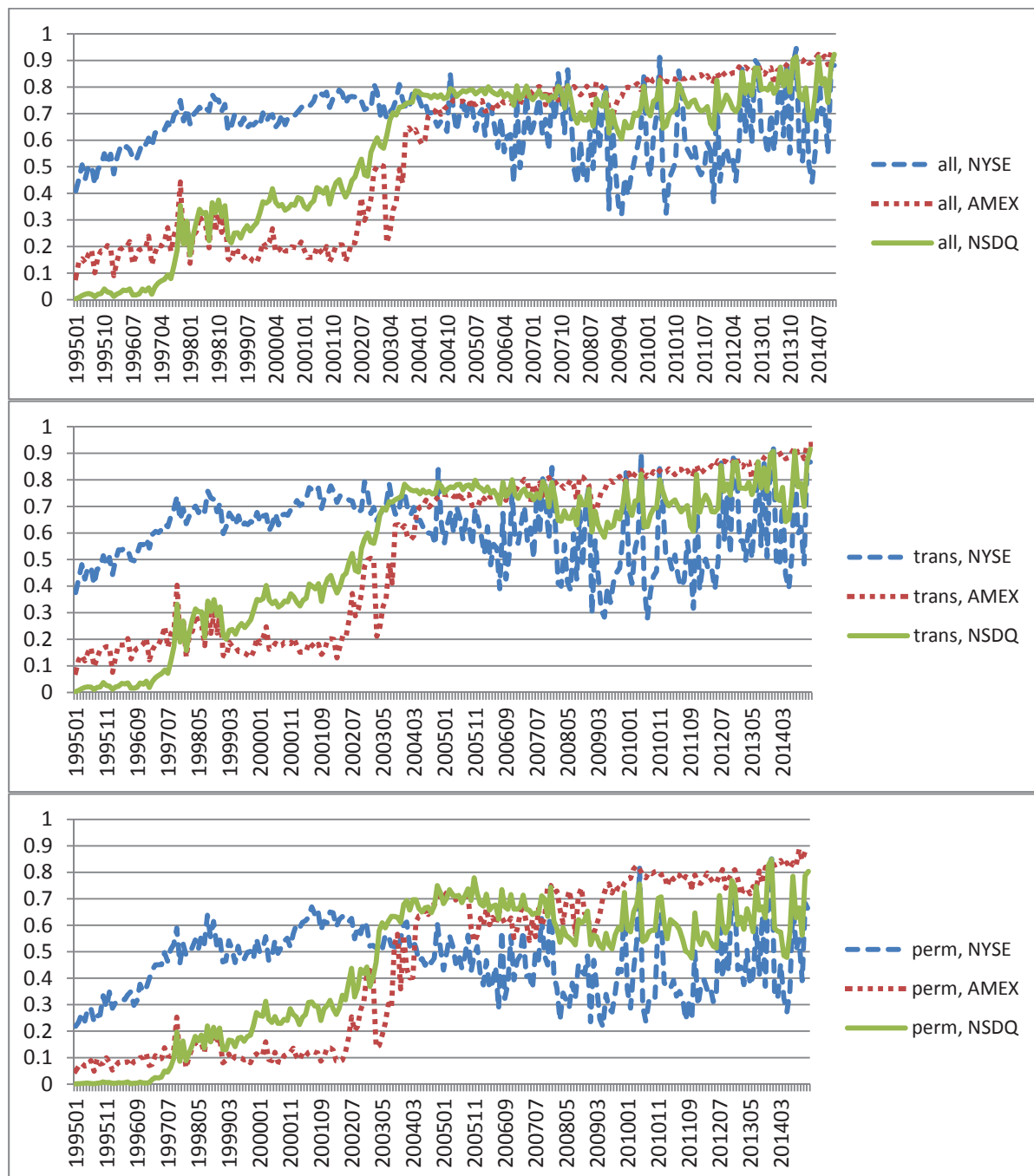
This table reports the number of jump days in a month. The jump stocks are grouped into buckets based on the exchanges on which the stock are listed: the exchange codes, which are respectively 1, 2, and 3 for NYSE, AMEX, and Nasdaq. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.20: Average numbers of jumps per jump day in a month for exchange buckets



This table reports the number of jumps per jump day in a month. The jump stocks are grouped into buckets based on the exchanges on which the stock are listed: the exchange codes, which are respectively 1, 2, and 3 for NYSE, AMEX, and Nasdaq. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Figure 4.21: The percentage of jump stocks for different exchanges buckets



This table reports the percentage of jump stocks for different exchanges. The jump stocks are grouped into buckets based on the exchanges on which the stock are listed: the exchange codes, which are respectively 1, 2, and 3 for NYSE, AMEX, and Nasdaq. The percentage of jump stocks is the number of the jump stocks divided by the total number of stocks. First subfigure shows the results of all jumps, including both transient and permanent jumps. Second subfigure shows the results of the transient jumps. Third subfigure shows the results for the permanent jumps. The sample period is January 1995 to December 2014.

Chapter 5

Concluding Remarks

This thesis covers three different projects in the areas of market microstructure and exchange rate. These original studies shed some new light on their respective research topics. The first project on central bank intervention, from a market microstructure perspective. The second calibrates the exchange rate predictability. The third project focuses on price jumps, especially the transient jumps, in the high frequency data, and the related market stability issue.

The paper on central bank intervention evaluates the effects of the central bank intervention on different variables. In particular, it stresses the severe endogeneity problem and the importance of instrumental variable. With the help of the instrumental variable, we conduct a thorough empirical investigation on the effects of the central bank intervention. The empirical results show that the instrumental variable does make a difference in the case where the intervention is endogenous with respect to the dependent variable. Moreover, we measure the persistence of the intervention's price impact, and also confirm that the intervention's price impact would be high when the volatility is high based on Markov switching models. These results have very practical application for the central banks.

The second study on calibrating the exchange rate predictability aims at measuring the information transfer from the predictors to the future return, thus provides us with an upper bound for any forecasting model based on the given set of predictors. In this way, we can distinguish whether the poor forecasting performance is due to lack of information in the predictors or the misspecification of the model. The importance is that in the last three decades, the model-

dependent studies constantly fail to outperform the random walk benchmark systematically. In our paper, we find that the issue is due to lack of information. Trying different forecasting model would not help. However, the intraday exchange rate movement is systematically predictable. The nonparametric model-independent approach proposed in this paper is also very useful for predictors and forecasting model selection. For future research, this method can be applied to other time series predictability problems.

The third study focuses on the jumps in intraday price movements, especially the transient jumps, the flash crash type of extreme price fluctuations. By identifying all the transient and permanent jumps in the TAQ data, we document cross-sectional variations in the price jumps and the structural change of the jump properties around 2003. While the most liquid, large cap, and high-price stocks become more stable, the illiquid, thinly-traded, small cap, and low-priced stocks suffer from more transient jumps after 2003. The empirical findings have never been documented before. We also discuss the theoretical mechanism: for the thinly-traded small cap stocks, traditional market makers are crowded out by HFTs after the latency reduction, and as endogenous liquidity providers, HFTs have the less long-term risk-bearing capacity and they might withdraw liquidity or stop making the market under adverse market conditions. The results are useful for the policy makers in charge of the market design. This project studies the U.S. equity market, further research can be extended to the price stability in the international stocks across different countries. Another direction is to identify the direct association between HFT and the increased stability in the illiquid, small cap, and low-priced stocks.

Bibliography

Aguilar, J. and S. Nydahl (2000), “Central bank intervention and exchange rates: The case of Sweden,” *Journal of International Financial Markets, Institutions, and Money*, 10, 303-322.

Aït-Sahalia, Y. (2002), “Telling from discrete data whether the underlying continuous time model is a diffusion, *The Journal of Finance* 57, 2075-2112.

Aït-Sahalia, Y. (2004), “Disentangling Diffusion from Jumps, *Journal of Financial Economics* 74, 487-528.

Aït-Sahalia, Y., J. Jacod (2009), “Testing for jumps in a discretely observed process, *Annals of Statistics* 37, 184-222.

Alquist, R. and M.D. Chinn (2008), “Conventional and Unconventional Approaches to Exchange Rate Modelling and Assessment,” *International Journal of Finance and Economics* 13, 2-13.

Andersen, T.G., L. Benzoni (2009), “Realized Volatility, In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, Th. (Eds.) *Handbook of Financial Time Series*, Springer.

Andersen, T., Bondarenko, O., 2014a. “VPIN and the flash crash.” *Journal of Financial Markets* 17, 1-46.

Andersen, T., Bondarenko, O., 2014b. “Reflecting on the VPIN dispute.” *Journal of Financial Markets* 17, 53-64.

Barndorff-Nielsen, O., and N. Shephard (2004), “Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics*, 2 (1), 1-37.

Barndorff-Nielsen, O., and N. Shephard (2006), “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation, *Journal of Financial Econometrics* 4 (1), 1-30.

Barnett, L., A.B. Barnett and A.K. Seth (2009), “Granger causality and transfer entropy are

equivalent for Gaussian variables,” *Physical review letters* 103 (23), 238701.

Beine, M., J. Lahaye, S. Laurent, C. J. Neely, and Palm, F. C. (2007), “Central bank intervention and exchange rate volatility, its continuous and jump components,” *International Journal of Finance and Economics*, 12, 201-223.

Biais, B. and T. Foucault (2014), “HFT and market quality, *Bankers, Markets & Investors* 128 (1), 5-19.

Boehmer, E., Fong, K. and Wu, J. (2014), “International evidence on algorithmic trading”, AFA 2013 San Diego Meetings Paper, available at SSRN: <http://ssrn.com/abstract=2022034>.

Brogaard, J. (2011), “HFT and Volatility, Working Paper, Washington University.

Brogaard, J., A. Carrion, T. Moyaert, R. Riordan, A. Shkilko, K. Sokolov (2016) “High-Frequency Trading and Extreme Price Movements, Available at SSRN: <https://ssrn.com/abstract=2531122> or <http://dx.doi.org/10.2139/ssrn.2531122>

Carlson, J. A. and M. Lo (2006), “One Minute in the Life of the DM/USD: Public News in an Electronic Market,” *Journal of International Money and Finance*, 25(7), 1090-1102.

Carr, P., and L. Wu (2003), “What type of process underlies options? A simple robust test. *Journal of Finance* 58, 2581-2610.

Chaboud, A. P. and O. F. Humpage (2005), “An Assessment of the Impact of Japanese Foreign-Exchange Intervention: 1991-2004,” International Finance Discussion Paper No. 824, Board of Governors of the Federal Reserve System.

Chamon, M., M. Garcia, and L. Candido de Souza (2015), “FX Interventions in Brazil: A Synthetic Control Approach,” Textos para discussao 630, PUC Department of Economics: Rio de Janeiro.

Cheung, Y.W., M.D. Chinn, A. Garcia Pascual (2005), “Empirical exchange rate models of the nineties: are any fit to survive?” *Journal of International Money and Finance* 24, 1150-1175.

Chinn, M.D. and R. Meese (1995), “Banking on Currency Forecasts: Is Change in Money Predictable?,” *Journal of International Economics* 38 (1-2), 161-78.

Chinn, M. D. and Moore, M. J. (2011), “Order flow and the monetary model of exchange rates:

- evidence from a novel data set,” *Journal of Money, Credit and Banking* 43(8), 1599-1624.
- Christensen, K., R. C. A. Oomen, and M. Podolskij (2014), “Fact or friction: Jumps at ultra high-frequency, *Journal of Financial Economics* 114, 576-599.
- Clark, T.E. and K.D. West (2006), “Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis,” *Journal of Econometrics* 135, 155-186.
- Dominguez, K. M. (1998), “Central bank intervention and exchange rate volatility,” *Journal of International Money and Finance* 17, 161-190.
- Dominguez, K. M. (2003), “The Market Microstructure of Central Bank Intervention,” *Journal of International Economics* 59(1), 25-45.
- Dominguez, K. M. (2006), “When do central bank interventions influence intra-daily and longer-term exchange rate movements?” *Journal of International Money and Finance* 25, 7, 1051-1071.
- Easley, D., M. Lopez de Prado, M. O’Hara (2012), “Flow toxicity and liquidity in a high-frequency world.” *Review of Financial Studies* 25, 1457-1493.
- Easley, D., M. Lopez de Prado, M. O’Hara (2015), “Optimal Execution Horizon.” *Mathematical Finance* 25(3), 640-672.
- Edison, H. J., P. A. Cashin, and H. Liang (2003), “Foreign Exchange Intervention and the Australian Dollar: Has It Mattered?” Working Paper No. 03/99, *International Monetary Fund*.
- Engel C., Hamilton J. (1990), “Long swings in the dollar: Are they in the data and do markets know it?” *American Economic Review* 80, 689-713.
- Engel, C., N.C. Mark, K.D. West (2008), “Exchange rate models are not as bad as you think,” In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), *NBER Macroeconomics Annual*, 2007. MIT Press, Cambridge, MA.
- Evans, M. D. D. and R. K. Lyons (2000), “The Price Impact of Currency Trades: Implications for Intervention,” *working paper*, U.C. Berkeley.
- Evans, M. D. D. and R. K. Lyons (2002), “Order Flow and Exchange Rate Dynamics,” *Journal of Political Economy* 110, 170-180.

- Evans, M. D. D. and R. K. Lyons (2005), "Meese-Rogoff Redux: Micro-Based Exchange-Rate Forecasting." *American Economic Review*, 95(2), 405-414.
- Evans, M. D. D. and R. K. Lyons (2006), "Understanding Order Flow." *International Journal of Finance and Economics* 11, 3-23.
- Evans, M. D. D. and R. K. Lyons (2008), "How is Macro News Transmitted to Exchange Rates?" *Journal of Financial Economics* 88(1), 26-50.
- Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25:2, 383-417.
- Fatum, R. and M. M. Hutchison (2003), "Is sterilised foreign exchange intervention effective after all? An event study approach." *Economic Journal* 113(487), 390-411.
- Fatum, R. and M. M. Hutchison (2006), "Effectiveness of official daily foreign exchange market intervention operations in Japan." *Journal of International Money and Finance* 25, 199-219.
- Fatum, R. and M. M. Hutchison (2010), "Evaluating foreign exchange market intervention: Self-selection, counterfactuals and average treatment effects," *Journal of International Money and Finance* 29, 3, 570-584
- Fatum, R. and M.R. King (2005), "Rules versus discretion in foreign exchange intervention: evidence from official Bank of Canada high-frequency data," Working Paper, University of Alberta.
- Fischer, A. M. and M. Zurlinden (1999), "Exchange Rate Effects of Central Bank Interventions: An Analysis of Transaction Prices," *Economic Journal* 109(458), 662-76.
- Fischer, A. (2006), "On the inadequacy of newswire reports for empirical research on foreign exchange interventions," *Journal of International Money and Finance* 25, 8, 1226-1240.
- Frankel, J.A. and A.K. Rose (1995), "Empirical Research on Nominal Exchange Rates," in: Grossman, G. and K.S. Rogoff (eds.), *Handbook of International Economics* Vol. 3, Elsevier-North Holland.
- Fratzscher M. (2008), "Oral interventions versus actual interventions in FX markets - An event study approach," *Economic Journal* 118, 1-28.

- Frenkel, M., C. Pierdzioch, G. Stadtmann (2004), “The Accuracy of Press Reports Regarding the Foreign Exchange Interventions of the Bank of Japan,” *Journal of International Financial Markets, Institutions and Money* 14 (1), 25-36.
- Frenkel, M., C. Pierdzioch, G. Stadtmann (2005), “The effects of Japanese foreign exchange market interventions on the yen/U.S. dollar exchange rate volatility,” *International Review of Economics and Finance* 14, 27-39.
- Friederich, S. and R. Payne (2015). “Order-to-trade ratios and market liquidity.” *Journal of Banking and Finance* 50 (0), 214-223.
- Gencay, R., F. Selcuk (2004), “Volatility-return dynamics across different timescales.” *Econometric Society World Congress 2005*, Congress paper.
- Gencay, R., F. Selcuk, and B. Whitcher (2005), “Multiscale systematic risk.” *Journal of International Money and Finance* 24, 55-70.
- Giacomini, R. and B. Rossi (2010), “Forecast Comparisons in Unstable Environments,” *Journal of Applied Econometrics* 25(4), 595-620.
- Gray, R. M. (2011), *Entropy and Information Theory*, Springer.
- Hamilton J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*. 57, 357-384
- Hagstromer, B., and L. Norden (2013), “The diversity of high-frequency traders,” *Journal of Financial Markets* 16, 741-770.
- Hasbrouck, J. (2007) *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press, New York.
- Hasbrouck, J. and G. Saar (2013), “Low-latency trading, *Journal of Financial Markets* 16, 646-679.
- Hendershott, J., and A. Menkveld (2011), “Does algorithmic trading improve liquidity? *Journal of Finance* 66.
- Henkel, S.J., Martin, J.S., Nardari, F., 2011. Time-varying short-horizon predictability. *Journal of Financial Economics*. 99:560-580.

- Hung, J. (1997), "Intervention strategies and exchange rate volatility: A noise trading perspective," *Journal of International Money and Finance* 16, 779-93.
- In, F. and Kim, S. (2006), "Multiscale hedge ratio between the Australian stock and futures markets: Evidence from wavelet analysis." *Journal of International Financial Management* 16, 411-423.
- In, F. and Kim, S. (2007), "A note on the relationship between Fama-French risk factors and innovations of ICAPM state variables." *Finance Research Letters* 4, 165-171.
- Inoue, A. and B. Rossi (2012), "Out-of-Sample Forecast Tests Robust to the Choice of Window Size," *Journal of Business and Economic Statistics* 30(3), 432-453.
- Ito, Takatoshi (2002), "Is Foreign Exchange Intervention Effective? The Japanese Experience in the 1990," NBER Working Paper 8914, *National Bureau of Economic Research*.
- Ito, Takatoshi (2007), "Myths and reality of foreign exchange interventions: An application to Japan." *International Journal of Finance and Economics* 12(2), 133-154.
- Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., (2009), "Microstructure noise in the continuous case: the pre-averaging approach, *Stochastic Processes and their Applications* 119, 2249-2276.
- Kaiser, A. and T. Schreiber (2002), "Information Transfer in Continuous Processes," *Physica D* 166, 43-62.
- Killeen, W., R. Lyons, and M. Moore, (2006), Fixed versus Flexible: lessons from EMS order Flow. *Journal of International Money and Finance*, 25(4), 551-579.
- Kim, S.-J., T. Kortian, and J. Sheen (2000), "Central Bank Intervention and Exchange Rate Volatility Australian Evidence," *Journal of International Financial Markets Institutions and Money* 10(3-4), 381-405.
- Kim, S.-J. (2007), "Intraday evidence of efficacy of 1991-2004 yen interventions by the Bank of Japan," *Journal of International Financial Markets, Institutions and Money* 17, 341-360.
- Kirilenko, A., Kyle, A. S., Samadi, M. and Tuzun, T. (2017), "The Flash Crash: High-Frequency Trading in an Electronic Market, *Journal of Finance* 72: 967C998.

- Kozachenko, L., and N. Leonenko (1987), "Sample Estimate of the Entropy of a Random Vector," *Problems of Information Transmission* 23(2), 95-101.
- Kozhan, R., M. Moore and R. Payne (2014), "Market Order Flows, Limit Order Flows and Exchange Rate Dynamics," Available at *SSRN*.
- Kraskov, H. Stogbauer, and P. Grassberger (2004), "Estimating Mutual Information," *Physical Review E* 69, 066138.
- Lee, S.S., Hannig, J. (2010), "Detecting jumps from Levy jump diffusion processes, *Journal of Financial Economics* 96, 271-290.
- Lee, S. S. and Mykland, P. A. (2008), "Jumps in Financial Markets: a New Nonparametric Test and Jump Dynamics," *Review of Financial studies* 21, 2535-2563.
- Love, R. and R. Payne (2008), "Macroeconomic News, Order Flows and Exchange Rates," *Journal of Financial and Quantitative Analysis* 43, 467-488.
- Mark, N.C. (1995), "Exchange rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review* 85, 201-218.
- Malkiel, B. G. (2003), "The Efficient Market Hypothesis and Its Critics, *Journal of Economic Perspectives* 17(1): 59-82.
- Meese, R.A. and K.S. Rogoff (1983a), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?," *Journal of International Economics* 14(1-2), 3-24.
- Meese, R.A. and K.S. Rogoff (1983b), "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Mis-specification?" in: *Exchange Rates and International Macroeconomics*, Jacob Frenkel, eds., Chicago: NBER and University of Chicago Press.
- Meese, R.A. and K.S. Rogoff (1988), "Was it Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating Rate Period," *Journal of Finance* 43, 923-948.
- Menkhoff, L. (2010), "High-Frequency Analysis of Foreign Exchange Interventions: What Do We Learn?" *Journal of Economic Surveys* 24(1), 85-112.
- Menkveld, A. J. and M. Zoican (2017) "Need for Speed? Exchange Latency and Liquidity, *Review of Financial Studies* 30 (2017), 1188-1228.

- Molodtsova, T. and D.H. Papell (2009), "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals," *Journal of International Economics*, 77(2), 167-180.
- Nanto, Dick K. (2007), "Japans Currency Intervention: Policy Issues." *CRS Report for Congress*, Washington, D.C. Congressional Research Service.
- Neely, C. J. (2005), "An Analysis of Recent Studies of the Effect of Foreign Exchange Intervention," *Federal Reserve Bank of St. Louis Review* 87(6), 685-717.
- Newman, V., C. Potter, and M. Wright (2011), "Foreign exchange market intervention," *Reserve Bank of Australia Bulletin* December, 67-76.
- O'Hara, M. (1995), *Market Microstructure Theories*. Blackwell Publishers, Cambridge, MA.
- Pasquariello, P. (2007), "Informative trading or just costly noise? An analysis of central bank interventions." *Journal of Financial Markets* 10, 107-143.
- Payne, Richard and P. Vitale (2003), "A Transaction Level Study of the Effects of Central Bank Intervention on Exchange Rates," *Journal of International Economics* 61(2), 331-52.
- Perez-Quiros, G., Timmermann, A. 2000. Firm Size and Cyclical Variations in Stock Returns. *Journal of Finance*, 55, 1229-1262.
- Podolskij, M., Vetter, M. (2009), "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps, *Bernoulli* 15, 634-658.
- Ramsey, J. and C. Lampart (1998), "The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income," *Studies in Nonlinear Dynamics & Econometrics* 3, 1-22.
- Ramsey, J. (2002), "Wavelets in Economics and Finance: Past and Future," *Studies in Nonlinear Dynamics & Econometrics* 6, 3, 1-29.
- Rime, D., L. Sarno, E. Sojli (2010), "Exchange rate forecasting, order flow and macro economic information," *Journal of International Economics* 80, 72-88.
- Rogoff, K. (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature* 34, 647-668.

- Rossi, B. (2013), “Exchange rate predictability,” *Journal of Economic Literature* 51 (4), 1063-1119.
- Sarno, L. and M. P. Taylor (2001), “Official Intervention in the Foreign Exchange Market: Is It Effective and, If So, How Does It Work?” *Journal of Economic Literature* 39(3), 839-68.
- Schreiber, T (2000), “Measuring Information Transfer,” *Physical Review Letters* 85(2), 461-464.
- Shannon, C.E. (1948), “A Mathematical Theory of Communication,” *Bell System Technical Journal* 27, 379-423, 623-656.
- Shannon, C.E. and W. Weaver (1949), *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.
- Sims, C., Zha, T., 2006. Were there regime switches in US monetary policy? *American Economic Review*. 96, 54-81.
- Taylor, M. (2005), “Official foreign exchange intervention as a coordinating signal to the dollar-yen market,” *Pacific Economic Review* 10, 73-82.
- Taylor, M.P. and L. Sarno (1998), “The Behavior of Real Exchange Rates During the Post-Bretton Woods Period,” *Journal of International Economics* 46(2), 281-312.
- U.S. Securities and Exchange Commission. (2013) “Equity market structure literature review part I: Market fragmentation.” *Securities and Exchange Commission Working Paper*.
- Vacha, L. and Barunik, J. (2012), “Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis.” *Energy Economics* 34(1), 241-247
- Vicente, R. and M. Wibral, (2014), “Efficient Estimation of Information Transfer,” *Directed Information Measures in Neuroscience*, edited by M. Wibral, R. Vicente, and J. T. Lizier, Springer, Berlin/Heidelberg.
- Vicente, R., M. Wibral, M. Lindner, and G. Pipa (2011), “Transfer entropy-a model-free measure of effective connectivity for the neurosciences,” *Journal of Computational Neuroscience* 30, 45.
- Wibral, M., R. Vicente, and J. T. Lizier, eds. (2014), “Directed Information Measures in Neuroscience,” Springer, Berlin/ Heidelberg.

Wright, J. (2008), “Bayesian Model Averaging and Exchange Rate Forecasting,” *Journal of Econometrics* 146, 329-341.

Yao, C. and M. Ye (2014), “Tick Size Constraints, Market Structure, and Liquidity, Available at SSRN: <https://ssrn.com/abstract=2359000> or <http://dx.doi.org/10.2139/ssrn.2359000>