

Point Process Based High Frequency Volatility Estimation: Theory and Applications



Yifan Li

Supervisors: Prof. Ingmar Nolte
Dr. Sandra Nolte

Department of Accounting and Finance
Lancaster University

Thesis submitted in partial fulfilment of the requirements for the
degree of
Doctor of Philosophy in Finance

September 2018

To my parents and beloved wife.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Yifan Li
September 2018

Acknowledgements

My PhD has been a challenging yet rewarding experience. I would like to express my gratitude and appreciation to many people who helped me directly and/or indirectly in developing this study and my academic career. It would be impossible for me to complete this work without the invaluable support and encouragement from you.

Firstly, I would like to thank my supervisors Prof. Ingmar Nolte and Dr. Sandra Nolte for introducing me to the field of financial econometrics. I feel extremely lucky to have you as both my supervisors and friends. You provide tremendous support to my research and career development in every aspect with your erudite knowledge and heart-warming encouragement. You set excellent examples for me to become not only a responsible and rigorous researcher, but also a helpful and enthusiastic scholar. I am very grateful for your kindness and patience to my capricious research and career choices, and I could have not achieved anything in the past four years without your guidance.

I thank Prof. Oliver Linton and Prof. Stephen Taylor for reviewing this lengthy thesis and providing insightful comments and corrections that greatly improved the thesis.

I would like to thank Lancaster University Management School (LUMS) for providing me with generous financial support to fund my PhD. I would also like to thank all the academic and administrative staffs in the Department of Accounting and Finance of LUMS to organize the PhD programme. With the support of the department I was able to participate in various seminars, conferences and exchange programmes, which are extremely important experiences for building my research and connections. I am grateful for all researchers that taught me various courses and offered me precious suggestions to my research and career plan.

I thank all my external co-authors, namely Simon Bodilsen, Dr. Rong Ding, Prof. Peter Reinhard Hansen, Ekaterina Kazak, Prof. Asger Lunde, Prof. Andrew Patton, Ye Zeng and Dr. Hang Zhou. I am very fortunate to know and collaborate with the above researchers. Working with them has significantly broadened my limited

research scope. Our collaborations also greatly strengthen my research portfolio and competitiveness in the job market, which is of crucial importance for me to secure a Lecturer in Finance position at Alliance Manchester Business School (AMBS).

Many friends and colleagues have accompanied me during my PhD journey in Lancaster. There are too many names to mention here, but the past four years certainly have been a very joyful and memorable experience for me because of you all. Specially, I would like to thank Wenwen Ke and Yiming Xiao for their unconditional understanding and support half the world away. Thanks for sharing our lives and standing by my side after all these years despite our geographical distances, and I am really looking forward to our reunion in the future.

I would like to thank my parents and parents-in-law for encouraging me to chase my dreams and providing invaluable support to me both financially and emotionally. During the past five years, my time at home was very limited, and I knew by heart that you missed me every single day when I was away. As much as I would like to spend more time with you, my curiosity and desire to explore the unknowns and expand human knowledge spur me into a lifelong journey that unfortunately drives us further apart. I am forever indebted to your everlasting love, and I hope that you will be proud of my growth and achievements.

Finally, I would like to express my deepest gratitude and love to my wife, Diansha Wang. It was an absolute miracle that we met each other in Lancaster, and I feel extremely blessed to have you in my life. With your tender love and sweet ways, you saved me from my childish self-pity and made me a better man. We had numerous exciting and enjoyable times together during the past four years, but I am sure that much more are yet to be discovered in the future. I am not very good at expressing my feelings towards you, but the following quotes from the Spider-man movie that we watched together summarize it pretty accurately: “When I look in your eyes, and you’re looking back in mine, everything feels not quite normal. I feel stronger and weaker at the same time. I feel excited, and at the same time, terrified. The truth is I don’t know what I feel, except I know what kind of man I want to be.”

Abstract

This thesis is a compilation of three main studies with the common theme: point process based high-frequency volatility estimation. The first chapter introduces a new class of high-frequency volatility estimators and examines its asymptotic properties. The second chapter studies the relative importance of market microstructure (MMS) variables on high-frequency volatility estimation. The third chapter proposes a Markov-switching model for high-frequency volatility estimation and provides intraday measures of information contents in the trading process using the proposed model.

In the first chapter, we propose a novel class of volatility estimators named the Renewal Based Volatility (*RBV*) estimator, and derive its asymptotic properties. This class of estimators is motivated by the work of Engle and Russell (1998), Gerhard and Hautsch (2002), Andersen, Dobrev, and Schaumburg (2008), Tse and Yang (2012), Nolte, Taylor, and Zhao (2018), which use price durations to construct high-frequency volatility estimators. We show that our *RBV* estimator nests the volatility estimator using price duration, thus providing a theoretical framework to analyse its asymptotic properties. Our theoretical results support the simulation and empirical findings in Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018) that: (1) both the non-parametric duration (*NPD*) based and the parametric duration (*PD*) based volatility estimators are more efficient than the Realized Volatility (*RV*) estimator; (2) a parametric design can greatly improve the efficiency of volatility estimation; (3) the *PD* estimator can provide accurate intraday volatility estimates. We provide simulation evidence for the performance of the *NPD* estimator and propose an exponentially smoothed version that can outperform noise-robust *RV*-type estimators under general market microstructure noise and jumps.

In the second chapter, we augment the *PD* estimator by including MMS variables in the parametric model. Specifically, we use a lognormal version of the Autoregressive Conditional Duration (*ACD*) model by Engle and Russell (1998), and include trading volume, bid-ask spread, total quote depth, quote depth difference, number of trades, order imbalance and order flow in the *ACD* model. Moreover, we use a best subset regression (*BSR*) approach to rank and select the included MMS

variables. Our empirical study based on high-frequency trade and quote data from 29 highly liquid securities and a market index ETF shows that, by benchmarking on a Realized Kernel measure, the inclusion of MMS covariates significantly improves the performance of volatility estimates on both daily and intraday levels. The BSR approach is very effective in selecting the most relevant MMS covariates for volatility estimation, and it suggests that contemporaneous number of trades and order flow are the most important variables for intraday volatility estimation. More importantly, intraday volatility estimates can be constructed from the ACD model even in the case when the RV-type estimators cannot be reliably constructed due to a lack of data.

In the third chapter, we extend the Autoregressive Conditional Intensity (ACI) model (Russell, 1999) with a Markov-switching (MS) structure. We propose to use the Stochastic Approximation Expectation Maximization (SAEM) (Celeux and Diebolt, 1992) to estimate the MS-ACI model, and provide simulation evidence supporting the validity of the estimation procedure. We apply our MS-ACI model to high-frequency trade and quote data from 9 highly liquid securities and a market index ETF. Our empirical findings suggest that the MS-ACI model captures two distinct volume-volatility regimes in the high-frequency data: a dominant regime that spreads evenly throughout the trading day with strong correlation between cumulative trading volume and price duration, and a minor regime that concentrates at the beginning and end of a trading day with much weaker correlation between cumulative trading volume and price duration. We link this phenomenon to the firm-specific information arrival process into the market, and provide a measure of intraday information content of the transaction process.

Table of Contents

List of Figures	xiv
List of Tables	xviii
Introduction	1
1 Asymptotic Theory for Renewal Based High-Frequency Volatility Estimation	7
1.1 Introduction	7
1.2 Prerequisites: Renewal Theory	10
1.3 Renewal Based Volatility Estimator	12
1.4 Parametric Renewal Based Volatility Estimator	15
1.5 Some Examples	19
1.6 The <i>NPD</i> Volatility Estimator Under Market Frictions	25
1.6.1 Drift Effect	25
1.6.2 Jump Effect	25
1.6.3 A More Realistic Model	27
1.6.4 Bias of the <i>NPD</i> estimator	29
1.6.5 Price Discretization	30
1.6.6 A Possible Bias Correction Method for the <i>NPD</i> Estimator	31
1.7 Simulation Study	32
1.7.1 Simulation Design	32
1.7.2 1FSV Model Without Noise and Price Discretization	37
1.7.3 Full 1FSV Model: Primal Volatility Estimators	39
1.7.4 Full 1FSV Model: Bias Corrected Estimators	43
1.8 Concluding Remarks	48
2 High-Frequency Volatility Estimation and the Relative Importance of Market Microstructure Variables	51
2.1 Introduction	51
2.2 Literature Review	53
2.3 Price Duration and Volatility Estimation	55

2.4	Price Duration Modelling	60
2.4.1	Inclusion and Selection of Variables	61
2.5	Data and Descriptive Statistics	65
2.6	Empirical Results	68
2.6.1	Analysis of the Relative Importance of MMS Covariates	68
2.6.2	Estimation of the LL-ACD model	71
2.6.3	Intraday Volatility Estimation with the LL-ACD Model	80
2.7	Concluding Remarks	88
3	High-Frequency Volatility Modelling: A Markov-Switching Autoregressive Conditional Intensity Model	91
3.1	Introduction	91
3.2	Conditional Intensity Modelling	94
3.2.1	Basic Point Process Theory	94
3.2.2	The Original ACI Model	96
3.2.3	Stationarity of the ACI Model	97
3.3	Markov-Switching ACI Model	99
3.3.1	Specification of the MS-ACI Model	99
3.3.2	Stationarity Condition of the MS-ACI Model	100
3.3.3	Model Estimation	101
3.3.4	Post-Estimation Diagnostics	109
3.4	Monte Carlo Simulation Study	111
3.5	Application to High-Frequency Stock Price Duration Data	116
3.5.1	Data	117
3.5.2	Main Empirical Results	122
3.6	Concluding Remarks	127
4	Conclusion	129
4.1	Summary of Main Findings	129
4.2	Implications	131
4.3	Limitations and Future Researches	132
4.4	Final Remarks	133
	Bibliography	135
	Appendix A Appendix to Chapter 1	147
A.1	The Time Changed Compounded Poisson Process	147
A.2	Proof to Theorem 1.5	147
A.3	Asymptotic Properties of the <i>RBV</i> Estimator under Infill Asymptotics	148
A.3.1	Relationship to the RV Estimator	150

A.3.2	End-of-Sample Bias	152
A.4	Proof of Corollary 1.3	152
A.5	Proof of Corollary 1.1	153
A.6	Simulation of $\rho^{(\delta)}$ and $\rho^{(r)}$ for the <i>PD</i> and <i>PR</i> estimators	154
A.7	An Approximated Time Discretization Bias	155
A.8	Determinants of the Bias of the <i>NPD</i> Model	156
A.9	Moment conditions for r_j of the GST model	162
A.10	Implementation details for RK, <i>NPD^z</i> , PRV and PBip estimators	162
A.11	Additional Tables and Figures	165
Appendix B Appendix for Chapter 2		171
B.1	Analysis of Co-movements Between RK and ICV Estimates	171
B.2	Additional Tables and Figures	174
B.3	Best Subset Selection Using Mixed Integer Optimization	188
B.4	Robustness Checks of the LL-ACD(1,1) Model Estimation	188
B.4.1	Gains in Information Criteria using BSR Selection	189
B.4.2	Residual Diagnostics	190
B.4.3	Goodness-of-Fit	193
B.5	Daily Volatility Estimation with the LL-ACD Model	196
B.6	Construction of the Intraday ICV Estimator	201
Appendix C Appendix for Chapter 3		203
C.1	Proof to Proposition 3.1	203
C.2	Approximated Single Move Sampler	204
C.3	Estimating the Variance-Covariance Matrix of Estimated Parameters and the Most Probable State Vector	205
C.4	Deseasonalization of the Price Duration and Volume	207
C.5	Robustness Checks of the MS(2)-ACI(2,1)-V Model	211
C.6	A Market Microstructure Model for Informed Trading and the Volume- Volatility Relationship	215
C.7	Tables and Figures	219

List of Figures

1.1	Discrepancy between the density of $R_i^{(\cdot)}$ and $\tilde{D}_i^{(\cdot)}$	24
1.2	Simulated diurnal pattern of intraday volatility and transaction arrival rate	36
1.3	Histogram and correlogram for the simulated price change with moderate level of MMS noise and no jumps	36
1.4	An example of simulated price path of the IFSV _J model with moderate level of noise	37
1.5	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD , $RV^{(\delta)}$, RV and $RBip$ for 1FSV model without noise and price discretization	38
1.6	Average sampling frequency of the NPD estimator for the 1FSV and 1FSV _J models	40
1.7	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD , $RV^{(\delta)}$, RV and $RBip$ for 1FSV model with moderate level of MMS noise	41
1.8	Average correlogram of calendar time returns and price duration returns	42
1.9	Average sampling frequency of the NPD and NPD^z estimator for the 1FSV and 1FSV _J models under optimal γ	44
1.10	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z , RK , PRV and $PBip$ for 1FSV model with moderate level of MMS noise	45
2.1	Daily price change threshold δ for all thirty stocks from 2011-Jan to 2014-Dec	66
2.2	An example of the price duration process	67
2.3	Correlogram and histogram of log price durations for SPY	67
2.4	Estimated diurnal patterns from models in Table 2.5	73
2.5	Summary of estimated contemporaneous MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 1	75
2.6	Summary of estimated contemporaneous MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 2	76

2.7	Summary of estimated contemporaneous MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 3	77
2.8	Example of intraday volatility estimates based on ICV and RK for SPY	82
2.9	Tick-by-tick price change for SPY on 09-Aug-2011	83
2.10	15-minute volatility estimates for AA and SPY on 29-Mar-2013	86
3.1	Convergence Diagram for Spec. 1 in Table 3.1	115
3.2	Scatter plots with regression line for AIG 2016-03, INTC 2016-08, SPY 2016-05	120
3.3	R^2 Difference for the volume-duration regressions in the first hour of the day and the rest of the day	121
3.4	Distribution of estimated regimes over time for AIG 2016-03, INTC 2016-08 and SPY 2016-05	124
3.5	Scatter plots with regression line for AIG 2016-03, INTC 2016-08, SPY 2016-05 based on estimated regimes	125
3.6	R^2 difference for the volume-duration regressions between observations in regime 1 and 2 using the estimated state vector	126
A.1	Simulated volatility signature plot for the <i>NPD</i> estimator on the GST model with no MMS noise	158
A.2	Simulated volatility signature plot for the <i>NPD</i> estimator on the GST model with i.i.d. MMS noise	159
A.3	Simulated volatility signature plot for the <i>NPD</i> estimator on the GST model with AR(1) MMS noise	160
A.4	Simulated volatility signature plot for the <i>NPD</i> estimator on the GST model with AR(1) MMS noise and price discretization	161
A.5	Simulated volatility signature plot for the <i>NPD</i> estimator on the GST model with price discretization and jumps	162
A.6	Optimal θ s of PRV and PBip estimators	164
A.7	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from <i>NPD</i> , $RV^{(\delta)}$, RV and RBip for 1FSV model with high level of MMS noise	166
A.8	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from <i>NPD</i> , $RV^{(\delta)}$, RV and RBip for 1FSV model with low level of MMS noise	167
A.9	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from <i>NPD</i> ^z , RK, PRV and PBip for 1FSV model with high level of MMS noise	168

A.10	Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z , RK, PRV and PBip for 1FSV model with low level of MMS noise	169
B.1	Correlogram and histogram of log price durations for IBM, JPM and PFE	182
B.2	Summary of estimated dynamic parameters of the LL-ACD(1,1)-K model for all stock-months, part 1	183
B.3	Summary of estimated dynamic parameters of the LL-ACD(1,1)-K model for all stock-months, part 2	184
B.4	Summary of estimated one-duration lagged MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 1	185
B.5	Summary of estimated one-duration lagged MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 2	186
B.6	Summary of estimated one-duration lagged MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 3	187
B.7	Comparison between d^L and d^O for all stock-months	189
B.8	Correlograms and quantile-quantile plots of the estimated residuals from model estimations in Table 2.5	191
B.9	Example of daily volatility estimates based on ICV and RK for SPY .	197
B.10	Illustration of the boundary problem of the intraday ICV estimator .	202
C.1	Examples of Deseasonalization: Raw and Deseasonalized Price Duration and Volume from AIG 2016-03, INTC 2016-08, SPY 2016-05 . . .	209
C.2	Lomb-Scargle Periodogram for Raw and Deseasonalized Price Durations, Volume and Bid-Ask Spread Covariates for AIG 2016-03, INTC 2016-08, and SPY 2016-05	210
C.3	Monthly choices of δ for 10-by-12 stock-month datasets	225
C.4	Yearly distribution of estimated regimes over time for all stock-months for the MS(2)-ACI(2,1)-V model	225
C.5	Difference in the estimated \hat{b}_1 s from the volume-duration regressions between observations in regime 1 and 2 using the estimated state vector	226
C.6	Estimated SoR for MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models . .	226
C.7	Yearly distribution of estimated regimes over time for all stock-months for the MS(2)-ACI(2,1) model	227
C.8	Quantile-Quantile Plots of the residuals obtained from ACI(2,1), ACI(2,1)-V, MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models for AIG 2016-03, INTC 2016-08 and SPY 2016-05	228

C.9 Correlograms of the residuals obtained from ACI(2,1), ACI(2,1)-V, MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models for AIG 2016-03, INTC 2016-08 and SPY 2016-05	229
C.10 Bayesian Information Criterion of ACI(2,1), ACI(2,1)-V, MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models for all stock-months	230
C.11 Distribution of estimated regimes over time from the MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-08 and SPY 2016-05	230

List of Tables

- 1.1 List of all volatility estimators considered in the simulation study . . . 33
- 1.2 Comparison of the optimal MSEs for all volatility estimators in Table 1.1 for the 1FSV and 1FSV_J models with low, moderate and high levels of noise 47
- 2.1 Description of MMS variables 62
- 2.2 Cross-correlation table between log price durations and MMS covariates 68
- 2.3 Best-subset regression outputs for SPY, 2011-01 69
- 2.4 Average of monthly relative importance of MMS covariates 70
- 2.5 Examples of LL-ACD(1,1) estimation outputs for 2011-01, SPY 72
- 2.6 Out-of-sample performance of the LL-ACD(1,1) models 80
- 2.7 Averaged correlation table of RK and monthly estimated intraday ICV volatility estimates for SPY 84
- 2.8 Comparison of MSEs of intraday ICV volatility estimates 84
- 3.1 Monte Carlo Simulation Results of Parameter Estimates of MS(2)-ACI(1,1) Models and the Corresponding Complete Models for 100 Random Draws of Data 113
- 3.2 Monte Carlo Simulation Results of Parameter Estimates of a MS(3)-ACI(1,1) Model and its Complete Version for 100 Random Draws of Data 115
- 3.3 Yearly Descriptive Statistics for $\hat{x}_{i,d}^\delta$ and $\ln \hat{Vol}_{i,d}^\delta$ 119
- 3.4 Parameter Estimates of MS(2)-ACI(2,1)-V Model for AIG 2016-03, INTC 2016-08 and SPY 2016-05 123
- A.1 Simulated moments for $\tilde{D}^{(\cdot)}$ and $R_i^{(\cdot)}$ and the simulated $\rho^{(\cdot)}$ 155
- A.2 Comparison of the optimal QLIKEs for all volatility estimators in Table 1.1 for the 1FSV and 1FSV_J models with low, moderate and high levels of noise 165
- B.1 Fixed-effect regressions for the comovement analysis between RK and ICV estimates 173

B.2	Daily descriptive statistics of the transaction data	174
B.3	Descriptive statistics of the dataset	175
B.4	Descriptive statistics of the dataset	176
B.5	Average of quarterly rankings of relative importance of MMS covariates	177
B.6	Average of half-yearly rankings of relative importance of MMS covariates	178
B.7	Average of yearly rankings of relative importance of MMS covariates .	179
B.8	Average MMS parameter estimates for the LL-ACD(1,1)-K model . .	180
B.9	Average MMS parameter estimates for the LL-ACD(1,1)-A model . .	181
B.10	Summary of diagnostic test results	192
B.11	Average adjusted R-squared for models with varying estimation windows	194
B.12	Summary of likelihood ratio test results	195
B.13	Averaged correlation table of RK and monthly estimated daily ICV volatility estimates	198
B.14	Comparison of MSEs of ICV volatility estimates within each estimation window	199
B.15	Comparison of MSEs of ICV volatility estimates across different esti- mation windows	200
C.1	Yearly Descriptive Statistics for $x_{i,d}^\delta$ and $\ln Vol_{i,d}^\delta$	219
C.2	Comprehensive estimation outputs for AIG 2016-03	220
C.3	Comprehensive estimation outputs for INTC 2016-08	221
C.4	Comprehensive estimation outputs for SPY 2016-05	222
C.5	Summary of Cramér-von-Mises, Andersen-Darling and Ljung-Box test results	223
C.6	Estimation Outputs of MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-8 and SPY 2016-05	224
C.7	Comparison of regime classification of the MS(2)-ACI(2,1)-V and the MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-08 and SPY 2016-05	224

Introduction

Volatility estimation is an important topic in the field of finance and financial econometrics, as it is a crucial input for asset pricing, portfolio allocation, risk management, etc. The availability of high-frequency data recently has led to a shift from volatility modelling at low frequency (e.g. the GARCH model by Engle (1982) and Bollerslev (1986) and its extensions) to high-frequency volatility measures. Since the seminal work of Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001), the Realized Volatility (RV)-type measures are popularized and became one of the most widely applied high-frequency volatility measures.

The popularization of the RV estimator is not surprising, as it possesses many desired properties of a volatility measure. Firstly, it is very easy to construct, as it only involves summing up squared intraday returns sampled at equidistant intervals. Secondly, assuming the log-price process to be a continuous semimartingale, the RV measure converges to the integrated variance of the price process with well-established asymptotic properties (see e.g. Barndorff-Nielsen and Shephard (2002)). Thirdly, a large number of extensions to the RV estimator are developed to address the issue that the RV measure is biased in the presence of observation errors and price discontinuities (jumps). For example, in order to correct for the bias introduced by market microstructure (MMS) noise, Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006) advocate the use of subsampling and more than one sampling frequency, Hansen and Lunde (2006) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a) popularize the Realized Kernel estimator, and Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Hautsch and Podolskij (2013) propose a pre-averaging approach. Examples of jump-robust RV-type estimators can be found in Barndorff-Nielsen and Shephard (2003) and Andersen, Dobrev, and Schaumburg (2012).

Nevertheless, despite the aforementioned advantages of the RV-type estimator, it suffers from the problems that it relies heavily on the amount of data available within the estimation window, and that only price information is incorporated in the

construction of RV measures. These two problems arise due to the non-parametric nature of the RV-type estimators, and confine the use of the RV-type estimators in situations where the amount of data is limited, i.e. less frequently traded stock or intraday volatility estimation, or incorporating other observable information in high-frequency volatility estimation.

An alternative high-frequency volatility estimator that has the potential to overcome the above problems is the point process based volatility estimator initially proposed by Engle and Russell (1998), and developed by Gerhard and Hautsch (2002), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018). As Tse and Yang (2012) summarize, this estimator enjoys a full parametric design, thus one can use data beyond the window of volatility estimation to improve the quality of parameter estimates which in turn leads to more precise volatility estimation. The parametric structure also facilitates the inclusion of other MMS covariates, which can not only further improve the quality of volatility estimation, but also provide a framework to analyse the relationship between volatility and other MMS covariates on a high-frequency level. These two properties are the key advantages of this estimator over the RV approach, and evidence also supports the superior performance of this estimator over the RV approach. Via simulation, Tse and Yang (2012) show that the point process based volatility estimator is more efficient than the RV-type estimators, and Nolte, Taylor, and Zhao (2018) find that it has better forecasting performance than the RV-type estimators. However, this estimator did not receive equal attention as the RV-type estimators, partly due to the fact that its theoretical properties are largely unknown.

This thesis is motivated by the potential of the point process based approach and attempts to popularize this approach for wider applications. The thesis consists of three individual chapters which contribute to the existing literature from both theoretical and empirical perspectives. More importantly, the thesis aims to advocate the use of the point process based approach by demonstrating that this approach is superior to the RV approach in both quality of volatility estimation and the provision of a powerful framework for MMS studies involving high-frequency volatility.

In the first chapter, we generalize the original estimator in Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018) and propose the Renewal Based Volatility (*RBV*) class of estimators. We also establish a theoretical framework to derive the asymptotic properties of the *RBV* estimators. The *RBV* estimators are constructed based on the idea that in the original estimator of Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018), the price change point process constructed in calendar time becomes a

renewal process in business time under some mild assumptions. Asymptotic results can be derived based on the renewal process in business time. Our theoretical findings suggest that, firstly, the non-parametric point process based volatility estimator is more efficient than existing RV-type estimators, and the *RBV* approach has the potential to outperform RV-type estimators under any sampling scheme. Secondly, we corroborate the advantages of using a parametric structure over the non-parametric estimators such as the RV. Specifically, by augmenting the *RBV* estimators with a parametric structure, the efficiency of volatility estimates can be further improved without increasing the resolution of price process. This chapter provides a theoretical foundation for the point process based volatility estimators in Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018), which will also be implemented in Chapters 2 and 3 of this thesis.

In addition to the theoretical considerations discussed above, Chapter 1 also provides an in-depth analysis of the Non-Parametric Duration (*NPD*) based volatility estimator by Nolte, Taylor, and Zhao (2018) in the presence of market frictions and jumps. We quantify the effect of jumps, time discretization, MMS noise and price discretization on the *NPD* estimator by a comprehensive simulation study. In our simulation results, we demonstrate that the *NPD* estimator is indeed more efficient than calendar time RV estimators sampled equally frequently when the sampling frequency is low, and is very robust to jumps. However, as the sampling frequency increases, the *NPD* estimator is more sensitive to MMS noise and thus more biased compared to the calendar time RV estimators. To overcome this problem, we propose an exponentially smoothed *NPD* estimator and show that it can mitigate the MMS noise bias and has the potential to outperform common noise robust RV-type methods such as the Realized Kernel and the pre-averaged RV.

In Chapter 2, we focus on exploiting information in the MMS covariates to improve the quality of high-frequency volatility estimation with the point process based approach, which is a feature that is not supported by the traditional RV approach. In detail, we include trading volume, bid-ask spread, total quote depth, quote depth difference, number of trades, order imbalance and order flow in a log specification of the Autoregressive Conditional Duration (ACD) model (Engle and Russell, 1998) to construct daily and intraday volatility estimates. We implement a Best Subset Regression (BSR) approach to assess the relative importance of each variable and select the optimal number of variables.

Using trade and quote data from 29 highly liquid stocks and a stock index ETF, our

empirical findings suggest that, firstly, the inclusion of MMS covariates can to a great extent improve the goodness-of-fit of the parametric model, and the BSR approach can effectively select the most relevant variables to include in the ACD model. This result is further reflected in the quality of daily and intraday volatility estimates. Using the Realized Kernel measure as a benchmark, we show that the inclusion of optimally selected MMS covariates significantly reduces the distance between the volatility estimates from the ACD models and the benchmark on both daily and intraday levels, and including all MMS covariates does not further improve the result. More importantly, we demonstrate that with the inclusion of MMS covariates, we can construct reliable intraday volatility estimates, even when the amount of data is considered inadequate for a reliable RV-type estimator to be constructed. Therefore, our results in the second chapter provide empirical evidence supporting the theoretical findings in Chapter 1.

In Chapter 3, we develop the Markov-Switching Autoregressive Conditional Intensity (MS-ACI) model by extending the ACI model originally proposed by Russell (1999) with a Markov-switching structure. We augment the stationarity condition of the ACI and MS-ACI model, and show that the MS-ACI model can be reliably estimated by the Stochastic Approximation Expectation Maximization (SAEM) algorithm (Celeux, Chauveau, and Diebolt, 1996; Celeux and Diebolt, 1992). Our results contribute to the literature by providing a frequentists' method to solve the path dependency problem in estimating the Markov-switching autoregressive models without any simplification to the parametric structure of the model.

By combining the MS-ACI model with the point process based volatility estimator, we are able to investigate the intraday regime-switching relationship between volatility and trading volume empirically. This application of the MS-ACI model is motivated by the related MMS literature that the trading volume submitted by traders who possess information will have a much higher impact on price volatility than those traders who trade only for liquidity purposes (see e.g. Copeland and Galai (1983), Glosten, Jagannathan, and Runkle (1993), Andersen (1996), Easley, Kiefer, O'Hara, and Paperman (1996) among others). Our empirical findings based on 9 highly liquid stocks and a market index ETF are consistent with the literature as we detect two distinct regimes in the intraday volume-volatility relationship by implementing the MS-ACI model. Specifically, we detect a minor regime that concentrates at around the beginning and end of trading days which corresponds to the regime of high information content, and a major regime that spreads evenly across trading days, which is considered to have lower information content. We cannot observe this effect

on the market index, and this suggests that what we capture may be associated with firm-specific information arrivals into the market. Based on our empirical investigation, we propose to use the posterior probability of regime classification as a measure of intraday informativeness of the market.

The rest of this thesis is structured as follows: Chapter 1: Asymptotic Theory for Renewal Based High-Frequency Volatility Estimation. Chapter 2: High-Frequency Volatility Estimation and the Relative Importance of Market Microstructure Variables. Chapter 3: High-Frequency Volatility Modelling: A Markov-Switching Autoregressive Conditional Intensity Model. Chapter 4 concludes and discusses the limitations and future researches from the thesis.

Chapter 1

Asymptotic Theory for Renewal Based High-Frequency Volatility Estimation

1.1 Introduction

Since the seminal paper by Engle and Russell (1998), a point process based high-frequency volatility estimator provides an important alternative to the Realized Volatility (RV)-type estimator as popularized by Andersen, Bollerslev, Diebold, and Labys (2001). The main argument supporting the point process based volatility estimator is its parametric structure and ability to provide intraday inference on local volatility, as opposed to an integrated volatility estimator from the RV estimator. The quality of volatility estimates from point process based estimators has been verified by Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018). In these papers, Tse and Yang (2012) show that the volatility estimates from fitting an Autoregressive Conditional Duration (ACD) (Engle and Russell, 1998) to the absolute price change point process can outperform RV-type estimators under the assumption of various stochastic volatility models. With the same volatility estimator, Nolte, Taylor, and Zhao (2018) show that volatility estimates from the point process can provide better predictability compared to those from the RV and RV variants. Despite these promising results showing a clear advantage of the point process based volatility estimators over the RV-type estimators, its theoretical properties have not yet been established.

Closely linked to the parametric point process based volatility estimator, Andersen, Dobrev, and Schaumburg (2008) and Nolte, Taylor, and Zhao (2018) propose two different non-parametric volatility estimators that use the price duration, that is, the time for the cumulative price change to surpass a given threshold, as a measure of

volatility. They demonstrate that the duration-based volatility estimator can easily outperform the RV-type estimator in ideal conditions with a smaller mean squared error (MSE). Much of the theoretical properties of these non-parametric estimators have been discussed in these papers respectively, but none of them generalize the properties of these non-parametric estimators to a setting where both time-varying volatility and a general market microstructure noise (MMS) are present. Moreover, the duration based approach suffers from a truncation bias, when the price change is not exactly the value of the threshold. Together with the market microstructure noise, the consistency and asymptotic behaviour of these non-parametric estimators are largely unknown, which greatly hinders their applications in empirical studies.

We propose a general class of volatility estimators that we will refer to as the Renewal Based Volatility (*RBV*) estimators, which provides a theoretical framework for the aforementioned point process based volatility estimators (with the exception of the estimators in Andersen, Dobrev, and Schaumburg (2008)). This class of volatility estimators is constructed based on a renewal process in business time, which is a time change that treats the integrated variance as a measure of time. Based on this renewal process and the fact that the counts of events are shared by both business and calendar clocks, we can construct an estimator that estimates the time elapse in business time, which corresponds to the integrated variance in calendar time. As we do not require any knowledge about the dynamics of the volatility process, this estimator is by construction non-parametric. Moreover, we show that, by specifying a dynamic structure on the observed point process in the calendar time and defining a link function that maps the durations in calendar time to its counterparts in business time, one can construct parametric *RBV*-class estimators that can achieve a higher efficiency than their non-parametric counterparts. This includes the parametric duration-based volatility estimator as in Engle and Russell (1998), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018), and the intensity-based volatility estimator (Gerhard and Hautsch, 2002; Li, Nolte, and Nolte, 2018c). We derive the asymptotic distribution of both the non-parametric and parametric *RBV* estimators, and show that they are consistent as long as one can construct a renewal process in business time. One desirable property of this class of estimators is that, the asymptotic variance can be constructed without another estimation step (such as the estimation of integrated quarticity in the RV framework, see e.g. Barndorff-Nielsen and Shephard (2004)), which allows one to construct more precise confidence bounds.

We examine Nolte, Taylor, and Zhao's (2018) non-parametric duration-based volatility estimator (*NPD*) in the *RBV* framework as a complement to our theoretical discussion.

We formalize the properties of the *NPD* estimator for a general semimartingale setting in the presence of jumps, time-varying volatility, irregular arrivals of observations, price discretization and MMS noise. Our findings suggest that, firstly, the *NPD* estimator is more robust to jumps than a realized bipower variation estimator. Secondly, although the *NPD* estimator has a smaller asymptotic variance than the calendar time RV-type estimators in the absence of noise, it is very sensitive to MMS noise. Consequently, the *NPD* estimator will be biased upwards more heavily compared to a calendar time RV-type estimators of similar sampling frequency, which significantly weakens its relative performance.

By correcting the biases for the *NPD* estimator and exploiting its smaller asymptotic variance, we propose to construct the *NPD* estimator on the exponentially smoothed price process, which we will refer to as the exponentially smoothed *NPD* estimator, denoted by NPD^z . In our simulation we show that, if the smoothing parameter is chosen optimally, the truncation bias due to time discretization can approximately offset the smoothed MMS noise bias at moderate to large sampling frequencies. At these sampling frequencies, the NPD^z estimator exhibits a significantly higher efficiency compared to the commonly used bias corrected calendar time sampling volatility estimators, including the Realized Kernel (Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008a), the pre-averaged RV and pre-averaged bipower variation (Hautsch and Podolskij, 2013) estimators. Additionally, we demonstrate that, although the optimal sampling frequency of the NPD^z estimator is much smaller than its calendar time competitors, its optimal efficiency is still better than the optimal performance from its competitors, which requires a much larger sampling frequency.

The main contributions of this chapter are three-folded: Firstly, we develop a theoretical framework on which the asymptotic properties of the aforementioned point process based volatility estimators can be derived. Specifically, we show that, the duration-based volatility estimator is indeed superior to RV-type estimators in ideal conditions, and a parametric structure can lead to a substantial increase in the efficiency of volatility estimation. Secondly, we propose a range-duration based estimator that in theory is more efficient than any RV estimator under a stochastic sampling scheme discussed in Fukasawa (2010b) and Fukasawa and Rosenbaum (2012). However, the properties of this estimator in a more general setup is yet to be verified. Finally, we evaluate the theoretical properties of the non-parametric duration-based volatility estimator under a very general model. We propose the exponentially smoothed *NPD* estimator which shows a clear efficiency advantage over the commonly used bias corrected calendar time sampling volatility estimators.

The rest of the chapter is structured as follows: Section 1.2 describes the general theory for the renewal process and renewal reward process. Section 1.3 and 1.4 introduces the renewal based volatility estimator and the parametric renewal based volatility estimator respectively. Section 1.5 gives some examples on both the non-parametric and parametric estimators that belong to the class of renewal based estimators. In Section 1.6, we examine the *NPD* estimator under a general semi-martingale in the presence of various market imperfections. We conduct a Monte Carlo simulation study in Section 1.7. Section 1.8 concludes.

1.2 Prerequisites: Renewal Theory

This section summarizes the related renewal theory used in constructing the renewal based volatility estimator. For a more comprehensive discussion, please refer to standard point process textbooks, e.g. Wolff (1989), Ross (1996), etc.

We start with the definition of a renewal process:

Definition 1.1. Renewal Process: *Let $\{D_i\}_{i=1,2,\dots}$ be a sequence of positive i.i.d. random variables with $0 < \mu = \mathbf{E}[D_i] < \infty$ which represents the inter-event arrival time, and let t_i denote the arrival time of the i -th event (renewal epoch) given by:*

$$t_i = \sum_{j=1}^i D_j. \quad (1.1)$$

A renewal process $X(t)$ is defined as a random variable that counts the number of event arrivals in the interval $(0, t]$:

$$X(t) \equiv \sum_{i=1}^{\infty} \mathbb{1}_{\{t_i \leq t\}}. \quad (1.2)$$

A renewal process has the following asymptotic properties:

Theorem 1.1. Elemental Renewal Theorem: *Let $X = \{X(t)\}_{t \geq 0}$ be a renewal process with mean inter-arrival time $0 < \mu < \infty$ and renewal function $m(t) = \mathbf{E}[X(t)]$, then*

$$\lim_{t \rightarrow \infty} \frac{X(t)}{t} \xrightarrow{a.s.} \frac{1}{\mu}, \quad (1.3)$$

$$\lim_{t \rightarrow \infty} \frac{m(t)}{t} \rightarrow \frac{1}{\mu}. \quad (1.4)$$

Proof. See e.g. Feller (1941), Doob (1948) Theorem 3.3.4, Chapter 3 in Ross (1996). \square

A seemingly trivial result from the above theorem is that for a given $0 < \mu < \infty$, $\lim_{t \rightarrow \infty} X(t) \rightarrow \infty$. The renewal function, $m(t) = E[X(t)]$, has the following second order asymptotic expansion as $t \rightarrow \infty$:

Proposition 1.1. *Let $X(t)$ be a renewal process defined in Definition 1 with mean and variance of the inter-event arrival time denoted as $0 < \mu < \infty$ and $0 < \sigma^2 < \infty$ respectively. Let $m(t) = E[X(t)]$ denote the renewal function. The process $m(t)$ has the following asymptotic expansion as $t \rightarrow \infty$:*

$$m(t) = \frac{t}{\mu} + \frac{\sigma^2}{2\mu^2} - 0.5 + o(1). \quad (1.5)$$

Proof. E.g. Corollary 3.4.7 in Ross (1996) \square

It is useful to consider the distribution of time elapses since the last renewal epoch. This is known as the age process of a renewal process, formally defined as follows:

Definition 1.2. Age Process of A Renewal Process: *Let $X(t)$ denote a renewal process defined in Definition 1.1. The age process of a renewal process is defined as:*

$$A(t) = t - t_{X(t)} \quad (1.6)$$

The moments of $A(t)$ can be derived from the moments of the renewal process if they exist:

Theorem 1.2. *For an age process $A(t)$ defined in Definition 1.2, and let the n -th moments of the inter-epoch duration of the underlying be denoted by $E[D_i^n] = \mu_n$. Provided that all μ_n exist, the moments of the age process $A(t)$ can be expressed as:*

$$E[A^n(t)] = \frac{\mu_{n+1}}{(n+1)\mu} \quad (1.7)$$

Proof. See, e.g. Coleman (1982). \square

We will also use the property of a renewal reward process, which is defined as follows:

Definition 1.3. Renewal Reward Process: *Let $X(t)$ denote a renewal process with i.i.d. inter-event duration $\{D_i\}_{i=1,2,\dots}$ that has mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $\{R_i\}_{i=1,2,\dots}$ denote a sequence of real-valued i.i.d. random variables with mean*

$v < \infty$ and variance $\sigma_r^2 < \infty$ associated with each D_i . Then the renewal reward process is defined as:

$$R(t) = \sum_{i=1}^{X(t)} R_i. \quad (1.8)$$

The expectation of this process, $r(t) = \mathbb{E}[R(t)]$, is defined as the reward function.

A renewal reward process has the following asymptotic properties:

Theorem 1.3. Renewal Reward Theorem: *For a renewal reward process defined in Definition 1.3, the following results hold:*

$$\lim_{t \rightarrow \infty} \frac{R(t)}{t} \xrightarrow{a.s.} \frac{v}{\mu}, \quad (1.9)$$

$$\lim_{t \rightarrow \infty} \frac{r(t)}{t} \rightarrow \frac{v}{\mu}. \quad (1.10)$$

Proof. E.g. Theorem 3.6.1, Chapter 3 in Ross (1996). □

1.3 Renewal Based Volatility Estimator

We are now in the position of constructing the renewal based volatility (*RBV*) estimator for financial price processes. We start with an assumption about the price process and the associated volatility process of interest:

Assumption 1.1. Price Processes: *Let the log price process $\{P(t)\}_{t>0}$ be a stochastic process with an adapted, càdlàg and strictly positive integrated variance (IV) process defined by $IV(0, t) = \int_0^t \sigma_p^2(s) ds$ with $IV(0, t) \rightarrow \infty$ as $t \rightarrow \infty$. We define a time change $\tau(t) = IV(0, t)$ that converts the calendar time to the integrated variation time, which is also known as the business time. We assume that the time changed price process $\tilde{P}(\tau(t)) = P(t)$ is a Lévy process in business time.*

In the above assumption, we do not need to specify a particular form for the process $\sigma_p(t)$ as long as it satisfies Assumption 1.1. In this section we will assume that the complete trajectory of $P(t)$ can be observed, and the effect of discrete observation of the price process will be analysed in Section 1.6.

We can reverse the time change $t \mapsto \tau(t)$ by using $t = \inf\{u \in \mathbb{R}^+ : IV(0, u) \geq \tau(t)\}$. It is clear that $\tau(t)$ is a stopping time for any t . Also, the time changed information set has the relationship $\mathcal{F}_t = \tilde{\mathcal{F}}_{\tau(t)}$. For a more rigorous discussion on the change of time method, please refer to Chapter 1 in Barndorff-Nielsen and Shiryaev (2010).

Assumption 1.1 may seem strict, but it is satisfied by a wide range of stochastic processes that are used in modelling financial price processes. We give two simple examples.

Example 1: Any continuous local martingale satisfies this assumption due to the following theorem.

Theorem 1.4. (Dambis-Dubin Schwarz): *Let $(M(t))_{t \geq 0}$ be a continuous \mathcal{F}_t -local martingale such that its quadratic variation $\langle M \rangle_\infty = +\infty$, then there exists a Brownian motion $(B(t))_{t \geq 0}$, such that for every $t \geq 0$, $M(t) = B(\langle M \rangle_t)$.*

Since the quadratic variation and integrated variance of $M(t)$ coincide, the resulting Lévy process in business time is a standard Wiener process. Note that Theorem 1.4 still holds when the stochastic volatility and the price process are correlated, which is known as the ‘leverage effect’ that is commonly observed in practice (see e.g. Bollerslev, Litvinova, and Tauchen (2006)).

Example 2: A (inhomogeneous) compounded Poisson process as in Oomen (2005) satisfies this assumption. The resulting Lévy process is a homogeneous compounded Poisson process. See Appendix A.1 for details.¹

The connection from the Lévy process in business time and the renewal theory in the previous section is established by the following proposition:

Proposition 1.2. *Let $\{Y(t)\}_{t \geq 0}$ be a Lévy process on the filtered probability space $\{\Omega, \mathcal{F}, P\}$. Define a stopping time process that automatically renews once stopped as:*

$$t_i = \inf_{t \geq t_{i-1}} \{Y(t) \in \mathcal{S}(t_{i-1})\}, \quad (1.11)$$

in which $\mathcal{S}(t_i)$ is the stopping condition for t_i as a function of \mathcal{F}_{t_i} . If, for any i, j and $t > 0$, $\text{Prob}(Y(t_i + t) \in \mathcal{S}(t_i)) = \text{Prob}(Y(t_j + t) \in \mathcal{S}(t_j))$, then the sequence $\{t_i\}_{i=1,2,\dots}$ corresponds to arrivals of a renewal process.

Proof. The condition $\text{Prob}(Y(t_i + t) \in \mathcal{S}(t_i)) = \text{Prob}(Y(t_j + t) \in \mathcal{S}(t_j))$ ensures that the stopping condition is equivalent to the paths of the Lévy process originating from all the possible starting points $t_i \in (0, \infty)$ regardless of when the previous event occurred. Then clearly the distribution of $t_i - t_{i-1}$ is i.i.d., which follows from the property of the Lévy process. As a result, $\{t_i\}_{i=1,2,\dots}$ is by definition a renewal process. \square

¹Relying on Theorem 1.4, we can account for the leverage effect if the latent price process follows a continuous local martingale. However, it is not clear if this claim is still valid in the case of this example, or other alternative specifications.

Consequently, when the price process $P(t)$ follows Assumption 1.1, we can obtain a Lévy process $\tilde{P}(\tau(t))$ in business time. According to Proposition 1.2, we can construct a renewal process $\{\tau(t_i)\}_{i=1,2,\dots}$ in business time by choosing an appropriate $\mathcal{S}(\tau(t_i))$ for each i . Effectively, we sample the price process at $\{t_i\}_{i=1,2,\dots}$ in calendar time in such way that the business time counterpart $\{\tau(t_i)\}_{i=1,2,\dots}$ is renewal. We therefore refer to this sampling scheme as renewal sampling:

Definition 1.4. Renewal Sampling: *For a price process $P(t)$ satisfying Assumption 1.1, a renewal sampling scheme samples $P(t)$ at $0 < t_1 < t_2 < \dots$ where the arrivals in business time $\{\tau(t_i)\}_{i=1,2,\dots}$ is a renewal process in business time. Denote the unobservable renewal process in business time as $\tilde{X}(\tau(t)) = \sum_{i>0} \mathbb{1}_{\{\tau(t_i) \leq \tau(t)\}}$ and its observable calendar counterpart as $X(t) = \sum_{i>0} \mathbb{1}_{\{t_i \leq t\}}$.*

Note that the càdlàg property of the integrated variance guarantees that $X(t) = \tilde{X}(\tau(t))$. Using Proposition 1.2, we can construct $X(t)$ in calendar time if the stopping condition in calendar time is only a function of the paths of $P(t)$, but not a function of time. Heuristically, by observing the path of the price process in calendar time, we can decide where to ‘stop’ the price process and obtain a sample. If the condition in Proposition 1.2 for $\mathcal{S}(t_i)$ is satisfied, then the stopping times in business time is by construction a renewal process.

The central contribution of this chapter is the following novel volatility estimator by sampling the price process $P(t)$ with a renewal sampling scheme:

Definition 1.5. Renewal Based Volatility (RBV) Estimator: *Let $\{P(t)\}_{t>0}$ be a price process that satisfies Assumption 1.1. Choose a $\mathcal{S}(t_i)$ according to Proposition 1.2, and apply renewal sampling on $\tilde{P}(\tau(t))$ to obtain the renewal sampling times $\{t_i\}_{i=1,2,\dots}$ and the point process $X(t) = \sum_{i>0} \mathbb{1}_{\{t_i \leq t\}}$, which has a business time counterpart $\tilde{X}(\tau(t))$ that is a renewal process. Let $0 < \mu < \infty$ and $0 < \sigma^2 < \infty$ denote the first two moments of the inter-epoch duration in business time, then the RBV estimator is defined by:*

$$RBV(0, t) = X(t)\mu. \tag{1.12}$$

The RBV estimator has the following asymptotic distribution:

Theorem 1.5. *The Renewal Based Volatility estimator as defined in Definition 1.5 has the following asymptotic distribution:*

$$\lim_{t \rightarrow \infty} \frac{RBV(0, t) - IV(0, t)}{\sqrt{X(t)}\sigma} \xrightarrow{d} \mathcal{N}(0, 1) \tag{1.13}$$

Proof. See Appendix A.2. □

One remark on the *RBV* estimator is that we can compute standard errors of the estimator without estimating the integrated quarticity as in the RV literature, which implies less estimation bias for the standard errors and confidence bounds. Similar to the RV-type estimators, the *RBV* estimator does not require any parametric assumption on the IV process in calendar time. The obvious problem here is that μ is not explicitly specified, and is dependent on the assumption of $P(t)$ and the stopping condition $\mathcal{S}(t_i)$. In Section 1.5 we show that in some special cases μ is available in closed form. Also, the process $\tilde{P}(\tau(t))$ is usually very simple (for example, a standard Wiener process). In this case the moments of the renewal process can be simulated easily.

We would like to point out that in (1.13), the limiting distribution is obtained when μ is fixed and $t \rightarrow \infty$. This is known as the sprawl asymptotics, or the long-span asymptotics, which is typical in the context of point processes. However, this is different to the infill asymptotics usually applied in the RV context where the time frame is fixed and the sampling frequency increases. To derive counterpart of Theorem 1.5 in the infill asymptotics setting, more assumptions are required for the asymptotic behaviour of $P(t)$ and $X(t)$, which is presented in Appendix A.3.

To distinguish between the two asymptotic settings, for the rest of the chapter, the sprawl asymptotics is involved when the construction of the price durations is fixed. Consequently, the moments of the price durations in business time, namely, μ , σ^2 , are fixed. Asymptotic results are derived by expanding the sampling window and letting $t \rightarrow \infty$. For the infill case, the time span is fixed and we allow the moments of the price durations in business time to change. Asymptotic results in this case are typically derived by letting $\mu \rightarrow 0$ for a fixed time frame.

1.4 Parametric Renewal Based Volatility Estimator

The duration in business time \tilde{D}_i is not directly observable, but we can observe its calendar time counterpart D_i . Using the fact that \tilde{D}_i is i.i.d., the connection between D_i and the integrated variance process is that:

$$\int_{t_{i-1}}^{t_{i-1}+D_i} \sigma_p^2(s) ds = \tilde{D}_i. \quad (1.14)$$

If we can specify a parametric model $g(t|\mathcal{F}_t)$ that uses all the information available in such a way that the following variable is i.i.d:

$$R_i = \int_{t_{i-1}}^{t_{i-1}+D_i} g(s|\mathcal{F}_s)ds, \quad (1.15)$$

then we can use the quantity $\frac{R_i\mu}{\mathbb{E}[R_i]}$ as an estimator for \tilde{D}_i . Without any loss of generality we set $\mathbb{E}[R_i] = \mu$ to simplify notation. We will refer to this estimator as the parametric renewal based volatility (*PRBV*) estimator, formally defined as follows:

Definition 1.6. Parametric Renewal Based Volatility (*PRBV*) Estimator:

Let $\{P(t)\}_{t>0}$, $X(t)$, $\tilde{X}(\tau(t))$, μ and σ^2 be defined identically to Definition 1.5. Define a parametric model $g(t|\mathcal{F}_t)$ and an i.i.d. variable R_i that follows (1.15) with $0 < \mathbb{E}[R_i] = \mu < \infty$ and $0 < \mathbb{V}[R_i] = \sigma_r^2 < \infty$. Then the *PRBV* estimator is defined as:

$$PRBV(0,t) = \sum_{i=1}^{X(t)} R_i = \int_0^t g(s|\mathcal{F}_s)ds. \quad (1.16)$$

Recall that the *RBV* estimator is already consistent, therefore for any i.i.d. R_i with finite moments, the *PRBV* estimator will still be consistent. However, the randomness in R_i may introduce extra noise in the *PRBV* estimator, unless there exists a substantial amount of positive correlation between R_i and \tilde{D}_i , which requires that $g(t|\mathcal{F}_t)$ is a good proxy of $\sigma_p^2(t)$ for all t . Thus, we can assess the efficiency of the *PRBV* estimator by using the *RBV* estimator as a benchmark.

Conditioning on that we can observe the i.i.d. variable R_i , the asymptotic distribution of the *PRBV* estimator can be derived analogously to the derivation of Theorem 1.5 noting that $R_i - \tilde{D}_i$ is a zero-mean i.i.d. variable, and the pair $\{R_i, \tilde{D}_i\}$ forms a renewal reward process.

Theorem 1.6. *The Parametric Renewal Based Volatility estimator as defined in Definition 1.6 has the following asymptotic distribution:*

$$\lim_{t \rightarrow \infty} \frac{PRBV(0,t) - IV(0,t)}{\sqrt{X(t)(\sigma^2 + \sigma_r^2 - 2\rho\sigma\sigma_r)}} \xrightarrow{d} \mathcal{N}(0,1) \quad (1.17)$$

where ρ is the correlation between R_i and \tilde{D}_i .

Proof. This follows similarly from the proof in Appendix A.2 by using that the variable $R_i - \tilde{D}_i$ is i.i.d. with zero mean. Note that $\mathbb{V}[R_i - \tilde{D}_i] = \sigma^2 + \sigma_r^2 - 2\rho\sigma\sigma_r$. \square

Note that the variance of *PRBV* is zero when $R_i = \tilde{D}_i$ for all i , indicating that the *PRBV*(0,t) can in theory be a perfect estimator for the integrated variance when R_i

is known. The variance of the *PRBV* estimator can be written as:

$$V[PRBV(0,t)] = V[RBV(0,t)] + X(t)(\sigma_r^2 - 2\rho\sigma\sigma_r), \quad (1.18)$$

and as long as $\sigma_r^2 - 2\rho\sigma\sigma_r < 0$, that is, $\rho \in (\frac{\sigma_r}{2\sigma}, 1]$, the *PRBV* estimator will always be more efficient than the *RBV* estimator. Obviously, the value of ρ is determined by the distance between $g(t|\mathcal{F}_t)$ and $\sigma_p^2(t)$, which is unfortunately model dependent.

We provide an example of $g(t|F_t)$ which allows us to examine ρ directly. Initially proposed by Gerhard and Hautsch (2002) derived from the instantaneous volatility estimator of Engle and Russell (1998), the conditional intensity process of $X(t)$ is used as a proxy of the instantaneous volatility. We define $g(t|F_t)$ as follows:

$$g(t|F_t) = \mu\lambda(t|\mathcal{F}_t), \quad (1.19)$$

where $\lambda(t|\mathcal{F}_t)$ is the \mathcal{F}_t -conditional intensity of the process $X(t)$ defined as:

$$\lambda(t|\mathcal{F}_t) \equiv \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{E}[X(t+\Delta) - X(t)|\mathcal{F}_t]. \quad (1.20)$$

The corresponding renewal reward variable R_i is then defined as:

$$R_i = \mu \int_{t_{i-1}}^{t_i} \lambda(s|\mathcal{F}_s) ds \equiv \mu\Lambda(t_{i-1}, t_i). \quad (1.21)$$

The i.i.d.-ness of R_i is guaranteed by the following theorem:

Theorem 1.7. Random Time Change Theorem (RTCT): *Let $X(t)$ be a simple point process adapted to a history \mathcal{F}_t with bounded, strictly positive \mathcal{F}_t -conditional intensity $\lambda(t|\mathcal{F}_t)$ and \mathcal{F}_t -compensators $\Lambda(t) = \int_0^t \lambda(u|\mathcal{F}_u) du$ with $\Lambda(\infty) = \infty$ almost surely. Under the random time change $t \mapsto \Lambda(t)$, the transformed process*

$$\tilde{X}(t) = X(\Lambda^{-1}(t))$$

is a Poisson process with unit rate.

Conversely, suppose there is given a history \mathcal{G}_t , a \mathcal{G}_t -adaptive cumulative process $M(t)$ with a.s. finite, monotonically increasing and continuous trajectories, and a \mathcal{G}_t -adapted simple Poisson process $X_0(t)$. Let \mathcal{F}_t denote the history of σ -algebras $\mathcal{F}_t = \mathcal{G}_{M(t)}$. Then $X(t) = X_0(M(t))$ is a simple point process that is \mathcal{F}_t -adapted and has \mathcal{F}_t -compensator $M(t)$.

Proof. See the proof in Theorem 7.4I in Daley and Vere-Jones (2003), Brown and Nair (1988) and Bowsher (2007). □

Theorem 1.7 suggests that $\Lambda(t_{i-1}, t_i) \sim i.i.d. \exp(1)$, so we have $R_i \sim i.i.d. \exp(\mu^{-1})$. The mean and variance of R_i are $\nu = \mu$ and $\sigma_r^2 = \mu^2$ respectively. We derive the following important proposition that characterizes the relationship between the conditional intensity processes in calendar time and business time:

Proposition 1.3. *Let $X(t)$ be a simple point process with conditional intensity process $\lambda(t|\mathcal{F}_t)$, and let $t \mapsto \tau(t)$ be a change of time from calendar time to business time. The conditional intensity process $\tilde{\lambda}(\tau(t)|\tilde{\mathcal{F}}_{\tau(t)})$ of the time-changed point process $\tilde{X}(\tau(t))$ follows:*

$$\tilde{\lambda}(\tau(t)|\tilde{\mathcal{F}}_{\tau(t)})\sigma_p^2(t) = \lambda(t|\mathcal{F}_t). \quad (1.22)$$

Proof. See Appendix A.4. □

Proposition 1.3 has some very powerful implications that provide theoretical foundations for intensity and duration based volatility estimation.

Corollary 1.1. *In Definition 1.6 with $g(t|\mathcal{F}_t) = \mu\lambda(t|\mathcal{F}_t)$, the rank correlation between R_i and \tilde{D}_i is 1. Additionally, if \tilde{D}_i is i.i.d. exponentially distributed, then $g(t|\mathcal{F}_t) = \sigma_p^2(t)$.*

Proof. See Appendix A.5. □

Corollary 1.1 suggests that, firstly, the *PRBV* estimator is likely to perform well due to the monotonic non-linear relationship between R_i and \tilde{D}_i . Secondly, the optimal renewal sampling scheme for $g(t|F_t) = \mu\lambda(t|\mathcal{F}_t)$ is a homogeneous Poisson sampling scheme in business time. In this case, the conditional intensity of $X(t)$ in calendar time is proportional to the spot volatility, so that the conditional intensity is a perfect estimator of instantaneous volatility for all t . However, the assumption that \tilde{D}_i is i.i.d. exponentially distributed requires further assumptions on the price process (e.g. the compounded Poisson process in Oomen (2006)), which is not desirable.

Alternatively, we can also correct for the discrepancy between the density of \tilde{D}_i and R_i :

Corollary 1.2. *In Definition 1.6 with $g(t|F_t) = \mu\lambda(t|\mathcal{F}_t)$, let $F_{\tilde{D}}(x)$ and $F_{\tilde{D}}^{-1}(x)$ denote the CDF of \tilde{D}_i and its inverse correspondingly. The following relationship holds for all i :*

$$\tilde{D}_i = F_{\tilde{D}}^{-1}(1 - \exp(-R_i/\mu)) \quad (1.23)$$

Proof. This is straightforward from the proof in Appendix A.5. □

The expression $F_{\tilde{D}}^{-1}(1 - \exp(-R_i/\mu))$ is effectively an exponential inverse probability integral transformation of \tilde{D}_i , which is a perfect estimator for the volatility between the two points $(t_{i-1}, t_i]$. However, this is a weaker result compared to Corollary 1.1 because in general $g(t|\mathcal{F}_t) \neq \sigma_p^2(t)$. In this case, inference based on R_i only reflects \tilde{D}_i in expectation with $E[R_i] = E[\tilde{D}_i]$ with a rank correlation of 1, and $g(t|\mathcal{F}_t)$ does not estimate the actual spot volatility.

From above, it is clear that regardless of the value of μ , if we know the true conditional intensity process in calendar time, we in principle know the underlying integrated variance process. Therefore, in practice one does not need to sample at ultra high-frequency to improve the precision of the volatility estimates, which is the common approach in the RV literature. Instead, one only needs to append the estimation window of the econometric model of the conditional intensity process to obtain a more precise estimate of the conditional intensity, which in turn leads to a more precise estimate of \tilde{D}_i for each i . This is in stark contrast with the RV-type estimators which relies heavily on the availability of data within the volatility estimation window. We stress that this is a very important property of the *PRBV* estimator that validates the intraday volatility estimates as in Engle and Russell (1998) and Tse and Yang (2012), and also renders the *PRBV* estimator advantageous over the RV-type estimator in the situation where the availability of data is limited.

To summarize our findings on the *PRBV* estimators, we have shown that, it is possible to construct a *PRBV* estimator as in Corollary 1.2 that always has zero variance if R_i is known. However, these properties are unlikely to hold in practice as we do not observe $g(t|\mathcal{F}_t)$ and have to use a model to estimate $\hat{g}(t|\mathcal{F}_t)$ and \hat{R}_i instead. This will inevitably introduce estimation noise in the model, even if the specification of $g(t|\mathcal{F}_t)$ is correct. As this is more related to the properties of the econometric model used for the observed point process that deserves individual investigations, we will leave it for future research.

1.5 Some Examples

We give some concrete examples of *RBV* and *PRBV* in this section and summarize their properties. Assume the efficient log-price follows a semi-martingale of the following form:

$$dP(t) = \alpha(t)dt + \sigma_p(t)dW(t), \quad (1.24)$$

where $\alpha(t)$ is a continuous \mathcal{F}_t -predictable process and $\sigma(t)$ is assumed to be càdlàg and strictly positive with $\int_0^t \sigma^2(s)ds \rightarrow \infty$ when $t \rightarrow \infty$. For now, we assume $\alpha(t) = 0$

and no discontinuities in the diffusion process for simplicity, and will discuss the effect of the drift term and jumps in the next section. The quantity of interest here is the integrated variance of the process over an interval $(0, T)$:

$$IV(0, T) = \int_0^T \sigma_p^2(s) ds. \quad (1.25)$$

Example 1: The first example of an *RBV* estimator, which will also be examined in detail in later sections, is the non-parametric duration-based (*NPD*) volatility estimator proposed by Nolte, Taylor, and Zhao (2018). We start by defining the absolute price change point process, firstly introduced by Engle and Russell (1998):

Definition 1.7. The Absolute Price Change Point Process: *The absolute price change point process $\{t_i^{(\delta)}\}_{i=0,1,\dots}$ for an observed price process $P(t)$ and a given price change threshold δ is constructed as follows:*

1. Set $t_0^{(\delta)} = 0$ and choose a threshold δ .
2. For $i = 1, 2, \dots$, compute the first exit time, $t_i^{(\delta)}$, of $P(t_{i-1}^{(\delta)})$ through the double barrier $[P(t_{i-1}^{(\delta)}) - \delta, P(t_{i-1}^{(\delta)}) + \delta]$ as:

$$t_i^{(\delta)} = \inf_{t > t_{i-1}^{(\delta)}} \{|P(t) - P(t_{i-1}^{(\delta)})| \geq \delta\}.$$

Iterate until the sample is depleted.

The arrivals of $t_i^{(\delta)}$ are referred to as price events. In the *RBV* framework, we can write $\mathcal{S}^{(\delta)}(t_i^{(\delta)}) = \{P(t_i^{(\delta)}) - \delta, P(t_i^{(\delta)}) + \delta\}$ and clearly it satisfies the condition in Proposition 1.2. Define the time change as $\tau(t) = \int_0^t \sigma^2(s) ds = IV(0, t)$, and $P(\tau(t))$ is a standard Brownian motion by Theorem 1.4. As a result from Theorem 1.2, under business time, $\{\tau(t_i^{(\delta)})\}_{i=1,2,\dots}$ forms a renewal process, denoted by $X^{(\delta)}(\tau(t))$.

Let $D_i^{(\delta)} = t_i^{(\delta)} - t_{i-1}^{(\delta)}$ and $\tilde{D}_i^{(\delta)} = \tau(t_i^{(\delta)}) - \tau(t_{i-1}^{(\delta)})$ denote the duration under calendar time and business time respectively. Note that $\tilde{D}_i^{(\delta)}$ is the stopping time for a Wiener process (starting at zero) to exit a symmetric interval $[-\delta, \delta]$. We can retrieve its moments from its moment generating function (see Table 1 in Andersen, Dobrev, and Schaumburg (2008)). The first three moments are:

$$E[\tilde{D}_i^{(\delta)}] = \delta^2, \quad E[(\tilde{D}_i^{(\delta)})^2] = \frac{5}{3}\delta^4, \quad E[(\tilde{D}_i^{(\delta)})^3] = \frac{61}{15}\delta^6. \quad (1.26)$$

The *NPD* estimator in Nolte, Taylor, and Zhao (2018) is of the following form:

$$NPD(0, t) = X^{(\delta)}(t)\delta^2 = X^{(\delta)}(t)\mu(\delta). \quad (1.27)$$

Note we use the notation $\mu(\delta)$ and $\sigma^2(\delta)$ to denote the mean and variance of the price duration in business time for some δ . Therefore it is clear that the *NPD* estimator belongs to the class of *RBV* estimators. The asymptotic distribution of the *NPD* estimator can be derived easily from (1.13):

$$\lim_{t \rightarrow \infty} \frac{NPD(0,t) - IV(0,t)}{\sqrt{\frac{2}{3}X^{(\delta)}(t)\delta^4}} \xrightarrow{d} \mathcal{N}(0,1) \quad (1.28)$$

Using the asymptotic relationship $\delta^2 = \frac{IV(0,t)}{X^{(\delta)}(t)}$, we see that $V[NPD(0,t)] \rightarrow \frac{2IV(0,t)^2}{3X^{(\delta)}(t)}$. This suggests that, given a common sampling frequency, on average the *NPD* estimator will be more than six times as efficient as the RV sampled in calendar time, exactly six times as efficient as the RV sampled in business time, and more efficient than the RV under tick time sampling due to that $IV(0,t)^2 \leq IQ(0,t)$ from Jensen's inequality (Fukasawa, 2010a).

The efficiency gain from the RV estimator is not surprising. Since the *NPD* estimator uses information in the path of the prices, it effectively uses more data than the RV estimator under the same sampling frequency. Additionally, as discussed in Section A.3.1, the *NPD* estimator is both an *RBV* estimator and a renewal RV estimator. It achieves the optimal efficiency for the renewal RV estimators due to the fact that the kurtosis of the return is 1.

Example 2: Inspired by Christensen and Podolskij (2007) and Andersen, Dobrev, and Schaumburg (2008) and following the idea of the *NPD* estimator, we can also construct a range duration-based *RBV*-type volatility estimator. Let r denote a fixed range size, then the following sequence of stopping times forms a renewal process in business time:

$$t_i^{(r)} = \inf_{t > t_{i-1}^{(\delta)}} \{P(t) \in \mathcal{S}^{(r)}(t_i^{(r)})\}, \quad (1.29)$$

where $\mathcal{S}^{(r)}(t_i^{(r)}) = \{P(t) : \sup_{t_i^{(r)} < s < t} P(s) - \inf_{t_i^{(r)} < s < t} P(s) \geq r\}$. Similar to the *NPD* estimator, let $X^{(r)}(\tau(t))$ denote the renewal process under business time. The first three moments of $\tilde{D}_i^{(r)} = \tau(t_i^{(r)}) - \tau(t_{i-1}^{(r)})$ is as follows (Andersen, Dobrev, and Schaumburg, 2008)):

$$E[\tilde{D}_i^{(r)}] = \frac{1}{2}r^2, \quad E[(\tilde{D}_i^{(r)})^2] = \frac{1}{3}r^4, \quad E[(\tilde{D}_i^{(r)})^3] = \frac{17}{60}r^6, \quad (1.30)$$

and the non-parametric range duration-based volatility (*NPR*) estimator is simply:

$$NPR(0, t) = \frac{1}{2} X^{(r)} r^2, \quad (1.31)$$

which has the following asymptotic distribution as $t \rightarrow \infty$:

$$\lim_{t \rightarrow \infty} \frac{NPR(0, t) - IV(0, t)}{\sqrt{\frac{1}{12} X^{(r)}(t) r^4}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (1.32)$$

Using the asymptotic relationship $r^2 = \frac{2IV(0, t)}{X^{(r)}(t)}$, we have $V[NPR(0, t)] = \frac{IV^2(0, t)}{3X^{(r)}(t)}$. So the *NPR* estimator is twice as efficient as the *NPD* estimator for a common sampling frequency.

The efficiency gain of the range-based estimators compared to the RV-based estimators has been addressed by Christensen and Podolskij (2007) and Andersen, Dobrev, and Schaumburg (2008), as price ranges exploit both the supremum and infimum of the price process, which can measure volatility more precisely than using price changes. We would like to stress that the asymptotic variance of the *NPR* estimator is smaller than the asymptotic variance of a general RV estimator under any sampling scheme (Fukasawa, 2010b; Fukasawa and Rosenbaum, 2012). With this *NPR* example, it is clear that the *RBV*-class of estimators are in essence different from the RV-type estimators.

Example 3: The parametric duration (intensity) based volatility estimator, initially proposed by Engle and Russell (1998) and further developed by Gerhard and Hautsch (2002), Tse and Yang (2012), Nolte, Taylor, and Zhao (2018) and Li, Nolte, and Nolte (2018b) is an example of a *PRBV* estimator. Specifically, it specifies the dynamics of $D_i^{(\delta)}$ with a fully parametric model (for example, the Autoregressive Conditional Duration model by Engle and Russell (1998)), and defines

$$g^{(\delta)}(t|F_t) = \mu \lambda^{(\delta)}(t|\mathcal{F}_t) = \delta^2 \lambda^{(\delta)}(t|\mathcal{F}_t), \quad (1.33)$$

in which $\lambda^{(\delta)}(t|\mathcal{F}_t)$ is the conditional intensity process of $X^{(\delta)}(t)$ defined in (1.20). Gerhard and Hautsch (2002) propose an instantaneous volatility estimator defined as $InsV^{(\delta)}(t) = g^{(\delta)}(t|F_t)$, and an estimator of the IV between the arrival of two price events can be constructed as follows:

$$R_i^{(\delta)} = \int_{t_{i-1}^{(\delta)}}^{t_i^{(\delta)}} g(s|\mathcal{F}_s) ds = \delta^2 \Lambda_i^{(\delta)} \sim i.i.d. \exp(\delta^{-2}), \quad (1.34)$$

with $E[R_i^{(\delta)}] = \delta^2$ and $V[R_i^{(\delta)}] = \delta^4$. As this quantity is i.i.d. from Theorem 1.7, the parametric duration (intensity) based (PD) estimator of the following form:

$$PD(0, t) = \sum_{i=1}^{X^{(\delta)}} R_i^{(\delta)}, \quad (1.35)$$

is by definition a *PRBV*-class estimator. The asymptotic properties of the *PRBV* estimator discussed in Theorem (1.6) and Proposition 1.3 can be applied directly to derive the asymptotic distribution of the *PD* estimator:

$$\lim_{t \rightarrow \infty} \frac{PD(0, t) - IV(0, t)}{\sqrt{C \cdot X^{(\delta)}(t) \delta^4}} \xrightarrow{d} \mathcal{N}(0, 1), \quad (1.36)$$

in which C is a constant which cannot be solved analytically. From Proposition 1.3, since $\tilde{D}_i^{(\delta)}$ can be easily simulated based on a Wiener process, we can simulate the constant C easily. Details of this simulation can be found in Appendix A.6. Based on 1000000 replications, we found that $C \approx 0.034$. Therefore, the asymptotic variance of the *PD* estimator is roughly one-twentieth of the *NPD* counterpart. It shows that, if the parametric model of $\lambda^{(\delta)}(t | \mathcal{F}_t)$ is well-specified, then there can be a substantial efficiency gain from the parametric estimation.

Based on the *NPR* estimator, we can construct a parametric range (*PR*) based volatility estimator by defining the renewal variable $R_i^{(r)}$ as:

$$R_i^{(r)} = 0.5r^2 \int_{t_{i-1}^{(r)}}^{t_i^{(r)}} \lambda^{(r)}(s | \mathcal{F}_s) ds. \quad (1.37)$$

The *PR* estimator is defined analogously to the *PD* estimator as:

$$PR(0, t) = \sum_{i=1}^{X^{(r)}} R_i^{(r)} = R^{(r)}(t), \quad (1.38)$$

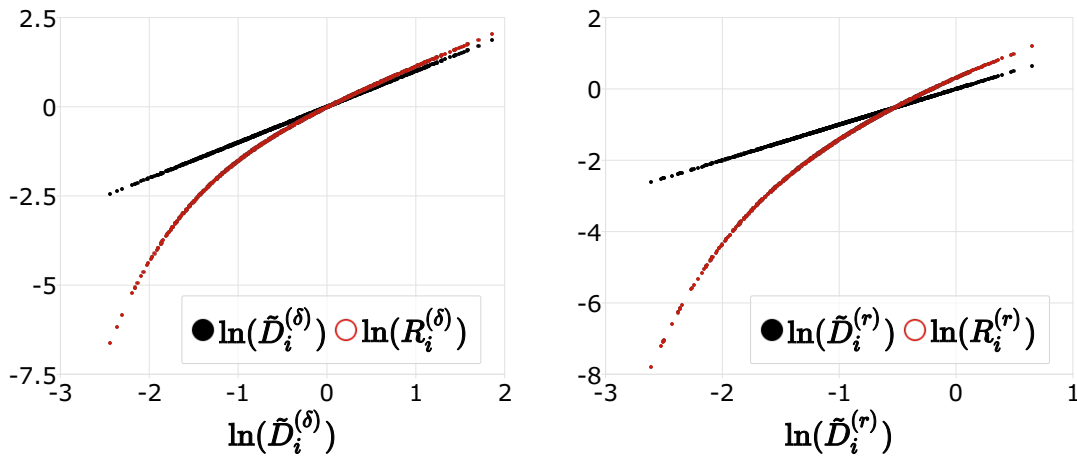
From the simulation in Appendix A.6 and the property of the *PRBV* estimator, we can derive the asymptotic distribution of *PD* given $R_i^{(r)}$:

$$\lim_{t \rightarrow \infty} \frac{PR(0, t) - IV(0, t)}{\sqrt{C \cdot X^{(r)}(t) r^4}} \xrightarrow{d} \mathcal{N}(0, 1), \quad (1.39)$$

where $C \approx 0.024$. Actually the *PR* estimator has a larger asymptotic variance than the *PD* estimator if we control for the same sampling frequency (when $r^2 = 2\delta^2$). This is due to the fact that the density $\tilde{D}_i^{(r)}$ deviates further from the exponential distribution.

We would like to stress that, although the *PD* and *PR* estimators for the integrated variance are unbiased and consistent in the sense of expectation, using $R_i^{(\cdot)}$ as an estimator of $\tilde{D}_i^{(\cdot)}$ will introduce a non-zero error due to the discrepancy between $R_i^{(\cdot)}$ and $\tilde{D}_i^{(\cdot)}$. We plot the simulated $\ln(\tilde{D}_i^{(\cdot)})$ against $\ln(R_i^{(\cdot)})$ in Figure 1.1. The figure suggests that, as the discrepancy between $R_i^{(r)}$ and $\tilde{D}_i^{(r)}$ is larger than that of $R_i^{(\delta)}$ and $\tilde{D}_i^{(\delta)}$, the *PR* estimator will be less efficient compared to the *PD* estimator. Also, based on the simulated $\tilde{D}_i^{(\cdot)}$, one can correct this discrepancy by the method in Corollary 1.2. After the correction, both estimators will have zero variance conditioning on the knowledge of $R_i^{(\cdot)}$.

Figure 1.1 Discrepancy between the density of $R_i^{(\cdot)}$ and $\tilde{D}_i^{(\cdot)}$



Note: $N = 1000000$. Descriptive statistics of $\{\tilde{D}_i^{(\cdot)}\}_{i=1:N}$ and $\{R_i^{(\cdot)}\}_{i=1:N}$ can be seen in Table A.1.

The results above suggest that, if the parametric model to estimate $R_i^{(\cdot)}$ performs equally well, then the two parametric estimators will have equal performance. This is in contrast to the efficiency difference between the *NPD* and *NPR* estimators as the *NPD* estimator is half as efficient as the *NPR* estimator. Because the variance of the reward variable $R_i^{(\cdot)}$ offsets the variance of $\tilde{D}_i^{(\cdot)}$ completely, the advantage of a lower variance for $\tilde{D}_i^{(r)}$ for the *NPR* estimator disappears. However, the *PR* estimator might be still preferred over the *PD* estimator because in a finite sample, one can obtain a larger sample size with range-based renewal sampling, resulting in more precise estimates for $R^{(\cdot)}$. Finally, as a result from Corollary 1.1, the instantaneous volatility estimator proposed in Gerhard and Hautsch (2002) does not hold for all t if the price process is assumed to follow (1.24), and can only serve as a proxy for the instantaneous volatility.

1.6 The *NPD* Volatility Estimator Under Market Frictions

This section discusses the theoretical properties of the *NPD* volatility estimator defined in (1.27) in the presence of drift, jumps, time discretization, market microstructure noise, and rounding effect.

1.6.1 Drift Effect

This section aims to clarify that the drift will not bias our estimator in an infill setting.² As the drift effect is very small in empirical high frequency applications, we will follow the approach by Barndorff-Nielsen and Shephard (2002) and discuss the drift effect in this section and assume it to be zero in other sections.

Firstly, as discussed in previous section, the *NPD* estimator is also a renewal RV estimator, and the quadratic variation theory can be applied. Therefore under the assumption of (1.24) and as $\delta \rightarrow 0$, the drift term will not bias our estimator. Moreover, we would like to note that there always exists a probability measure where the price process does not possess a drift by the use of Girsanov-Maruyama transformation. As the volatility remains unchanged after the change of measure and the *NPD* estimator can also be constructed on that probability measure, the presence of a drift is not a main concern under the infill setting.

1.6.2 Jump Effect

This section discusses the possible effect of jumps on the *NPD* estimator in the infill asymptotics setting. The *NPD* estimator is by construction very robust to large jumps, as pointed out by Andersen, Dobrev, and Schaumburg (2008), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018), because of its truncation feature. For simplicity, we consider the following diffusion process with jumps on the interval $(0, T]$:

$$P(t) = P(0) + \int_0^t \sigma_p(s) dW(s) + \sum_{j=1}^{J(t)} L_j, \quad (1.40)$$

where $J(t)$ is a counting process independent of $W(t)$, and L_j is the size of the j -th jump. We assume that $|L_j - \delta| > 0$ almost surely for all j as $\delta \rightarrow 0$, so that

²Note that a drift in (1.24) can in general be estimated in the long-span asymptotics, and it will also introduce a positive bias to the *NPD* estimator. We were unable to derive a bias correction with the presence of drift in this setting.

each arrival of jump will almost surely trigger a price event. For a simple RV-type estimator under any sampling scheme, the $IV(0, t)$ estimates will be positively biased and include the jump variations. Let us consider the point process $\tilde{X}^{(\delta)}(\tau(t))$ under business time with $\tau(t) = IV(0, t)$. Denote the number of jumps in the duration $\tilde{D}_i^{(\delta)}$ by J_i , then conditional on that there is no jumps in the duration, the conditional mean and variance are δ^2 and $\frac{2}{3}\delta^4$ respectively. For the durations that contain a jump, it will always end the duration with the jump.

We can split the durations in business time by whether they contain a jump. For the durations that do not contain a jump, we have $E[\tilde{D}_i^{(\delta)} | J_i = 0] = \delta^2$. For the durations that contain a jump, we interpret the jump as a random inspection time to a duration in business time, and the renewal process is immediately renewed when it is inspected. The time travelled on the business clock till the inspection time but before the actual price event would have occurred is therefore the length of the duration in business time. The density of the duration that contains a jump then can be interpreted as the age process of the renewal process defined in Definition 1.2. According to Theorem 1.2, we have:

$$E[\tilde{D}_i^{(\delta)} | J_i = 1] = \frac{\mu^2(\delta) + \sigma^2(\delta)}{2\mu(\delta)} = \frac{5}{6}\delta^2 < \delta^2 \quad (1.41)$$

This suggests that each jump will on average shorten the distance travelled on the business clock by $\frac{1}{6}\delta^2$. For a total of $\tilde{X}^{(\delta)}(\tau(T))$ events in which $J(T)$ of them are jump-induced, the expected business time elapse is therefore:

$$IV(0, T) = \tilde{X}^{(\delta)}(\tau(T))\delta^2 - \frac{1}{6}E[J(T)]\delta^2. \quad (1.42)$$

It is then clear that the bias introduced by a jump is just $\frac{1}{6}E[J(T)]\delta^2$, which goes to zero as $\delta \rightarrow 0$ given the jump process is of finite activity. It is interesting to see that the *NPD* estimator is less affected by jumps when $\delta \rightarrow 0$, in contrast to an RV estimator which is not robust to jumps at all regardless of the sampling scheme. As empirically price jumps are found to be very infrequent (on average less than one per week as documented in Andersen, Bollerslev, and Dobrev (2007) and Lee and Hannig (2010)), we can safely conclude that the estimator is robust to jumps in infill setting. Note that following the same notion, the *NPR* estimator is also very robust to jumps in the infill setting.

It is worth noting that in the sprawl asymptotics setting, the presence of jumps biases the *NPD* estimator upwards, and the bias is unbounded if $E[J(t)] \rightarrow \infty$ as $t \rightarrow \infty$.

1.6.3 A More Realistic Model

Real data does not follow the model specified in (1.24), as it possesses various types of market imperfections, including irregularly spaced observations, market microstructure noise, price discretization, etc. It greatly complicates the analysis of the theoretical properties of the *NPD* estimator as the properties of the *RBV*-class estimators may not apply in some cases. In this section we attempt to derive some asymptotic results for the *NPD* estimator under a general setting with random arrival times of observations and a very general structure of market microstructure noise.

Our strategy here is to add features to the pure diffusion model in (1.24). We firstly define the latent efficient log-price process as:

$$P^*(t) = P^*(0) + \int_0^t \sigma_p(t) dW(t). \quad (1.43)$$

To account for the random arrival of observations, we define a sequence of random arrival times of the tick changes³ (or revisions for quote data) $0 = t_0 < t_1 < t_2 \cdots$, and assume that the process $P^*(t)$ is only observed at these random arrival times. The sequence $\{t_j\}$ and the arrival times in business time $\{\tau(t_j)\}$ with $\tau(t) = \int_0^t \sigma_p^2(s) ds$ are natural stopping times.

At each t_j , the observed process $P(t)$ is measured with noise V_j , commonly referred to as the MMS noise:

$$P_j = P_j^* + V_j \quad (1.44)$$

Whenever no confusion is caused, we suppress the notation of $P(t)$ as a function of calendar time and use $P_j = P(t_j)$ to denote the j -th observed price. We build our assumptions of the MMS noise based on the noise assumptions in Zhang (2006), Bandi and Russell (2008) and Ait-Sahalia, Mykland, and Zhang (2011):

Assumption 1.2. The Market Microstructure Noise: *The MMS noise component V_j in (1.44) is assumed to possess the following properties:*

1. V_j is strictly stationary with mean 0.
2. All moments of V_j exist and are finite.
3. V_j is mixing.
4. $V_j \perp P_j^*$.

³Note that the *NPD* estimator will always sample data in tick time, and we only consider the arrival of tick changes as the flat trades are irrelevant in the discussion.

We exclude the case where the noise is correlated with the efficient price movement as argued by Hansen and Lunde (2006). This is a common assumption in the existing literature mentioned above, and to a large extent simplifies our analysis. Note that the MMS noise in Assumption 1.2 does not include price discretization effect caused by a minimum tick size. This will be discussed separately in Section 1.6.5.

The literature suggest that the trade durations in calendar time $d_j = t_j - t_{j-1}$ have seasonality patterns, are very persistent and are correlated with the volatility of the efficient price (e.g. Easley and O'Hara (1992), Chen, Diebold, and Schorfheide (2013)). However, we are more interested in the properties of the trade durations in business time denoted by $\tilde{d}_j = \tau(t_j) - \tau(t_{j-1})$, which are more relevant to our analysis. Since $P^*(\tau(t))$ is a standard Wiener process, by the martingale stopping theorem, P_j^* is a martingale, and the martingale difference sequence (MDS) $r_j^* = P_j^* - P_{j-1}^*$ is mixture normally distributed:

$$r_j^* \sim \mathcal{MN}(0, \tilde{d}_j). \quad (1.45)$$

It is therefore clear that the property of \tilde{d}_j is embedded in the property of the tick returns of the efficient price r_j^* . We make the following assumption on the MDS process r_j^* :

Assumption 1.3. Tick Return of the Efficient Price Process: *The tick return of the efficient price process r_j^* is strongly mixing and strictly stationary with finite moments.*

Assumptions 1.2 and 1.3 will be used to quantify the influence of MMS noise on the bias of the *NPD* estimator.

The *NPD* estimator is constructed by Definition 1.7 and (1.27) on the observed price process P_j with the following form:

$$NPD(0, t) = X^{(\delta)}(t) \delta^2, \quad (1.46)$$

where $X^{(\delta)}(t) = \sum_{i=1}^{\infty} \mathbb{1}_{\{t_i^{(\delta)} \leq t\}}$ and $t_i^{(\delta)}$ is the arrival time of the i -th price event.

Deviations from the continuous martingale setting results in a biased *NPD* estimator, since the mean duration in business time is not δ^2 any more, and the point process in business time ceases to be renewal due to the existence of MMS noise. Fortunately the mixing assumptions for the MMS noise and the tick returns ensure that when δ is large enough, \tilde{d}_j can be regarded as the stopping time from a Wiener process due to the functional central limit theorem via martingale approximation in

e.g. Gordin and Peligrad (2011). In the following section we analyse the bias of the *NPD* estimator in detail based on our assumptions of the price process above, and show that the bias diminishes as δ increases.

1.6.4 Bias of the *NPD* estimator

To derive the bias of the *NPD* estimator in the presence of MMS noise and time discretization, we start from the renewal RV estimator based on $X^{(\delta)}(t)$:

$$RV^{(\delta)}(0,t) = \sum_{i=1}^{X^{(\delta)}(t)} (r_i^{(\delta)})^2 \quad (1.47)$$

Since the *NPD* estimator simply truncates $(r_i^{(\delta)})^2$ to δ^2 , we must have that $(r_i^{(\delta)})^2 \geq \delta^2$. Intuitively, if there is no MMS noise, $RV^{(\delta)}$ would be unbiased, and the difference between the two estimators is the bias of the *NPD* estimator caused by time discretization. We will therefore use $Bias_{TD}^{(\delta)}(0,t) = NPD(0,t) - RV^{(\delta)}(0,t)$ to denote the time discretization bias of the *NPD* estimator, which is always negative.

In the presence of MMS noise, let us decompose $r_i^{(\delta)}$ as:

$$r_i^{(\delta)} = r_i^{(*,\delta)} + V_i^{(\delta)} - V_{i-1}^{(\delta)}, \quad (1.48)$$

in which $r_i^{(*,\delta)} = P^*(t_i^{(\delta)}) - P^*(t_{i-1}^{(\delta)})$ denotes the return of the efficient price. A well-established result (see e.g. Hansen and Lunde (2006), Bandi and Russell (2008)) of the RV estimator under autocorrelated noise is that:

$$E[RV^{(\delta)}(0,t)] = IV(0,t) + \sum_{i=1}^{X^{(\delta)}(t)} E[(V_i^{(\delta)} - V_{i-1}^{(\delta)})^2] \quad (1.49)$$

And therefore:

$$E[NPD(0,t)] = IV(0,t) + E[Bias_{MMS}^{(\delta)}(0,t)] + E[Bias_{TD}^{(\delta)}(0,t)], \quad (1.50)$$

in which $Bias_{MMS}^{(\delta)}(0,t) = \sum_{i=1}^{X^{(\delta)}(t)} (V_i^{(\delta)} - V_{i-1}^{(\delta)})^2$ is the bias induced by the market microstructure noise, which is strictly positive. The above results suggest that the *NPD* estimator is generally biased with two sources of biases: the truncation bias introduced by time discretization and the MMS noise bias.

For the *TD* bias, we are unable to derive an explicit expression in the general case. We show in Appendix A.7 that an approximated version of $Bias_{TD}^{(\delta)}$ converges

to zero in the absence of MMS noise with a rate of δ^{-1} . If we believe that $Bias_{TD}^{(\delta)}$ is of the order δ^{-1} in the general case, then $Bias_{TD}^{(\delta)}$ decays much slower than $Bias_{MMS}^{(\delta)}$ as $Bias_{MMS}^{(\delta)}$ is of the order δ^{-2} . Also, $Bias_{TD}^{(\delta)}$ will always bias the *NPD* estimator towards zero when $\delta \rightarrow 0$, which is due to the fact that $X^\delta(t)$ is capped at the number of observed tick changes. To give a graphical illustration of the bias of the *NPD* estimator under the two sources of biases, we simulate a simple price model and analyse the bias of the *NPD* estimator by adding the features to the price model. The results are present in Appendix A.8.

The discussion above also suggests that, the *NPD* estimator will be less biased compared to the $RV^{(\delta)}$ estimator if $Bias_{MMS}^{(\delta)}$ dominates $Bias_{TD}^{(\delta)}$, but will perform worse than the $RV^{(\delta)}$ estimator if there is no MMS noise at all. Interestingly, we may find cases where $Bias_{MMS}^{(\delta)}$ approximately offsets $Bias_{TD}^{(\delta)}$ when δ is large (as in Figure A.3 in Appendix A.8 for example). In this case the *NPD* estimator will have a bias close to zero and thus very efficient, although the $Bias_{MMS}^{(\delta)}$ is not zero. This suggests a potential bias correction technique if one can ‘adjust’ $Bias_{MMS}^{(\delta)}$ or $Bias_{TD}^{(\delta)}$ in a way that the two biases approximately cancels as δ increases. We will exploit this property in Section 1.6.6 to construct bias corrected *NPD* estimators.

1.6.5 Price Discretization

The observed price in practice is not continuously distributed, due to price discretization. The minimum allowed quote change is known as the tick size, which is typically 1 cent for securities in the US market that are traded above \$1. This is known as the round-off error discussed in the RV literature (see e.g. Delattre and Jacod (1997), Li and Mykland (2015) and the reference therein). We show, that this noise will also have a very special impact on the *NPD* estimator with simulation evidence in Appendix A.8.

We write the discretized return as $\hat{r}_j = h_\epsilon(P_j) - h_\epsilon(P_{j-1})$, where $h_\epsilon(x)$ is a rounding function for the log price P_j . If we compare the discretized return \hat{r}_j and the return r_j without discretization, we have the following expression:

$$\begin{aligned}\hat{r}_j &= r_j + \bar{\mathcal{E}}_j - \bar{\mathcal{E}}_{j-1} \\ \bar{\mathcal{E}}_j &= h_\epsilon(P_j) - P_j,\end{aligned}\tag{1.51}$$

and $\bar{\mathcal{E}}_j$ is thus the price discretization error. Depending on the assumed rounding function $h_\epsilon(x)$, the theoretical property of $\bar{\mathcal{E}}_j$ will differ. Note that the bid-ask bounce effect is already incorporated in the MMS noise component V_j , and the price discretization error $\bar{\mathcal{E}}_j$ merely reflects the effect of rounding.

To simplify our analysis, we use the rounding function: $h_\varepsilon(x) = \varepsilon \text{nint}(\frac{x}{\varepsilon})$ and $\text{nint}(x)$ returns the nearest integer of x . This basically assumes that the price discretization is in the log scale, which is reasonable if the price level is roughly constant within the time period. We can interpret ε as the log tick size. The distribution of Ξ_j is then roughly identical for all j but can be potentially autocorrelated if ε is large, and we also have $\Xi_j \in (-0.5\varepsilon, 0.5\varepsilon)$. Thus we can regard Ξ_j as another noise term in the price process similar to the MMS noise component V_j and incorporate this in $\text{Bias}_{MMS}^{(\delta)}$.

The price discretization has a more profound impact on the sampling scheme. Specifically, when one takes δ to be between $(x\varepsilon, (x+1)\varepsilon]$ for some integer x , the resulting sampling scheme $X^{(\delta)}(t)$ will be exactly the same due to the discreteness in \hat{r}_j . As a result, choosing multiple δ in the range $(x\varepsilon, (x+1)\varepsilon]$ does not effectively change the asymptotic property of $X^{(\delta)}(t)$. An implication of this is that one can influence the level of the truncation bias for a fixed sampling scheme $X^{(\delta)}(t)$ by choosing a δ within the range $((x-1)\varepsilon, x\varepsilon]$ for some integer x . When $\text{Bias}_{TD}^{(\delta)}$ dominates, we should always choose $\delta = x\varepsilon$ to minimize $\text{Bias}_{TD}^{(\delta)}$. When $\text{Bias}_{MMS}^{(\delta)}$ is large, one can choose $\delta \rightarrow (x-1)\varepsilon$ to inflate $\text{Bias}_{TD}^{(\delta)}$ and counterbalance the positive $\text{Bias}_{MMS}^{(\delta)}$. As is shown in Figure A.4 in Appendix A.8, there can be a δ in the range of $((x-1)\varepsilon, x\varepsilon]$ that corrects the bias of the *NPD* estimator completely. However, this requires the knowledge of $\text{Bias}_{TD}^{(\delta)}$ at any δ , which can be very difficult to estimate empirically.

1.6.6 A Possible Bias Correction Method for the *NPD* Estimator

In this section we propose a bias correction method for the *NPD* estimator, and compare the performance of this bias correction method in a simulation study in Section 1.7 against some commonly used calendar time volatility estimators.

Inspired by the pre-averaging estimator in Jacod, Li, Mykland, Podolskij, and Vetter (2009), we propose to smooth the transaction price before constructing the *NPD* estimator. In detail, instead of constructing the *NPD* estimator based on the observed discrete price $h_\varepsilon(P_j)$, we construct a smoothed price process Z_j , and construct the *NPD* estimator based on Z_j instead. We choose a simple exponential smoothing structure for the process Z_j :

$$\begin{aligned} Z_1 &= h_\varepsilon(P_1) \\ Z_j &= (1 - \gamma)Z_{j-1} + \gamma h_\varepsilon(P_j), \quad \gamma \in [0, 1] \end{aligned} \tag{1.52}$$

where γ is a smoothing factor. Clearly when $\gamma = 1$, $Z_j = h_\varepsilon(P_j)$ so the process is not smoothed, and when $\gamma = 0$, $Z_j = h_\varepsilon(P_1)$ for all j . Intuitively, the variation of the noise is diminished by this exponential smoothing to some extent, thus the *NPD* estimator constructed on Z_j is less affected by MMS noise. We will denote the exponentially smoothed *NPD* estimator constructed on Z_j as NPD^z .

The exponentially smoothed price process Z_j , is still contaminated by noise, albeit the magnitude of noise is reduced by the smoothing. Intuitively, the larger the γ , the larger the impact of MMS noise on the NPD^z estimator. Thus the exponential smoothing provides a way to alter the impact of MMS noise on the NPD^z estimator. As is shown in the previous section, if we can choose a γ so that $Bias_{MMS}^{(\delta)} + Bias_{TD}^{(\delta)}$ is approximately zero for some moderate to large δ , we can greatly improve the performance of the *NPD* estimator.

The price smoothing approach has two additional advantages over the original *NPD* estimator: firstly, it is a natural solution to the price discretization, and the sampling frequency will change more smoothly with respect to δ . Secondly, empirical data contains a very large amount of flat trades which will be completely ignored by the price change point process. By smoothing the price process, we can sample the data at every transaction instead of every tick change, which greatly increases the maximum sampling frequency.

Nevertheless, in this chapter we do not provide an analytical solution to choose γ optimally, as the truncation bias is not available in closed form. In practice, we can choose γ by benchmarking the NPD^z estimator on some unbiased volatility estimator and minimize the MSE, as documented in Hautsch and Podolskij (2013). Moreover, the estimator will be less robust to jumps compared to the *NPD* estimator, simply because the exponential smoothing distributes a jump to all previous transactions, which will have a larger chance to be absorbed into a price event. This is however not a significant problem if the jumps are assumed to be large and rare, so that the smoothed jumps still trigger price events as they occur.

1.7 Simulation Study

1.7.1 Simulation Design

We conduct a simulation study to demonstrate the properties of the price duration based volatility estimators (*NPD* and $RV^{(\delta)}$) and compare their performance to existing calendar time methods. We list all volatility estimators considered in this

chapter in Table 1.1.

Table 1.1 List of all volatility estimators considered in the simulation study

Acronym	Description	Type	MMS	Jump
<i>NPD</i>	See Section 1.5	δ	N	Y
<i>RV</i> ^(δ)	Renewal RV	δ	N	N
<i>NPD</i> ^z	See Section 1.6.6	δ	N	Y
RV	Realized Variance	CTS	N	N
RBip	Realized Bipower Variation	CTS	N	Y
RK	Realized Kernel	CTS	Y	N
PRV	Pre-averaged Realized Variance	CTS	Y	N
PBip	Pre-averaged Bipwer Variation	CTS	Y	Y

Note: The column Type shows the type of sampling schemes: δ stands for the δ -associated price change point process sampling and CTS refers to calendar time sampling. The column MMS describes whether the estimator is robust in the presence of MMS noise, and the column Jump shows the robustness to jumps for the volatility estimators.

We consider a one-factor stochastic volatility (1FSV) model⁴ with jumps to simulate the efficient price process, a model commonly used in this literature (see e.g. Huang and Tauchen (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a), etc.). The log-efficient price is specified as:

$$\begin{aligned} dP^*(t) &= \mu dt + \sigma_p(t)dW(t) + dJ(t), \quad \sigma_p(t) = \exp(\beta_0 + \beta_1 v(t))s_p(t) \\ dv(t) &= \alpha v(t)dt + dB(t), \quad \text{corr}(dW(t), dB(t)) = \varphi, \end{aligned} \quad (1.53)$$

in which $J(t) = \sum_{i=0}^{N_J(t)} J_i$ is a pure jump process. We assume that $N_J(t)$ is a homogeneous Poisson process with rate λ_J , and J_i is i.i.d. normal with zero mean and variance σ_J^2 . Note that we augment the original 1FSV model by a time deterministic function $s_p(t)$ to accommodate the well-documented L or U-shaped pattern of intra-day volatility. In the simulation study we set $t \in [0, 1]$ to represent fractions of time from a trading day from 9:30 to 16:00, and the process $\tau(t)$ is initialized by a random draw from its unconditional distribution. The function $s_p(t)$ in our simulation study is specified as:

$$s_p(t) = \frac{1}{a_1 t + a_2} - \frac{\ln(\frac{a_1}{a_2} + 1)}{a_1} + 1, \quad a_1 > 0, a_2 > 0. \quad (1.54)$$

This function has the property that $\int_0^1 s(t)dt = 1$. When a_1 and a_2 are properly chosen, the function can produce a L-shaped pattern.

⁴We will use the subscript 1FSV_J to denote a 1FSV model with jump.

We build the MMS noise component in transaction time instead of calendar time. Specifically, we assume that the point process of transaction arrivals (or quote updates), denoted by $N(t)$, follows an inhomogeneous Poisson process where the intensity function $\lambda(t)$ is specified as a time-deterministic function to mimic the empirical U-shaped pattern of transaction arrivals. We specify the intensity function as:

$$\lambda(t) = \frac{1}{\Delta t} \left(b_1(b_2 t - 1)^{b_3} - \frac{b_1(b_2 - 1)^{b_3+1} + 1}{b_2(b_3 + 1)} + \lambda_0 \right), \quad (1.55)$$

in which $b_1 > 0$, $b_2 > 0$ and $b_3 = 2, 4, \dots$, λ_0 is the baseline arrival rate of transactions. The quantity Δt is the discretization step size of the simulation. The expected number of transactions in the interval $(0, 1)$ is therefore $E[N(t)] = \int_0^1 \lambda(t) dt = \frac{\lambda_0}{\Delta t}$.

Let t_j denote the j -th arrival of transaction, and $P_j^* = P^*(t_j)$ denote the efficient price at the j -th transaction. Empirically we cannot observe P_j^* due to the presence of MMS noise, and the following decomposition is frequently used in the literature:

$$P_j = P_j^* + V_j, \quad (1.56)$$

in which V_j is a MMS noise term satisfying Assumption 1.2, and P_j is the log-price process measured with error. We assume that V_j follows a Gaussian AR(1) process specified as follows:

$$V_j = \rho V_{j-1} + v_j, \quad v_j \sim \mathcal{N}(0, (1 - \rho^2) \sigma_v^2) \quad (1.57)$$

For the sake of stationarity we require that $|\rho| < 1$. The unconditional variance of the noise is therefore $V[V_j] = \sigma_v^2$.

Empirically, the transaction returns contain a large amount of flat trades where the transaction price do not move at all. For example, in Liesenfeld, Nolte, and Pohlmeier (2006), the proportion of flat trades for two stocks traded in NYSE is over 60%. Jacod, Li, and Zheng (2017) reports an over 70% of flat trades in the transaction data from Citigroup. For the mid-quote data the proportion of flat price changes will be even larger, as the best quotes can remain constant even when the transaction price moves. As a result, the empirical transaction returns are typically found to have excess kurtosis due to the amount of flat trades that cannot be reasonably explained by the normal assumption. To account for this effect, we follow the approach of Griffin and Oomen (2008) and assume that the tick change of price process is governed by a first order Markov chain. Let S_j be a stationary and recurrent two-state first order Markov chain with transitional parameters $P(S_{j+1} = n | S_j = m) = p_{mn}$ where

$m, n \in \{0, 1\}$. We rewrite (1.56) as:

$$P_j = \begin{cases} P_j^* + V_j, & S_j = 1 \\ P_{j-1}, & S_j = 0 \end{cases} \quad (1.58)$$

Therefore, when $S_j = 1$, the observed price change is updated by the rounded efficient price process plus noise, and remains constant whenever $S_j = 0$.

The observed log-price process, $h_\varepsilon(P_j)$ is specified as follows:

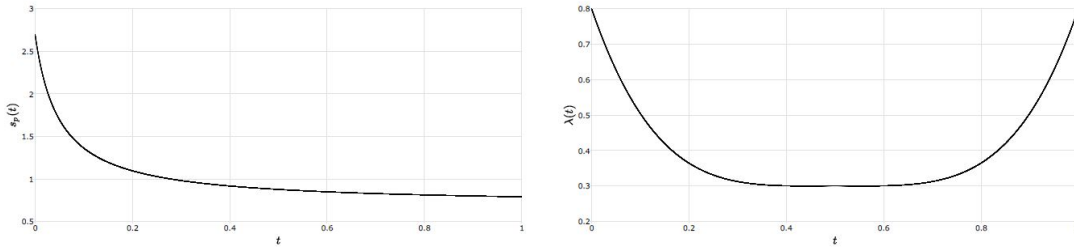
$$h_\varepsilon(P_j) = \left(\varepsilon \text{nint}\left(\frac{P_j}{\varepsilon}\right) \right), \quad (1.59)$$

in which $\text{nint}(x)$ returns the nearest integer of a real number x . Note that the rounding will introduce additional flat trades to the observed price process when the price change is rounded to zero. We set $\varepsilon = \ln(P_0 + 0.01) - \ln(P_0)$ to represent the log tick size.

The parameters for the 1FSV model we use are: $\mu = 0$, $\beta_0 = -4.3711$, $\beta_1 = 0.05934$, $\alpha = -0.011$, $\varphi = -0.3$, $\lambda_J = 2$, $a_1 = 10$, $a_2 = 0.5$, and $\sigma_J = 0.01$. The unconditional mean of the annualized daily volatility is roughly 27%, and the expected jump variation is about 0.0002 per day. The transaction and tick arrival parameters are set as $b_1 = 0.5$, $b_2 = 2$, $b_3 = 4$, $\lambda_0 = 0.4$, $p_{11} = 0.6$ and $p_{22} = 0.8$. We set the Euler step size of the simulation to be $\Delta t = \frac{1}{23400}$, so that the expected number of transactions within a trading day is 9360. The diurnal patterns of intraday volatility and the arrivals of transactions are plotted in Figure 1.2. From the figure we can clearly see that the intraday volatility has a L-shaped pattern and the arrivals of observations possess a U-shaped pattern. An example of a simulate price path of the 1FSV_J is presented in Figure 1.4.

For the MMS noise parameters, we set $\rho = -0.5$ and $\sigma_v^2 = \omega IV$, where ω is the noise-to-signal ratio. Empirically ω is found to be quite small (typically smaller than 0.1% as documented in Hansen and Lunde (2006)). We therefore choose $\omega = 0.005, 0.001$ and 0.0002 to represent high, moderate and low noise scenarios. The resulting $\sigma_v^2 \approx 0.00115^2, 0.00052^2$ and 0.00023^2 respectively. The expected number of flat trades implied by the Markov chain is about 67% of the total transactions. The actual amount of flat trades in $h_\varepsilon(P_j)$ depends on the initial price $P(0)$, as the rounding error is smaller when $P(0)$ is large, and vice versa. We set $P(0) = 20$, and the resulting proportion of flat trades is approximately 70%. We plot a histogram

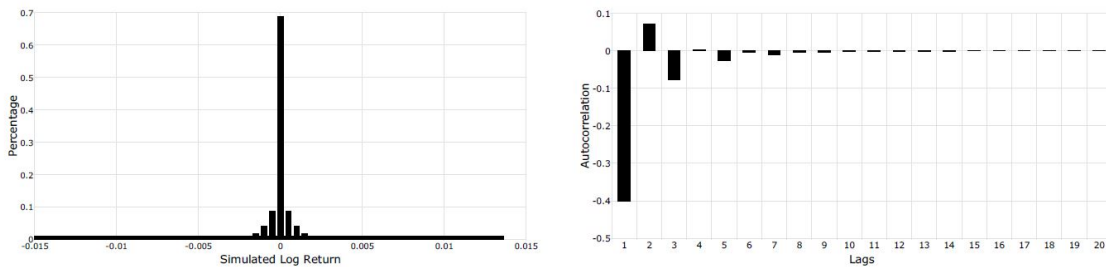
Figure 1.2 Simulated diurnal pattern of intraday volatility and transaction arrival rate



Note: $s_p(t)$ is specified in (1.54) with $a_1 = 10, a_2 = 0.5$. $\lambda(t)$ is specified in (1.55) with $b_1 = 0.5, b_2 = 2, b_3 = 4, \lambda_0 = 0.4$ and $\Delta t = 1$. t is the fraction of time in a trading day.

and the correlogram for the simulated price change $h_\epsilon(P_j) - h_\epsilon(P_{j-1})$ in Figure 1.3 with $J(t) = 0$ and $\Delta t = \frac{1}{23400}$ for the moderate noise case. It is clear that the observed price change is leptokurtic with a sample kurtosis of approximately 15. This closely resembles the empirical density of the price changes as in Liesenfeld, Nolte, and Pohlmeier (2006). The autocorrelation for price changes suggests that the price changes follow an ARMA-type process with negative first order autocorrelation, which is consistent with the findings in e.g. Oomen (2006).

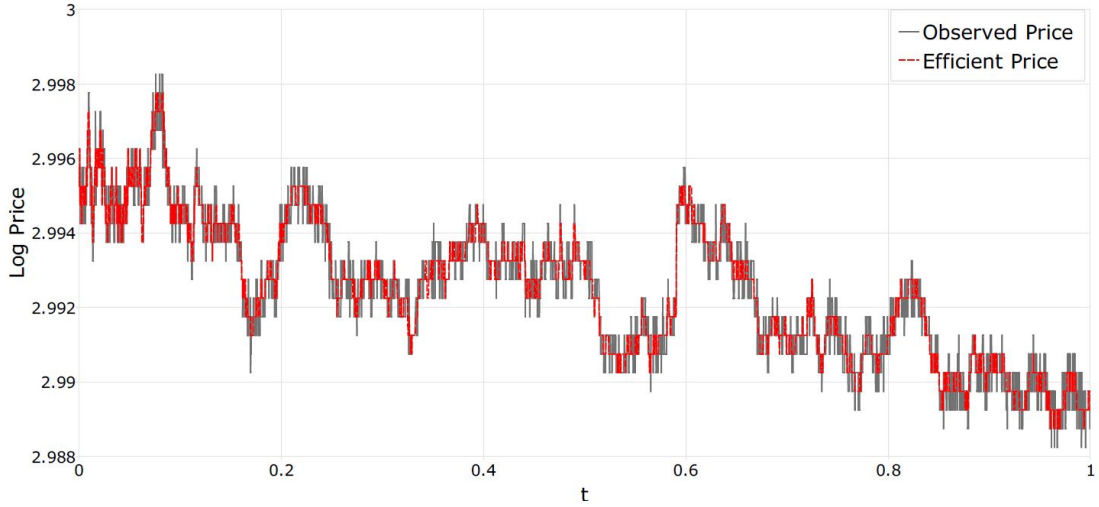
Figure 1.3 Histogram and correlogram for the simulated price change with moderate level of MMS noise and no jumps



Note: The histogram is on the left side and the correlogram is on the right.. The histogram is constructed based on simulated price changes for 10000 days without jumps. The correlogram is constructed based on the average autocorrelation for 10000 days without jumps. $\Delta t = \frac{1}{23400}$ in the simulation. The noise-to-signal ratio is set to be $\omega = 0.001$.

We use the bias, the mean squared error (MSE) and the QLIKE measure to compare the performance among estimators. For the true integrated variance $IV(0,t)$ and an

Figure 1.4 An example of simulated price path of the IFSV_J model with moderate level of noise



Note: t is the fraction of time in a trading day. $\Delta t = \frac{1}{23400}$ in the simulation. The noise-to-signal ratio is set to be $\omega = 0.001$.

estimate of IV denoted by $\widehat{IV}(0,t)$, the three measures are defined as follows:

$$Bias(\widehat{IV}(0,t)) = E[\widehat{IV}(0,t) - IV(0,t)], \quad (1.60)$$

$$MSE(\widehat{IV}(0,t)) = E[(\widehat{IV}(0,t) - IV(0,t))^2], \quad (1.61)$$

$$QLIKE(\widehat{IV}(0,t)) = E\left[\frac{\widehat{IV}(0,t)}{IV(0,t)} - \ln \frac{\widehat{IV}(0,t)}{IV(0,t)} - 1\right]. \quad (1.62)$$

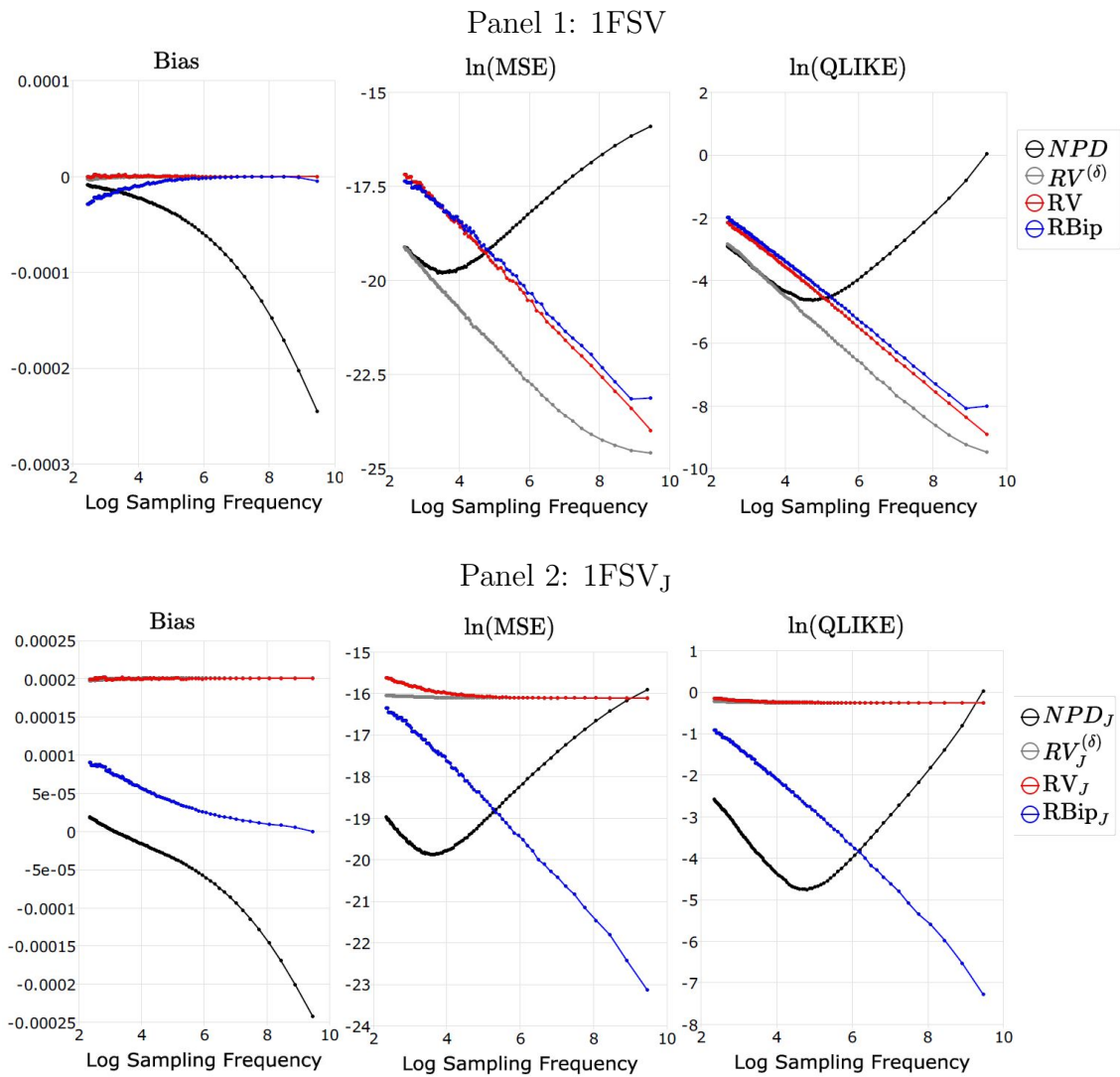
1.7.2 1FSV Model Without Noise and Price Discretization

Firstly, we would like to show that price duration based volatility estimators, namely NPD and $RV^{(\delta)}$ are indeed superior to calendar time RV and realized bipower (RBip) estimators when we can observe $P^*(t)$ in continuous time without noise or price discretization.⁵ We simulate 10000 replications of $P^*(t)$ for $t \in (0,1)$ with and without jump. We construct the NPD and $RV^{(\delta)}$ estimators on a grid of δ s with $\delta = x\delta_0$, in which $\delta_0 = 0.1\epsilon$ and $x \in \mathbb{Z}^+$. The calendar time sampled (CTS) RV and RBip estimators are constructed based on the average sampling frequency of the NPD estimator for each x , so that CTS estimators will have a fixed sampling frequency that is comparable to that of the NPD estimator.⁶ The Bias, MSE and QLIKE of the four estimators are plotted in Figure 1.5 against the log sampling frequency.

⁵Technically, when $P^*(t)$ is observed in continuous time, NPD and $RV^{(\delta)}$ coincide in the absence of jump. Due to that we use an Euler method to approximate the continuous time $P^*(t)$, NPD will be different from $RV^{(\delta)}$ even in the absence of jump as a result of time discretization.

⁶Note that it is not always possible to construct a NPD estimator from a sample if the maximum range of the price is smaller than the threshold. A similar issue arises when constructing kernel and

Figure 1.5 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD , $RV^{(\delta)}$, RV and $RBip$ for 1FSV model without noise and price discretization



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and $RBip$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The Euler discretization step $\Delta t = \frac{1}{23400}$.

From the plots on the first column in Figure 1.5, we can see a strong negative bias for the *NPD* estimator at the maximum frequency when δ is small due to time discretization in the simulation. In theory the *NPD* estimator should converge to the integrated variance as δ decreases, but in simulation whenever we use a discrete approximation to the continuous efficient price process, the truncation bias will affect the performance of the *NPD* estimator when δ is small. Since $RV^{(\delta)}$ is unaffected by this truncation, it converges to the theoretical quadratic variation as δ decreases.

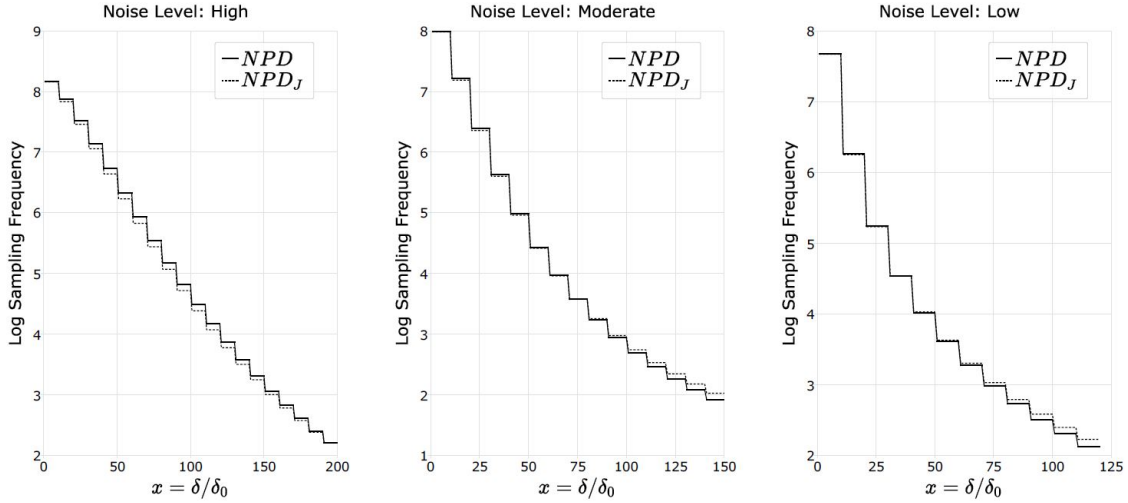
Comparing the efficiency of $RV^{(\delta)}$ with RV and RBip in the absence of jump, we can see clearly that $RV^{(\delta)}$ is indeed superior to RV and RBip at any sampling frequency considered in this simulation, as discussed in Section A.3.1. *NPD* and $RV^{(\delta)}$ have similar efficiency when δ is large, but the performance of *NPD* deteriorates as δ shrinks and the truncation bias becomes larger. However, even in the presence of truncation bias the *NPD* estimator is still more efficient than CTS estimators for sampling frequencies less than roughly 140 per day. When the jumps are present, $RV^{(\delta)}$ and RV are not robust to jumps and their efficiency drops sharply. We also see that the *NPD* estimator is more robust to jumps compared to RBip estimator as the jump variation for the *NPD* estimator is of a much smaller magnitude. Consequently, the efficiency advantage of the *NPD* estimator is even larger in the presence of jumps.

1.7.3 Full 1FSV Model: Primal Volatility Estimators

We proceed to add irregular transaction arrivals, price discretization and MMS noise to the 1FSV model, and compare the performance of price duration based volatility estimators *NPD* and $RV^{(\delta)}$ to the calendar time estimators RV and RBip. Note that these estimators are all ‘primal’ estimators without any correction for MMS bias. The average (log) sampling frequency for the *NPD* estimator is presented in Figure 1.6 for the 1FSV and 1FSV_J model.

From Figure 1.6 we see that the sampling frequency of *NPD* estimators always decreases at multiples of $10\delta_0 = \varepsilon$ due to price discretization. The sampling frequency ranges from roughly 3000 ($\exp(8)$) which is the average number of tick returns per day, to roughly 7 ($\exp(2)$) for all three levels of noise. The presence of jumps does not have a large impact on the average sampling frequency for small δ s as expected, and will increase the sampling frequency slightly when δ is large. Similar to the previous case, we use the average sampling frequency of the *NPD* estimator to construct

pre-averaging estimators as they are not guaranteed to be positive. The computation of Bias, MSE and QLIKE is only based on valid volatility estimates and ignores all invalid volatility estimates.

Figure 1.6 Average sampling frequency of the NPD estimator for the 1FSV and 1FSV $_J$ models

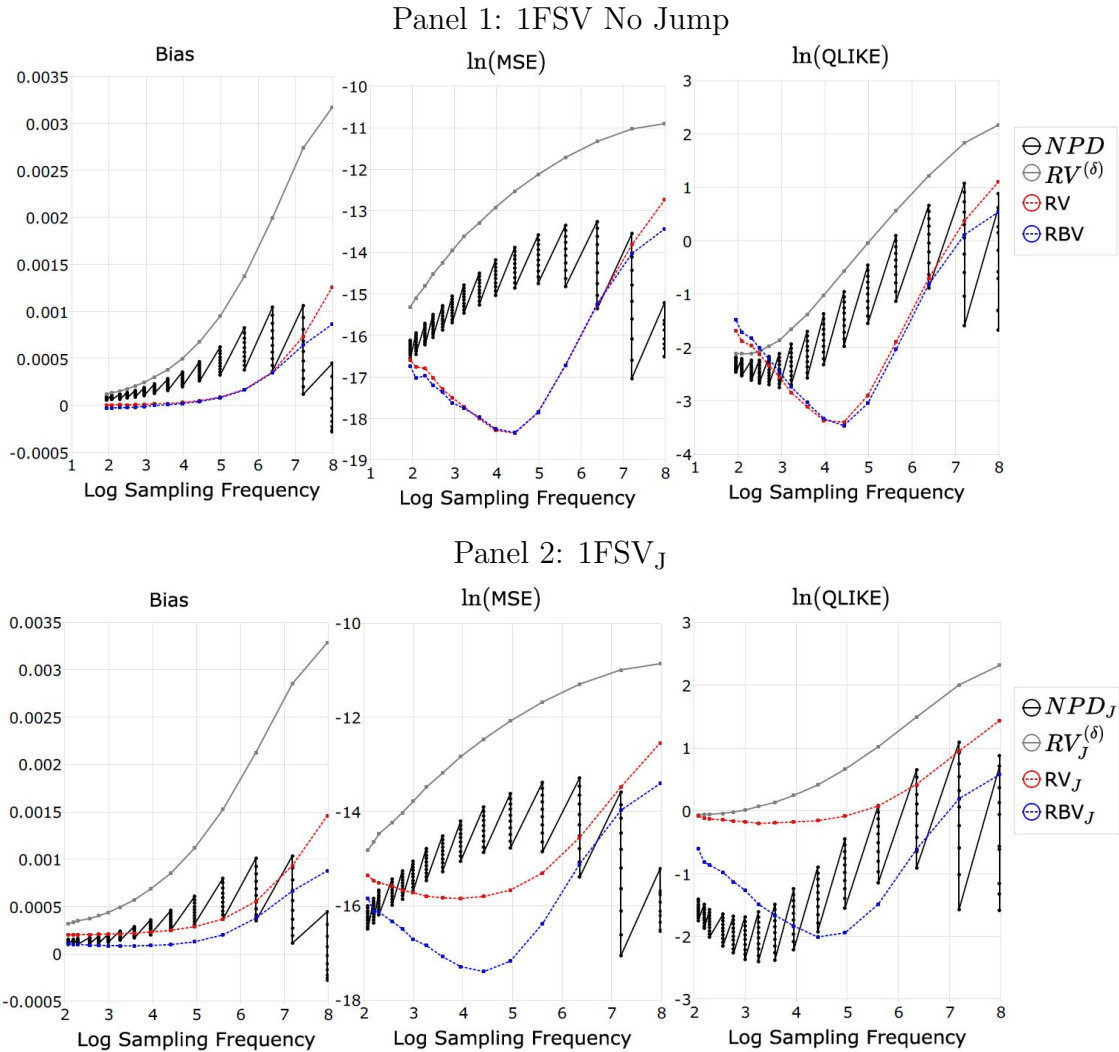
Note: For the high, moderate and low levels of noise, the δ ranges from δ_0 to $200\delta_0$, $150\delta_0$ and $120\delta_0$ correspondingly. The step size is set to be $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. For each δ , we compute the average sampling frequency by averaging the number of price durations over 10000 Monte Carlo draws of 1FSV and 1FSV $_J$ model. The noise-to-signal ratios for the high, moderate and low levels of noise are $\omega = 0.005$, 0.001 and 0.0002 respectively.

the calendar time RV and RBip estimators for each δ . The performance of these estimators under moderate noise can be viewed in Figure 1.7, and results for the high and low levels of noise cases can be found in Figure A.7 and A.8 in Appendix A.11.

From Figure 1.7 we can observe that, due to the price discretization, for $\delta \in ((x-1)\epsilon, x\epsilon]$ the sampling scheme does not change. As a result, there will be multiple volatility estimates from the NPD for a given sampling frequency as δ changes within the range $((x-1)\epsilon, x\epsilon]$. It is clear that the $RV^{(\delta)}$ is the worst estimator among all 4 estimators that has a significantly larger bias and is not robust to jumps at all. Although NPD performs better than $RV^{(\delta)}$, the efficiency advantage of NPD over RV and RBip is greatly weakened by the MMS noise bias as calendar time estimators outperforms the NPD estimator for a very large range of δ . For RV and RBip, we see that the optimal sampling frequency is around $exp(4.4)$, which corresponds to a sampling frequency of 84 per day. It is evident that RBip has the overall best performance when sampled optimally due to its smallest MSE and QLIKE and its robustness to jumps. Note that the optimal sampling frequency is close to the theoretical optimal sampling frequency as proposed by Bandi and Russell (2008): $(2\omega)^{-2/3} \approx 63$.

The inferior performance of price duration based estimators to the calendar time estimators is due to that the price duration returns have a much more pronounced

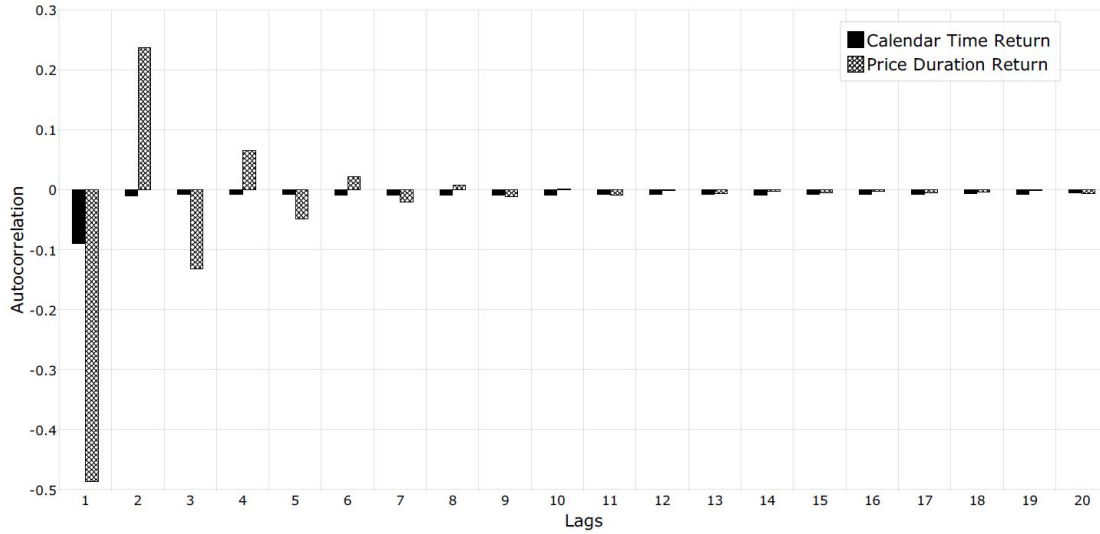
Figure 1.7 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD , $RV^{(\delta)}$, RV and $RBip$ for 1FSV model with moderate level of MMS noise



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and $RBip$. The truncation threshold δ ranges from $\delta_1/50$ to δ_0 with a step size of $\delta_0 = 0.1\varepsilon$, with $\varepsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.001$.

autocorrelation structure than the calendar time returns with the same sampling frequency. We plot the average correlogram for the calendar time returns and price duration returns sampled at RV 's optimal sampling frequency in Figure 1.8.

Figure 1.8 Average correlogram of calendar time returns and price duration returns



Note: The results are based on averaging the first 20 autocorrelations of calendar time and price duration returns from 10000 replications of the 1FSV model with moderate noise. The sampling frequency for the calendar time return is 84 per day. The corresponding threshold of price duration is $\delta = 51\delta_0$.

Figure 1.8 shows a MA(1) dependence structure for the calendar time returns, and an ARMA-type dependence structure for the price duration returns that clearly has a higher magnitude. This suggests that the MMS noise under calendar time sampling can be regarded as i.i.d. when we sample sparsely, thus the calendar time estimators are much less affected by the MMS noise. For the renewal based estimators, we see that the $RV^{(\delta)}$ performs the worst due to the dependence in the noise structure, and NPD performs better simply because the truncation bias mitigates part of the MMS noise bias. More importantly, the performance of NPD is more sensitive to the size of MMS noise than calendar time methods when the sampling frequency is on the same level. The sensitivity to the size of noise for the NPD estimator can also be seen from Figure A.7 and A.8 in Appendix A.11. In the low noise case NPD performs significantly better than the calendar time methods with smaller MSE and QLIKE if the sampling frequency is smaller than 84, similar to the no noise case. This advantage quickly diminishes as the size of the MMS noise increases, and in the large noise case the performance of NPD is completely dominated by the CTS methods for any sampling frequency smaller than 1000.

Interestingly, when size of the noise is large, one may choose a very small δ in

such a way that the truncation bias exactly offsets the MMS noise bias, which explains why the *NPD* estimator has better performance when δ is small. However, even if we can reliably choose such a δ , the performance of this *NPD* estimator is still inferior to an optimally sampled CTS estimator. Moreover, it is difficult to choose a δ that can maximize MSE or QLIKE for a $\delta \in ((x-1)\varepsilon, x\varepsilon]$. If the goal is to choose an estimator that has a smaller MSE or QLIKE, then for the *NPD* estimator one needs to choose a large δ that are less affected by the truncation bias, and hopes that the MMS bias does not outweigh the smaller asymptotic variance of renewal sampling. As a result, CTS primal estimators are preferred over the *NPD* estimator due to that the optimal sampling frequency already has closed form approximations (See e.g. Bandi and Russell (2008) and Hansen and Lunde (2006)) and their optimal performance dominate the *NPD* estimator in the presence of moderate to high level of MMS noise.

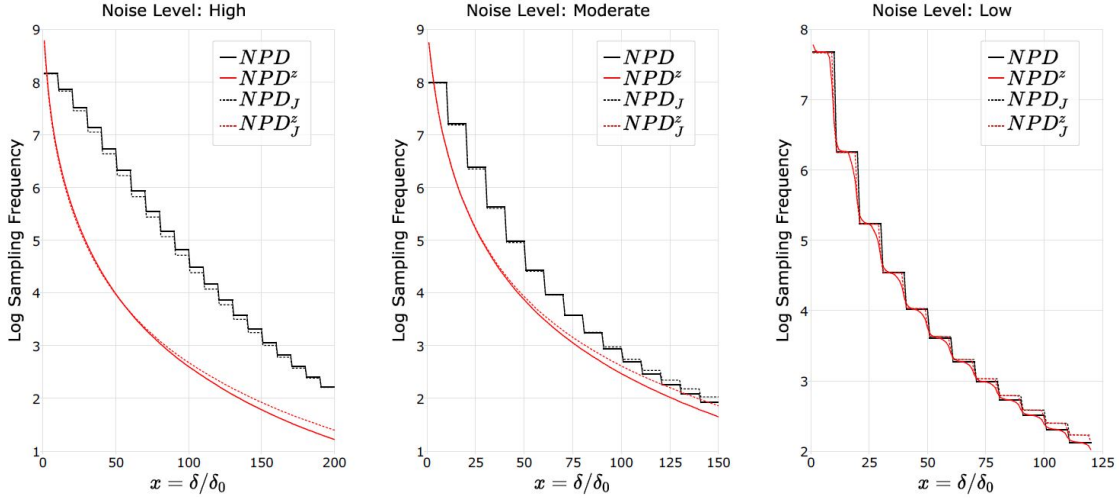
1.7.4 Full 1FSV Model: Bias Corrected Estimators

The discussion above suggests that, to fully exploit the smaller asymptotic variance of the price duration based estimators, it is necessary to mitigate impact of the MMS noise bias for the *NPD* estimator. To this end, we compare the performance of the exponentially smoothed *NPD*^z estimator to calendar time bias corrected methods, namely RK, PRV and PBip estimators, which are state-of-the-art calendar time volatility estimators that are known to be highly efficient and robust to MMS noise (also robust to jumps for PBip). Similar to the previous comparison, we compare the *NPD*^z estimator to the calendar time rivals with the same average sampling frequency.

The choice of tuning parameters for these estimators are non-trivial, as they have a very large impact on the performances of these estimators. Our aim here is to compare the optimal performance of all these estimators, therefore we will use optimized tuning parameters assuming they are known in advance. For the RK estimator the optimal choice of the bandwidth is provided in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a), but there is no analytical solution to the optimal tuning parameters for *NPD*^z, PRV and PBip estimators. We therefore choose the tuning parameters for *NPD*^z, PRV and PBip by a grid search method that minimizes the simulated MSE of the estimators. Details of tuning parameter selection and implementation of all estimators considered is presented in Appendix A.10.

Figure 1.9 shows the average sampling frequency of the *NPD*^z estimator under optimal γ compared to that of the *NPD* estimator. It is clear that as the level of noise increases, the sampling frequency of the *NPD*^z estimator deviates from that

Figure 1.9 Average sampling frequency of the NPD and NPD^z estimator for the 1FSV and 1FSV_J models under optimal γ

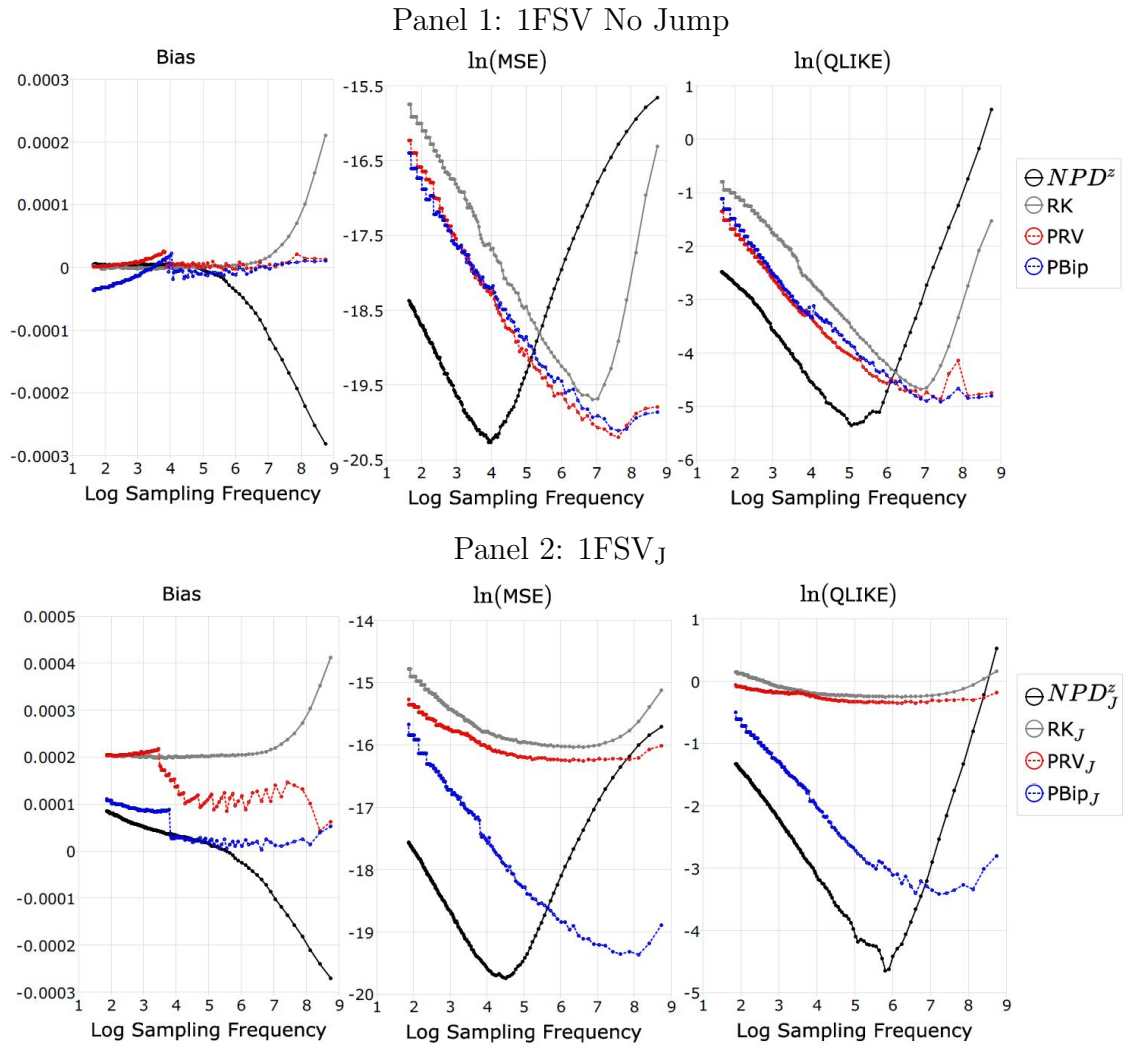


Note: For the high, moderate and low levels of noise, the δ ranges from δ_0 to $200\delta_0$, $150\delta_0$ and $120\delta_0$ correspondingly. The step size is set to be $\delta_0 = 0.1\varepsilon$, with $\varepsilon = \ln(20.01) - \ln(20)$. For each δ , we compute the average sampling frequency by averaging the number of price durations over 10000 Monte Carlo draws of 1FSV and 1FSV_J model. The noise-to-signal ratio is set to be $\omega = 0.001$. See Appendix A.10 for the values of the tuning parameter γ under different levels of noise.

of the NPD estimator. As the impact of noise is alleviated by the smoothing, it is expected that the sampling frequency for the NPD^z estimator is smaller than that of the NPD estimator to reduce the positive MMS bias. It is also interesting to see that the sampling frequency of NPD^z can exceed the average number of ticks in a day as smoothing removes all the flat trades. The sampling frequency is also a smoother function of δ due to exponential smoothing. Finally note that in the low level of noise case, we can still observe a step-shaped sampling frequency curve for the NPD^z , as the optimal γ s are very close to 1. This suggests that smoothing does not improve the MSE of the NPD estimator in the low level of noise case, similar to the optimal θ s for the pre-averaged estimators for sparsely sampled returns.

We plot the Bias, MSE and QLIKE of NPD^z , RK, PRV and PBip for the moderate level of noise case in Figure 1.10, and the other two cases can be viewed in Figure A.9 and A.10 in Appendix A.11. In Figure 1.10, we see that the bias correction leads to a significant improvement in the MSE and QLIKE of volatility estimates compared to the primal estimators, especially at larger sampling frequencies. The RK estimator performs worse than the pre-averaged estimators because the window length H is optimized for convergence rate instead of MSE, thus the MMS noise bias is not fully corrected for at high sampling frequencies. For PRV and PBip, we see that these two estimators are very robust to MMS noise in the absence of

Figure 1.10 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^ε , RK, PRV and PBip for 1FSV model with moderate level of MMS noise



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for the NPD^ε model, or the log sampling frequency of the equidistant intraday return per day for RK, PRV and PBip. The truncation threshold δ ranges from $\delta_1 50$ to δ_0 with a step size of $\delta_0 = 0.1\varepsilon$, with $\varepsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.001$.

jump. The performance of PRV is however affected by jumps as θ^* is optimized to minimize the MSE which uses the actual IV instead of QV. As a result, optimal PRV_J underestimates the QV so that it is less biased.

Comparing the MSE and QLIKE for the four estimators in Figure 1.10 we can see that, the NPD^z estimator has a clear advantage of efficiency at any sampling frequency smaller than approximately $\exp(5) \approx 150$. The NPD^z is biased towards zero when the sampling frequency is large due to the truncation bias. Similar to a NPD estimator, the truncation bias diminishes as δ increases. As is discussed in Section 1.6.4, the optimal γ shrinks the MMS bias in a way that it approximately offsets the truncation bias when one samples relatively sparsely. As a result, the smaller asymptotic variance of the RBV -class estimators leads to a more efficient NPD^z estimator compared to its calendar time rivals for a moderate to small sampling frequency. Interestingly, the smoothed price process Z_j itself is not noise free, so constructing RV-type estimators based on Z_j is still inferior to the NPD^z estimator.

We provide a comprehensive comparison of the optimal MSEs of all volatility estimators considered under various model settings in Table 1.2. A similar comparison of optimal QLIKEs can be found in Table A.2 in Appendix A.11. From Tables 1.2 and A.2, we see that despite a much smaller optimal sampling frequency of the NPD^z estimator compared to the pre-averaged estimators, its optimal MSE and QLIKE still outperform those of the pre-averaged estimators. Moreover, the exponential smoothing to some extent preserves the robustness to jumps of the NPD estimator, and it is evident that the efficiency advantage of the NPD^z estimator over the calendar time competitors is more pronounced in the presence of jumps.

From the discussion above we can conclude that NPD^z has the overall best MSE and QLIKE which is also very robust to jumps. Its performance is closely followed by the pre-averaging estimators PRV and PBip in the absence of noise. It is suboptimal to use a very high sampling frequency for the NPD^z estimator due to the truncation error, but the NPD^z estimator under a sparse sampling frequency can still beat the pre-averaged estimators that uses much more observations in terms of efficiency. Also note that the optimal MSE and QLIKE for the NPD^z is even lower than the optimal MSE and QLIKE of NPD in the absence of noise. This is because the smoothed MMS noise bias serves as a bias correction to the truncation bias, which reduces the bias of the NPD estimator without greatly affecting its variance.

Table 1.2 Comparison of the optimal MSEs for all volatility estimators in Table 1.1 for the 1FSV and 1FSV_J models with low, moderate and high levels of noise

Estimator	<i>NPD</i>	<i>RV</i> ^(δ)	<i>RV</i>	<i>RBip</i>	<i>NPD</i> ^c	<i>RK</i>	<i>PRV</i>	<i>PBip</i>
1FSV model with low level of noise								
Optimal log MSE	-20.9544	-18.1988	-19.3120	-19.3286	-20.9757	-20.2790	-20.7407	-20.8603
$\delta/\hat{\delta}_0$	34	101	21	21	35	6	5	1
Sampling Freq.	93	10	189	189	93	2160	2160	2160
1FSV model with moderate level of noise								
Optimal log MSE	-17.0324	-15.3145	-18.3553	-18.3411	-20.2637	-19.6876	-20.1954	-20.1100
$\delta/\hat{\delta}_0$	11	141	51	51	47	9	5	5
Sampling Freq.	1356	7	84	84	146	2955	2955	2955
1FSV model with high level of noise								
Optimal log MSE	-16.5537	-11.6995	-17.2449	-17.3405	-20.1142	-18.6052	-19.4332	-19.4418
$\delta/\hat{\delta}_0$	6	191	151	151	49	12	7	6
Sampling Freq.	3529	9	21	21	841	2618	3529	3529
1FSV _J model with low level of noise								
Optimal log MSE	-21.0187	-15.8231	-15.9375	-18.3767	-21.0195	-16.0064	-16.2491	-20.0609
$\delta/\hat{\delta}_0$	34	111	31	21	34	10	17	8
Sampling Freq.	93	9	93	187	93	2142	517	2142
1FSV _J model with moderate level of noise								
Optimal log MSE	-17.0643	-14.8169	-15.8456	-17.3982	-19.7464	-15.9891	-16.2552	-19.3642
$\delta/\hat{\delta}_0$	11	141	61	51	37	11	11	3
Sampling Freq.	1322	8	52	83	272	1322	1322	2929
1FSV _J model with high level of noise								
Optimal log MSE	-16.5834	-11.7460	-15.5769	-16.4333	-19.6157	-15.8562	-16.1501	-18.5307
$\delta/\hat{\delta}_0$	6	191	151	151	39	19	7	8
Sampling Freq.	3491	9	20	20	1159	2524	3491	3491

Note: Optimal log MSE for an estimator is the smallest log MSE among all the sampling frequencies considered. The smallest value is highlighted in bold. The entries for the rows $\delta/\hat{\delta}_0$ represents the value of the threshold as multiples of $\hat{\delta}_0 = 0.1\varepsilon$, with $\varepsilon = \ln(20.01) - \ln(20)$. The sampling frequency is the average sampling frequency at the optimal δ s for *NPD*, *RV*^(δ) and *NPD*^c, and is the calendar time sampling frequency for *RV*, *RBip*, *RK*, *PRV* and *PBip*.

1.8 Concluding Remarks

This chapter proposes the class of renewal based volatility estimator for high frequency volatility estimation, and develops its asymptotic theory of the estimator based on renewal theory. The renewal based volatility estimator differs from RV-type estimators as it does not require an equidistant deterministic sampling grid and does not rely on computing squared returns. Our theory opens up a wide range of possibilities to construct alternative volatility estimators such as range duration-based *RBV*-type estimators with more efficiency compared to RV-type estimators, while providing consistency and asymptotic distribution for the entire class of renewal based volatility estimators. Moreover, the stochastic sampling duration in calendar time is allowed to be parametrized, which can potentially lead to a significant efficiency gain compared to non-parametric renewal based volatility estimators.

Using the theory of renewal based volatility estimators, we prove theoretically the consistency and provide the asymptotic distribution for the point process based volatility estimator as in Engle and Russell (1998), Gerhard and Hautsch (2002), Tse and Yang (2012), Nolte, Taylor, and Zhao (2018) and Li, Nolte, and Nolte (2018b) under a continuous martingale setting. We examine Nolte, Taylor, and Zhao's (2018) *NPD* estimator in detail, showing its robustness to drifts and jumps, and establishing its bias structure under MMS noise, time discretization and price discretization. In our simulation study we show that: (1) it is suboptimal to choose a very small δ due to truncation bias. (2) When the MMS noise level is small, the *NPD* estimator is more efficient than the calendar time estimators. (3) The *NPD* estimator in general is more robust to jumps than the RBip estimator. (4) The *NPD* estimator is much more sensitive to the level of noise compared to the calendar time methods. (5) Exponentially smoothing the contaminated price process can yield an approximately unbiased NPD^z estimator that provides high efficiency compared to optimized RK and pre-averaged estimators while preserving the robustness to jumps.

This chapter has several limitations that provide rooms for future research. Firstly, the idea of a range duration-based volatility estimator can be further developed as it is showing some very promising properties under the pure diffusion assumption. Different from the realized range estimator proposed by Christensen and Podolskij (2007), the normalizing coefficient for the *NPR* estimator is just 0.5, and the asymptotic properties follow directly from our theory. However, the properties of this estimator under various noise structures are yet to be verified, but it is promising that its properties can be analysed following the same approach for the *NPD* estimator presented in this paper. Secondly, the properties of the *PRBV* estimator require

further analysis, as we assume that the renewal reward process R_i is known. Therefore it is also helpful to examine the impact of estimation noise of R_i on the efficiency of the *PRBV* estimator. Finally, theoretical properties of the *NPD*^z estimator and a data-driven method to select the optimal smoothing parameter γ are also worth separate investigation.

Chapter 2

High-Frequency Volatility Estimation and the Relative Importance of Market Microstructure Variables

2.1 Introduction

Volatility is an important topic in financial econometrics and a crucial input for any asset pricing, portfolio allocation and risk management framework. It is considered as a latent process that describes the variability of the return process over a given interval, and thus requires estimation from the observed price process. The availability of high-frequency financial data has led to a shift of volatility estimators from low frequency volatility models such as GARCH-type models (Bollerslev, 1986; Engle, 1982) to high-frequency volatility measures, with the realized volatility (RV)-type estimators (Andersen, Bollerslev, Diebold, and Labys, 2001) being the most representative and widely applied high-frequency volatility measures. The RV-type volatility estimators have the advantages of well-established mathematical properties and can be modified to provide precise volatility measures that are robust to various market frictions. However, this type of estimators suffer from the problem of heavy reliance on the availability of data and its non-parametric design also restricts the user to only use price information for volatility estimation.

Restricting ourselves to only price data results in large information loss for intraday volatility estimation, due to the fact that information about volatility contained in observable market microstructure (MMS) variables is completely ignored. To overcome this problem, we apply the duration-based volatility estimator initially

proposed by Engle and Russell (1998) and subsequently developed by Tse and Yang (2012), Nolte, Taylor, and Zhao (2018) and Li, Nolte, and Nolte (2018a). This estimator provides a flexible parametric structure to incorporate other explanatory variables and also produces precise volatility estimates on both daily and intraday levels.

In this chapter, we consider the following MMS variables: trading volume, bid-ask spread, total quote depth, quote depth difference, number of trades, order flow and order imbalance¹. By coupling a variant of the Autoregressive Conditional Duration (ACD) model (Engle and Russell, 1998) with the best subset selection regression, we conduct a novel analysis of the relative importance of the MMS variables based on their contribution to volatility estimation and provide guidance on the optimal selection of MMS variables for volatility estimation, followed by a study of the effect of inclusion of MMS variables on the quality of volatility estimates.

Our main empirical findings based on 29 highly liquid securities and a market index ETF (SPY) suggest that, firstly, order flow and number of trades possess the most important information for volatility estimation, followed by total quote depth, quote depth difference, bid-ask spread, order imbalance and trading volume. Both the ranking and the optimal choice of MMS covariates vary considerably across stocks. More importantly, we demonstrate that, by benchmarking on a realized kernel (RK) estimator, the volatility estimates obtained from the ACD model with the inclusion of optimally selected MMS covariates significantly outperform those obtained from an ACD model without any MMS covariates on both daily and intraday levels. Moreover, including all MMS covariates does not further improve the results. These findings have two important implications: (1) our variable selection procedure can to a great extent extract the most relevant information for volatility estimation; (2) MMS covariates indeed contain invaluable information about the latent price volatility process and should not be overlooked.

The contribution of this chapter is three-folded. Firstly, we develop a framework for ranking and choosing various MMS variables for high-frequency volatility estimation. Secondly, we are among the first to provide rankings and optimal choices of a universe of MMS variables based on their relative importance for volatility estimation, which provides new insights into the relationship between intraday volatility and other variables. Thirdly, we demonstrate that it is possible to obtain more precise volatility estimates by including MMS covariates, which is an important alternative to the

¹The definitions of these MMS variables are given later on in Table 2.1

RV-type volatility estimators, especially when the number of observations are limited.

The rest of this chapter is structured as follow: A literature review on the related theoretical and empirical findings on the relationship between volatility and other MMS variables is given in Section 2.2. Section 2.3 introduces the duration-based volatility estimator. Section 2.4 describes the main methodology used in this chapter. Data and descriptive statistics are given in Section 2.5. Our main findings are presented in Section 2.6. Section 2.7 concludes.

2.2 Literature Review

This section discusses the rationale for our choice of MMS variables. The intraday relationship between price change dynamics and MMS variables has been studied extensively in the past two decades. These studies typically model intraday price dynamics by incorporating variables that summarize high-frequency trading mechanisms (e.g. trades, trade durations, information asymmetry) into the high-frequency price formation process based on some standard market microstructure models (e.g. Glosten and Milgrom (1985), Kyle (1985), among others). Since the seminal work by Brennan and Subrahmanyam (1996) and Madhavan, Richardson, and Roomans (1997), various models have been proposed that examine the market microstructure effects on a transaction level, for example Ghysels and Jasiak (1998), Dufour and Engle (2000), Rydberg and Shephard (2003), Manganelli (2005), Sadka (2006), Nolte (2008), Jondeau, Lahaye, and Rockinger (2015) amongst others. These studies mainly focus on a comprehensive structure and trading mechanism rather than focusing on volatility estimation. However, this research is closely linked to our study as it also examines the relationships between market microstructure covariates and volatility in a high-frequency context, and rationalizes our inclusion of MMS covariates by providing both theoretical and empirical evidence to the relationship.

In the existing literature, the relationship between volatility and trading volume has been discussed in various contexts. Since the seminal work of Copeland (1976), the relationship between trading volume and return volatility has attracted a lot of attention (see Epps and Epps (1976), Tauchen and Pitts (1983), Lamoureux and Lastrapes (1990), Andersen and Bollerslev (1997a), Suominen (2001), Kalev, Liu, Pham, and Jarnecic (2004), Darrat, Zhong, and Cheng (2007) among others). In these papers, trading volume is regarded as a proxy for information arrival which drives the price volatility. In a high-frequency context, information-based MMS models suggest that a high trading volume is associated with a subsequent trade that has a high

return variance (Easley and O'Hara, 1992). Manganeli (2005) supports the results in Easley and O'Hara (1992) for only the most frequently traded stocks. Using a copula approach, Nolte (2008) documents a positive simultaneous effect between price change and trading volume. Concluding from the above studies, it is clear that volume is considered as an important determinant of volatility at both low and high frequencies.

Trading frequency, or the number of trades in a given interval, connects to price volatility in a different way. Diamond and Verrecchia (1987) assert that given a short-selling constraint, informed traders cannot sell short when the bad news arrives. This results in a reduced trading frequency, but larger price adjustments afterwards. On the other hand, Easley and O'Hara (1992) suggest that higher trading frequency is associated with more informed traders participating, which subsequently triggers price movements. In a different context, Oomen (2006) proposes a model for the observed price process where the volatility within a time interval is proportional to the number of transactions within the interval. Engle and Russell (1998) document a negative relationship between price durations and the number of trades per second in a duration.

Order flow and order imbalance, which measures the information flow revealed through trading (Easley, Kiefer, O'Hara, and Paperman, 1996; Hasbrouck, 1991; Opschoor, Taylor, van der Wel, and van Dijk, 2014), are also closely linked to return volatility. In our study, order imbalance is defined as the difference between the number of buy and sell orders in a specific time interval, and order flow is referred to as the difference between the buyer- and seller-initiated volume (or volume-weighted order imbalance). The measure has a close link to market liquidity, and MMS theories predict that return volatility is induced by net order flow (Admati and Pfleiderer, 1988; Kyle, 1985), which has been confirmed empirically in various studies (e.g. Madhavan, Richardson, and Roomans (1997), Chan and Fong (2000), Opschoor, Taylor, van der Wel, and van Dijk (2014)). Chordia, Roll, and Subrahmanyam (2002) show that a daily order imbalance is strongly related to contemporaneous volatility and the effect is asymmetric for order imbalances with excess buy or sell orders. These studies, however, mainly focus on relationships between volatility and order imbalance at a daily or lower frequency. Following these studies, we examine the relationship between a measure of order flow/imbalance and local volatility in a high-frequency context. It is worth noting that the order flow/imbalance measure is also related to the Probability of Informed Trading (PIN) measure, as order imbalance and order flow reveal the information content in the trades which, in turn, updates the price and triggers price volatility (Easley, López de Prado, and O'Hara, 2012; Easley and O'Hara, 1992).

Bid-ask spreads and quote depths link to price volatility from the liquidity perspective. Classic information-based MMS theory (e.g. Copeland and Galai (1983), Easley and O'Hara (1992)) predicts a positive relationship between the bid-ask spread and return volatility, due to the arrival of private information, which increases the degree of information asymmetry in the market. Kyle (1985) and Rahman, Krishnamurti, and Lee (2005) assert that market depths are negatively affected by the information asymmetry, and for larger depths, it is less likely that orders can 'walk the book', and therefore are associated with lower return volatility. Parlour (1998) suggests that, when the bid (ask) side of the limit order book has excess liquidity over the ask (bid) side, traders will submit market buy (sell) orders instead of limit buy (sell) orders for prompt execution, which will subsequently move up (down) the price and cause a contemporaneous increase in return volatility. Thus, these theories suggest that bid-ask spread and the difference between bid depth and ask depth should move in the same direction as price volatility, while the quote depths per se will move in the opposite direction. These theoretical findings are supported by many empirical studies, for instance Bollerslev and Melvin (1994), Handa and Schwartz (1996), Foucault (1999), Ahn, Bae, and Chan (2001), Næs and Skjeltorp (2006), Nolte (2008) and Hussain (2011), among others.

Concluding from above, all MMS variables contain information about price volatility. We expect a positive relationship between volume, order imbalance, order flow, bid-ask spread and volatility, and a negative relationship between quote depth difference with volatility. The relationship between number of trades and volatility is mixed, but the results in Engle and Russell (1998) are very close to our approach, and suggest a positive relationship. In this chapter, we provide a clear picture of the relationship between these variables and local volatility in a high-frequency setting, and assess their relative informativeness in volatility estimation.

2.3 Price Duration and Volatility Estimation

We follow the general framework of point process-based volatility estimation in Engle and Russell (1998), Gerhard and Hautsch (2002), Tse and Yang (2012) and Nolte, Taylor, and Zhao (2018). Let $P(t)$ denote an observed price process, and suppose a decision maker in need of a risk measure is concerned about the size of a significant price change, δ . Construct the absolute price change point process (Engle and Russell, 1998) as follows:

Definition 2.1 (The Absolute Price Change Point Process). *Based on a realization of the observed price process in level $P(t)$, we construct a point process by:*

1. Set $t_0^{(\delta)} = 0$ and choose a threshold δ .
2. For $i = 1, 2, \dots$, compute the first exit time, $t_i^{(\delta)}$, of $P(t_{i-1}^{(\delta)})$ through the double barrier $[P(t_{i-1}^{(\delta)}) - \delta, P(t_{i-1}^{(\delta)}) + \delta]$ as:

$$t_i^{(\delta)} = \inf_{t > t_{i-1}^{(\delta)}} \{|P(t) - P(t_{i-1}^{(\delta)})| \geq \delta\}.$$

Iterate until the sample is depleted.

The point process $\{t_i^{(\delta)}\}$ describes the arrival times of transactions that result in a price change of at least δ . We will refer to the arrival times $\{t_i^{(\delta)}\}$ as the δ -related price events, or price events for short. Intuitively, for a given price series, more frequent arrivals of price events in a given time frame can be translated into higher price volatility.

To construct a volatility measure from the absolute price change point process, we introduce three related concepts that are considered as equivalent representations of the point process.

Definition 2.2 (Counting Representation). *For the absolute price change point process defined by the arrival times of price events $\{t_i^{(\delta)}\}$, the counting process of this point process is defined by:*

$$N^{(\delta)}(t) \equiv \sum_{i=1}^{\infty} \mathbb{1}_{\{t \geq t_i\}}. \quad (2.1)$$

Definition 2.3 (Price Duration Representation). *The price duration process of $\{t_i^{(\delta)}\}$ is defined by:*

$$x_i^{(\delta)} \equiv t_i^{(\delta)} - t_{i-1}^{(\delta)}. \quad (2.2)$$

Definition 2.4 (Conditional Intensity Representation). *Let \mathcal{F}_t denote the natural filtration of the point process $\{t_i^{(\delta)}\}$. The \mathcal{F}_t -conditional intensity process of $\{t_i^{(\delta)}\}$ is defined by:*

$$\lambda^{(\delta)}(t|\mathcal{F}_t) \equiv \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{E}[N^{(\delta)}(t+\Delta) - N^{(\delta)}(t)|\mathcal{F}_t]. \quad (2.3)$$

Definitions 2.2 to 2.4 are three equivalent characterizations of the point process $\{t_i^{(\delta)}\}$. Specially, the \mathcal{F}_t -conditional intensity process can be interpreted as the expected number of price events for the next instant, which has a close connection to an instantaneous volatility measure as given by the formula below (Hautsch, 2012):

$$\sigma_{(\delta)}^2(t) = \lambda^{(\delta)}(t|\mathcal{F}_t) \left(\frac{\delta}{P(t)} \right)^2. \quad (2.4)$$

Intuitively, each price event is associated with $\left(\frac{\delta}{P(t)}\right)^2$ amount of price volatility. The above formula simply uses the expected number of price events multiplied by the price volatility contribution of each price event as an instantaneous volatility measure.

To model the latent object $\lambda^{(\delta)}(t|\mathcal{F}_t)$ using the observable price durations $x_i^{(\delta)}$, we introduce an alternative definition of the conditional intensity (Daley and Vere-Jones, 2003):

$$\lambda^{(\delta)}(t|\mathcal{F}_t) = \frac{f_x(t-t_{i-1}^{(\delta)}|\mathcal{F}_{t_{i-1}^{(\delta)}})}{1-F_x(t-t_{i-1}^{(\delta)}|\mathcal{F}_{t_{i-1}^{(\delta)}})}, \quad t \in (t_{i-1}^{(\delta)}, t_i^{(\delta)}], i=1,2,\dots \quad (2.5)$$

in which $f_x(\cdot|\mathcal{F}_{t_{i-1}^{(\delta)}})$ and $F_x(\cdot|\mathcal{F}_{t_{i-1}^{(\delta)}})$ are the conditional densities of $x_i^{(\delta)}$ conditioning on the information set $\mathcal{F}_{t_{i-1}^{(\delta)}}$. Thus, by modelling the conditional cumulative density function (CDF) of $x_i^{(\delta)}$, we can make inference about instantaneous volatility within the spell of a price duration.²

Usually we are more interested in the integrated conditional variance over some period $(0, T)$. It is then natural to integrate (2.5) to obtain an estimate of the integrated conditional variance (ICV). Suppose that $t_l^{(\delta)} \leq T$ is the last arrival of price events in the dataset, then an estimate of the ICV can be constructed as:

$$ICV(0, T) \equiv \int_0^T \sigma_{(\delta)}^2(t) dt = - \sum_{i=1}^l \ln \left(1 - F_x(x_i^{(\delta)}|\mathcal{F}_{t_{i-1}^{(\delta)}}) \right) \left(\frac{\delta}{P(t_{i-1}^{(\delta)})} \right)^2 \quad (2.6)$$

In practice, we will replace the conditional CDF $F_x(\cdot|\mathcal{F}_{t_{i-1}^{(\delta)}})$ in the above estimator by an empirical estimate. The performance of the ICV estimator then depends crucially on the goodness-of-fit of the model for the conditional density of $x_i^{(\delta)}$.

In the discussion above, when we consider the conditional distribution of $x_i^{(\delta)}$, we restrict ourselves to only condition on the natural filtration $\mathcal{F}_{t_{i-1}^{(\delta)}}$, which contains all the internal history of the point process up to time $t_{i-1}^{(\delta)}$. However, as discussed in the literature section, there are various MMS covariates that are considered to be related to price volatility. Therefore, it is expected that by conditioning on an extended information set, we can model the conditional distribution of $x_i^{(\delta)}$ with better accuracy, which in return improves the quality of volatility estimates. In our

²The information about the conditional intensity is updated upon arrival of every price events, but not between two price events.

study, we model the following conditional density by extending the information set to:

$$F_x(x_i^{(\delta)} | \mathcal{F}_{t_{i-1}}^{(\delta)} \cup G_{t_i}^{(\delta)}), \quad (2.7)$$

in which $G_{t_i}^{(\delta)}$ is the information set of some MMS covariates up to time $t_i^{(\delta)}$. We will use $\widetilde{\mathcal{F}}_{t_i}^{(\delta)} = \mathcal{F}_{t_{i-1}}^{(\delta)} \cup G_{t_i}^{(\delta)}$ to denote the extended information set. Note that we are essentially using contemporaneous information in other MMS covariates to fit the conditional density of $x_i^{(\delta)}$. As our main interest is to provide a precise *ex post* price duration-based volatility estimator, exploiting information in the contemporaneous covariates can to a large extent improve the goodness-of-fit of the duration density, which in turn yields a more precise volatility estimator. We would like to stress that although contemporaneous information is not permitted in a forecasting setting, here it does not contradict our method. In the same way as RV estimates, we focus on the construction of an input for a forecasting model rather than the specification of the forecasting model, and volatility estimates obtained from our model can always be used in volatility forecasting specifications such as the HAR model (Corsi, 2009).

The use of $\widetilde{\mathcal{F}}_{t_i}^{(\delta)}$ in volatility estimation is also motivated by the fact that we can analyse the interactions between contemporaneous MMS variables and price volatility. To see this, firstly note that under correct specification of $F_x(x_i^{(\delta)} | \widetilde{\mathcal{F}}_{t_i}^{(\delta)})$, the following holds:

$$-\ln \left(1 - F_x(x_i^{(\delta)} | \widetilde{\mathcal{F}}_{t_i}^{(\delta)}) \right) \sim i.i.d.exp(1). \quad (2.8)$$

The above relationship is known as the exponential probability integral transform. Let us define $ICV_i \equiv ICV(t_{i-1}^{(\delta)}, t_i^{(\delta)})$, then from (2.6) and (2.8) it is immediate that ICV_i is a process of i.i.d. exponential random variables. If we define the average instantaneous volatility within the i -th price duration as:

$$\overline{\sigma}_i^2 = ICV_i / x_i^{(\delta)}, \quad (2.9)$$

then by taking logarithm on both sides of (2.9), it follows that:

$$\mathbb{E}[\ln \overline{\sigma}_i^2] + \mathbb{E}[\ln x_i^{(\delta)}] = \ln C + \gamma, \quad (2.10)$$

$$\text{Cov}(\ln \overline{\sigma}_i^2, \ln x_i^{(\delta)}) = \frac{\pi^2}{12} - \frac{\text{V}[\ln \overline{\sigma}_i^2] + \text{V}[\ln x_i^{(\delta)}]}{2}, \quad (2.11)$$

where $C = \delta^2 / P(t_{i-1}^{(\delta)})^2$ and $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.³ Since empirically $\text{Cov}(\ln \bar{\sigma}_i^2, \ln x_i^{(\delta)})$ is almost always negative due to the large variations in the price durations (in our data, $\text{V}[\ln x_i^{(\delta)}]$ is around 2 for all securities), we can interpret price durations as an inverse measure of the average instantaneous volatility within the duration. Intuitively, the longer the price duration, the longer it takes for the price to change by δ amount, and consequently the lower is the average instantaneous volatility therein.

By examining the correlation between price durations and other MMS covariates, we can therefore infer the correlation between other MMS covariates and the average instantaneous volatility. For example, if the trading volume per second in a price duration is negatively correlated with price duration, then we would expect it to be positively correlated with the average instantaneous volatility, and vice versa. However, there are some caveats in the interpretation of the correlation between covariates and price durations. Since a price duration is also a measure of time, if an included MMS covariate also scales up with time, then the correlation between that covariate and price duration will be strongly positive. These MMS covariates include trading volume, number of trades and total quote depths submitted within a price duration. To ensure that we can translate the relationship between covariates and price durations into the relationship between covariates and average instantaneous volatility, we use a per second measure of these variables by dividing them by the length of the corresponding price duration.

The (price) duration-based volatility estimator has an intrinsic link to the popular RV approach. The RV approach typically relies on the assumption that the log-price process follows a jump-diffusion process with observation error (discretization, MMS noise⁴, etc.), and estimates the integrated variance of the log-price process non-parametrically. Thus, based on a specific log-price model, asymptotic properties of the RV-type estimators can be derived, and biases introduced by observation error and jumps can be corrected. On the contrary, the price-duration based volatility estimator does not impose any assumptions on the log-price process, and simply treats the price events as a proxy of volatility. This approach is apparently less vulnerable to specification error of the log-price model, but no theoretical results can be derived without any assumptions about the underlying price process. Li,

³Note that to derive the above relationships, we used the fact that if ε_i is i.i.d. unit exponential, then $\text{E}[\ln \varepsilon_i] = \gamma$ and $\text{V}[\ln \varepsilon_i] = \frac{\pi^2}{12}$.

⁴In the RV literature, MMS noise or market microstructure noise refers to the deviation of observed price from the latent efficient price process caused by microstructure effects such as the bid-ask bounce effect, strategic trading and imperfections in the trading mechanisms, etc.

Nolte, and Nolte (2018a) prove that if we assume that the log-price process follows a jump-diffusion process, then the ICV estimator consistently estimates integrated variance of the jump-diffusion process with a much higher efficiency compared to the RV-type estimators, if the conditional intensity process is known.⁵ This result establishes the theoretical foundation of the ICV estimator, and suggests that the quality of the volatility estimates of the ICV estimator depends crucially on the goodness-of-fit of the parametric model of the absolute price change point process.

We would like to conclude our discussion of the duration-based volatility estimator by summarizing the advantages of the ICV estimator over the RV estimator. Firstly, a parametric structure can be specified for the price durations or the conditional intensity process. This allows us to include more data outside the window for which volatility needs to be estimated. For example, we could use monthly data to construct daily volatility estimators. This can lead to efficiency gains as the RV-type approach is confined to the data within the estimation window of volatility. Moreover, this parametric structure allows the inclusion of other MMS covariates, which can potentially improve the accuracy of intraday volatility estimates. Also, intraday volatility estimates or even instantaneous volatility estimates can be constructed based on the price durations or conditional intensity process. This can be particularly challenging for RV-type estimators as smaller estimation windows results in a much smaller amount of data, which greatly affects the performance of the RV-type estimators.

2.4 Price Duration Modelling

We model $x_i^{(\delta)}$ using a modified version of the Lognormal Log-ACD (LL-ACD) model proposed by Allen, Chan, McAleer, and Peiris (2008). The LL-ACD(p,q) model is specified as follows:

$$x_i = \exp(c + \gamma'Z_i + \Psi_i), \quad (2.12)$$

$$\Psi_i = \sum_{j=1}^p \beta_j \Psi_{j-1} + \sum_{j=1}^q \alpha_j \varepsilon_{i-j} + \varepsilon_i, \quad (2.13)$$

$$\varepsilon_i \sim i.i.d. \mathcal{N}(0, \sigma_\varepsilon^2), \quad (2.14)$$

in which Z_i and γ are some covariates and the associated parameter vector. It is obvious that the model is equivalent to a log-linear regression of x_i on Z_i with ARMA-type

⁵In Li, Nolte, and Nolte (2018a), it is proven that a non-parametric duration-based volatility estimator is more than 6 times more efficient than a RV estimator, while the parametric duration-based volatility estimator is more than 20 times more efficient than the non-parametric duration-based volatility estimator.

error terms Ψ_i . Let us define the parameter vector $\theta = \{c, \gamma', \beta, \dots, \beta_p, \alpha_1, \dots, \alpha_q, \sigma_\varepsilon^2\}'$. Based on the dataset $\{\mathbf{X}, \mathbf{Z}\}$ where $\mathbf{X} = \{x_i^{(\delta)}\}_{i=1:I}$ and $\mathbf{Z} = \{Z_i\}_{i=1:I}$, the model can be easily estimated via Quasi Maximum Likelihood (QML) with the following conditional log-likelihood function:

$$\ln \mathcal{L}(\theta; \mathbf{X}, \mathbf{Z}) = -\frac{I}{2} \ln 2\pi - \frac{I}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2} \sum_{i=1}^I \frac{\varepsilon_i^2}{\sigma_\varepsilon^2}. \quad (2.15)$$

It is then clear that the density $F_x(\cdot | \mathcal{F}_{t_{i-1}}^{(\delta)}, \theta)$ estimator is simply a log-normal density. We can then construct the ICV estimator based on the estimated parameter vector $\hat{\theta}$ and the estimated error term $\hat{\varepsilon}_i$ as:

$$ICV(0, T) = -\sum_{i=1}^I \ln \left(1 - \Phi(\hat{\varepsilon}_i / \hat{\sigma}_\varepsilon) \right) \left(\frac{\delta}{P(t_{i-1}^{(\delta)})} \right)^2, \quad (2.16)$$

in which $\Phi(\cdot)$ is the CDF of a standard normal distribution.

The LL-ACD model is chosen because of the following rationales: Firstly, the log-linear form allows us to include seasonality components and other explanatory variables freely with a guaranteed positive fitted duration $\hat{x}_i^{(\delta)}$. Secondly, the QML estimation of the LL-ACD model, as derived by Allen, Chan, McAleer, and Peiris (2008) using the results in Bollerslev and Wooldridge (1992), ensures that the parameter estimates are consistent even when the log-normal density is misspecified. This is a crucial property of the model that validates our analysis especially when explanatory variables are included. It is worth noting that estimates of γ are still vulnerable to endogeneity bias due to simultaneity or omitted variables, which is likely to be the case in our analysis. However, since we mainly focus on the goodness-of-fit of the model in general to improve the quality of volatility estimates instead of the marginal effects of the included variables, this is not our major concern. Thirdly, (2.12) permits a method to estimate γ conveniently via OLS regressions, since the dynamic structure in the residuals does not alter the unbiasedness of the OLS estimator. This allows us to examine the relative importance of explanatory variables using simple OLS regressions without estimating the dynamic parameters. Also, the OLS estimates can be used to initialize the LL-ACD estimation, which can speed up the estimation significantly.

2.4.1 Inclusion and Selection of Variables

Empirically price durations are subject to diurnal effects which are usually filtered out prior to model estimation (Engle and Russell, 1998). The LL-ACD model allows

Table 2.1 Description of MMS variables

Name	Notation	Parameter	Description
Volume	VOL_i	γ_{VOL}	Logarithm of the total trading volume per second in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.
Total quote depth	TQ_i	γ_{TQ}	Logarithm of the sum of the best bid and ask depth per second in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.
Number of transactions	NT_i	γ_{NT}	Logarithm of the number of transactions per second in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.
Bid-ask spread	BAS_i	γ_{BAS}	Mean bid-ask spread in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.
Quote depth difference	QD_i	γ_{QD}	The absolute difference between the logarithm of the sum of the best bid and the (sum of) best ask depth in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.
Order imbalance	OI_i	γ_{OI}	The absolute difference between the logarithm of the sum of the number of buyer- and seller-initiated orders in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.
Order flow	OF_i	γ_{OF}	The absolute difference between the logarithm of the sum of the buyer- and seller-initiated volume in $(t_{i-1}^{(\delta)}, t_i^{(\delta)})$.

for joint estimation of seasonality parameters and the ACD parameters, which can potentially lead to efficiency gains in parameter estimation (Hautsch, 2012).

We specify the seasonality regressors in a flexible Fourier form as proposed by Andersen and Bollerslev (1997b):

$$s_i = \sum_{j=1}^P v_j \bar{t}_i^j + \sum_{j=1}^Q \left(v_{c,j} \cos(\bar{t}_i \cdot 2\pi j) + v_{s,j} \sin(\bar{t}_i \cdot 2\pi j) \right), \quad (2.17)$$

in which P and Q are predetermined degrees of the polynomials, and \bar{t}_i is the calendar time of the i -th event divided by the total length of a trading day. v_j , $v_{c,j}$ and $v_{s,j}$ are parameters to be estimated. We can include each component of s_i into Z_i to estimate the seasonality parameters. In all of our model estimations, the degree of polynomials for the flexible Fourier regressors are set to be $P = 1$ and $Q = 3$.

A key contribution of this chapter is the inclusion and selection of MMS covariates in intraday volatility estimation. Based on our discussion in the literature review section, we include the following covariates in the estimation procedure: trading volume, total quote depth, quote depth difference, order imbalance, order flow and the number of transactions. The definitions of these variables are summarized in Table 2.1. As discussed in Section 2.3, for the VOL_i , TQ_i and NT_i variables, we use a measure of accumulated speed instead of the exact quantity to ensure a meaningful

interpretation of the associated parameters. We include both the contemporaneous MMS covariates and their one-duration lagged version in the LL-ACD model to control for potential lead-lag relationships between price durations and these covariates. However, we would like to note that, in this setting there is no clear economic interpretation of the coefficients of the lagged covariates due to the randomness of the length of the past duration.

As a summary to the discussion above, the $\gamma'Z_i$ component in (2.12) can be written as follows:

$$\begin{aligned} \gamma'Z_i = & s_i + \gamma_{VOL,0}VOL_i + \gamma_{TQ,0}TQ_i + \gamma_{NT,0}NT_i + \gamma_{BAS,0}BAS_i + \gamma_{QD,0}QD_i + \gamma_{OI,0}OI_i + \gamma_{OF,0}OF_i \\ & + \gamma_{VOL,1}VOL_{i-1} + \gamma_{TQ,1}TQ_{i-1} + \gamma_{NT,1}NT_{i-1} + \gamma_{BAS,1}BAS_{i-1} + \gamma_{QD,1}QD_{i-1} + \gamma_{OI,1}OI_{i-1} + \gamma_{OF,1}OF_{i-1}. \end{aligned} \quad (2.18)$$

In empirical studies, one can include a richer set of explanatory variables in $\gamma'Z_i$ that may further improve the goodness-of-fit of the model. However, as the number of parameter increases, the efficiency of the parameter estimates deteriorates, which leads to less efficient volatility estimates for a given sample size. To improve the performance of volatility estimation, we propose to only include the most relevant covariates in the estimation of the LL-ACD model. The relevance of the covariates is determined adaptively using an OLS regression as specified in (2.12) by treating Ψ_i as a correlated error term.

We use the best subset regression (BSR) to select the MMS covariates that are most relevant to price duration modelling. BSR is a classical statistical method (see, e.g. Beale, Kendall, and Mann (1967) and Hocking and Leslie (1967)) that is frequently used in variable selection problems. It has recently regained attention due to the development of optimization methods in Bertsimas, King, and Mazumder (2016). To discuss this selection method in detail, let $\gamma = \{\mathbf{v}, \boldsymbol{\gamma}_Z\}$ where \mathbf{v} is the $(P + 2Q)$ -by-1 seasonality parameters, and $\boldsymbol{\gamma}_Z$ is the 14-by-1 vector of the MMS parameters. Starting with the regression model:

$$\ln x_i^{(\delta)} = c + \gamma'Z_i + \Psi_i. \quad (2.19)$$

For each K ranging from 1 to 14, we solve the following nonconvex problem:

$$\min_{c, \boldsymbol{\gamma}} \|\ln x_i^{(\delta)} - c + \gamma'Z_i\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\gamma}_Z\|_0 \leq K \quad (2.20)$$

where $\|(\cdot)\|_2$ is the l_2 norm and $\|\boldsymbol{\gamma}_Z\|_0$ denotes the pseudo-norm of $\boldsymbol{\gamma}_Z$ that counts the number of non-zero elements in $\boldsymbol{\gamma}_Z$. The problem above can be expressed as a Mixture Integer Optimization (MIO) problem, as suggested by Bertsimas, King, and

Mazumder (2016), and can be solved very efficiently using MIO optimizers. The detailed optimization setup is documented in Appendix B.3.

In essence, BSR for a given K finds the optimal combination of K different covariates that minimizes the mean squared error. We will refer to the optimized model for a given K as the K -optimal model. Intuitively, as K ranges from 1 to 14, more important MMS covariates will be included first in K -optimal models and less important MMS covariates will only be included when more important ones are already in the model. Therefore, the number of inclusions of each MMS covariate in the K -optimal model serves as a natural ranking of relative importance of each MMS variable. To determine the overall optimal combination of MMS covariates to be included, we can pick the best model among all K -optimal models using information criteria, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). We will refer to the overall best model among all the K -optimal models as K^* -optimal model, where K^* is the number of regressors that optimizes some model selection criteria.

Other choices of variable selection schemes are also available, for example, regularization methods such as the LASSO by Tibshirani (1994) or the elastic net by Zou and Hastie (2005), dimension reduction methods such as principal component regression (PCR) or partial least square (PLS) regressions (Wold, Sjöström, and Eriksson, 2001). The best subset selection approach has the advantage that it has a very straightforward economic interpretation without many tuning parameters. The relative performance of BSR and shrinkage estimators remain an open debate (see, e.g. Bertsimas, King, and Mazumder (2016) and Hastie, Tibshirani, and Tibshirani (2017)), but the BSR approach provides a simple solution to rank the relative importance of the variables based on their inclusions in the optimal model. For shrinkage estimators, it is not immediately clear what criteria should be used to rank the contribution of the variables, and the performance of these estimators depends crucially on the choice of the tuning parameters. As to the PCR and PLS approaches, one may argue that these estimators can extract some latent factors in the system of regressors, but latent factors lack a clear economic interpretation as they are just linear combinations of the regressors. A summary in Hastie, Tibshirani, and Friedman (2009) shows that the variable selection methods described above have similar in-sample performance, and we choose the best subset regression method due to its simplicity and more straightforward economic interpretations.

To summarize our estimation strategy of the LL-ACD model, we firstly ‘pre-select’

the MMS covariates via BSR in (2.19), and include the selected MMS covariates to construct the LL-ACD model. This two-step approach may not be necessary, as one can attempt to jointly select the covariates and estimate the LL-ACD model via penalized maximum likelihood. Our two-step approach has the advantage over the joint approach that BSR can be estimated very efficiently via MIO optimizers, while the joint approach can be very computationally intensive since the likelihood function needs to be optimized numerically for every different combination of MMS covariates.

2.5 Data and Descriptive Statistics

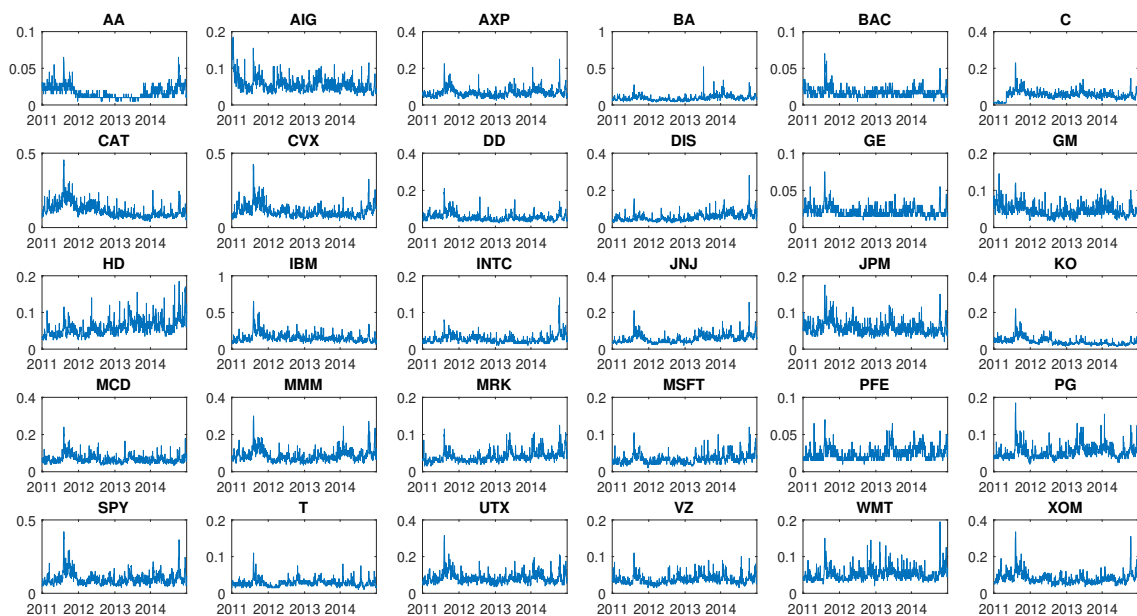
The data used in our empirical investigation is the most recent tick-by-tick trade and quote data from 29 assets, namely AA, AIG, AXP, BA, BAC, C, CAT, CVX, DD, DIS, GE, GM, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, PFE, PG, T, UTX, VZ, WMT, XOM and a stock index, SPY.⁶ The sample period is from January 2011 to December 2014, including 1006 trading days starting from 9:30 in the morning to 16:00 in the afternoon. Both trade and quote data have time stamps in seconds. The trade data consists of trade prices and volumes at all timestamps and the best quote price contains bid and ask prices and depths at their own timestamps. To determine whether a trade price is buyer- or seller-initiated, and the exact bid-ask spread for certain trade prices, it is necessary to merge the quote data with the trade data according to trade timestamps. We apply a refined Lee and Ready (1991) algorithm, as proposed by Nolte (2008), to determine the trade direction. We present some descriptive statistics of the daily transaction data in Table B.2 in Appendix B.2. The table summarizes the liquidity of the 30 securities within the sampling period. The table shows that SPY is the most liquid security with the largest amount of transactions per day, dollar volume per day and smallest Amihud's (2002) measure, whereas AA is the most illiquid stock.

We then construct the absolute price change process from the transaction price data according to Definition 2.1. We choose δ on a daily basis by letting the daily mean duration to be as close to 5 minutes as possible, in the spirit of the choice in Tse and Yang (2012) and the 5-minute RV measure (see E.g. Liu, Patton, and Sheppard (2015)). The rationale of our choice of δ is that, firstly, the price process is sampled equally sparsely for each day, so that the MMS noise does not have a large

⁶We follow the choice of assets in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), and only consider data from the primary listing of security exchanges where these stocks are traded. We use data from the New York Stock Exchange for all the stock listed except for INTC, MSFT and SPY, for which data from NASDAQ is used.

impact on the price durations. Secondly, this choice provides comparability across securities and trading days. We plot the daily selected δ for each security in Figure 2.1. From the figure we see that the chosen δ is very different across stocks, and also varies considerably over time. Generally, for each day, securities with a smaller bid-ask spread, lower price volatility and less liquidity tend to exhibit a smaller δ . Note that the stock C had a one-for-ten stock split on 9th May, 2011, which results in a drastic change in the δ chosen.

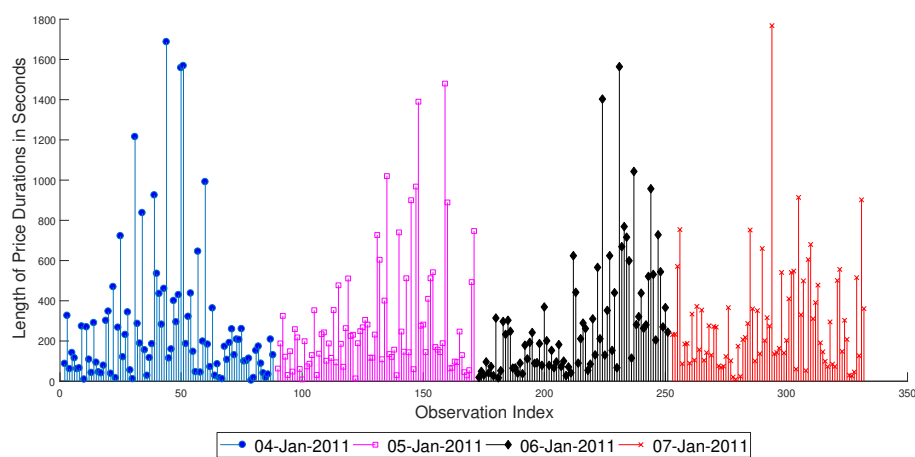
Figure 2.1 Daily price change threshold δ for all thirty stocks from 2011-Jan to 2014-Dec



Note: The x-axis represents a day index for the sample period mounting up to 1006 days. The y-axis is the value of δ . The value of δ on a day is calculated by the value that makes the average duration to be closest to 5 minutes.

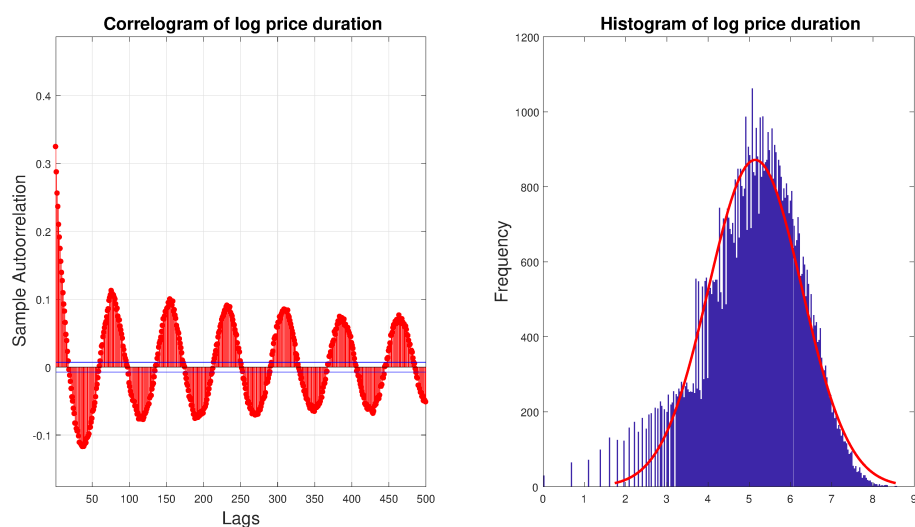
We present an example of price durations in Figure 2.2. The figure shows that there is a clear diurnal pattern in the price duration process as documented in Engle and Russell (1998). We plot the correlogram and histogram for the log price durations for SPY in Figure 2.3. Figure 2.3 confirms the seasonal effect in the price duration, with a cycle of approximately 78 lags, corresponding to a daily seasonal effect. It also suggests that the log price duration is very persistent if we only look at the peaks of the correlograms, which is also a common feature of the price duration data. From the histogram we see that the unconditional distribution of log price durations appears skewed with a discrete left tail, but is generally not leptokurtic. We present some additional correlograms and histograms in Figure B.1 in Appendix B.2, which shows that the unconditional density and the correlograms of price durations are similar across different securities.

Figure 2.2 An example of the price duration process



Note: The example is extracted from SPY, 4/1/2011 to 7/1/2011, consisting of 4 consecutive trading days.

Figure 2.3 Correlogram and histogram of log price durations for SPY



Note: The correlogram and histogram are constructed based on the log price duration of SPY for the whole sample. In the histogram, the red solid line represents a fitted normal density.

We conclude our data section by providing some descriptive statistics for the price durations and the MMS covariates. A detailed table of descriptive statistics of all variables can be found in Tables B.3 and B.4 in Appendix B.2. We present a cross-correlation table between the log price durations and the MMS covariates averaged over 30 securities in Table 2.2. From the first row of Table 2.2 we see that, all the variables have a negative contemporaneous relationship with log price duration. Our predictions in the literature review section are in line with the findings except for the TQ_i variable. It is also evident that the MMS covariates are highly correlated, especially for VOL_i , NT_i and TQ_i which are all measured in arrivals per second, and OF_i and OI_i which are measures of imbalance between order arrivals. The lagged variables have a considerably weaker cross-correlation with their contemporaneous counterparts.

Table 2.2 Cross-correlation table between log price durations and MMS covariates

	$\ln x_i^{(\delta)}$	VOL_i	QD_i	BAS_i	OI_i	OF_i	NT_i	TQ_i	
$\ln x_{i-1}^{(\delta)}$	0.4826	-0.4590	-0.2430	-0.4343	-0.2774	-0.5344	-0.5994	-0.2121	$\ln x_i^{(\delta)}$
VOL_{i-1}	-0.2587	0.6151	0.1143	0.1846	0.0653	0.1838	0.8041	0.7167	VOL_i
QD_{i-1}	-0.1119	0.0797	0.2088	0.1354	0.2304	0.2449	0.1016	0.2357	QD_i
BAS_{i-1}	-0.3609	0.2220	0.1226	0.6571	0.0520	0.2203	0.3078	-0.0233	BAS_i
OI_{i-1}	-0.0294	0.0143	0.0500	0.0279	0.0911	0.6523	0.1055	0.1351	OI_i
OF_{i-1}	-0.1721	0.0748	0.0887	0.1502	0.0576	0.1213	0.1558	0.0824	OF_i
NT_{i-1}	-0.3859	0.5947	0.0726	0.2384	0.0329	0.0848	0.6941	0.6082	NT_i
TQ_{i-1}	-0.0443	0.4667	0.0691	-0.0821	0.0641	0.0137	0.3354	0.6827	TQ_i

Note: The table reports sample autocorrelation between the row variables and column variables for the whole sampling period averaged over 30 securities. Numbers in shaded cells use the column variables on the right hand side. Numbers in bold are the first order autocorrelations of each variables.

2.6 Empirical Results

2.6.1 Analysis of the Relative Importance of MMS Covariates

In this section we analyse the relative importance of all the MMS covariates in modelling price durations. We use the best-subset regression as described in (2.19) and (2.20) in Section 2.4.1. We compute monthly K -optimal models for each stock-month and obtain monthly rankings of each MMS covariate and a monthly overall K^* -optimal model chosen as the K -optimal model with minimum BIC⁷. We present an example of the BSR output in Table 2.3. Table 2.3 shows that for SPY 2011-01, the contemporaneous volume is the most important MMS covariate as $\gamma_{VOL,0}$ is included

⁷We standardize all the MMS covariates and the seasonality variables to ensure that the magnitude of the parameter estimates are comparable.

in all the K -optimal models, while the one-duration lagged quote difference is the least important variable. BIC suggests that $K^* = 8$, so the optimal model only uses 8 out of all 14 MMS covariates.

Table 2.3 Best-subset regression outputs for SPY, 2011-01

K	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	$\gamma_{TQ,0}$	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	$\gamma_{TQ,1}$	BIC
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3906.24
2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	3357.14
3	1	0	0	1	1	0	0	0	0	0	0	0	0	0	3268.17
4	1	0	0	1	1	1	0	0	0	0	0	0	0	0	3221.35
5	1	0	0	1	1	1	0	0	0	0	0	0	1	0	3198.42
6	1	1	0	1	1	1	0	0	0	0	0	0	1	0	3181.67
7	1	1	0	1	1	1	1	0	0	0	0	0	1	0	3175.03
8	1	1	0	1	1	1	1	0	0	0	0	1	1	0	3169.44
9	1	1	0	1	1	1	1	0	0	0	0	1	1	1	3174.27
10	1	1	0	1	1	1	1	1	0	0	0	1	1	1	3178.52
11	1	1	0	1	1	1	1	1	0	0	1	1	1	1	3183.34
12	1	1	0	1	1	1	1	1	0	1	1	1	1	1	3189.13
13	1	1	1	1	1	1	1	1	0	1	1	1	1	1	3196.15
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3203.46
Sum	14	9	2	13	12	11	8	5	1	3	4	7	10	6	

Note: for each K , we present the parameters included in the K -optimal model as discussed in Section 2.4.1. The K -optimal models are obtained via (2.19) and (2.20). BIC is the Bayesian Information Criterion. The K^* -optimal model is highlighted in red.

The relative importance of MMS covariates and the specification of K^* -optimal models vary over time and across securities. To summarize our findings, for each security we present the average number of inclusions of every MMS covariate in the monthly K -optimal model and the average monthly K^* in Table 2.4. In this table, we see that $\gamma_{OF,0}$ and $\gamma_{NT,0}$ are on average included in more than 13 K -optimal models, and are considered the most important MMS covariates for monthly price duration modelling. $\gamma_{TQ,0}$ and $\gamma_{QD,0}$ follow closely behind, suggesting that the quote depth information is also very important in modelling price durations. The one-duration lagged variables are much less important than the contemporaneous variables as expected. Interestingly, $\gamma_{VOL,0}$ is the least important variable among the contemporaneous variables overall. $\overline{K^*}$ suggests that on average we would only include 7 to 8 MMS covariates in the monthly regression, which implies that most of the one-duration lagged variables are actually discarded.

To see the effect of a larger estimation window, we present the relative importance of MMS variables based on quarterly, half-yearly and yearly regressions in Tables B.5 to B.7 in Appendix B.2. Tables B.5 to B.7 show that the relative importance of the variables are fairly consistent across different lengths of estimation window. We also observe that as the estimation window expands, $\overline{K^*}$ increases as a result of a larger sample size.

Table 2.4 Average of monthly relative importance of MMS covariates

Ticker	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	$\gamma_{TQ,0}$	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	$\gamma_{TQ,1}$	\bar{K}^*
AA	9.44	10.04	4.73	11.42	13.35	13.52	7.75	5.40	2.92	2.65	5.69	6.56	5.40	6.13	7.71
AIG	4.56	10.88	10.29	6.79	13.38	13.63	11.10	3.42	3.94	4.10	3.40	5.94	7.56	6.02	6.96
AXP	5.85	10.58	10.83	7.00	13.29	13.67	10.81	3.00	3.98	4.19	3.54	5.92	7.04	5.25	7.17
BA	7.56	10.42	10.77	8.13	13.27	13.73	10.52	3.46	3.46	5.42	3.38	5.73	5.19	3.85	6.90
BAC	9.48	10.42	2.96	11.77	13.92	12.71	8.29	5.58	3.04	2.96	5.67	7.48	4.46	6.27	7.35
C	7.06	9.56	9.19	10.04	13.35	13.31	9.98	4.02	3.81	2.88	3.00	5.83	7.42	5.52	8.44
CAT	5.04	10.65	10.79	7.00	13.48	13.48	10.92	3.73	3.42	6.38	3.38	5.13	6.56	5.04	6.77
CVX	5.19	10.46	11.13	7.90	13.42	13.33	10.69	3.44	4.38	5.15	2.90	4.96	6.50	5.56	6.92
DD	5.25	10.75	10.27	7.56	13.35	13.56	10.96	3.79	4.10	3.73	3.35	5.08	7.10	6.10	6.90
DIS	5.60	10.56	10.65	6.88	13.35	13.65	10.94	3.71	3.79	3.69	3.23	5.92	7.31	5.71	7.56
GE	9.56	10.27	5.67	11.29	13.81	13.08	9.79	3.52	2.56	3.10	3.96	6.38	6.21	5.77	7.21
GM	5.42	10.38	9.50	7.33	13.29	13.71	11.29	3.21	3.33	2.90	3.75	6.08	8.33	6.35	7.79
HD	5.40	10.75	10.50	5.69	13.48	13.52	10.83	4.31	3.56	5.31	3.19	5.42	6.88	6.15	6.98
IBM	6.75	10.17	11.63	8.33	13.65	13.35	10.31	3.69	3.31	5.83	3.73	5.60	4.40	4.23	6.52
INTC	6.10	10.54	4.58	11.06	13.17	13.83	10.42	3.90	3.10	3.48	4.75	6.81	7.29	5.90	6.50
JNJ	6.06	10.38	10.77	6.54	13.29	13.65	11.21	3.50	3.65	3.48	3.77	5.71	6.56	6.42	7.29
JPM	7.56	10.15	9.44	8.44	13.42	13.42	10.85	3.08	3.42	3.58	3.21	5.73	7.42	5.23	7.85
KO	7.17	9.81	10.15	8.88	13.29	13.65	10.56	2.81	3.48	3.58	3.44	5.31	6.83	6.02	7.81
MCD	3.96	10.85	10.65	6.56	13.31	13.63	11.19	3.52	4.10	5.27	3.71	5.77	6.38	6.08	6.44
MMM	6.54	10.63	10.67	7.85	13.10	13.77	10.69	3.67	3.40	4.75	4.10	5.77	4.79	5.19	6.79
MRK	6.88	10.00	10.10	9.38	13.23	13.71	10.54	3.10	3.02	2.98	3.46	5.60	7.00	5.98	7.73
MSFT	5.08	10.63	5.46	10.77	13.06	13.94	10.52	3.08	3.17	3.02	4.96	7.27	7.69	6.33	6.56
PFE	9.69	9.73	7.48	11.19	13.56	13.19	10.60	3.08	2.96	2.77	3.79	5.42	5.98	5.52	7.88
PG	5.38	10.48	10.65	6.90	13.35	13.65	11.13	3.44	3.38	4.17	3.46	5.56	7.04	6.42	7.21
SPY	12.27	9.50	3.96	13.23	11.31	11.42	8.08	4.69	3.15	3.54	3.65	5.40	9.65	5.13	7.92
T	9.79	9.79	6.44	11.00	13.38	13.35	10.63	3.13	3.04	3.10	3.54	5.67	6.23	5.90	8.21
UTX	7.83	10.69	10.52	8.21	13.23	13.77	10.52	2.96	3.15	4.02	3.75	5.90	5.54	4.90	6.63
VZ	6.77	10.23	9.63	9.21	13.31	13.67	10.83	2.73	2.75	2.98	3.77	6.42	7.00	5.71	8.02
WMT	5.90	10.52	10.83	5.96	13.33	13.65	11.13	3.15	4.02	4.06	3.56	5.52	7.21	6.17	7.19
XOM	5.81	10.23	10.13	9.52	12.90	13.44	11.08	3.19	3.63	4.63	3.02	4.52	7.21	5.65	7.81
Average	6.83	10.33	9.01	8.73	13.29	13.47	10.47	3.58	3.43	3.92	3.74	5.81	6.67	5.68	7.30

Note: For each security, we report the averaged number of inclusions in the monthly K -optimal models for every MMS covariate. \bar{K}^* is the average number of MMS covariates in the monthly K^* -optimal model, averaged over 48 months. Covariates that receive top five average inclusions are highlighted in bold.

Our findings regarding the relative importance of MMS variables show that the rankings of variables can be very different cross-sectionally, especially for the contemporaneous volume and order imbalance. Although it is not the focus of this chapter, we provide some possible explanations to the heterogeneous rankings in our results. Findings in Li, Nolte, and Nolte (2018c) suggest that there exists regime-switching behaviour in the intraday volume-volatility relationship for individual stocks, but not for the SPY. Therefore volume information is more noisy and contain less information for other stocks, but is more important for the SPY. Since this effect is likely to be driven by the firm-specific information content according to Li, Nolte, and Nolte (2018c), we suspect that the rankings of these MMS variables are highly related to the information content of each variables. Also, NT_i and OF_i are almost always the most important variables. Thus if one is not interested in selecting the optimal MMS variables, we recommend to just include these two variables for volatility estimation.

2.6.2 Estimation of the LL-ACD model

After identifying the most relevant MMS covariates, our next step is to build the LL-ACD model based on the K^* model by specifying an ARMA structure for the residuals Φ_i from the least square regressions used in the variable selection process. The LL-ACD model is summarized in (2.12) to (2.14). In our empirical analysis we choose $p = q = 1$, so that the residual Φ_i is estimated with an ARMA(1,1) structure, and we will refer to this model as the LL-ACD(1,1) model. The model is estimated via QML with the likelihood function specified in (2.15).

We estimate the LL-ACD(1,1) model monthly for all 30×48 stock months. To compare between the performance of LL-ACD(1,1) models with or without MMS covariates, we estimate the plain LL-ACD model (LL-ACD(1,1)-P) with only seasonality covariates, LL-ACD(1,1)-K model with the optimal number of MMS covariates obtained from BSR, and the LL-ACD(1,1)-A model which includes all 14 MMS covariates⁸. We also compare the LL-ACD models with the OLS regressions used in the variable selections, and denote these models as OLS-P, OLS-K and OLS-A respectively. To conserve space, we only present an example of estimation outputs for the stock-months 2011-01 SPY in Table 2.5.⁹

Table 2.5 presents the estimated dynamic parameters and MMS parameters for

⁸Throughout this chapter, we will use the capital letters in suffixes or superscripts ‘P’ to denote a plain model without MMS covariates, ‘K’ to denote a model with optimally selected MMS covariates, and ‘A’ to denote a model which includes all MMS covariates.

⁹Full estimation outputs from all models for each stock-month are available upon request.

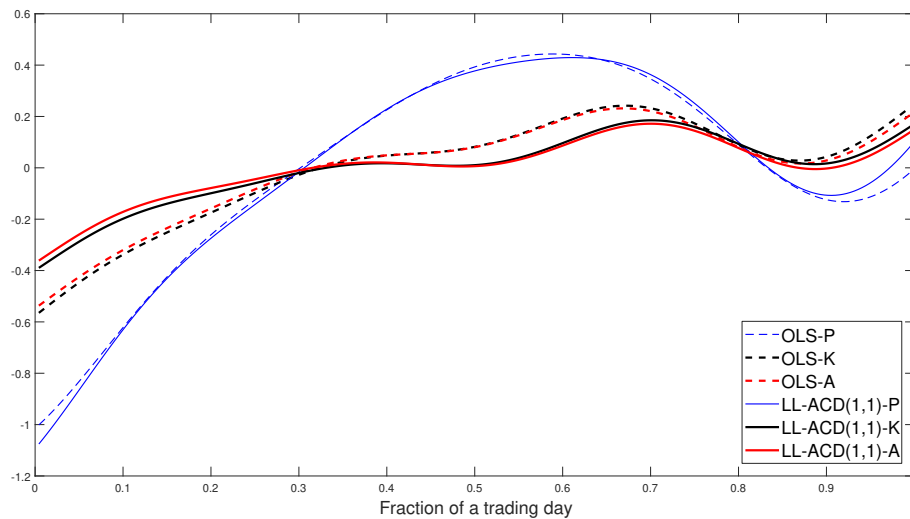
Table 2.5 Examples of LL-ACD(1,1) estimation outputs for 2011-01, SPY

Model	OLS-P	LL-ACD(1,1)-P	OLS-K	LL-ACD(1,1)-K	OLS-A	LL-ACD(1,1)-A
Dynamic Parameters						
c	5.1756*** (0.0339)	5.1746*** (0.0385)	5.1756*** (0.0218)	5.1863*** (0.0621)	5.1756*** (0.0219)	5.1890*** (0.0599)
β_1		0.7885*** (0.1297)		0.9742*** (0.0112)		0.9729*** (0.0119)
α_1		-0.6657*** (0.1660)		-0.8970*** (0.0202)		-0.8955*** (0.0212)
σ_ε^2	0.9787 -	0.9704*** (0.0199)	0.4007 -	0.6152*** (0.0105)	0.3982 -	0.6130*** (0.0105)
MMS Parameters						
$\gamma_{VOL,0}$			-0.3188*** (0.0375)	-0.3227*** (0.0364)	-0.3352*** (0.0376)	-0.3313*** (0.0366)
$\gamma_{QD,0}$			-0.0975*** (0.0173)	-0.0995*** (0.0173)	-0.1034*** (0.0173)	-0.1044*** (0.0175)
$\gamma_{BAS,0}$					0.0115 (0.0192)	0.0117 (0.0197)
$\gamma_{OI,0}$			-0.2554*** (0.0227)	-0.2502*** (0.0223)	-0.2578*** (0.0230)	-0.2476*** (0.0222)
$\gamma_{OF,0}$			-0.2349*** (0.0340)	-0.2090*** (0.0301)	-0.2317*** (0.0337)	-0.2105*** (0.0302)
$\gamma_{NT,0}$			-0.4119*** (0.0500)	-0.6109*** (0.0581)	-0.4618*** (0.0521)	-0.6370*** (0.0586)
$\gamma_{TQ,0}$			0.1142*** (0.0367)	0.2629*** (0.0455)	0.1746*** (0.0434)	0.2898*** (0.0470)
$\gamma_{VOL,1}$					0.0576* (0.0345)	0.0662* (0.0341)
$\gamma_{QD,1}$					-0.0041 (0.0177)	0.0011 (0.0163)
$\gamma_{BAS,1}$					0.0233 (0.0226)	0.0158 (0.0215)
$\gamma_{OI,1}$					0.0356 (0.0217)	0.0484** (0.0213)
$\gamma_{OF,1}$			-0.0589*** (0.0192)	-0.0277 (0.0203)	-0.0888*** (0.0255)	-0.0679** (0.0264)
$\gamma_{QD,1}$			0.1717*** (0.0285)	0.1343*** (0.0276)	0.1764*** (0.0377)	0.1191*** (0.0349)
$\gamma_{TQ,1}$					-0.0794** (0.0371)	-0.0655** (0.0302)
Diagnostic Statistics						
#Obs.	1587	1587	1587	1587	1587	1587
LL	-2234.27	-2202.74	-1525.76	-1479.90	-1520.66	-1474.25
BIC	4527.49	4486.55	3169.44	3099.83	3203.46	3132.73
R^2	0.1481	0.1804	0.6512	0.6706	0.6534	0.6729
JB	115.3318***	121.4586***	2.1850	7.1971**	1.9947	5.8690*
LL(20)	116.8915***	21.5502	140.4901***	12.9367	149.0443***	13.4882

Note: For the OLS models, the Newey-West standard errors are reported in parentheses. For the LL-ACD models, QML standard errors are in parentheses. For all estimated parameters and tests, ***: p-value<0.01, **: p-value<0.05, *: p-value<0.1. Seasonality parameters are suppressed. LL: maximized log-likelihood. BIC: Bayesian Information Criterion. R^2 is computed by $\rho(\ln x_i^{(\delta)}, \ln \hat{x}_i^{(\delta)})^2$, where $\ln \hat{x}_i^{(\delta)}$ is the fitted log price duration from the model. JB: Jacque-Bera test. LB(20): Ljung-Box test at 20 lags.

three example stock-months and the post-estimation statistics. Seasonality parameters \mathbf{v} are suppressed as they do not have meaningful interpretations. We present the diurnal patterns implied by the estimated seasonality parameters from the six models in Figure 2.4. Figure 2.4 clearly shows a reverse U-shaped pattern when no MMS covariates are included, as is documented by Engle and Russell (1998). The inclusion of MMS covariates dampens the diurnal pattern of price durations and appears to have an increasing trend over the day. The OLS and the corresponding ACD estimates are very similar, suggesting that the estimated seasonality parameters are not largely affected by the dynamic specifications.

Figure 2.4 Estimated diurnal patterns from models in Table 2.5



Note: The x-axis denote the fraction of time computed as $x/23400$, where x is the seconds since the beginning of the trading day. Diurnal pattern curves are reconstructed by plugging in the parameter estimates from the six models into (2.17).

From the dynamic parameter panels in Table 2.5, we observe a large $\hat{\beta}_1$ that is close to 1, especially when the MMS covariates are included. This shows that the price durations are very persistent, and this persistence cannot be explained by the MMS covariates. Interestingly, $\hat{\alpha}_1$ is largely negative. This suggests that an over-prediction of the previous price duration is associated with a shorter price duration, and vice versa. The estimated variance of the error terms are reduced by the inclusion of MMS covariates, as can be seen from $\hat{\sigma}_\varepsilon^2$. For the MMS parameters, it is clear that both parameter estimates and standard error estimates from OLS models are close to those obtained from the LL-ACD models for most of the variables. Also, for those MMS parameters included in the -K models, the corresponding estimates in the -A models are also similar. The signs of the contemporaneous parameters all match our

predictions in the literature review, and are highly significant except for $\widehat{\gamma}_{BAS,0}$.

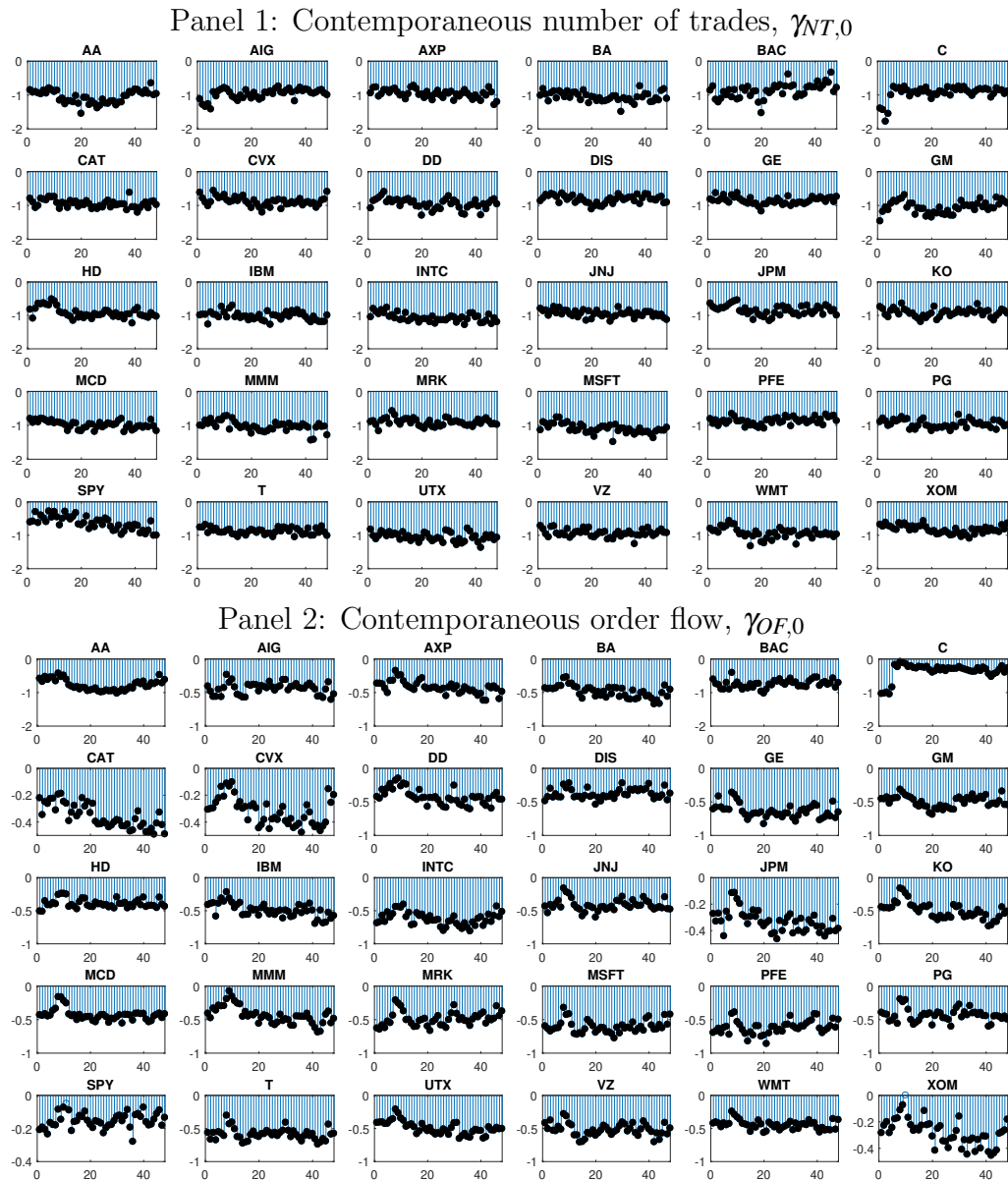
As to the diagnostic statistics, it is evident that models with more parameters have higher log-likelihood and R^2 statistics as expected, and the inclusion of MMS covariates greatly improves the performance of both the OLS and the LL-ACD models. From the BIC we see that indeed the LL-ACD(1,1)-K model outperforms all 5 other models. Diagnostic tests of the residuals show that, the inclusion of MMS covariates can greatly improve the Jacque-Bera test statistics, and the residuals are much closer to be normally distributed for the -K and -A models compared to the -P models. The LL-ACD structure seems to slightly worsen the fit of the log-normal density, but it can to a great extent capture the dependence structure in the residuals of the OLS models, as can be viewed from the Ljung-Box statistics. The dynamic structure also improves the LL, BIC and R^2 .

We proceed to summarize our main empirical findings on the MMS parameters for all stock-month estimations, which shed new lights on the intraday interaction between price volatility and the MMS covariates considered. To be concise we only present results for the contemporaneous MMS variables in the main text as their one-duration lagged counterparts are generally less important and are usually not included in the -K model. For all 30×48 stock-months, we present parameter estimates (in the order of their overall relative importance) of $\gamma_{NT,0}$, $\gamma_{OF,0}$, $\gamma_{TQ,0}$, $\gamma_{QD,0}$, $\gamma_{BAS,0}$, $\gamma_{OI,0}$, and $\gamma_{VOL,0}$ from the LL-ACD(1,1)-K model in Figures 2.5 to 2.7. Estimation results for other parameters can be found in Figures B.2 to B.6 in Appendix B.2. We also present summaries of average MMS parameter estimates for each stock based on -K and -A model in Tables B.8 and B.9.

Figure 2.5 presents results for $\gamma_{NT,0}$ and $\gamma_{OF,0}$, the two most important parameters that are almost always included in every K -optimal model. We clearly see that the two parameters are always included in the K^* -optimal model and the LL-ACD(1,1)-K model. The parameter estimates are all negative and highly significant, suggesting a positive contemporaneous relationship between the trade arrivals and order flow with price volatility, which is consistent with the theoretical prediction of Easley and O'Hara (1992). Estimates of the two parameters are relatively stable for most of the stocks, suggesting that the relationships we detect do not vary significantly over time. Interestingly we see that the stock split in C results in a structural break in the parameter estimates in Figures 2.5 to 2.7.

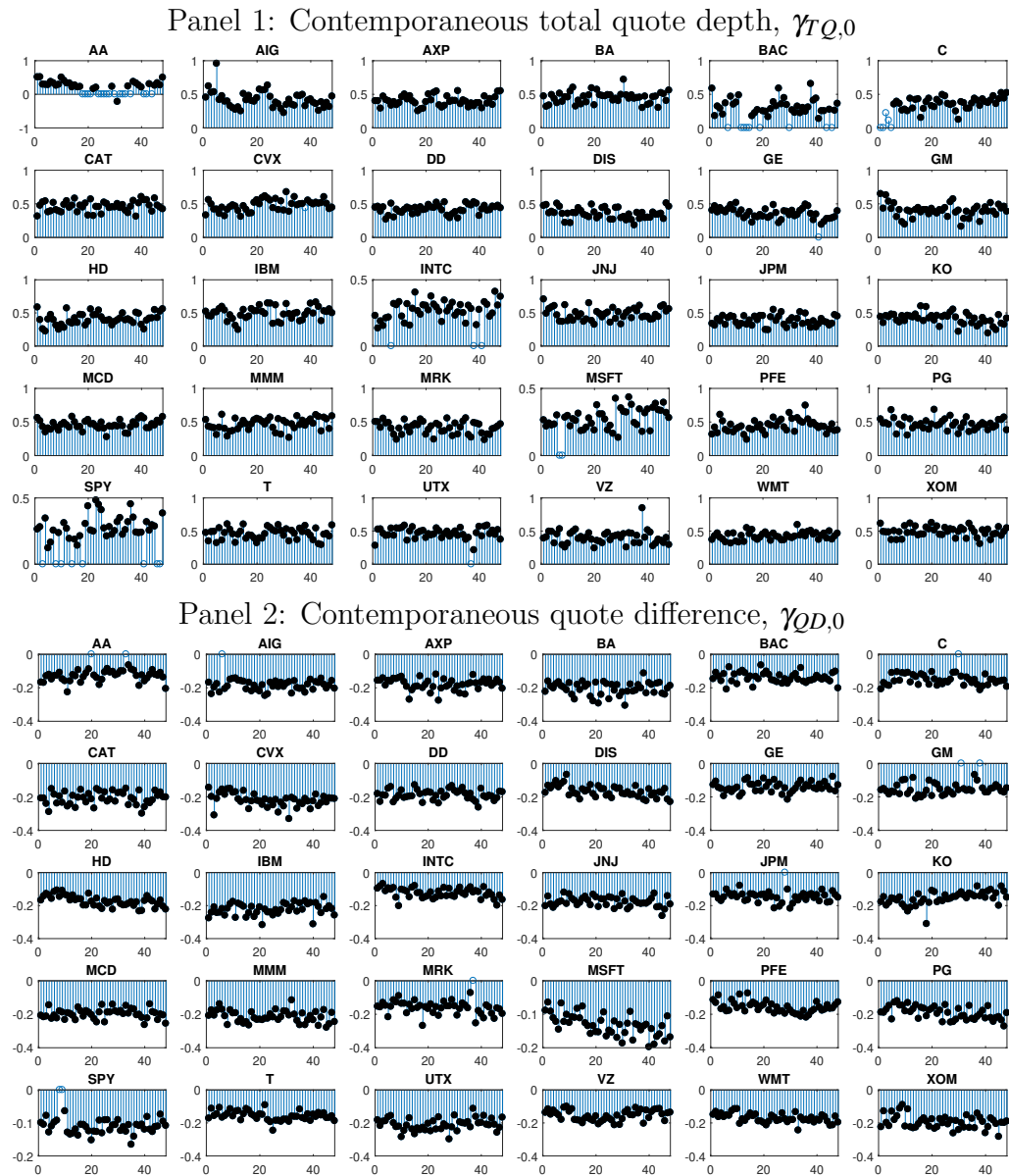
In Figure 2.6 we present the parameter estimates for the quote depth informa-

Figure 2.5 Summary of estimated contemporaneous MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 1



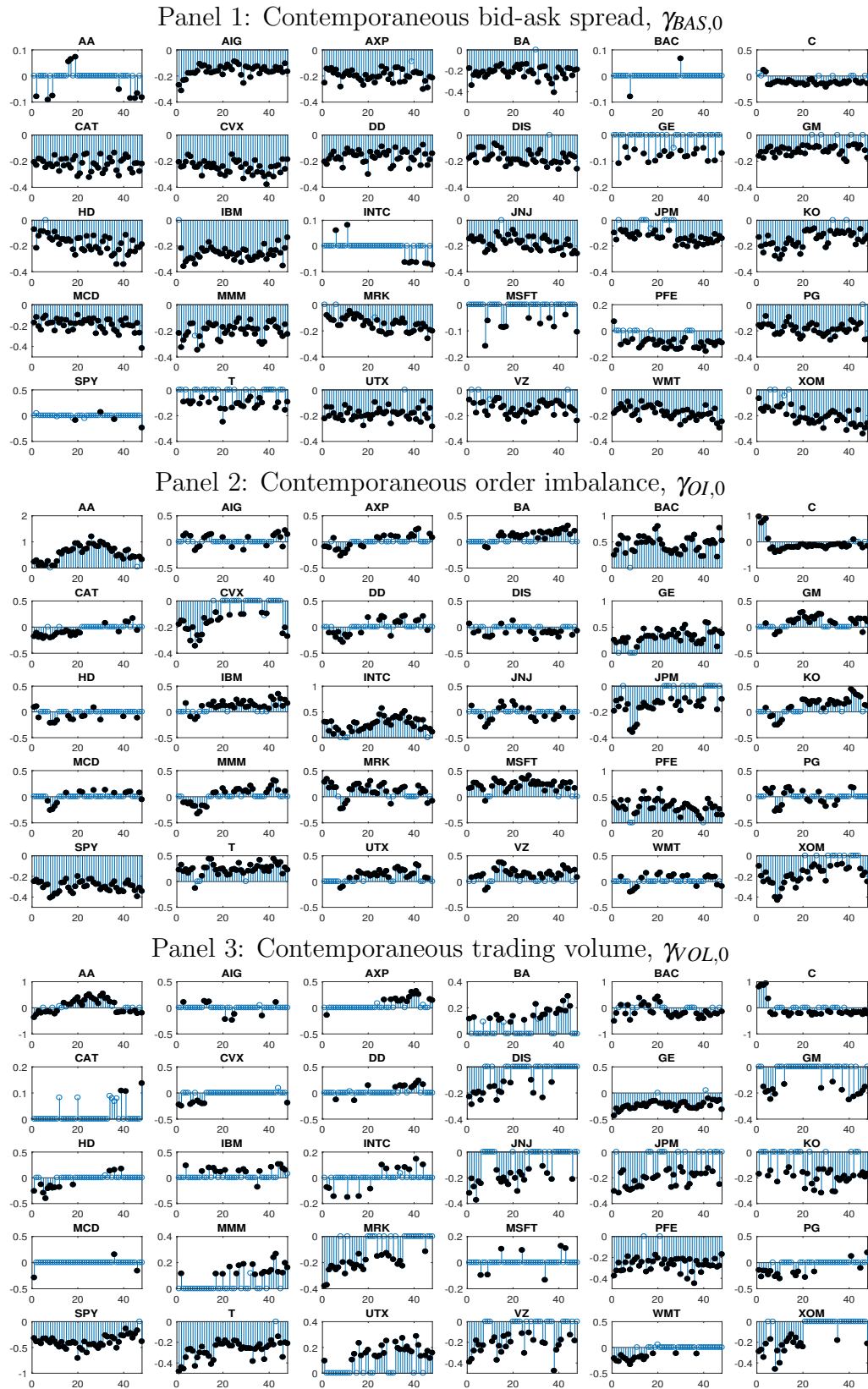
Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

Figure 2.6 Summary of estimated contemporaneous MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 2



Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

Figure 2.7 Summary of estimated contemporaneous MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 3



Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

tion, $\gamma_{TQ,0}$ and $\gamma_{QD,0}$, which are ranked in the third and fourth places among all the parameter estimates. The parameter estimates for $\gamma_{TQ,0}$ and $\gamma_{QD,0}$ generally follow our prediction that total quote depths per second co-move positively with price durations, implying a negative relation with average instantaneous volatility, while quote difference has the exact opposite effect on price durations and volatility. Similar to the results in Figure 2.5, parameter estimates of $\gamma_{TQ,0}$ and $\gamma_{QD,0}$ are almost always included in the LL-ACD(1,1)-K model and are highly significant with a relatively stable path. However, the magnitude of these parameter estimates are smaller compared to the estimated $\gamma_{NT,0}$ and $\gamma_{OF,0}$, and we observe more exclusions of $\gamma_{TQ,0}$ and $\gamma_{QD,0}$ in the LL-ACD(1,1)-K model, suggesting that they are not as important as the parameters in Figure 2.5. These results are also consistent with the aforementioned theoretical predictions by e.g. Kyle (1985) and Parlour (1998).

Figure 2.7 shows three parameters that are considered less important among all the contemporaneous parameters, namely $\gamma_{BAS,0}$, $\gamma_{OI,0}$ and $\gamma_{VOL,0}$. Our general observation is that the relationship between price durations and these covariates varies both cross-sectionally and in time. Our findings with the contemporaneous bid-ask spread are generally consistent with our prediction that a larger bid-ask spread is associated with shorter price durations and higher price volatility within the price duration. However, some securities do not possess this characteristic at all and they are simply dropped from the LL-ACD(1,1)-K model. For example, for most of the LL-ACD(1,1)-K model estimates of AA, BAC, INTC, MSFT and SPY, the parameter $\gamma_{BAS,0}$ is excluded. For $\gamma_{OI,0}$ and $\gamma_{VOL,0}$, both the sign and the significance of the parameters vary considerably across stocks and time. This is probably also the reason why these two parameters are the least important among all contemporaneous parameters. Since the correlation between the two variables and price durations are highly negative, it is likely that the information in the contemporaneous volume and order imbalance is influenced by the more important variables, such as the contemporaneous number of trades and the order flow.

To validate our empirical findings above, we perform a battery of robustness checks and present the results in Appendix B.4. Our findings generally suggest that the use of an ARMA-type structure to a very large extent captures the autoregressive structure in the residuals of the OLS model. The inclusion of the MMS covariates in the LL-ACD(1,1) model greatly improves the goodness-of-fit of the Gaussian distribution, and contributes significantly to the log-likelihood and R^2 of the LL-ACD model. The performances of the -K and the -A model are very similar. This result is consistent across different sizes of estimation windows. The performance of the

LL-ACD generally deteriorates as the size of the estimation window expands, but the -K and -A models are more robust to changes in the estimation window.

We conclude our discussion on the performance of the LL-ACD models with a pseudo¹⁰ out-of-sample test for the LL-ACD models. For a given size of the estimation window (monthly, quarterly, half-yearly or yearly), we use the estimated parameters from the previous estimation window to construct error terms and log-likelihoods using data of the current estimation window. For example, for quarterly estimated data, we simply plug in parameter estimates of the LL-ACD(1,1)-P, -K and -A models estimated on data from the first quarter into the corresponding LL-ACD models constructed using data from the second quarter. Note that for the -K model, the optimally selected MMS covariates are also based on the data from the first quarter. We then roll this process over to the next quarter till the end of the sample.

We evaluate the out-of-sample performance of the models by comparing the mean squares of the residuals (MSR) and the log-likelihoods with the MSR and log-likelihoods obtained from in-sample estimation. For each size of the estimation window, we compute the pooled MSR and log-likelihood per price duration for the three specifications of the LL-ACD(1,1) model. We also compute the difference between the in-sample and out-of-sample MSRs and log-likelihoods for each model. Intuitively this difference measures the sensitivity of the model to intertemporal changes in the data generating process (DGP) parameters. The results are presented in Table 2.6.

Table 2.6 shows that the LL-ACD(1,1)-K and LL-ACD(1,1)-A models outperform the LL-ACD(1,1)-P model in both the in-sample and out-of-sample settings for all sizes of estimation windows. The in-sample performance of the MSR and LL/obs. are superior to the out-of-sample results for all models and estimation windows as expected. The in-sample results generally suggest that the -A model has the best fit as it uses all of the MMS covariates. Nevertheless, the out-of-sample performance of the LL-ACD(1,1)-K model is significantly better than the LL-ACD(1,1)-A model based on monthly estimation windows. Also, the in-minus-out results suggest that the -P model is least sensitive to changes in the DGP parameters, as a result of its most parsimonious specification. The -K model performs better than the -A model, especially for the monthly results, as it uses only half of the MMS covariates. The results between the -K and -A models are closer as the size of the estimation window

¹⁰We would like to note that this is not a genuine out-of-sample study because the contemporaneous information is not available in an out-of-sample study. We also cannot determine the threshold value δ in an out-of-sample setting.

Table 2.6 Out-of-sample performance of the LL-ACD(1,1) models

LL-ACD(1,1)	Out-of-sample			In-sample			In minus out		
	-P	-K	-A	-P	-K	-A	-P	-K	-A
Panel 1: Monthly Estimation									
\overline{MSR}	1.1048	0.5301***	0.5315 ^{ooo}	1.0799	0.4938***	0.4897***	-0.0249	-0.0363	-0.0419
LL/obs.	-1.4649	-1.1018	-1.1035	-1.4508	-1.0639	-1.0597	0.0141	0.0379	0.0438
Panel 2: Quarterly Estimation									
\overline{MSR}	1.0991	0.5265***	0.5259***	1.0867	0.5018***	0.5006***	-0.0124	-0.0246	-0.0253
LL/obs.	-1.4637	-1.0972	-1.0967	-1.4558	-1.0726	-1.0714	0.0079	0.0246	0.0254
Panel 3: Half-Yearly Estimation									
\overline{MSR}	1.0994	0.5287***	0.5283***	1.0888	0.5059***	0.5053***	-0.0106	-0.0228	-0.0230
LL/obs.	-1.4652	-1.0990	-1.0986	-1.4575	-1.0770	-1.0764	0.0077	0.0220	0.0222
Panel 4: Yearly Estimation									
\overline{MSR}	1.1125	0.5339***	0.5340 ^{ooo}	1.1034	0.5094***	0.5091***	-0.0091	-0.0245	-0.0250
LL/obs.	-1.4720	-1.1040	-1.1041	-1.4645	-1.0807	-1.0804	0.0074	0.0232	0.0236

Note: Out-of-sample (in-sample) results refer to the residuals and log-likelihoods of the LL-ACD models constructed using parameter estimates from the previous (current) estimation window. Within each panel and for each model specification, \overline{MSR} is the pooled mean squares of the residuals obtained from all 30 securities. LL/obs is the log-likelihood per price duration computed for the whole dataset. The three columns under the ‘In minus out’ header report the difference between the in-sample and out-of-sample \overline{MSR} and LL/obs. The best performing model in terms of out-of-sample, in-sample and in minus out \overline{MSR} and LL/obs. is highlighted in bold. For each panel, we perform modified Diebold-Mariano tests to test whether \overline{MSRs} differ significantly between models. The asterisks (circles) superscripts on the \overline{MSRs} indicate significantly smaller (larger) \overline{MSRs} compared against the \overline{MSRs} to the left. One to three symbols correspond to significance at 10%, 5% and 1%, respectively.

widens, since the optimal number of covariates in the -K model increases with the size of the estimation window.

2.6.3 Intraday Volatility Estimation with the LL-ACD Model

Tse and Yang (2012) show that ICV estimates obtained from ACD models can lead to precise intraday volatility estimates from their simulation and empirical findings. Recently, Li, Nolte, and Nolte (2018a) provide theoretical support to Tse and Yang’s (2012) findings, and prove that in a continuous martingale setting, the ICV estimator in (2.21) is a very good proxy of the integrated variance of the underlying price process over the price duration. In this section, we aim to show that the ICV estimator can indeed estimate intraday volatility, and the inclusion of MMS covariates can improve the quality of the intraday ICV volatility estimates. We also provide insights into modelling ultra high-frequency local volatility processes when the RV-type estimators cannot be reliably constructed due to a lack of data.

The ICV volatility estimates are constructed based on the residuals of the LL-ACD(1,1) models estimated from the previous section. We introduce some notation to describe the ICV estimates obtained from different models on each trading day. Let the subscript d denote the day index, which ranges from 1 to 1006 in our sample. Let

$\{t_{i,d}^{(\delta)}\}$ denote the arrivals of price events at day d , and let I_d denote the set of indices of the price events within day d . To construct an ICV estimates for a price duration on day d , we estimate the LL-ACD(1,1)-P, LL-ACD(1,1)-K and LL-ACD(1,1)-A models monthly, quarterly, half-yearly and yearly. For each stock, the above estimation frequencies correspond to 48, 12, 8 and 4 model estimations respectively. We use the superscript $M \in \{P, K, A\}$ to denote LL-ACD(1,1)-P, LL-ACD(1,1)-K and LL-ACD(1,1)-A models correspondingly. The superscript $m \in \{1, 3, 6, 12\}$ represents the size of the estimation window in months.

Let $\hat{\varepsilon}_{i,d}^{M,m}$ denote the i -th estimated residual at day d from model M with an estimation window of m months. For example, $\hat{\varepsilon}_{i,d}^{P,4}$ is the i -th estimated residual at day d from the plain LL-ACD(1,1) model estimated with a four-month estimation window. We use the notation $\hat{\theta}^{M,m}(d)$ to denote the parameter estimates from model M with estimation window size m at day d , which is identical for d within the same estimation window. ICV estimates for the i -th price duration at day d from model M with an estimation window of m months is defined as:

$$ICV_{i,d}^{M,m} \equiv -\ln \left(1 - \Phi(\hat{\varepsilon}_{i,d}^{M,m} / \hat{\sigma}_{\varepsilon}^{M,m}(d)) \right) \left(\frac{\delta}{P(t_{i-1,d}^{(\delta)})} \right)^2. \quad (2.21)$$

To assess the quality of the intraday ICV volatility estimates, we need to construct a valid benchmark intraday volatility estimator. In theory, the ICV estimator can provide integrated variance estimates for every price duration, which in our setting is roughly once every 5 minutes. However, we are unable to evaluate the performance of the ICV estimators at such a high frequency because the construction of a benchmark intraday RV-type estimator becomes very problematic due to a lack of data. To circumvent this problem, we split a trading day into 6 equally spaced intervals of 3900 seconds, and construct an intraday RK estimator¹¹ based on the transactions within each interval. We only consider the security SPY because it is the most liquid security in our dataset. Each hourly window on average has 1500 transactions, which is sufficient to construct an RK measure. We denote this intraday RK estimator using $RK_{j,d}$, in which $j = 1 : 6$ refers to each 3900-second interval of the trading day. Thus $RK_{j,d}$ is an estimator of the integrated variance of the interval $(3900(j-1), 3900j]$ for which we measure calendar time in seconds.

We then compare the performance of the ICV estimates with the intraday RK measures by constructing the ICV measures for each j . Denote $I_{j,d}$ the set of

¹¹Following Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), we use a non flat-top Parzen kernel with optimal bandwidth selection and tick-by-tick sampling for each hourly interval. The RK estimates are computed using Kevin Sheppard's MFE toolbox.

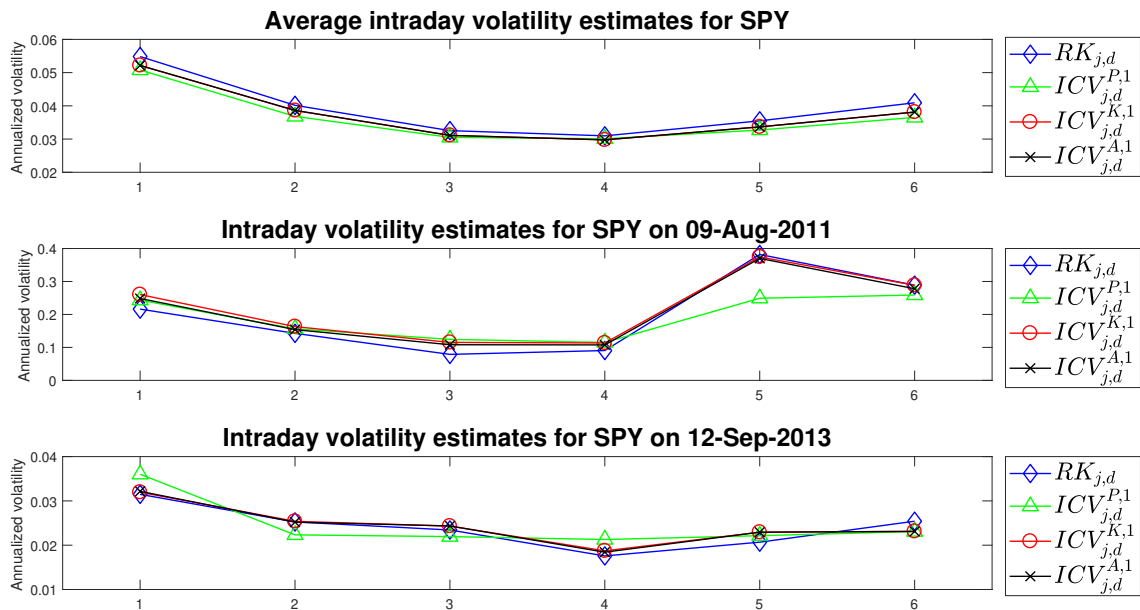
price event index such that $t_i^{(\delta)} \in (3900(j-1), 3900j]$, $\forall i \in I_{j,d}$. We construct the corresponding $ICV_{j,d}^{M,m}$ estimator for $M \in \{P, K, A\}$ and $m \in \{1, 3, 6, 12\}$:

$$ICV_{j,d}^{M,m} \equiv \sum_{i \in I_{j,d}} ICV_{i,d}^{M,m} + Q_{j,d}, \tag{2.22}$$

where $Q_{j,d}$ is a boundary correction term which ensures that $ICV_{j,d}^{M,m}$ is an estimate of the integrated variance over the interval $(3900(j-1), 3900j]$. We present the construction of $Q_{j,d}$ in Appendix B.6. Note that the following analyses are based on the annualized version of intraday volatility estimates computed by the square root of a given intraday volatility measure multiplied by 252.

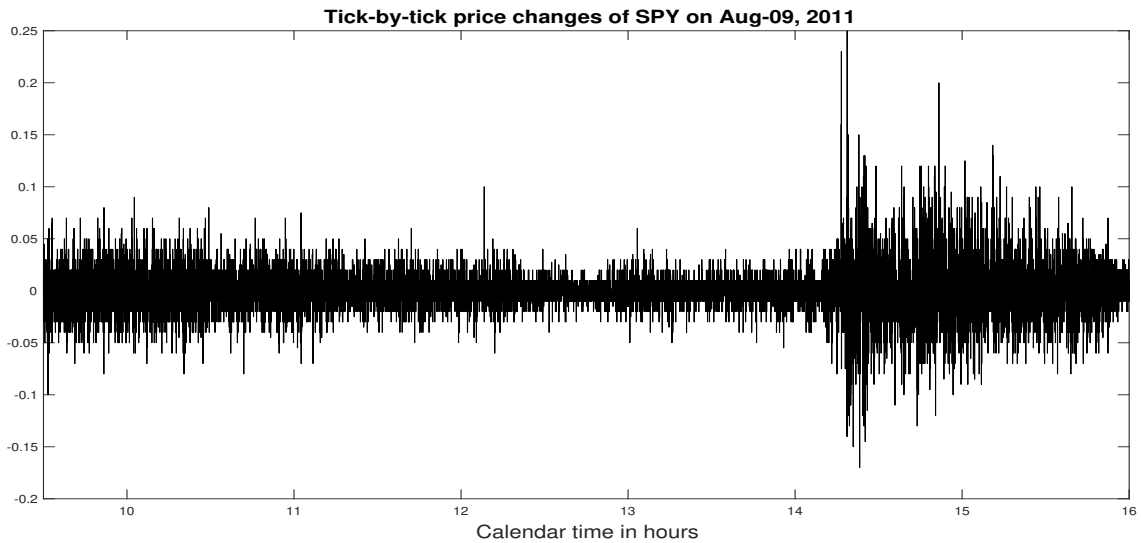
We plot some examples of the intraday volatility estimates from $RK_{j,d}$ and $ICV_{j,d}^{M,m}$ considered here using monthly estimations in Figure 2.8. The upper panel shows the average intraday volatility estimates averaged for each hourly interval. It is obvious that the average intraday volatility estimates reveal a U-shaped diurnal pattern. In the middle and lower panels, we select two representative days with relatively higher and lower daily volatility estimates. The middle panel is of particular interest, as we identify a sudden volatility increase in interval 5, which corresponds to the time period 13:50-14:55. We present a plot of the tick-by-tick price change for SPY on 09-Aug-2011 in Figure 2.9 to illustrate this abrupt change in volatility in the afternoon period.

Figure 2.8 Example of intraday volatility estimates based on ICV and RK for SPY



Note: The x-axis denotes the index of the intraday intervals j . In the first panel, we average the volatility estimates for each j and each model.

Figure 2.9 Tick-by-tick price change for SPY on 09-Aug-2011



Note: The x-axis denotes the calendar time in hours.

Figure 2.8 has some important implications. Firstly, from the upper panel it is clear that the ICV estimators are capable of providing accurate intraday volatility estimates, even without the inclusion of MMS covariates. This corroborates the findings in Tse and Yang (2012) and provides empirical evidence to support the theoretical considerations in Li, Nolte, and Nolte (2018a). Secondly, From the middle panel, we see that the RK measure reflects the sudden volatility burst as observed in Figure 2.9. Volatility estimates from $ICV_{j,d}^{K,1}$ and $ICV_{j,d}^{A,1}$ follow this sudden increase in RK accurately, while the $ICV_{j,d}^{P,1}$ estimates fail to capture this sudden change in intraday volatility. This is strong supporting evidence that the inclusion of MMS covariates indeed improves the intraday volatility estimates from the ICV estimators. This result, however, is not surprising because the plain ICV estimator does not use any contemporaneous information, so the abrupt price change is unpredictable from the past data. Finally, in the lower panel, all three ICV estimators closely follow the path of RK, but the ICV estimators with the MMS covariates seem to have a higher correlation with the RK estimates. To summarize the co-movements between the four intraday estimators, we present a correlation table for the four estimators in Table 2.7, which confirms that all three ICV estimators have very high correlations with RK and the ICV estimators with the inclusion of MMS covariates perform better than the plain version.

To formally assess the performance of the ICV estimators, we compute the mean squared errors (MSEs) of the ICV estimators with respect to the RK for each interval

Table 2.7 Averaged correlation table of RK and monthly estimated intraday ICV volatility estimates for SPY

	RK_d	$ICV_{j,d}^{P,1}$	$ICV_{j,d}^{K,1}$	$ICV_{j,d}^{A,1}$
$RK_{j,d}$	1.0000			
$ICV_{j,d}^{P,1}$	0.9415	1.0000		
$ICV_{j,d}^{K,1}$	0.9753	0.9702	1.0000	
$ICV_{j,d}^{A,1}$	0.9767	0.9693	0.9985	1.0000

Note: for SPY, we compute the correlation matrix of the matrix $\{RK_{j,d}, ICV_{j,d}^{P,1}, ICV_{j,d}^{K,1}, ICV_{j,d}^{A,1}\}_{j=1:6, d=1:1006}$.

Table 2.8 Comparison of MSEs of intraday ICV volatility estimates

$j =$	$MSE_j^{P,1}$	$MSE_j^{K,1}$	$MSE_j^{A,1}$	$MSE_j^{P,3}$	$MSE_j^{K,3}$	$MSE_j^{A,3}$	$MSE_j^{P,6}$	$MSE_j^{K,6}$	$MSE_j^{A,6}$	$MSE_j^{P,12}$	$MSE_j^{K,12}$	$MSE_j^{A,12}$
1	7.075	6.127**	5.986	6.800	5.995**	5.880	6.717	5.728**	5.739 ^{oo}	6.639	5.854*	5.846
2	3.886	2.448***	2.371*	3.889	2.422***	2.387	3.916	2.405***	2.403	3.875	2.480***	2.477
3	3.879	2.230***	2.115*	3.992	2.151***	2.041*	4.070	2.045***	2.042	4.149	2.090***	2.091
4	3.434	2.309***	2.248	3.462	2.298***	2.171***	3.586	2.165***	2.159**	3.649	2.222***	2.220
5	8.439	2.877***	2.794**	8.600	3.073***	2.965**	9.070	2.971***	2.967	9.503	3.194**	3.194
6	5.936	3.022***	2.952**	6.157	3.089***	3.047	6.265	3.223***	3.217***	6.833	3.361***	3.353***
Overall	5.441	3.169***	3.078**	5.483	3.171***	3.082***	5.604	3.089***	3.088	5.775	3.200***	3.197

Note: The MSEs are multiplied by 10^5 . For each j , we compute $MSE_j^{M,m}$ for each $M \in \{P, K, A\}$ and $m \in \{1, 3, 6, 12\}$ according to (2.23). For each m , we perform modified Diebold-Mariano tests to test whether $MSE_j^{P,m} = MSE_j^{K,m}$ and $MSE_j^{K,m} = MSE_j^{A,m}$. The asterisks (circles) superscripts on the MSEs indicate significantly smaller (larger) MSEs compared against the MSE to the left. One to three symbols correspond to significance at 10%, 5% and 1%, respectively.

using:

$$MSE_j^{M,m} = \frac{1}{1006} \sum_{d=1}^{1006} (ICV_{j,d}^{M,m} - RK_{j,d})^2. \quad (2.23)$$

We present $MSE_j^{M,m}$ for all j , M and m in Table 2.8, and test whether for each m and j , $MSE_j^{P,m} = MSE_j^{K,m}$ and $MSE_j^{K,m} = MSE_j^{A,m}$, with the modified Diebold-Mariano test (Diebold and Mariano, 1995; Harvey, Leybourne, and Newbold, 1997). The table provides interesting insights into the impact of MMS covariates on the intraday ICV estimates. Firstly we see that for interval 5, $MSE_5^{P,m}$ is considerably larger than $MSE_5^{K,m}$ and $MSE_5^{A,m}$ due to the effect of afternoon volatility spikes as described in Figures 2.8 and 2.9. This clearly indicates that the inclusion of contemporaneous MMS covariates reflects the contemporaneous signal of intraday volatility shifts, which greatly improves the performance of intraday ICV volatility estimates.

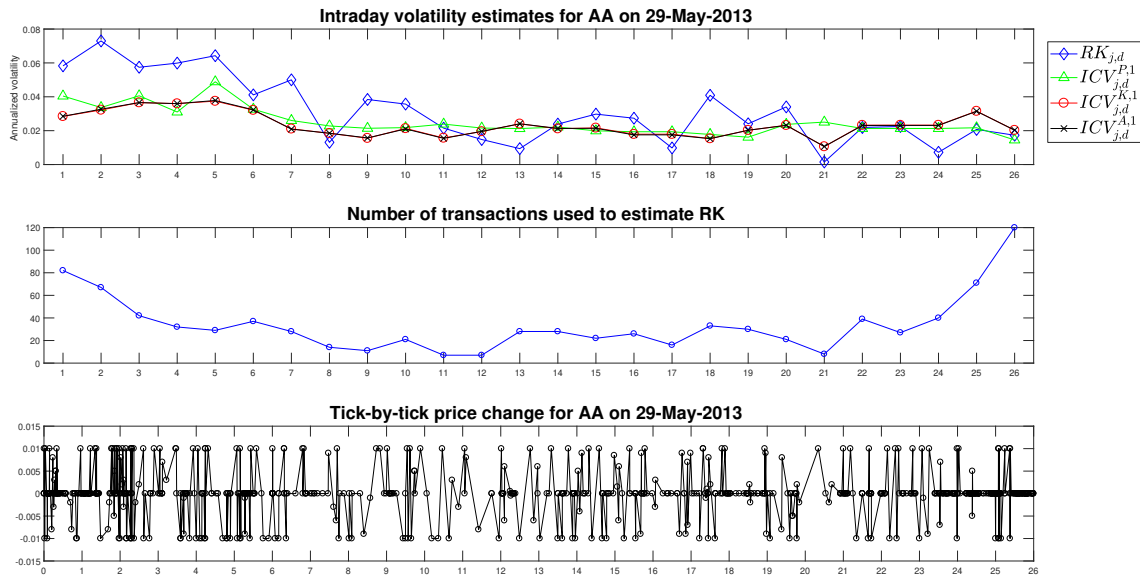
Interestingly, $MSE_j^{K,m}$ and $MSE_j^{A,m}$ also seem to possess some diurnal patterns, as $MSE_1^{K,m}$ and $MSE_1^{A,m}$ are generally larger than those MSEs for other intervals. We suspect that this is due to possible intraday structural breaks in the MMS parameters. As pointed out by Li, Nolte, and Nolte (2018c), price durations can have intraday regime-switching relationships with MMS covariates. These relationships can behave very differently at the start of the trading day and switch to another regime for the rest of the trading day. Therefore it is possible that our specification of the MMS parameters is prone to intraday parameter instability for the observations within interval 1, resulting in less precise volatility estimates for the period 9:30-10:35. The ICV estimates are reasonably stable across different m , and overall, intraday volatility estimates from LL-ACD(1,1)-K and -A models are generally very similar and both outperform those from the plain LL-ACD(1,1) model. Also it is clear that the intraday ICV estimates from the -K and -A models are more robust to changes in the estimation window by comparing the MSEs across different estimation window sizes. This is a reflection of the robustness check results in Appendix B.4.

We would like to stress that the liquidity of the securities limits the use of RV-type nonparametric volatility estimators due to a lack of data, while the ICV estimators can always be constructed as long as we can obtain a time series of price durations and the associated MMS covariates. To illustrate this point, we construct 15-minute intraday volatility estimates for all stocks, models and sizes of estimation windows. We plot an example of the 15-minute volatility estimates in Figure 2.10.

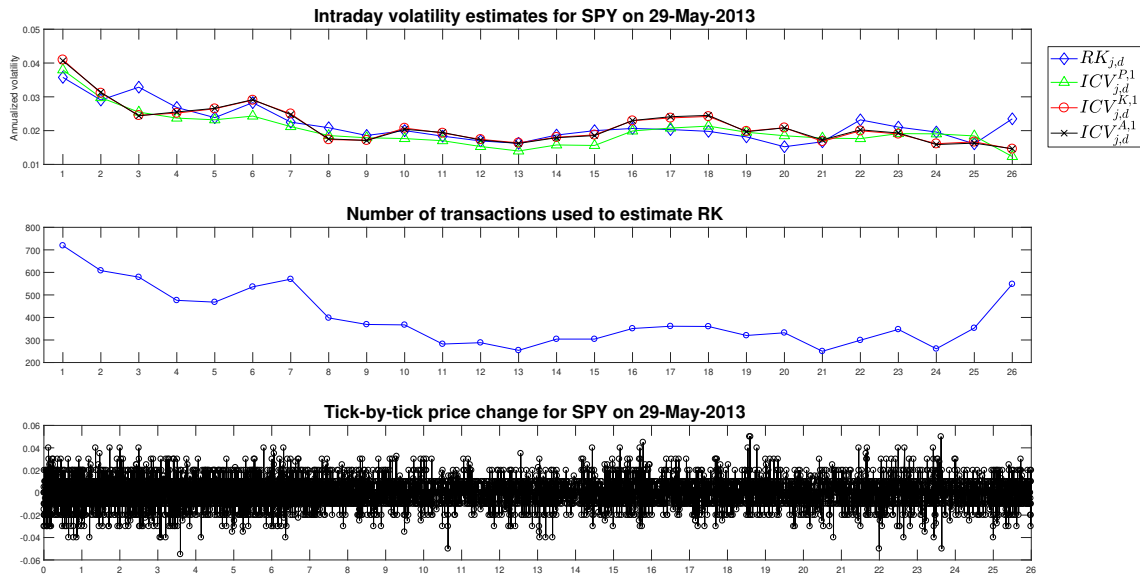
From Figure 2.10, we see that the RK estimates diverge from the ICV estimates and appear more noisy for AA, but are much closer to the ICV estimates for SPY. This

Figure 2.10 15-minute volatility estimates for AA and SPY on 29-Mar-2013

Panel 1: AA



Panel 2: SPY



Note: For each panel, in the upper two plots, the x-axis denotes each j representing 26 fifteen-minute interval used to construct the volatility measures. In the third plot, the interval $(j - 1, j]$ refers to the j -th 15-minute interval.

is related to the number of transactions used to obtain the RK estimates, shown in the middle figure within each panel. For AA, the majority of RK estimates use less than 50 observations, thus are much less precise than those for SPY. For the ICV estimates, we assert that their precisions depend on the number of price durations within the estimation window and the goodness-of-fit of the ACD model rather than the number of observations within each 15-minute window. Thus, even with a monthly estimation window, we can obtain more precise intraday volatility estimates compared to the RV-type estimators as more data is involved in the estimation of local volatility for each interval.

As the true volatility process is not observable, we cannot prove that the ICV estimates are more precise than the RK at a very high frequency, apart from the theoretical evidence in Li, Nolte, and Nolte (2018a). However, since RK can be regarded as a good proxy of volatility when the number of observations used is large, we would expect that: (1) The co-movement between RK and ICV estimates increases as the number of observations used to construct RK increases. (2) The inclusion of MMS covariates improves the co-movement between RK and ICV estimates only when RK is precise. (3) In addition to the optimally selected MMS covariates, adding all MMS covariates does not have a big impact on the co-movement between RK and ICV estimates. To conserve space, we defer this analysis to Appendix B.1. Our fix-effect regression results are highly consistent with the above arguments, which provide indirect evidence supporting our claim that the ICV estimates are more precise than RK when the number of observations is small, and that the inclusion of MMS covariates indeed improves the precision of intraday volatility estimates.

As a summary to this section, we list three key implications of our results. Firstly, contemporaneous information in the MMS covariates reflects sudden shocks to the price volatility, which can significantly improve intraday ICV estimates. This is driven by improvements in the in-sample goodness-of-fit of the LL-ACD model, and also leads to more accurate daily volatility estimates.¹² Secondly, with the inclusion of MMS covariates, the LL-ACD models are more robust to changes in the estimation window, thus one can potentially use a larger estimation window to obtain more precise volatility estimates. Finally, the ICV estimators can be constructed even when the RK estimator cannot be constructed reliably, which provides intraday volatility estimates at a much higher resolution.

¹²Detailed analysis on a daily level can be found in Appendix B.5.

2.7 Concluding Remarks

This chapter examines the intraday dynamics of a local volatility measure and its interactions with a universe of MMS covariates, including both contemporaneous and lagged version of trading volume, trading activity, bid-ask spread, total quote depths, quote difference, order imbalance and order flow. We adopt the LL-ACD model of Allen, Chan, McAleer, and Peiris (2008), which is a parametric model that can provide high-frequency volatility estimation with the inclusion of MMS covariates, and also enjoys the QML property.

We use TAQ data for thirty securities from 2011 to 2014. By applying the best subset regression, we find that in descending order, contemporaneous versions of number of trades, order flow, total quote depths, quote difference, bid-ask spread, order imbalance and volume are the first 7 most important variables for volatility estimation over all securities. The one duration-lagged variables are in general less important than their contemporaneous counterparts, but the rankings vary across securities. Our best subset regression approach also selects the most important MMS covariates that should be used for volatility estimation.

The estimation of the LL-ACD models shows that the model successfully captures the intraday pattern of the price durations. Most of the signs of the estimates of contemporaneous MMS variables are consistent with theoretical predictions and highly significant. We find that the inclusion of optimally selected MMS covariates significantly improves both the in-sample and out-of-sample fit of the LL-ACD model, while including all MMS covariates does not further improve the in-sample fit of the LL-ACD model substantially. Also, models with the inclusion of MMS covariates are more robust to a larger estimation window.

We proceed to show that the inclusion of MMS covariates improves the precision of intraday volatility estimates by benchmarking on a realized kernel estimator. We find that the information in the contemporaneous MMS covariates can correctly reflect sudden changes in the intraday volatility process, leading to more precise intraday volatility estimates. We also provide evidence supporting the argument that the ICV estimators are reliable even when the RK estimator cannot be constructed due to a lack of data. In general we find that RK estimates are close to ICV estimates whenever RK is precise. The inclusion of optimally chosen MMS covariates significantly improve the co-movement between RK and ICV estimates, while including all MMS covariates does not further improve the results.

The key message from our study is that the ICV volatility estimator is superior to the RV-type estimator in estimating intraday volatility, especially when the number of transactions to construct a realized measure is scarce. The parametric structure of the ICV estimator allows us to overcome the limitation of intraday data by using data beyond the volatility estimation window. Moreover, it allows us to incorporate contemporaneous observation information in the MMS covariates to provide more accurate intraday volatility estimates. To sum up, the ICV volatility estimator based on the ACD model coupled with MMS covariates is an important approach to intraday volatility modelling that should not be overlooked.

Chapter 3

High-Frequency Volatility Modelling: A Markov-Switching Autoregressive Conditional Intensity Model

3.1 Introduction

Since the seminal work of Engle and Russell (1998), a growing amount of the literature has emerged on parametric modelling of intraday financial data. An important strand of the literature concentrates on the parametric modelling of intraday price volatility. Contrary to the Realized Volatility (RV)-type estimators (Andersen, Bollerslev, Diebold, and Labys, 2001) which are non-parametric by design, Engle and Russell's (1998) parametric approach relies on construction of volatility estimators based on the conditional intensity process of the absolute price change point process (Gerhard and Hautsch, 2002). This type of volatility estimator has been shown to perform at least as good as RV-type estimators both theoretically and empirically (Nolte, Taylor, and Zhao, 2018; Tse and Yang, 2012). More importantly, the parametric design allows for the inclusion of other market microstructure covariates in the estimation of intraday volatility. This feature enables researchers to analyse the interaction between volatility and other aspects of the market on an intraday level without compromising the quality of volatility estimation, as existing models are either not capable of incorporating other variables (e.g. RV framework), or are considered inappropriate for intraday volatility estimation (e.g. intraday GARCH framework).

The models used for the intensity-based volatility estimation, namely variants of the Autoregressive Conditional Duration (ACD) model by Engle and Russell (1998)

and the Autoregressive Conditional Intensity (ACI) model by Russell (1999), are constructed by specifying an autoregressive structure to the duration or the intensity process. These specifications successfully capture the high persistence and over-dispersion feature of empirical price durations, and can be extended to accommodate long memory and asymmetric features, similar to the extensions of GARCH models (see Hautsch (2012) for a comprehensive summary of the extensions of the ACD and ACI models). However, the existing extensions of ACD and ACI models do not provide a satisfactory solution to incorporate a Markov-Switching structure.¹

In this chapter, we develop the Markov-Switching Autoregressive Conditional Intensity (MS-ACI) model. A well-known problem in the estimation of a Markov-switching GARCH-type model is that the observed likelihood becomes dependent on the full path of the latent states. This renders standard maximum likelihood estimation of MS-GARCH model intractable and existing solutions either rely on a simplification of the structure (MS-ARCH by Cai (1994) and Hamilton and Susmel (1994)), a collapsing approximation of the variance equation (Dueker, 1997; Gray, 1996; Haas, Mittnik, and Paoletta, 2004; Klaassen, 2002), or Bayesian estimation (Bauwens, Dufays, and Rombouts, 2014; Bauwens, Preminger, and Rombouts, 2010; Billio, Casarin, and Osuntuyi, 2014). By applying the Stochastic Approximation Expectation Maximization (SAEM) algorithm (Celeux and Diebolt, 1992), we show that our model can be reliably estimated within a frequentist's approach without any structural simplification. The algorithm allows us to overcome the path-dependency problem and obtain maximum likelihood estimates of the model.

We apply the MS-ACI model to examine the volume-volatility relationship on a high-frequency level. This is motivated by a large body of market microstructure literature that suggests a regime-switching feature in the observed transaction process. Both asymmetric information trading model (Copeland and Galai, 1983; Easley, Kiefer, O'Hara, and Paperman, 1996) and strategic trading model (Admati and Pfleiderer, 1988; Kyle, 1985) assume that there are two types of market participants: the informed traders who only trade when they possess advantageous information, and the uninformed traders who trade for liquidity or other reasons. The trading volume is frequently used as a proxy of the unobservable information flow in both empirical and theoretical investigations (e.g. Copeland (1976), Epps and Epps (1976), Tay,

¹To the best of our knowledge, the MS-ACD model was proposed by Hujer, Vuletic, and Kokot (2002) and the MS-ACI model has not yet been documented. However, Hujer, Vuletic, and Kokot's (2002) method relies on a simplification of the model using Gray's (1996) approximation, which has a serious drawback of analytical intractability, and the quality of this approximation has not yet been verified.

Ting, Tse, and Warachka (2009), Easley, López de Prado, and O'Hara (2012)). From the information market setting, volume submitted by the informed traders will have a much larger impact on price volatility than those from uninformed traders. The decomposition of trading volume based on its interaction with volatility has been documented by Bessembinder and Seguin (1993), Andersen (1996), Aragó and Nieto (2005) and Hussain (2011). Nevertheless, empirical findings in these studies are limited by their GARCH framework.

In our empirical analysis we apply the MS-ACI model on the price duration data computed from the Trade and Quote (TAQ) data of ten frequently traded securities (including a market index ETF, SPY) for the year 2016, timestamped at milliseconds. We identify two distinct regimes in the high-frequency relationship between price duration and the associated trading volume: a dominant regime spreading evenly across the trading day, and a minor regime with an L- or U-shaped diurnal pattern that resembles the well-documented pattern of price volatility (Andersen and Bollerslev, 1997b). We establish the connection between volume-volatility relationship and correlation between volume and price durations, and show that the minor regime corresponds to price durations with high information content. Interestingly, the regime-switching behaviour is not observable on the SPY data, suggesting that our findings are related to firm-specific news arrivals. Our post estimation diagnostics demonstrate that the two regimes are identified with very high confidence, and the Markov switching structure greatly improves the goodness-of-fit of the model compared to a single regime setup.

The main contributions of this chapter are three-folded: firstly we develop the MS-ACI model, which allows the detection of structural breaks on an intraday level with the inclusion of other covariates. The stationarity and moment conditions for the ACI and MS-ACI model are updated in this chapter, which augments the stationarity condition given in Hautsch (2012). Secondly, we are among the first to implement the SAEM algorithm to the Markov-switching autoregressive models. We provide simulation evidence on the reliability of this algorithm, and examine some theoretical properties of the MS-ACI model through the significance of the regimes. Thirdly, our application to high-frequency price duration data provides new insights in the volume-volatility relationship and the information arrival theory in a high-frequency context. We provide a more intuitive method to disentangle the heterogeneous information content in trading volume, as well as a measure of the probability of information shocks to the market, and of the volatility response to the information shocks.

The rest of this chapter is structured as follows: Section 3.2 introduces some basic point process theories and the original ACI model. In Section 3.3, the specification and estimation technique of the MS-ACI model are discussed. Simulation evidence is provided in Section 3.4, with the empirical application following in Section 3.5. Section 3.6 concludes.

3.2 Conditional Intensity Modelling

3.2.1 Basic Point Process Theory

This section briefly summarizes fundamental point process theory following the discussion in Hautsch (2012). For an in-depth textbook treatment we refer to Karr (1991) among others.

Let t denote the physical time and let the sequence $\{t_i\}_{i=1:T}$ denote the arrival time of the i -th event, subject to $0 < t_i < t_{i+1}, \forall i$.² For a sample size T , the complete observed sequence of the point process can be denoted as $\{t_i\}_{i=1}^T$. We only consider a univariate point process for the conciseness of notation.

A point process $\{t_i\}_{i=1}^T$ can be characterized by three different processes. The first one is the counting process, denoted as $N(t) := \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}}$ for the right-continuous version and $\check{N}(t) := \sum_{i \geq 1} \mathbb{1}_{\{t_i < t\}}$ for the left-continuous one. Thus, these two functions are step functions, and jump upward at (or after) the occurrence of event at time t_i . The second process is the duration process defined as $x_i = t_i - t_{i-1}$ for $i > 1$ and $x_1 = t_1$. The backward recurrence time is therefore defined as the function $x(t) = t - t_{\check{N}(t)}$, which is the time elapsed since the last event. Let \mathcal{F}_t denote the information set available till time t , which is also known as the filtration of the point process. The \mathcal{F}_t -conditional intensity representation of a point process is given by:

Definition 3.1. (Hautsch 2012:pp71) *Let $N(t)$ be a simple point process on $[0, \infty)$ that is adapted to some history \mathcal{F}_t and assume that $\lambda(t; \mathcal{F}_t)$ is a positive-valued process with sample paths that are left-continuous and have right-hand limits. Then the process*

$$\lambda(t; \mathcal{F}_t) \approx \lambda(t+, \mathcal{F}_t) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{E}[N(t+\Delta) - N(t) | \mathcal{F}_t], \quad \lambda(t+) > 0, \forall t, \quad (3.1)$$

²We assume that events do not occur simultaneously throughout this chapter, and this type of point process is referred to as *simple point process*.

with $\lambda(t+, \mathcal{F}_t) := \lim_{\Delta \downarrow 0} \lambda(t + \Delta, \mathcal{F}_t)$, is called the \mathcal{F}_t -conditional intensity process of the counting process $N(t)$.

The \mathcal{F}_t -conditional intensity process is of particular importance because it is closely related to the likelihood function of a point process. Another important feature of the conditional intensity process is that, it is directly connected to martingale-based point process theory. This is due to the fact that any \mathcal{F}_t -adapted simple point process as described above is a \mathcal{F}_t -submartingale³, and according to the Doob-Meyer decomposition, the following equation holds:

$$N(t) = M(t) + \Lambda(t), \quad (3.2)$$

in which $M(t)$ is a zero mean \mathcal{F}_t -martingale, and $\Lambda(t)$ is a unique \mathcal{F}_t -predictable cumulative process called the *compensator* of the semimartingale. This *compensator* is the integrated conditional intensity defined as:

$$\Lambda(t) = \int_0^t \lambda(s; \mathcal{F}_s) ds. \quad (3.3)$$

More importantly, according to the Random Time Change theorem (RTCT hereafter, see e.g. Bousher (2007)), the time-changed compensator process $\{\Lambda(t_i)\}_{i=1:T}$ is a unit rate independent Poisson process. Let us define the inter-event compensator as $\Lambda(s, t) = \int_s^t \lambda(u; \mathcal{F}_u) du$, then this transformed process $\{\Lambda(t_{i-1}, t_i)\}$ is the duration of a unit rate independent Poisson process and it follows that:

$$\Lambda(t_{i-1}, t_i) \sim i.i.d. Exp(1). \quad (3.4)$$

This property serves as a crucial tool in constructing intensity-based models, and is used in generating residuals and diagnostic tests for point processes.

The log-likelihood function of a point process based solely on the conditional intensity process derived by Karr (1991) is given by:

$$\ln \mathcal{L}(\boldsymbol{\theta}; \mathbb{Y}) = \sum_{i=1}^T [-\Lambda(t_{i-1}, t_i) + \ln \lambda(t_i | \mathcal{F}_{t_i})]. \quad (3.5)$$

Here $\mathbb{Y} = \{t_i\}_{i=1:T}$ and $\boldsymbol{\theta}$ is the general parameter vector for some parametrized conditional intensity process. The log-likelihood function has a very intuitive interpretation. Heuristically, it is the log of the probability of no events occurring

³A \mathcal{F}_t -submartingale is defined as an \mathcal{F}_t -adapted process $N(t)$ with $E[N(t)] < \infty$ satisfying the condition $E[N(t) | \mathcal{F}_s] \geq N(s)$ almost surely. When the inequality is replaced by equality, $N(t)$ is said to be a \mathcal{F}_t -martingale.

between time t_{i-1} and t_i given by $\exp[-\Lambda(t_{i-1}, t_i)]$ multiplied by the probability of events occurring at exactly t_i given by $\lambda(t_i | \mathcal{F}_t)$.

3.2.2 The Original ACI Model

In the original paper, Russell (1999) documents a bivariate ACI model with no seasonality adjustment to the raw data. In this section, we describe the main ingredients of the original ACI model in a univariate framework as in Bauwens and Hautsch (2006).

The ACI model is a fully parametric model, which specifies the conditional intensity in a multiplicative form as follows:

$$\lambda(t | \mathcal{F}_t) = \Phi(t) \lambda_0(t). \tag{3.6}$$

The $\Phi(t)$ is an autoregressive component that could include time-varying covariates, and $\lambda_0(t)$ is the baseline intensity function⁴. To ensure the non-negativity of the conditional intensity, $\Phi(t)$ is usually parametrized as the exponential form of an ARMA-type structure, as an example:

$$\Phi(t) = e^{\tilde{\Phi}_{N(t)+1} + \eta' Z(t)}, \tag{3.7}$$

$$\tilde{\Phi}_i = \sum_{j=1}^q \alpha_j \tilde{\epsilon}_{i-j} + \sum_{k=1}^p \beta_k \tilde{\Phi}_{i-k}, \tag{3.8}$$

in which $Z(t)$ is a matrix of covariates (can include both time-varying and time-invariant covariates) and η is the corresponding parameter vector. $\tilde{\Phi}_i$ is a zero mean ARMA-type process, and the weak stationarity condition for $\tilde{\Phi}_i$ is that all the roots of the polynomial $\beta(z) = 1 - \sum_{k=1}^q \beta_k z^k$ lie outside the unit circle. The innovation terms, $\tilde{\epsilon}_i$, can be defined as:

$$\tilde{\epsilon}_i = -\gamma - \ln \epsilon_i = -\gamma - \ln \Lambda(t_{i-1}, t_i) = -\gamma - \ln \int_{t_{i-1}}^{t_i} \lambda(u | \mathcal{F}_u) du, \tag{3.9}$$

in which γ is the Euler-Mascheroni constant. According to (3.4), since $\Lambda(t_{i-1}, t_i)$ is i.i.d. exponential if the ACI model is correctly specified, the logarithm of an exponential variable follows an extreme value type-I distribution with mean γ hence

⁴Additional components can be included multiplicatively to capture other effects of interest on the conditional intensity, e.g. a seasonality component which is a deterministic function of calendar time, a daily effect which only changes interdaily, etc. To discuss the fundamental structure of the ACI model, we ignore these functional components for the moment.

$\tilde{\varepsilon}_i$ is a zero mean martingale⁵.

The baseline intensity component $\lambda_0(t)$ can be specified in various ways. Because a closed form solution of the integrated conditional intensity and thus the error term is more computationally convenient, the following specifications are popular: (1) the exponential baseline $\lambda_0(t) = e^w$, (2) the Weibull baseline $\lambda_0(t) = ae^{wa}x(t)^{a-1}$, (3) the Burr-type baseline $\lambda_0(t) = e^\kappa \frac{ax(t)^{a-1}}{e^{-wa} + x(t)^a}$. The model can be easily estimated by maximum likelihood with the log-likelihood function given in (3.5).

3.2.3 Stationarity of the ACI Model

The stationarity concept discussed in this chapter refers to weak (covariance) stationarity unless stated otherwise. Previous studies focus on discussing the stationarity of the $\tilde{\Phi}_i$ component (see Russell (1999), Hautsch (2012) for instance), and the stationarity condition of $\tilde{\Phi}_i$ is considered as a sufficient stationarity condition for the ACI model. We show that with the error term specified as in (3.9), the stationarity of $\tilde{\Phi}_i$ is actually necessary instead of sufficient for the conditional intensity process $\lambda(t_i|\mathcal{F}_{t_i})$ or the duration process x_i to be stationary. To illustrate this, consider an ACI(0,1) model with $\tilde{\Phi}_i = 0.5\tilde{\varepsilon}_{i-1}$ and an exponential baseline function $\lambda_0(t) = e^w$. Clearly $\tilde{\Phi}_i$ is both strictly and weakly stationary because it is an i.i.d. zero mean martingale with finite second moment, therefore $\lambda(t_i|\mathcal{F}_{t_i})$ is strictly stationary. We can write $\tilde{\Phi}_i = -0.5\gamma - \ln \varepsilon_{i-1}^{0.5}$, in which $\varepsilon_{i-1} = \Lambda(t_{i-2}, t_{i-1}) \sim \text{Exp}(1)$ according to the RTCT, and $\lambda(t_i|\mathcal{F}_{t_i}) = e^{\tilde{\varepsilon}_{i-1}+w} = \frac{e^{-\gamma+w}}{\varepsilon_{i-1}^{0.5}}$. The second unconditional moment of $\lambda(t_i|\mathcal{F}_{t_i})$ is thus given by:

$$\mathbf{E}[\lambda(t_i|\mathcal{F}_{t_i})^2] = \mathbf{E}\left[\frac{e^{2(-\gamma+w)}}{(\varepsilon_{i-1}^{0.5})^2}\right] = e^{2(-\gamma+w)} \mathbf{E}[\varepsilon_{i-1}^{-1}]. \quad (3.10)$$

The expectation can be evaluated analytically as:

$$\mathbf{E}[\varepsilon_{i-1}^{-1}] = \int_0^\infty x^{-1} e^{-x} dx = \Gamma(0) = \infty, \quad (3.11)$$

in which $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ is the Gamma function. Thus, $\lambda(t_i|\mathcal{F}_{t_i})$ does not have a finite second moment and is not weakly stationary. The duration process, however,

⁵Alternative specifications are also possible, for example $\tilde{\varepsilon}_i = 1 - \ln \Lambda(t_{i-1}, t_i)$. However, taking the log of the integrated intensity stabilizes the model in situations in which there are potential underflow or overflow risks. Moreover, it is easier to derive theoretical properties for the log integrated intensity residuals, as will be shown later.

is stationary in this case. This is because:

$$\int_{t_{i-1}}^{t_i} \lambda(t_i | \mathcal{F}_{t_i}) = \varepsilon_{i-1}^{-0.5} e^{-\gamma+w} x_i = \varepsilon_i \sim i.i.d. Exp(1), \quad (3.12)$$

which is due to the RTCT. Therefore, it is easy to show that:

$$E[x_i^2] = e^{-2(-\gamma+w)} E[\varepsilon_{i-1}] E[\varepsilon_i^2] = e^{-2(-\gamma+w)}, \quad (3.13)$$

which proves that x_i is weakly stationary. If we instead let $\tilde{\Phi}_i = -0.5\tilde{\varepsilon}_{i-1}$, then the weak stationarity for duration and intensity process is reversed. This example has two main implications: (1) a weakly stationary intensity process does not imply a weakly stationary duration process, and vice versa; (2) extra constraints on the ARMA parameters are needed for both the duration and intensity processes to be stationary in addition to the stationarity condition for $\tilde{\Phi}_i$, thus the covariance stationarity condition serves only as a necessary but not sufficient condition for the ACI model. Concluding from the above example, we present the following stationarity condition for both the intensity and duration process from the ACI model with an exponential baseline:

Proposition 3.1. *For an ACI(p,q) model defined as in (3.7), (3.8), (3.9), and $\lambda_0(t) = e^w$, the necessary and sufficient conditions for the weak stationarity of the conditional intensity and duration process are:*

1. *All roots of the polynomial $\beta(z)$ lie outside the unit circle, where $\beta(z)$ is the polynomial in the lag operator form of $\tilde{\Phi}_i$: $\beta(L)\tilde{\Phi}_i = \alpha(L)\tilde{\varepsilon}_{i-1}$, in which $\beta(L) = 1 - \sum_{i=1}^p \beta_i L^i$ and $\alpha(L) = \sum_{j=1}^q \alpha_j L^j$.*
2. *If condition 1 holds, then $\tilde{\Phi}_i$ can be rewritten as an MA(∞) representation $\tilde{\Phi}_i = \psi(L)\tilde{\varepsilon}_{i-1}$, in which $\psi(L) = \frac{\alpha(L)}{\beta(L)}$. Let $\{\psi_j\}_{j=1,\dots,\infty}$ denote the j -th coefficient for the polynomial $\psi(L)$, then the following condition is required: $|\psi_j| < 0.5, \forall j$.*

Proof. See Appendix C.1. □

Intuitively, condition 1 can be viewed as the stationarity condition for the ARMA component and condition 2 is the moment condition since power transformed unit exponential variables do not necessary have finite second moments. For ACI models with a general baseline of the form $\lambda_0(t) = e^w g(x(t))$, a closed form anti-derivative $G(x) = \int g(x) dx$ with a well-defined inverse G^{-1} , the conditional intensity and duration

processes can be written respectively as:

$$\lambda(t_i | \mathcal{F}_{t_i}) = Cg(x_i) \prod_{j=1}^{\infty} \varepsilon_{i-j}^{-\psi_j}, \quad (3.14)$$

$$x_i = G^{-1} \left(C \varepsilon_i \prod_{j=1}^{\infty} \varepsilon_{i-j}^{\psi_j} \right). \quad (3.15)$$

Substituting (3.15) into (3.14) yields:

$$\lambda(t_i | \mathcal{F}_{t_i}) = Cg \left(G^{-1} \left(C \varepsilon_i \prod_{j=1}^{\infty} \varepsilon_{i-j}^{\psi_j} \right) \right) \prod_{j=1}^{\infty} \varepsilon_{i-j}^{-\psi_j}. \quad (3.16)$$

Thus in this general baseline case, non-constant baseline parameters can be involved in the moment condition which varies for different baselines and may not have an explicit solution. However, as long as the baseline function only involves power transformation of the duration (e.g. a Weibull baseline), analytical solutions of condition 2 can be derived for that particular specification of the ACI model.

3.3 Markov-Switching ACI Model

We propose the Markov-Switching ACI (MS-ACI) model as the counterpart of a Markov Switching GARCH and Markov Switching ACD models in an ACI framework. The specification, estimation and model diagnostics will be introduced in this section.

3.3.1 Specification of the MS-ACI Model

Let $\mathbb{S} = \{s_i\}_{i=1:T}, s_i \in \mathcal{M} = \{1 : M\}$ denote a sequence of M different discrete latent states attached to each duration process x_i of the univariate simple point process $\{t_i\}_{i=1:T}$. These states are assumed to follow a first order ergodic Markov chain with the transition probability $\mathbf{P}(s_i = m | s_{i-1} = l) = \pi_{lm}$ for $l, m \in \mathcal{M}$ and an invariant probability measure π_i . The MS(M)-ACI(p, q) is therefore specified as:

$$\lambda(t; \mathcal{F}_t) = \Phi(t) \lambda_0(t), \quad (3.17)$$

$$\Phi(t) = e^{\tilde{\Phi}_{\tilde{N}(t)+1}(s_{\tilde{N}(t)+1}) + \eta(s_{\tilde{N}(t)+1})Z(t)}, \quad (3.18)$$

$$\tilde{\Phi}_i(s_i) = \sum_{j=1}^q \alpha_j(s_i) \tilde{\varepsilon}_{i-j}(s_{i-j}) + \sum_{k=1}^p \beta_k(s_i) \tilde{\Phi}_{i-k}(s_{i-k}), \quad (3.19)$$

$$\tilde{\varepsilon}_i(s_i) = -\gamma - \ln \int_{t_{i-1}}^{t_i} \lambda(u; \mathcal{F}_u) du = -\gamma - \tilde{\Phi}_i(s_i) \int_{t_{i-1}}^{t_i} \ln \lambda_0(u) du, \quad (3.20)$$

$$\mathbf{P}(s_i = m | s_{i-1} = l) = \pi_{lm}, \quad l, m \in \mathcal{M} = \{1 : M\}, \quad (3.21)$$

in which γ is the Euler-Mascheroni constant, $\mathbf{Z}(t)$ is a matrix of some possible covariates, and $\boldsymbol{\eta}(s_{\check{N}(t)+1})$ is the corresponding regime-specific coefficient vector. The component $\boldsymbol{\eta}(s_{\check{N}(t)+1})\mathbf{Z}(t)$ enables state-specific relationships between the covariates and the intensity process. The regime-specific baseline function can be specified as an exponential baseline function:

$$\lambda_0(t) = e^{w(s_{\check{N}(t)+1})}, \quad (3.22)$$

or as a Weibull baseline function:

$$\lambda_0(t) = a(s_{\check{N}(t)+1})e^{w(s_{\check{N}(t)+1})a(s_{\check{N}(t)+1})}x(t)^{a(s_{\check{N}(t)+1})-1}, \quad (3.23)$$

in which $\check{N}(t) + 1$ for $t_{i-1} < t \leq t_i$ refers to the event counter i . We will denote all the dynamic and baseline parameters with the parameter vector $\boldsymbol{\theta}$ and the transition parameters of the Markov chain with the $(M \times M)$ matrix $\boldsymbol{\Pi}$. We restrict ourselves to these two types of baselines because these two types of baselines are in the exponential family, and the convergence of the Stochastic Approximation EM algorithm has been established by Delyon, Lavielle, and Moulines (1999) and Allasonnière, Kuhn, and Trouvé (2010) under this condition.

3.3.2 Stationarity Condition of the MS-ACI Model

Similar to a plain ACI model, weak stationarity of a MS-ACI model requires both the stationarity of the MS-ARMA component and a moment condition for the conditional intensity and duration. Francq and Zakoïan (2001) and Stelzer (2009) provide the strict and weak stationarity conditions for $\check{\Phi}_i$. For conciseness we do not present this condition, and refer the reader to Theorem 2.1 in Stelzer (2009) for a rigorous definition and proof of this stationarity condition. For the moment condition, consider a MS(M)-ACI(p, q) model with an exponential baseline function, $\lambda(t_i | \mathcal{F}_{t_i}) = e^{w(s_i) + \check{\Phi}_i}$. We write $\check{\Phi}_i$ in state space form $\bar{z}_i = \boldsymbol{\Theta}_i \bar{z}_{i-1} + \bar{\eta}_i$ with:

$$\bar{z}_i = \begin{bmatrix} \check{\Phi}_i \\ \check{\Phi}_{i-1} \\ \vdots \\ \check{\Phi}_{i-p+1} \\ \tilde{\boldsymbol{\varepsilon}}_i \\ \tilde{\boldsymbol{\varepsilon}}_{i-1} \\ \vdots \\ \tilde{\boldsymbol{\varepsilon}}_{i-q+1} \end{bmatrix}_{(p+q) \times 1}, \quad \bar{\eta}_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \tilde{\boldsymbol{\varepsilon}}_i \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(p+q) \times 1}, \quad (3.24)$$

and

$$\Theta_i = \begin{bmatrix} \beta_1(s_i) & \cdots & \beta_p(s_i) & \alpha_1(s_i) & \cdots & \alpha_q(s_i) \\ & & 0 & 0 & \cdots & 0 \\ & I_{p-1} & \vdots & \vdots & \ddots & \vdots \\ & & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & & & 0 \\ \vdots & \ddots & \vdots & & I_{q-1} & \vdots \\ 0 & \cdots & 0 & & & 0 \end{bmatrix}_{(p+q) \times (p+q)}, \quad (3.25)$$

in which I_k is the $k \times k$ identity matrix. Assume the stationarity condition of $\tilde{\Phi}_i$ is satisfied, then from Bougerol and Picard (1992), the unique stationary solution of \bar{z}_i is:

$$\bar{z}_i = \bar{\eta}_i + \sum_{j=1}^{\infty} \Theta_i \Theta_{i-1} \cdots \Theta_{i-j+1} \bar{\eta}_{i-j}. \quad (3.26)$$

Consider only the first element of the vector \bar{z}_i , it is again a product of an infinite number of power transformed i.i.d.-unit exponential variables. Let $\tilde{\psi}_{i,j}$ be the $(p+1)$ -th element on the first row of the matrix $\Theta_i \Theta_{i-1} \cdots \Theta_{i-j+1}$, then we have $\tilde{\Phi}_i = \sum_{j=1}^{\infty} \tilde{\psi}_{i,j} \tilde{\epsilon}_{i-j}$. Thus, conditioning on a particular realization of states $\{s_i\}_{i=1:T}$, it is required that $|\tilde{\varphi}_{i,j}| < 0.5, \forall i, j$ to guarantee the existence of the second moment for both the conditional intensity and duration processes. To derive the unconditional moment condition, one needs to integrate out all possible paths of the hidden states, which is analytically intractable. This is known as the path dependency problem, as will be introduced in detail in the next section. The stationarity condition for other baselines of the plain ACI model can be applied directly to the Markov switching case, thus we will not elaborate on this.

3.3.3 Model Estimation

Similar to the plain ACI model, one relies on a maximum likelihood estimator (MLE) to estimate the parameter vector θ and the state probabilities Π . A straightforward implementation of MLE calculates the argument that maximizes the observed log-likelihood of the data which is the marginal log-likelihood of the observed data \mathbb{Y} :

$$\ln \mathcal{L}(\theta, \Pi; \mathbb{Y}) = \ln \sum_{\mathbb{S}} \mathcal{L}(\theta, \Pi; \mathbb{Y}, \mathbb{S}). \quad (3.27)$$

(3.27) is empirically very difficult to maximize for two reasons: (1) the log of the sums of functional forms of the likelihood function is difficult to maximize by standard

gradient/score methods; (2) the dimensionality of \mathbb{S} may be so large that we have to sum over the entire state space which may be infeasible. We hence apply the Stochastic Approximation Expectation-Maximization (SAEM) algorithm developed by Celeux and Diebolt (1992) and further analysed by Kuhn and Lavielle (2004) to overcome the two difficulties in maximizing (3.27).

Expectation Maximization Algorithm and its Stochastic Extensions

The SAEM algorithm is a stochastic approximation version of the Expectation Maximization (EM) algorithm originally documented by Dempster, Laird, and Rubin (1977). The EM algorithm is extremely useful in the presence of missing data or latent variable models, in which the direct maximization of the observed likelihood function is usually very difficult. To formally introduce the EM algorithm, let $f(\mathbf{Z}) = f(\mathbf{X}, \mathbf{Y})$ be the density of the complete data with \mathbf{Y} being the observable and \mathbf{X} the discrete unobservable data (which is the usual situation in which the EM algorithm is applied) characterized by the true parameter vector $\tilde{\boldsymbol{\theta}}$ ($\tilde{\boldsymbol{\theta}}$ will be used henceforward to denote some parameter vector $\boldsymbol{\theta}$ for the true data generating process). One wishes to estimate $\tilde{\boldsymbol{\theta}}$ by the MLE $\hat{\boldsymbol{\theta}}_{MLE}$ that maximizes the observed (marginal) log-likelihood function:

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg \max_{\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta}; \mathbf{Y}) = \ln \sum_{\mathbf{X}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}). \quad (3.28)$$

Instead of maximizing this usually infeasible observed log-likelihood, the EM algorithm calculates an estimation of $\boldsymbol{\theta}$ by maximizing the expectation of the observed likelihood with respect to \mathbf{X} given a previous estimate of the parameter vector. Formally, the EM algorithm consists of two steps:

Expectation step (E-step): At the n -th iteration, calculate the expectation of the log-likelihood with respect to the conditional distribution of \mathbf{X} given \mathbf{Y} and a current parameter estimate $\boldsymbol{\theta}^{(n)}$:

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(n)}) = \mathbf{E}_{\mathbf{X} | \mathbf{Y}, \boldsymbol{\theta}^{(n)}} \left[\ln \mathcal{L}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}) \right]. \quad (3.29)$$

Maximization step (M-step): Maximize $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(n)})$ w.r.t $\boldsymbol{\theta}$ to obtain a new estimate of the parameter vector $\boldsymbol{\theta}^{(n+1)}$. Repeat the algorithm until $Q(\boldsymbol{\theta}^{(n^*)} | \boldsymbol{\theta}^{(n^*-1)}) - Q(\boldsymbol{\theta}^{(n^*-1)} | \boldsymbol{\theta}^{(n^*-2)}) < \varepsilon$ for some small ε and some n^* , and the resulting parameter estimate is $\hat{\boldsymbol{\theta}}_{EM} = \boldsymbol{\theta}^{(n^*)}$.

By transferring the objective function from the log of a sum (3.28) to the sum of logs (3.29), the maximization problem is significantly simplified. Moreover, the

EM algorithm has an appealing likelihood-ascent property such that for each iteration, $\ln \mathcal{L}(\boldsymbol{\theta}^{(i)}; \mathbf{Y}) \geq \ln \mathcal{L}(\boldsymbol{\theta}^{(i-1)}; \mathbf{Y})$, so that the marginal likelihood is non-decreasing and the sequence $\{\boldsymbol{\theta}_{i=1,2,\dots}^{(i)}\}$ converges to $\tilde{\boldsymbol{\theta}}$ under some regularity conditions (Boyles, 1983; Wu, 1983).

The EM algorithm has been applied in various situations, including estimating finite mixture of distribution and hidden Markov models. However, it is not directly applicable to the MS-ACI model, because the expectation step becomes intractable due to the path dependency problem. The path dependency problem is created by the autoregressive structure of the model. Specifically, when estimating autoregressive models via maximum likelihood, the joint likelihood function is often replaced by the conditional likelihood function and it is calculated recursively for the complete dataset from observation 1 to T . In the MS-ACI model ((3.19) and (3.20)), since the autoregressive component $\tilde{\boldsymbol{\Phi}}$ and the error term $\tilde{\boldsymbol{\epsilon}}$ are functions of s_i , and $\tilde{\boldsymbol{\Phi}}_i$ is a function of $\tilde{\boldsymbol{\Phi}}_{i-k}$ for $k = 1 : i - 1$, both $\tilde{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{\epsilon}}$ are effectively a function of the complete history of the states, which then enter into the conditional likelihood function for observation i . Therefore, when taking the expectation with respect to the conditional distribution of states \mathbb{S} , for M different regimes and a sample size T , the possible paths of states grow exponentially (T^M), which renders the calculation of the E-step intractable.

To evaluate the E-step when no analytical expression of the expectation is available, several algorithms have been proposed, including the Monte Carlo EM (MCEM) (Wei and Tanner, 1990), the stochastic EM (SEM) (Celeux, Chauveau, and Diebolt, 1996; Celeux and Diebold, 1985; Diebolt and Ip, 1996) and the SAEM (Celeux and Diebolt, 1992). All these algorithms involve drawing random samples from the posterior distribution of the missing data given the observable data and parameters, hence they are stochastic extensions to the EM algorithm. The MCEM solves the intractable E-step by replacing the E-step with a Monte Carlo approximation of the expectation:

Simulation step (S-step): At the n -th iteration, draw M_n realizations of the unobserved data $\{\mathbf{X}^{(m_n, n)}\}_{m_n=1:M_n}$ from its conditional density $p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$.

Integration Step (I-Step) Construct $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})$ as the Monte Carlo integration given by:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) = \frac{1}{M_n} \sum_{m_n=1}^{M_n} \left[\ln \mathcal{L}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}^{(m_n, n)}) \right]. \quad (3.30)$$

Maximization Step (M-step): Maximize the $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})$ w.r.t. $\boldsymbol{\theta}$ to obtain $\boldsymbol{\theta}^{(n+1)}$. Repeat until convergence.

As $M_n \rightarrow \infty$, the MCEM algorithm converges to the EM algorithm, therefore the MCEM inherits most of the appealing features of the original EM algorithm. Although this algorithm provides an elegant solution to the intractable E-step problem, there are several issues that render this algorithm unattractive. Firstly, the convergence of this algorithm requires M_n to be increased steadily with n , whereas the optimal method to choose M_n based on n is still an open question. Secondly, the MCEM algorithm is computationally intensive in terms of the amount of simulation of the missing data needed. For example, Jank (2005) reports that the total amount of simulated vectors can easily reach 1 million for a complete run of the MCEM algorithm.

Alternatively, the SEM algorithm attempts to ‘fill in’ the missing data by a single draw from the posterior density of the missing data. It can be viewed as a special case of the MCEM algorithm by setting $M_n = 1$ for all n in the S-step, so the I-step becomes a very poor approximation of the E-step of the original EM algorithm. Different from the MCEM algorithm, the SEM algorithm does not directly yield a point estimate. Instead, the sequence $\{\boldsymbol{\theta}^{(n)}\}_{n=1,2,\dots}$ is a time-homogeneous Markov chain. Diebolt and Ip (1996) and Nielsen (2000) show that, if the chain is ergodic, then $\{\boldsymbol{\theta}^{(n)}\}_{n=1,2,\dots}$ converges in distribution to a normally distributed random variable centred at $\tilde{\boldsymbol{\theta}}$ as $n \rightarrow \infty$ which can therefore be used to generate a point estimate by averaging over the converged Markov chain. This algorithm is applicable wherever the MCEM is, and it is generally less computationally intensive since it requires less simulations per iteration and the maximisation problem is much simpler. However, as discussed in Celeux, Chauveau, and Diebolt (1996), the SEM algorithm is vulnerable to small sample size, which gives rise to the SAEM algorithm.

The SAEM algorithm replaces the Monte Carlo integration in the MCEM by a stochastic approximation scheme inspired by the stochastic approximation method introduced by Robbins and Monro (1951).

Simulation step (S-step): At the n -th iteration, draw M_n realizations of the unobserved data $\{\mathbf{X}^{(m_n,n)}\}_{m_n=1:M_n}$ from its conditional density $p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}^{(n)})$.

Stochastic Approximation: Construct $Q_n(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)})$ as the stochastic approximation given by:

$$Q_n(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) = (1 - \gamma_n)Q_{n-1}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n-1)}) + \frac{\gamma_n}{M_n} \sum_{m_n=1}^{M_n} \left[\ln \mathcal{L}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}^{(m_n,n)}) \right], \quad (3.31)$$

with $\{\gamma_n\}$ being a sequence of positive real numbers (usually referred to as the step size) that decrease gradually to zero, satisfying the following property: $\gamma_0 = 1$, $\sum_{n=1}^{\infty} \gamma_n = \infty$ and $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$.

Maximization Step (M-step): Maximize the $Q_n(\theta|\theta^{(n)})$ w.r.t. θ to obtain $\theta^{(n+1)}$. Repeat until convergence.

Clearly when $\gamma_n = M_n = 1, \forall n$, the SAEM algorithm reduces to the SEM algorithm. The sequence $\{\theta^{(n)}\}$ converges almost surely to a local maximum of the observed likelihood, which has been proven by Delyon, Lavielle, and Moulines (1999) and Allasonnière, Kuhn, and Trouvé (2010) under very general conditions. In addition to the relatively stronger convergence comparing with the convergence in distribution as in the SEM case, it has a significant advantage over the MCEM algorithm in the efficiency of exploiting the simulated missing data due to the fact that, as opposed to the MCEM case, the SAEM algorithm converges with a fixed M_n . Moreover, the recursive structure of $Q_n(\theta|\theta^{(n)})$ suggests that all previously drawn states contribute to this quantity $Q_n(\theta|\theta^{(n)})$, with the earlier sample discounted more towards zero by the step size γ_n . This greatly improves the efficiency of using sampled missing values, as in the MCEM algorithm, all missing data drawn at the $(n-1)$ -th iteration will be dropped and re-sampled at the n -th iteration. Finally, it is straightforward to obtain an estimate of the variance-covariance matrix with the SAEM algorithm, as will be demonstrated later in this section.

Estimating the MS-ACI Model with the SAEM Algorithm

Details on the implementation of the SAEM algorithm for the estimation of the MS-ACI model are provided in this section. Before introducing the algorithm, all the relevant log-likelihood functions will be listed for demonstration purposes. The conditional log-likelihood of \mathbb{Y} given the state vector \mathbb{S} :

$$\ln \mathcal{L}(\theta; \mathbb{Y}|\mathbb{S}) = \sum_{i=1}^T \left[-\Lambda(t_{i-1}, t_i) + \ln \lambda(t_i | \mathcal{F}_{t_i}) \right] = \sum_{i=1}^T \left[-\varepsilon_i(s_i) + \Phi_i(s_i) + \ln \lambda_0(x_i) \right]. \quad (3.32)$$

This log-likelihood can be easily maximized since it is in a log-linear form. The complete data log-likelihood for the joint density of $\{\mathbb{Y}, \mathbb{S}\}$ can be factorized as:

$$\ln \mathcal{L}(\theta, \Pi; \mathbb{Y}, \mathbb{S}) = \ln \mathcal{L}(\theta; \mathbb{Y}|\mathbb{S}) + \ln \mathcal{L}(\Pi; \mathbb{S}), \quad (3.33)$$

in which $\ln \mathcal{L}(\Pi; \mathbb{S})$ is the marginal log-likelihood for the Markov chain, given by:

$$\ln \mathcal{L}(\Pi; \mathbb{S}) = \ln \pi_{s_1} + \sum_{i=2}^T \ln \pi_{s_{i-1}, s_i}. \quad (3.34)$$

Note that since Π is independent of $\mathcal{L}(\theta; \mathbb{Y} | \mathbb{S})$, the complete likelihood can be maximized by separately maximising $\mathcal{L}(\theta; \mathbb{Y} | \mathbb{S})$ and $\mathcal{L}(\Pi; \mathbb{S})$.⁶ The marginal likelihood of \mathbb{Y} , which cannot be directly maximized as discussed in the previous section, can be expressed as:

$$\ln \mathcal{L}(\theta; \mathbb{Y}) = \ln \sum_{\mathbb{S}} \mathcal{L}(\theta; \mathbb{Y} | \mathbb{S}) \mathcal{L}(\Pi; \mathbb{S}). \quad (3.35)$$

To apply the SAEM algorithm, we first show that we can obtain a random sample from the conditional density of the state given the current parameter estimate $\theta^{(n)}$ and the data \mathbb{Y} . Since the multivariate density $p(\mathbb{S} | \theta^{(n)}, \mathbb{Y})$ is not a standard distribution,

$$p(\mathbb{S} | \theta^{(n)}, \mathbb{Y}) \propto f(\mathbb{Y} | \mathbb{S}, \theta^{(n)}) p(\mathbb{S}), \quad (3.36)$$

it is inconvenient to draw a random sample of \mathbb{S} as a vector from this density. Bauwens, Preminger, and Rombouts (2010) develop a single move Gibbs sampling scheme to draw a random sample of the states by a state-by-state approach for the MS-GARCH model, which can also be implemented in our case. The single move sampler is formulated as follows:

$$p(s_i | s_{1:i-1}^{(n+1)}, s_{i+1:T}^{(n)}, \theta^{(n)}, \mathbb{Y}) \propto p(s_i | s_{i-1}^{(n+1)}, s_{i+1}^{(n)}, \Pi^{(n)}) f(y_{i:T} | s_i, s_{1:i-1}^{(n+1)}, s_{i+1:T}^{(n)}, \theta^{(n)}). \quad (3.37)$$

Note that $s_{1:i-1}^{(n+1)}$ have been drawn in previous iterations. A change of s_i leads to different updates of $\tilde{\Phi}_i(s_i)$ and $\varepsilon_i(s_i)$ assuming unequal state parameters, and the impact of this change will also affect their future values due to the autoregressive structure of the model. Hence we must include information in $y_{i:T}$, which in turn describes the path dependency problem. Moreover, we need to condition on $s_{1:i-1}^{(n+1)}$ and $s_{i+1:T}^{(n)}$ to construct the likelihood for every $i = 2 : T$. From (3.37), we can obtain a random s_i by calculating (3.37) for every $s_i \in \mathcal{M}$, normalizing it and drawing a random s_i as drawing from a multinomial density. We can thus obtain one draw of the vector $\mathbb{S}^{(n+1)}$ by drawing each $s_i^{(n+1)}$ for i from 1 to T conditioning on $s_{1:i-1}^{(n+1)}$, $s_{i-1:T}^{(n)}$, \mathbb{Y} and $\theta^{(n)}$, which corresponds to a Gibbs sampling scheme for the states that iteratively draws one state at a time.

The single move approach has been criticized by Bauwens, Dufays, and Rombouts

⁶We maximize $\mathcal{L}(\theta; \mathbb{Y} | \mathbb{S})$ using standard gradient-based algorithms such as Newton-Raphson, and the MLE of $\mathcal{L}(\Pi; \mathbb{S})$ is available in closed form.

(2014) and Billio, Casarin, and Osuntuyi (2014) for being computationally intensive because for each s_i , one has to compute the conditional likelihood of $y_{i:T}$, which amounts to $T - i + 1$ number of calculations of the conditional density for one observation. For a sample size of T , we need to perform $(M - 1) \frac{T(T+1)}{2}$ number of calculations for a single draw of \mathbb{S} , which grows quadratically with T . We propose a single move sampling scheme that approximates (3.37) by calculating the conditional likelihood of $y_{i:i+\Delta}$ for some Δ instead of T when T is large, which in return reduces the number of calculations to be $(M - 1)(T\Delta + \frac{\Delta - \Delta^2}{2})$ that only grows linearly with T . Details of this sampler are presented in Appendix C.2.

Concluding from above, we are able to obtain a random draw from the distribution $p(\mathbb{S}|\theta^{(n)}, \mathbb{Y})$, and therefore we explain the implementation of SAEM algorithm to estimate the MS-ACI model.

Simulation Step (S-step): At the n -th iteration, given current parameter estimate $\theta^{(n)}$, $\Pi^{(n)}$ and current draw of state $\mathbb{S}^{(n)}$, draw $\mathbb{S}^{(n+1)}$ from the following density for $i = 1 : T$:

$$p(s_i = l | s_{-i}^{(n)}, \theta^{(n)}, \mathbb{Y}) \approx \frac{p(s_i = l | s_{i-1}^{(n+1)}, s_{i+1}^{(n)}, \Pi^{(n)}) f(y_{i:i+\Delta} | s_i = l, s_{-i}^{(n)}, \theta^{(i)})}{\sum_m^{\mathcal{M}} p(s_i = m | s_{i-1}^{(n+1)}, s_{i+1}^{(n)}, \Pi^{(n)}) f(y_{i:i+\Delta} | s_i = m, s_{-i}^{(n)}, \theta^{(i)})}, \quad (3.38)$$

in which $s_{-i}^{(n)} = \{s_{1:i-1}^{(n+1)}\} \cup \{s_{i+1:T}^{(n)}\}$.

Stochastic Approximation: Let $\vartheta^{(n)}$ denote the vector that combines $\theta^{(n)}$ and $M^2 - M$ free probability parameters in $\Pi^{(n)}$, update the quantity:

$$Q_n(\vartheta | \vartheta^{(n)}) = (1 - \gamma_n) Q_{n-1}(\vartheta | \vartheta^{(n-1)}) + \gamma_n \ln \mathcal{L}(\vartheta; \mathbb{Y}, \mathbb{S}^{(n+1)}), \quad (3.39)$$

in which γ_n is a positive step size that gradually decreases to zero as $n \rightarrow \infty$.

Maximization Step (M-step): maximize $Q_n(\vartheta | \vartheta^{(n)})$ w.r.t. ϑ to obtain $\vartheta^{(n+1)}$.⁷ Repeat until the stopping rule is satisfied:

$$\max \left(\frac{|\vartheta^{(n+1)} - \vartheta^{(n)}|}{|\vartheta^{(n)} + \delta_2|} \right) < \delta_1, \quad (3.40)$$

in which δ_1 and δ_2 are predetermined small values. To avoid a premature termination from an unlucky Monte Carlo draw, the algorithm is stopped when the stopping rule

⁷This maximization is similar to the maximization of the complete log-likelihood since it is a weighted sum of complete log-likelihoods. Therefore it can be factorized into a weighted sum of conditional log-likelihoods of the data given the state vector which will be maximized using gradient-based algorithms, and a weighted sum of log-likelihoods of simulated Markov chains of which the closed form MLE is available.

is satisfied three successive times, as suggested by Booth and Hobert (1999). We choose $\delta_1 = 0.001$ and $\delta_2 = 0$ in the estimation of our models.

Note that our SAEM scheme implicitly chooses $M_n = 1$ for all iterations due to the usually large sample size of time series data. One of the appeals of the SAEM algorithm is that it only needs to determine the step size γ_n . The general guideline to choose this step size is that, as explained by Jank (2006), small (large) step sizes reduce (inflate) the Monte Carlo error and yield slower (faster) convergence. We use the SEM algorithm as a burn-in step for our SAEM algorithm, because the SEM algorithm can achieve a fast convergence in distribution to a local maxima close to the initial value. The step size γ_n is of the following form:

$$\gamma_n = \begin{cases} 1, & n \leq n_0, \\ \frac{1}{(n-n_0)^{0.9}}, & n > n_0. \end{cases} \quad (3.41)$$

Thus the initial n_0 step is effectively SEM iterations, which will produce a sequence of parameter estimates $\{\vartheta^{(i)}\}_{i=1:n_0}$ that will converge fast but have large Monte Carlo error. The initial values of the parameter estimates for the SAEM algorithm is therefore given by the last parameter estimates of the SEM iterations. We use a small SAEM step size to refine this initial estimates when the initial parameter estimates of the SEM iterations reach convergence. As a result, n_0 should be chosen to be large enough so that less SAEM iterations are needed for the algorithm to converge.

The initial values of the SEM steps, $\vartheta^{(0)}$ and $\mathbb{S}^{(0)}$, are important factors in the estimation procedure. We found that the estimation scheme is robust to choices of $\mathbb{S}^{(0)}$, $\Pi^{(0)}$ and all ARMA parameters, but depends crucially on the baseline parameters. This is due to the fact that the regime identification is to a large extent determined by the baseline parameters. Guidances on choices of the initial values are provided in Section 1.7.

The SAEM algorithm only provides a point estimate for the parameter vector. As suggested by Delyon, Lavielle, and Moulines (1999) and Kuhn and Lavielle (2004), we can also obtain variance-covariance matrix estimates for the parameter estimates. From the posterior probability of the states conditioning on the parameter estimates, we are able to provide an estimate of the most probable state vector. The detailed estimation procedures for the variance-covariance matrix and the most probable state vector are presented in Appendix C.3.

3.3.4 Post-Estimation Diagnostics

The residual series $\{\hat{\epsilon}_i\}_{i=1:T}$ is often used to perform diagnostic tests for the original ACI model. The main property exploited here is that according to the RTCT, $\{\hat{\epsilon}_i\}_{i=1:T}$ follows an i.i.d. unit exponential process if the model specification is correct. As a result, diagnostic tests of the original ACI model usually involve testing the unit exponentiality of $\{\hat{\epsilon}_i\}_{i=1:T}$ and the presence of autocorrelation in the series $\{\hat{\epsilon}_i\}_{i=1:T}$. Since the RTCT applies to the general intensity process of point processes, this applies to the MS-ACI model as well and similar tests can be constructed for our MS-ACI model accordingly. These tests, however, are not directly applicable to the MS-ACI model with only parameter estimates $\hat{\vartheta}$ from the SAEM algorithm. This is because a state vector \mathbb{S} is required for the residual series to be calculated, which implies that without conditioning on a particular realization of the state vector, the diagnostic tests cannot be performed due to the inability to obtain $\{\hat{\epsilon}_i\}_{i=1:T}$. Thanks to the estimation of the most probable state sequence, we are able to obtain the conditional residual series $\{\hat{\epsilon}_i|\hat{\mathbb{S}}\}_{i=1:T}$ by plugging in the estimated (most probable) state sequence $\hat{\mathbb{S}}$ in the calculation of the residual series. Similar tests such as the Ljung-Box test for autocorrelation or some empirical density function tests for unit exponentiality can be performed on $\{\hat{\epsilon}_i|\hat{\mathbb{S}}\}_{i=1:T}$, which renders model diagnostics and comparison possible for our MS-ACI model. Certainly, one can perform the diagnostic tests on some residual series conditioning on any arbitrary \mathbb{S} , but the estimated state vector $\hat{\mathbb{S}}$ ensures that the corresponding conditional residual series $\{\hat{\epsilon}_i|\hat{\mathbb{S}}\}_{i=1:T}$ has the largest complete likelihood in the state vectors drawn for its estimation. Thus, the test performances for the residual series conditional on the estimated state vector are likely to be better compared to conditioning on an arbitrary \mathbb{S} .

For the original ACI model, the maximized log-likelihood also provides important insights in model comparisons and can be used to construct a likelihood ratio test for nested models. In the MS-ACI case, however, the maximized observed log-likelihood is not available because the SAEM algorithm does not evaluate it directly, and the integral over all realizations of \mathbb{S} is intractable due to the dimensionality of \mathbb{S} . The conditional log-likelihood of the MS-ACI model $\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}|\hat{\mathbb{S}})$ does provide some information on the goodness-of-fit of the model, but is not decisive due to the lack of theoretical foundations to construct a likelihood-based test.

The residual tests described above assess the appropriateness of applying the ACI framework. To be specific, the autocorrelation tests examine whether the autoregressive structure $\tilde{\Phi}_i$ is correctly specified, and the empirical density tests assess the goodness-of-fit of the baseline specification. These tests, however, do not provide a

method to evaluate the validity of the regime-switching structure. Methods to assess the contribution of an extra regime to the model estimation is important because the inclusion of another regime can lead to over-parametrization which reduces the efficiency of parameter estimation when the data generating process (DGP) does not possess the regime-switching property assumed by the model. In a Bayesian framework, this can be performed by computing the Bayes factor from the marginal likelihood (Bauwens, Dufays, and Rombouts (2014) develop the algorithm for the MS-GARCH model, which can be applied to our case if a Bayesian approach is pursued). In the frequentist's approach, this is usually accomplished by the likelihood ratio test for the nested models, which is not available due to the intractable observed likelihood.

To provide a descriptive statistic that reflects the validity of the existence of M regimes, we focus on the $T \times M$ posterior probability matrix \mathcal{P}^Δ conditioning on the estimated parameters $\hat{\boldsymbol{\vartheta}}$ and the estimated state vector $\hat{\mathbb{S}}$. The element at the intersection of row i and column m in \mathcal{P}^Δ , denoted by $\mathcal{P}_{i,m}^\Delta = p^\Delta(s_i = m | \hat{s}_{1:i-1}, \hat{s}_{i+1:T}, \mathbb{Y}, \hat{\boldsymbol{\vartheta}})$, is the posterior probability of the i -th state being classified as state m conditioning on $\hat{s}_{1:i-1}$, $\hat{s}_{i+1:T}$, \mathbb{Y} and $\hat{\boldsymbol{\vartheta}}$ calculated similarly as in (C.7) with the truncation lag Δ being the last adapted Δ in the estimation. Based on this matrix, we construct a statistic named the ‘Significance of Regimes’ (*SoR* hereafter) which serves as an indicator of the overall significance of the regime-switching structure. It is calculated as follows:

$$SoR = T^{-1} \sum_{i=1}^T \max_{m \in \mathcal{M}} \mathcal{P}_{i,m}^\Delta. \quad (3.42)$$

Intuitively, *SoR* is the average of the largest probability in every row of \mathcal{P}^Δ . It measures the average (conditional) probability of each state being classified in the most probable states. The rationale behind this statistic is that, assuming the DGP consists of M distinct regimes with densities far apart from each other (which is often assumed in the MS-GARCH literature), the probability of any observation being classified into its corresponding true state will be close to one. On the contrary, when all M densities are identical, all the elements in the matrix \mathcal{P}^Δ reduce to M^{-1} . This gives the range of *SoR*, and measures the significance of the existing regimes, with 1 being very significant for all the regimes and M^{-1} being not significant at all. The *SoR* allows easy comparisons across models with different number of regimes and baseline specifications. Moreover, we can calculate *SoR* for each regime to compare their relative significance. The *SoR* for the l -th regime is defined as:

$$SoR(l) = T^{-1} \sum_{i=1}^T \max_{m \in \mathcal{M}} \mathcal{P}_{i,m}^\Delta \mathbb{1}_{\{\max_{m \in \mathcal{M}} \arg(\mathcal{P}_{i,m}^\Delta) = l\}}, \quad (3.43)$$

which basically calculates the average of the largest element in every row of \mathcal{P}^Δ if the largest element belongs to the l -th state. We will elaborate on this in the simulation section.

3.4 Monte Carlo Simulation Study

In this section, we conduct a Monte Carlo simulation study to highlight the estimation performance of our SAEM algorithm under various parameter specifications. We start with MS(2)-ACI(1,1) models without covariates, and mainly focus on the impact of different baseline parameters and transition parameters. The reason to ignore a more complicated ARMA-type structure is because the ARMA-type structure is designed to be of zero mean. It will therefore have a very limited effect on the likelihood of the model given stationarity and identical baselines. We will consequently show that the quality of the parameter estimates of our model is dominated by the differences in the baseline parameters. Moreover, since the inclusion of other covariates serves as a regime-specific stochastic baseline function for each regime, our simulation evidence on the baseline function also reveals the importance of including covariates on the quality of the parameter estimates.

The selection of parameter constellations is designed to examine the aforementioned relationship between specifications and quality of estimates. The size of the Monte Carlo study for each specification is 100. For each specification, we choose $T = 1000$, $n_0 = 60$ and update Δ every 50 iterations in the estimation process of every Monte Carlo draw of the latent state vector \mathbb{S} and observable data vector \mathbb{Y} . We keep the initial values of the ARMA parameters to be zero and the initial Markov chain $\pi_{lm}^{(0)} = 0.5, \forall l, m$ and $\mathbb{S}^{(0)}$ to be a random draw of T repeated independent fair coin tosses. The initial values of the scale parameter $w(s_i)^{(0)}$ for each regime should be set carefully to allow for the convergence of the model. Specifically, in the two-regime case, assume the true scale parameters $\tilde{w}(1) > \tilde{w}(2)$, for $w(1)$ to converge to $\tilde{w}(1)$ (instead of $\tilde{w}(2)$, or to avoid being eliminated by the algorithm), we recommend setting $w^{(0)}(1) > \tilde{w}(1) > \tilde{w}(2) > w^{(0)}(2)$ or $\tilde{w}(1) > w^{(0)}(1) \gg w^{(0)}(2) > \tilde{w}(2)$. In this case, the initial values for each baseline are closer to their own true values and will converge to the corresponding maxima. Consequently, the label-switching problem is also alleviated.

The quality of the parameter estimates is assessed through the bias and the root mean squared error (RMSE) of the Monte Carlo parameter estimates. In detail, for each parameter specification, we benchmark the bias and the RMSE of the Monte Carlo

parameter estimates against the bias and the RMSE from a Monte Carlo simulation of the same model given the latent state vector (referred to as the complete model hereafter, and the incomplete model refers to the model with a latent state vector). The rationale of this comparison is that, given the state vector, the MS-ACI model reduces to a mixture of simple ACI models and the parameter estimates are the corresponding MLEs, which are consistent when the density of the data is correctly specified.⁸ This intuitively serves as an upper bound for the quality of the parameter estimates of the incomplete model. Thus, by comparing the quality of parameter estimates of the incomplete model to its corresponding complete version, we are able to depict the loss of accuracy associated with both parameter specifications and the observability of the Markov chain.

The main result for the case $M = 2$ is presented in Table 3.1. In Table 3.1, we base our simulation on six different specifications, with spec. 1-5 using exponential baselines and spec. 6 using a Weibull baseline. The first part of the table shows the bias and RMSE of the Monte Carlo parameter estimates of the incomplete model, with the bias and RMSE of the complete counterparts included in the second part of the table. The RMSE ratio calculated by dividing the RMSE of each parameter estimate of the incomplete model by that of the complete model allows easy comparison. Generally, the bias and RMSE of the incomplete model tend to fluctuate across specifications, while those for the complete model are much more consistent.

The immediate observation from the upper part of the table is that, by comparing spec. 1, 2 and 4, the closer the gap between the two regime-specific densities implied by the scale parameters $\tilde{w}(1)$ and $\tilde{w}(2)$, the smaller the mean SoR , which is as expected. According to the RMSE ratios, it is evident that the quality of the parameter estimates deteriorates with the mean SoR implied by the increasingly narrowing gap between the DGP scale parameters. The existence of a more persistent latent Markov chain can improve the mean SoR and the estimation performance by comparing the RMSE ratios of spec. 2 and 3. However, a persistent latent Markov chain alone cannot guarantee reliable parameter estimates, as is shown in spec. 5. It is worth mentioning that in spec. 5, the regime-specific scale parameters are identical, and the only information on the latent state vector implied by the observable data is the differences in the ARMA parameters, which does not generate enough discrepancy in the density for the latent Markov chain to be well identified. This is explained by the large downward biases in the transitional parameters in spec. 5. Moreover, the

⁸In fact, since the exponential baseline belongs to the linear-exponential family, the estimator is also the Quasi-MLE, which is robust to misspecification in the baseline function. For a detailed discussion on the consistency of this estimator, please refer to Hautsch (2012).

Table 3.1 Monte Carlo Simulation Results of Parameter Estimates of MS(2)-ACI(1,1) Models and the Corresponding Complete Models for 100 Random Draws of Data

Spec.	$\tilde{\alpha}_1(1)$ Bias RMSE	$\tilde{\alpha}_1(2)$ Bias RMSE	$\tilde{\beta}_1(1)$ Bias RMSE	$\tilde{\beta}_1(2)$ Bias RMSE	$\tilde{w}(1)$ Bias RMSE	$\tilde{w}(2)$ Bias RMSE	$\tilde{a}(1)$ Bias RMSE	$\tilde{a}(2)$ Bias RMSE	$\tilde{\pi}_{11}$ Bias RMSE	$\tilde{\pi}_{22}$ Bias RMSE	$\bar{\Delta}$ $\sigma(\text{SoR})$
Results for Incomplete Models											
1	0.05 -0.0046 0.0355	0.1 -0.0010 0.0332	0.8 -0.0429 0.1654	0.95 -0.0368 0.2248	0 0.0127 0.0584	-5 0.0154 0.0620	1 . .	1 . .	0.5 0.0001 0.0242	0.5 0.0038 0.0213	34.80 0.9803 0.0017
2	0.05 -0.0116 0.0542	0.1 0.0005 0.0465	0.8 -0.0844 0.3712	0.95 -0.0676 0.3573	0 -0.0077 0.0820	-3 0.0082 0.0668	1 . .	1 . .	0.5 -0.0008 0.0342	0.5 -0.0002 0.0357	38.44 0.9086 0.0082
3	0.05 -0.0019 0.0313	0.1 -0.0011 0.0391	0.8 -0.0877 0.3330	0.95 -0.0245 0.1103	0 0.0110 0.0646	-3 0.0091 0.0908	1 . .	1 . .	0.9 -0.0039 0.0136	0.9 -0.0012 0.0166	31.14 0.9761 0.0039
4	0.05 -0.0112 0.0556	0.1 0.0034 0.0706	0.8 -0.0079 0.3404	0.95 0.0263 0.4455	0 -0.1490 0.2122	-1 -0.0583 0.1351	1 . .	1 . .	0.5 0.0870 0.1430	0.5 -0.1578 0.2217	38.06 0.7427 0.0464
5	0.05 0.0183 0.0456	0.1 -0.0230 0.0515	0.8 0.1325 0.2962	0.95 -0.0118 0.3296	0 0.0355 0.1121	0 -0.0351 0.1311	1 . .	1 . .	0.9 -0.4107 0.4359	0.9 -0.5080 0.5331	34.53 0.6682 0.0592
6	0.05 -0.0193 0.0879	0.1 -0.0061 0.0564	0.8 -0.0759 0.4527	0.95 0.0213 0.4211	0 0.0611 0.2460	-1 0.0047 0.0606	1 0.0119 0.0555	2 0.0590 0.2300	0.5 -0.0208 0.1063	0.5 -0.0028 0.1242	42.51 0.7941 0.0351
Results for Complete Models											
1	0.05 -0.0009 0.0240	0.1 -0.0039 0.0243	0.8 -0.0322 0.1385	0.95 -0.0008 0.1268	0 0.0021 0.0518	-5 0.0101 0.0651	1 . .	1 . .	0.5 -0.0040 0.0225	0.5 -0.0071 0.0239	
2	0.05 -0.0020 0.0219	0.1 -0.0030 0.0280	0.8 -0.0080 0.1316	0.95 0.0006 0.1167	0 0.0048 0.0572	-3 0.0059 0.0618	1 . .	1 . .	0.5 0.0001 0.0228	0.5 -0.0020 0.0233	
3	0.05 -0.0010 0.0245	0.1 -0.0001 0.0233	0.8 -0.0464 0.2217	0.95 -0.0209 0.0600	0 -0.0061 0.0675	-3 -0.0039 0.0785	1 . .	1 . .	0.9 -0.0014 0.0136	0.9 0.0015 0.0119	
4	0.05 0.0012 0.0242	0.1 0.0005 0.0267	0.8 -0.0312 0.1409	0.95 -0.0007 0.1109	0 -0.0067 0.0561	-1 -0.0002 0.0674	1 . .	1 . .	0.5 0.0007 0.0185	0.5 -0.0021 0.0215	
5	0.05 -0.0045 0.0251	0.1 -0.0032 0.0270	0.8 -0.0762 0.2291	0.95 -0.0036 0.0465	0 0.0014 0.0538	0 -0.0016 0.0794	1 . .	1 . .	0.9 -0.0028 0.0147	0.9 0.0001 0.0159	
6	0.05 -0.0049 0.0247	0.1 0.0009 0.0295	0.8 -0.0157 0.1276	0.95 -0.0237 0.1255	0 0.0002 0.0568	-1 -0.0043 0.0370	1 0.0006 0.0346	2 0.0205 0.0711	0.5 0.0011 0.0212	0.5 -0.0001 0.0204	
RMSE Ratios											
1	1.4817	1.3645	1.1944	1.7724	1.1274	0.9526	.	.	1.0749	0.8908	
2	2.4817	1.6570	2.8210	3.0619	1.4343	1.0806	.	.	1.5012	1.5296	
3	1.2793	1.6789	1.5019	1.8394	0.9570	1.1562	.	.	1.0043	1.3881	
4	2.2982	2.6423	2.4167	4.0177	3.7799	2.0051	.	.	7.7378	10.3085	
5	1.8177	1.9104	1.2929	7.0868	2.0814	1.6511	.	.	29.6159	33.4576	
6	3.5616	1.9104	3.5486	3.3566	4.3310	1.6392	1.6029	3.2333	5.0072	6.0982	

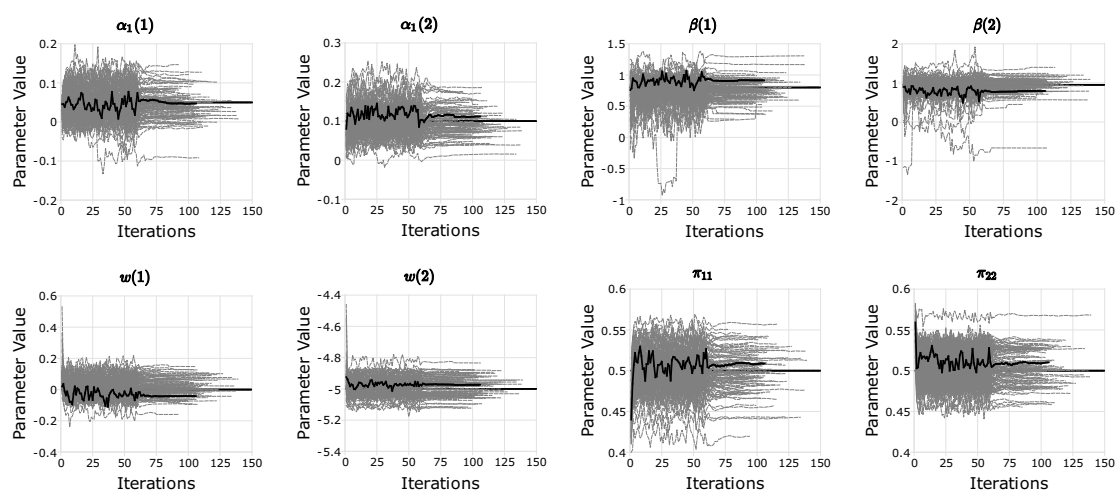
Note: $\overline{\text{SoR}}$ and $\sigma(\text{SoR})$ are the mean and standard deviation of the significance of regime defined in equation (3.42). RMSE is the root mean square error of the Monte Carlo parameter estimates. $\bar{\Delta}$ is the mean of the adapted Δ in each simulation. The RMSE ratio section is the RMSE of the corresponding parameter of the incomplete model over the RMSE of its complete counterpart.

mean parameter estimates for the ARMA parameters of the incomplete model are very close to each other, suggesting a label-switching problem caused by a common baseline. Finally, comparing spec. 4 and 6, we observe that a more complex baseline function can improve the mean *SoR* by allowing more flexible discrepancies of the regime-specific densities, which can potentially reinforce the estimation performance. In our example, by changing the Weibull parameter, the Markov chain parameters are estimated with much better accuracy (smaller bias and RMSE ratio).

An important message from Table 3.1 is that, since the latent Markov chain will inevitably result in a loss of information in the estimation process, the quality of the parameter estimates depends crucially on how much information can be extracted from the observed data itself, or how close the sampled state vector is to the true latent vector. Complexity of baselines and the persistence of the latent Markov chain will influence the amount of information available in the observed data, and will have a significant effect on the quality of the parameter estimates. However, the predominant factors are the DGP scale parameters of the baseline functions, which generally control the location of the distribution and the ability of our sampler to correctly classify the states. This rationalizes the use of the *SoR* as an indicator of the quality of the parameter estimates, as it is a measure of how confident our sampler is in classifying the observations. This also suggests that, the MS-ACI model is by construction inconsistent because of the missing state vector, and to ensure the performance of the MS-ACI model, we recommend a *SoR* of at least 90% so that the hidden Markov chain is on average properly identified (as can be observed from the RMSE ratios for the last three specifications). Also note that for all the specifications, the average adapted Δ is less than 50. This suggests that under these specifications, instead of taking all the observations into the sampler, less than 50 observations satisfy our tolerance of approximation error, which leads to a vast improvement in the computational efficiency of the single move sampler.

To assess the stability of our estimation scheme, we plot the convergence diagram for spec. 1 as an example in Figure 3.1. From the figure, the transition from SEM to SAEM can be clearly observed at $n_0 = 60$. Before $n_0 = 60$, the SEM algorithm provides trial parameter estimates with large Monte Carlo errors that are close to the maxima, and the SAEM algorithm drives the trial values to its convergence. The diagram shows that the majority of the parameter trajectories lie symmetrically around the DGP value. There are some erratic outliers in the ARMA parameters possibly due to the small sample size of the data. The convergence diagram provides evidence in favour of the stability and reliability of our SEM-SAEM estimation scheme.

Figure 3.1 Convergence Diagram for Spec. 1 in Table 3.1



Note: Each grey dashed line represents a trajectory of parameter estimates of the SEM-SAEM algorithm. DGP value is indicated by the horizontal black line. A sample trajectory is highlighted in each graph by a black solid line.

To analyse the effect of an extra regime, similar Monte Carlo simulation results for a MS(3)-ACI(1,1) model are presented in Table 3.2. The specification of the

Table 3.2 Monte Carlo Simulation Results of Parameter Estimates of a MS(3)-ACI(1,1) Model and its Complete Version for 100 Random Draws of Data

ϑ	DGP Values	Bias	RMSE	Bias	RMSE	RMSE Ratio
		Incomplete		Complete		
$\alpha_1(1)$	0.25	0.0024	0.0610	-0.0008	0.0423	1.4415
$\alpha_1(2)$	0.15	0.0075	0.0578	-0.0019	0.0361	1.5982
$\alpha_1(3)$	0.05	0.0053	0.0640	0.0031	0.0323	1.9781
$\beta_1(1)$	0.9	-0.0320	0.1645	-0.0004	0.0928	1.7727
$\alpha_1(2)$	0.8	-0.0616	0.2074	0.0006	0.1188	1.7460
$\alpha_1(3)$	0.6	-0.0131	0.2267	-0.0257	0.1298	1.7464
$w(1)$	-5	0.0059	0.0936	0.0116	0.0907	1.0321
$w(2)$	-1	-0.0050	0.0600	0.0053	0.0501	1.1973
$w(3)$	0	0.0011	0.0420	0.0043	0.0355	1.1828
$a(1)$	1	0.0194	0.0545	0.0047	0.0445	1.2246
$a(2)$	1.5	0.0286	0.1014	-0.0065	0.0607	1.6700
$a(3)$	2	0.0234	0.1119	0.0023	0.0805	1.3903
π_{11}	0.5	-0.0037	0.0310	0.0002	0.0285	1.0876
π_{12}	0.25	0.0033	0.0340	-0.0007	0.0221	1.5339
π_{21}	0.25	0.0023	0.0348	0.0016	0.0251	1.3834
π_{22}	0.5	-0.0107	0.0434	-0.0064	0.0298	1.4587
π_{31}	0.25	-0.0009	0.0310	0.0021	0.0241	1.2878
π_{32}	0.25	0.0009	0.0385	0.0021	0.0261	1.4735
\overline{SoR}	0.8610	$\sigma(\overline{SoR})$	0.0122	$\overline{\Delta}$	30.84	
$\overline{SoR(1)}$	0.9916	$\sigma(\overline{SoR(1)})$	0.0032			
$\overline{SoR(2)}$	0.8298	$\sigma(\overline{SoR(2)})$	0.0172			
$\overline{SoR(3)}$	0.7760	$\sigma(\overline{SoR(3)})$	0.0252			

Note: \overline{SoR} and $\sigma(\overline{SoR})$ are the mean and standard deviation of the significance of regime defined in equation (3.42). The statistics for $\overline{SoR}(l)$ are the significance of regime for the individual regime l defined in equation (3.43). RMSE is the root mean square error of the Monte Carlo parameter estimates. $\overline{\Delta}$ is the mean of the adapted Δ in each simulation.

MS(3)-ACI(1,1)-Weibull model in Table 3.2 consists of two Weibull regimes (regime 2 and 3) with densities that are close to each other and one exponential regime further apart from the two other regimes. The transitional parameters suggest that the probability of staying in the same regime is 0.5. The most interesting observation is for the individual mean *SoR*s. Since the first regime distinguishes itself from the other two, mean *SoR*(1) is larger than mean *SoR*(2) and mean *SoR*(3), approaching the theoretical maximum 1. The lower mean *SoR*(2) and mean *SoR*(3) are in turn explained by their closeness in distance. The smaller mean *SoR*(3) suggests that there are possibly more overlaps between regime 3 and regime 1, rendering it the least significant regime of the three. This shows that the inclusion of a regime that is similar to one of the existing regimes reduces the overall *SoR* and introduces noise to the model. The RMSE ratios also indicate that parameter estimates for the first regime perform slightly better than the other two. To sum up from Table 3.2, in the case $M = 3$, our SAEM algorithm provides reliable parameter estimates when at least one of the regimes is outstanding. The *SoR* for the single regime is able to capture the relative significance of each regime in comparison with the other two, which will provide inference for the selection of M .

On the choice of M in the estimation stage of our MS-ACI model, we recommend to start with small M and compare its performance to models with a larger M . Due to the fact that there will be more overlaps between multiple densities, models with larger M may be inferior to those with a smaller M in terms of *SoR*. For example, if a MS(2)-ACI(1,1) model were implemented to the data generated by the model in Table 3.2, the model would identify two distinct regimes with a mean *SoR* of over 97%. This is because the distance between densities of regime 2 and 3 are much closer in comparison to their distance to regime 1, and the two-state model combines regime 2 and 3 into a single regime. This suggests that *SoR* should not be the sole criteria for model selection. In this case, the overall goodness-of-fit of the baseline function of a three-regime model will outperform that of a two-regime model, and residual diagnostics should also be taken into account in the model selection procedure.

3.5 Application to High-Frequency Stock Price Duration Data

In the previous section we were able to evaluate the quality of the parameter estimates from the SEM-SAEM estimation scheme of our MS-ACI model with various specifications for $M \leq 3$. This section introduces the dataset for the application of our MS-ACI model, and analyses the empirical results in detail.

3.5.1 Data

We choose 9 highly liquid stocks and a stock index ETF in our empirical analysis to demonstrate our findings, namely AIG, CVX, GM, INTC, JPM, PFE, T, VZ, WMT and SPY. The data is obtained from the Trade and Quote (TAQ) dataset.⁹ The sample period ranges from 1 Jan 2016 to 31 Dec 2016. The trade dataset consists of prices and trading volumes, timestamped to milliseconds. To support the argument of regime shifts in intraday volatility, we apply our MS-ACI model to the point process of absolute price change events. The point process is constructed as follows: We start with an observed transaction price series $P(t_{j,d}^*)$, in which t_j^* refers to the arrival time of the j -th transaction in day d , the subscript $j = 1 : J_d$ is the transaction counter for day d with a total of J_d transactions at day d , and $d = 1 : D$ is the index for trading days for a total of D trading days. An arbitrary price change threshold δ (typically in multiples of ticks) is chosen in order to construct the ‘price event’, where the cumulative price change from the previous event is equal or larger than δ . Specifically, the event time $t_{i,d}$ for event i on day d is obtained by the following algorithm:

1. From $j = 1$ for every d , set $t_{0,d} = t_{1,d}^*$. Set the value of δ .
2. Let $t_{i,d} = \inf_{t_{j,d}^* > t_{i-1,d}} \{|P(t_{j,d}^*) - P(t_{i-1,d})| \geq \delta\}$.
3. Stop if $t_{I_d,d} \leq t_{J_d,d}^*$ and $t_{I_d+1,d} > t_{J_d,d}^*$.

The process $\{t_{i,d}^\delta\}_{i=1:I_d, d=1:D}$ records the arrival times of each price event, and is known as the δ -related absolute price change process, in which I_d is the number of price events in day d . An instantaneous volatility measure can be constructed based on the conditional intensity representation of this point process, as proposed by Engle and Russell (1998) and Gerhard and Hautsch (2002):

$$\sigma_\delta^2(t_{i,d}) = \lambda^\delta(t_{i,d} | \mathcal{F}_{t_{i,d}}) \left[\frac{\delta}{P(t_{i,d})} \right]^2, \quad (3.44)$$

in which $\lambda^\delta(t_{i,d} | \mathcal{F}_{t_{i,d}})$ is the conditional intensity of $\{t_{i,d}^\delta\}$ as defined in (3.1). Notice that the volatility process is proportional to the conditional intensity process.

The choice of δ is crucial in constructing this volatility estimator, because it determines the (random) sampling frequency of the raw dataset. Generally, a small δ samples the raw dataset more frequently, which provides more precise volatility

⁹The dataset is cleaned according to Holden and Jacobsen (2014) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). The stocks are chosen for illustrative purpose. We have generated results for up to 30 highly liquid stocks, and they are available upon request.

estimates, but can lead to a very noisy price event process due to market microstructure noise. On the contrary, a large δ samples the raw dataset more sparsely, and the resulting price event process is less affected by market microstructure noise, at the expense of the precision of the volatility estimates. We construct the point process monthly and choose δ to be the minimum multiples of \$0.005 that on average samples the raw data every 5 minutes.¹⁰ Tse and Yang (2012) show that under this sampling frequency, intensity-based volatility estimates performs very well against the RV estimates. Moreover, this choice of δ results in similar sample sizes across different securities, enabling an easier cross-sectional comparison. Other choices of δ are available. For example, Nolte, Taylor, and Zhao (2018) suggests to set δ to be three times the average bid-ask spread of the previous sampling period. However, this requires information on the intraday quoted prices, and the cross-sectional sample size will differ significantly. We present the δ used for all 120 stock-month datasets in Figure C.3 in Appendix C.7.

For demonstration purposes it is more convenient to analyse the statistical properties of the duration representation of the absolute price change process defined as $x_{i,d}^\delta = t_{i,d}^\delta - t_{i-1,d}^\delta$, which will be referred to as price durations. To investigate the regime-switching volatility-volume relationship based on the price duration data, we compute a volume measure by the log cumulative trading volume within each price duration, denoted by $\ln Vol_{i,d}^\delta$. It is well documented that empirical price durations and the price duration based covariates exhibit diurnal patterns (Andersen and Bollerslev, 1997b; Engle and Russell, 1998). To ensure that our results are not driven by these time deterministic effects, we filter out the seasonality pattern from the raw price durations and duration based volume measure, and obtain their deseasonalized versions $\ln \hat{x}_{i,d}^\delta$ and $\ln \hat{Vol}_{i,d}^\delta$. The detailed deseasonalization procedure is presented in Appendix C.4. Yearly descriptive statistics for the deseasonalized price durations and volume are presented in Table 3.3 below. Table 3.3 shows that, firstly, the mean duration is roughly 300 seconds which corresponds to our choice of δ . The distribution of duration is skewed to the right and has over-dispersion, as the mean is much larger than the median, and the standard deviation is larger than the mean for almost every stock. The minimum duration is not exactly zero but smaller than 0.01. The minimum volume, however, can be zero because it is in log. The log volume distribution is very symmetric. For comparison we also present the yearly descriptive statistics for raw price durations and volumes in Table C.1 in Appendix C.7. Comparing Tables 3.3 and C.1, it is clear that the mean duration and volume

¹⁰The length of the sample window is chosen to contain sufficient data and trading days to allow for a reliable seasonality estimate, but not to an excessive extent so that the estimation time of the model is manageable and potential intertemporal parameter instability can be avoided.

do not change too much. The main differences are changes in the maxima and that the deseasonalized variables have less standard deviation, which is expected as the intraday variations have been removed from the variables.

Table 3.3 Yearly Descriptive Statistics for $\dot{x}_{i,d}^\delta$ and $\ln \dot{Vol}_{i,d}^\delta$

Ticker	Obs.	$\dot{x}_{i,d}^\delta$					$\ln \dot{Vol}_{i,d}^\delta$				
		Mean	σ	Min	Median	Max	Mean	σ	Min	Median	Max
AIG	17925	321.11	346.17	0.00	212.80	5383.79	10.52	1.21	0.00	10.56	17.87
CVX	18014	316.99	334.58	0.00	211.34	3692.72	10.77	1.20	0.00	10.81	18.64
GM	17240	327.96	375.66	0.00	203.13	5801.45	11.24	1.12	0.00	11.27	16.80
INTC	16541	339.97	384.96	0.00	214.40	6015.84	11.79	1.24	2.30	11.83	19.52
JPM	17270	331.75	367.36	0.00	212.47	6267.21	11.61	1.09	0.00	11.64	17.47
PFE	17416	340.70	455.78	0.00	203.95	12422.01	12.03	1.28	1.33	12.07	19.10
SPY	18530	313.15	397.74	0.00	185.29	9944.04	13.52	1.01	3.89	13.56	17.17
T	17008	335.65	386.79	0.00	211.85	6575.75	11.83	1.18	0.00	11.86	18.40
VZ	16853	336.58	365.23	0.00	217.63	4564.45	11.27	1.21	0.00	11.30	18.54
WMT	17082	336.05	380.97	0.00	209.53	4774.98	10.83	1.19	0.00	10.87	18.43

Note: The table presents the descriptive statistics for the deseasonalized price durations $\dot{x}_{i,d}^\delta$ and the deseasonalized volume $\ln \dot{Vol}_{i,d}^\delta$ from 30 securities for the year 2016. Obs. denotes the total number of observations. σ is the standard deviation.

To briefly describe the intraday volume-volatility relationship, we regress $\ln \dot{Vol}_{i,d}^\delta$ on $\ln \dot{x}_{i,d}^\delta$ using a linear regression model:

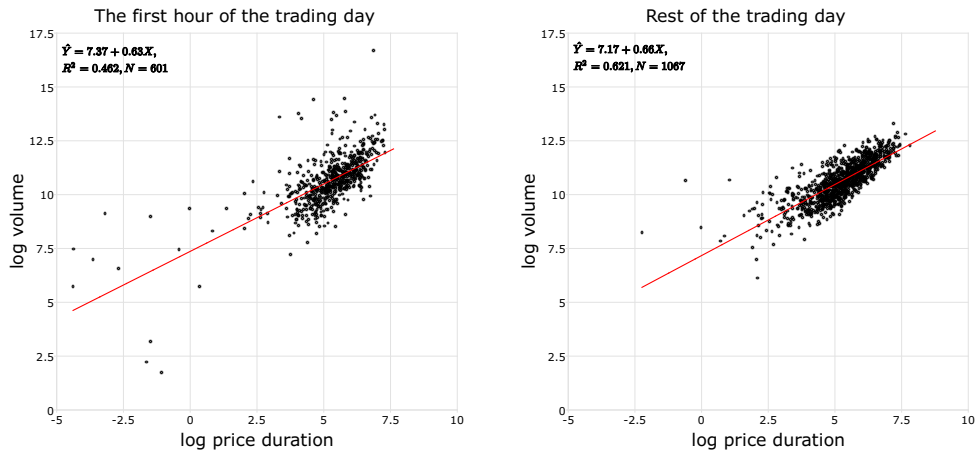
$$\ln \dot{Vol}_{i,d}^\delta = b_0 + b_1 \ln \dot{x}_{i,d}^\delta + \varepsilon_{i,d} \quad (3.45)$$

Intuitively, the regression describes the elasticity between duration and volume, as b_1 is the percentage increase in volume per 1 percent increase in price duration. If there is a regime-switching relationship between volume and volatility, we should be able to observe different estimates of the parameters and of the R^2 for different subsamples of the dataset. As a preliminary descriptive analysis, we split the data into two subsamples based on calendar time: the first subsample includes all price durations within the first hour of a trading day ($t_{i,d} \leq 3600s$), and the second subsample consists of the rest of the observations. We present three examples in Figure 3.2.

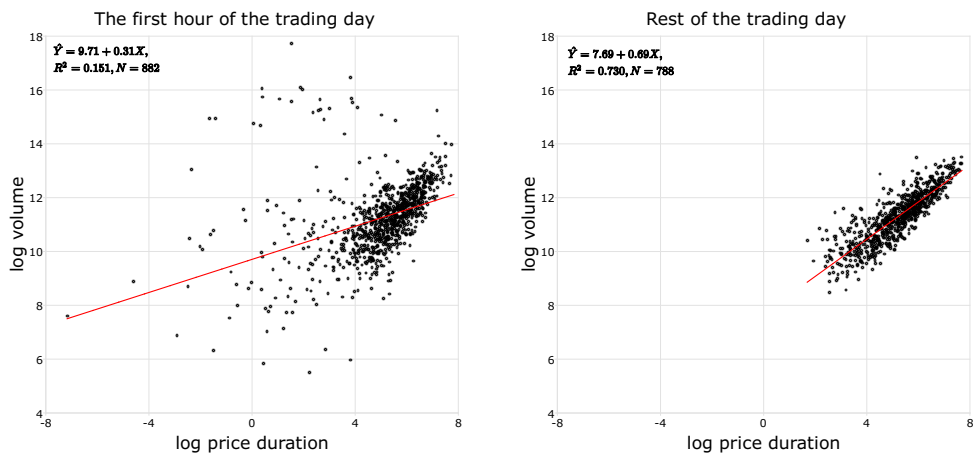
From Figure 3.2, it is evident that for individual stocks, the volume-duration relationship for observations during the first hour of the trading day is very different from those from the rest of the day. For observations in the first hour, the estimated R^2 is much smaller. We can see from the right panel of the graph that for the rest of the day, $\ln \dot{Vol}_{i,d}^\delta$ and $\ln \dot{x}_{i,d}^\delta$ are highly linearly dependent and cluster symmetrically along the regression line. During the first hour, the regression line deviates from the cluster, which is driven by observations that have disproportionately large volume in short price durations. However, we cannot observe this effect for the SPY, as from

Figure 3.2 Scatter plots with regression line for AIG 2016-03, INTC 2016-08, SPY 2016-05

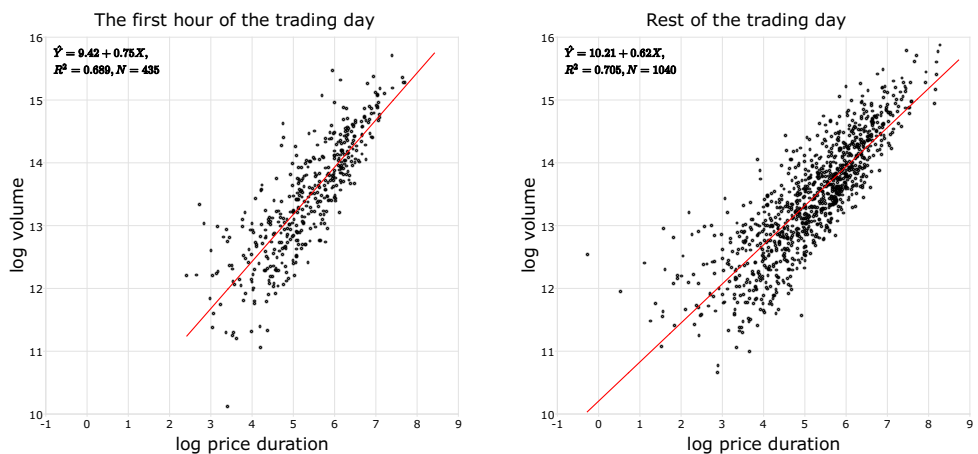
Panel 1: AIG 2016-03



Panel 2: INTC 2016-08



Panel 3: SPY 2016-05

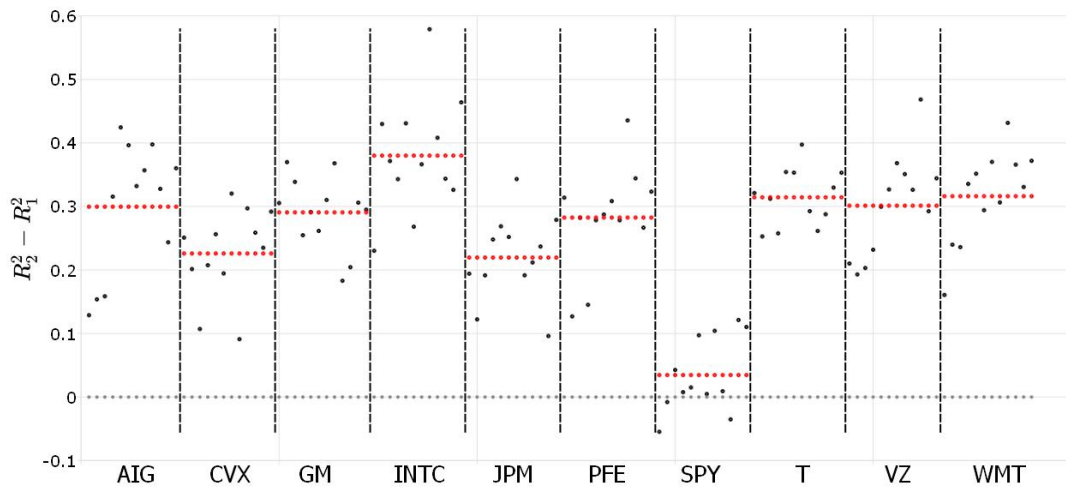


Note: Three 1-by-2 graphs from top to bottom: AIG 2016-03, INTC 2016-08, SPY 2016-05. X-axis: $\ln \hat{x}_{i,d}^\delta$. Y-axis: $\ln \hat{Vol}_{i,d}^\delta$. The regression line is obtained by (3.45).

the figure, the regression outputs are very close for the two subsamples.

To give a complete picture of our dataset, we denote R_1^2 , the R^2 from the regression on the first hour subsample, and R_2^2 , the R^2 from the other subsample. We plot $\hat{R}_2^2 - \hat{R}_1^2$ for all 10×12 stock-month datasets in Figure 3.3. From Figure 3.3, it is evident that the average R^2 difference is between 0.2 to 0.4 for all individual securities except SPY, whose average R^2 difference is much closer to zero.

Figure 3.3 R^2 Difference for the volume-duration regressions in the first hour of the day and the rest of the day



Note: The figure plots $\hat{R}_2^2 - \hat{R}_1^2$ obtained from regression (3.45) for 10×12 stock-month datasets, with \hat{R}_1^2 the estimated R^2 from the observations in the first hour of the trading day and \hat{R}_2^2 from the rest of the observations. Each black dot represents the R^2 difference for one stock-month dataset. The vertical black dashed lines split observations from each stock, and between two vertical red lines, the R^2 differences are ordered chronologically. The horizontal red dots represents the yearly average R^2 difference for each stock.

Concluding our data section, we have found a regime-switching volume-volatility relationship for all the individual stocks considered in our analysis by a simple static OLS regression. However, the simple OLS regression has two major drawbacks: (1) it does not consider the dynamics of the intraday volatility and (2) dividing the data at the first hour is somewhat arbitrary in the sense that there can be informed traders trading at other times during the day, and uninformed traders can also trade within the first hour. This motivates our MS-ACI model as it is tailored for capturing the dynamics of intraday volatility through a fully parametrized specification. Moreover, the regime identification is based solely on the correlation between volume and volatility, without any arbitrarily chosen threshold.

3.5.2 Main Empirical Results

This section discusses our main empirical findings. From our previous results in the data section, we use a two-state MS-ACI model with the inclusion of the trading volume to detect a regime-switching volume-volatility relationships in the data. The full model specification is summarized as follows:

$$\begin{aligned}
 \lambda(t; \mathcal{F}_t) &= \Phi(t) \lambda_0(t), \\
 \Phi(t) &= e^{\tilde{\Phi}_{\tilde{N}(t)+1}(s_{\tilde{N}(t)+1}) + \eta_{vol}(s_{\tilde{N}(t)+1}) \ln Vol_{\tilde{N}(t)+1,d}^\delta}, \\
 \tilde{\Phi}_i(s_i) &= \sum_{j=1}^2 \alpha_j(s_i) \tilde{\epsilon}_{i-j}(s_{i-j}) + \beta(s_i) \tilde{\Phi}_{i-1}(s_{i-1}), \\
 \lambda_0(t) &= a(s_{\tilde{N}(t)+1}) e^{w(s_{\tilde{N}(t)+1}) a(s_{\tilde{N}(t)+1})} x(t)^{a(s_{\tilde{N}(t)+1})-1}, \\
 \tilde{\epsilon}_i(s_i) &= -\gamma - \tilde{\Phi}_i(s_i) - a(s_i) w(s_i) - a(s_i) \ln \dot{x}_{i,d}^\delta, \\
 \mathbf{P}(s_i = m | s_{i-1} = l) &= \pi_{lm}, \quad l, m \in \mathcal{M} = \{1, 2\}.
 \end{aligned} \tag{3.46}$$

The parameters $\alpha_1(s_i)$, $\alpha_2(s_i)$ and $\beta(s_i)$ capture the state-specific ARMA(2,1)-type structure between durations. As in the simulation section, we specify a regime-switching Weibull baseline for the model with location parameters $w(s_i)$ and shape parameters $a(s_i)$. The parameter $\eta_{vol}(s_i)$ accounts for the state specific volume-volatility relationship of interest. Note that for the duration $t \in (t_{i-1}^\delta, t_i^\delta]$, we have $t_{\tilde{N}(t)+1}^\delta = t_i^\delta$, therefore the model captures the contemporaneous volume-volatility relationship for state s_i as desired. The parameters π_{lm} are the transition probabilities between state l and m . Estimation of π_{11} and π_{22} suffices, as the other two parameters are implicitly estimated.

We refer to the above model as the MS(2)-ACI(2,1)-V model, with the V representing the inclusion of the volume covariate. We estimate the above model using the deseasonalized dataset $(\dot{x}_{i,d}^\delta, \ln Vol_{i,d}^\delta)$ to obtain estimates of the parameters, the variance-covariance matrix of the estimated parameters and the most probable state sequence. Multiple starting values are used in the estimation to guarantee a stable convergence of the algorithm.

We present three examples of parameter estimates for the stock-months AIG 2016-03, INTC 2016-08 and SPY 2016-05 in Table 3.4. The estimated $\hat{\beta}(m)$ suggests that the empirical intensity process is highly persistent, which is consistent with the literature. The baseline parameters from the two regimes differ considerably, which partly determines the identification of the regimes. There is a considerable difference between $\hat{\eta}_{vol}(1)$ and $\hat{\eta}_{vol}(2)$ for AIG and INTC, but this difference is not visible in

the SPY. In general, for the individual stocks we observe two distinct regimes: a dominant regime (regime 2) with more than 80% observations and a large $|\hat{\eta}_{vol}(2)|$, and a minor regime (regime 1) with $|\hat{\eta}_{vol}(1)|$ much closer to zero. For the SPY, the difference between $|\hat{\eta}_{vol}(1)|$ and $|\hat{\eta}_{vol}(2)|$ is not as clear, so we will use regime 1 to denote the regime with smaller number of observations.

From the diagnostic statistics panel, we see that both the Cramér-von-Mises test and the Anderson-Darling test cannot reject the unit exponentiality hypothesis of the residual $\hat{\epsilon}_i$. More importantly, $\widehat{SoR}(m)$ and \widehat{SoR} suggest that for the individual stocks, the regimes are classified with high confidence, while for the SPY, the classification of regimes is much weaker. To conserve space, we leave extensive robustness checks on the justification of our model selection strategy to Appendix C.5, and only present key implications of our findings in the main text.

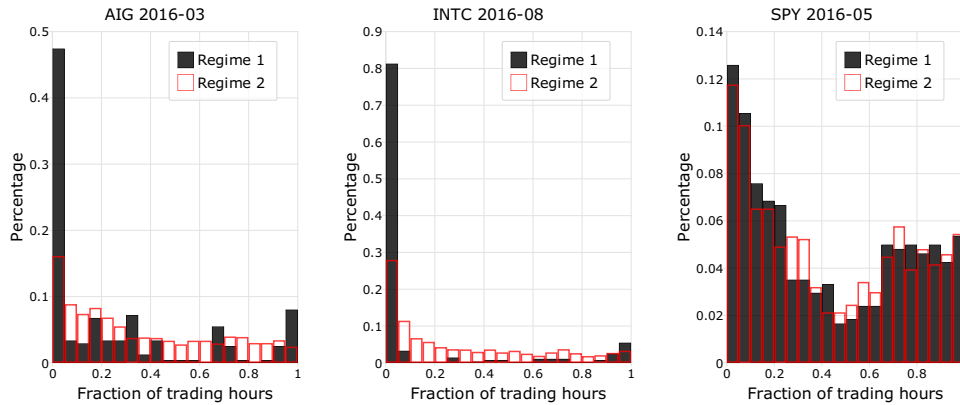
Table 3.4 Parameter Estimates of MS(2)-ACI(2,1)-V Model for AIG 2016-03, INTC 2016-08 and SPY 2016-05

$m =$	AIG 2016-03		INTC 2016-08		SPY 2016-05	
	1	2	1	2	1	2
Dynamic Parameters						
$\alpha_1(m)$	0.0223 (0.0278)	0.2310*** (0.0201)	0.3258*** (0.0385)	0.2192*** (0.0206)	0.4526*** (0.0240)	0.1942*** (0.0199)
$\alpha_2(m)$	0.0514* (0.0277)	-0.0988*** (0.0206)	0.1747*** (0.0580)	-0.0807*** (0.0222)	-0.1742*** (0.0229)	-0.2371*** (0.0214)
$\beta(m)$	1.0572*** (0.0128)	0.9388*** (0.0104)	0.4121*** (0.1294)	0.9241*** (0.0185)	1.1183*** (0.0193)	0.8992*** (0.0132)
Baseline Parameters						
$w(m)$	1.2721** (0.5477)	3.0143*** (0.1181)	-0.6130 (0.5590)	4.1465*** (0.1482)	6.1217*** (0.1868)	8.3208*** (0.1897)
$a(m)$	1.0137*** (0.0460)	2.6856*** (0.0512)	0.6473*** (0.0297)	2.2720*** (0.0477)	3.1820*** (0.0783)	2.6130*** (0.0560)
Other Parameters						
$\eta_{vol}(m)$	-0.5982*** (0.0580)	-2.1542*** (0.0487)	-0.2173*** (0.0349)	-1.9628*** (0.0486)	-2.7864*** (0.0836)	-2.6960*** (0.0647)
π_{mm}	0.8364*** (0.0233)	0.9720*** (0.0044)	0.8338*** (0.0215)	0.9635*** (0.0052)	0.2646*** (0.0184)	0.5347*** (0.0166)
Diagnostic Statistics						
Obs. in $\hat{s}_i = m$	236	1432	273	1397	540	935
$E[\hat{\epsilon}_i]$	1.0011		0.9982		0.9658	
$\sigma[\hat{\epsilon}_i]$	1.0004		1.0706		0.9508	
CvM	0.0968		0.1971		0.1529	
AD	0.5669		1.2970		1.2566	
LB(50)	44.55		79.82***		53.17*	
$\widehat{SoR}(m)$	0.9508	0.9798	0.9582	0.9814	0.6581	0.7598
\widehat{SoR}	0.9759		0.9776		0.7298	

Note: Standard errors are in parentheses. ***, ** and * represent significance at 1%, 5% and 10% respectively. Observation counts and residual statistics are based on the estimated most probable state sequence. The CvM and AD are Cramér-von-Mises and Andersen-Darling statistics (Stephens, 2013) for unit exponential distribution, with the critical values bootstrapped from 100000 simulated unit exponential vectors. LB(50) is the Ljung-Box (Ljung and Box, 1978) test statistics at lag 50. Definitions of \widehat{SoR} and $\widehat{SoR}(m)$ can be found in (3.42) and (3.43).

To show that the two regimes estimated by the MS(2)-ACI(2,1)-V model indeed

Figure 3.4 Distribution of estimated regimes over time for AIG 2016-03, INTC 2016-08 and SPY 2016-05



Note: The x-axis represents fraction of trading hours with. The y-axis is the percentage of data in each regime. Regime 1 and 2 correspond to the most probable state vector estimated in Table 3.4. Each bar counts the percentage of data falling into a roughly 20-minute bin.

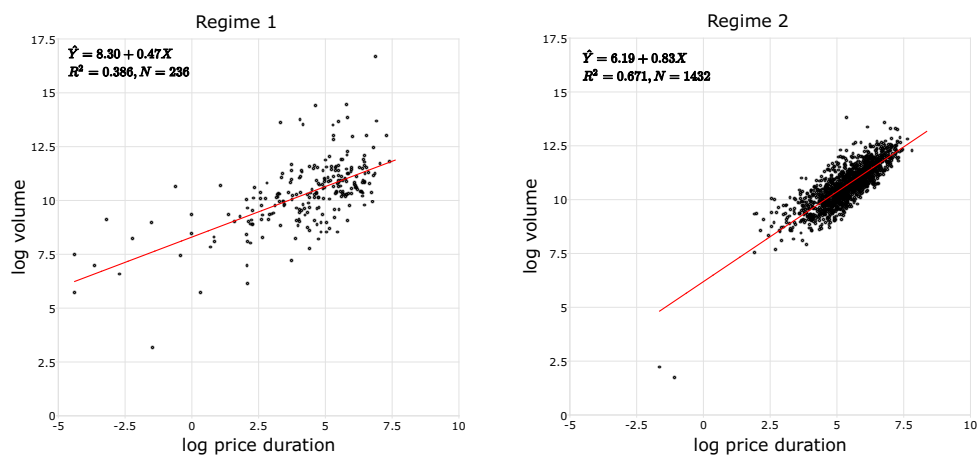
correspond to the regime-switching behaviour captured by the OLS regression in the data section, we firstly plot the distribution of regimes against calendar time in Figure 3.4. In the figure it is clear that for individual stocks, most of the observations in regime 1 concentrate during the first 20 minutes of a trading day, and have almost no occurrences in the middle of the trading hours. Observations in regime 2 are more evenly spread throughout the day. Moreover, we could not detect such a pattern for SPY, since the two regimes have similar distributions throughout the trading hours. Figure 3.4 shows that the Markov chain assumption for the state vector of the individual stocks are invalid due to the diurnal pattern, and the model can be potentially improved by modelling the diurnal pattern of the transitional probabilities explicitly.

More interestingly, the above results provide strong evidence for firm-specific information content being captured by regime 1 for the individual stocks, as the distribution of regime 1 across time resembles the well-documented diurnal pattern of price volatility (see, e.g. Andersen and Bollerslev (1997b)). This is also in line with Chordia, Roll, and Subrahmanyam (2005) who find that new information is usually updated to the price process within 30 minutes. To show that our findings are consistent on all the stock-months considered, for each security we aggregate the observations classified as regime 1 and 2 for the year 2016, and plot a yearly version of Figure 3.4 for each security in Figure C.4 in Appendix C.7.

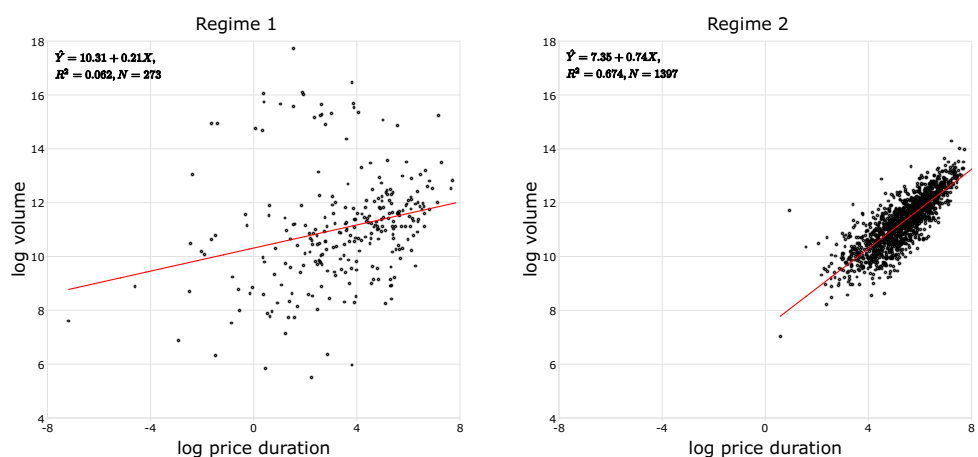
We then produce scatter plots (Figure 3.5) and R^2 difference plots (Figure 3.6)

Figure 3.5 Scatter plots with regression line for AIG 2016-03, INTC 2016-08, SPY 2016-05 based on estimated regimes

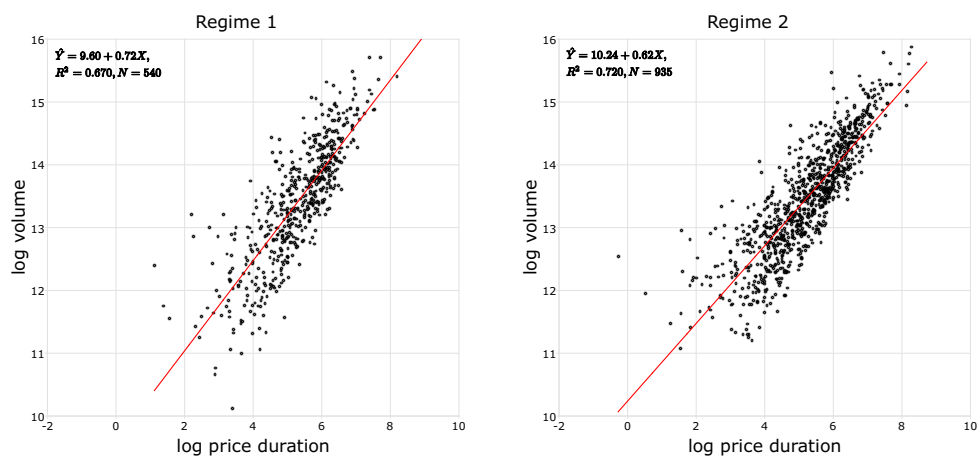
Panel 1: AIG 2016-03



Panel 2: INTC 2016-08



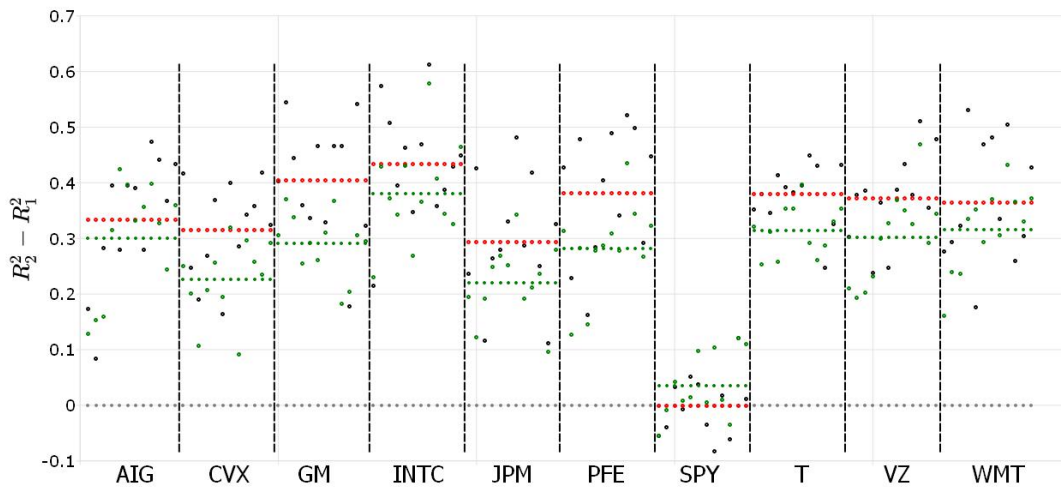
Panel 3: SPY 2016-05



Note: Three 1-by-2 graphs from top to bottom: AIG 2016-03, INTC 2016-08, SPY 2016-05. X-axis: $\ln x_{i,d}^\delta$. Y-axis: $\ln \hat{Vol}_{i,d}^\delta$. The regression line is obtained by (3.45), with regime 1 and 2 corresponding to the estimated regimes in Table 3.4.

similar to Figures 3.2 and 3.3 using the estimated states. Comparing Figure 3.5 with Figure 3.2, we can clearly observe the effect of the state classification of the MS-ACI model. For AIG and INTC, the estimated \hat{b}_1 and R^2 for regime 1 are smaller compared to regime 2.¹¹ This is also reflected in Figure 3.6 as the R^2 differences between regime 1 and 2 enlarge for all individual securities considered, compared to those in Figure 3.3. For the SPY, we do not see any significance difference for the scatter plots in Figures 3.2 and 3.5. The average R^2 difference even decreases to nearly zero in Figure 3.6, indicating that the MS-ACI model does not detect a regime-switching behaviour in the volume-volatility relationship of the SPY.

Figure 3.6 R^2 difference for the volume-duration regressions between observations in regime 1 and 2 using the estimated state vector



Note: The figure plots $\hat{R}_2^2 - \hat{R}_1^2$ obtained from regression (3.45) for 10×12 stock-month datasets, with \hat{R}_1^2 and \hat{R}_2^2 being the estimated R^2 from the observations classified as regime 1 and 2 respectively. The estimated state vector is obtained by the MS(2)-ACI(2,1)-V model. Each black dot represents the R^2 difference for one stock-month dataset. The vertical black dashed lines split observations from each stock, and between two vertical red lines, the R^2 differences are ordered chronologically. The horizontal red dots represents the yearly average R^2 difference for each stock. Figure 3.3 is also plotted as green dots for comparison.

Concluding our empirical analysis, our MS-ACI model detects two distinct regimes in the high-frequency volume-volatility relationship of 9 individual securities. Observations from regime 1 cluster at the start and the end of a trading day, which matches the typical diurnal pattern of information arrivals. Observations from regime 2 distributes evenly across the trading hours, and price durations and volume within this regime has a much stronger correlation compared to the other regime. This regime-switching behaviour cannot be found on the stock index ETF, SPY. We claim that regime 1 is related to arrivals of firm-specific information into the market

¹¹The findings on the differences between \hat{b}_1 for regime 1 and 2 are also consistent across all stock-months, as is shown in Figure C.5 in Appendix C.7.

based on the findings above. We propose a simple market microstructure-based model that explains the lower $\hat{\beta}_1$ and R^2 from the regime 1 as a result of informed trading. The detailed model is presented in Appendix C.6. To summarize the model briefly, the presence of informed trading can be viewed as an omitted explanatory variable in the residual that is negatively correlated with price duration, thus resulting in lower $\hat{\beta}_1$ and R^2 in the volume-duration regression for the informed regime.

Based on the above arguments, our model has several important implications. Firstly, our model brings new insight to the information arrival theory on a high-frequency level. With the estimates of most probable states, we can measure the speed of overnight news updates and the informed volume impact of volatility directly in a fully parametric framework. Secondly, the most probable state vector and the posterior probabilities of the state vector can serve as a high-frequency measure of the most probable state of the market (with or without information content) and the probability of the state. Lastly, the difference $|\hat{\eta}_{vol}(1) - \hat{\eta}_{vol}(2)|$ can be used to assess the average impact of information flow on the volatility process.

3.6 Concluding Remarks

This chapter develops the Markov-Switching ACI (MS-ACI) model by extending the Markov-switching structure to the original ACI model (Russell, 1999). The stationarity and moment conditions for both the original ACI model and the MS-ACI model under the exponential baseline are derived, which augments the stationarity condition provided by Hautsch (2012). To overcome the path dependency problem in the estimation of Markov-switching autoregressive models, we propose to estimate the model via the SAEM algorithm combined with an approximated version of Bauwens, Preminger, and Rombouts's (2010) single move sampler. By introducing the concept of Significance of Regimes (*SoR*), we demonstrate in our simulation that our SAEM estimation scheme is capable of providing reliable parameter estimates for the MS-ACI model in both the two- and three-regime case when there are significant structural breaks in the data generating process. The quality of parameter estimates of the MS-ACI model depends crucially on the magnitude of the discrepancy between the regime-switching baselines of the intensity, which is summarized by our *SoR* statistic. Our simulation analysis subsequently shows that the *SoR* statistic can serve as an indicator of the relative quality of parameter estimates of the MS-ACI model in comparison with the complete model assuming the latent vector is given, which provides a diagnostic tool in the empirical application. The simulation also presents the reliability of our estimation scheme and the improvement in efficiency of our

approximated single step sampler.

The existing literature suggests that only the informed part, or the unanticipated part of the trading volume should have an impact on price volatility (Admati and Pfleiderer, 1988; Andersen, 1996; Aragó and Nieto, 2005; Bessembinder and Seguin, 1993; Clark, 1973; Copeland and Galai, 1983; Copeland, 1976; Easley, Kiefer, O'Hara, and Paperman, 1996; Epps and Epps, 1976; Hussain, 2011; Kyle, 1985). We propose a novel approach to split high-frequency transaction data based on the correlation between trading volume and price duration, and relate this correlation to the information content of the price duration through a simple market microstructure model. Based on our MS-ACI model, we detect two distinct regimes for all 9 stocks, namely, a dominant regime with low information content and a minor regime with high information content. The absence of this regime-switching behaviour in the SPY data supports that this effect is mainly due to firm-specific news arrivals. As a result, our model provides measures of information content for each price duration as well as the impact of informed trading volume on price volatility.

The regimes we detect only exploit correlations between trading volume and price duration, and the conclusions on information content of regimes are based on certain market microstructure models, such as Glosten and Milgrom (1985) or Easley, Kiefer, O'Hara, and Paperman (1996). Therefore, our results are rather indirect, which is somewhat similar to the PIN-type measures initially proposed by Easley, Kiefer, O'Hara, and Paperman (1996). It would be helpful to address this issue by including some physical news data to further test the regime classifications. This, however, is beyond the scope of this chapter. In all, the validity of the information content argument is irrelevant to the econometric properties of the model itself, as we demonstrate that our model is fully capable of modelling regime-switching behaviour in point processes.

Chapter 4

Conclusion

4.1 Summary of Main Findings

This thesis expands existing literature on point process based volatility estimator by developing the following aspects of the estimator: (1) the asymptotic property of the estimator; (2) the roles and relative importances of MMS covariates in high-frequency volatility estimation with the estimator; (3) using the estimator as a platform to model intraday regime-switching volume-volatility relationships.

In the first chapter, we propose the *RBV* class of estimators by generalizing the original duration-based volatility estimator (Engle and Russell, 1998; Nolte, Taylor, and Zhao, 2018; Tse and Yang, 2012). We prove its asymptotic properties and show that the *NPD* estimator (Nolte, Taylor, and Zhao, 2018) can be regarded as both an RV estimator and a *RBV* estimator that is more efficient than RV estimators under calendar time or tick time sampling. We also introduce the *NPR* estimator which is in theory twice as efficient as the *NPD* estimator. More importantly, we demonstrate that imposing a parametric structure on the *RBV* model can potentially further improve its efficiency in volatility estimation. These results establish theoretical foundation for the simulation and empirical results in Engle and Russell (1998), Tse and Yang (2012), and Nolte, Taylor, and Zhao (2018).

We examine the properties of the *NPD* estimator in detail, and analyse its bias in the presence of MMS noise, jumps, time discretization and price discretization. We show that the theoretical superior efficiency of the *NPD* estimator over calendar time RV can only be achieved when we sample sparsely enough, as the *NPD* estimator is more sensitive to MMS noise than calendar time RV. To resolve this problem, we propose the *NPD^z* estimator which smooth the observed price exponentially, and we show that by choosing the smoothing parameter optimally, the *NPD^z* estimator

outperforms common calendar time noise-robust RV estimators such as RK and pre-averaged RV when the sampling frequency is moderate to small.

In the second chapter, we include MMS covariates in the LL-ACD model, and show that the inclusion leads to more precise volatility estimates. Specifically, we include both contemporaneous and one duration lagged versions of trading volume, bid-ask spread, order flow, order imbalance, number of trades, quote depth and quote difference measures in the LL-ACD model, and use a BSR approach to rank and select the optimal number of variables based on their contribution to the goodness-of-fit of the model. Our findings suggest that the contemporaneous number of trades and order flow are the two variables that are considered most important in intraday volatility modelling, and that contemporaneous variables are generally more important than the one duration lagged ones. The rankings of the variables do vary considerably cross-sectionally, but are fairly stable in different periods. Moreover, we find that including the optimally selected MMS covariates can significantly improve the goodness-of-fit of the LL-ACD model in both in-sample and out-of-sample settings, while including all MMS covariates lead to a deterioration in the out-of-sample performance of the model. This result validates the reliability of the variable selection from the BSR method.

Using trade and quote data from 29 highly liquid stocks and 1 market index ETF, we construct daily and intraday volatility estimates from the plain LL-ACD model (-P model), the LL-ACD model with optimally chosen MMS covariates (-K model) and the LL-ACD model with all MMS covariates (-A model). Using daily and intraday Realized Kernels as a benchmark, our results suggest that volatility estimates from the -K model always significantly outperform those from the -P model by having a smaller distance from the benchmark, whereas the performance of the -A model does not differ significantly from the -K model. More importantly, we show that the LL-ACD-K model can deliver accurate volatility estimate even when the Realized Kernel cannot be reliably constructed due to a lack of data.

In Chapter 3, we develop the MS-ACI model and updates the stationarity condition for the ACI and the MS-ACI model. We propose to use the SAEM algorithm to estimate our model, and discuss details of the implementation of the algorithm and evaluation of the goodness-of-fit of the model. Specifically, we introduce the *SoR* measure and argue that it can be used to assess the necessity of a regime switching structure in the model. Our simulation results show that the MS-ACI model can be reliably estimated by the SAEM algorithm, and that the MS-ACI model performs

best when the underlying regime-specific densities are distinct from each other.

We apply the MS-ACI model to the trade and quote data of 9 highly liquid stocks and a stock index ETF to examine whether there exist intraday regime-switching volume-volatility relationships in the data. Our empirical findings suggest that there exist two distinct regimes for the stocks, but not for the stock index. One regime concentrates at the beginning and end of a trading day, which we believe is associated with a higher information content. Another regime spreads evenly through out trading days, and can be interpreted as a regime with low information content. Our MS-ACI model provides new insights into intraday dynamics of volatility and trading process, and measures of the informativeness of the market can be constructed based on the posterior probability of regime classification.

4.2 Implications

This thesis has several important implications for both academic researchers and practitioners. Firstly, results in Chapter 1 indicate that the renewal sampling scheme is a superior sampling scheme than other conventional sampling schemes for RV, such as calendar time or tick time. The main advantage of the renewal sampling scheme is that one uses the price path to construct the renewal sampling scheme, which effectively uses all the observations without any discards. Chapter 1 also explains the reason why the *PD* estimator in Tse and Yang (2012) outperforms the RV-type estimators, as the parametric structure improves the precision of volatility estimates for each price duration.

Chapter 2 highlights that information in the MMS variables are indeed highly relevant to high-frequency volatility estimation. As the MMS variables are almost always discarded in the RV approach, it further illustrates the importance of a parametric structure, which enables the use of a richer information set to pin down the intraday volatility process. Concluding from above, we highly advocate the use of the parametric point process based approach over the RV approach, as the point process based approach is less dependent on the availability of intraday data, thus can be applied to a wider set of assets.

Finally, Chapter 3 presents some interesting possibilities on the application of the point process based volatility estimator to MMS research. Although volatility is a crucial input in MMS models, its latent nature presents a big challenge for researchers to include it in a high-frequency MMS model. The parametric point process based

approach enables researchers to analyse relationship between MMS covariates and a valid volatility estimator, which renders the empirical applications of MMS theories more convincing.

4.3 Limitations and Future Researches

As with all studies, this thesis has various limitations which provide grounds for future research.

One of the limitations that applies to all three chapters is a good rule to select the truncation threshold δ for the duration-based volatility estimator. In Chapter 1, we conclude that δ should be large enough so that the MMS noise effect can be ignored, similar to the RV-type measures. Therefore, in Chapter 2 and 3 we apply a similar criterion as in the RV literature and choose a δ such that the sampling frequency corresponds to once every five minutes. However, the performance of this choice is not assessed empirically in this thesis. This choice has two more drawbacks that requires further clarification: (1) By choosing δ depending on the average length of price durations, the chosen δ is itself an endogenous random variable that depends on volatility of the same period, and the effect of a random δ on the *RBV* and *PRBV* estimators is yet to be verified. (2) This choice of δ does not consider the size of the MMS noise in general. From Chapter 1, it may be possible to choose δ based on the autocorrelation of the returns, but this is also subject to future research.

In Chapter 1, we do not provide the asymptotic properties when the intensity function of the *PRBV* estimator needs to be estimated by a parametric model instead of assuming to be known. This result is actually derived theoretically and will be covered by the following paper which focuses on the *PRBV* estimator. Due to length constraints, it is not included in the paper. Chapter 1 also leaves some interesting questions that calls for future research. Firstly, the properties of the range duration based volatility estimators are to be verified. Secondly, the existence of a simple method to correct for the MMS and time discretization bias for the *NPD* estimator is still unclear. Finally, the use of exponentially smoothing the observed price process to mitigate MMS noise can be verified in detail.

Chapter 2 also leaves some open questions to be addressed. Firstly, we need to extend the theoretical framework in Chapter 1 to account for the explanatory variables in the ACD model. This is because the volatility estimates are derived by conditioning on a richer information set, and the renewal property of the point process

in business time may not hold in this situation. Also, it would be interesting to include other sources of information in the model, such as public news data and high-frequency option data, and assess whether they provide extra information to model high-frequency volatility in addition to the MMS variables used in this chapter.

In Chapter 3, a potential limitation of the empirical findings is that the link from the regime classifications to the information content of the regimes requires further verification, due to the latent nature of information arrivals into the market. One could further augment the model by using physical news arrival data to examine the validity of this link.

4.4 Final Remarks

To sum up, this thesis presents important developments to the literature of point process based volatility estimators both in theory and applications. The findings show that the point process based estimator uses data more efficiently than the widely-applied RV approach, and has the advantage to use information other than the price process in volatility modelling and estimation. This thesis establishes the theoretical foundation of the point process based volatility estimator and advocates the use of point process based approach in future volatility modelling and MMS research.

Bibliography

- ADMATI, A. R., AND P. PFLEIDERER (1988): “A Theory of Intraday Patterns: Volume and Price Variability,” *The Review of Financial Studies*, 1(1), 3–40.
- AHN, H.-J., K.-H. BAE, AND K. CHAN (2001): “Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong,” *The Journal of Finance*, 56(2), 767–788.
- AÏT-SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2011): “Ultra high frequency volatility estimation with dependent microstructure noise,” *Journal of Econometrics*, 160(1), 160–175.
- ALLASSONNIÈRE, S., E. KUHN, AND A. TROUVÉ (2010): “Construction of Bayesian Deformable Models via Stochastic Approximation Algorithm: A Convergence Study,” *Bernoulli*, 16(3), 641–678.
- ALLEN, D., F. CHAN, M. MCALEER, AND S. PEIRIS (2008): “Finite sample properties of the QMLE for the Log-ACD model: Application to Australian stocks,” *Journal of Econometrics*, 147(1), 163–185.
- AMIHUD, Y. (2002): “Illiquidity and stock returns: Cross-section and time-series effects,” *Journal of Financial Markets*, 5(1), 31–56.
- ANDERSEN, T. G. (1996): “Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility,” *The Journal of Finance*, 51, 169.
- ANDERSEN, T. G., AND T. BOLLERSLEV (1997a): “Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns,” *The Journal of Finance*, 52(3), 975–1005.
- ANDERSEN, T. G., AND T. BOLLERSLEV (1997b): “Intraday periodicity and volatility persistence in financial markets,” *Journal of Empirical Finance*, 4(2-3), 115–158.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): “The distribution of realized stock return volatility,” *Journal of Financial Economics*, 61, 43–76.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2000): “Great Realizations,” *Risk*, 13(3), 105–108.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): “The Distribution of Realized Exchange Rate Volatility,” *Journal of the American Statistical Association*, 96(453), 42–55.

- ANDERSEN, T. G., T. BOLLERSLEV, AND D. DOBREV (2007): “No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications,” *Journal of Econometrics*, 138(1), 125–180.
- ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2008): “Duration-Based Volatility Estimation,” *Working Paper, Northwestern University*.
- ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2012): “Jump-robust volatility estimation using nearest neighbor truncation,” *Journal of Econometrics*, 169(1), 75–93.
- ARAGÓ, V., AND L. NIETO (2005): “Heteroskedasticity in the returns of the main world stock exchange indices: Volume versus GARCH effects,” *Journal of International Financial Markets, Institutions and Money*, 15(3), 271–284.
- BANDI, F. M., AND J. R. RUSSELL (2008): “Microstructure noise, realized variance, and optimal sampling,” *Review of Economic Studies*, 75(2), 339–369.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008a): “Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76(6), 1481–1536.
- (2008b): “Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76(6), 1481–1536.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): “Realized kernels in practice: Trades and quotes,” *Econometrics Journal*, 12(3).
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2011): “Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading,” *Journal of Econometrics*, 162, 149–169.
- BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002): “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(2), 253–280.
- (2003): “Realized power variation and stochastic volatility models,” .
- (2004): “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, 2, 1–48.
- BARNDORFF-NIELSEN, O. E., AND A. SHIRYAEV (2010): *Change of Time and Change of Measure*. World Scientific Publishing Company, Singapore.
- BAUWENS, L., A. DUFAYS, AND J. V. K. ROMBOUTS (2014): “Marginal likelihood for Markov-switching and change-point GARCH models,” *Journal of Econometrics*, 178, 508–522.
- BAUWENS, L., AND N. HAUTSCH (2006): “Modelling Financial High Frequency Data using Point Processes,” *Social Science Research Network Working Paper Series*.

- BAUWENS, L., A. PREMINGER, AND J. V. K. ROMBOUTS (2010): “Theory and inference for a Markov switching GARCH model,” *Econometrics Journal*, 13(2), 218–244.
- BEALE, E. M. L., M. G. KENDALL, AND D. W. MANN (1967): “The discarding of variables in multivariate analysis,” *Biometrika*, 54(3-4), 357–366.
- BERTSIMAS, D., A. KING, AND R. MAZUMDER (2016): “Best subset selection via a modern optimization lens,” *Annals of Statistics*, 44(2), 813–852.
- BESSEMBINDER, H., AND P. SEGUIN (1993): “Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets,” *The Journal of Financial and Quantitative Analysis*, 28(1), 21–39.
- BILLINGSLEY, P. (2009): *Convergence of Probability Measures*. Wiley, New York.
- BILLIO, M., R. CASARIN, AND A. OSUNTUYI (2014): “Efficient Gibbs sampling for Markov switching GARCH models,” *Computational Statistics & Data Analysis*, *Accepted manuscript*.
- BOLLERSLEV, T. (1986): “Generalised Autoregressive Conditional Heteroscedasticity,” *Journal of Econometrics*, 31, 307–327.
- BOLLERSLEV, T., J. LITVINOVA, AND G. TAUCHEN (2006): “Leverage and volatility feedback effects in high-frequency data,” *Journal of Financial Econometrics*, 4(3), 353–384.
- BOLLERSLEV, T., AND M. MELVIN (1994): “Bid-Ask Spreads and Volatility in the Foreign Exchange Market: An Empirical Analysis,” *Journal of International Economics*, 36, 355–372.
- BOLLERSLEV, T., AND J. M. WOOLDRIDGE (1992): “Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances,” *Econometric Reviews*, 11(2), 143–172.
- BOOTH, J. G., AND J. P. HOBERT (1999): “Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1), 265–285.
- BOUGEROL, P., AND N. PICARD (1992): “Strict stationarity of generalized autoregressive processes,” *The Annals of Probability*, 20(4), 1714–1730.
- BOWSER, C. G. (2007): “Modelling security market events in continuous time: Intensity based, multivariate point process models,” *Journal of Econometrics*, 141(2002), 876–912.
- BOYLES, R. A. (1983): “On the Convergence of the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(1), 47–50.
- BRENNAN, M. J., AND A. SUBRAHMANYAM (1996): “Market microstructure and asset pricing: On the compensation for illiquidity in stock returns,” *Journal of Financial Economics*, 41(3), 441–464.
- BROWN, T. C., AND M. G. NAIR (1988): “A Simple Proof of the Multivariate Random Time Change Theorem for Point Processes,” *Journal of Applied Probability*, 25(1), 210–214.

- CAI, J. (1994): “A Markov Model of Switching-Regime ARCH,” *Journal of Business & Economic Statistics*, 12(3), 309–316.
- CELEUX, G., D. CHAUVEAU, AND J. DIEBOLT (1996): “Stochastic versions of the em algorithm: an experimental study in the mixture case,” *Journal of Statistical Computation and Simulation*, 55(January 2015), 287–314.
- CELEUX, G., AND J. DIEBOLD (1985): “The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problems,” *Computational Statistics Quarterly*, 2, 73–82.
- CELEUX, G., AND J. DIEBOLT (1992): “A stochastic approximation type EM algorithm for the mixture problem,” *Stochastics and Stochastic Reports*, 41(1-2), 119–134.
- CHAN, K., AND W.-M. FONG (2000): “Trade size, order imbalance, and the volatility-volume relation,” *Journal of Financial Economics*, 57(2), 247–273.
- CHEN, F., F. X. DIEBOLD, AND F. SCHORFHEIDE (2013): “A Markov-switching multifractal inter-trade duration model, with application to US equities,” *Journal of Econometrics*, 177(2), 320–342.
- CHORDIA, T., R. ROLL, AND A. SUBRAHMANYAM (2002): “Order imbalance, liquidity, and market returns,” *Journal of Financial Economics*, 65(1), 111–130.
- (2005): “Evidence on the speed of convergence to market efficiency,” *Journal of Financial Economics*, 76(2), 271–292.
- CHRISTENSEN, K., AND M. PODOLSKIJ (2007): “Realized range-based estimation of integrated variance,” *Journal of Econometrics*, 141(2), 323–349.
- CLARK, P. K. (1973): “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices,” *Econometrica*, 41, 135–155.
- COLEMAN, R. (1982): “The Moments of Recurrence Time,” *European Journal of Operational Research*, 9(2), 181–183.
- COPELAND, T., AND D. GALAI (1983): “Information effects on the bid-ask spread,” *The Journal of Finance*, 38, 1457–1469.
- COPELAND, T. E. (1976): “A Model of Asset Trading under the Assumption of Sequential Information Arrival,” *The Journal of Finance*, 31(4), 1149–1168.
- CORSI, F. (2009): “A simple approximate long-memory model of realized volatility,” *Journal of Financial Econometrics*, 7, 174–196.
- DALEY, D. J., AND D. VERE-JONES (2003): *An introduction to the theory of point processes*, vol. I. Springer Science & Business Media.
- DARRAT, A. F., M. ZHONG, AND L. T. W. CHENG (2007): “Intraday volume and volatility relations with and without public news,” *Journal of Banking and Finance*, 31, 2711–2729.
- DELATTRE, S., AND J. JACOD (1997): “A central limit theorem for normalized functions of increments of a diffusion process, in the presence of round-off errors,” *Bernoulli*, 3(1), 1–28.

- DELYON, B., M. LAVIELLE, AND E. MOULINES (1999): "Convergence of a Stochastic Approximation Version of the EM Algorithm," *The Annals of Statistics*, 27(1), 94–128.
- DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- DIAMOND, D. W., AND R. E. VERRECCHIA (1987): "Constraints on short-selling and asset price adjustment to private information," *Journal of Financial Economics*, 18(2), 277–311.
- DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 20(1), 134–144.
- DIEBOLT, J., AND E. IP (1996): "A stochastic EM algorithm for approximating the maximum likelihood estimate," in *Markov Chain Monte Carlo in Practice*, ed. by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Chapman and Hall, London.
- DOOB, J. L. (1948): "Renewal Theory From the Point of View of the Theory of Probability," *Transactions of the American Mathematical Society*, 63(3), 422–438.
- DUEKER, M. (1997): "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility," *Journal of Business & Economic Statistics*, 15(1), 26–34.
- DUFOUR, A., AND R. F. ENGLE (2000): "Time and the Price Impact of a Trade," *The Journal of Finance*, 55(6), 2467–2498.
- EASLEY, D., N. KIEFER, M. O'HARA, AND J. PAPERMAN (1996): "Liquidity, information, and infrequently traded stocks," *Journal of Finance*, 51, 1405–1436.
- EASLEY, D., M. M. LÓPEZ DE PRADO, AND M. O'HARA (2012): "Flow Toxicity and Liquidity in a High Frequency World," *Review of Financial Studies*, 25, 1457–1493.
- EASLEY, D., AND M. O'HARA (1992): "Time and the Process of Security Price Adjustment," *The Journal of Finance*, 47(2), 577–605.
- ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.
- ENGLE, R. F., AND J. R. RUSSELL (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.
- EPPS, T. W., AND L. M. EPPS (1976): "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implication For The Mixture-of-Distributions Hypothesis," *Econometrica*, 44, 305–321.
- FELLER, W. (1941): "On the Integral Equation of Renewal Theory," *Annals of Mathematical Statistics*, 12, 243–267.
- FOUCAULT, T. (1999): "Order flow composition and trading costs in a dynamic limit order market," *Journal of Financial Markets*, 2(2), 99–134.
- FRANCQ, C., AND J.-M. ZAKOÏAN (2001): "Stationarity of multivariate Markov-switching ARMA models," *Journal of Econometrics*, 102, 339–364.

- FUKASAWA, M. (2010a): “Central limit theorem for the realized volatility based on tick time sampling,” *Finance and Stochastics*, 14(2), 209–233.
- (2010b): “Realized volatility with stochastic sampling,” *Stochastic Processes and their Applications*, 120(6), 829–852.
- FUKASAWA, M., AND M. ROSENBAUM (2012): “Central limit theorems for realized volatility under hitting times of an irregular grid,” *Stochastic Processes and their Applications*, 122(12), 3901–3920.
- GERHARD, F., AND N. HAUTSCH (2002): “Volatility estimation on the basis of price intensities,” *Journal of Empirical Finance*, 9, 57–89.
- GHYSELS, E., AND J. JASIAK (1998): “GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model,” *Studies in Nonlinear Dynamics & Econometrics*, 2(4), 133.
- GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks,” *The Journal of Finance*, 48, 1779–1801.
- GLOSTEN, L. R., AND P. R. MILGROM (1985): “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,” *Journal of Financial Economics*, 14, 71–100.
- GORDIN, M., AND M. PELIGRAD (2011): “On the functional central limit theorem via martingale approximation,” *Bernoulli*, 17(1), 424–440.
- GRAY, S. F. (1996): “Modeling the conditional distribution of interest rates as a regime-switching process,” *Journal of Financial Economics*, 42(1), 27–62.
- GRIFFIN, J. E., AND R. C. A. OOMEN (2008): “Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?,” *Econometric Reviews*, 27(1-3), 230–253.
- GUT, A. (2012): “Anscombe’s Theorem 60 Years Later,” *Sequential Analysis*, 31(3), 368–396.
- HAAS, M., S. MITTNIK, AND M. S. PAOLELLA (2004): “A New Approach to Markov-Switching GARCH Models,” *Journal of Financial Econometrics*, 2(4), 493–530.
- HAMILTON, J. D., AND R. SUSMEL (1994): “Autoregressive conditional heteroskedasticity and changes in regime,” *Journal of Econometrics*, 64(1-2), 307–333.
- HANDA, P., AND R. A. SCHWARTZ (1996): “Limit Order Trading,” *The Journal of Finance*, 51(5), 1835–1861.
- HANSEN, P. R., AND A. LUNDE (2006): “Realized Variance and Market Microstructure Noise,” *Journal of Business & Economic Statistics*, 24(2), 127–161.
- HARVEY, D., S. LEYBOURNE, AND P. NEWBOLD (1997): “Testing the equality of prediction mean squared errors,” *International Journal of Forecasting*, 13(2), 281–291.

- HASBROUCK, J. (1991): “Measuring the Information Content of Stock Trades,” *The Journal of Finance*, XLVI(1), 179–207.
- HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Predictions*. Springer, New York, 2nd editio edn.
- HASTIE, T., R. TIBSHIRANI, AND R. J. TIBSHIRANI (2017): “Extended Comparisons of Best Subset Selection, Forward Stepwise Selection, and the Lasso,” *Stanford University Working Paper*.
- HÄUSLER, E., AND H. LUSCHGY (2015): *Stable Convergence and Stable Limit Theorems*. Springer International Publishing, Switzerland.
- HAUTSCH, N. (2012): *Econometrics of Financial High-Frequency Data*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- HAUTSCH, N., AND M. PODOLSKIJ (2013): “Preaveraging-Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence,” *Journal of Business and Economic Statistics*, 31(2), 165–183.
- HOCKING, R. R., AND R. N. LESLIE (1967): “Selection of the Best Subset in Regression Analysis,” *Technometrics*, 9(4), 531–540.
- HOLDEN, C. W., AND S. JACOBSEN (2014): “Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions,” *Journal of Finance*, 69(4), 1747–1785.
- HUANG, X., AND G. TAUCHEN (2005): “The relative contribution of jumps to total price variance,” *Journal of Financial Econometrics*, 3(4), 456–499.
- HUJER, R., S. VULETIC, AND S. KOKOT (2002): “The Markov Switching ACD Model,” *Working Paper Series: Finance and Accounting, Johann Wolfgang Goethe-Universität Frankfurt a.A.*, No.90.
- HUSSAIN, S. M. (2011): “The Intraday Behaviour of Bid-Ask Spreads , Trading Volume and Return Volatility : Evidence from DAX30,” *Journal of Economics and Finance*, 3, 23–34.
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): “Microstructure noise in the continuous case: The pre-averaging approach,” *Stochastic Processes and their Applications*, 119(7), 2249–2276.
- JACOD, J., Y. LI, AND X. ZHENG (2017): “Statical Property of Market Microstructure Noise,” *Econometrica*, 8(4), 1133–1174.
- JANK, W. (2005): “Quasi-Monte Carlo sampling to improve the efficiency of Monte Carlo EM,” *Computational Statistics & Data Analysis*, 48(4), 685–701.
- (2006): “Implementing and Diagnosing the Stochastic Approximation EM Algorithm,” *Journal of Computational and Graphical Statistics*, 15(4), 803–829.
- JONDEAU, E., J. LAHAYE, AND M. ROCKINGER (2015): “Estimating the price impact of trades in a high-frequency microstructure model with jumps,” *Journal of Banking and Finance*, 61, S205–S224.

- KALEV, P. S., W. M. LIU, P. K. PHAM, AND E. JARNECIC (2004): “Public information arrival and volatility of intraday stock returns,” *Journal of Banking and Finance*, 28(6), 1441–1467.
- KARR, A. (1991): *Point Processes and Their Statistical Inferences*. Dekker, New York.
- KLAASSEN, F. (2002): “Improving GARCH volatility forecasts with regime-switching GARCH,” *Empirical Economics*, 27(2), 363–394.
- KUHN, E., AND M. LAVIELLE (2004): “Coupling a stochastic approximation version of EM with an MCMC procedure,” *ESAIM: Probability and Statistics*, 8(August), 115–131.
- KYLE, A., AND A. OBIZHAEVA (2012): “Market Microstructure Invariants: Theory and Implications of Calibration,” *Portfolio The Magazine Of The Fine Arts*.
- KYLE, A. S. (1985): “Continuous Auctions and Insider Trading,” *Econometrica*, 53(6), 1315–1335.
- LAMOUREUX, C. G., AND W. D. LASTRAPES (1990): “Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects,” *The Journal of Finance*, 45(1), 221–229.
- LEE, C. M. C., AND M. J. READY (1991): “Inferring Trade Direction from Intraday Data,” *The Journal of Finance*, 46, 733–746.
- LEE, S. S., AND J. HANNIG (2010): “Detecting jumps from Lévy jump diffusion processes,” *Journal of Financial Economics*, 96(2), 271–290.
- LI, Y., AND P. A. MYKLAND (2015): “Rounding errors and volatility estimation,” *Journal of Financial Econometrics*, 13(2), 478–504.
- LI, Y., I. NOLTE, AND S. NOLTE (2018a): “Asymptotic Theory for Renewal Based High-Frequency Volatility Estimation,” in *Point Process Based High-Frequency Volatility Estimation: Theory and Applications (Doctoral Thesis)*, chap. 1. Lancaster University, Lancaster, UK.
- (2018b): “High-Frequency Volatility Estimation and the Relative Importance of Market Microstructure Effects,” in *Point Process Based High-Frequency Volatility Estimation: Theory and Applications (Doctoral Thesis)*, chap. 2. Lancaster University, Lancaster, UK.
- (2018c): “High-Frequency Volatility Modelling : A Markov-Switching Autoregressive Conditional Intensity Model,” in *Point Process Based High-Frequency Volatility Estimation: Theory and Applications (Doctoral Thesis)*, chap. 3. Lancaster University, Lancaster, UK.
- LIESENFELD, R., I. NOLTE, AND W. POHLMEIER (2006): “Modelling financial transaction price movements: A dynamic integer count data model,” *Empirical Economics*, 30(4), 795–825.
- LIU, L. Y., A. J. PATTON, AND K. SHEPPARD (2015): “Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes,” *Journal of Econometrics*, 187(1), 293–311.

- LJUNG, G. M., AND G. E. P. BOX (1978): “On a measure of lack of fit in time series models,” *Biometrika*, 65, 297–303.
- LOMB, N. R. (1976): “Least-Squares Frequency-Analysis of Unequally Spaced Data,” *Astrophysics and Space Science*, 39(2), 447–462.
- LOTOV, I. V. (1996): “On Some Boundary Crossing Problems for Gaussian Random Walks,” *The Annals of Probability*, 24(4), 2154–2171.
- LOUIS, T. A. (1982): “Finding the Observed Information Matrix when Using the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2), 226–233.
- MADHAVAN, A., M. RICHARDSON, AND M. ROOMANS (1997): “Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks,” *The Review of Financial Studies*, 10(4), 1035–1064.
- MANGANELLI, S. (2005): “Duration, volume and volatility impact of trades,” *Journal of Financial Markets*, 8(4), 377–399.
- NÆS, R., AND J. A. SKJELTORP (2006): “Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market,” *Journal of Financial Markets*, 9(4), 408–432.
- NIELSEN, F. (2000): “The Stochastic EM Algorithm: Estimation and Asymptotic Results,” *Bernoulli*, 6(3), 457–489.
- NOLTE, I. (2008): “Modeling a multivariate transaction process,” *Journal of Financial Econometrics*, 6, 143–170.
- NOLTE, I., S. TAYLOR, AND X. ZHAO (2018): “More Accurate Volatility Estimation and Forecasts Using Price Durations,” *Working Paper, Lancaster University Management School*.
- OOMEN, R. C. A. (2005): “Properties of bias-corrected realized variance under alternative sampling schemes,” *Journal of Financial Econometrics*, 3(4), 555–577.
- OOMEN, R. C. A. (2006): “Properties of realized variance under alternative sampling schemes,” *Journal of Business and Economic Statistics*, 24(2), 219–237.
- OPSCHOOR, A., N. TAYLOR, M. VAN DER WEL, AND D. VAN DIJK (2014): “Order flow and volatility: An empirical investigation,” *Journal of Empirical Finance*, 28, 185–201.
- PARLOUR, C. A. (1998): “Price Dynamics in Limit Order Markets,” *Review of Financial Studies*, 11(4), 789–816.
- PELIGRAD, M. (1986): “Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey),” in *Dependence in Probability and Statistics: A Survey of Recent Results*, ed. by E. Eberlein, and M. S. Taqqu. Birkhäuser, Boston.
- RAHMAN, S., C. KRISHNAMURTI, AND A. C. LEE (2005): “The dynamics of security trades, quote revisions, and market depths for actively traded stocks,” *Review of Quantitative Finance and Accounting*, 25(2), 91–124.

- ROBBINS, H., AND S. MONRO (1951): “A Stochastic Approximation Method,” *The Annals of Mathematical Statistics*, 22(3), 400–407.
- ROSS, S. M. (1996): *Stochastic Processes*. John Wiley & Sons, Inc., New York.
- RUSSELL, J. R. (1999): “Econometric Modeling of Multivariate Irregularly-Spaced High-Frequency Data,” *Working Paper, University of Chicago*.
- RYDBERG, T. H., AND N. SHEPHARD (2003): “Dynamics of Trade-by-Trade Price Movements: Decomposition and Models,” *Journal of Financial Econometrics*, 1, 2–25.
- SADKA, R. (2006): “Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk,” *Journal of Financial Economics*, 80(2), 309–349.
- SCARGLE, J. D. (1982): “Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data,” *The Astrophysical Journal*, 263, 835–853.
- STELZER, R. (2009): “On Markov-Switching ARMA Processes-Stationarity, Existence of Moments, and Geometric Ergodicity,” *Econometric Theory*, 25(1), 43–62.
- STEPHENS, M. A. (2013): “EDF Statistics for Goodness of Fit and Some Comparisons,” *Journal of the American Statistical Association*, 69(347), 730–737.
- SUOMINEN, M. (2001): “Trading Volume and Information Revelation in Stock Markets,” *The Journal of Financial and Quantitative Analysis*, 36(4), 545.
- TAUCHEN, G. E., AND M. PITTS (1983): “The Price Variability-Volume Relationship on Speculative Markets,” *Econometrica*, 51(2), 485–505.
- TAY, A., C. TING, Y. K. TSE, AND M. WARACHKA (2009): “Using high-frequency transaction data to estimate the probability of informed trading,” *Journal of Financial Econometrics*, 7, 288–311.
- TIBSHIRANI, R. (1994): “Regression Selection and Shrinkage via the Lasso,” *Journal of the Royal Statistical Society B*, 58, 267–288.
- TSE, Y.-K., AND T. T. YANG (2012): “Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach,” *Journal of Business & Economic Statistics*, 30(4), 533–545.
- WEI, G. C., AND M. A. TANNER (1990): “A Monte Carlo Implementation of the EM Algorithm and the Poor Man’s Data Augmentation Algorithms,” *Journal of the American Statistical Association*, 85(411), 699–704.
- WOLD, S., M. SJÖSTRÖM, AND L. ERIKSSON (2001): “PLS-regression: A basic tool of chemometrics,” *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.
- WOLFF, R. W. (1989): *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs NJ.
- WU, C. F. J. (1983): “On the convergence properties of the EM algorithm,” *The Annals of Statistics*, 11(1), 95–103.

- ZHANG, L. (2006): “Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach,” *Bernoulli*, 12(6), 1019–1043.
- ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100, 1394–1411.
- ZOU, H., AND T. HASTIE (2005): “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2), 301–320.

Appendix A

Appendix to Chapter 1

A.1 The Time Changed Compounded Poisson Process

Following Oomen (2006), the price process $\{P(t)\}_{t>0}$ is specified as:

$$P(t) = \sum_{i=1}^{N(t)} r_i, \quad r_i \sim i.i.d. \mathcal{N}(0, \sigma^2), \quad (\text{A.1})$$

where $N(t)$ is a inhomogeneous Poisson process with time-varying intensity function $\lambda(t|\mathcal{F}_t) = \mathbb{E}[N(t)]$. The integrated variance of this process is defined by:

$$IV(0, t) = \sigma^2 \int_0^t \lambda(s|\mathcal{F}_s) ds. \quad (\text{A.2})$$

It then follows directly from Theorem 1.7 that the time changed counting process $\tilde{N}(\tau(t)) = N(t)$ where $\tau(t) = IV(0, t)$ follows a homogeneous Poisson process with constant intensity σ^{-2} . Since r_i is i.i.d., $\tilde{P}(\tau(t)) = P(t)$ is by definition a Lèvy process.

A.2 Proof to Theorem 1.5

We start by listing some important facts about the renewal process $\tilde{X}(\tau(t))$ and the inter-epoch durations \tilde{D}_i in business time. Firstly, from Assumption 1.1, we have $\lim_{t \rightarrow \infty} \tau(t) \rightarrow \infty$, so the two limiting conditions $t \rightarrow \infty$ and $\tau(t) \rightarrow \infty$ can be used interchangeably. Next, from Theorem 1.1 we have:

$$\lim_{t \rightarrow \infty} \frac{\tilde{X}(\tau(t))}{\tau(t)} \xrightarrow{a.s.} \frac{1}{\mu} \quad (\text{A.3})$$

Since $0 < \mu < \infty$, by applying Theorem 2.2 in Gut (2012) we see that:

$$\lim_{\mu \rightarrow 0} \frac{\tilde{X}^{(1)}(t')\mu}{\tau(T)} = \frac{\tilde{X}^{(\mu)}(\tau(T))\mu}{\tau(T)} \xrightarrow{a.s.} 1, \quad (\text{A.4})$$

Now from Theorem 2.3 in Gut (2012) we have:

$$\lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{\tilde{X}(\tau(t))} (\tilde{D}_i - \mu)}{\sigma \sqrt{\tilde{X}(\tau(t))}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (\text{A.5})$$

Applying (A.4) and substituting $X(t) = \tilde{X}(\tau(t))$ and $RBV(0, t) = \tilde{X}(\tau(t))\mu$ into the above equation yield the desired result.

A.3 Asymptotic Properties of the *RBV* Estimator under Infill Asymptotics

The reason why the sprawl asymptotics is preferred in the renewal literature is that usually we assume the data generating parameters μ and σ^2 to be fixed, and we estimate these parameters by an infinitely long sample. In our case, we can actually change μ arbitrarily by altering the stopping criteria $\mathcal{S}(t_i)$. In this section we use the superscript (μ) to distinguish between the renewal sampling schemes with different μ .

We consider the asymptotic properties of the *RBV* estimator defined in Definition 1.5 under a fixed sampling period $(0, T)$. We assume that the price process follows the assumptions in Assumption 1.1. The quantity of interest is therefore $IV(0, T) = \tau(T) = \int_0^T \sigma_P^2(s) ds$, which is a random variable. The durations in business time $\{\tilde{D}_i^{(\mu)}\}_{i=1: X(T)}$, are still i.i.d., so the point process $\tilde{X}^{(\mu)}(\tau(t))$ is still a renewal process. We can think of the quantity $\tilde{X}^{(\mu)}(\tau(T))$ as the counts of renewal epochs when the renewal process is stopped randomly at time $\tau(T)$.

To derive the counterpart result of Theorem 1.5 under infill asymptotics, we require the following additional assumptions on $P(t)$ and $X^{(\mu)}(t)$:

Assumption A.1. For a fixed time period $(0, T)$:

1. (Divergence of the sampling frequency) We assume that $\lim_{\mu \rightarrow 0} X^{(\mu)}(T) \rightarrow \infty$.
2. (Convergence of the age density) We assume that $\lim_{\mu \rightarrow 0} \frac{\mu_{n+1}}{(n+1)\mu} \rightarrow 0$ for all $n = 1, 2, \dots$, where μ_n is the n -th moment of $\tilde{D}_i^{(\mu)}$.

Assumption A.1.1 ensures that by sampling with an infinitesimally small μ in business time, the renewal sampling frequency goes to infinity. This implies that the price process in business time must contain either a Brownian motion component or an infinitely active jump component, or both. A direct consequence of Assumption A.1.2 is that:

$$\lim_{\mu \rightarrow 0} \sum_{i=1}^{\tilde{X}^{(\mu)}(\tau(T))} D_i^{(\mu)} \xrightarrow{a.s.} \tau(T). \quad (\text{A.6})$$

This is due to the fact that the age process defined in Definition 1.2 converges uniformly to a point mass at zero as μ approaches zero, so the arrival time of the last epoch $\tau(t_{X(T)})$ converges almost surely to a random variable $\tau(T)$. Another implication of Assumption A.1 is that:

$$\lim_{\mu \rightarrow 0} \tilde{X}^{(\mu)}(\tau(T))\mu \xrightarrow{a.s.} \tau(T). \quad (\text{A.7})$$

To prove this result, note that for any μ , we can standardize the renewal process $\tilde{X}^{(\mu)}(\tau(T))$ by scaling the time by a factor of $1/\mu$. The resulting process is a renewal process with unit mean duration and all other moments proportional to the original process, which we denote by $\tilde{X}^{(1)}(t')$ where $t' = \tau(T)/\mu$. Note that $\lim_{\mu \rightarrow 0} t' \rightarrow \infty$, and from the Elementary Renewal Theorem in Theorem 1.1, we have:

$$\lim_{t' \rightarrow \infty} \frac{\tilde{X}^{(1)}(t')}{t'} \xrightarrow{a.s.} 1, \quad (\text{A.8})$$

Revert the scaling, we see that:

$$\lim_{\mu \rightarrow 0} \frac{\tilde{X}^{(1)}(t')\mu}{\tau(T)} = \frac{\tilde{X}^{(\mu)}(\tau(T))\mu}{\tau(T)} \xrightarrow{a.s.} 1 \quad (\text{A.9})$$

The asymptotic result of the *RBV* estimator in the infill asymptotics case is derived by a direct application of Corollary 6.4 in Häusler and Luschgy (2015). Since μ can be chosen arbitrarily, we choose $\mu(n) = n^{-1}$ with $n = 1, 2, \dots$, so that $\mu(n) \rightarrow 0$ is equivalent to $n \rightarrow \infty$. We then construct the following random variable:

$$Z_{ni} = \frac{D_i^{(\mu(n))} - \mu(n)}{\mu(n)^{-0.5} \sigma(n)}. \quad (\text{A.10})$$

Note that Z_{ni} is a square integrable martingale difference array w.r.t. its natural filtration \mathcal{F}_{nk} , and $\mathbb{E}[Z_{ni}^2 | \mathcal{F}_{n,i-1}] = \mu(n)$. Additional technical assumptions are required for Corollary 6.4 in Häusler and Luschgy (2015) to hold:

Assumption A.2. *Technical assumptions for the stable convergence of the *RBV* estimator:*

1. (Finiteness) $X^{(\mu^{(n)})}(\tau(T))$ is a finite stopping time w.r.t. \mathcal{F}_{nk} for every $n \in \mathbb{N}$.
2. (Measurability) $\tau(T)$ is a measurable random variable in $\mathcal{G} = \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{G}_{n\infty})$ where $\mathcal{G}_{n\infty} = \sigma(\bigcup_{k=0}^{\infty} \mathcal{G}_{nk})$ and $\mathcal{G}_{nk} = \bigcap_{m>n} \mathcal{F}_{mk}$.
3. (Conditional Lindeberg's condition):

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{\tilde{X}^{(\mu^{(n)})}(\tau(T))} \mathbb{E}[Z_{ni}^2 \mathbb{1}_{\{|Z_{ni}| \geq \varepsilon\}} | \mathcal{F}_{n,i-1}] \xrightarrow{P} 0 \quad (\text{A.11})$$

for every $\varepsilon > 0$.

We can derive from $\mathbb{E}[Z_{ni}^2 | \mathcal{F}_{n,i-1}] = \mu(n)$ and (A.7) that:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{\tilde{X}^{(\mu^{(n)})}(\tau(T))} \mathbb{E}[Z_{ni}^2 | \mathcal{F}_{n,i-1}] \xrightarrow{P} \tau(T), \quad (\text{A.12})$$

Therefore we can apply Corollary 6.4 in Häusler and Luschgy (2015) to Z_{ni} , which yields:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{\tilde{X}^{(\mu^{(n)})}(\tau(T))} Z_{ni} \xrightarrow{\text{s.t.}} \sqrt{\tau(T)} \mathcal{N}(0, 1), \quad (\text{A.13})$$

where s.t. refers to stable convergence in law. (A.13) leads to the following asymptotic distribution of *RBV*:

$$\lim_{\mu \rightarrow 0} \frac{RBV(0, T) - IV(0, T)}{\sqrt{IV(0, T)} \mu^{-1} \sigma^2} \xrightarrow{d} \mathcal{N}(0, 1), \quad (\text{A.14})$$

Thus, similar asymptotic results to Theorem 1.5 also holds under the setting of infill asymptotics, in the expense of additional assumptions in Assumptions 4 and 5.

A.3.1 Relationship to the RV Estimator

The infill asymptotics results for the *RBV* estimator can be linked naturally to the RV estimator, as we can interpret the renewal sampling scheme as a stochastic sampling scheme for the RV estimator. We start with the assumption that $P(t)$ is a continuous local martingale to which Theorem 1.4 can be applied. For a given μ , let us denote the renewal sampling scheme as $X^{(\mu)}(t)$, the sampling times as $\{t_i^{(\mu)}\}_{i=1,2,\dots}$ and the inter-event return as $r_i^{(\mu)} = P(t_i^{(\mu)}) - P(t_{i-1}^{(\mu)})$. We define the renewal RV and the *RBV*

estimator as

$$\begin{aligned} RV^{(\mu)}(0, T) &= \sum_{i=1}^{X^{(\mu)}(T)} (r_i^{(\mu)})^2, \\ RBV^{(\mu)}(0, T) &= X^{(\mu)}(T)\mu. \end{aligned} \tag{A.15}$$

From the theory of quadratic variation and (A.14) we know that both estimators are consistent, and converge to $IV(0, T)$. Specially, for the $RV^{(\mu)}$ estimator, due to the i.i.d.-ness of the inter-event arrival time in business time denoted by $\tilde{D}_i^{(\mu)} = \tau(t_i^{(\mu)}) - \tau(t_{i-1}^{(\mu)})$, $r_i^{(\mu)}$ is also i.i.d. From the martingale property of the Wiener process we have:

$$E[r^{(\mu)}] = 0, E[(r_i^{(\mu)})^2 | \tilde{D}_i] = \tilde{D}_i, E[(r_i^{(\mu)})^2] = \mu. \tag{A.16}$$

This suggests that a natural and consistent estimator of μ is just the sample moment of the squared return, $\hat{\mu} = \frac{1}{X^{(\mu)}(T)} \sum_{i=1}^{X^{(\mu)}(T)} (r_i^{(\mu)})^2$. Obviously, by using $\hat{\mu}$ instead of μ , the $RBV^{(\mu)}$ estimator coincides with the $RV^{(\mu)}$ estimator. The cost of using $\hat{\mu}$ in the $RBV^{(\mu)}$ estimator is then a larger asymptotic variance. Using Corollary 3.11 in (Fukasawa, 2010b) and Assumption A.1, we see that as $\mu \rightarrow 0$:

$$V[RV^{(\mu)}(0, T)] \rightarrow \frac{2}{3} \sum_{i=1}^{X^{(\mu)}(T)} (r_i^{(\mu)})^4. \tag{A.17}$$

When the unconditional kurtosis $\kappa^{(\mu)}$ of $r_i^{(\mu)}$ exists, the above asymptotic variance converges to $\frac{2}{3}X^{(\mu)}(T)\kappa^{(\mu)}\mu^2$, which is due to the i.i.d.-ness of $r_i^{(\mu)}$.

The asymptotic variance $\frac{2}{3}X^{(\mu)}(T)\kappa^{(\mu)}\mu^2$ has some very interesting implications. Firstly, if $\kappa = 3$ and $r_i^{(\mu)}$ is normally distributed, we have:

$$V[RV^{(\mu)}](0, T) \rightarrow 2IV(0, T)^2/X^{(\mu)}(T),$$

which is identical to the asymptotic variance of the RV estimator sampled in business time (e.g. Hansen and Lunde (2006), Oomen (2006)). The business time RV can indeed be considered as a *RBV* estimator with a constant duration in business time. Moreover, if we can sample $r_i^{(\mu)}$ by setting $\kappa^{(\mu)} = 1$, then the asymptotic variance of the $RV^{(\mu)}$ estimator can be minimized, and is equal to $2IV(0, T)^2/3X^{(\mu)}(T)$. This implies that the optimal renewal RV estimator must have $r_i^{(\mu)}$ following a two-point distribution. We show in Section 1.5 that, the non-parametric duration-based volatility estimator in Nolte, Taylor, and Zhao (2018) is both a *RBV*-class estimator and an optimal renewal RV estimator.

A.3.2 End-of-Sample Bias

In practice, we do not have data of infinite length, and the sample has to stop somewhere. Therefore, there will be a small End-of-Sample (*EoS*) bias for the renewal process when the last renewal epoch is before the end of the sample. The correction of this bias can be obtained from the second order asymptotic expansion of the renewal function as in Proposition 1.1. Let T and $\tau(T)$ denote the endpoint of the sampling interval in calendar and business time respectively, the *EoS* bias correction is:

$$EoS = \tau(T) - \mathbb{E}[\tilde{X}^{(\mu)}(\tau(T))]\mu = 0.5\mu - \frac{\sigma^2}{2\mu}. \quad (\text{A.18})$$

Therefore the bias correction is smaller than 0.5μ , and can even be negative when $\sigma^2 > \mu^2$. In theory one should always add this bias correction to the *RBV* and *PRBV* estimator. Nevertheless, in the infill asymptotics case, when $\frac{\sigma^2}{\mu} \rightarrow 0$ as $\mu \rightarrow 0$, we have $EoS \rightarrow 0$.

A.4 Proof of Corollary 1.3

To prove the corollary, we use the Doob-Meyer decomposition of a point process. In detail, any \mathcal{F}_t -adapted point process $X(t)$ is a submartingale, and for a submartingale, the following decomposition is unique:

$$X(t) = \Lambda(t) + M(t), \quad (\text{A.19})$$

where $\Lambda(t)$ is a \mathcal{F}_t -predictable increasing process called the compensator of $X(t)$, and $M(t)$ is a \mathcal{F}_t martingale. The compensator process and the intensity process is linked via the following relationship:

$$\Lambda(t) = \int_0^t \lambda(t|\mathcal{F}_t) dt \quad (\text{A.20})$$

Therefore, to prove the corollary, we firstly show that under business time, $\tilde{\Lambda}(\tau(t)) = \Lambda(t)$ is the compensator of the process $\tilde{X}(\tau(t))$. Note that under business time, we have the following decomposition for $\tilde{X}(\tau(t))$:

$$\tilde{X}(\tau(t)) = \check{\Lambda}(\tau(t)) + \check{M}(\tau(t)), \quad (\text{A.21})$$

in which $\check{\Lambda}(\tau(t))$ is the compensator of $\tilde{X}(\tau(t))$ and $\check{M}(\tau(t))$ is a martingale in business time. Moreover, if we change $X(t)$ from calendar time to business time, we have that:

$$\tilde{X}(\tau(t)) = \tilde{\Lambda}(\tau(t)) + \tilde{M}(\tau(t)). \quad (\text{A.22})$$

Importantly, $\tilde{M}(\tau(t))$ is also a martingale due to the fact that the time change preserves the martingale property according to the optional stopping theorem. Then from the uniqueness of the Doob-Meyer decomposition we see that for all t , $\tilde{M}(\tau(t)) = \check{M}(\tau(t))$ and therefore $\check{\Lambda}(\tau(t)) = \tilde{\Lambda}(\tau(t))$.

By the definition of conditional intensity we see that:

$$\tilde{\Lambda}(\tau(t)) = \int_{\tau(0)}^{\tau(t)} \tilde{\lambda}(\tau(s)|\mathcal{F}_s) d\tau(s) = \int_0^t \lambda(s|\mathcal{F}_s) ds, = \Lambda(t) \quad (\text{A.23})$$

and it is therefore clear that $\tilde{\lambda}(\tau(t)|\mathcal{F}_t)$ is the conditional intensity process of $\tilde{X}(\tau(t))$. Now, since the above equation holds for an arbitrary t , it must also hold that:

$$\tilde{\lambda}(\tau(t)|\mathcal{F}_t) d\tau(t) = \lambda(t|\mathcal{F}_s) dt. \quad (\text{A.24})$$

Substituting $d\tau(t) = \sigma_p^2(t) dt$ into the above equation yields the desired result.

A.5 Proof of Corollary 1.1

To prove the first part of the corollary, we only need to show that R_i is a monotonically increasing function of \tilde{D}_i in the sense that for any $\tilde{D}_i > \tilde{D}_j$, $R_i > R_j$.

From the proof of Proposition 1.3, we can write R_i in terms of $\tilde{\lambda}(\tau(t)|\mathcal{F}_t)$:

$$R_i = \mu \int_{t_{i-1}}^{t_{i-1} + D_i} \lambda(s|\mathcal{F}_s) ds = \mu \int_{\tau(t_{i-1})}^{\tau(t_{i-1}) + \tilde{D}_i} \tilde{\lambda}(\tau(s)|\mathcal{F}_s) d\tau(s). \quad (\text{A.25})$$

Note that by the definition of conditional intensity and due to that the process $\tilde{X}(\tau(t))$ is renewal, we have $\tilde{\lambda}(\tau(t_i) + s|\mathcal{F}_{t_i}) = h_{\tilde{D}}(s)$, where $h_{\tilde{D}}(s)$ is the hazard function of the renewal process $\tilde{X}(\tau(t))$ defined by:

$$h_{\tilde{D}}(s) = -\frac{d \ln(1 - F_{\tilde{D}}(s))}{ds}, \quad (\text{A.26})$$

in which $F_{\tilde{D}}(s)$ is the CDF of \tilde{D}_i . The cumulative hazard function $H_{\tilde{D}}(x)$ is defined as:

$$H_{\tilde{D}}(s) = \int_0^s h_{\tilde{D}}(u) du = -\ln(1 - F_{\tilde{D}}(s)). \quad (\text{A.27})$$

The equivalence between $h_{\tilde{D}}(s)$ and $\tilde{\lambda}(\tau(t_i) + s | \mathcal{F}_{t_i})$ suggests the following relationship which holds true for all t :

$$\int_{\tau(t_{i-1})}^{\tau(t_{i-1}+t)} \tilde{\lambda}(\tau(s) | \mathcal{F}_s) d\tau(s) = \int_{t_{i-1}}^{t_{i-1}+t} \lambda(s | \mathcal{F}_s) ds = -\ln(1 - F_{\tilde{D}}(\tau(t_{i-1} + t) - \tau(t_{i-1}))). \quad (\text{A.28})$$

Taking $t = D_i$ and substitute into the equation above:

$$R_i = -\mu \ln(1 - F_{\tilde{D}}(\tilde{D}_i)). \quad (\text{A.29})$$

Note that the term $-\ln(1 - F_{\tilde{D}}(\tilde{D}_i))$ in the above equation is the exponential inverse probability integral transform of \tilde{D}_i which follows an i.i.d. unit exponential distribution. This is consistent with the result that R_i is i.i.d. exponential. More importantly, $-\ln(1 - F_{\tilde{D}}(\tilde{D}_i))$ is a monotonically increasing function of \tilde{D}_i , which completes the proof of the first part.

To prove the second part of the corollary, we note that when \tilde{D}_i is i.i.d. exponential with mean μ and variance μ^2 , $\tilde{\lambda}(\tau(t) | \mathcal{F}_t) = \mu^{-1}$. Apply Proposition 1.3 and observe that $\sigma_p^2(t) = \mu \lambda(t | \mathcal{F}_t) = g(t | \mathcal{F}_t)$ as desired.

A.6 Simulation of $\rho^{(\delta)}$ and $\rho^{(r)}$ for the PD and PR estimators

To simulate $\rho^{(\delta)}$ and $\rho^{(r)}$, we firstly simulate a standard Wiener process. Let $\Delta W_i \sim \mathcal{N}(0, \Delta)$, and the (discrete) Wiener process is simulated as:

$$W_j = \sum_{k=1}^{\infty} \Delta W_k. \quad (\text{A.30})$$

In the simulation we set $\Delta = 10^{-5}$. The stopping times $\{\tilde{D}_i^{(\delta)}\}_{i=1:N}$ and $\{\tilde{D}_i^{(r)}\}_{i=1:N}$ are then constructed by setting $\delta = r = 1$ based on this Wiener process as follows:

$$\begin{aligned} \tilde{D}_i^{(\delta)} &= \frac{1}{\Delta} \inf_{j>i-1} \{W_j : |W_j - W_{i-1}| \geq 1\}, \\ \tilde{D}_i^{(r)} &= \frac{1}{\Delta} \inf_{j>i-1} \{W_j : \sup_{i-1<s<j} (W_s) - \inf_{i-1<s<j} (W_s) \geq 1\}. \end{aligned} \quad (\text{A.31})$$

We choose $N = 1000000$. Note that there will be a small truncation bias due to the discreteness of the simulated Wiener process. This will cause the simulated $\tilde{D}_i^{(\delta)}$ and $\tilde{D}_i^{(r)}$ to be biased upward slightly, and the bias vanishes as $\Delta \downarrow 0$. This will not have

a significant impact as long as Δ is relatively small compared to δ or r . Based on the simulated $\tilde{D}_i^{(\delta)}$ and $\tilde{D}_i^{(r)}$, we can construct $R_i^{(\delta)}$ and $R_i^{(r)}$ as:

$$R_i^{(\cdot)} = -\mathbb{E}[\tilde{D}^{(\cdot)}] \ln(1 - \hat{F}_{\tilde{D}^{(\cdot)}}(\tilde{D}_i^{(\cdot)})), \quad (\text{A.32})$$

in which $\hat{F}_{\tilde{D}^{(\cdot)}}(x)$ is the empirical CDF of $\tilde{D}^{(\cdot)}$. We do not use the theoretical CDF because it is not available in closed form. The correlation $\rho^{(\cdot)}$ is then computed based on $\{\tilde{D}_i^{(\cdot)}\}_{i=1:N}$ and $\{R_i^{(\cdot)}\}_{i=1:N}$. We plot the simulated moments for $\{\tilde{D}_i^{(\cdot)}\}_{i=1:N}$ and $\{R_i^{(\cdot)}\}_{i=1:N}$ and the simulated $\rho^{(\cdot)}$ in Table A.1. $\mathbb{V}[\tilde{D}_i^{(\cdot)} - R_i^{(\cdot)}]$ of an arbitrary δ

Table A.1 Simulated moments for $\tilde{D}_i^{(\cdot)}$ and $R_i^{(\cdot)}$ and the simulated $\rho^{(\cdot)}$

	$\mathbb{E}[\tilde{D}^{(\delta)}]$	$\mathbb{V}[\tilde{D}^{(\delta)}]$	$\mathbb{E}[R_i^{(\delta)}]$	$\mathbb{V}[R_i^{(\delta)}]$	$\rho^{(\delta)}$	$\mathbb{V}[\tilde{D}_i^{(\delta)} - R_i^{(\delta)}]$
Simulated	1.0033	0.6707	1.0000	0.9999	0.9998	0.0330
Theoretical	1	0.6667	1	1	-	0.0340
	$\mathbb{E}[\tilde{D}^{(r)}]$	$\mathbb{V}[\tilde{D}^{(r)}]$	$\mathbb{E}[R_i^{(r)}]$	$\mathbb{V}[R_i^{(r)}]$	$\rho^{(r)}$	$\mathbb{V}[\tilde{D}_i^{(r)} - R_i^{(r)}]$
Simulated	0.5036	0.0844	0.5000	0.2500	0.9915	0.0463
Theoretical	0.5	0.0833	0.5	0.25	-	0.0471

Note: $\delta = r = 1$. $N = 1000000$. Theoretical values of the simulated moments can be found in (1.26) and (1.30). Note that for theoretical moments of $\mathbb{V}[\tilde{D}_i^{(\cdot)} - R_i^{(\cdot)}]$ we plug in the simulated $\rho^{(\cdot)}$ in the relationship: $\mathbb{V}[\tilde{D}_i^{(\cdot)}] + \mathbb{V}[R_i^{(\cdot)}] - 2\rho^{(\cdot)}\sqrt{\mathbb{V}[\tilde{D}_i^{(\cdot)}]\mathbb{V}[R_i^{(\cdot)}]}$.

or r can be easily obtained by scaling the corresponding variables.

A.7 An Approximated Time Discretization Bias

Throughout this section we assume $r_j = r_j^*$, that is, the MMS noise is absent in the price process. We start by decomposing $r_i^{(\delta)}$ as:

$$r_i^{(\delta)} = \sum_{j=j_{i-1}^{(\delta)}}^{j_{i-1}^{(\delta)} + M_i^{(\delta)}} r_j, \quad (\text{A.33})$$

where $j_i^{(\delta)}$ is the observation index of $t_i^{(\delta)}$, and $M_i^{(\delta)}$ is the number of observations in the i -th price duration (excluding the starting point). We see that since we assume r_j to be strongly mixing and strictly stationary with finite moments, from the central limit theorem (e.g. Peligrad (1986), Billingsley (2009)) it holds that:

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N r_j \sim \mathcal{N}(0, NV[r_j]), \quad (\text{A.34})$$

where $V[r_j] = \mathbf{E}[\tilde{d}_j]$ is the unconditional variance of the tick return. Now consider N_i being a sufficiently large random variable, so that $\sum_{n=1}^{N_i} r_j$ is approximately mixture normal. The absolute price change point process truncates this random variable $\sum_{n=1}^{N_i} r_j$ whenever $|\sum_{n=1}^{N_i} r_j| \geq \delta$, and the distribution of $r_i^{(\delta)}$ becomes very complicated.

To provide an approximated result, we treat the sequence r_j as i.i.d. normal variates with variance $V[r_j]$. Let $S_i = \sum_{j=1}^{N_i} r_j$ denote the partial sum of the returns till step N_i , the process S_i is then a Gaussian random walk. For a truncation threshold δ , we use the joint distribution $\{N_i^{(\delta)}, S_{N_i^{(\delta)}}\}$ to approximate $\{M_i^{(\delta)}, r_i^{(\delta)}\}$. The asymptotic expansion of $\mathbf{E}[S_{N_i^{(\delta)}}^2]$ as $\delta \rightarrow \infty$ is given in Lotov (1996):

$$\mathbf{E}[S_{N_i^{(\delta)}}^2] = \delta^2 + 2\delta\sqrt{V[r_j]}\mathcal{K} + V[r_j]\mathcal{K}^2 + \frac{1}{4} + o(1), \quad (\text{A.35})$$

where $\mathcal{K} \approx 0.58258087$ is defined through: $\mathcal{K} = \frac{1}{\sqrt{2\pi}} \lim_{n \rightarrow \infty} [2\sqrt{n} - \sum_{m=1}^n m^{-1}]$. From Wald's identity we also have that: $\mathbf{E}[S_{N_i^{(\delta)}}^2] = V[r_j]\mathbf{E}[N_i^{(\delta)}]$.

In this Gaussian random walk setting, $S_{N_i^{(\delta)}}$ can be interpreted as the return for the i -th price duration. As a result, the expected TD bias is just: $TD^{(\delta)}(0, t) = \sum_{i=1}^{X(t)} S_{N_i^{(\delta)}}^2 - X(t)\delta^2$. Apply Wald's identity once again, we have $\mathbf{E}[TD^{(\delta)}(0, t)] \rightarrow \mathbf{E}[X(t)](\mathbf{E}[S_{N_i^{(\delta)}}^2] - \delta^2)$ in the limit. Also, $\mathbf{E}[X(t)]$ in the limit converges to $\frac{IV(0, t)}{\mathbf{E}[S_{N_i^{(\delta)}}^2]}$ which is from the property of renewal processes. The approximated \widetilde{TD} bias is therefore:

$$\begin{aligned} \lim_{\delta \rightarrow \infty} \widetilde{TD}^{(\delta)}(0, t) &\rightarrow \frac{IV(0, t)}{\mathbf{E}[S_{N_i^{(\delta)}}^2]} \mathbf{E}[S_{N_i^{(\delta)}}^2 - \delta^2] \\ &= IV(0, t) \left(1 - \frac{1}{1 + O(\delta^{-1})}\right) \end{aligned} \quad (\text{A.36})$$

which converges to zero as $\delta \rightarrow \infty$ with a rate of δ^{-1} .

A.8 Determinants of the Bias of the *NPD* Model

We provide a simple example to illustrate the bias of the *NPD* estimator as a function of δ via simulation. Based on (1.43), we assume that the arrival of $\tau(t_j)$ in business time follows a homogeneous Gamma process with intensity measure $\nu(x) = \frac{\gamma}{xe^{\lambda x}}$. The inter-observation durations in business time are then i.i.d. Gamma distributed:

$\tilde{d}_j \sim \Gamma(\gamma, \lambda)$. Let z_j^* denote an i.i.d. standard normal variable, we have:

$$r_j^* = z_j^* \sqrt{\tilde{d}_j} \sim \mathcal{N}(0, \tilde{d}_j). \quad (\text{A.37})$$

so the tick return is unconditionally mixture normal. This simple structure allows for a leptokurtic distribution of r_j^* with the following sample moments: $\mathbf{V}[r_j^*] = \gamma\lambda$, $\mathbf{K}[r_j^*] = 3 + \frac{3}{\gamma}$. For the noise term V_j , we assume that it follows an AR(1) process:

$$V_j = \rho V_{j-1} + v_j, \quad v_j \sim \mathcal{N}\left(0, \frac{\sigma_v^2}{1 - \rho^2}\right). \quad (\text{A.38})$$

To ensure that V_j complies with Assumption 1.2, we further require that $|\rho| < 1$ and $v_j \perp r_j^*$. The unconditional variance V_j is therefore σ_v^2 for any $\rho \in (-1, 1)$. The tick return r_j is therefore conditionally normally distributed with an ARMA-type autoregressive structure. We will refer to this model as the Gamma subordinated transaction (GST) model. Some moment conditions for r_j are summarized in Appendix A.9.

To illustrate the asymptotic properties of the *NPD* estimator in this setting, we construct the $X^{(\delta)}(t)$ process for various parameter settings and a range of δ based on the simulated P_i . We then compare the simulated $\mu(\delta)$ with δ^2 , which describes the bias of the *NPD* estimator. To show this difference graphically, we plot the volatility signature plot (Andersen, Bollerslev, Diebold, and Labys, 2000) of the *NPD* estimator for a theoretical interval using the asymptotic property of $X^{(\delta)}(t)$. The volatility signature plot is constructed by plotting $\mathbf{E}[NPD(0, t)] = \frac{IV(0, t)}{\mu(\delta)}$ against δ for some finite $IV(0, t)$, and comparing it to the true integrated variance. The mean duration in business time can be simulated by collecting the number of transactions $M_i^{(\delta)}$ required to trigger the i -th price duration, and the mean duration $\mu(\delta)$ can be obtained as:

$$\mu(\delta) = \frac{\gamma\lambda}{N} \sum_{i=1}^k M_i^{(\delta)}, \quad (\text{A.39})$$

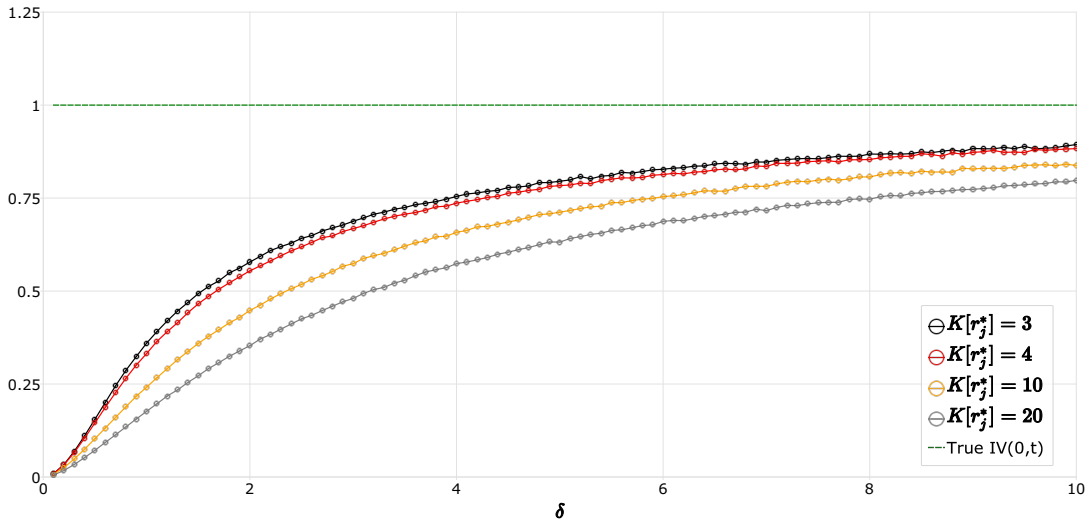
in which k is the size of the simulation. Alternatively, it can be simulated by the renewal *RV* estimator (with a larger simulation error) as:

$$\mu(\delta) = \frac{1}{N} \sum_{i=1}^k (r_i^{(\delta)})^2. \quad (\text{A.40})$$

For each δ , we choose $k = 100000$. All the parameters of the GST model are set for illustrative purposes only.

The first case we examine is the case where $V_i = 0$ and $V[r_j^*] = 1$. We set the kurtosis $K[r_j^*]$ to be 20, 10, 4 and 3 to examine the effect of an excess kurtosis on the *NPD* estimator.¹ The volatility signature plots of the *NPD* estimator under these parameter settings are presented in Figure A.1. In the simulation we set $IV(0,t) = 1$ with δ ranging from 0.1 to 10 with a step size of 0.1. From Figure A.1 we see that, as discussed in the previous section, the *NPD* estimator is downwardly biased in the absence of MMS noise due to $Bias_{TD}^{(\delta)}$, which is a function of $V[r_j^*]$ and kurtosis. Generally, holding the variance constant, r_j^* with heavier tails will have a larger truncation bias on average, as is shown in Figure A.1. It is also clear that the bias decays slowly as δ increases, which corroborates our result in A.7.

Figure A.1 Simulated volatility signature plot for the *NPD* estimator on the GST model with no MMS noise



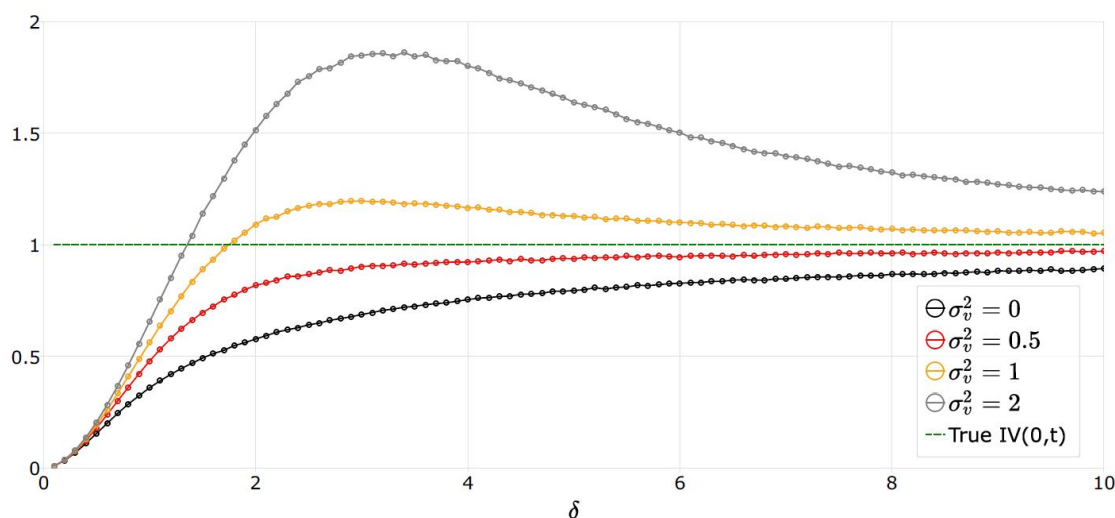
Note: We simulate $\mu(\delta)$ based on (A.39) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ , the volatility signature curve is computed by $IV(0,t) \frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. Each circle represents a volatility estimate from the *NPD* estimator computed at the value of δ .

In the next case, we examine the effect of various sizes of i.i.d. MMS noise by choosing $\rho = 0$ with $\sigma_v^2 \in \{0, 0.5, 1, 2\}$. We will use the same parameter settings from the previous case with $K[r_j^*] = 4$ for illustration, as the effect of kurtosis is similar for both cases. The volatility signature plots are presented in Figure A.2. The figure corroborates our previous discussion on the truncation bias and the MMS noise bias. From the graph, we see that when the size of the noise is small ($\sigma_v^2 \leq 0.5$), the MMS noise bias is smaller than the truncation bias and the volatility signature curve converges from below. When the size of the MMS bias is large enough to

¹For the first three cases, the corresponding parameter values for (γ, λ) are $(\frac{3}{47}, \frac{47}{3})$, $(\frac{3}{17}, \frac{17}{3})$ and $(3, \frac{1}{3})$ respectively. When $K[r_j^*] = 3$, $\tilde{d}_j = 1$ for all j so that r_j^* is i.i.d. normal.

compensate for the truncation bias, the volatility signature curve has a hump shape and converges from above. This result is consistent with Figures 2, and 3 in Nolte, Taylor, and Zhao (2018), which document a similar curve with a different setting. Also, in the case where the MMS bias is large enough, we see that the volatility signature curve intersects the true $IV(0,t)$ at some finite δ so the *NPD* estimator is unbiased. Unfortunately, we are unable to derive an analytical form for this particular *NPD* estimator as we cannot estimate the amount of $Bias_{TD}^{(\delta)}$.

Figure A.2 Simulated volatility signature plot for the *NPD* estimator on the GST model with i.i.d. MMS noise

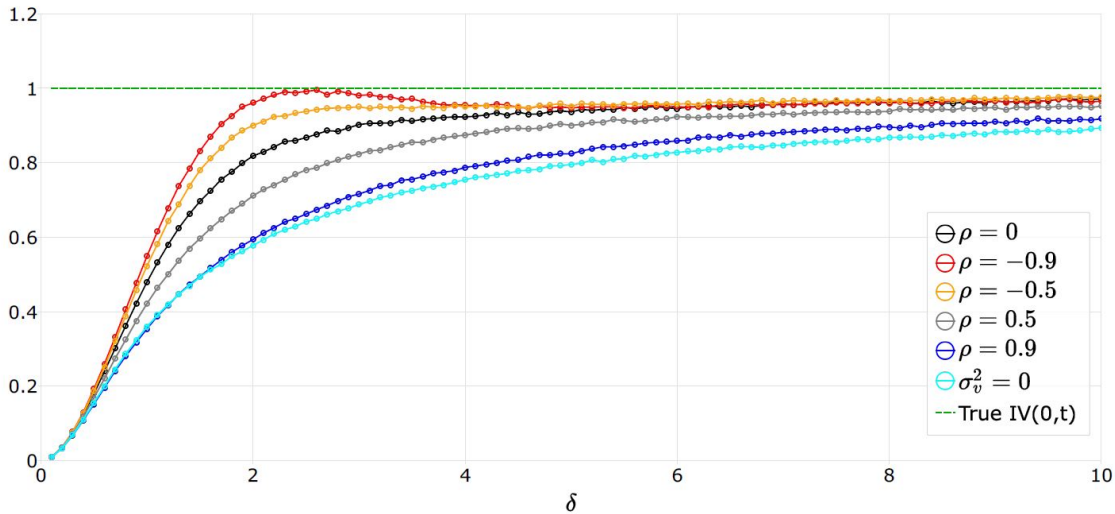


Note: We simulate $\mu(\delta)$ based on (A.39) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ , the volatility signature curve is computed by $IV(0,t) \frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$ and $\lambda = 1/3$ for all four cases. Each circle represents a volatility estimate from the *NPD* estimator computed at the value of δ .

Figure A.3 shows the case with AR(1) MMS noise. In the simulation we use the settings from the previous case with $\sigma_v^2 = 0.5$ and $\rho \in \{-0.9, -0.5, 0, 0.5, 0.9\}$, so that the unconditional variance of the noise remains unchanged. The figure shows that, negative autocorrelation inflates $Bias_{MMS}^{(\delta)}$ when δ is small, and affects the shape of the volatility signature plot. We can clearly see a hump-shaped volatility signature curve for $\rho = -0.9$. The effect of negatively correlation decays as δ increases, and the volatility signature curves converge to the i.i.d. noise case. The impact of positively correlated noise is more persistent and has less of an impact on the *NPD* estimator. However, in the positively correlated noise case the volatility signature curve deviates from the i.i.d. case as ρ increases.

We also include price discretization in the example. As discussed in Section 1.6.5, we assume that r_j is discrete with the support $\{\dots, -2\epsilon, -\epsilon, 0, \epsilon, 2\epsilon, \dots\}$. We construct

Figure A.3 Simulated volatility signature plot for the *NPD* estimator on the GST model with AR(1) MMS noise

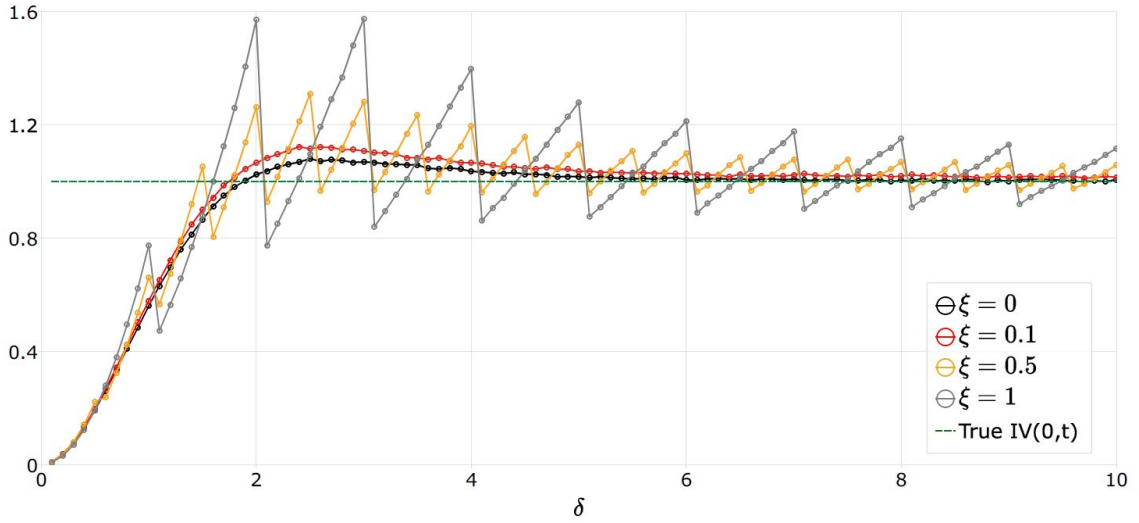


Note: We simulate $\mu(\delta)$ based on (A.39) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ , the volatility signature curve is computed by $IV(0, t) \frac{\delta^2}{\mu(\delta)}$ where $IV(0, t) = 1$. The GST model parameters are $\gamma = 3$, $\lambda = 1/3$ and $\sigma_v^2 = 0.5$ for all five cases. Each circle represents a volatility estimate from the *NPD* estimator computed at the value of δ .

the same volatility signature plots for $\varepsilon \in [0, 0.1, 0.5, 1]$, and construct the discrete log-price process $h_\varepsilon(P_j)$, where $h_\varepsilon(x) = \varepsilon \text{nint}(\frac{x}{\varepsilon})$ and $\text{nint}(x)$ is the nearest integer function. A slight complication arises in this situation. As the *NPD* estimator always samples in tick time, all the zero entries in r_j are completely disregarded. We choose the parameter settings for P_j from the previous AR(1) noise case with $\rho = -0.5$ and examine the effect of different levels of ε on the *NPD* estimator. The volatility signature plots in this case are presented in Figure A.4.

Figure A.4 reveals some very interesting features of the *NPD* estimator under price discretization. Comparing the case with $\varepsilon = 0$ and $\varepsilon = 0.1$, we see that the bias increases slightly as a result of the price discretization. When $\varepsilon = 0.5$ or 1, the volatility signature curves have a zigzag pattern. As discussed in Section 1.6.5, this is due to the invariant sampling scheme for $\delta \in ((n-1)\varepsilon, n\varepsilon]$, so that $\mu(\delta)$ is also constant within the range. As a result, the *NPD* volatility estimates for $\delta \in (n-1)\varepsilon, n\varepsilon]$ will become a quadratic function of δ peaking at every $n\varepsilon$. By sampling at $n\varepsilon$, we obtain the volatility signature curve that has the least truncation bias, and this bias can be artificially increased by letting $\delta \downarrow (n-1)\varepsilon$ without changing the properties of the sampling scheme. Therefore, if the magnitude of the MMS noise is large enough, one may be able to obtain solutions of δ^* for multiple n , represented by the multiple intersections between the volatility signature curves and the true IV for $\varepsilon \geq 0.5$.

Figure A.4 Simulated volatility signature plot for the *NPD* estimator on the GST model with AR(1) MMS noise and price discretization



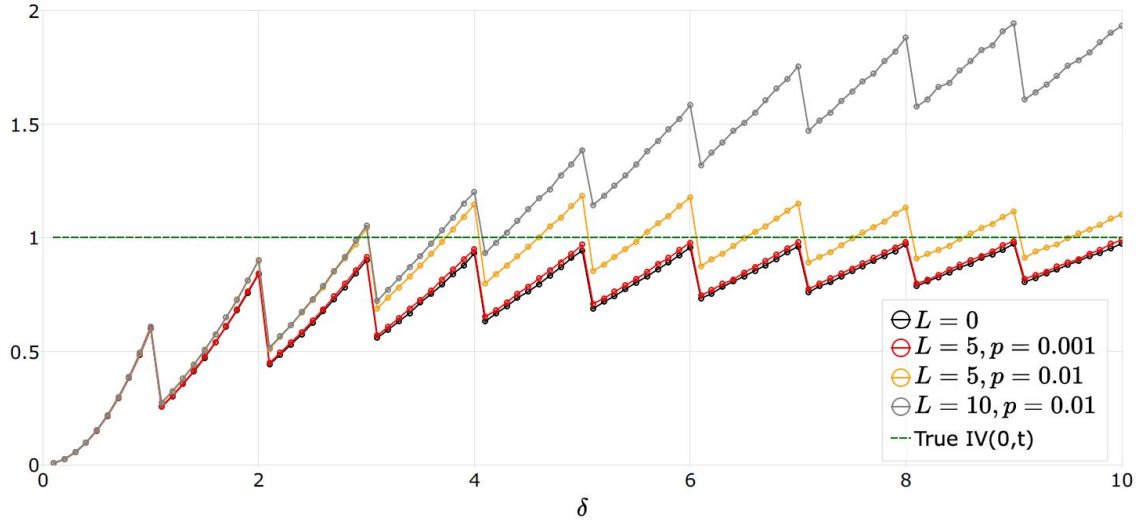
Note: We simulate $\mu(\delta)$ based on (A.39) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ , the volatility signature curve is computed by $IV(0,t) \frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$, $\lambda = 1/3$, $\rho = -0.5$ and $\sigma_v^2 = 0.5$ for all five cases. Each circle represents a volatility estimate from the *NPD* estimator computed at the value of δ .

In the last case, we examine the effect of jumps and price discretization on the *NPD* estimator and assume that $V_i = 0$ for simplicity. The discrete price process with jumps is specified as follows:

$$\begin{aligned} \tilde{P}_j &= h_\varepsilon(P_j^*) + L \cdot L_j \cdot J_j \\ L_j &\sim i.i.d. \text{Bernoulli}(p), J_j \sim i.i.d. \text{Rademacher} \end{aligned} \quad (\text{A.41})$$

In this simple setting, L is the size of each jump which is assumed to be a constant, L_j is a Bernoulli draw on each arrival of transaction representing the arrivals of jumps, and J_j determines the direction of the jump. We plot the simulated volatility signature plot in this case in Figure A.5. From the figure, we see that both L and p influences the bias of the *NPD* estimator. We see that when δ is very small, the four curves coincide, which proves our previous theoretical result on the jump effect. As the jump size and jump intensity increase, the *NPD* estimator absorbs more jump variation and are also affected. In the extreme case with $L = 10$, we see that the *NPD* estimator diverges from the true *IV*. However, the jump intensity used here (one per 100 transaction) is highly unlikely in reality (as opposed to less than one per week as documented in Andersen, Bollerslev, and Dobrev (2007) and Lee and Hannig (2010)).

Figure A.5 Simulated volatility signature plot for the NPD estimator on the GST model with price discretization and jumps



Note: We simulate $\mu(\delta)$ based on (A.39) for $\delta \in [0.1, 10]$ with a step size of 0.1. For each δ , the volatility signature curve is computed by $IV(0,t) \frac{\delta^2}{\mu(\delta)}$ where $IV(0,t) = 1$. The GST model parameters are $\gamma = 3$, $\lambda = 1/3$ and $V_j = 0$ for all four cases. Jumps are specified as (A.41). Each circle represents a volatility estimate from the NPD estimator computed at the value of δ .

A.9 Moment conditions for r_j of the GST model

$$E[r_j] = 0, \quad (\text{A.42})$$

$$V[r_j] = \gamma\lambda + \frac{2\sigma_v^2}{1+\rho}, \quad (\text{A.43})$$

$$E[r_j r_{j-k}] = \frac{\rho-1}{\rho+1} \rho^{k-1} \sigma_v^2, \quad (\text{A.44})$$

$$E[r_j^4] = 3\gamma\lambda^2(1+\gamma) + \frac{12\gamma\lambda\sigma_v^2}{1+\rho} + \frac{12\sigma_v^4}{(1+\rho)^2}. \quad (\text{A.45})$$

A.10 Implementation details for RK, NPD^z , PRV and PBip estimators

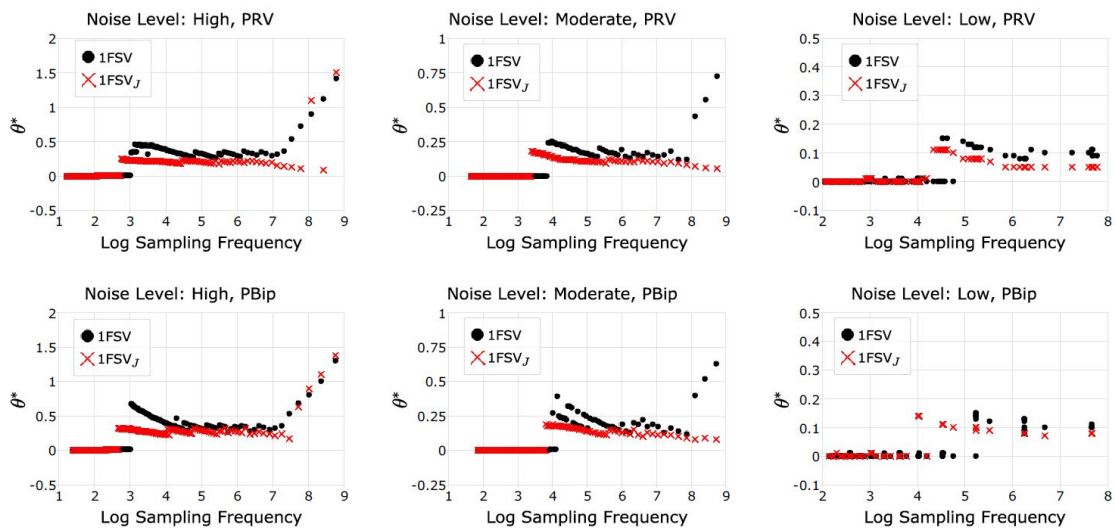
For the RK estimator, we use a Tukey-Hannings₂ kernel, with the optimal bandwidth $H = 5.74\xi N^{0.5}$, in which $\xi = \frac{\sigma_v^2}{\sqrt{\int_0^1 \sigma_p^4(s) ds}}$ and N is the sampling frequency, as given by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a). In the simulation, we use the true value of σ_v^2 and $\int_0^1 \sigma_p^4(s) ds$ for each random draw of the 1FSV model to construct the RK estimator. The RK estimator is then constructed based on the calendar time sampled returns with the sampling frequency given by the average

sampling frequency of the NPD^z estimator (or the NPD_j^z estimator in the presence of jump) for each $\delta = x\delta_0$. An example of the sampling frequencies for the moderate level of noise case is shown in Figure 1.9.

For the NPD^z estimator, we choose the optimal smoothing parameter by computing the MSE of the NPD^z estimator based on 10000 random draws of the 1FSV and 1FSV_J models for a grid of $\delta = x\delta_0$, and choose the γ that minimizes the MSE of the NPD^z estimator for some δ . The resulting optimized γ^* s for the 1FSV model with high, moderate and low levels of noise are 0.1, 0.30 and 0.92 respectively. For the 1FSV_J model, the corresponding γ^* s are: 0.1, 0.31 and 0.99. The γ^* s for the 1FSV model are very close to those of the 1FSV_J model with the same noise level. We also see that the smoothing parameter is reversely related to the size of noise as expected. The NPD^z estimator is then constructed on a grid of $\delta = x\delta_0$ on the smoothed price process Z_j .

For the PRV and PBip estimators, we need to determine the tuning parameter θ that controls for the window width of pre-averaging (see e.g. Jacod, Li, Mykland, Podolskij, and Vetter (2009) or Hautsch and Podolskij (2013)), and σ_v^2 to correct for the pre-averaged MMS bias. As the optimal value of θ varies with the sampling frequency according to Hautsch and Podolskij (2013), we optimize θ for each sampling frequency used in order to obtain optimized performance for the PRV and PBip estimators at each sampling frequency. In detail we use a grid of $\theta \in [0, 2]$ to construct both estimators and to choose an optimal θ^* that minimizes the MSE of the estimator at each sampling frequency. Note that when $\theta^* = 0$, we use RV and RBip instead. We plot the optimal θ^* s of PRV and PBip for both the 1FSV and 1FSV_J models under three different levels of noise in Figure A.6:

Figure A.6 shows that the optimal θ s indeed vary with the sampling frequency. Generally, a much larger θ is required for the highest sampling frequency, and for the sampling frequency within $exp(4)$ to $exp(7)$, θ is very stable. When the sampling frequency decreases further, θ quickly drop to zero, as the simple RV and RBip estimators have better efficiency when the impact of MMS noise is small. The presence of jump seems to decrease the optimal θ^* slightly, but the optimal θ^* s have a similar pattern with or without jumps.

Figure A.6 Optimal θ s of PRV and PBip estimators

Note: The results are based on 10000 replications of the 1FSV and 1FSV_J models. For each black dot, the x-axis shows the log sampling frequency used to construct PRV and PBip estimators and the y-axis represents the optimized value for θ . For each sampling frequency, θ^* is computed by a grid search method for $\theta \in [0, 2]$ that minimizes the simulation MSE.

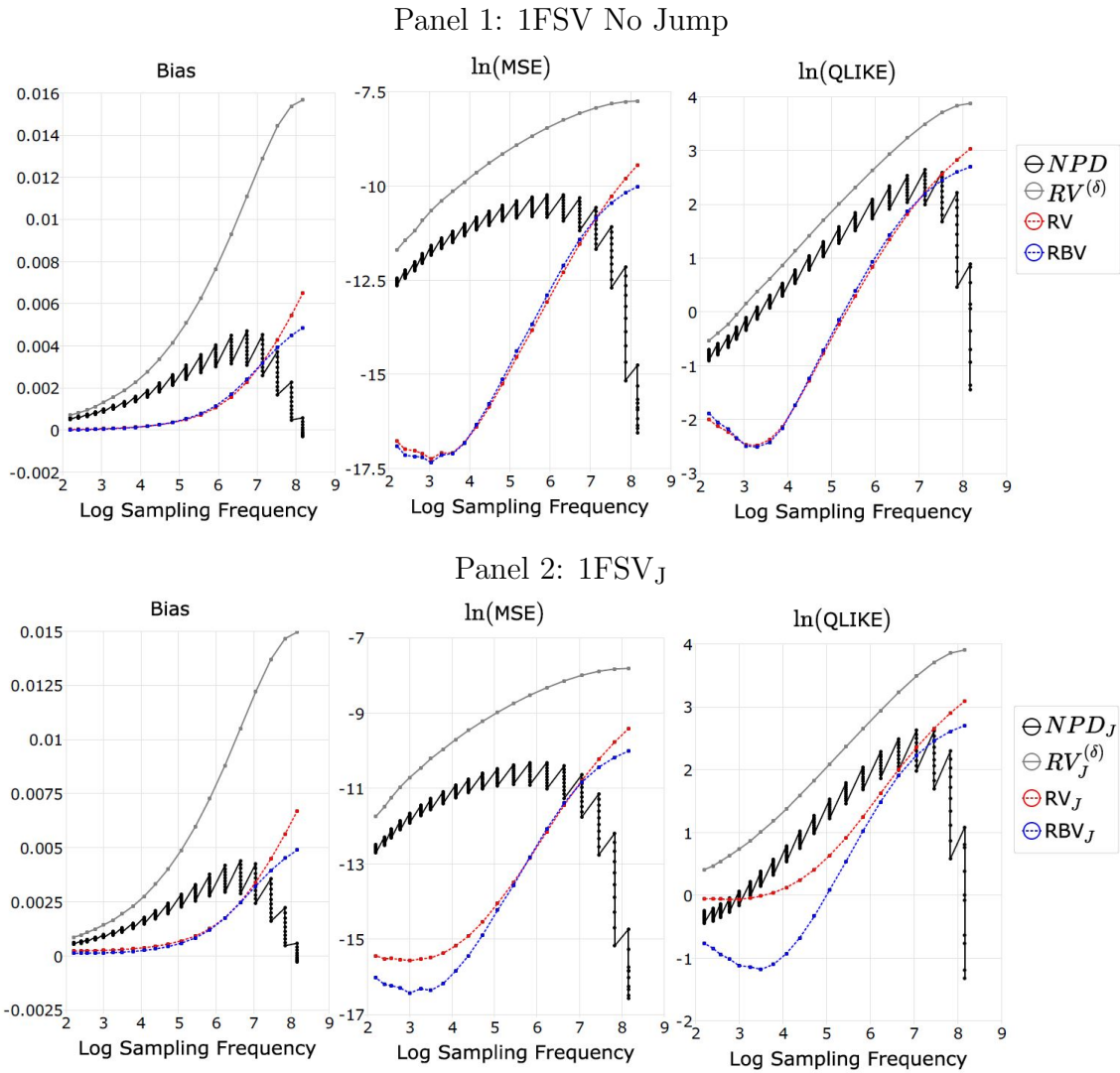
A.11 Additional Tables and Figures

Table A.2 Comparison of the optimal QLIKEs for all volatility estimators in Table 1.1 for the 1FSV and 1FSV_J models with low, moderate and high levels of noise

Estimator	<i>NPD</i>	<i>RV</i> ^(δ)	<i>RV</i>	<i>RBip</i>	<i>NPD</i> ^c	<i>RK</i>	<i>PRV</i>	<i>PBip</i>
1FSV model with low level of noise								
Optimal log QLIKE	-5.3841	-3.1229	-4.1165	-4.3328	-5.4041	-5.1137	-4.8457	-5.5014
δ/δ_0	24	61	21	21	25	9	11	8
Sampling Freq.	189	26	189	189	189	2160	524	2160
1FSV model with moderate level of noise								
Optimal log QLIKE	-2.7444	-2.1163	-3.3983	-3.4633	-5.3551	-4.6682	-4.8649	-4.9156
δ/δ_0	91	131	51	51	27	9	6	6
Sampling Freq.	19	8	84	84	594	2955	2955	2955
1FSV model with high level of noise								
Optimal log QLIKE	-1.4389	-0.5337	-2.4789	-2.5070	-5.1677	-3.5907	-4.4197	-4.3951
δ/δ_0	5	191	141	141	30	14	6	6
Sampling Freq.	3529	9	27	27	1845	2618	3529	3529
1FSV _J model with low level of noise								
Optimal log QLIKE	-5.4757	-0.1705	-0.2015	-2.8499	-5.4833	-0.2122	-0.3157	-4.0977
δ/δ_0	24	111	51	11	24	14	17	1
Sampling Freq.	187	9	38	517	187	517	517	2142
1FSV _J model with moderate level of noise								
Optimal log QLIKE	-2.3990	-0.0655	-0.2012	-2.0102	-4.6502	-0.1925	-0.3497	-3.4153
δ/δ_0	81	141	81	51	18	18	15	7
Sampling Freq.	26	8	26	83	1322	1322	1322	2929
1FSV _J model with high level of noise								
Optimal log QLIKE	-1.3215	0.4063	-0.0663	-1.1812	-4.0293	-0.0827	-0.1327	-2.7092
δ/δ_0	4	191	161	131	20	19	7	7
Sampling Freq.	3491	9	16	33	2524	2524	3491	3491

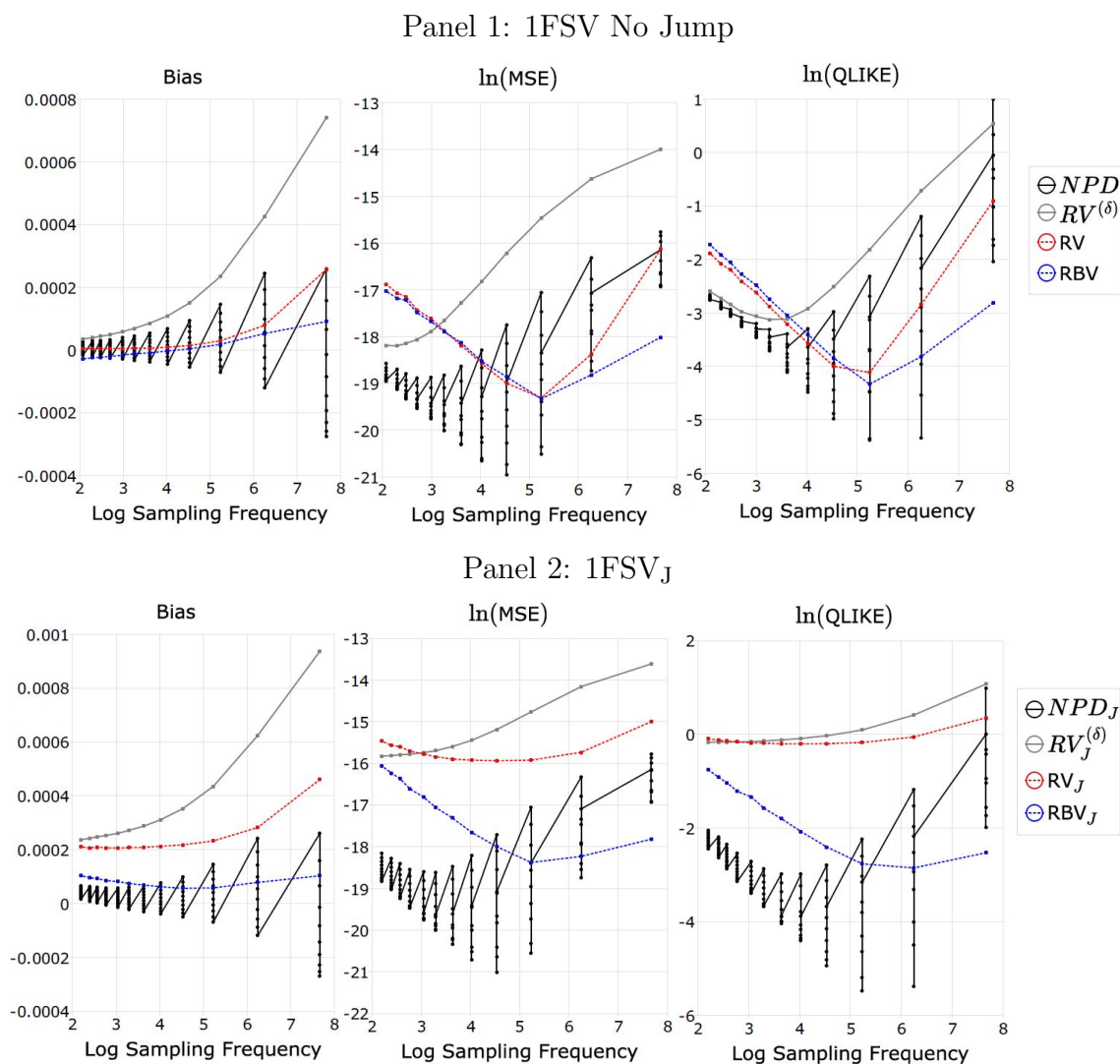
Note: Optimal log QLIKE for an estimator is the smallest log QLIKE among all the sampling frequencies considered. The smallest value is highlighted in bold. The entries for the rows $\delta = x\delta_0$ represents the value of the threshold as multiples of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The sampling frequency is the average sampling frequency at the optimal δ s for *NPD*, *RV*^(δ) and *NPD*^c, and is the calendar time sampling frequency for *RV*, *RBip*, *RK*, *PRV* and *PBip*.

Figure A.7 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD , $RV^{(\delta)}$, RV and RBip for 1FSV model with high level of MMS noise



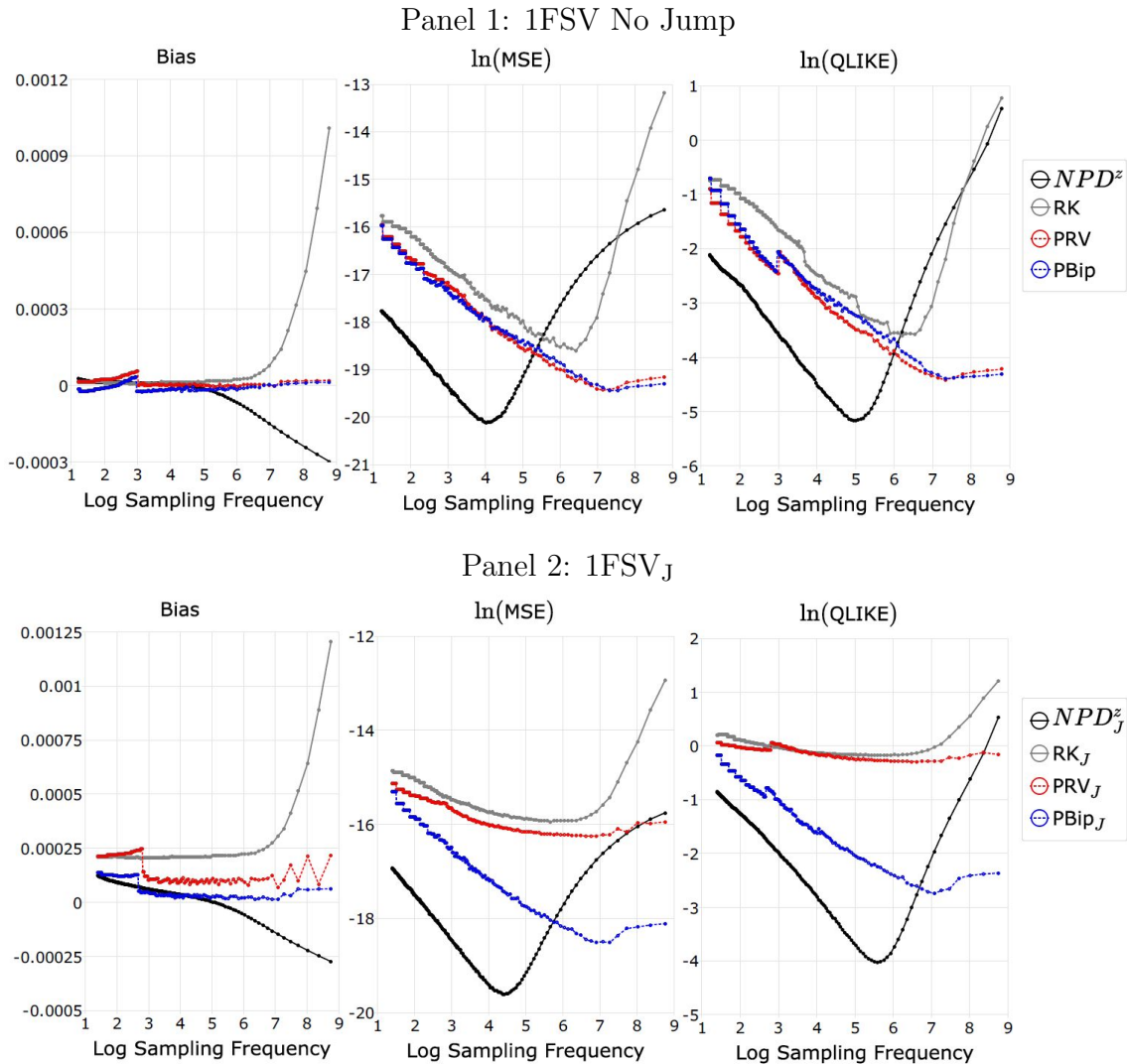
Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and RBip. The truncation threshold δ ranges from $200\delta_0$ to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.005$.

Figure A.8 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD , $RV^{(\delta)}$, RV and $RBip$ for 1FSV model with low level of MMS noise



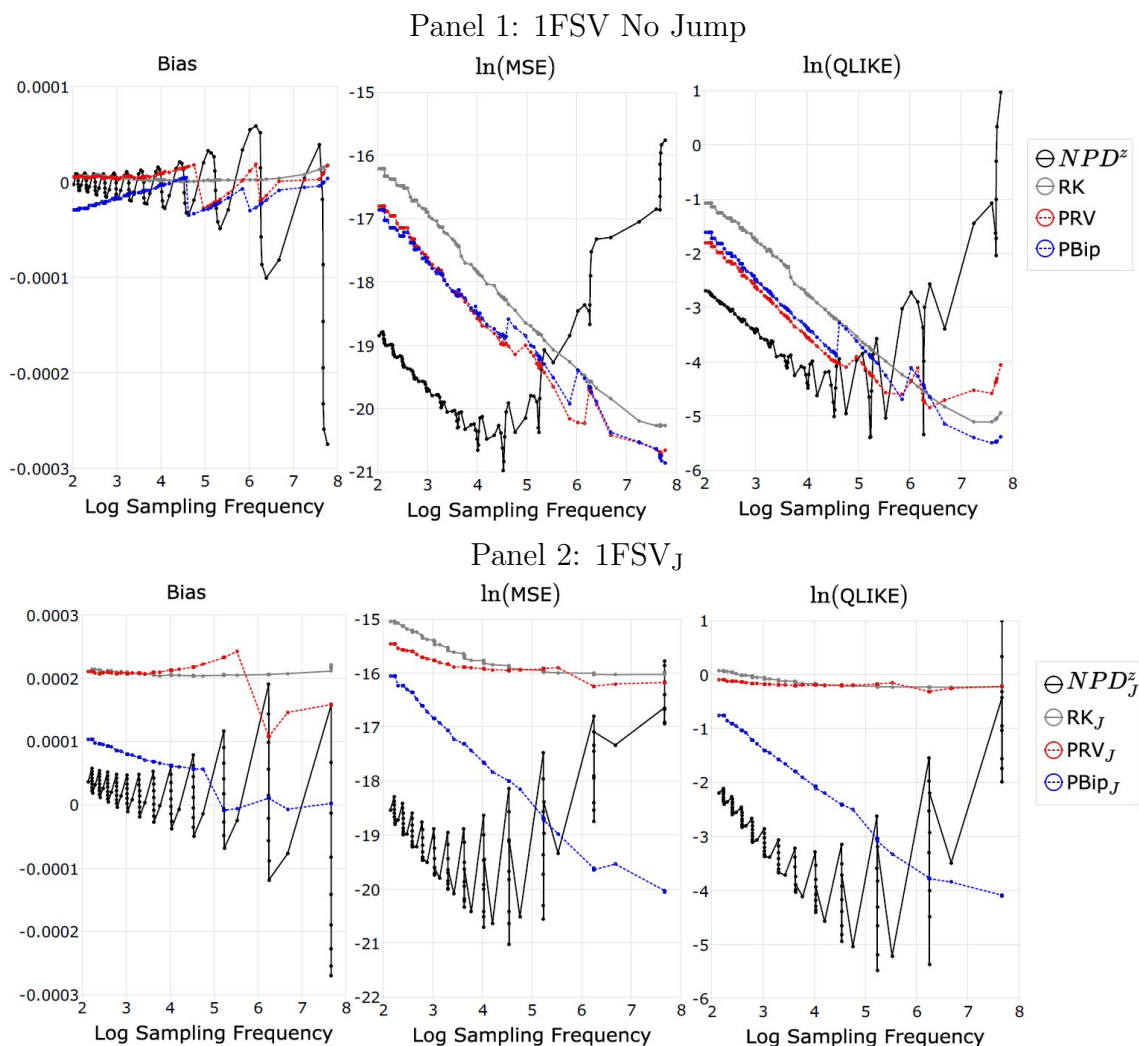
Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for NPD and $RV^{(\delta)}$, or the log sampling frequency of the equidistant intraday return per day for RV and $RBip$. The truncation threshold δ ranges from δ_{120} to δ_0 with a step size of $\delta_0 = 0.1\varepsilon$, with $\varepsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.0002$.

Figure A.9 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z , RK, PRV and PBip for 1FSV model with high level of MMS noise



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for the NPD^z model, or the log sampling frequency of the equidistant intraday return per day for RK, PRV and PBip. The truncation threshold δ ranges from δ_{150} to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.005$.

Figure A.10 Simulated Bias, MSE and QLIKE for daily volatility estimates obtained from NPD^z , RK, PRV and PBip for 1FSV model with low level of MMS noise



Note: The results are based on 10000 replications of the 1FSV model with and without jumps. The x-axis denotes the average log sampling frequency for a given δ for the NPD^z model, or the log sampling frequency of the equidistant intraday return per day for RK, PRV and PBip. The truncation threshold δ ranges from δ_{150} to δ_0 with a step size of $\delta_0 = 0.1\epsilon$, with $\epsilon = \ln(20.01) - \ln(20)$. The subscript J represents an estimator constructed on the 1FSV model with jumps. The noise-to-signal ratio is set to be $\omega = 0.0002$.

Appendix B

Appendix for Chapter 2

B.1 Analysis of Co-movements Between RK and ICV Estimates

In this section we provide empirical evidence for the following arguments in Section 2.6.3: (1) The co-movement between RK and ICV estimates increases as the number of observation used to construct RK increases. (2) The inclusion of MMS covariates improves the co-movement between RK and ICV estimates only when RK is precise. (3) In addition to the optimally selected MMS covariates, adding all MMS covariates does not have a big impact on the co-movement between RK and ICV estimates.

We define a measure of co-movements between RK and ICV estimates as follows. Let $j = 1 : 26$ denote the index of 26 fifteen-minute intraday intervals, $RK_{k,j,d}$ denote the realized kernel measure of the j -th interval at day d for stock $k = 1 : 30$, and $ICV_{k,j,d}^{M,m}$ the ICV estimates of the j -th interval for stock k based on model M with a size of estimation window m . For each d and stock k , we compute the co-movement measure between RK and ICV estimates as:

$$CM_{k,d}^{M,m} = \frac{1}{2} \ln \left(\frac{1 + \rho_{k,d}^{M,m}}{1 - \rho_{k,d}^{M,m}} \right), \quad (\text{B.1})$$

where $\rho_{k,d}^{M,m}$ is the sample correlation between $RK_{k,j,d}$ and $ICV_{k,j,d}^{M,m}$ for $j = 1 : 26$. The measure $CM_{k,d}^{M,m}$ is in essence the Fisher transformation of the daily correlation between the 15-minute RK and a 15-minute ICV estimate. Intuitively, the larger $CM_{k,d}^{M,m}$, the higher the co-movements between $RK_{k,j,d}$ and $ICV_{k,j,d}^{M,m}$ for stock k on day d . A zero $CM_{k,d}^{M,m}$ represents no correlation between $RK_{k,j,d}$ and $ICV_{k,j,d}^{M,m}$ on the day d . Also, denote $NoT_{k,d} = \frac{1}{26} \sum_{j=1}^{26} NoT_{k,j,d}$ where $NoT_{k,j,d}$ is the number of transactions

in the j -th interval on day d for stock k , so $NoT_{k,d}$ measures the average number of observations available for the RK estimator on day d .

To test the three aforementioned arguments, we propose the following set of fixed-effect regressions for each m :

$$CM_{k,d}^{P,m} = a_k + \tau_d + c^{P,m} + b^{P,m} \ln NoT_{k,d} + \varepsilon_{k,d} \quad (B.2)$$

$$CM_{k,d}^{K,m} = a_k + \tau_d + c^{K,m} + b^{K,m} \ln NoT_{k,d} + \varepsilon_{k,d} \quad (B.3)$$

$$CM_{k,d}^{A,m} = a_k + \tau_d + c^{A,m} + b^{A,m} \ln NoT_{k,d} + \varepsilon_{k,d} \quad (B.4)$$

$$CM_{k,d}^{K,m} - CM_{k,d}^{P,m} = a_k + \tau_d + c^{KP,m} + b^{KP,m} \ln NoT_{k,d} + \varepsilon_{k,d} \quad (B.5)$$

$$CM_{k,d}^{A,m} - CM_{k,d}^{K,m} = a_k + \tau_d + c^{AK,m} + b^{AK,m} \ln NoT_{k,d} + \varepsilon_{k,d}, \quad (B.6)$$

in which a_k and τ_d are firm-specific fixed effect and time fixed effect for each day. In the first three equations, the parameter $c^{M,m}$ can be interpreted as the baseline level of comovement between RK and ICV estimates, and $b^{M,m}$ represents the improvement of comovements per 1% increase of $NoT_{k,d}$. If argument (1) holds true, we expect $\hat{c}^{M,m}$ and $\hat{b}^{M,m}$ to be positive and highly significant for all M and m . In (B.5), $c^{KP,m}$ represents the baseline difference between the comovement measures of ICV estimates from the -K model and the -P model to RK, and $b^{KP,m}$ is the effect of the number of transactions on the difference of the comovement measures. If argument (2) holds true, we would expect $\hat{c}^{KP,m}$ to be insignificant, but $\hat{b}^{KP,m}$ to be positive and significant. Similarly, if argument (3) holds true, we would expect both $\hat{c}^{AK,m}$ and $\hat{b}^{AK,m}$ to be insignificant.

We present the fixed-effect regression results in Table B.1. Our findings strongly support our previous prediction that for Equations (B.2) to (B.4), both $\hat{c}^{M,m}$ and $\hat{b}^{M,m}$ are positive and significant for all m . This suggests that the comovements between RK and ICV estimates are indeed higher when the RK is more precise. It is worth noting that $\hat{c}^{P,m}$ and $\hat{b}^{P,m}$ are smaller than $\hat{c}^{K,m}$, $\hat{c}^{A,m}$ and $\hat{b}^{K,m}$, $\hat{b}^{A,m}$. This suggests that the inclusion of MMS covariates increases both the baseline comovement and the effect of the precision of RK. Regression outputs for (B.5) and (B.6) corroborate arguments (2) and (3) with insignificant $\hat{c}^{KP,m}$, $\hat{c}^{AK,m}$ and $\hat{b}^{AK,m}$, and a significantly positive $\hat{c}^{KP,m}$. These results show that when the RK measure is precise, the comovement between RK and ICV estimates is strengthened by the inclusion of optimally selected MMS covariates, and including all MMS covariates does not further improve the results.

Table B.1 Fixed-effect regressions for the comovement analysis between RK and ICV estimates

Equation	(B.2)	(B.3)	(B.4)	(B.5)	(B.6)
Panel 1: Monthly Estimation Window, $m = 1$					
$\widehat{c}^{M,m} / \widehat{c}^{KP,m} / \widehat{c}^{AK,m}$	0.7502***	0.7561***	0.7700***	0.0059	0.0139
Std. Err.	(4.81)	(4.15)	(4.18)	(0.10)	(1.89)
$\widehat{b}^{M,m} / \widehat{b}^{KP,m} / \widehat{b}^{AK,m}$	0.1336***	0.2010***	0.2004***	0.0675***	-0.0006
Std. Err.	(3.81)	(5.41)	(5.37)	(5.81)	(-0.55)
Panel 2: Quarterly Estimation Window, $m = 3$					
$\widehat{c}^{M,m} / \widehat{c}^{KP,m} / \widehat{c}^{AK,m}$	0.7274***	0.7603***	0.7618***	0.0328	0.0015
Std. Err.	(4.62)	(4.20)	(4.21)	(0.55)	(0.39)
$\widehat{b}^{M,m} / \widehat{b}^{KP,m} / \widehat{b}^{AK,m}$	0.1318***	0.2000***	0.2002***	0.0682***	0.0003
Std. Err.	(3.74)	(5.34)	(5.37)	(5.46)	(0.34)
Panel 3: Half-Yearly Estimation Window, $m = 6$					
$\widehat{c}^{M,m} / \widehat{c}^{KP,m} / \widehat{c}^{AK,m}$	0.7205***	0.7821***	0.7778***	0.0617	-0.0043
Std. Err.	(4.59)	(4.26)	(4.25)	(0.94)	(-1.39)
$\widehat{b}^{M,m} / \widehat{b}^{KP,m} / \widehat{b}^{AK,m}$	0.1306***	0.1967***	0.1976***	0.0662***	0.0009
Std. Err.	(3.70)	(5.20)	(5.26)	(4.84)	(1.41)
Panel 4: Yearly Estimation Window, $m = 12$					
$\widehat{c}^{M,m} / \widehat{c}^{KP,m} / \widehat{c}^{AK,m}$	0.7112***	0.7745***	0.7715***	0.0633	-0.0030
Std. Err.	(4.47)	(4.29)	(4.27)	(0.99)	(-1.19)
$\widehat{b}^{M,m} / \widehat{b}^{KP,m} / \widehat{b}^{AK,m}$	0.1301**	0.1973***	0.1981***	0.0672***	0.0008
Std. Err.	(3.65)	(5.39)	(5.42)	(5.06)	(1.53)

Note: Firm and time fixed effects are omitted from the output. The standard errors are robust to firm-specific clustering effect and heteroscedasticity. **: significant at 1% significance level. ***: significant at 0.1% significance level. Number of observations used in each regression is 30180.

B.2 Additional Tables and Figures

Table B.2 Daily descriptive statistics of the transaction data

Ticker	NT/day	VOL/day	DVOL/day	BAS/day	Amihud	RK/day
AA	1674.588	2.697	3.198	0.011	4.574	0.260
AIG	2658.873	1.584	6.124	0.014	2.959	0.267
AXP	2586.435	0.974	5.911	0.018	1.443	0.184
BA	2260.438	0.708	6.300	0.029	1.411	0.182
BAC	2591.780	14.939	17.166	0.011	0.872	0.278
C	4322.931	8.306	18.117	0.012	0.671	0.265
CAT	3104.053	1.026	9.689	0.027	0.984	0.212
CVX	3687.456	1.228	13.415	0.025	0.536	0.167
DD	2459.300	0.945	5.046	0.015	1.609	0.178
DIS	3174.753	1.818	9.671	0.014	0.844	0.171
GE	2516.661	5.802	12.360	0.011	0.615	0.179
GM	2940.327	2.513	7.725	0.012	1.953	0.265
HD	2873.177	1.427	7.752	0.015	1.000	0.170
IBM	2456.244	0.721	13.175	0.055	0.514	0.143
INTC	1692.897	4.219	10.527	0.011	0.968	0.195
JNJ	3364.024	1.721	12.948	0.014	0.399	0.124
JPM	4352.591	3.882	16.843	0.012	0.569	0.220
KO	2939.903	2.260	11.247	0.012	0.526	0.133
MCD	2879.529	1.058	9.590	0.017	0.579	0.124
MMM	2404.032	0.666	6.762	0.026	0.977	0.146
MRK	2810.784	2.008	8.757	0.012	0.796	0.154
MSFT	1985.356	4.664	15.236	0.011	0.641	0.182
PFE	2654.956	4.848	12.053	0.011	0.590	0.169
PG	2995.692	1.523	10.837	0.014	0.504	0.124
SPY	9141.895	30.491	441.940	0.012	0.011	0.114
T	2687.274	3.507	11.560	0.011	0.521	0.141
UTX	2243.252	0.651	5.814	0.026	1.321	0.163
VZ	2755.626	2.053	8.870	0.012	0.781	0.147
WMT	2958.746	1.537	9.906	0.014	0.603	0.131
XOM	4791.709	2.793	23.870	0.015	0.283	0.157

Note: NT/day: number of transactions per day. VOL/day: Total trading volume per day $\times 10^{-6}$. DVOL/day: total dollar volume per day $\times 10^{-7}$. BAS/trade: average bid-ask spread for every transaction within the trading day. Amihud: Amihud's (2002) illiquidity measure $\times 10^{10}$. RK/day: realized kernel estimates of daily annualized volatility per day.

Table B.3 Descriptive statistics of the dataset

Variables	Ticker	AA	AIG	AXP	BA	BAC	C	CAT	CVX	DD	DIS	GE	GM	HD	IBM	INTC
$\ln x_i^{(\delta)}$	Mean	4.909	4.951	4.983	4.942	4.933	4.882	4.986	5.020	4.995	4.983	4.965	4.880	4.982	4.982	4.943
	Std. Dev.	1.485	1.342	1.334	1.355	1.393	1.423	1.296	1.293	1.317	1.352	1.384	1.417	1.324	1.329	1.400
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	9.354	9.097	8.874	9.109	9.640	9.886	8.922	8.821	9.173	8.789	9.195	9.374	9.140	8.815	9.332
VOL_i	Mean	4.751	4.091	3.623	3.360	6.557	5.375	3.760	3.933	3.579	4.330	5.575	4.718	4.022	3.387	5.053
	Std. Dev.	1.227	1.076	1.011	1.047	1.150	1.145	0.935	0.921	1.020	0.864	1.064	1.058	0.998	1.054	1.149
	Min	-1.641	-1.887	-3.664	-4.883	-1.423	-2.028	-2.741	-2.526	-3.006	-1.946	-2.639	-1.792	-1.386	-3.450	-5.194
	Max	11.832	11.614	9.872	9.856	12.822	14.217	10.671	11.264	10.477	11.305	13.218	10.969	10.567	9.915	17.431
QD_i	Mean	0.700	0.523	0.447	0.466	0.615	0.476	0.400	0.393	0.438	0.528	0.582	0.489	0.438	0.454	0.449
	Std. Dev.	0.616	0.501	0.427	0.468	0.508	0.496	0.431	0.414	0.443	0.467	0.512	0.466	0.416	0.470	0.422
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	6.557	6.337	5.445	6.297	6.033	5.915	6.188	6.585	5.640	5.872	6.240	5.504	5.534	8.700	7.597
BAS_i	Mean	0.011	0.018	0.024	0.043	0.011	0.013	0.036	0.033	0.020	0.017	0.011	0.014	0.019	0.076	0.011
	Std. Dev.	0.002	0.012	0.021	0.041	0.001	0.005	0.026	0.027	0.016	0.017	0.002	0.007	0.014	0.059	0.005
	Min	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	Max	0.100	0.230	1.190	2.370	0.075	0.205	0.540	0.660	0.310	2.000	0.140	0.180	0.380	4.350	0.180
OI_i	Mean	0.930	0.646	0.600	0.628	0.968	0.611	0.547	0.520	0.602	0.629	0.891	0.702	0.625	0.628	0.809
	Std. Dev.	0.751	0.544	0.503	0.512	0.781	0.555	0.480	0.461	0.511	0.537	0.722	0.584	0.536	0.511	0.652
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	6.277	5.273	5.460	3.584	6.118	6.795	4.317	3.892	4.263	4.443	5.468	4.890	4.564	4.190	5.075
OF_i	Mean	2.820	1.469	1.383	1.568	2.689	1.385	1.222	1.101	1.407	1.340	2.213	1.590	1.352	1.482	2.412
	Std. Dev.	3.192	2.031	1.908	2.024	3.437	2.076	1.747	1.608	1.950	1.920	2.935	2.218	1.918	1.954	3.038
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	13.443	12.480	11.452	10.762	15.337	15.084	11.365	11.957	11.755	13.793	13.389	12.502	11.613	11.195	17.431
NT_i	Mean	-2.464	-1.981	-2.075	-2.201	-2.030	-1.458	-1.881	-1.720	-2.136	-1.815	-2.026	-1.828	-1.911	-2.138	-2.353
	Std. Dev.	0.929	0.702	0.691	0.721	0.832	0.701	0.642	0.609	0.711	0.615	0.794	0.746	0.661	0.722	0.877
	Min	-6.312	-5.523	-5.371	-5.765	-6.028	-7.431	-5.182	-4.543	-7.555	-7.186	-5.525	-4.844	-4.668	-5.493	-8.084
	Max	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TQ_i	Mean	7.245	5.233	4.500	4.049	9.432	7.042	4.422	4.521	4.595	5.493	8.084	6.108	5.122	3.906	7.569
	Std. Dev.	1.031	1.018	0.975	0.938	0.889	1.775	0.805	0.796	0.961	0.892	0.863	0.843	1.025	0.937	0.964
	Min	2.066	1.109	0.446	-0.344	3.240	0.667	0.904	0.755	-2.257	-1.888	3.219	1.238	1.291	0.081	-1.687
	Max	12.962	12.868	10.670	10.574	14.744	16.072	10.795	11.411	11.310	11.760	13.075	10.956	12.207	10.522	13.907
Obs./Day		69.804	77.298	77.297	77.827	73.753	75.233	77.459	77.426	77.167	76.733	74.507	77.455	77.063	77.655	75.477

For the definition of the variables, see Table 2.1.

Table B.4 Descriptive statistics of the dataset

Variables	Ticker	JNJ	JPM	KO	MCD	MMM	MRK	MSFT	PFE	PG	SPY	T	UTX	VZ	WMT	XOM
$\ln x_i^{(\delta)}$	Mean	4.928	4.947	4.896	4.929	4.988	4.930	4.904	4.956	4.923	5.143	4.879	4.966	4.889	4.914	5.051
	Std. Dev.	1.416	1.347	1.440	1.392	1.330	1.416	1.436	1.417	1.410	1.135	1.451	1.348	1.431	1.411	1.263
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	5.081	5.063	5.056	5.069	5.106	5.088	5.050	5.106	5.075	5.220	5.017	5.088	5.030	5.050	5.165
VOL_i	Mean	4.300	5.079	4.623	3.813	3.291	4.457	5.088	5.380	4.182	7.258	5.118	3.247	4.507	4.159	4.676
	Std. Dev.	0.980	0.992	0.978	0.990	0.959	0.942	1.110	0.982	0.980	0.934	0.966	1.022	1.029	1.015	0.978
	Min	-3.192	-2.001	-1.526	-3.434	-4.159	-2.401	-5.318	-0.182	-3.434	-3.002	-1.402	-3.932	-1.179	-1.516	0.134
	Max	4.178	5.012	4.514	3.704	3.209	4.351	5.031	5.289	4.058	7.235	5.010	3.158	4.386	4.058	4.602
QD_i	Mean	0.457	0.416	0.542	0.432	0.403	0.527	0.417	0.633	0.505	0.288	0.608	0.416	0.557	0.477	0.498
	Std. Dev.	0.442	0.416	0.509	0.430	0.413	0.489	0.400	0.539	0.501	0.248	0.520	0.438	0.491	0.449	0.487
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	0.339	0.302	0.405	0.315	0.288	0.400	0.310	0.495	0.372	0.226	0.476	0.295	0.429	0.360	0.355
BAS_i	Mean	0.018	0.014	0.014	0.023	0.038	0.015	0.011	0.012	0.018	0.012	0.012	0.040	0.014	0.018	0.019
	Std. Dev.	0.015	0.007	0.011	0.020	0.037	0.010	0.005	0.006	0.015	0.002	0.004	0.039	0.009	0.013	0.014
	Min	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	Max	0.013	0.012	0.012	0.017	0.026	0.012	0.011	0.011	0.013	0.012	0.011	0.028	0.012	0.013	0.015
OI_i	Mean	0.637	0.601	0.735	0.607	0.583	0.743	0.764	0.928	0.651	0.390	0.868	0.622	0.746	0.659	0.522
	Std. Dev.	0.552	0.532	0.620	0.511	0.486	0.623	0.629	0.748	0.549	0.385	0.704	0.506	0.620	0.554	0.471
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	0.488	0.452	0.582	0.470	0.460	0.588	0.610	0.731	0.511	0.276	0.693	0.511	0.588	0.511	0.395
OF_i	Mean	1.350	1.169	1.636	1.382	1.397	1.602	2.123	2.135	1.448	0.741	1.985	1.516	1.660	1.449	1.031
	Std. Dev.	1.952	1.802	2.307	1.934	1.877	2.230	2.800	2.850	2.017	1.281	2.680	1.986	2.290	2.022	1.515
	Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	0.667	0.593	0.753	0.691	0.703	0.760	0.945	0.934	0.726	0.436	0.885	0.760	0.784	0.717	0.597
NT_i	Mean	-1.723	-1.458	-1.848	-1.922	-2.169	-1.893	-2.164	-1.970	-1.834	-0.787	-1.903	-2.234	-1.889	-1.856	-1.451
	Std. Dev.	0.660	0.634	0.733	0.669	0.696	0.700	0.863	0.785	0.667	0.499	0.757	0.725	0.728	0.686	0.587
	Min	-5.383	-4.091	-4.610	-5.004	-5.541	-7.699	-8.026	-5.432	-5.485	-7.251	-6.701	-6.930	-4.732	-4.500	-3.899
	Max	-1.768	-1.472	-1.880	-1.976	-2.219	-1.943	-2.241	-2.019	-1.889	-0.733	-1.954	-2.276	-1.937	-1.910	-1.466
TQ_i	Mean	5.430	6.460	6.117	4.694	3.936	5.935	7.492	7.574	5.236	8.891	7.064	3.971	6.048	5.311	5.546
	Std. Dev.	0.953	0.840	0.989	0.867	0.846	0.955	0.934	0.925	0.922	0.576	0.898	0.918	0.950	0.961	0.902
	Min	0.734	2.163	1.888	1.050	0.046	0.274	-2.728	2.996	1.518	0.237	2.139	-1.632	1.661	1.537	2.154
	Max	5.333	6.430	6.060	4.605	3.871	5.871	7.438	7.526	5.123	8.893	7.011	3.900	5.991	5.218	5.447
Obs./Day		76.492	77.099	76.828	77.359	77.475	76.642	76.346	74.268	77.083	77.555	76.207	77.599	76.651	76.826	77.405

For the definition of the variables, see Table 2.1.

Table B.5 Average of quarterly rankings of relative importance of MMS covariates

Ticker	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	γ_{TQ_0}	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	γ_{TQ_1}	\bar{K}^*
AA	9.81	10.13	4.31	12.06	13.31	13.63	8.00	5.13	2.19	2.25	5.19	6.50	5.75	6.75	10.25
AIG	4.13	10.88	10.56	5.00	13.44	13.56	11.25	2.69	3.75	4.56	3.13	6.75	8.25	7.06	8.75
AXP	6.50	10.44	10.38	5.94	13.44	13.56	10.50	2.38	3.81	4.13	3.06	6.56	8.13	6.19	9.63
BA	8.06	10.63	11.00	7.94	13.38	13.63	10.69	2.50	2.75	6.56	3.13	6.00	5.06	3.69	8.81
BAC	8.50	10.38	2.31	12.00	14.00	12.69	8.00	6.25	2.38	2.38	5.75	8.19	5.50	6.69	9.44
C	8.06	9.56	8.88	9.56	13.63	13.19	9.69	4.56	3.75	2.38	2.31	6.31	7.56	5.56	10.81
CAT	3.50	10.88	10.63	6.88	13.81	13.19	11.06	2.63	3.50	7.38	2.44	5.31	8.00	5.81	9.00
CVX	3.31	10.63	10.88	8.19	13.81	13.19	10.75	2.63	4.06	5.81	2.44	5.88	7.44	6.00	9.13
DD	4.56	11.00	10.13	6.38	13.56	13.44	11.13	2.81	4.88	3.13	3.13	5.25	8.38	7.25	9.63
DIS	5.19	10.75	10.81	5.69	13.38	13.63	11.06	3.19	3.81	2.56	3.25	6.38	8.25	7.00	9.75
GE	9.50	10.50	5.00	11.56	13.75	13.19	9.75	2.88	1.94	2.19	4.38	6.63	7.06	6.50	9.63
GM	5.06	10.31	10.06	7.31	13.31	13.69	11.13	2.50	3.13	2.31	3.50	6.75	8.88	7.00	9.81
HD	4.44	10.81	11.13	4.63	13.56	13.44	10.69	4.13	3.25	4.94	2.31	5.88	8.56	7.25	10.00
IBM	7.50	10.44	11.56	8.75	13.69	13.31	10.44	2.44	2.31	5.44	4.25	6.88	4.06	3.81	8.56
INTC	5.69	10.69	4.00	11.94	13.06	13.94	9.81	2.81	2.63	3.00	5.38	7.44	7.94	6.69	8.75
JNJ	6.06	10.56	10.94	6.06	13.38	13.63	10.69	2.38	4.25	2.75	3.06	6.13	7.88	7.25	10.38
JPM	6.94	10.13	9.06	8.94	13.69	13.31	10.69	2.63	3.69	2.19	3.00	5.94	8.38	6.38	9.94
KO	7.56	9.50	10.75	8.94	13.44	13.56	10.31	2.56	4.25	2.25	2.56	5.25	7.69	6.38	10.31
MCD	3.19	11.06	10.50	4.63	13.44	13.56	11.31	3.13	4.25	5.56	2.50	6.56	7.75	7.50	8.94
MMM	6.25	10.63	10.69	7.81	13.38	13.63	10.81	3.00	3.00	4.94	3.69	6.81	5.44	4.94	8.44
MRK	7.13	9.81	10.50	8.75	13.44	13.50	10.19	2.69	3.38	2.25	3.13	6.13	7.44	6.69	9.69
MSFT	3.94	10.81	5.50	11.50	13.13	13.88	10.31	2.31	2.69	2.38	5.25	7.94	8.06	7.25	8.25
PFE	9.56	9.88	7.50	11.81	13.69	13.13	10.31	3.00	2.06	1.63	3.69	6.06	6.56	6.06	9.88
PG	4.63	10.50	10.94	5.88	13.31	13.69	11.13	2.19	3.88	3.75	2.81	6.44	8.38	7.50	9.38
SPY	12.81	9.69	2.25	13.31	11.69	11.06	6.94	6.00	2.25	3.25	3.31	6.81	9.56	5.94	9.50
T	10.50	9.56	6.63	11.44	13.56	13.00	10.38	2.25	3.06	2.25	3.25	6.00	6.69	6.44	10.63
UTX	8.69	10.56	10.81	7.75	13.19	13.81	10.19	2.13	2.50	4.06	4.06	6.13	5.94	5.19	8.88
VZ	7.31	10.31	9.44	9.56	13.50	13.50	10.63	2.25	2.56	2.31	3.25	6.81	7.38	6.13	10.06
WMT	4.63	10.69	10.88	5.56	13.44	13.56	11.13	2.75	4.25	3.44	3.06	6.25	8.25	7.13	9.69
XOM	5.75	10.06	9.94	9.81	13.44	13.50	10.94	2.25	4.13	4.88	1.44	5.31	7.25	6.25	10.38
Average	6.63	10.39	8.93	8.52	13.43	13.39	10.33	3.03	3.28	3.56	3.39	6.38	7.38	6.34	9.54

Note: For each security, we report the averaged number of inclusions in the quarterly K -optimal models for every MMS covariate. \bar{K}^* is the average number of MMS covariates in the yearly K^* -optimal model, averaged over 4 years. Covariates that receive top five average inclusions are highlighted in bold.

Table B.6 Average of half-yearly rankings of relative importance of MMS covariates

Ticker	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	γ_{rQ_0}	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	γ_{rQ_1}	\bar{K}^*
AA	8.75	10.25	4.00	12.00	13.38	13.63	6.75	5.63	2.13	2.25	5.75	6.75	6.00	7.75	11.00
AIG	3.88	11.13	10.13	4.50	13.38	13.63	11.38	2.75	3.38	4.25	3.63	7.50	8.25	7.25	10.13
AXP	5.88	10.38	10.50	7.00	13.38	13.63	10.50	1.88	4.25	3.00	2.88	6.25	8.38	7.13	11.50
BA	8.50	10.63	10.75	8.13	13.38	13.63	10.63	1.75	2.50	5.63	3.13	6.38	6.13	3.88	10.25
BAC	8.13	10.75	2.50	12.00	14.00	12.88	7.25	6.50	2.00	2.38	6.38	8.13	5.38	6.63	10.63
C	6.25	9.13	9.50	9.88	13.38	13.25	9.88	3.75	4.25	2.13	1.25	7.13	8.75	6.38	11.38
CAT	4.75	10.88	10.63	6.38	13.75	13.25	10.88	3.00	2.88	7.50	1.88	5.88	8.00	5.38	11.25
CVX	3.13	10.75	10.75	7.00	13.75	13.25	10.75	1.88	4.75	5.88	2.00	6.25	8.13	6.63	10.88
DD	4.63	11.00	10.25	7.25	13.63	13.38	11.00	2.38	5.00	2.50	2.75	5.63	8.38	7.25	10.75
DIS	5.25	10.75	9.88	4.88	13.50	13.50	11.25	2.63	4.50	2.88	3.00	7.00	8.38	7.63	10.63
GE	9.00	10.75	6.38	11.75	14.00	12.88	9.38	2.63	2.00	1.75	4.25	7.13	6.75	6.25	10.75
GM	4.25	10.50	10.25	8.00	13.38	13.63	11.38	2.50	3.38	1.63	3.75	6.63	8.63	7.13	11.00
HD	4.50	10.88	11.25	4.25	13.50	13.50	10.75	3.88	3.63	4.50	2.38	6.25	8.13	7.63	11.25
IBM	7.88	10.13	11.75	7.88	13.88	13.13	10.50	2.13	2.00	6.25	4.25	7.00	4.50	3.75	9.75
INTC	5.38	10.75	4.13	11.50	13.00	14.00	10.50	2.00	2.50	2.25	5.88	8.00	8.13	7.00	9.75
JNJ	4.88	10.00	11.63	5.88	13.50	13.50	10.63	2.63	4.75	2.13	2.88	7.00	8.13	7.50	10.50
JPM	7.13	10.13	9.75	8.75	13.50	13.50	10.63	2.50	4.00	2.25	1.88	6.25	8.13	6.63	11.13
KO	6.63	9.63	10.75	9.75	13.38	13.63	9.63	2.13	4.00	1.63	3.25	5.75	7.63	7.25	11.13
MCD	3.50	11.00	10.50	5.63	13.38	13.63	11.25	2.13	4.13	6.13	3.00	6.63	7.38	6.75	10.63
MMM	5.75	10.75	11.00	8.25	13.13	13.88	10.75	2.25	3.25	4.88	3.25	6.63	6.00	5.25	10.13
MRK	5.50	10.13	10.63	8.63	13.25	13.75	10.50	2.25	3.75	2.25	3.25	6.13	7.75	7.25	11.38
MSFT	3.13	10.75	5.13	11.63	13.00	14.00	10.25	2.25	3.00	2.38	5.50	8.13	8.38	7.50	9.13
PFE	9.50	10.00	7.75	12.00	13.63	13.25	10.00	2.63	1.75	1.75	3.88	5.75	7.00	6.00	11.00
PG	4.75	10.38	11.38	5.25	13.38	13.63	11.13	1.75	5.00	3.25	2.50	6.63	8.25	7.75	10.13
SPY	12.25	9.88	4.13	13.38	11.75	11.25	6.50	5.25	2.00	2.75	2.00	7.75	9.88	6.25	11.13
T	9.88	9.63	7.50	11.13	13.63	13.25	10.38	2.25	3.38	1.25	3.38	6.25	6.88	6.25	11.75
UTX	8.25	10.63	10.63	7.50	13.25	13.75	10.63	2.13	2.75	3.50	3.88	6.25	6.38	5.38	10.88
VZ	6.50	10.13	9.88	9.38	13.50	13.50	10.63	1.75	2.38	2.50	3.75	6.75	7.75	6.50	10.25
WMT	4.13	10.50	11.13	6.00	13.38	13.63	11.25	2.25	4.50	3.00	3.38	5.88	8.63	7.38	11.13
XOM	5.25	10.25	9.38	9.75	13.25	13.50	11.13	2.25	4.00	4.63	1.00	5.75	8.00	6.88	11.38
Average	6.24	10.41	9.13	8.51	13.40	13.43	10.27	2.72	3.39	3.30	3.33	6.65	7.60	6.60	10.75

Note: For each security, we report the averaged number of inclusions in the half-yearly K -optimal models for every MMS covariate. \bar{K}^* is the average number of MMS covariates in the half-yearly K^* -optimal model, averaged over 4 years. Covariates that receive top five average inclusions are highlighted in bold.

Table B.7 Average of yearly rankings of relative importance of MMS covariates

Ticker	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	γ_{TQ_0}	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	γ_{TQ_1}	\bar{K}^*
AA	9.00	10.50	3.75	12.00	13.50	13.50	6.50	5.50	2.00	1.75	6.00	7.25	6.50	7.25	11.50
AIG	3.00	11.00	10.25	4.25	13.50	13.50	11.50	2.75	4.00	4.50	3.50	7.50	8.50	7.25	11.50
AXP	6.50	10.75	9.75	6.25	13.50	13.50	10.75	1.75	4.25	3.75	2.75	6.50	8.00	7.00	12.25
BA	8.25	10.50	10.50	8.00	13.25	13.75	10.75	1.50	2.50	5.00	3.75	6.50	6.25	4.50	11.25
BAC	9.00	11.00	2.00	12.00	14.00	12.75	7.00	6.75	1.75	2.25	6.25	7.75	5.50	7.00	11.50
C	7.50	9.25	9.25	9.25	13.75	13.25	10.00	4.50	3.75	1.50	1.75	6.75	8.00	6.50	12.25
CAT	4.50	11.00	10.50	6.00	13.75	13.25	11.00	3.00	3.50	7.50	1.50	5.75	8.00	5.75	12.25
CVX	2.50	10.75	10.75	7.75	14.00	13.00	10.75	1.50	4.75	5.75	2.50	6.00	8.00	7.00	11.25
DD	5.50	11.00	10.50	6.75	13.50	13.50	11.25	2.00	4.75	2.50	2.00	5.75	8.50	7.50	12.00
DIS	5.00	11.00	9.50	4.50	13.50	13.50	11.00	2.25	5.00	2.00	3.00	7.50	9.00	8.25	11.75
GE	9.00	11.00	7.25	12.00	14.00	13.00	9.00	2.75	2.00	1.50	3.75	7.00	6.75	6.00	11.50
GM	4.25	10.75	10.75	7.00	13.25	13.75	11.25	2.25	3.50	1.50	4.50	6.75	8.75	6.75	11.50
HD	4.25	11.00	11.25	3.25	13.25	13.75	10.75	3.75	4.25	4.25	2.25	6.75	8.50	7.75	11.75
IBM	8.25	10.25	11.75	7.50	14.00	13.00	10.50	2.00	1.50	6.50	4.50	6.75	4.75	3.75	11.00
INTC	4.25	10.75	5.25	12.00	13.00	14.00	9.50	2.00	2.75	1.75	6.25	8.25	7.75	7.50	10.00
JNJ	5.50	9.75	11.75	5.00	13.50	13.50	10.50	1.50	5.50	2.00	2.75	7.00	8.75	8.00	12.00
JPM	7.00	10.00	9.50	8.50	13.75	13.25	10.50	2.50	4.00	1.50	2.25	6.75	9.00	6.25	11.25
KO	6.75	9.50	11.00	10.25	13.50	13.50	9.50	1.75	4.50	1.75	2.50	5.50	7.75	7.25	12.00
MCD	2.50	11.25	10.00	5.25	13.25	13.75	11.00	1.25	4.50	5.00	2.75	7.50	8.75	8.25	11.75
MMM	6.50	11.00	10.75	8.25	13.25	13.75	11.00	2.00	3.50	4.00	3.25	7.25	5.50	5.00	11.75
MRK	6.25	10.00	10.50	8.25	13.50	13.50	10.25	2.00	4.50	2.25	2.75	6.00	8.00	7.25	11.50
MSFT	2.25	10.75	5.50	12.00	13.00	14.00	10.00	2.75	2.75	2.00	5.50	7.75	8.50	8.00	9.50
PFE	9.50	10.00	6.25	12.00	13.75	13.25	10.25	3.25	2.00	1.75	4.00	6.00	6.75	6.25	11.75
PG	5.00	10.25	11.75	5.00	13.50	13.50	11.00	1.50	5.00	3.25	2.25	6.50	8.50	8.00	11.25
SPY	12.25	9.50	4.00	13.50	12.00	11.50	5.75	6.00	1.75	2.50	2.25	7.50	10.25	6.25	11.50
T	10.00	9.75	6.75	11.75	13.50	13.25	10.50	2.00	3.50	1.25	3.25	5.75	7.00	6.75	12.75
UTX	8.50	10.75	10.50	7.25	13.25	13.75	10.75	2.00	2.75	3.75	3.50	6.50	6.25	5.50	12.25
VZ	6.75	10.25	10.50	9.25	13.50	13.50	10.50	2.00	2.50	1.75	3.75	6.50	7.50	6.75	11.25
WMT	3.50	10.75	11.25	5.50	13.50	13.50	11.00	2.25	5.50	2.50	3.75	6.00	8.25	7.75	12.00
XOM	5.75	10.50	8.75	9.75	13.75	13.25	11.25	1.50	4.25	4.50	1.75	5.50	7.50	7.00	11.75
Average	6.29	10.48	9.06	8.33	13.48	13.39	10.18	2.62	3.56	3.06	3.35	6.69	7.70	6.80	11.58

Note: For each security, we report the averaged number of inclusions in the yearly K -optimal models for every MMS covariate. \bar{K}^* is the average number of MMS covariates in the yearly K^* -optimal model, averaged over 4 years. Covariates that receive top five average inclusions are highlighted in bold.

Table B.8 Average MMS parameter estimates for the LL-ACD(1,1)-K model

Ticker	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	γ_{TQ_0}	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	γ_{TQ_1}
AA	0.066	-0.126	-0.009	0.555	-0.781	-1.063	0.188	0.034	0.000	0.000	0.030	-0.023	0.017	-0.052
AIG	-0.003	-0.184	-0.162	0.024	-0.441	-0.977	0.418	0.000	0.001	0.003	0.000	-0.008	0.027	-0.021
AXP	0.055	-0.179	-0.199	0.001	-0.426	-0.970	0.399	0.000	0.005	0.005	0.004	-0.007	0.039	-0.029
BA	0.082	-0.207	-0.212	0.086	-0.489	-1.057	0.450	-0.004	0.001	0.018	0.000	-0.008	0.014	-0.006
BAC	-0.102	-0.140	0.000	0.441	-0.749	-0.877	0.261	0.032	0.000	0.000	0.022	-0.035	0.004	-0.039
C	-0.041	-0.151	-0.103	-0.070	-0.367	-0.944	0.335	0.018	0.013	0.001	0.005	-0.021	0.059	-0.031
CAT	0.017	-0.209	-0.223	-0.050	-0.358	-0.954	0.458	0.002	0.002	0.032	0.000	-0.002	0.029	-0.007
CVX	-0.036	-0.220	-0.248	-0.103	-0.322	-0.884	0.476	0.009	0.005	0.016	-0.002	-0.006	0.025	-0.022
DD	0.027	-0.188	-0.167	-0.001	-0.426	-0.957	0.418	-0.001	0.003	0.005	-0.001	-0.003	0.035	-0.022
DIS	-0.071	-0.168	-0.172	-0.053	-0.379	-0.825	0.355	0.002	0.004	0.014	0.006	-0.018	0.036	-0.026
GE	-0.228	-0.142	-0.033	0.302	-0.640	-0.865	0.349	0.002	0.000	0.002	0.007	-0.011	0.012	-0.009
GM	-0.062	-0.149	-0.108	0.073	-0.499	-1.056	0.398	-0.003	0.004	0.001	0.008	-0.017	0.065	-0.037
HD	-0.035	-0.173	-0.181	-0.032	-0.394	-0.928	0.407	-0.001	0.003	0.020	0.000	-0.006	0.046	-0.025
IBM	0.055	-0.224	-0.256	0.099	-0.494	-1.024	0.491	0.000	0.000	0.016	0.000	-0.006	0.004	0.000
INTC	0.000	-0.125	-0.007	0.251	-0.626	-1.073	0.252	-0.002	0.000	0.003	0.007	-0.017	0.025	-0.010
JNJ	-0.103	-0.178	-0.176	-0.017	-0.418	-0.955	0.480	-0.003	0.006	0.005	0.002	-0.013	0.045	-0.035
JPM	-0.151	-0.145	-0.114	-0.114	-0.331	-0.854	0.381	0.007	0.007	0.011	0.000	-0.013	0.054	-0.019
KO	-0.130	-0.157	-0.149	0.095	-0.501	-0.928	0.407	-0.002	0.004	0.002	0.005	-0.013	0.038	-0.026
MCD	-0.006	-0.199	-0.193	-0.001	-0.433	-0.987	0.454	0.000	0.000	0.020	-0.001	-0.010	0.017	-0.012
MMM	0.063	-0.215	-0.203	0.015	-0.439	-1.035	0.466	-0.004	0.001	0.016	0.000	-0.009	0.015	-0.007
MRK	-0.117	-0.160	-0.142	0.093	-0.491	-0.918	0.415	-0.005	0.002	0.004	0.003	-0.013	0.047	-0.034
MSFT	0.002	-0.133	-0.018	0.202	-0.606	-1.104	0.268	0.000	0.000	0.001	0.008	-0.023	0.049	-0.021
PFE	-0.260	-0.159	-0.077	0.298	-0.616	-0.874	0.440	0.000	0.000	0.003	0.005	-0.011	0.034	-0.020
PG	-0.059	-0.193	-0.184	-0.003	-0.439	-0.965	0.467	-0.004	0.002	0.006	0.000	-0.009	0.029	-0.027
SPY	-0.384	-0.108	-0.009	-0.299	-0.153	-0.642	0.236	0.018	0.000	-0.001	-0.003	-0.010	0.150	-0.015
T	-0.271	-0.158	-0.061	0.231	-0.589	-0.890	0.457	0.000	0.004	0.001	0.007	-0.017	0.046	-0.040
UTX	0.102	-0.210	-0.185	0.091	-0.496	-1.055	0.456	0.002	0.000	0.005	0.002	-0.006	0.011	-0.009
VZ	-0.136	-0.155	-0.134	0.123	-0.522	-0.941	0.402	0.001	0.000	0.002	0.003	-0.020	0.043	-0.022
WMT	-0.073	-0.173	-0.177	0.016	-0.446	-0.947	0.425	-0.004	0.004	0.010	0.002	-0.008	0.052	-0.030
XOM	-0.102	-0.195	-0.192	-0.156	-0.282	-0.830	0.488	-0.003	0.009	0.019	-0.002	-0.007	0.060	-0.041
Overall	-0.063	-0.171	-0.136	0.070	-0.472	-0.946	0.396	0.003	0.003	0.008	0.004	-0.012	0.038	-0.023

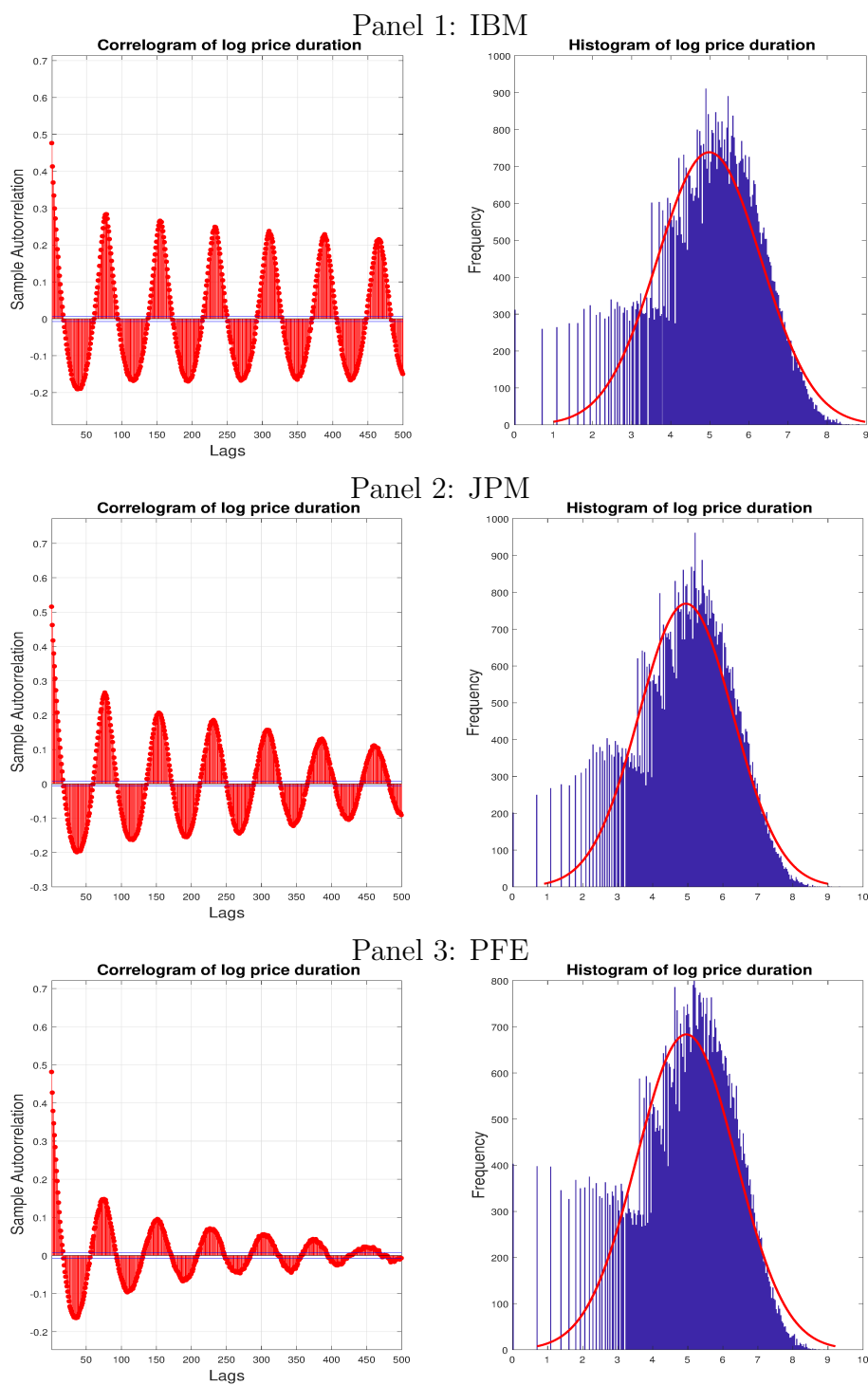
Note: The table presents average estimates of MMS parameter estimates for each stock. For each stock, we estimate the LL-ACD(1,1)-K model on a monthly basis to obtain 48 parameter estimates per each MMS variable, and report the average of the 48 parameter estimates per MMS variable per stock.

Table B.9 Average MMS parameter estimates for the LL-ACD(1,1)-A model

Ticker	$\gamma_{VOL,0}$	$\gamma_{QD,0}$	$\gamma_{BAS,0}$	$\gamma_{OI,0}$	$\gamma_{OF,0}$	$\gamma_{NT,0}$	γ_{TQ_0}	$\gamma_{VOL,1}$	$\gamma_{QD,1}$	$\gamma_{BAS,1}$	$\gamma_{OI,1}$	$\gamma_{OF,1}$	$\gamma_{NT,1}$	γ_{TQ_1}
AA	0.051	-0.131	-0.017	0.552	-0.776	-1.089	0.236	0.056	0.015	0.013	0.049	-0.043	0.029	-0.092
AIG	-0.017	-0.188	-0.169	0.014	-0.433	-0.977	0.437	-0.001	0.013	0.026	0.018	-0.028	0.050	-0.053
AXP	0.046	-0.182	-0.203	-0.002	-0.424	-0.969	0.410	0.014	0.020	0.029	0.024	-0.042	0.065	-0.070
BA	0.113	-0.208	-0.216	0.094	-0.499	-1.075	0.445	0.001	0.012	0.042	0.018	-0.033	0.027	-0.036
BAC	-0.116	-0.143	-0.006	0.439	-0.745	-0.898	0.300	0.050	0.011	0.007	0.047	-0.054	0.004	-0.080
C	-0.079	-0.161	-0.112	-0.074	-0.358	-0.935	0.368	0.038	0.039	0.026	0.011	-0.037	0.070	-0.074
CAT	0.020	-0.210	-0.231	-0.058	-0.353	-0.959	0.465	0.025	0.011	0.054	0.011	-0.024	0.041	-0.044
CVX	-0.060	-0.225	-0.257	-0.106	-0.316	-0.876	0.497	0.007	0.020	0.040	0.008	-0.023	0.048	-0.055
DD	0.020	-0.190	-0.172	0.000	-0.426	-0.958	0.432	0.016	0.021	0.029	0.017	-0.026	0.054	-0.068
DIS	-0.103	-0.172	-0.177	-0.060	-0.372	-0.816	0.379	0.017	0.017	0.032	0.022	-0.043	0.046	-0.053
GE	-0.228	-0.143	-0.049	0.302	-0.639	-0.873	0.367	0.001	0.014	0.020	0.028	-0.034	0.036	-0.051
GM	-0.123	-0.157	-0.123	0.065	-0.488	-1.039	0.437	0.007	0.016	0.026	0.029	-0.041	0.080	-0.064
HD	-0.053	-0.177	-0.190	-0.038	-0.388	-0.929	0.429	0.027	0.019	0.042	0.018	-0.035	0.062	-0.073
IBM	0.081	-0.224	-0.261	0.101	-0.499	-1.037	0.489	0.000	0.008	0.045	0.022	-0.038	0.016	-0.030
INTC	-0.014	-0.127	-0.020	0.250	-0.624	-1.085	0.280	-0.004	0.012	0.015	0.028	-0.038	0.063	-0.046
JNJ	-0.138	-0.181	-0.184	-0.028	-0.407	-0.944	0.509	0.017	0.023	0.025	0.022	-0.041	0.062	-0.083
JPM	-0.187	-0.149	-0.130	-0.127	-0.317	-0.842	0.411	0.013	0.023	0.031	0.016	-0.034	0.068	-0.050
KO	-0.160	-0.162	-0.163	0.094	-0.497	-0.919	0.434	0.007	0.021	0.035	0.027	-0.035	0.053	-0.059
MCD	-0.035	-0.201	-0.206	-0.010	-0.423	-0.978	0.477	0.013	0.015	0.047	0.014	-0.033	0.035	-0.052
MMM	0.071	-0.216	-0.200	0.018	-0.442	-1.044	0.474	0.004	0.010	0.029	0.019	-0.035	0.032	-0.051
MRK	-0.152	-0.168	-0.147	0.090	-0.485	-0.913	0.447	0.011	0.019	0.017	0.027	-0.039	0.061	-0.067
MSFT	-0.010	-0.135	-0.032	0.201	-0.603	-1.113	0.291	-0.005	0.010	0.016	0.032	-0.048	0.085	-0.058
PFE	-0.263	-0.163	-0.085	0.298	-0.615	-0.882	0.458	0.004	0.014	0.018	0.025	-0.032	0.049	-0.056
PG	-0.095	-0.198	-0.199	-0.007	-0.431	-0.955	0.496	0.018	0.015	0.041	0.022	-0.039	0.038	-0.068
SPY	-0.401	-0.113	-0.005	-0.300	-0.150	-0.682	0.286	0.062	0.008	0.002	0.005	-0.025	0.136	-0.041
T	-0.276	-0.161	-0.086	0.229	-0.586	-0.892	0.474	0.017	0.026	0.020	0.033	-0.042	0.044	-0.069
UTX	0.121	-0.212	-0.189	0.094	-0.500	-1.076	0.465	0.008	0.009	0.029	0.025	-0.040	0.035	-0.050
VZ	-0.172	-0.160	-0.148	0.123	-0.518	-0.931	0.428	0.010	0.008	0.028	0.030	-0.046	0.044	-0.039
WMT	-0.105	-0.177	-0.181	0.015	-0.443	-0.941	0.450	0.014	0.019	0.028	0.025	-0.038	0.070	-0.071
XOM	-0.141	-0.199	-0.210	-0.170	-0.267	-0.812	0.514	-0.005	0.023	0.052	0.008	-0.027	0.078	-0.061
Overall	-0.080	-0.174	-0.146	0.067	-0.468	-0.948	0.419	0.015	0.016	0.029	0.023	-0.036	0.053	-0.059

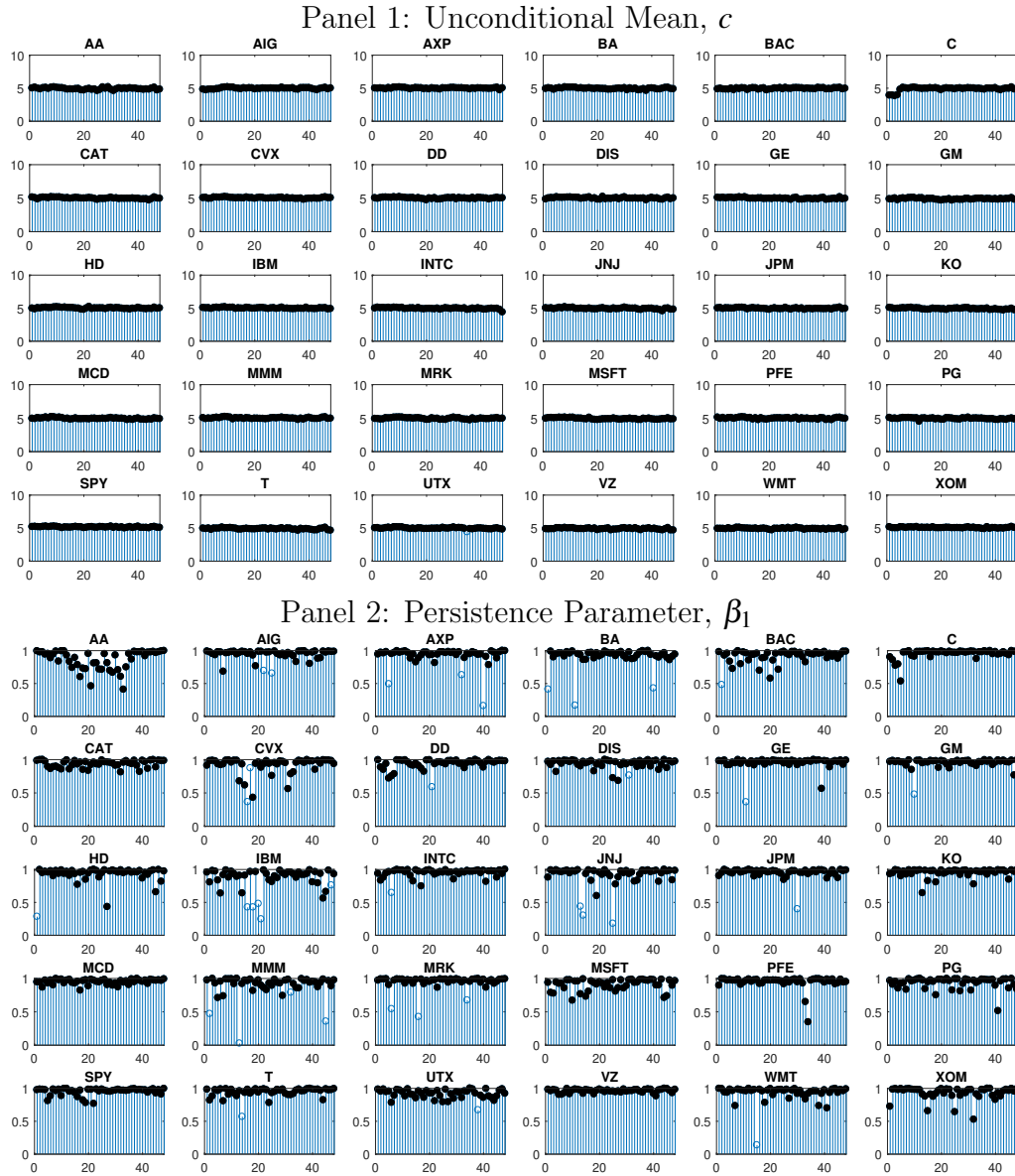
Note: The table presents average estimates of MMS parameter estimates for each stock. For each stock, we estimate the LL-ACD(1,1)-A model on a monthly basis to obtain 48 parameter estimates per each MMS variable, and report the average of the 48 parameter estimates per MMS variable per stock.

Figure B.1 Correlogram and histogram of log price durations for IBM, JPM and PFE



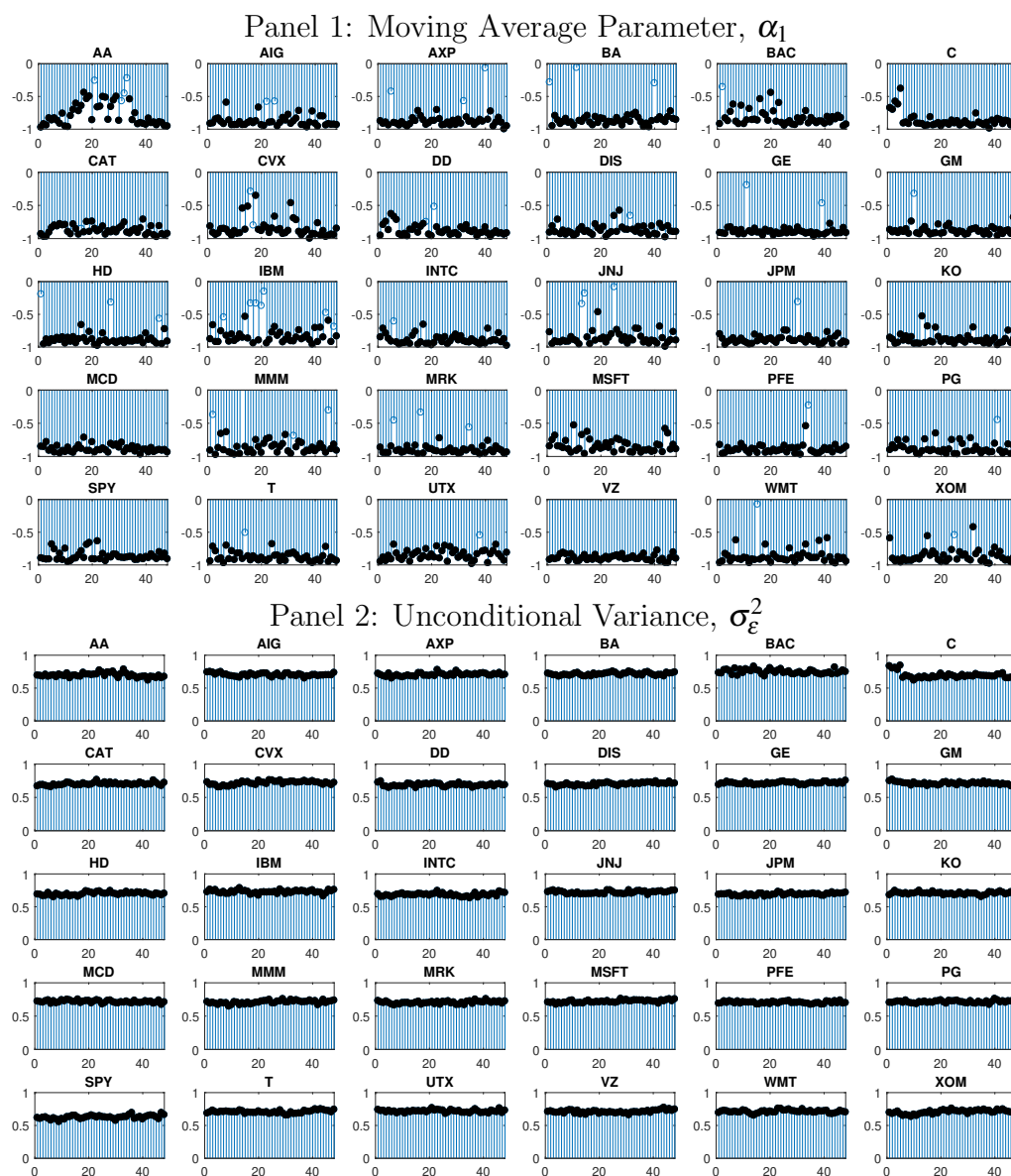
Note: The correlogram and histogram are constructed based on the log price duration of IBM, JPM and PFE for the complete sampling period. In the histogram, the red solid line represents a fitted normal density.

Figure B.2 Summary of estimated dynamic parameters of the LL-ACD(1,1)-K model for all stock-months, part 1



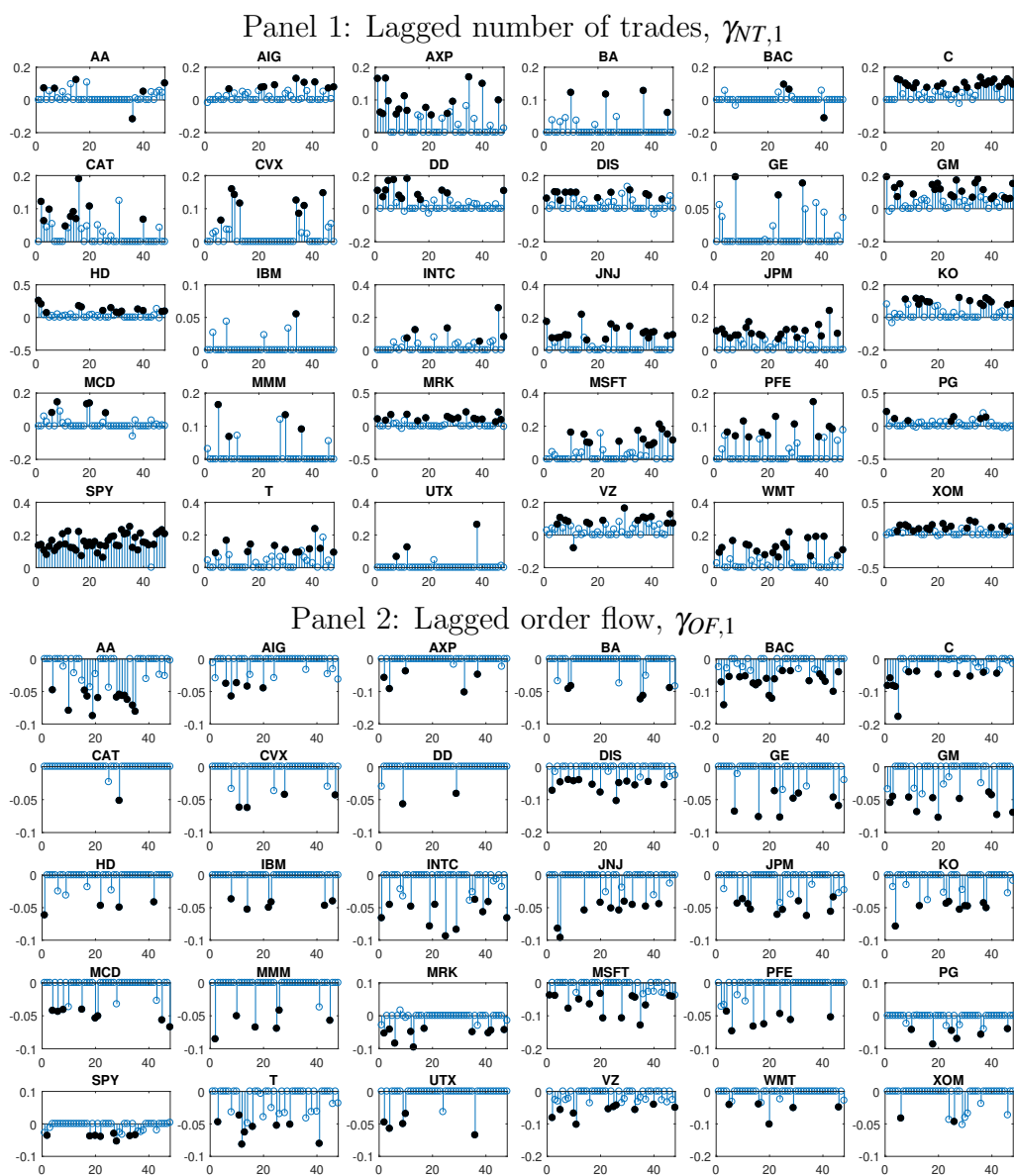
Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

Figure B.3 Summary of estimated dynamic parameters of the LL-ACD(1,1)-K model for all stock-months, part 2



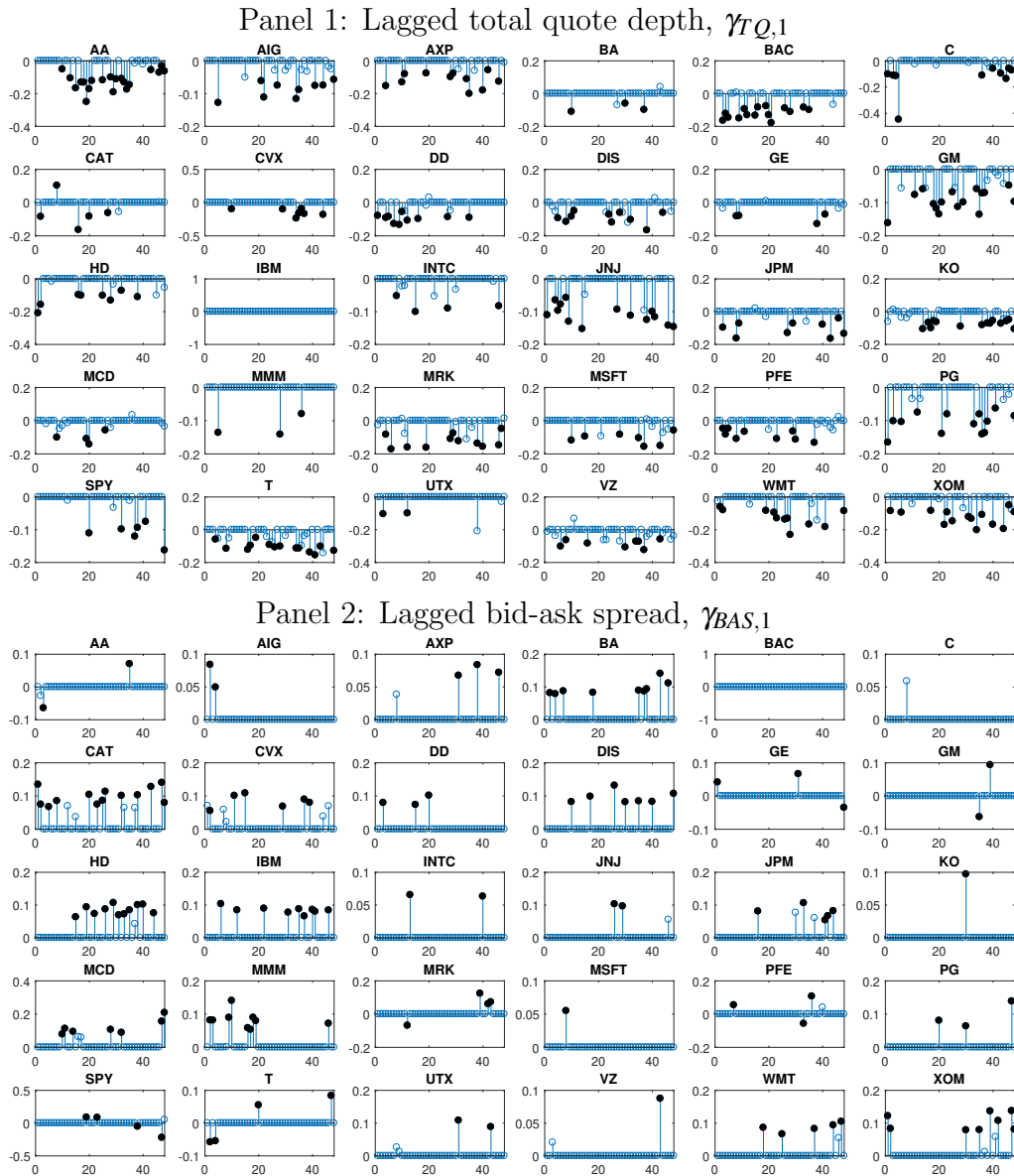
Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

Figure B.4 Summary of estimated one-duration lagged MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 1



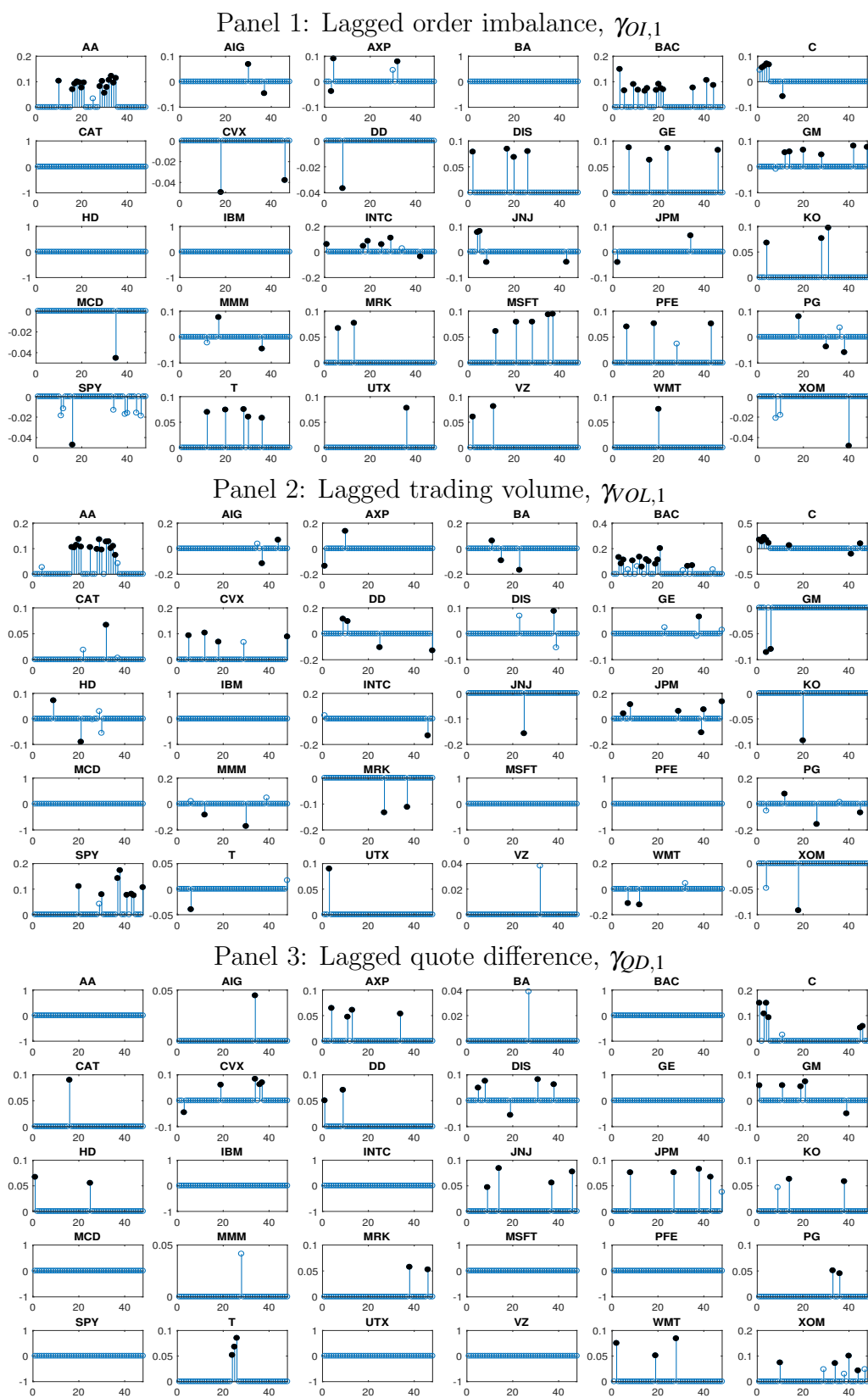
Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

Figure B.5 Summary of estimated one-duration lagged MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 2



Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

Figure B.6 Summary of estimated one-duration lagged MMS parameters of the LL-ACD(1,1)-K model for all stock-months, part 3



Note: The parameter estimates are obtained from monthly estimation of the LL-ACD(1,1)-K model. For each stock, parameter estimates are ordered chronologically with the x-axis representing the month index. Each circle denotes the value of a parameter estimate, with solid black circles highlighting significance at 5%.

B.3 Best Subset Selection Using Mixed Integer Optimization

This section outlines the optimization procedure of the best subset selection regression using Mixed Integer Optimization (MIO) as proposed by Bertsimas, King, and Mazumder (2016).

The best subset selection regression is the solution to the following nonconvex problem:

$$\min_{c, \gamma} \|\ln x_i^{(\delta)} - c + \gamma' Z_i\|_2^2 \quad \text{subject to} \quad \|\gamma_Z\|_0 \leq K, \quad (\text{B.7})$$

where $\|(\cdot)\|_2$ is the l_2 norm and $\|\gamma_Z\|_0$ denotes the pseudo-norm of γ_Z that counts the number of non-zero elements in γ_Z . Let the dimension of γ_Z be N-by-1, and denote $z = \{z_1, \dots, z_N\}$ an N-by-1 binary vector where each element takes value 1 if the corresponding element in γ_Z is non-zero, and zero otherwise. According to Bertsimas, King, and Mazumder (2016), this can be rewritten as a MIO problem as follows:

$$\begin{aligned} & \min_{c, \gamma, z} \|\ln x_i^{(\delta)} - c + \gamma' Z_i\|_2^2 \\ & \text{subject to} \quad -\mathcal{M}_U z \leq \gamma_Z \leq \mathcal{M}_U z \\ & \quad z_i \in \{0, 1\} \\ & \quad \sum_{i=1}^N z_i \leq K. \end{aligned} \quad (\text{B.8})$$

The constant \mathcal{M}_U should be taken to be a sufficiently large number (in magnitude larger than the largest element in the optimized γ_Z) to ensure that the bounds on γ_Z are valid. In our analysis we standardize all the MMS covariates and choose \mathcal{M}_U to be 5. This problem can then be optimized easily via MIO optimizers with very high efficiency, such as Gurobi, BARON, etc., even when the dimension of γ_Z is large.

B.4 Robustness Checks of the LL-ACD(1,1) Model Estimation

In this section, we present in-sample and out-of-sample diagnostic tests for the goodness-of-fit of LL-ACD(1,1)-P, LL-ACD(1,1)-K and LL-ACD(1,1)-A models. We will mainly focus on monthly estimated models and discuss the implications of expanding the estimation window.

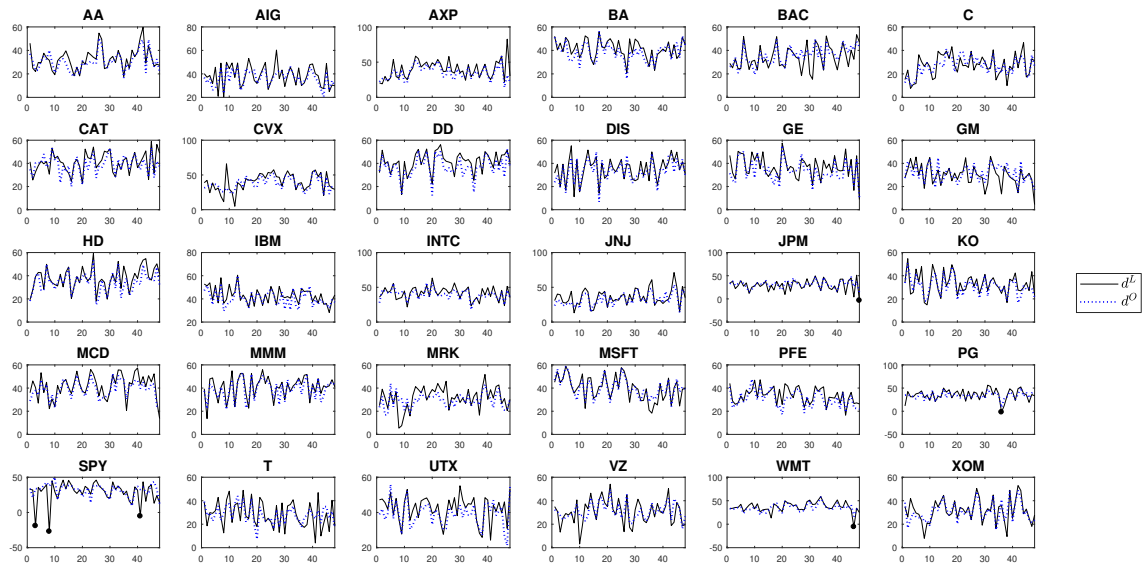
B.4.1 Gains in Information Criteria using BSR Selection

We firstly show that the MMS covariates selected by the K^* -optimal model also give smaller BIC for LL-ACD(1,1)-K model. To verify this, we firstly compute the difference in the BICs of OLS-A model and OLS-K model for each stock-month, and compare it with the same difference obtained by comparing LL-ACD(1,1)-A with -K model. Let us use the superscripts O and L to distinguish the BICs obtained from OLSs model and LL-ACD(1,1) models, we compute the following quantities for each stock month:

$$\begin{aligned} d^O &= BIC^{O,A} - BIC^{O,K}, \\ d^L &= BIC^{L,A} - BIC^{L,K}. \end{aligned} \tag{B.9}$$

We do not consider $BIC^{O,P}$ or $BIC^{L,P}$ in this comparison because they are far worse than the models with the inclusion of MMS covariates. The quantity d^O will be strictly non-negative for all stock-months, because the K^* -optimal models always choose the model with the minimum BIC, and $d^O = 0$ if and only if $K^* = 14$, which is unlikely from the results in Table 2.4. The quantity d^L , however, can be negative if the LL-ACD(1,1)-A model receives a smaller BIC, due to the interaction between the MMS covariates and the dynamic structure of the LL-ACD model. We plot d^O and d^L for each stock-month in Figure B.7.

Figure B.7 Comparison between d^L and d^O for all stock-months



Note: See (B.9) for the definition of d^O and d^L . For each stock, d^O and d^L are computed monthly and ordered chronologically with the x-axis representing the month index. Black dots indicate that $d^L < 0$ for a particular month.

Figure B.7 clearly shows that the d^L s are in general very similar to the d^O s and are mostly above zero for each stock-month. This suggests that the favourable BIC of the OLS-K model over the OLS-A model indeed carries over to the LL-ACD models. We observe 5 cases out of 1440 where $d^L < 0$ so that the LL-ACD(1,1)-A model actually has the smallest BIC. Also, we do not guarantee that LL-ACD(1,1)-K model possesses the smallest BIC over all combinations of MMS covariates for all stock-months. However, from Figure B.7 it is clear that the selected MMS covariates from BSR performs equally well in the LL-ACD(1,1)-K model compared with the OLS regressions. Therefore, obtaining the overall optimal combination of MMS covariates for the LL-ACD model is unlikely to have a significant improvement over this result, but could impose unnecessary computational burden on model estimation.

B.4.2 Residual Diagnostics

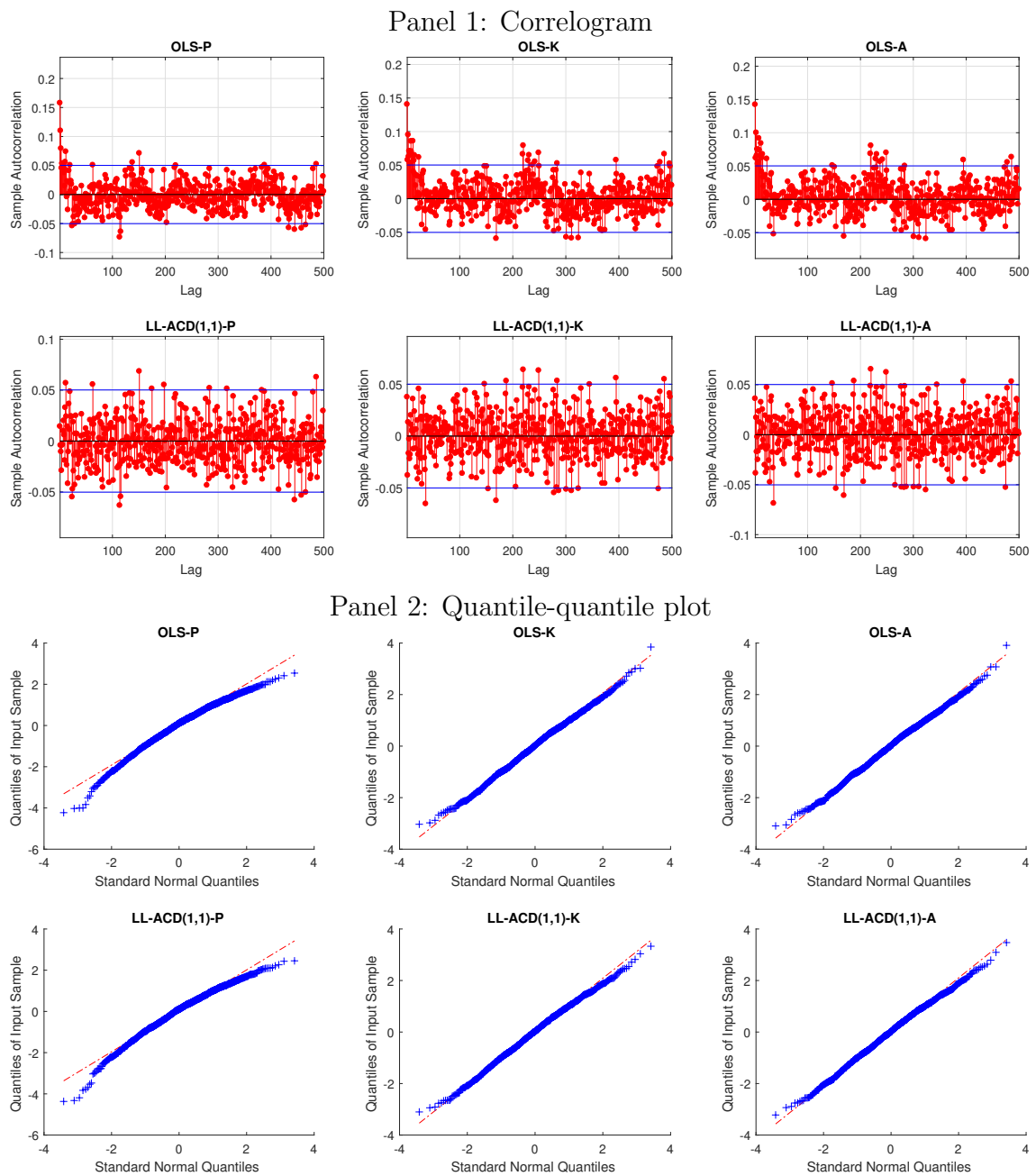
We proceed to compare the performances with the OLS models and the LL-ACD models. We examine whether the fitted residuals $\hat{\varepsilon}_i$ ($\hat{\Psi}_i$ in the OLS model) from each model in Table 2.5 satisfies the i.i.d. normality assumption of the LL-ACD model. We firstly present a graphical example of the goodness-of-fit of all models in Table 2.5 for 2011-01, SPY in Figure B.8.

Figure B.8 is consistent with the LB(20) and JB statistics in Table 2.5. From the correlograms, all models successfully capture the diurnal pattern in the raw price durations in Figure 2.3. For the OLS models, we see significant autocorrelation at small lags and around 220 lags, while the autocorrelation pattern in the residuals of the LL-ACD models are much less pronounced. From the Q-Q plots, it is evident that the inclusion of MMS covariates improves the goodness-of-fit of both OLS and LL-ACD models.

To give a comprehensive evaluation of the model specifications of all six models, we summarize diagnostic test results for all stock-months. We also compare the test results for the models estimated quarterly, half-yearly and yearly to analyse the impact of a larger estimation window. We consider the Ljung-Box test at lag 20 as an indicator of the goodness-of-fit of the ARMA(1,1) structure of the residual, and the Jacque-Bera test to test the normal assumption of the estimated residuals. For each test, we present the average test statistics and percentage of the rejections among all model estimates in Table B.10.¹

¹Total numbers of monthly, quarterly, half-yearly and yearly models for all securities estimations are 1440, 480, 240 and 120 respectively.

Figure B.8 Correlograms and quantile-quantile plots of the estimated residuals from model estimations in Table 2.5



Note: Title of each figure refers to the model where the residuals are estimated. For definition of the models, see Section 2.6.2. In the correlograms, the two blue horizontal lines are 95% confidence bounds. In the quantile-quantile (Q-Q) plot, we plot the standardized residuals ($\hat{\epsilon}_i/\hat{\sigma}_\epsilon$) against the quantile of a standard normal distribution.

Table B.10 Summary of diagnostic test results

Model	OLS			LL-ACD(1,1)		
Specification	-P	-K	-A	-P	-K	-A
Panel 1: Monthly Estimation						
Average LB(20)	217.91	135.50	124.13	18.88	20.70	20.72
% of Sig. at 10%	99.93%	99.10%	98.47%	8.33%	13.26%	12.71%
% of Sig. at 5%	99.93%	98.40%	97.85%	4.31%	7.92%	8.13%
% of Sig. at 1%	99.79%	97.01%	96.04%	1.11%	3.33%	2.99%
Average JB	122.18	8.86	8.81	95.15	7.82	7.85
% of Sig. at 10%	100.00%	40.69%	39.51%	99.79%	38.61%	38.13%
% of Sig. at 5%	100.00%	27.99%	27.92%	99.72%	26.04%	26.88%
% of Sig. at 1%	100.00%	10.35%	10.42%	99.65%	9.72%	9.58%
Panel 2: Quarterly Estimation						
Average LB(20)	682.47	473.10	460.60	20.30	26.59	26.82
% of Sig. at 10%	100.00%	100.00%	100.00%	12.29%	36.67%	38.13%
% of Sig. at 5%	100.00%	100.00%	100.00%	7.08%	27.92%	28.75%
% of Sig. at 1%	100.00%	100.00%	100.00%	2.50%	11.88%	12.29%
Average JB	402.29	20.29	20.08	283.00	18.03	17.84
% of Sig. at 10%	100.00%	68.33%	68.33%	100.00%	70.83%	69.79%
% of Sig. at 5%	100.00%	60.00%	58.33%	99.79%	58.75%	58.96%
% of Sig. at 1%	100.00%	37.50%	35.63%	99.79%	38.75%	38.13%
Panel 3: Half-Yearly Estimation						
Average LB(20)	1378.77	1039.36	1026.61	21.78	38.64	38.84
% of Sig. at 10%	100.00%	100.00%	100.00%	19.58%	74.58%	75.83%
% of Sig. at 5%	100.00%	100.00%	100.00%	11.25%	63.33%	65.42%
% of Sig. at 1%	100.00%	100.00%	100.00%	2.50%	42.50%	43.33%
Average JB	824.05	29.61	29.48	569.52	27.06	26.98
% of Sig. at 10%	100.00%	83.33%	83.33%	100.00%	86.67%	84.58%
% of Sig. at 5%	100.00%	78.75%	77.50%	100.00%	80.42%	77.50%
% of Sig. at 1%	100.00%	63.75%	63.75%	100.00%	64.17%	62.92%
Panel 4: Yearly Estimation						
Average LB(20)	2893.67	2256.54	2243.83	25.31	66.07	66.15
% of Sig. at 10%	100.00%	100.00%	100.00%	34.17%	96.67%	96.67%
% of Sig. at 5%	100.00%	100.00%	100.00%	24.17%	94.17%	94.17%
% of Sig. at 1%	100.00%	100.00%	100.00%	9.17%	89.17%	89.17%
Average JB	1699.74	59.91	59.60	1148.67	60.24	59.56
% of Sig. at 10%	100.00%	89.17%	89.17%	100.00%	93.33%	93.33%
% of Sig. at 5%	100.00%	89.17%	88.33%	100.00%	92.50%	93.33%
% of Sig. at 1%	100.00%	85.83%	85.00%	100.00%	87.50%	86.67%

Note: LB(20) and JB stand for the Ljung-Box test statistic at lag 20 and the Jacque-Bera test statistic respectively. In panel 1 to 4, we report the LB(20) and JB statistics averaged over 1440, 480, 240 and 120 model estimations with monthly, quarterly, half-yearly and yearly estimation window respectively. For each test, % of Sig. at $\alpha\%$ reports the percentage of tests rejected at $\alpha\%$ significance level.

Firstly we compare the performance of all six models within each panel in Table B.10. The LB(20) statistics show that for the same model specification, the LL-ACD models always outperform the OLS models significantly, which is not surprising. LB(20) for the -K and -A specifications are better than the -P specification for the OLS models, but it is the opposite case for the LL-ACD model. This is because the MMS covariates can to some extent explain the dependence structure in the log price durations, but it is a very noisy measure. Once we control for the ARMA effect in the residuals, the inclusion of MMS covariates only adds noise to the dynamic structure, which results in larger LB(20) statistics. This effect is more pronounced when the estimation window expands. As to the JB statistics, we can confirm that the inclusion of the MMS covariates greatly improves the fit of the normal density.

Comparing across different panels in Table B.10, the goodness-of-fit of both the OLS and LL-ACD models generally deteriorate as the estimation window expands. However, we can see that the -K and -A models are more robust to changes in the size of the estimation window. This is more obvious in terms of the Jacque-Bera statistics. The average JB statistics increases from 95 to over 1000 for the LL-ACD(1,1)-P model, while for the -K and -A models the increase is much less pronounced. The LB(20) statistics from the LL-ACD(1,1)-P generally outperform the LL-ACD(1,1)-K and -A model in both average test statistics and rejection rate, and all the LL-ACD models significantly outperform their OLS counterparts in terms of the LB(20) statistics.

Our findings in Table B.10 clearly indicate that the LL-ACD(1,1) structure to a large extent captures the dependence structure in the price durations. The inclusion of the MMS covariates can greatly improve the fit of the normal density, but will add noise to the dependence structure of the residuals. The overall performance of the model specifications decreases as the estimation window widens, and monthly estimated models have the best performance in terms of both average test statistics and percentage of test rejections.

B.4.3 Goodness-of-Fit

We assess the overall goodness-of-fit of the models using the adjusted R^2 s of the log price duration models. For each security, we present the average adjusted R^2 s for each model with four different sizes of estimation windows in Table B.11. Table B.11 corroborates our findings in Table B.10 that for a given size of estimation window, all LL-ACD models outperform their OLS counterparts. The inclusion of MMS covariates improves the adjusted R-squares by more than 30% in all cases. All the adjusted R-squares decrease as the estimation window widens, except for

the LL-ACD(1,1)-P models where the adjusted R-squares actually increase slightly. However, the changes are all very small (typically less than 1%) and may not be statistically significant.

Table B.11 Average adjusted R-squared for models with varying estimation windows

Model	OLS			LL-ACD(1,1)		
Specification	-P	-K	-A	-P	-K	-A
$m = 1$	36.3456%	71.4521%	71.5967%	41.1081%	72.6787%	72.7905%
$m = 3$	36.0217%	70.9915%	71.0374%	41.2141%	72.6306%	72.6753%
$m = 6$	35.9057%	70.6796%	70.6988%	41.2376%	72.5579%	72.5807%
$m = 12$	35.7544%	70.4041%	70.4129%	41.2658%	72.4722%	72.4848%

Note: m denotes the size of the estimation window in months. For each model specification, we report the adjusted R-squared averaged over 1440, 480, 240 and 120 model estimations with monthly, quarterly, half-yearly and yearly estimation window respectively.

To formally assess whether the improvements of the ACD models over the OLS models are significant, and whether the MMS covariates contribute significantly to the likelihood of the model, we perform likelihood ratio tests for the OLS models versus LL-ACD models with the same specification, the -P models versus -K models and -K models versus -A models. For each estimation window, this amounts to seven likelihood ratio tests. We present the results in Table B.12. Our findings suggest that the autoregressive structure contributes significantly to the log-likelihood for all model estimations as expected, the -K models significantly outperform the -P models with a substantial average LR statistics and 100% rejection rate at 1%. The -A models still seem to provide some improvements to the likelihood, but this effect is much weaker.

To sum up, in this section we firstly demonstrate that the optimal MMS covariates selection via the OLS models also improve BICs for the LL-ACD(1,1)-K models over the -P models and the -A models. In-sample diagnostics and goodness-of-fit tests suggest that the inclusion of MMS covariates can significantly improve the performance of the models in terms of residual diagnostics and model likelihoods, and the -K models perform very close to the -A models. Analysis based on the estimation window shows that the goodness-of-fit of the LL-ACD models generally decreases as the estimation window expands, but the -K and the -A models are more robust to changes in the estimation window compared to the -P models.

Table B.12 Summary of likelihood ratio test results

Unrestricted Restricted	OLS-K OLS-P	OLS-A OLS-K	LL-ACD-K LL-ACD-P	LL-ACD-A LL-ACD-K	LL-ACD-P OLS-P	LL-ACD-K OLS-K	LL-ACD-A OLS-A
Panel 1: Monthly Estimation							
Average LR	1294.16	14.97	1240.77	13.44	127.44	74.05	72.52
% of Sig. at 10%	100.00%	68.40%	100.00%	54.72%	100.00%	99.72%	99.79%
% of Sig. at 5%	100.00%	53.19%	100.00%	41.53%	100.00%	99.65%	99.72%
% of Sig. at 1%	100.00%	27.71%	100.00%	20.69%	100.00%	99.03%	99.24%
Panel 2: Quarterly Estimation							
Average LR	3813.76	12.07	3688.23	12.29	404.48	278.95	279.18
% of Sig. at 10%	100.00%	68.54%	100.00%	57.08%	100.00%	100.00%	100.00%
% of Sig. at 5%	100.00%	56.25%	100.00%	45.21%	100.00%	100.00%	100.00%
% of Sig. at 1%	100.00%	30.42%	100.00%	27.50%	100.00%	100.00%	100.00%
Panel 3: Half-Yearly Estimation							
Average LR	7526.44	9.67	7335.12	11.31	823.31	632.00	633.63
% of Sig. at 10%	100.00%	68.75%	100.00%	60.83%	100.00%	100.00%	100.00%
% of Sig. at 5%	100.00%	51.67%	100.00%	52.08%	100.00%	100.00%	100.00%
% of Sig. at 1%	100.00%	31.67%	100.00%	32.08%	100.00%	100.00%	100.00%
Panel 4: Yearly Estimation							
Average LR	14897.45	8.18	14586.07	11.27	1695.26	1383.88	1386.97
% of Sig. at 10%	100.00%	60.83%	100.00%	66.67%	100.00%	100.00%	100.00%
% of Sig. at 5%	100.00%	56.67%	100.00%	54.17%	100.00%	100.00%	100.00%
% of Sig. at 1%	100.00%	35.00%	100.00%	36.67%	100.00%	100.00%	100.00%

Note: LR stand for the likelihood ratio test statistics computed as two times the difference in the estimated log-likelihoods from the unrestricted and the restricted models. In panel 1 to 4, we report LR statistics averaged over 1440, 480, 240 and 120 model estimations with monthly, quarterly, half-yearly and yearly estimation window respectively. For each test, % of Sig. at $\alpha\%$ reports the percentage of tests rejected at $\alpha\%$ significance level.

B.5 Daily Volatility Estimation with the LL-ACD Model

In this section, we construct daily volatility estimates from the LL-ACD models and compare them against a RV-type benchmark. The purpose of this section is to show that the ICV estimator estimates the same object as the RV-type estimator, and the inclusion of MMS covariates improves the performance of the ICV estimator by reducing the deviation from the benchmark. On a daily level, the RV-type estimators can be constructed relatively precisely due to the large amount of daily transactions (the least liquid stock AA in our sample on average has more than 1500 transactions per day), which justifies our daily comparison between the ICV estimators.

Daily ICV estimates can therefore be constructed based on $ICV_{i,d}^{M,m}$ as:

$$ICV_d^{M,m} \equiv \sum_{i \in I_d} ICV_{i,d}^{M,m} + \frac{1}{6} \left(\frac{\delta}{P(t_{I_d,d}^{(\delta)})} \right)^2. \quad (\text{B.10})$$

Note that the term $\frac{1}{6} \left(\frac{\delta}{P(t_{I_d,d}^{(\delta)})} \right)^2$ is added to correct for the end-of-sample bias as documented in Li, Nolte, and Nolte (2018a).

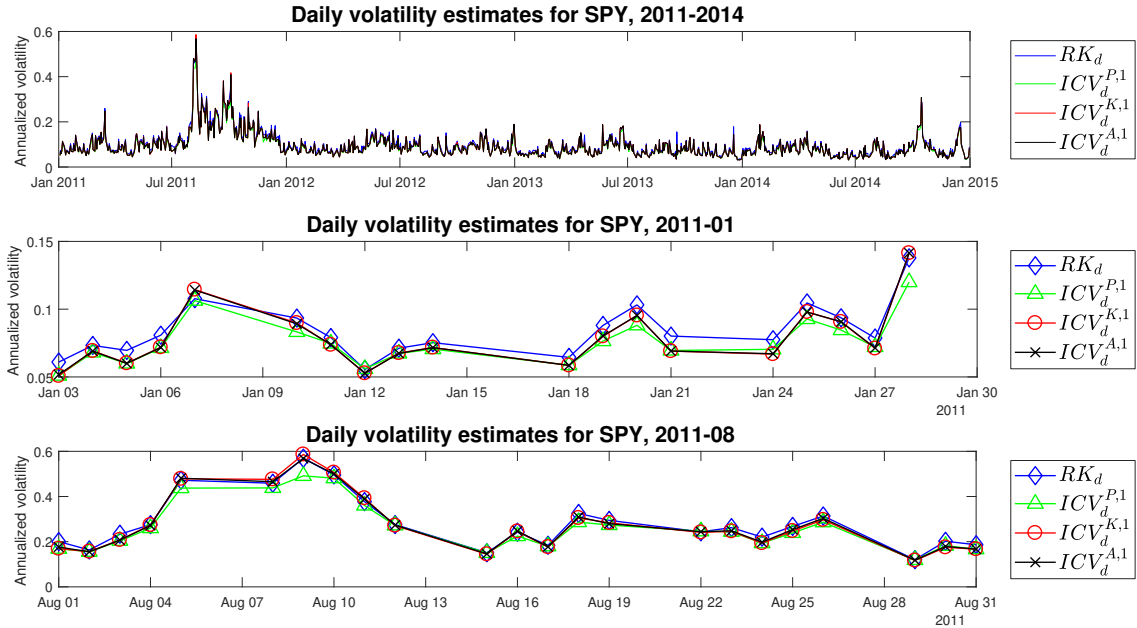
In order to evaluate the ICV volatility estimates, we compare our daily ICV estimates to a benchmark RV-type volatility estimator: the realized kernel (RK) estimates², which is a widely applied estimator that is robust to MMS noise and capable of incorporating all observations to compute volatility estimates (Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008b). Following Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), we use a non flat-top Parzen kernel with optimal bandwidth selection and tick-by-tick sampling. To give a graphical illustration of the RK and ICV estimates, we plot examples of monthly ICV estimates and RK estimates from SPY in Figure B.9.

In Figure B.9 we present a line plot for the whole sampling period in the upper panel and two zoomed-in plots for the two months 2011-01 and 2011-08, which correspond to a calmer period and a more volatile period³. Our first observation is that the four estimators provide very similar volatility estimates, as can be seen from

²As a convention, we compute all volatility measures as annualized square root of the variance measures. Daily ICV and RK refer to their daily annualized volatility measures instead of the daily variance measures whenever no confusion is caused.

³The high volatility estimates of August 2011 is believed to be a result from the European sovereign debt crisis and the downgrade of U.S. credit rating.

Figure B.9 Example of daily volatility estimates based on ICV and RK for SPY



Note: The x-axis denotes the date on which the daily volatility estimates are computed. We report annualized volatility measures instead of estimated variances. The middle and lower panels are zoomed-in versions of the upper panel for 2011-01 and 2011-08.

the upper panel. In the zoomed-in plots, we see that the RK estimates are always slightly above the ICV estimates. This is due to the truncation feature of the ICV estimators that regards any price changes of larger than δ as δ . This truncation provides robustness to jumps in the price process, but also introduces a truncation bias that converges to zero as the sampling frequency decreases. From the zoomed-in plots, it is evident that despite that the truncation threshold δ is shared among all three ICV estimators, $ICV_d^{K,1}$ and $ICV_d^{A,1}$ seem to be less affected by the truncation bias and are in closer proximity to the RK_d estimates than the $ICV_d^{P,1}$ estimates.

To assess how close the four volatility estimates are related to each other, we present the average correlation table between the four volatility estimators in Table B.13, which shows clearly that the four estimators are indeed highly correlated and $ICV_d^{K,1}$ and $ICV_d^{A,1}$ have higher correlation with RK_d compared to $ICV_d^{P,1}$. Also, ICV estimates with the inclusion of MMS covariates are extremely highly correlated. This can be a sign that $ICV_d^{K,1}$ has already incorporated most of the information through the MMS covariates, and $ICV_d^{A,1}$ does not add much additional information to the volatility estimates. Results from Figure B.9 and Table B.13 provide strong empirical evidence supporting the findings in Li, Nolte, and Nolte (2018a) that ICV estimators indeed estimate the same quantity as RV-type estimators.

Table B.13 Averaged correlation table of RK and monthly estimated daily ICV volatility estimates

	RK_d	$ICV_d^{P,1}$	$ICV_d^{K,1}$	$ICV_d^{A,1}$
RK_d	1.0000			
$ICV_d^{P,1}$	0.9705	1.0000		
$ICV_d^{K,1}$	0.9721	0.9904	1.0000	
$ICV_d^{A,1}$	0.9723	0.9906	0.9997	1.0000

Note: for each security, we compute the correlation matrix of the matrix $\{RK_d, ICV_d^{P,1}, ICV_d^{K,1}, ICV_d^{A,1}\}_{d=1:1006}$, and average the correlation matrix over 30 securities.

We compute the mean squared error (MSE) for each ICV estimate with respect to the RK measure, and test the difference using the modified Dieblod-Mariano test in Table B.14.

In Table B.14, we clearly observe that $MSE^{K,m}$ outperforms $MSE^{P,m}$ for all securities and m . This indicates that the inclusion of MMS covariates can significantly improve the performance of the $ICV^{P,m}$ estimator. We also see that $MSE^{K,m}$ and $MSE^{A,m}$ are close to each other both numerically and statistically, especially when m is large. This result is partly driven by the fact that when m is large, the K^* -optimal models tend to select more MMS covariates and the difference between LL-ACD(1,1)-K and LL-ACD(1,1)-All is small. However, even with a small m , we can safely conclude that including additional MMS covariates in excess of the covariates selected in the LL-ACD(1,1)-K models do not have a substantial impact on the performance of the $ICV^{K,m}$ estimator, as the majority of the test results are insignificant.

Table B.14 also shows some interesting pattern when we compare the MSEs across different sizes of estimation windows. Comparing $MSE^{P,1}$ with $MSE^{MP,3}$, $MSE^{P,6}$ and $MSE^{P,12}$, we see that $MSE^{P,m}$ is increasing as the estimation window expands. We therefore also compare the performance of the ICV estimators across estimation windows by producing a re-arranged version of Table B.14 and test whether $MSE^{M,1} = MSE^{M,3}$, $MSE^{M,3} = MSE^{M,6}$ and $MSE^{M,6} = MSE^{M,12}$ for each M . We present the results in Table B.15. From Table B.15 we can clearly observe that $MSE^{P,m}$ indeed deteriorates significantly as the estimation window extends, whereas results of $MSE^{K,m}$ and $MSE^{A,m}$ are much more robust to changes in the estimation window. We only see a decline of the performance of $ICV^{K,m}$ and $ICV^{A,m}$ when we use a one-year estimation window. In fact we even see some significant improvements after expanding the estimation window.

Table B.14 Comparison of MSEs of ICV volatility estimates within each estimation window

Ticker	$MSE^{P,1}$	$MSE^{K,1}$	$MSE^{A,1}$	$MSE^{P,3}$	$MSE^{K,3}$	$MSE^{A,3}$	$MSE^{P,6}$	$MSE^{K,6}$	$MSE^{A,6}$	$MSE^{P,12}$	$MSE^{K,12}$	$MSE^{A,12}$
AA	3.650	3.362***	3.360	3.661	3.377***	3.380	3.678	3.386***	3.389 ^o	3.685	3.398***	3.401 ^{oo}
AIG	1.589	1.356***	1.355	1.609	1.356***	1.361	1.616	1.367***	1.368	1.644	1.392***	1.392
AXP	0.782	0.613***	0.615	0.790	0.607***	0.607	0.791	0.612***	0.612	0.797	0.614***	0.614
BA	0.974	0.804***	0.801	0.987	0.804***	0.803	0.985	0.798***	0.800	0.990	0.803***	0.802
BAC	2.178	1.862***	1.862	2.212	1.841***	1.844	2.208	1.878***	1.878	2.215	1.910***	1.909
C	2.432	1.663***	1.666	2.467	1.667***	1.667	2.447	1.588***	1.590 ^{oo}	2.717	1.783***	1.787 ^{oo}
CAT	0.846	0.722***	0.713**	0.852	0.706***	0.705*	0.854	0.702***	0.702	0.856	0.692***	0.692
CVX	0.425	0.341***	0.340	0.430	0.343***	0.341**	0.429	0.337***	0.337	0.435	0.337***	0.337
DD	0.785	0.637***	0.635	0.794	0.655***	0.654	0.793	0.664***	0.664	0.798	0.671***	0.671
DIS	0.661	0.545***	0.544	0.673	0.557***	0.556	0.672	0.569***	0.569	0.677	0.575***	0.574
GE	1.023	0.880***	0.875*	1.032	0.877***	0.876	1.033	0.882***	0.882	1.041	0.888***	0.888
GM	1.686	1.346***	1.343	1.689	1.364***	1.359***	1.688	1.361***	1.360*	1.691	1.392***	1.392
HD	0.697	0.607***	0.602**	0.699	0.605***	0.605	0.703	0.609***	0.608	0.707	0.611***	0.611
IBM	0.476	0.412***	0.412	0.480	0.413***	0.412	0.482	0.421***	0.421	0.484	0.431**	0.431
INTC	1.607	1.391***	1.383**	1.617	1.398***	1.397	1.623	1.401***	1.400	1.630	1.417***	1.417
JNJ	0.327	0.257***	0.258	0.328	0.255***	0.255	0.331	0.251***	0.251 ^{oo}	0.333	0.253***	0.253 ^{oo}
JPM	0.729	0.573***	0.572	0.730	0.576***	0.577	0.729	0.572***	0.572	0.732	0.574***	0.574 ^{oo}
KO	0.411	0.335***	0.333	0.415	0.335***	0.336	0.418	0.339***	0.340	0.420	0.336***	0.336
MCD	0.366	0.292***	0.291	0.367	0.288***	0.288	0.368	0.287***	0.287	0.370	0.285***	0.285
MMM	0.484	0.415***	0.415	0.485	0.414***	0.413	0.487	0.415***	0.415	0.493	0.409***	0.409
MRK	0.560	0.472***	0.468	0.563	0.479***	0.480	0.564	0.481***	0.481	0.568	0.484***	0.485 ^{oo}
MSFT	1.037	0.945***	0.935***	1.041	0.946***	0.946	1.043	0.956***	0.955*	1.048	0.969***	0.966***
PFE	1.002	0.855***	0.853	1.010	0.857***	0.857	1.019	0.864***	0.864	1.020	0.862***	0.862
PG	0.388	0.316***	0.313*	0.392	0.313***	0.313	0.395	0.315***	0.315	0.398	0.314***	0.314
SPY	0.140	0.081***	0.080	0.144	0.080***	0.078***	0.145	0.077***	0.077*	0.146	0.078***	0.078*
T	0.525	0.436***	0.435	0.531	0.436***	0.435	0.537	0.434***	0.434	0.540	0.444***	0.444
UTX	0.769	0.654***	0.650*	0.778	0.650***	0.649	0.778	0.654***	0.654	0.781	0.647***	0.646**
VZ	0.607	0.496***	0.494	0.612	0.503***	0.502	0.615	0.496***	0.496	0.617	0.506***	0.506
WMT	0.460	0.383***	0.383	0.465	0.386***	0.386	0.467	0.388***	0.389	0.467	0.389***	0.389
XOM	0.306	0.239***	0.237	0.307	0.236***	0.236	0.307	0.236***	0.236	0.310	0.244**	0.244**
Overall	0.931	0.776***	0.774***	0.939	0.778***	0.777	0.940	0.778***	0.778	0.954	0.790***	0.790

Note: The MSEs are multiplied by 10^3 . For each security, we compute $MSE^{M,m}$ for each $M \in \{P, K, A\}$ and $m \in \{1, 3, 6, 12\}$ according to (2.23). For each m , we perform modified Diebold-Mariano tests to test whether $MSE^{P,m} = MSE^{K,m}$ and $MSE^{K,m} = MSE^{A,m}$. The asterisks (circles) superscripts on the MSEs indicate significantly smaller (larger) MSEs compared against the MSE to the left. One to three symbols correspond to significance at 10%, 5% and 1%, respectively.

Table B.15 Comparison of MSEs of ICV volatility estimates across different estimation windows

Ticker	$MSE^{P,1}$	$MSE^{P,3}$	$MSE^{P,6}$	$MSE^{P,12}$	$MSE^{K,1}$	$MSE^{K,3}$	$MSE^{K,6}$	$MSE^{K,12}$	$MSE^{A,1}$	$MSE^{A,3}$	$MSE^{A,6}$	$MSE^{A,12}$
AA	3.650	3.661	3.678 ^{oo}	3.685	3.362	3.377	3.386	3.398	3.360	3.380	3.389	3.401
AIG	1.589	1.609 ^{ooo}	1.616 ^o	1.644 ^{ooo}	1.356	1.356	1.367	1.392 ^{oo}	1.355	1.361	1.368	1.392 ^{oo}
AXP	0.782	0.790 ^{oo}	0.791	0.797 ^o	0.613	0.607	0.612	0.614	0.615	0.607	0.612	0.614
BA	0.974	0.987 ^o	0.985	0.990 ^{oo}	0.804	0.804	0.798	0.803	0.801	0.803	0.800	0.802
BAC	2.178	2.212 ^{oo}	2.208	2.215 ^o	1.862	1.841	1.878 ^o	1.910 ^{oo}	1.862	1.844	1.878 ^o	1.909 ^{oo}
C	2.432	2.467	2.447	2.717 ^{ooo}	1.663	1.667	1.588 ^{***}	1.783 ^{ooo}	1.666	1.667	1.590 ^{***}	1.787 ^{ooo}
CAT	0.846	0.852 ^o	0.854	0.856	0.722	0.706	0.702	0.692 ^{***}	0.713	0.705	0.702	0.692 ^{***}
CVX	0.425	0.430 ^o	0.429	0.435 ^o	0.341	0.343	0.337 [*]	0.337	0.340	0.341	0.337	0.337
DD	0.785	0.794	0.793	0.798 ^{oo}	0.637	0.655	0.664	0.671	0.635	0.654	0.664	0.671
DIS	0.661	0.673 ^{oo}	0.672	0.677 ^{oo}	0.545	0.557	0.569 ^{oo}	0.575	0.544	0.556	0.569 ^{oo}	0.574
GE	1.023	1.032 ^{oo}	1.033	1.041 ^{ooo}	0.880	0.877	0.882	0.888	0.875	0.876	0.882	0.888
GM	1.686	1.689	1.688	1.691 ^o	1.346	1.364	1.361	1.392 ^{ooo}	1.343	1.359	1.360	1.392 ^{ooo}
HD	0.697	0.699	0.703 ^o	0.707 ^{oo}	0.607	0.605	0.609	0.611	0.602	0.605	0.608	0.611
IBM	0.476	0.480 ^{ooo}	0.482	0.484	0.412	0.413	0.421 ^o	0.431	0.412	0.412	0.421 ^o	0.431
INTC	1.607	1.617 ^o	1.623 ^{oo}	1.630 ^{ooo}	1.391	1.398	1.401	1.417 ^{oo}	1.383	1.397 ^o	1.400	1.417 ^{oo}
JNJ	0.327	0.328	0.331 ^{ooo}	0.333 ^{oo}	0.257	0.255	0.251 [*]	0.253	0.258	0.255	0.251	0.253
JPM	0.729	0.730	0.729	0.732	0.573	0.576	0.572	0.574	0.572	0.577	0.572	0.574
KO	0.411	0.415 ^{oo}	0.418 ^o	0.420 ^{ooo}	0.335	0.335	0.339	0.336	0.333	0.336	0.340	0.336
MCD	0.366	0.367	0.368 ^o	0.370 ^o	0.292	0.288 [*]	0.287	0.285	0.291	0.288	0.287	0.285
MMM	0.484	0.485	0.487	0.493 ^{ooo}	0.415	0.414	0.415	0.409 ^{**}	0.415	0.413	0.415	0.409 ^{**}
MRK	0.560	0.563	0.564	0.568 ^{ooo}	0.472	0.479	0.481	0.484	0.468	0.480 ^{oo}	0.481	0.485
MSFT	1.037	1.041	1.043	1.048 ^{ooo}	0.945	0.946	0.956	0.969 ^o	0.935	0.946	0.955	0.966 ^o
PFE	1.002	1.010 ^o	1.019 ^{oo}	1.020	0.855	0.857	0.864 ^o	0.862	0.853	0.857	0.864 ^o	0.862
PG	0.388	0.392 ^{oo}	0.395 ^{ooo}	0.398 ^{oo}	0.316	0.313	0.315	0.314	0.313	0.313	0.315	0.314
SPY	0.140	0.144 ^{ooo}	0.145	0.146 ^o	0.081	0.080	0.077	0.078 ^o	0.080	0.078	0.077	0.078 ^o
T	0.525	0.531 ^o	0.537	0.540 ^{ooo}	0.436	0.436	0.434	0.444 ^{ooo}	0.435	0.435	0.434	0.444 ^{ooo}
UTX	0.769	0.778 ^{ooo}	0.778	0.781 ^{oo}	0.654	0.650	0.654	0.647 ^{**}	0.650	0.649	0.654	0.646 ^{**}
VZ	0.607	0.612 ^{oo}	0.615 ^{oo}	0.617	0.496	0.503	0.496 [*]	0.506 ^{ooo}	0.494	0.502 ^o	0.496	0.506 ^{ooo}
WMT	0.460	0.465	0.467 ^{oo}	0.467	0.383	0.386	0.388	0.389	0.383	0.386	0.389	0.389
XOM	0.306	0.307	0.307	0.310	0.239	0.236	0.236	0.244 ^{ooo}	0.237	0.236	0.236	0.244 ^{ooo}
Overall	0.931	0.939 ^{oo}	0.940	0.954 ^{oo}	0.776	0.778	0.778	0.790 ^{oo}	0.774	0.777	0.778	0.790 ^{oo}

Note: The MSEs are multiplied by 10^3 . For each security, we compute $MSE^{M,m}$ for each $M \in \{P, K, A\}$ and $m \in \{1, 3, 6, 12\}$ according to (2.23). For each M , we perform modified Diebold-Mariano tests to test whether $MSE^{M,1} = MSE^{M,3}$, $MSE^{M,3} = MSE^{M,6}$ and $MSE^{M,6} = MSE^{M,12}$. The asterisks (circles) superscripts on the MSEs indicate significantly smaller (larger) MSEs compared against the MSE to the left. One to three symbols correspond to significance at 10%, 5% and 1%, respectively.

The performance of the ICV estimators summarized in Tables B.14 and B.15 is closely linked to the goodness-of-fit of the corresponding LL-ACD(1,1) model specifications. The superior performance of $ICV^{K,m}$ and $ICV^{A,m}$ over $ICV^{P,m}$ is not surprising because the LL-ACD(1,1)-K and -A models have significantly better in-sample goodness-of-fit in comparison with the -P models. This also explains the similar performance of $ICV^{K,m}$ and $ICV^{A,m}$ and the robustness of these two estimators to changes in the estimation window. As the -A models do not provide substantial improvement to the goodness-of-fit compared to the -K model, we recommend to use the -K models to construct daily volatility estimates to improve the efficiency of volatility estimates.

B.6 Construction of the Intraday ICV Estimator

We present the full specification of the $ICV_{j,d}^{M,m}$ estimator for the interval $(3900(j-1), 3900j]$ as in (2.22):

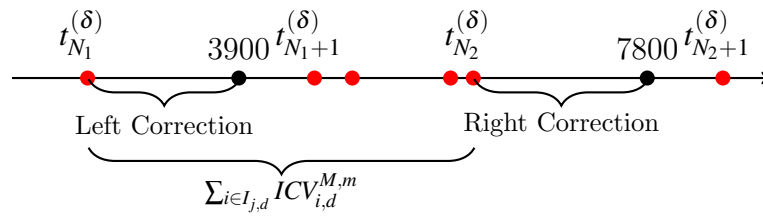
$$ICV_{j,d}^{M,m} \equiv \sum_{i \in I_{j,d}} ICV_{i,d}^{M,m} + Q_{j,d}, \quad (\text{B.11})$$

$$Q_{j,d} = \underbrace{-ICV_{N_{j-1}+1,d}^{M,m} \frac{3900(j-1) - t_{N_{j-1},d}^{(\delta)}}{x_{N_{j-1}+1,d}^{(\delta)}}}_{\text{Left Correction}} + \underbrace{ICV_{N_j+1,d}^{M,m} \frac{t_{N_j}^{(\delta)} - 3900j}{x_{N_j+1,d}^{(\delta)}}}_{\text{Right Correction}}, \quad (\text{B.12})$$

in which $N_j = N^{(\delta)}(3900j)$ is the number of price events at time $3900j$. Note that the term $\sum_{i \in I_{j,d}} ICV_{i,d}^{M,m}$ is an estimator of the integrated variance of the interval $(t_{N_{j-1},d}^{(\delta)}, t_{N_j,d}^{(\delta)})$. However, in general $t_{N_j}^{(\delta)} \leq 3900j$ since the last price event in the j -th interval does not necessarily arrive at the boundary of the interval. Therefore, in $\sum_{i \in I_{j,d}} ICV_{i,d}^{M,m}$, we have included a volatility estimate of $(t_{N_{j-1}}^{(\delta)}, 3900(j-1)]$ at the left boundary which does not belong to $(3900(j-1), 3900j]$, but failed to estimate the integrated variance of the interval $(t_{N_j,d}^{(\delta)}, 3900j]$ of the right boundary. We provide an illustration of this boundary problem in Figure B.10.

To resolve this boundary problem, we need to remove the volatility estimates of $(t_{N_{j-1},d}^{(\delta)}, 3900(j-1)]$ at the left boundary and add the volatility estimate of $(t_{N_j}^{(\delta)}, 3900j]$ at the right boundary to $\sum_{i \in I_{j,d}} ICV_{i,d}^{M,m}$. Taking the left boundary correction for example. In $Q_{j,d}$, we assign $ICV_{N_{j-1}+1,d}^{M,m}$ to $(t_{N_{j-1},d}^{(\delta)}, 3900(j-1)]$ and $(3900(j-1), t_{N_{j-1}+1,d}^{(\delta)})$ based on the proportion of the length of the two intervals to the price duration $x_{N_{j-1}+1,d}^{(\delta)}$, and remove the corresponding proportion of the volatility estimates for the

Figure B.10 Illustration of the boundary problem of the intraday ICV estimator



Note: The above graph presents the left and right boundary corrections for $j=2$. Each red dot represents an arrival of a price event. The black dots are boundaries of the 3900-second intervals.

interval $(t_{N_{j-1},d}^{(\delta)}, 3900(j-1)]$. Similarly the second term in $Q_{j,d}$ corrects for the right boundary. Also note that the set $I_{j,d}$ can be an empty set. In this case, the price duration is a superset of a 3900-second interval, and the left and right corrections can be computed similarly by removing the proportion of the volatility estimates that does not lie in the 3900-second interval.

Appendix C

Appendix for Chapter 3

C.1 Proof to Proposition 3.1

The necessity of condition 1 is obvious. The sufficiency of condition 1 is rejected by the example in Section 2.3. To prove that condition 2 ensures the stationarity of both duration and intensity processes, consider the following expansion:

$$\tilde{\Phi}_i = \sum_{j=1}^{\infty} \psi_j \tilde{\varepsilon}_{i-j} = -\gamma \sum_{j=1}^{\infty} \psi_j - \sum_{j=1}^{\infty} \psi_j \ln \varepsilon_{i-j}. \quad (\text{C.1})$$

The absolute summability of ψ_j is guaranteed by condition 1 so the constant $-\gamma \sum_{j=1}^{\infty} \psi_j$ is bounded. The conditional intensity process can therefore be expressed as:

$$\lambda(t_i | \mathcal{F}_{t_i}) = C \prod_{j=1}^{\infty} \varepsilon_{i-j}^{-\psi_j}, \quad (\text{C.2})$$

in which $C = w - \gamma \sum_{j=1}^{\infty} \psi_j$ is a bounded constant. Similarly, the duration process can be expressed as:

$$x_i = C \varepsilon_i \prod_{j=1}^{\infty} \varepsilon_{i-j}^{\psi_j}. \quad (\text{C.3})$$

Note that both the intensity and duration processes are in the form of products of an infinite number of power transformed i.i.d.-unit exponential random variables. The second moment of $\lambda(t_i | \mathcal{F}_{t_i})$ and x_i are, respectively:

$$\text{E}[\lambda(t_i | \mathcal{F}_{t_i})^2] = C^2 \prod_{j=1}^{\infty} \text{E}[\varepsilon_{i-j}^{-2\psi_j}], \quad (\text{C.4})$$

$$\text{E}[x_i^2] = C^2 \prod_{j=1}^{\infty} \text{E}[\varepsilon_{i-j}^{2\psi_j}]. \quad (\text{C.5})$$

Clearly, the second moment of $\lambda(t_i|\mathcal{F}_{t_i})$ and x_i exist iff $0 < \mathbb{E}[\varepsilon_{i-j}^{\pm 2\psi_j}] < \infty, \forall j$. Because $\varepsilon_{i-j} \sim i.i.d.Exp(1), \forall i, j$, the expectation can be expressed as:

$$\mathbb{E}[\varepsilon_{i-j}^{\pm 2\psi_j}] = \int_0^\infty x^{\pm 2\psi_j} e^{-x} dx = \Gamma(\pm 2\psi_j + 1). \quad (\text{C.6})$$

One property of real-valued $\Gamma(x)$ is that it converges for all $x \in \mathbf{R}^+$, and diverges otherwise. Thus $1 \pm 2\psi_j > 0$ is required to guarantee the existence of the second moment. This also suggests that if the second moment of both $\lambda(t_i|\mathcal{F}_{t_i})$ and x_i exist, we must have $1 \pm 2\psi_j > 0$. We therefore conclude that $|\psi_j| < 0.5, \forall j$ and condition 1 are the sufficient and necessary conditions for weakly stationarity of both conditional intensity and duration processes, which completes the proof.

C.2 Approximated Single Move Sampler

Instead of using (3.37) that conditions on all the information up to time T , we calculate the following quantity:

$$p(s_i = l | s_{-i}^{(n)}, \theta^{(n)}, \mathbb{Y}) \approx \frac{p(s_i = l | s_{i-1}^{(n+1)}, s_{i+1}^{(n)}, \Pi^{(n)}) f(y_{i:i+\Delta} | s_i = l, s_{-i}^{(n)}, \theta^{(n)})}{\sum_m^{\mathcal{M}} p(s_i = m | s_{i-1}^{(n+1)}, s_{i+1}^{(n)}, \Pi^{(n)}) f(y_{i:i+\Delta} | s_i = m, s_{-i}^{(n)}, \theta^{(n)})}, \quad (\text{C.7})$$

in which $s_{-i}^{(n)} = \{s_{1:i-1}^{(n+1)}\} \cup \{s_{i+1:T}^{(n)}\}$. For every $s_i \in \mathcal{M}$, the following relationship holds:

$$f(y_{i:T} | s_i = l, s_{-i}^{(n)}, \theta^{(n)}) = f(y_{i:i+\Delta} | s_i, s_{-i}^{(n)}, \theta^{(n)}) f(y_{i+\Delta+1:T} | s_i, s_{-i}^{(n)}, \theta^{(n)}). \quad (\text{C.8})$$

To validate our approximation, we thus only need to show that $\exists \Delta$ and $\forall l, m \in \mathcal{M}$:

$$f(y_{i+\Delta+1:T} | s_i = l, s_{-i}^{(n)}, \theta^{(n)}) \approx f(y_{i+\Delta+1:T} | s_i = m, s_{-i}^{(n)}, \theta^{(n)}). \quad (\text{C.9})$$

This in fact holds for any sequence of s_i given the stationarity condition of the MS-ACI model. A change in s_i only affects the conditional density of $y_{i:T}$ through $\tilde{\Phi}_i(s_i)$ and $\tilde{\varepsilon}_i(s_i)$, which can be interpreted as innovation impulses at t_i and t_{i+1} . Due to the stationarity of the Markov chain and the $\tilde{\Phi}_i$ component which is a zero mean MS-ARMA structure, the innovation impulses will decay exponentially and $\tilde{\Phi}_{i+\Delta}$ will converge to its stationary distribution for $\Delta \rightarrow \infty$. Thus by eliminating the term (C.9) from the denominator and numerator of the standardized multinomial density derived from (3.37), our approximate sampler can be derived. The empirical choice of Δ depends on the persistence of the MS-ACI model, and we select Δ adaptively to ensure a precise approximation and an improvement in computational efficiency.

To choose Δ adaptively, at the (n) -th iteration, we calculate the posterior probabilities of $s_i = l$ conditioning on $\vartheta^{(n)}$, $s_{1:i-1}^{(n)}$ and $s_{i+1:T}^{(n)}$ with $\Delta = T$ for all i , and compute the same quantity with $\Delta = 1, 2, \dots$. We denote the $T \times 1$ vector of probabilities $\bar{p}(l)^n$ formed from the conditional posterior probabilities of each state belonging to state l with $\Delta = n$ calculated from equation (C.7) (the condition is omitted for conciseness), and the $T \times M$ probability matrix $\mathcal{P}^n = \{\bar{p}(1)^n, \dots, \bar{p}(M)^n\}$. Thus, \mathcal{P}^T is the exact conditional density of \mathbb{S} from the original single move sampler evaluated without sampling, and the autoregressive structure of our model predicts that $\lim_{n \rightarrow T} \mathcal{P}^n = \mathcal{P}^T$. We therefore pick Δ to be the value that satisfies the following criteria:

$$\Delta = \arg \min_n \left[\max(|\mathcal{P}^T - \mathcal{P}^n|) \leq \delta_3 \right], \quad n = 1 : T. \quad (\text{C.10})$$

This ensures that the maximum distance between \mathcal{P}^T and \mathcal{P}^Δ is smaller or equal to a predetermined small value δ_3 , which guarantees the quality of our approximation. It is obviously not efficient to evaluate Δ for every iteration due to the evaluation of \mathcal{P}^T , as there is a trade-off between accuracy and efficiency. In our analysis we choose $\delta_3 = 0.001$ and update Δ every 20 iterations.

C.3 Estimating the Variance-Covariance Matrix of Estimated Parameters and the Most Probable State Vector

As suggested by Delyon, Lavielle, and Moulines (1999) and Kuhn and Lavielle (2004), the variance-covariance matrix for parameter estimates can be estimated directly from the iterations of the SAEM algorithm. The inverse of the negative information matrix for the incomplete log-likelihood is a natural estimator for the variance-covariance matrix, which cannot be computed directly due to the difficulty in computing the incomplete log-likelihood. Louis (1982) derives an important identity to resolve this problem. Let $\mathcal{I}(\vartheta; \mathbb{Y})$ denote the observed information matrix, $\mathcal{I}(\vartheta; \mathbb{Y}, \mathbb{S})$ denote the complete information matrix, $s(\vartheta; \mathbb{Y})$ denote the observed score vector and $s(\vartheta; \mathbb{Y}, \mathbb{S})$ denote the complete score, then the following identity holds:

$$-\mathcal{I}(\vartheta; \mathbb{Y}) = -\mathbf{E}_{\mathbb{S}|\mathbb{Y}}[\mathcal{I}(\vartheta; \mathbb{Y}, \mathbb{S})|\mathbb{S}] - \mathbf{Cov}_{\mathbb{S}|\mathbb{Y}}[s(\vartheta; \mathbb{Y}, \mathbb{S})|\mathbb{S}]. \quad (\text{C.11})$$

Using the Fisher identity:

$$s(\vartheta; \mathbb{Y}) = \mathbf{E}_{\mathbb{S}|\mathbb{Y}}[s(\vartheta; \mathbb{Y}, \mathbb{S})|\mathbb{S}]. \quad (\text{C.12})$$

We can use the following stochastic approximation scheme to obtain an estimate of the observed information matrix:

$$S_n = (1 - \gamma_n)S_{n-1} + \gamma_n s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,K)}), \quad (\text{C.13})$$

$$D_n = (1 - \gamma_n)D_{n-1} + \gamma_n (\mathcal{J}(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,K)}) + s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,K)})s(\vartheta^{(n+1)}; \mathbb{Y}, \mathbb{S}^{(n,K)})'), \quad (\text{C.14})$$

$$H_n = D_n - S_n S_n'. \quad (\text{C.15})$$

Delyon, Lavielle, and Moulines (1999) show that $-H_n$ converges to $-\mathcal{J}(\vartheta, \mathbb{Y})$ when ϑ converges to the limiting value and the function $\ln \mathcal{L}(\vartheta; \mathbb{Y})$ is smooth enough. Thus the inverse of $-H_n$ will be used as an estimate of the variance-covariance matrix of ϑ .

As to the estimation of the most probable state vector, we rely on the posterior density of \mathbb{S} given the data and the estimated parameter vector $\hat{\vartheta}$. Due to the dimensionality of \mathbb{S} , it is difficult to directly maximize the likelihood of \mathbb{S} conditioning on $\hat{\vartheta}$ and \mathbb{Y} , and standard algorithms (e.g. Viterbi's algorithm) for the Hidden Markov Model are not applicable in this case because of the path dependency issue. Bauwens, Preminger, and Rombouts (2010) use the smoothed posterior probabilities for each state to obtain an estimate of this state sequence. This estimate, however, is not the most possible state sequence in the sense that it does not account for the dependence structure in the multivariate distribution of \mathbb{S} . We hereby propose a direct Monte Carlo search method to obtain $\hat{\mathbb{S}}$ by exploiting the following relationship:

$$p(\mathbb{S}|\hat{\vartheta}, \mathbb{Y}) \propto f(\mathbb{Y}|\mathbb{S}, \hat{\vartheta})p(\mathbb{S}) = f(\mathbb{Y}, \mathbb{S}|\hat{\vartheta}). \quad (\text{C.16})$$

Theoretically, one can use an arbitrary multivariate multinomial density $g(\mathbb{S})$ ¹ that corresponds to the dimension of the distribution of \mathbb{S} to simulate a set of N_s trial state sequences denoted as $\{\tilde{\mathbb{S}}_1, \dots, \tilde{\mathbb{S}}_{N_s}\}$, and a natural estimate is obtained by:

$$\hat{\mathbb{S}} = \arg \max_{n_s} f(\tilde{\mathbb{S}}_{n_s}, \mathbb{Y}|\hat{\vartheta}), \quad n_s = 1 : N_s. \quad (\text{C.17})$$

Intuitively, as $N_s \rightarrow \infty$ and $g(\tilde{\mathbb{S}}) > 0$ for all $\tilde{\mathbb{S}}$, this algorithm will exploit all possible realizations of $\tilde{\mathbb{S}}$ and obtain the one that maximizes the joint likelihood. An obvious problem is that this algorithm is extremely inefficient if $g(\mathbb{S})$ is not chosen properly, and a good kernel should account for the information in the observed data for the trial $\tilde{\mathbb{S}}$ to be simulated closely to where the likelihood is concentrated. Our solution

¹A possible choice of $g(\mathbb{S})$ can be, for example, a plain M -state Markov chain with the transition parameters $\hat{\Pi}$.

is to continue performing the S-step N_s times after the convergence of the SAEM algorithm with parameters fixed to the converged value $\hat{\vartheta}$, so that $g(\mathbb{S})$ in our case becomes the single move sampler. The underlying reason is that the conditional posterior probabilities used in the single move sampler include information of the observed data, therefore is much more efficient than an uninformed sampler such as a plain Markov chain. Other possible choices include the particle filter sampler described in Bauwens, Dufays, and Rombouts (2014) and the Metropolis-Hastings sampler in Billio, Casarin, and Osuntuyi (2014).

C.4 Deseasonalization of the Price Duration and Volume

This section explains the deseasonalization procedure of the price duration data $\{x_{i,d}, \ln Vol_{i,d}\}$. We apply a time-deterministic flexible Fourier regression as suggested by Andersen and Bollerslev (1997b), and filter out the diurnal pattern before model estimation.² Taking the duration series as an example, we firstly demean the raw price duration series on a daily basis:

$$\tilde{x}_{i,d}^\delta = \frac{x_{i,d}^\delta}{\bar{x}_d^\delta}, \quad (\text{C.18})$$

in which \bar{x}_d^δ is the daily average of the raw price durations. Then we estimate the following regression on the de-meaned duration series with OLS:

$$\tilde{x}_{i,d}^\delta = c + \sum_{p=1}^P c_p (\bar{t}_{i,d}^\delta)^p + \sum_{q=1}^Q c_{c,q} \cos(2q\pi\bar{t}_{i,d}^\delta) + c_{s,q} \sin(2q\pi\bar{t}_{i,d}^\delta) + u_{i,d}, \quad (\text{C.19})$$

in which $\bar{t}_{i,d}^\delta \in (0, 1]$ is the fraction of the time elapse of event $t_{i,d}^\delta$ from the beginning of a trading day, and $u_{i,d}$ is assumed to be a Gaussian zero mean error term. In the above regression, c is the constant term, c_p captures the time-deterministic mean duration with a polynomial of degree P , and $c_{s,q}$ and $c_{c,q}$ account for the curvature of this seasonality pattern in addition to the polynomial shape, with the maximum polynomial degree of this flexible Fourier form given by Q . We choose P and Q by running this regression with P ranging from 0 to 3 and Q ranging from 0 to 5, and pick the P and Q with the smallest Bayesian Information Criteria. The fitted

²The ACI model is capable of estimating the seasonality with the intensity jointly by adding a multiplicative seasonality component. However, this will greatly increase the number of parameters to be estimated for the MS-ACI model, which is undesirable. We therefore pre-filter the diurnal pattern before estimating the model, which is a common approach in the ACD literature.

value of this regression, $\hat{x}_{i,d}^\delta$ is therefore the estimated seasonality component. The deseasonalized duration, $\dot{x}_{i,d}^\delta$, is computed by:

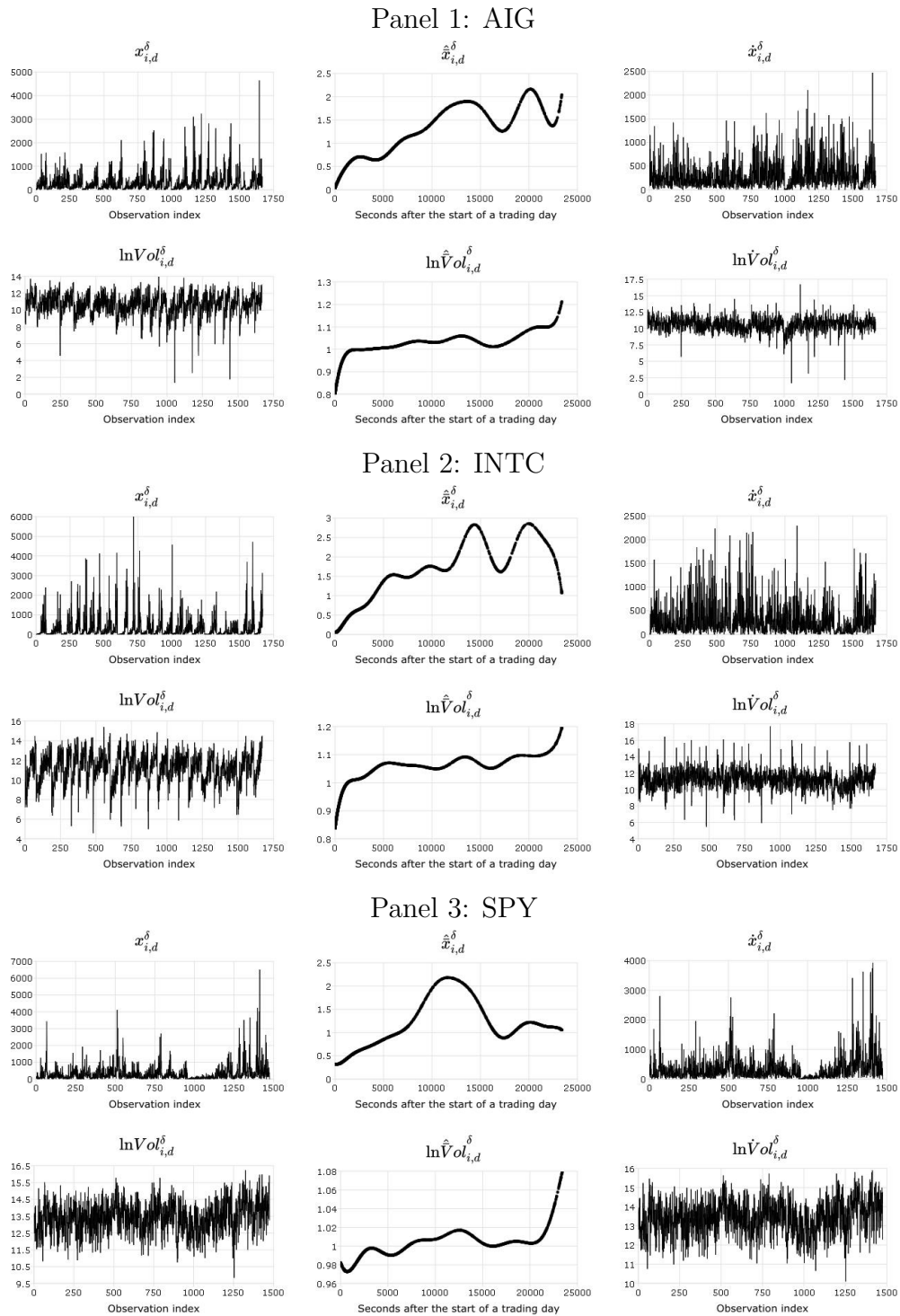
$$\dot{x}_{i,d}^\delta = \frac{x_{i,d}^\delta}{\hat{x}_{i,d}^\delta}. \quad (\text{C.20})$$

Note that the seasonality filter only controls for deterministic time effects, and the resulting deseasonalized price duration is not de-meanned. In the ACD literature it is common to obtain a deseasonalized price duration with unit mean, but clearly in our case $\mathbf{E}[\dot{x}_{i,d}^\delta]$ does not have a mean of one, simply because $\hat{x}_{i,d}^\delta$ is estimated on a de-meanned series and is of unit mean. This guarantees that the seasonality we capture is not affected by the daily mean shifts. Moreover, the daily means of the raw price durations are included in the deseasonalized price duration. It is because changes in the daily mean can also contain important information on the regime shifts.

We compute the deseasonalized price duration $\dot{x}_{i,d}^\delta$ and the deseasonalized log volume $\ln \dot{Vol}_{i,d}^\delta$ with the procedure above. To provide a graphical illustration of the deseasonalization process, we present some examples of the raw data and the corresponding deseasonalized data in Figure C.1 on the next page. From Figure C.1, we can clearly see the decomposition of the raw series into the unit mean seasonality component and the deseasonalized series. We find a reverse U-shape diurnal pattern for the price duration generally, which is consistent with the literature. We also discover a distinct pattern for the volume within each price duration. The trading volume starts very low at the beginning of the day which probably results from the small price duration at the start of the trading day. Interestingly volume usually peaks at the end of the trading day when the duration is also small. The diurnal pattern for the stock market index SPY is generally less prominent compared to individual stocks.

To show that our flexible Fourier regression successfully captures the diurnal pattern in the raw series, we present the Lomb-Scargle (Lomb, 1976; Scargle, 1982) periodogram for the raw and deseasonalized series in Figure C.2, which is a standard diagnostic tool for periodic pattern detection in irregularly sampled data. In Figure C.2, it is clear that for the raw series for all securities, the power spectral peaks at frequencies that are multiples of approximately 0.43×10^{-4} , which corresponds to the reciprocal of the total number of seconds in a trading day, $\frac{1}{23400}$. These spikes correspond to a daily recurring pattern and its harmonics, which is a strong evidence for the diurnal pattern of the raw series. After the deseasonalization, all the spikes at multiples of 0.43×10^{-4} are removed from the raw series, which suggests that our

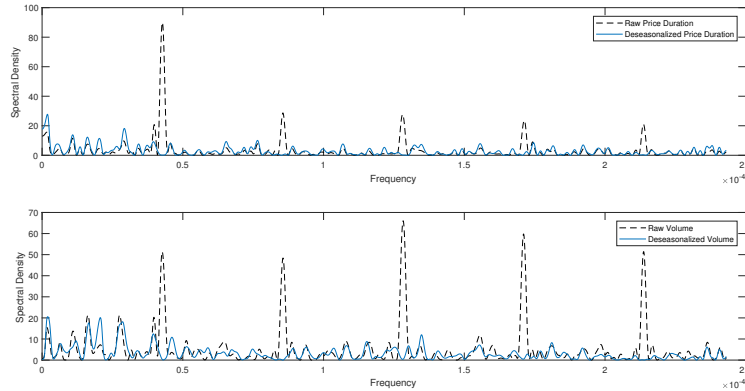
Figure C.1 Examples of Deseasonalization: Raw and Deseasonalized Price Duration and Volume from AIG 2016-03, INTC 2016-08, SPY 2016-05



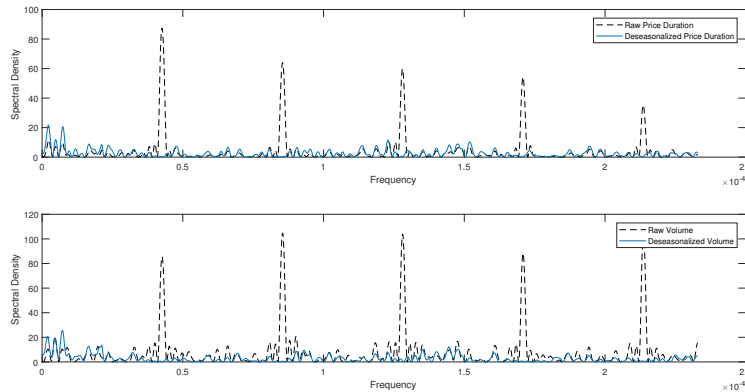
Note: Three 2-by-3 graphs from top to bottom: AIG 2016-03, INTC 2016-08, SPY 2016-05. The x-axis is the observation index, the y-axis is the value of the series. $x_{i,d}^\delta$ and $\ln Vol_{i,d}^\delta$ denote the raw price duration and raw log cumulative trading volume within the corresponding duration. $\hat{x}_{i,d}^\delta$ and $\hat{\ln Vol}_{i,d}^\delta$ are the fitted value from the flexible Fourier regression in (C.19) computed from the corresponding raw series. Finally, $\tilde{x}_{i,d}^\delta$ and $\tilde{\ln Vol}_{i,d}^\delta$ are the deseasonalized version of the two series calculated by (C.20).

Figure C.2 Lomb-Scargle Periodogram for Raw and Deseasonalized Price Durations, Volume and Bid-Ask Spread Covariates for AIG 2016-03, INTC 2016-08, and SPY 2016-05

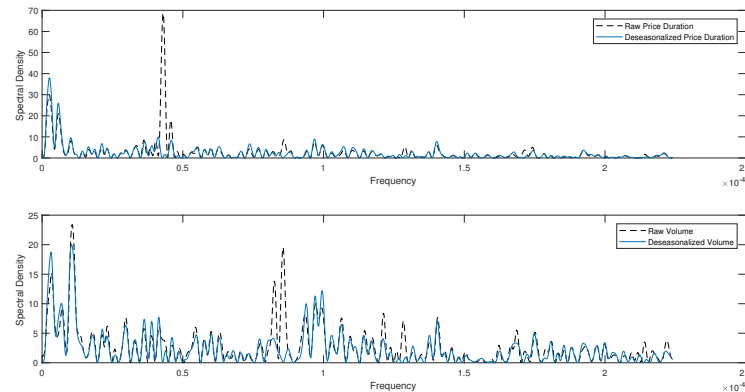
Panel 1: AIG



Panel 2: INTC



Panel 3: SPY



Note: Three 2-by-1 graphs from top to bottom: AIG 2016-03, INTC 2016-08, SPY 2016-05. x-axis is the frequency of the data considered and y-axis is the estimated power spectral density corresponding to the frequency. A daily frequency is equivalent to $\frac{1}{23400} \approx 0.46 \times 10^{-4}$ in the x-axis. The dashed line represents the periodogram for the raw series and the solid line is its deseasonalized version.

deseasonalization regression successfully captures the diurnal pattern in the raw series. Moreover, the spectral density spikes at other frequencies remain in the periodogram. To sum up, our deseasonalization procedure successfully captures and removes only the diurnal patterns from both the raw price durations and volumes. The quality of this deseasonalization is corroborated by the Lomb-Scargle periodogram.

C.5 Robustness Checks of the MS(2)-ACI(2,1)-V Model

In this section, we justify our choice of the MS(2)-ACI(2,1)-V model. In detail, we compare the performance of our model to the original ACI(2,1) model, the ACI(2,1)-V model, and the MS(2)-ACI(2,1) model. We present comprehensive estimation outputs for all 4 models considered above for the three example stock-months: AIG 2016-03, INTC 2016-08 and SPY 2016-05. The results can be found in Tables C.2, C.3 and C.4 in Appendix C.7.

From Tables C.2, C.3 and C.4, it is evident that the estimated persistence parameters $\beta(m)$, the baseline parameters $w(m)$ and the volume parameters $\eta_{vol}(m)$ of the one-regime model are weighted averages of the corresponding parameters across two regimes from the two-regime model, as expected. As discussed in Section 1.7, the validity of the Markov-switching structure depends crucially on the discrepancy between the regime-specific location parameters of the baseline function. Therefore, the differences in the estimated $\hat{w}(m)$ and $\hat{\eta}_{vol}(m)$ between regimes are the first indicators of the performance of the regime-switching structure. The general observation for the three example stock-months and our entire dataset is that, the difference between $\hat{w}(m)$ s are very small for the MS(2)-ACI(2,1) model. By adding the volume covariate to the two- and three-regime models, the difference in the estimated $\hat{w}(m)$ widen significantly for the individual stocks, and there also exists a large discrepancy in the estimated $\hat{\eta}_{vol}(m)$ for those stocks. For the SPY, however, the difference in the estimated $\hat{\eta}_{vol}(m)$ is much less pronounced.

The validity of the regime-switching structure is better summarized in the \widehat{SoR} statistics, as discussed in Section 3.3.4. In Tables C.2, C.3 and C.4, we can clearly see that when volume is not included, \widehat{SoR} is around 70% for the example stock-months. The inclusion of the volume covariate greatly improves the \widehat{SoR} for AIG and INTC to more than 95%, but the \widehat{SoR} for the SPY is still around 70%. From Figure C.6, it is evident that almost all \widehat{SoR} estimated from the MS(2)-ACI(2,1)-V model for the 9 stocks are over 95%, which shows that the MS(2)-ACI(2,1)-V model classifies the

regimes with a very high confidence. The \widehat{SoR} s for the MS(2)-ACI(2,1) model are considerably lower, and rarely exceeds 90%. For the SPY, we see that the \widehat{SoR} s are significantly smaller than the other stocks, which corroborates our findings in Section 3.5.2 that the regime classification is much weaker for the stock index ETF.

The discussion above provides strong evidence supporting our key findings in Chapter 3, as there indeed exists regime-switching behaviour in the intraday volume-duration relationship that can be identified with strong confidence. However, the substantial \widehat{SoR} of the MS(2)-ACI(2,1) model also requires investigation, as it appears that we may be able to detect the regime-switching behaviour in the volatility process itself without the inclusion of the volume covariate. To address this concern, we plot the yearly distribution of estimated regimes for the MS(2)-ACI(2,1) model in Figure C.7. Comparing Figures C.4 and C.7, we clearly see that we can only observe the diurnal pattern of the informed regime (regime 1) when the volume covariate is included, as in Figure C.4. In the absence of the volume covariate, the captured two regimes have similar distributions across the trading hours. This clearly indicates that the regimes detected by the MS(2)-ACI(2,1) model is very different from the MS(2)-ACI(2,1)-V model. A possible explanation to this is that the one-regime model with a Weibull baseline cannot fully capture the empirical density of the durations, and the two regimes identified in the MS(2)-ACI(2,1) model simply attempt to fit the empirical data with a mixture of two Weibull distributions.

We proceed to show that the two-regime models are superior to the one-regime counterparts as they provide a much more flexible baseline and autoregressive structure for the conditional intensity. To examine the goodness-of-fit of the models, we examine the distributional assumption and the degree of autocorrelation in the residuals $\hat{\epsilon}_i$. Examples in Tables C.2 and C.3 show that residuals from the ACI(2,1) exhibit over-dispersion, i.e. $\sigma(\hat{\epsilon}_i) > 1$. This results in high Cramér-von-Mises and Andersen-Darling statistics that reject the null hypothesis of unit exponentiality of the residuals. Consequently, the Weibull baseline fails to correctly capture the distribution of empirical price durations. This problem is exacerbated when the volume is added to the ACI(2,1) model, with residuals deviating further from the unit exponential distribution. The volume covariate also introduces noise to the autoregressive structure of the ACI(2,1) model, leading to a larger Ljung-Box test statistics. By introducing another regime, we observe significant improvements in all statistics considered.

To illustrate the properties of the residuals from each model in Tables C.2, C.3

and C.4, we present Quantile-Quantile plots and correlograms for the residuals in Figures C.8 and C.9. Figure C.8 shows that the Weibull baseline cannot capture the tail behaviour of the empirical distribution of the price durations, and an additional regime can to a large extent mitigate this issue. In Figure C.9, we clearly see that the inclusion of volume introduces noise to the autocorrelation structure of the residuals in the one-regime model, potentially due to the diurnal pattern of the regime-switching volume-duration relationship. This structure can be captured by the MS(2)-ACI(2,1)-V model, resulting in the less significant autocorrelation structure in the residuals of the two-regime model.

In order to show that the two-regime models are superior to the one-regime models in terms of goodness-of-fit for all the stock-months considered, we present a summary of Cramér-von-Mises, Andersen-Darling and Ljung-Box tests results for all stock-months in Table C.5. From the table it is evident that both the MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V model outperform their one-regime counterparts in all the cases considered. The inclusion of volume covariate in both the one- and two-regime models leads to a deterioration of the goodness-of-fit of the model. However, almost all CvM and AD tests are strongly significant for the ACI(2,1)-V model, while the MS(2)-ACI(2,1)-V has much fewer strong significant results.

We examine the log-likelihoods from the four models as an overall measure of the goodness-of-fit of the models. As discussed in Section 3.3.4, the observed likelihood for the MS-ACI models are not available, and we can only rely on the conditional likelihoods $\mathcal{L}(\hat{\vartheta}; \mathbb{Y}|\hat{\mathbb{S}})$. These likelihoods cannot be used to construct a likelihood ratio test, but can still be informative on the relative improvements from a restricted model to a more flexible one. We construct Bayesian Information Criterion (BIC) based on the usual likelihood $\mathcal{L}(\hat{\vartheta}; \mathbb{Y})$ for the one-regime models and the conditional likelihood $\mathcal{L}(\hat{\vartheta}; \mathbb{Y}|\hat{\mathbb{S}})$ for the two-regime models, and present the ranking of each model for each stock-month in Figure C.10. The figure shows that the BIC suggests that the MS(2)-ACI(2,1)-V model outperforms the other three models for all 120 stock-months with the smallest BIC. Interestingly, we find that the ACI(2,1)-V model frequently outperforms the MS(2)-ACI(2,1) model. This is due to the fact that the contribution of another regime is offset by the penalty on the number of parameters. However, we see that both a regime-switching structure and the volume covariate are needed to achieve a better BIC.

Finally, we briefly discuss the effect of more than two regimes. We present examples of estimation outputs of MS(3)-ACI(2,1)-V model in Table C.6 to compare

against the performance of a MS(2)-ACI(2,1)-V model. In Table C.6, we see that the diagnostic statistics for residuals and likelihoods, such as BIC, CvM, AD and LB(50) actually suggest that the MS(3)-ACI(2,1)-V model performs slightly better than the MS(2)-ACI(2,1)-V model. This is not surprising as the three-regime model provides an even more flexible structure, and there are still room for improvements for the MS(2)-ACI(2,1)-V model.

The more important concern is the interpretation of the regimes captured by the three-regime model. From the magnitude of $\hat{\eta}_{vol}(m)$ of AIG and INTC, we can conclude that in addition to the high information regime (regime 3 for AIG and 1 for INTC) and a low information regime (regime 2 for AIG and INTC), there is also an intermediate regime. We compare the classification of regimes for the two-regime and three-regime model, and present the results in Table C.7. In the table, we see that for AIG, the majority of observations in regime 2 for the two-regime model remain in regime 2, and the high-information regime is split into regime 1 and 3 in the three-regime case. This is the opposite case for INTC, with the high-information regime remaining relatively stable and the low-information regime being separated. Finally for the SPY, there seems to be less association between the regime classifications of the two- and three-regime models. This transition in classifications from the two-regime model to the three-regime model can also be observed from the distribution of regimes in Figure C.11. It is clear that for AIG, regime 2 of the three-regime model correspond to the low-information regime of the two-regime model, while regimes 2 and 3 add up to the low-information regime for INTC.

As the information content can change smoothly during the trading day, additional regimes in the model attempt to fit the data by classifying the durations that are less likely to be in either a high information content or a low information content regime into an intermediate regime. However, the higher the number of regimes, the more overlapped the regime-specific densities are, and the less accurate the regime classification is. As can be seen from the \widehat{SoR} of the MS(3)-ACI(2,1)-V model, there is a considerable decrease in the confidence of regime classification due to the additional regime. Therefore, for the purpose of regime classification, it is not optimal to include another regime onto the MS(2)-ACI(2,1)-V model.

To conclude our robustness check section, we demonstrate that the MS(2)-ACI(2,1)-V model is able to classify durations into two distinct regimes with very high confidence, and the volume covariate plays an indispensable role in the significance of regime classification. The regime-switching model shows a clear advantage over the

one-regime ACI model in terms of residual diagnostics, with the MS(2)-ACI(2,1)-V model having the lowest BIC for all stock-month combinations. We further show that it is possible to enhance the performance of the MS(2)-ACI(2,1)-V model by introducing another regime, but the confidence of regime classification can be greatly hampered by the additional regime.

C.6 A Market Microstructure Model for Informed Trading and the Volume-Volatility Relationship

We start by establishing the link between price durations and the intraday local volatility process. From (3.44), it is clear that the conditional intensity process is proportional to the instantaneous volatility process. Let us consider the average conditional intensity for the duration $(t_{i-1}, t_i]$ denoted by $\bar{\lambda}_i^\delta$. From the Random Time Change Theorem we have:

$$\Lambda_i^\delta \equiv \int_{t_{i-1}}^{t_i} \bar{\lambda}_i^\delta dt = \bar{\lambda}_i^\delta x_i^\delta \sim i.i.d.Exp(1) \quad (C.21)$$

Consequently:

$$E[\ln \bar{\lambda}_i^\delta] + E[\ln x_i^\delta] = -\gamma, \quad (C.22)$$

where γ is the Euler-Mascheroni constant. This simple relationship suggests that, the longer the expected duration, the smaller the expected average intensity within each duration. This expectation is also valid if one conditions on $\mathcal{F}_{t_{i-1}}$. Also, it is clear that:

$$\text{Cov}(\ln \bar{\lambda}_i^\delta, \ln x_i^\delta) = \frac{\pi^2}{12} - \frac{V[\ln \bar{\lambda}_i^\delta] + V[\ln x_i^\delta]}{2}. \quad (C.23)$$

Since empirically the R.H.S is almost always negative due to the large variance of log price duration (in our setting, $V[\ln x_i^\delta]$ is around 3), the log price duration and log average spot volatility are also negatively correlated. Note that the above relationship holds for any stationary process with finite moments, and deseasonalization does not affect the relationship above.

We proceed by imposing some assumptions about the market, which is commonly used in the existing literature (e.g. Copeland and Galai (1983), Easley, Kiefer, O'Hara, and Paperman (1996), Glosten and Milgrom (1985) and Kyle (1985)).

Assumption C.1 (Assumptions of the market). *We impose the following assumptions on the market:*

1. *News arrivals are strongly exogenous.*
2. *There exist two type of participants in the market: informed traders and uninformed traders.*
3. *Informed traders only trade when they possess private information about the fundamental value of the asset not incorporated in the price process.*
4. *Uninformed traders trade for liquidity and speculative purposes. They cannot predict price movements and do not alter the fundamental value of the asset.*

It is understood that the uninformed traders include the market makers. The liquidity supply and demand are related to the uncertainty in the market, such that when the market maker is unsure of the information content, the bid-ask spread will inflate, discouraging uninformed traders to participate. Thus, only informed traders who can bear the extra transaction costs are willing to trade. It is a stylized fact that the bid-ask spread and price volatility spike at the beginning of a trading day due to overnight information aggregation. Therefore trading activities of the liquidity traders will also be affected negatively by the uncertainty in the market as the transaction cost increases.

Supported by our empirical investigation of the data and the notion of Market Microstructure Invariants by Kyle and Obizhaeva (2012), we assume that when the informed trader is absent, the deseasonalized log price duration $\ln \dot{x}_i^\delta$ and the deseasonalized log cumulative volume over the same price duration $\ln \dot{Vol}_i^\delta$ follows a log-linear relationship:

$$\ln \dot{Vol}_i^\delta = b_0 + b_1 \ln \dot{x}_i^\delta + \varepsilon_i, \quad (\text{C.24})$$

in which b_0 is an intercept term accounting for the minimum volume requirements of an order, $b_1 > 0$ represents a constant arrival rate of volume, and ε_i is an i.i.d. zero mean innovation term with finite variance.

The above relationship suggests that, adjusted for seasonality effect, we would expect the volume of the uninformed traders to accumulate when the price duration increases. Since the price duration is negatively correlated with the average spot volatility, this actually suggests a negative correlation between accumulated volume within a duration and the average spot volatility within the duration. This is also consistent with the parameter estimates of $\hat{\eta}_{vol}(2)$ for the uninformed regime. In this relationship, the price duration is only interpreted as time for the uninformed trading

volume to accumulate. This is because from Assumption 1(d), uninformed traders do not change the fundamental value of the asset, and their volume accumulation is only relevant to the average spot volatility up to a seasonality effect in the absence of informed traders.

To include the impact of private information, we assume that the total volume can be decomposed as follows:

$$\ln Vol_i^\delta = b_0 + b_1 \ln x_i^\delta + \ln Z_i + \varepsilon_i. \quad (\text{C.25})$$

In which $\ln Z_i$ is the proportion of volume submitted by the informed traders in the i -th price duration. We impose the restriction $\ln Z_i \geq 0$, so that informed traders cannot remove volume from the total volume. For simplicity, we assume that Z_i also follows a power law:

$$\ln Z_i = s_i(\mu + c_1 \ln \bar{\lambda}_i^\delta + u_i), \quad (\text{C.26})$$

in which $s_i = 1$ if the informed traders participate in the market within the price duration i , and $s_i = 0$ otherwise. The parameter $\mu > 0$ serves as the minimum amount traded by the informed traders, c_1 is the interaction between the log informed volume and the log average intensity (volatility), and u_i is a zero mean weak white noise process subject to $u_i \geq \mu + c_1 \ln \bar{\lambda}_i^\delta$ for all i . Literature on both the volume-volatility relationship (e.g. Bessembinder and Seguin (1993), Arag3 and Nieto (2005), Hussain (2011), etc.) and the relevant market microstructure models (e.g. Glosten and Milgrom (1985), Easley, Kiefer, O'Hara, and Paperman (1996)) predict that $c_1 > 0$, such that the higher average volatility is associated with more volume submitted by informed traders. We can then expand the expression of the total volume as:

$$\ln Vol_i^\delta = (b_0 + s_i\mu + s_i c_1 E[\ln \bar{\lambda}_i^\delta]) + b_1 \ln x_i^\delta + s_i c_1 (\ln \bar{\lambda}_i^\delta - E[\ln \bar{\lambda}_i^\delta]) + s_i u_i + \varepsilon_i \quad (\text{C.27})$$

When $s_i = 0$, the above equation reduces to (C.24), and one can consistently estimate b_0 and b_1 via a simple OLS regression. In the $s_i = 1$ case, OLS estimates of b_1 will be biased due to the omitted variable $\ln Z_i$. We can derive the expected value of the OLS estimates \hat{b}_0 and \hat{b}_1 conditioning on $s_i = 1$:

$$\begin{aligned} E[\hat{b}_0 | s_i = 1] &= b_0 + E[\ln Z_i | s_i = 1] = b_0 + \mu + c_1 E[\ln \bar{\lambda}_i^\delta | s_i = 1], \\ E[\hat{b}_1 | s_i = 1] &= b_1 + \frac{c_1 \text{Cov}(\ln \bar{\lambda}_i^\delta, \ln x_i^\delta | s_i = 1)}{V[\ln x_i^\delta | s_i = 1]}. \end{aligned} \quad (\text{C.28})$$

We clearly see that \hat{b}_0 is biased upwards and \hat{b}_1 is biased downwards due to the omitted variable $\ln Z_i$ which is negatively correlated with $\ln x_i^\delta$. The bias for both

parameters is directly influenced by the parameter c_1 . Clearly, the larger c_1 , the more biased the OLS estimates of b_0 and b_1 . Thus, larger impact of informal trading volume on price volatility drives the total volume away from the log-linear relationship in (C.24). The additional residual term u_i in the regression will also reduce the R^2 of the regression when $s_i = 1$. To infer s_i from $\ln Vol_i^\delta$ and $\ln x_i^\delta$, one can rely on deterministic pattern of s_i as attempted in the data section. A better solution is to design a regime-switching model that captures the difference in the estimated \hat{b}_0 and \hat{b}_1 conditioning on s_i , which in essence is the identification strategy of our MS-ACI model.

To sum up, the simple model above establishes the relationship between price duration and the associated trading volume, and links it to the information content of the price duration. Results in Figures 3.5, 3.6 and C.5 correctly predict the theoretical differences in the estimated b_0 , b_1 and R^2 between the states of high and low information content. This justifies our interpretation of regime classifications based on levels of information content.

C.7 Tables and Figures

Table C.1 Yearly Descriptive Statistics for $x_{i,d}^\delta$ and $\ln Vol_{i,d}^\delta$

Ticker	Obs.	$x_{i,d}^\delta$					$\ln Vol_{i,d}^\delta$				
		Mean	σ	Min	Median	Max	Mean	σ	Min	Median	Max
AIG	17925	323.72	487.98	0.00	159.31	10299.68	10.52	1.48	0.00	10.67	14.86
CVX	18014	321.32	466.08	0.00	165.18	9471.82	10.77	1.36	0.00	10.89	14.71
GM	17240	331.83	597.59	0.00	127.35	11761.73	11.24	1.41	0.00	11.39	15.43
INTC	16541	349.34	571.09	0.00	149.33	9459.21	11.79	1.47	2.30	11.94	16.57
JPM	17270	332.57	552.04	0.00	142.91	10860.28	11.61	1.28	0.00	11.72	15.56
PFE	17416	334.16	542.23	0.00	154.52	11331.14	12.03	1.51	1.10	12.21	16.42
SPY	18530	314.17	460.01	0.00	165.88	10658.61	13.52	1.04	3.66	13.56	17.09
T	17008	340.42	559.00	0.00	146.19	10465.31	11.83	1.45	0.00	12.01	15.56
VZ	16853	342.79	563.45	0.00	145.60	10812.89	11.27	1.51	0.00	11.47	15.30
WMT	17082	338.55	570.51	0.00	140.41	10508.58	10.83	1.49	0.00	11.01	15.10

Note: The table presents the descriptive statistics for the raw price durations $x_{i,d}^\delta$ and the raw volume $\ln Vol_{i,d}^\delta$ from 10 securities for the year 2016. Obs. denotes the total number of observations. σ is the standard deviation.

Table C.2 Comprehensive estimation outputs for AIG 2016-03

Model	ACI(2,1)	ACI(2,1)-V	MS(2)-ACI(2,1)		MS(2)-ACI(2,1)-V	
$m =$	1	1	1	2	1	2
Dynamic Parameters						
$\alpha_1(m)$	0.0741*** (0.0235)	0.0473 (0.0432)	0.2766*** (0.0197)	0.0109 (0.0128)	0.0223 (0.0278)	0.2310*** (0.0201)
$\alpha_2(m)$	-0.0065 (0.0223)	-0.0152 (0.0449)	-0.0095 (0.0149)	-0.0994*** (0.0141)	0.0514* (0.0277)	-0.0988*** (0.0206)
$\beta(m)$	0.9731*** (0.0066)	0.9917*** (0.0074)	1.2146*** (0.0226)	0.7831*** (0.0174)	1.0572*** (0.0128)	0.9388*** (0.0104)
Baseline Parameters						
$w(m)$	-5.7144*** (0.0837)	1.9300*** (0.2817)	-5.3204*** (0.0329)	-5.9165*** (0.0357)	1.2721** (0.5477)	3.0143*** (0.1181)
$a(m)$	1.1940*** (0.0248)	1.7586*** (0.0513)	1.6332*** (0.0422)	1.2510*** (0.0276)	1.0137*** (0.0460)	2.6856*** (0.0512)
Other Parameters						
$\eta_{vol}(m)$		-1.2268*** (0.0518)			-0.5982*** (0.0580)	-2.1542*** (0.0487)
π_{1m}			0.3090*** (0.0180)		0.8364*** (0.0233)	
π_{2m}				0.5410*** (0.0157)		0.9720*** (0.0044)
Diagnostic Statistics						
Obs. In $\hat{s}_i = m$	1668	1668	665	1003	236	1432
$E[\hat{\epsilon}_i]$	0.9966	1.0179	0.969		1.0011	
$\sigma[\hat{\epsilon}_i]$	1.0473	1.2407	0.9433		1.0004	
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y} \hat{\mathbb{S}})$	-11046.42	-10379.48	-10768.4		-9895.10	
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}, \hat{\mathbb{S}})$			-11855.5		-10159.1	
BIC	13.2674	12.4721	12.9651		11.9269	
CvM	0.4213*	2.0902***	0.2113		0.0968	
AD	3.4050**	12.1077***	1.8107		0.5669	
LB(50)	51.9571	132.8224***	51.8145		44.5522	
\widehat{SoR}			0.7394		0.9759	
$\widehat{SoR}(m)$			0.6806	0.7766	0.9508	0.9798

Note: Standard errors are in parentheses. ***, ** and * represent significance at 1%, 5% and 10% respectively. Observation counts and residual statistics are based on the estimated most probable state sequence. BIC is computed as $BIC = T^{-1}(k \ln T - 2 \ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}|\hat{\mathbb{S}}))$ where T is the number of observations. The CvM and AD are Cramér-von-Mises and Andersen-Darling statistics (Stephens, 2013) for unit exponential distribution, with the critical values bootstrapped from 100000 simulated unit exponential vectors. LB(50) is the Ljung-Box (Ljung and Box, 1978) test statistics at lag 50. Definitions of \widehat{SoR} and $\widehat{SoR}(m)$ can be found in (3.42) and (3.43).

Table C.3 Comprehensive estimation outputs for INTC 2016-08

Model	ACI(2,1)	ACI(2,1)-V	MS(2)-ACI(2,1)		MS(2)-ACI(2,1)-V	
$m =$	1	1	1	2	1	2
Dynamic Parameters						
$\alpha_1(m)$	0.1670*** (0.0299)	0.1998*** (0.0259)	0.3603*** (0.0246)	-0.0014 (0.0198)	0.3258*** (0.0385)	0.2192*** (0.0206)
$\alpha_2(m)$	0.0308 (0.0474)	0.0484 (0.0457)	-0.0911*** (0.0257)	-0.0807*** (0.0173)	0.1747*** (0.0580)	-0.0807*** (0.0222)
$\beta(m)$	0.4335*** (0.1669)	0.3484** (0.1693)	1.3869*** (0.0252)	0.6826*** (0.0161)	0.4121*** (0.1294)	0.9241*** (0.0185)
Baseline Parameters						
$w(m)$	-5.7263*** (0.0419)	2.1623*** (0.3580)	-5.3420*** (0.0429)	-5.9541*** (0.0317)	-0.6130 (0.5590)	4.1465*** (0.1482)
$a(m)$	0.8836*** (0.0207)	1.1993*** (0.0358)	1.1188*** (0.0306)	1.0893*** (0.0235)	0.6473*** (0.0297)	2.2720*** (0.0477)
Other Parameters						
$\eta_{vol}(m)$		-0.8249*** (0.0470)		-0.2173***	-1.9628*** (0.0349)	(0.0486)
π_{1m}			0.2817*** (0.0186)		0.8338*** (0.0215)	
π_{2m}				0.6110*** (0.0148)		0.9635*** (0.0052)
Diagnostic Statistics						
Obs. In $\hat{s}_i = m$			552	1118	273	1397
$E[\hat{\epsilon}_i]$	1.0003	1.0002	0.9832		0.9982	
$\sigma[\hat{\epsilon}_i]$	1.1334	1.2273	0.9334		1.0706	
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y} \hat{\mathbb{S}})$	-11316.0	-10824.4	-10876.4		-10048.9	
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}, \hat{\mathbb{S}})$			-11923.7		-10352.3	
BIC	13.5743	12.9900	13.0790		12.0968	
CvM	0.4592*	3.7252***	0.1389		0.1971	
AD	3.2887**	23.7503***	1.4335		1.297	
LB(50)	110.2901***	159.6120***	52.2646		79.8233***	
\widehat{SoR}			0.7421		0.9776	
$\widehat{SoR}(m)$			0.6739	0.7654	0.9582	0.9814

Note: Standard errors are in parentheses. ***, ** and * represent significance at 1%, 5% and 10% respectively. Observation counts and residual statistics are based on the estimated most probable state sequence. BIC is computed as $BIC = T^{-1}(k \ln T - 2 \ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y} | \hat{\mathbb{S}}))$ where T is the number of observations. The CvM and AD are Cramér-von-Mises and Andersen-Darling statistics (Stephens, 2013) for unit exponential distribution, with the critical values bootstrapped from 100000 simulated unit exponential vectors. LB(50) is the Ljung-Box (Ljung and Box, 1978) test statistics at lag 50. Definitions of \widehat{SoR} and $\widehat{SoR}(m)$ can be found in (3.42) and (3.43).

Table C.4 Comprehensive estimation outputs for SPY 2016-05

Model	ACI(2,1)	ACI(2,1)-V	MS(2)-ACI(2,1)		MS(2)-ACI(2,1)-V	
$m =$	1	1	1	2	1	2
Dynamic Parameters						
$\alpha_1(m)$	0.0520* (0.0286)	0.2912*** (0.0265)	-0.0430 (0.0301)	0.0995*** (0.0294)	0.4526*** (0.0240)	0.1942*** (0.0199)
$\alpha_2(m)$	0.0330 (0.0291)	-0.1784*** (0.0250)	0.0317 (0.0284)	0.0055 (0.0323)	-0.1742*** (0.0229)	-0.2371*** (0.0214)
$\beta(m)$	0.9909*** (0.0044)	0.9784*** (0.0069)	0.9029*** (0.0150)	1.0604*** (0.0114)	1.1183*** (0.0193)	0.8992*** (0.0132)
Baseline Parameters						
$w(m)$	-5.5242*** (0.2158)	7.2711*** (0.2439)	-5.3643*** (0.0660)	-6.0901*** (0.0790)	6.1217*** (0.1868)	8.3208*** (0.1897)
$a(m)$	1.1561*** (0.0229)	2.4276*** (0.0535)	1.3925*** (0.0357)	1.2394*** (0.0289)	3.1820*** (0.0783)	2.6130*** (0.0560)
Other Parameters						
$\eta_{vol}(m)$		-2.3027*** (0.0644)			-2.7864*** (0.0836)	-2.6960*** (0.0647)
π_{1m}			0.4611*** (0.0190)		0.2646*** (0.0184)	
π_{2m}				0.5277*** (0.0178)		0.5347*** (0.0166)
Diagnostic Statistics						
Obs. In $\hat{s}_i = m$			673	802	540	935
$E[\hat{\epsilon}_i]$	0.9953	0.9906	1.0040		0.9658	
$\sigma[\hat{\epsilon}_i]$	1.1048	1.0784	1.0090		0.9508	
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y} \hat{\mathbb{S}})$	-9779.12	-8715.73	-9555.70		-8481.35	
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}, \hat{\mathbb{S}})$			-10571.1		-9414.42	
BIC	13.2846	11.8476	13.0162		11.5694	
CvM	0.5043**	0.3700*	0.0716		0.1529	
AD	3.7033**	2.3386*	0.8978		1.2566	
LB(50)	41.1280	33.6644	40.1028		53.1705*	
\widehat{SoR}			0.6892		0.7298	
$\widehat{SoR}(m)$			0.6493	0.7355	0.6581	0.7598

Note: Standard errors are in parentheses. ***, ** and * represent significance at 1%, 5% and 10% respectively. Observation counts and residual statistics are based on the estimated most probable state sequence. BIC is computed as $BIC = T^{-1}(k \ln T - 2 \ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}|\hat{\mathbb{S}}))$ where T is the number of observations. The CvM and AD are Cramér-von-Mises and Andersen-Darling statistics (Stephens, 2013) for unit exponential distribution, with the critical values bootstrapped from 100000 simulated unit exponential vectors. LB(50) is the Ljung-Box (Ljung and Box, 1978) test statistics at lag 50. Definitions of \widehat{SoR} and $\widehat{SoR}(m)$ can be found in (3.42) and (3.43).

Table C.5 Summary of Cramér-von-Mises, Andersen-Darling and Ljung-Box test results

Cramér-von-Mises Test												
Stocks	ACI(2,1)			ACI(2,1)-V			MS(2)-ACI(2,1)			MS(2)-ACI(2,1)-V		
$\alpha\%$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
AIG	4	8	10	12	12	12	0	0	1	1	5	7
CVX	7	10	10	12	12	12	0	1	2	2	4	5
GM	6	10	12	12	12	12	0	1	3	0	0	2
INTC	2	7	10	12	12	12	0	1	2	1	7	8
JPM	6	8	10	12	12	12	0	1	1	1	4	8
PFE	4	6	8	12	12	12	1	2	3	3	3	5
SPY	4	10	11	4	6	8	0	2	3	2	4	4
T	4	8	9	12	12	12	0	1	2	3	6	7
VZ	7	10	11	12	12	12	0	0	0	4	8	9
WMT	0	6	9	12	12	12	0	1	3	2	5	6
Average	4.4	8.3	10	11.2	11.4	11.6	0.1	1	2	1.9	4.6	6.1

Andersen-Darling Test												
Stocks	ACI(2,1)			ACI(2,1)-V			MS(2)-ACI(2,1)			MS(2)-ACI(2,1)-V		
$\alpha\%$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
AIG	8	10	12	12	12	12	0	4	5	1	6	7
CVX	9	10	11	12	12	12	1	5	6	2	4	5
GM	9	12	12	12	12	12	0	2	5	0	0	3
INTC	3	10	10	12	12	12	1	3	5	0	6	8
JPM	8	11	12	12	12	12	0	3	5	3	4	7
PFE	6	8	11	12	12	12	1	3	3	3	3	6
SPY	8	11	11	4	7	9	0	2	3	2	4	4
T	7	10	10	12	12	12	0	2	2	2	6	7
VZ	8	12	12	12	12	12	0	1	3	5	9	9
WMT	6	10	10	12	12	12	0	2	4	2	6	6
Average	7.2	10.4	11.1	11.2	11.5	11.7	0.3	2.7	4.1	2	4.8	6.2

Ljung-Box Test at Lag 50												
Stocks	ACI(2,1)			ACI(2,1)-V			MS(2)-ACI(2,1)			MS(2)-ACI(2,1)-V		
$\alpha\%$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
AIG	8	10	10	4	4	4	1	2	3	2	4	6
CVX	6	8	9	9	9	9	0	1	2	4	9	9
GM	7	9	9	8	10	10	0	2	3	6	6	7
INTC	10	12	12	11	12	12	0	2	2	3	8	8
JPM	7	7	7	6	6	6	2	3	3	5	7	10
PFE	8	10	10	8	8	9	0	1	3	1	2	4
SPY	4	5	6	6	8	9	1	3	5	3	4	5
T	6	7	8	7	7	7	0	1	1	1	4	4
VZ	8	8	9	5	5	6	1	3	3	4	7	7
WMT	8	9	9	7	7	7	2	3	3	2	5	8
Average	7.2	8.5	8.9	7.1	7.6	7.9	0.7	2.1	2.8	3.1	5.6	6.8

Note: $\alpha\%$ is the significance level of the test. Each number at $\alpha\%$ for each stock represents the number of times the test is significant at $\alpha\%$ in the year 2016.

Table C.6 Estimation Outputs of MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-8 and SPY 2016-05

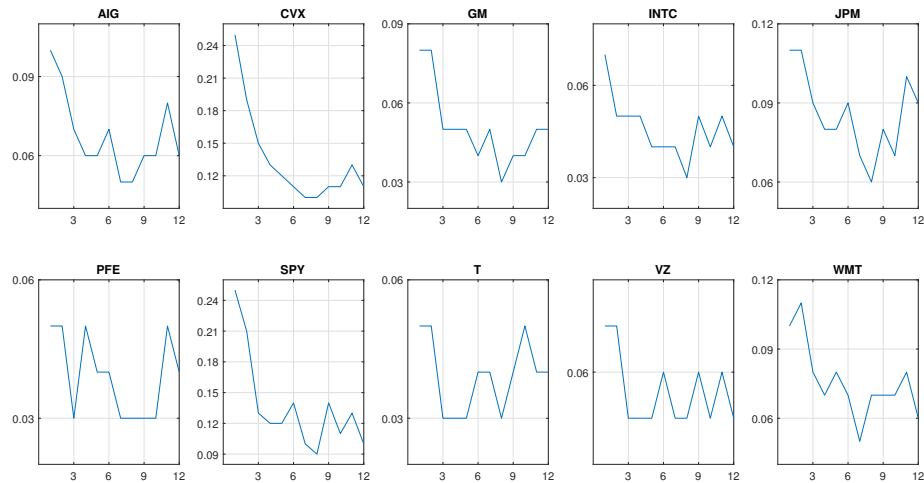
$m =$	AIG 2016-03			INTC 2016-08			SPY 2016-05		
	1	2	3	1	2	3	1	2	3
Dynamic Parameters									
$\alpha_1(m)$	0.6637*** (0.0527)	0.1625*** (0.0210)	-0.1066*** (0.0381)	0.4245*** (0.0491)	0.2665*** (0.0253)	0.1623*** (0.0247)	0.2655*** (0.0236)	0.4978*** (0.0217)	-0.0315 (0.0485)
$\alpha_2(m)$	-0.2271*** (0.0402)	-0.0377* (0.0222)	-0.1012*** (0.0386)	0.3540*** (0.0526)	0.0726** (0.0343)	-0.1398*** (0.0256)	-0.2030*** (0.0243)	-0.1995*** (0.0188)	-0.2091*** (0.0245)
$\beta(m)$	1.1296*** (0.0304)	0.9486*** (0.0081)	0.5922*** (0.0574)	0.1456 (0.1182)	0.4494*** (0.1001)	1.0010*** (0.0090)	0.8464*** (0.0095)	1.1216*** (0.0143)	0.9656*** (0.0126)
Baseline Parameters									
$w(m)$	2.7364*** (0.3082)	3.0557*** (0.1138)	-0.4410 (0.5721)	-1.4031** (0.6173)	6.8810*** (0.1708)	2.4194*** (0.1915)	10.0468*** (0.3057)	7.5401*** (0.1396)	5.2086*** (0.1170)
$a(m)$	1.8675*** (0.0768)	2.9142*** (0.0588)	0.9089*** (0.0478)	0.5591*** (0.0290)	2.7982*** (0.0803)	2.1172*** (0.0573)	2.4884*** (0.1614)	3.6700*** (0.2991)	4.4428*** (0.4239)
Other Parameters									
$\eta_{vol}(m)$	-1.3669*** (0.0778)	-2.3522*** (0.0555)	-0.3934*** (0.0546)	-1.1110*** (0.0301)	-3.0383*** (0.0976)	-1.5153*** (0.0553)	-2.8758*** (0.1592)	-3.5892*** (0.3022)	-3.6269*** (0.3338)
π_{1m}	0.4951*** (0.0366)	0.0102*** (0.0028)	0.4840*** (0.0388)	0.8093*** (0.0285)	0.0366*** (0.0072)	0.0149*** (0.0043)	0.5695*** (0.0201)	0.3804*** (0.0222)	0.2236*** (0.0206)
π_{2m}	0.2886*** (0.0339)	0.9463*** (0.0063)	0.0982*** (0.0237)	0.0032 (0.0035)	0.8764*** (0.0130)	0.1096*** (0.0118)	0.3744*** (0.0197)	0.4300*** (0.0230)	0.0875*** (0.0140)
Diagnostic Statistics									
Obs. in $\hat{s}_j = m$	212	1300	156	180	676	814	648	472	355
$E[\hat{\epsilon}_j]$	0.9829			1.0063			0.9843		
$\sigma[\hat{\epsilon}_j]$	1.0381			1.0056			0.9536		
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y} \hat{\mathbb{S}})$	-9788.59			-9858.42			-8269.41		
$\ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}, \hat{\mathbb{S}})$	-10436.3			-10512.9			-9527.54		
BIC	11.8437			11.9131			11.3315		
CvM	0.2586			0.1349			0.0292		
AD	1.6334			0.7422			0.2689		
LB(50)	44.1905			59.9984***			48.1202*		
\widehat{SoR}	0.924			0.9339			0.7501		
$\widehat{SoR}(m)$	0.7621	0.9645	0.8011	0.975	0.9122	0.9439	0.7405	0.6793	0.8479

Note: Standard errors are in parentheses. ***, ** and * represent significance at 1%, 5% and 10% respectively. Observation counts and residual statistics are based on the estimated most probable state sequence. BIC is computed as $BIC = T^{-1}(k \ln T - 2 \ln \mathcal{L}(\hat{\vartheta}; \mathbb{Y}|\hat{\mathbb{S}}))$ where T is the number of observations. The CvM and AD are Cramér-von-Mises and Andersen-Darling statistics (Stephens, 2013) for unit exponential distribution, with the critical values bootstrapped from 100000 simulated unit exponential vectors. LB(50) is the Ljung-Box (Ljung and Box, 1978) test statistics at lag 50. Definitions of \widehat{SoR} and $\widehat{SoR}(m)$ can be found in (3.42) and (3.43).

Table C.7 Comparison of regime classification of the MS(2)-ACI(2,1)-V and the MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-08 and SPY 2016-05

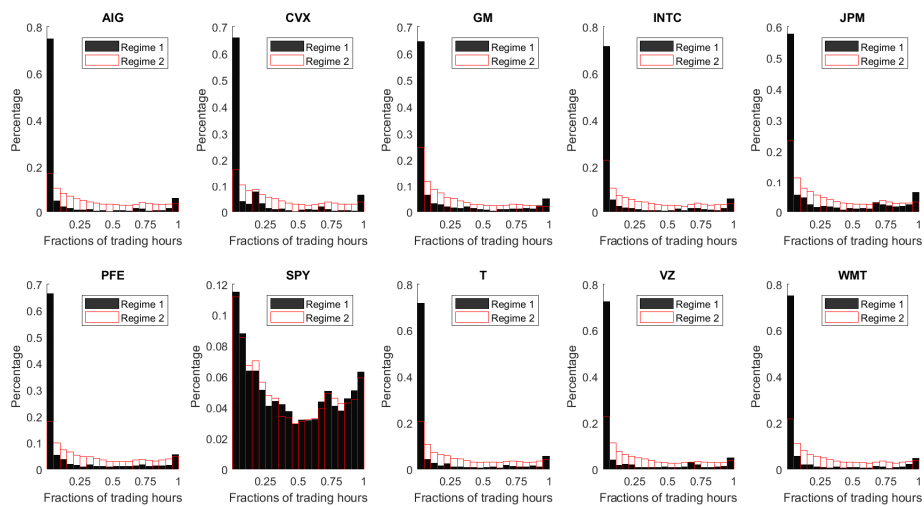
MS(2)-ACI(2,1)-V model										
	$m =$	AIG 2016-03			INTC 2016-08			SPY 2016-05		
		1	2	Sum	1	2	Sum	1	2	Sum
MS(3)-	1	91	121	212	166	14	180	187	461	648
ACI(2,1)	2	51	1249	1300	38	638	676	211	261	472
-V model	3	94	62	156	69	745	814	142	213	355
	Sum	236	1432	1668	273	1397	1670	540	935	1475

Note: This table presents the classification of observations under the estimated most probable state vector of the MS(2)-ACI(2,1)-V and the MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-08 and SPY 2016-05. E.g. the first entry, 91, represents the number of observations classified as regime 1 in MS(2)-ACI(2,1)-V and stayed in regime 1 in the MS(3)-ACI(2,1)-V model.

Figure C.3 Monthly choices of δ for 10-by-12 stock-month datasets

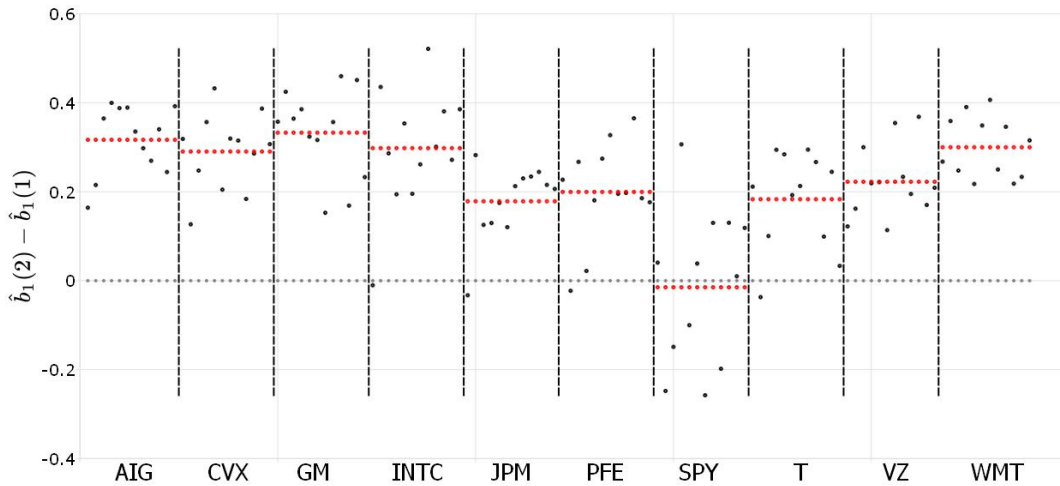
Note: the x-axis denotes the months and the y-axis gives the value of δ .

Figure C.4 Yearly distribution of estimated regimes over time for all stock-months for the MS(2)-ACI(2,1)-V model



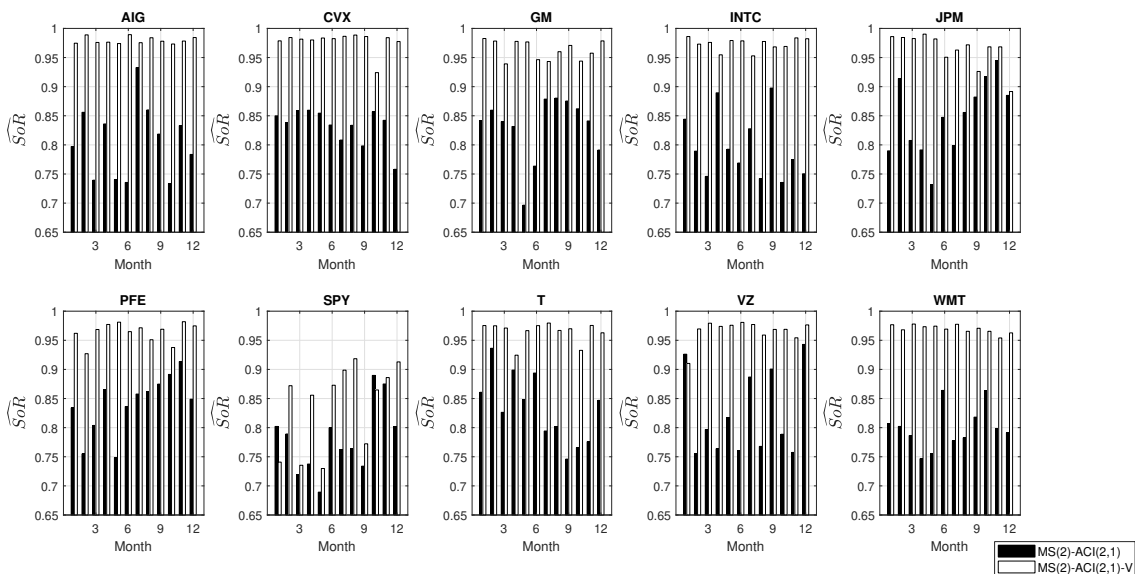
Note: The x-axis represents fraction of trading hours. The y-axis is the percentage of data in each regime. Regime 1 and 2 correspond to the most probable state vector estimated by the MS(2)-ACI(2,1)-V model. Each bar counts the percentage of data falling into a roughly 20-minute bin.

Figure C.5 Difference in the estimated \hat{b}_1 s from the volume-duration regressions between observations in regime 1 and 2 using the estimated state vector



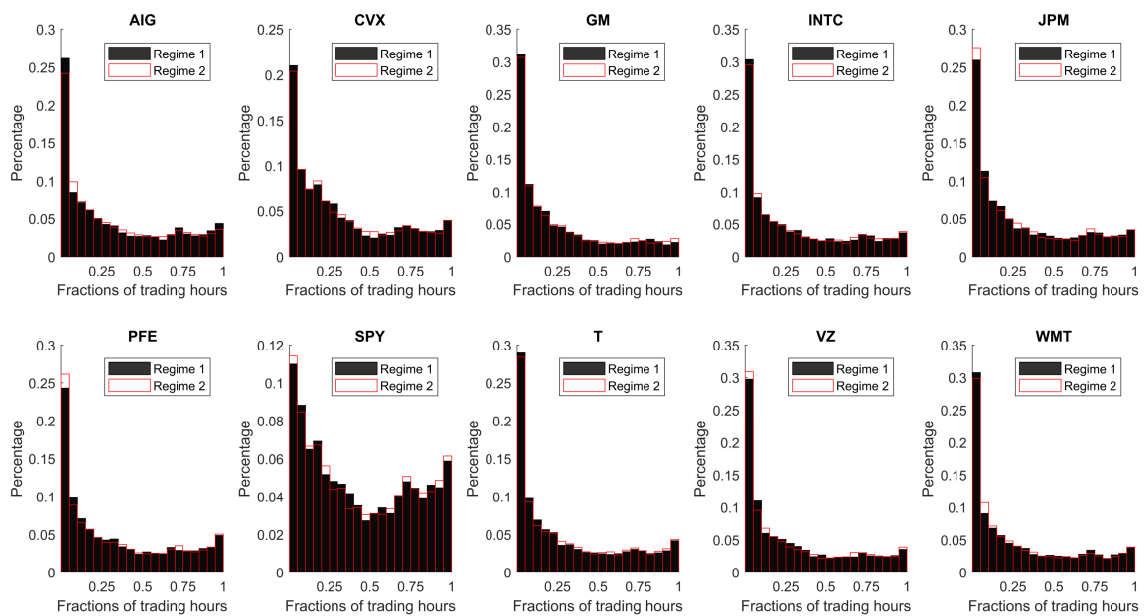
Note: The figure plots $\hat{b}_1(2) - \hat{b}_1(1)$ obtained from regression (3.45) for 10×12 stock-month datasets, with $\hat{b}_1(m)$ being the estimated \hat{b}_1 from the observations classified as regime m . The estimated state vector is obtained by the MS(2)-ACI(2,1)-V model. Each black dot represents the \hat{b}_1 difference for one stock-month dataset. The vertical black dashed lines split observations from each stock, and between two vertical red lines, the \hat{b}_1 differences are ordered chronologically.

Figure C.6 Estimated \widehat{SoR} for MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models



Note: The x-axis denotes the months from which the \widehat{SoR} is estimated. The definition of \widehat{SoR} can be found in (3.42).

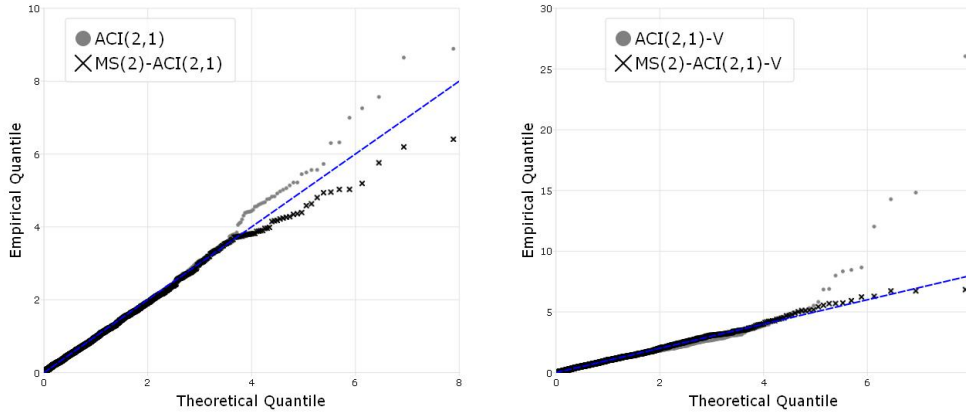
Figure C.7 Yearly distribution of estimated regimes over time for all stock-months for the MS(2)-ACI(2,1) model



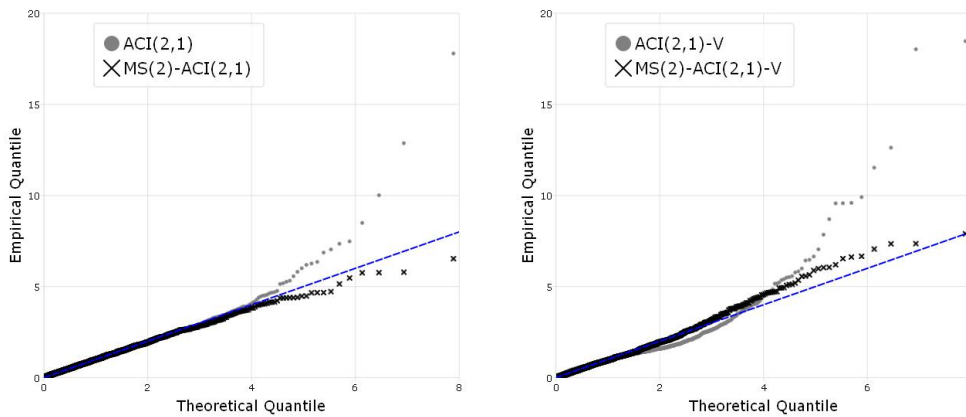
Note: The x-axis represents fraction of trading hours. The y-axis is the percentage of data in each regime. Regime 1 and 2 correspond to the most probable state vector estimated by the MS(2)-ACI(2,1) model. Each bar counts the percentage of data falling into a roughly 20-minute bin.

Figure C.8 Quantile-Quantile Plots of the residuals obtained from ACI(2,1), ACI(2,1)-V, MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models for AIG 2016-03, INTC 2016-08 and SPY 2016-05

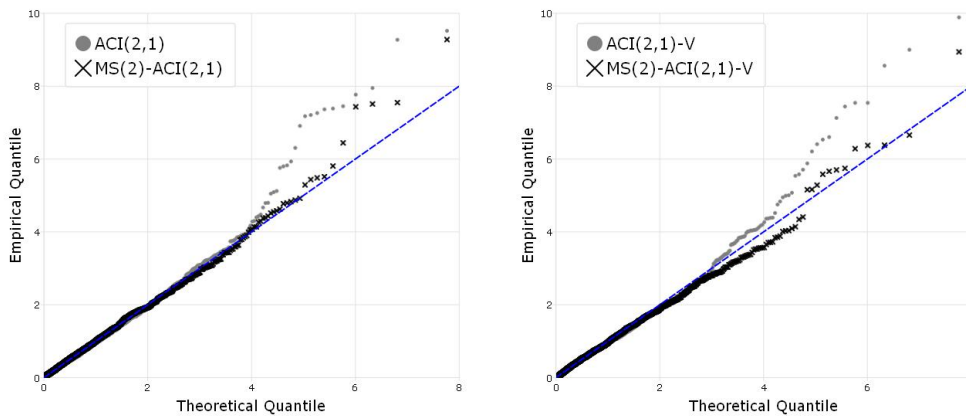
Panel 1: AIG 2016-03



Panel 2: INTC 2016-08



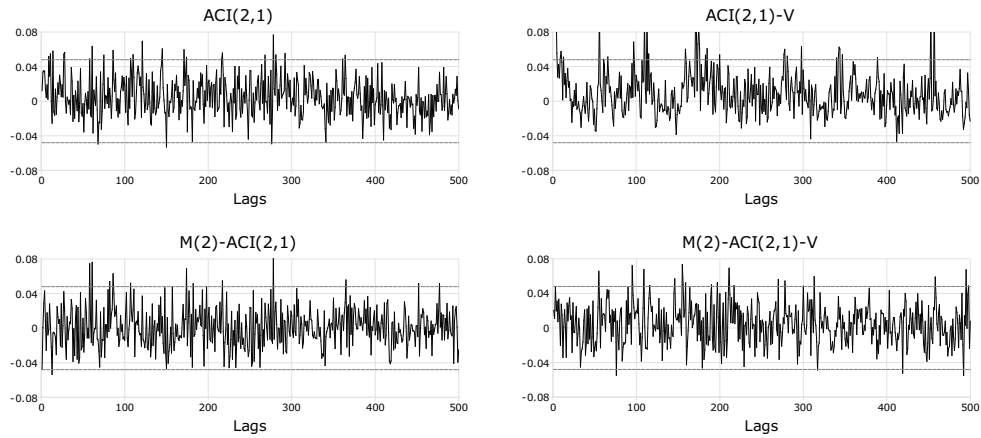
Panel 3: SPY 2016-05



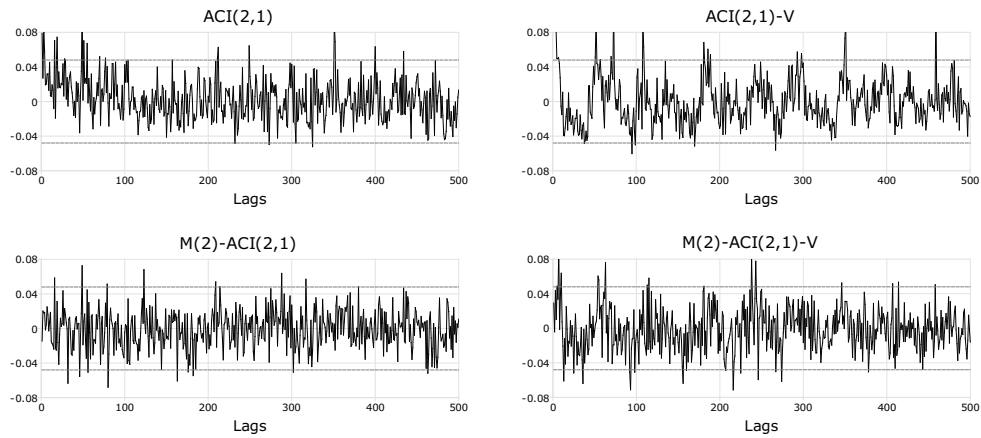
Note: The blue dashed lines represents the line $Y = X$. The residuals are obtained from estimation outputs in Tables C.2 to C.4.

Figure C.9 Correlograms of the residuals obtained from ACI(2,1), ACI(2,1)-V, MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models for AIG 2016-03, INTC 2016-08 and SPY 2016-05

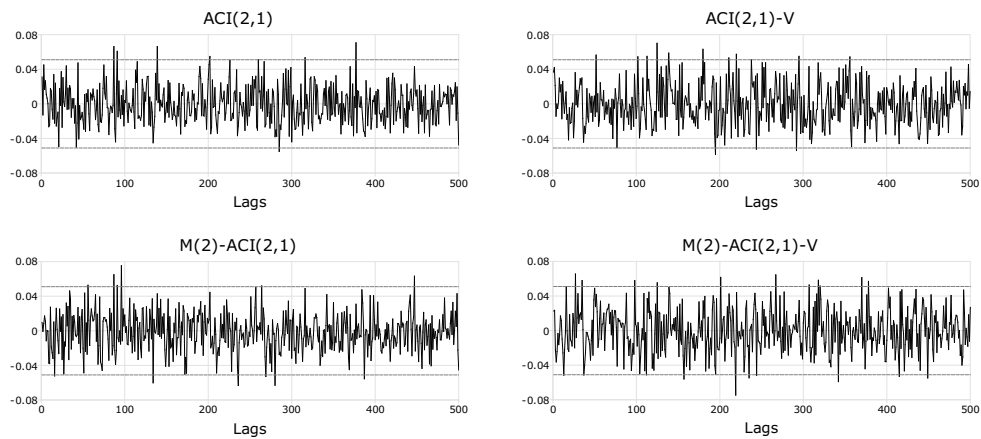
Panel 1: AIG 2016-03



Panel 2: INTC 2016-08

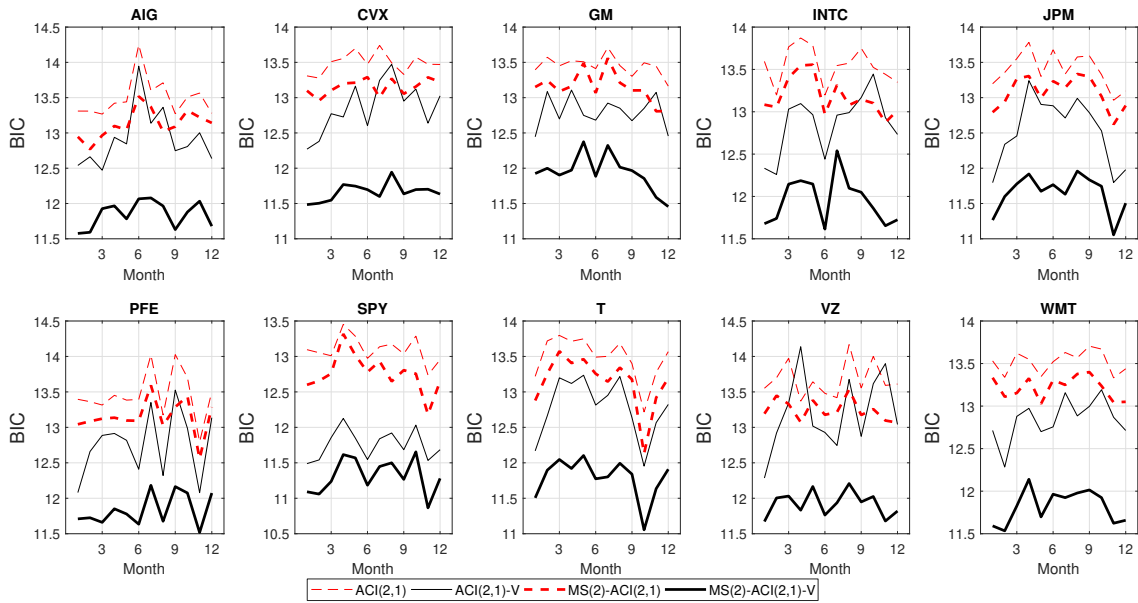


Panel 3: SPY 2016-05



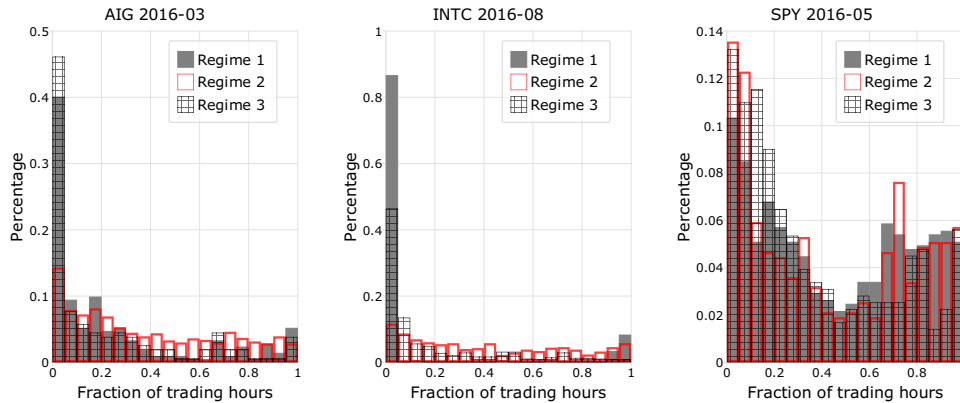
Note: The grey dashed lines are the 95% confidence bounds. The residuals are obtained from estimation outputs in Tables C.2 to C.4.

Figure C.10 Bayesian Information Criterion of ACI(2,1), ACI(2,1)-V, MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V models for all stock-months



Note: BIC is computed as $BIC = T^{-1}(k \ln T - 2 \ln \mathcal{L})$. For ACI(2,1) and ACI(2,1)-V, the maximized log-likelihood is used. For MS(2)-ACI(2,1) and MS(2)-ACI(2,1)-V, the conditional likelihood $\mathcal{L}(\hat{\phi}; \mathbb{Y}|\mathbb{S})$ is used.

Figure C.11 Distribution of estimated regimes over time from the MS(3)-ACI(2,1)-V model for AIG 2016-03, INTC 2016-08 and SPY 2016-05



Note: The x-axis represents fraction of trading hours with. The y-axis is the percentage of data in each regime. Regime 1, 2 and 3 correspond to the most probable state vector estimated in Table C.6. Each bar counts the percentage of data falling into a roughly 20-minute bin.