

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Economic Relevance	2
1.3	Scientific Relevance	5
1.4	Literature Gap	6
1.5	Contents	7
2	General Framework	11
2.1	Classes of Volatility Models	11
2.1.1	Time-Invariant Volatility Models	11
2.1.2	Stochastic Volatility Models	12
2.1.3	Integrated Volatility	14
2.2	Framework to Model the Integrated Volatility	15
2.2.1	Modelling Asset Prices	15
2.2.2	Modelling Integrated Volatility	18
2.2.3	Quadratic Variation of Stochastic Integrals	19
2.3	Sampling and Pre-processing Intraday Data	19
2.3.1	Sampling Schemes	20
2.3.2	Data Sampling	21
2.3.3	Data Pre-processing	23
3	Integrated Volatility	29
3.1	Framework to Estimate Integrated Volatility	29
3.1.1	Stylized Facts of Volatility	29
3.1.2	Microstructure Noise	31
3.1.3	Assumptions on the Noise Process	32
3.2	Estimating Integrated Volatility	32
3.2.1	Canonical Estimator	33
3.2.2	Averaging Estimator	35
3.2.3	Two-Time-Scales Estimator	37
3.2.4	Adjusted Two-Time-Scales Estimator	38
3.2.5	Area Adjusted Two-Time-Scales Estimator	39
3.3	Estimating Interday Periodicity	40
3.3.1	Modelling Interday Periodicity	40
3.3.2	Non-Parametric Periodicity Estimation	43
3.3.3	Parametric Periodicity Estimation	43

4	Zero-inflated Data Generation Processes	47
4.1	Descriptive Analysis	47
4.1.1	Data Description	47
4.1.2	Data Presentation	49
4.1.3	Characterising Zero-inflated Data	50
4.2	Zero-inflated Models	53
4.2.1	Semi-continuous Data	53
4.2.2	Discrete Data	54
4.2.3	Continuous Data	55
4.3	Modelling Zero-inflated Volatility Estimates	55
4.3.1	Markov Transition Probabilities	56
4.3.2	Markov Switching Mixture Model	58
4.3.3	Model Estimation and Computation	61
5	Volatility Shock Causing Incidents	65
5.1	Filtering Ad-hoc News	65
5.1.1	Selected Sample	66
5.1.2	Identifying the Data Generation Process	67
5.1.3	Testing for Volatility Shocks	72
5.2	Quantifying Financial Text Data	75
5.2.1	Sampling	75
5.2.2	Pre-processing	77
5.2.3	Quantification	78
5.3	Predictive Text Mining	82
5.3.1	The Pattern Recognition Task	82
5.3.2	Unsupervised Learning	84
5.3.3	Supervised Learning	85
6	Algorithmic Text Forecasting	87
6.1	Classification Algorithms	87
6.1.1	Multinomial Regression	88
6.1.2	Single-Hidden-Layer Neural Networks	88
6.1.3	Stabilized Linear Discriminant Analysis	89
6.2	Ensemble Learning Algorithms	90
6.2.1	Boosting	90
6.2.2	Bagging	91
6.2.3	Random Forests	92
6.3	Real-Time Classification	93
6.3.1	Labelling the Sample	93
6.3.2	Training the Algorithms	93
6.3.3	Classifying Unexpected Incidents	94

7	Benchmarking	99
7.1	Cross-Validation	99
7.1.1	K-Fold Cross-Validation	100
7.1.2	Cross-Validating the Algorithms	101
7.1.3	Testing for the Best Fitting Algorithms	105
7.2	Confusion Analytics	110
7.2.1	Confusion Matrix	110
7.2.2	Fail Forecasts and Economic Scenarios	110
7.2.3	Precision, Recall, F-measure, and AUC	112
7.3	Random Observations	119
7.3.1	Request	119
7.3.2	Choosing Random Sample	120
7.3.3	Determining the Monitoring Benchmark	121
8	Monitoring	123
8.1	System Evaluation	124
8.1.1	Effect of Ad-Hoc News	124
8.1.2	Discriminating Shocks from Non-Shocks	126
8.1.3	Errors of Shocks and Errors of Non-Shocks	127
8.2	Variation in Identification	131
8.2.1	Ask and Bid Quotes	132
8.2.2	Volatility Estimators	134
8.2.3	Coinciding Classification Algorithms	135
8.3	Economic Evaluation	137
8.3.1	Costs of Misclassification	137
8.3.2	Evaluating the Trading Scenario	139
8.3.3	Evaluating the Risk Management Scenario	144
9	Conclusion	147
9.1	Primary Findings	147
9.2	Weaknesses and Limitations	150
9.3	Proposals to Future Research	151
A	Software Libraries	155
A.1	R	155
A.2	C++	157
	Bibliography	159