

©Copyright 2012

Brian Donhauser



# Jump Variation From High-Frequency Asset Returns: New Estimation Methods

Brian Donhauser

A dissertation  
submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Eric W. Zivot, Chair

Donald B. Percival

Yu-chin Chen

Program Authorized to Offer Degree:  
Department of Economics



University of Washington

**Abstract**

Jump Variation From High-Frequency Asset Returns:  
New Estimation Methods

Brian Donhauser

Chair of the Supervisory Committee:  
Robert Richards Chaired Professor Eric W. Zivot  
Department of Economics

A large literature has emerged in the last 10 years using high-frequency (intraday) asset returns to estimate lower-frequency phenomena, several of which being conditional daily return variance and its components jump variation and integrated variance. We propose several new estimators of jump variation and integrated variance. We base the first set on the jump detection work of [Lee and Mykland \(2008\)](#). Simply, these estimate jump variation by summing the hard-thresholded or naively shrunk squared returns (and estimate integrated variance as the residual of realized variance). In the second set, we appeal to [Johnstone and Silverman \(2004, 2005\)](#) for their discrete Bayesian model of a sparse signal plus noise and argue for its use as a model of high-frequency asset returns with jumps. Within this model, we derive optimal estimators of jump variation and integrated variance. In a simulation analysis, we find outperformance of our estimators against the established. In an empirical analysis we find dramatically different estimates of jump variation and integrated variance based on the estimation scheme, suggesting careful attention must be paid to the method of decomposing conditional daily return variance into jump variation and integrated variance.



# CONTENTS

Contents . . . . .	iii
List of Figures . . . . .	v
List of Tables . . . . .	vi
List of Definitions, Assumptions, Models, Theorems, and Results . . . . .	viii
Chapter 1: An Overview of High-Frequency Finance . . . . .	1
1.1 Data Errors and Cleaning . . . . .	1
1.2 Empirical Characteristics . . . . .	2
1.3 Market Microstructure Noise . . . . .	4
1.4 Trading Process Frictions . . . . .	5
1.5 Informational Effects . . . . .	6
1.6 Structural Market Microstructure Models . . . . .	7
1.7 High-Frequency Financial Econometric Models . . . . .	7
1.8 Integrated Variance and Jump Variation from High-Frequency Financial Asset Returns . . . . .	8
Chapter 2: $JV$ and $IV$ Estimation in a Continuous Semimartingale Model . . . . .	14
2.1 Toward the $SM+$ Model . . . . .	14

2.2	<i>IV</i> and <i>JV</i> Estimation in Previous Work . . . . .	31
2.3	Jump Location Estimation in Previous Work . . . . .	38
2.4	<i>JV</i> and <i>IV</i> Estimation via Jump Location Estimation . . . . .	46
2.5	The Simulation Models . . . . .	52
2.6	Estimator Evaluation Criteria for Simulation Models . . . . .	52
2.7	Simulation Results . . . . .	53
2.8	Empirical Results . . . . .	58
2.9	Conclusion . . . . .	61
Chapter 3: <i>JV</i> and <i>IV</i> Estimation in a Discrete Bayesian Model . . . . .		67
3.1	Discrete Model from Continuous . . . . .	67
3.2	The Bayesian Model . . . . .	70
3.3	Theoretical Results in the Bayesian Model . . . . .	72
3.4	Jump Variation in the Bayesian Model . . . . .	80
3.5	The Simulation Models . . . . .	86
3.6	Estimator Evaluation Criteria for Simulation Models . . . . .	89
3.7	Simulation Results . . . . .	91
3.8	Empirical Results . . . . .	96
3.9	Conclusion . . . . .	100
Bibliography . . . . .		106

## LIST OF FIGURES

1.1	WMT, Jan 02, 2008: 1-minute prices $S_i$ and log-returns $r_i$ . . . . .	9
2.1	Time series: $\widehat{\mathcal{J}}V_{\text{BNS04}}$ , $\widehat{\mathcal{J}}V_{\text{NS}}$ , for WMT . . . . .	64
2.2	Scatter plot: $\widehat{\mathcal{J}}V_{\text{NS}}$ vs. $\widehat{\mathcal{J}}V_{\text{BNS04}}$ for WMT . . . . .	65
3.1	Johnstone (2011) Illustration . . . . .	75
3.2	Time series: 250 simulated returns . . . . .	78
3.3	Scatter plot: 250 simulated $ \mu_i $ vs. $ \hat{\mu}(x_i) $ . . . . .	79
3.4	Time series: $\widehat{\mathcal{J}}V_{\text{BNS04}}$ , $\widehat{\mathcal{J}}V_{\text{EB},\ell_2}$ , for WMT . . . . .	103
3.5	Scatter plot: $\widehat{\mathcal{J}}V_{\text{EB},\ell_2}$ vs. $\widehat{\mathcal{J}}V_{\text{BNS04}}$ for WMT . . . . .	104
3.6	Scatter plot: $\widehat{\mathcal{J}}V_{\text{EB},\ell_2}$ vs. $\widehat{\mathcal{J}}V_{\text{NS}}$ for WMT . . . . .	105

## LIST OF TABLES

2.1	Example values for thresholds $t_l$ and $t_r$ . . . . .	45
2.2	MPE for $s = 4$ . . . . .	56
2.3	MAPE for $s = 4$ . . . . .	57
2.4	MPE for $s = 7$ . . . . .	58
2.5	MAPE for $s = 7$ . . . . .	59
2.6	MPE for $s = 10$ . . . . .	60
2.7	MAPE for $s = 10$ . . . . .	61
2.8	MPE for $s = 15$ . . . . .	62
2.9	MAPE for $s = 15$ . . . . .	63
2.10	Summary: $\widehat{JV}_{\text{BNS04}}$ , $\widehat{JV}_{\text{NS}}$ for WMT . . . . .	66
3.1	MPE for $s = 4$ . . . . .	93
3.2	MAPE for $s = 4$ . . . . .	94
3.3	MPE for $s = 7$ . . . . .	95
3.4	MAPE for $s = 7$ . . . . .	96
3.5	MPE for $s = 10$ . . . . .	97
3.6	MAPE for $s = 10$ . . . . .	98
3.7	MPE for $s = 15$ . . . . .	99
3.8	MAPE for $s = 15$ . . . . .	100

3.9	$\widehat{\mathcal{J}}_{\text{EB},\ell_2}, \widehat{I}_{\text{EB},\ell_2}$ , across volatility estimation windows, WMT . . . . .	101
3.11	Summary: $\widehat{\mathcal{J}}_{\text{BNS04}}, \widehat{\mathcal{J}}_{\text{NS}}, \widehat{\mathcal{J}}_{\text{EB},\ell_2}$ , for WMT . . . . .	102

# LIST OF DEFINITIONS, ASSUMPTIONS, MODELS, AND THEOREMS

Definition 2.1.1	General stochastic process: $\mathcal{GSP}$ . . . . .	14
Definition 2.1.2	Log-returns: $r$ . . . . .	15
Definition 2.1.3	Partition: $\mathcal{P}$ . . . . .	16
Definition 2.1.4	Log-returns on a partition: $r$ . . . . .	16
Definition 2.1.5	Semimartingale: $\mathcal{SM}$ . . . . .	17
Theorem 2.1.6	No arbitrage $\rightarrow \mathcal{SM}$ . . . . .	17
Definition 2.1.7	Mesh: $\ \mathcal{P}\ $ . . . . .	18
Definition 2.1.8	Quadratic Variation: $[X]$ . . . . .	18
Definition 2.1.9	Realized Variance: $RV$ . . . . .	19
Theorem 2.1.10	$RV \rightarrow [X]$ . . . . .	20
Definition 2.1.11	Special semimartingale: $\mathcal{SSM}$ . . . . .	20
Theorem 2.1.12	$\mathcal{SSM} : X = A + M$ canonical . . . . .	20
Theorem 2.1.13	$\mathcal{SSM} \rightarrow$ local risk premia exist . . . . .	21
Theorem 2.1.14	$\mathcal{SM} +$ bounded jumps $\rightarrow \mathcal{SSM}$ . . . . .	21
Theorem 2.1.15	$\mathcal{SM}^c \rightarrow \mathcal{SSM}^c$ . . . . .	22
Assumption 2.1.16	$\sum  \Delta X  < \infty$ . . . . .	22
Theorem 2.1.17	$X = X^c + J$ . . . . .	23

Theorem 2.1.18	$\mathcal{SM} : X = A^c + M^c + J$ . . . . .	23
Definition 2.1.19	Jump variation: $JV$ . . . . .	24
Theorem 2.1.20	$\mathcal{SM} : [X] = [X^c] + \Sigma(\Delta X)^2$ . . . . .	25
Theorem 2.1.21	$\mathcal{SM}^c : [X] = [M]$ . . . . .	25
Theorem 2.1.22	$\mathcal{M}_{loc}^2 : E\{M^2\} = E\{[M]\}$ . . . . .	26
Theorem 2.1.23	$\mathcal{SSM}, \mathcal{DET} : \text{Var}\{r\} = E\{[X]\}$ . . . . .	27
Definition 2.1.24	$\mathcal{SM}+, \mathcal{SM}^{c+}$ . . . . .	28
Theorem 2.1.25	$\mathcal{M}_{loc}^2 : M = \int \sigma dW$ . . . . .	28
Theorem 2.1.26	$\mathcal{FV}^c : A^c = \int a du$ . . . . .	29
Theorem 2.1.27	$\mathcal{SM}+ : X = \int a du + \int \sigma dW + J$ . . . . .	29
Theorem 2.1.28	$\mathcal{SM}^{c+} : X = \int a du + \int \sigma dW$ . . . . .	30
Theorem 2.2.1	$\mathcal{M}_{loc}^2 : [M] = \int \sigma^2 dt$ . . . . .	31
Definition 2.2.2	Integrated variance: $IV$ . . . . .	32
Theorem 2.2.3	$\mathcal{SM}+ : RV \rightarrow IV + JV$ . . . . .	33
Theorem 2.2.4	$\mathcal{SM}^{c+} : RV \rightarrow IV$ . . . . .	34
Theorem 2.2.5	$\mathcal{SM}^{c+} : RV \rightarrow \text{Var}\{r\}$ . . . . .	34
Theorem 2.2.6	$\mathcal{SM}+ : RV \rightarrow \text{Var}\{r\} + \text{error}$ . . . . .	35
Definition 2.2.7	Semimartingale plus plus: $\mathcal{SM}++$ . . . . .	36
Definition 2.2.8	$RBPV$ . . . . .	36
Definition 2.2.9	$\widehat{IV}, \widehat{JV}$ . . . . .	37
Theorem 2.2.10	$\mathcal{SM}++ : \widehat{IV} \rightarrow IV, \widehat{JV} \rightarrow JV$ . . . . .	37
Definition 2.3.1	$LM_l$ . . . . .	39

Theorem 2.3.2	$\mathcal{SM}++$ : convergence of $LM_t$ . . . . .	40
Theorem 2.3.3	Jump location estimation: local . . . . .	41
Definition 2.3.4	$LM_g$ . . . . .	42
Theorem 2.3.5	$\mathcal{SM}++$ : convergence of $LM_g$ . . . . .	42
Definition 2.3.6	$\beta_\alpha$ . . . . .	43
Theorem 2.3.7	Jump location estimation: global . . . . .	43
Definition 2.3.8	$\mathbf{t}_t, \mathbf{t}_g$ . . . . .	44
Definition 2.3.9	$\mathcal{J}, \mathcal{I}, N, \widehat{\mathcal{I}}, \widehat{N}$ . . . . .	44
Theorem 2.3.10	Jump location estimation: alternative . . . . .	46
Definition 2.4.1	$\widehat{\mathcal{J}}_{\text{HT}}, \widehat{\mathcal{I}}_{\text{HT}}$ . . . . .	47
Theorem 2.4.2	$\widehat{\mathcal{J}}_{\text{HT}} \rightarrow \mathcal{J}$ . . . . .	47
Definition 2.4.3	$\widehat{\mathcal{J}}_{\text{NS}}, \widehat{\mathcal{I}}_{\text{NS}}$ . . . . .	49
Theorem 2.4.4	$0 \leq \widehat{\mathcal{J}}_{\text{NS}} \leq RV$ . . . . .	49
Theorem 2.4.5	$\widehat{\mathcal{J}}_{\text{NS}} \rightarrow \mathcal{J}$ . . . . .	50
Assumption 3.1.1	. . . . .	68
Definition 3.1.2	Continuous $\rightarrow$ Discrete . . . . .	68
Model 3.1.3	Discrete, Stochastic Volatility . . . . .	68
Definition 3.1.4	Discrete Approximations/Definitions for $IV, JV, QV$ . . . . .	69
Assumption 3.1.5	Constant Volatility . . . . .	69
Result 3.1.6	Simplified, Redefined Interval Volatility . . . . .	70
Model 3.1.7	Discrete, Constant Volatility . . . . .	70
Model 3.2.1	Discrete, Stochastic Volatility, Empirical Bayesian . . . . .	70

Model 3.2.2	Discrete, Constant Volatility, Empirical Bayesian . . . . .	71
Result 3.3.1	Marginal Density of $x   \{\mu \neq 0\}$ : $g(x)$ . . . . .	72
Result 3.3.2	Marginal Density of $x$ : $p(x)$ . . . . .	72
Result 3.3.3	Posterior Density of $\mu   x$ : $\pi(\mu   x)$ . . . . .	73
Result 3.3.4	Posterior Non-Zero Jump Probability: $w(x)$ . . . . .	73
Result 3.3.5	Posterior Density of $\mu   (x, \mu \neq 0)$ : $\gamma(\mu   x)$ . . . . .	74
Result 3.3.6	Posterior Density of $\mu   x$ as a Function of $w(x)$ : $\pi(\mu   x)$ . . .	74
Result 3.3.7	Posterior Density of $\mu   (x, \mu \neq 0)$ for Laplace prior: $\gamma(\mu   x)$ .	75
Result 3.3.8	Posterior Mean of $\mu   (x, \mu \neq 0)$ for Laplace Prior: $\hat{\mu}_\gamma(x)$ . . .	76
Result 3.3.9	Posterior Mean of $\mu   x$ : $\hat{\mu}_\pi(x)$ . . . . .	77
Definition 3.4.1	Naïve Posterior Mean and Median of $JV   \mathbf{x}$ . . . . .	81
Definition 3.4.2	Naïve Posterior Mean and Median of $IV   \mathbf{x}$ . . . . .	82
Definition 3.4.3	Posterior Mean and Median of $JV   \mathbf{x}$ . . . . .	82
Result 3.4.4	$\widehat{JV}_{NEB, \ell_2}(\mathbf{x})$ vs. $\widehat{JV}_{EB, \ell_2}(\mathbf{x})$ . . . . .	82
Result 3.4.5	Posterior Mean of $JV   \mathbf{x}$ : $\widehat{JV}_{EB, \ell_2}(\mathbf{x})$ . . . . .	83
Result 3.4.6	Posterior Median of $JV   \mathbf{x}$ : $\widehat{JV}_{EB, \ell_1}(\mathbf{x})$ . . . . .	86
Definition 3.4.7	Posterior Mean and Median of $IV   \mathbf{x}$ . . . . .	86
Model 3.5.1	Simulation Models . . . . .	87
Definition 3.6.1	MPE and MAPE . . . . .	89
Result 3.6.2	Simulation Approximations to MPE and MAPE . . . . .	90

## ACKNOWLEDGMENTS

I wish to thank Eric Zivot for his encouragement and belief in my ability. I thank the rest of my committee members Donald B. Percival, Yu-chin Chen, James V. Burke, and Charles R. Nelson, for their advice as I begin my career. Also, I acknowledge my classmates Christian Henn, Drew Creal, and Scott Payseur whose correspondences helped guide me academically. As a friend and colleague, Dustin Lennon went above and beyond the call of duty. I may not have finished if it were not for the regular conversations he and I had. Finally, I must mention my girlfriend Shanna Dulin. She selflessly put her needs aside throughout this process so that she could be there for me emotionally. I cannot imagine having a warmer, more caring, more empathetic person by my side. I most certainly would not have finished without her.

# DEDICATION

to Meira



## Chapter 1

### AN OVERVIEW OF HIGH-FREQUENCY FINANCE

Not until the late 1980's had the study of high-frequency financial data—taken, roughly, to be data observed at daily or higher frequencies—been possible for more than a handful of researchers. Developments in the storage and vending of such data along with increases in computing power laid the groundwork for research in high-frequency finance. [Wood \(2000\)](#) gives a detailed account of the early days of high-frequency financial data.

Fast-forwarding to the current time, with a tractable dataset in hand, the modern high-frequency financial econometrician (1) tests theoretical market microstructure models, (2) models prices, volumes, and durations at high-frequencies directly (i.e., statistical modeling), or (3) models lower-frequency phenomena (e.g., daily integrated volatility) using high-frequency data. Brief mention of the former two will be given here while the latter will take the focus. But before any modeling can take place, data cleaning methods, empirical characteristics, and market microstructure noise must be considered. Discussions immediately follow here.

#### ***1.1 Data Errors and Cleaning***

The financial practitioner or researcher working at traditional, lower frequencies can find clean data sets with relative ease. This is not the case with high-frequency financial data which can typically be distributed with a significant amount of recorded errors. Because of this, special care must be taken in cleaning and filtering such data.

[Dacorogna et al. \(2001\)](#) dedicate a lengthy, involved chapter to cleaning aspects. Recently, [Barndorff-Nielsen et al. \(2009\)](#) discuss cleaning issues and walk the reader

through a step-by-step cleaning procedure—using NYSE TAQ data—with their end goal being the estimation of daily integrated volatility using their realized kernel methods. [Brownlees and Gallo \(2006\)](#) demonstrate significant improvements in the estimation of duration models after implementing their carefully-explained cleaning procedure.

Cleaning methods and procedures obviously depend on the financial instrument, exchange, data distribution, etc., with each possibly possessing their own idiosyncratic issues. But speaking in a broad sense, errors one may run into include:

- prices entered as zero
- misplaced decimal points
- other general typos
- “test” data submitted to the system
- scaling problems (e.g., stock splits, changes in quoting convention, etc.)
- multiple prices for a given time stamp
- general computer system errors

[Aït-Sahalia et al. \(2011\)](#) lump these errors—in the observation of the true asset price process—together with market microstructure effects from trading process frictions and asymmetric information. Here, I handle them separately because I consider these errors to be fundamentally different in that they are typically filtered out prior to modeling considerations. A detailed explanation of the filtering process is outside the scope of this paper and, instead, I refer the reader to the aforementioned citations.

## **1.2 Empirical Characteristics**

With a clean set of high-frequency financial data, one still must consider the number of unique characteristics which distinguish it from the more traditional, simpler, lower-frequency. One may coerce the data set into one devoid of such characteristics—being simpler to handle with the traditional tools of financial econometrics—or model these characteristics outright. Depending on the desired modeled phenomena (e.g.,

integrated volatility, durations, prices, etc.) different irregularities require more focus as to how they potentially complicate a model. Discussions of these irregularities will appear again as my particular model calls for them in later chapters. For a more involved discussion see [Dacorogna et al. \(2001, ch. 5\)](#), [Tsay \(2005, ch. 5.1–5.3\)](#), and [Russell and Engle \(2009\)](#). Here, I only give brief attention to these irregularities.

- Unequally spaced time intervals: Also goes under titles *variable trading intensity* or *non-synchronous trading*. One can easily take for granted the uniform spacing of data in calendar time when using low-frequency data. [Hautsch \(2004\)](#) argues this irregular spacing is the main property distinguishing high-frequency data. The simplest method for handling is to coerce the data to regular spacing using a *previous tick* or *linear interpolation* (most studies prefer previous tick). [Hansen and Lunde \(2006\)](#) provide a careful discussion. The other option is to model this phenomena outright—the subject of a large literature on *autoregressive conditional duration* (henceforth, *ACD*) modeling and discussed in section 1.7.
- Discrete valued prices: This certainly exists regardless of data frequency. However, the size of a minimum tick represents a larger percentage of price movements at the higher frequencies. With this level of granularity, the typical continuous diffusion assumption for the asset price process becomes untenable. I discuss various methods for modeling prices as discrete random variables in section 1.7.
- Zero price-change transactions: The previous item should be mentioned with caveat: if the average tick size is a few multiples larger than the minimum tick size, the continuous diffusion assumption might still be tenable (this can be the case with certain foreign exchange markets). However, this is not the case with large-cap liquid assets where around two-thirds of transactions are associated with zero price change ([Tsay, 2005](#)). For the discrete price models to be discussed, this presents an additional challenge.
- Daily trading U-shaped pattern: A familiar daily pattern appears in asset prices:

high price volatility, high trading intensity, low trade duration, and large transaction volume at the beginning and end of trading sessions—with the opposite occurring in the middle of the trading session (Gourieroux and Jasiak, 2001; Tsay, 2005). This pattern can lead one to deduce an inappropriately high level of volatility—which can be corrected for by treating the pattern as deterministic.

- Simultaneous transactions: Because of the structure of the limit order book, multiple prices may be stamped to the same time. This is not an error. For example, a buy limit order may come onto the market and get filled by the sell-side of the book at several transaction prices. Methods for handling this—so that each time is associated with only one price—include taking the last price, averaging over the multiple prices, and performing a volume-weighted average over the multiple prices. One would obviously skip this step if modeling the limit order book were specifically of interest (e.g., Roşu, 2009)

### 1.3 Market Microstructure Noise

Because of market microstructure and trading data realities, many empirical irregularities show up in high-frequency financial data. A number of those irregularities—and methods for coping with them—were discussed in the last section. Here, I discuss *market microstructure noise*—a set of empirical irregularities fundamentally different from those of section 1.2 in that they obscure observation of the true, underlying, latent price of an asset. More specifically, one could represent market microstructure noise as  $U_t$  in the simple model  $X_t = Y_t + U_t$ , where  $X_t$  represents the empirically observed log price and  $Y_t$  the latent log price. Here,  $Y_t$  includes the empirical characteristics in section 1.2 and is cleaned of the errors in section 1.1.

Aït-Sahalia et al. (2011) partition microstructure noise into that caused by *a*) trading process frictions and *b*) informational effects, and I follow that here.

## 1.4 Trading Process Frictions

This market microstructure noise comes from the natural frictions associated with trading and the exchange set-up. The financial econometrics literature identifies several types of market microstructure noise from trading process frictions, with a few mentioned here:

- Bid-Ask bounce: This is the most frequently discussed among the types of market microstructure noise. Simply put, the transaction price tends to *bounce* between the bid and the ask prices depending on the previous transaction being seller or buyer initiated. This induces negative lag-1 autocorrelation in the empirical price-change series. As the observation frequency increases along with localization in time, the latent price behaves more like a constant and, thus, the *bid-ask bounce* comes to dominate the observed price process. Methods for coping with this include sub-sampling and different forms of filtering. See [Roll \(1984\)](#); [Tsay \(2005\)](#) for a more detailed discussion.
- Non-Synchronous trading: I discussed this earlier in section [1.2](#) under the heading *unequally spaced time intervals*. Here, I mention it as a potential source of market microstructure noise—which is not as intuitively understandable as bid-ask bounce.

Suppose an asset has, say, a positive drift component. Suppose, also, one has just observed a series of zero price-change transactions. One might think of the observed price series as not having *caught up* to the latent price and that a large empirical price change is *building up*. Thus a series of zero price-change transactions (negative when adjusting for the mean price-change resulting from drift) indicates a large, *positive* price-change—inducing negative lag- $n$  serial-correlation. See [Campbell et al. \(1996\)](#); [Tsay \(2005\)](#) for a more detailed discussion.

As an aside, not necessarily as a source of noise but still of tremendous modeling

difficulty is the existence of non-synchronous trading in multivariate modeling. One can see how this relates to integrated *co-volatility* estimation in [Griffin and Oomen \(2010\)](#); [Voev and Lunde \(2007\)](#); [Zhang \(2010\)](#); [Hayashi and Yoshida \(2005\)](#).

- Rounding noise: I mentioned discrete valued prices in section 1.2 as an empirical characteristic that presents a challenge to traditional financial models. In addition to this, one can think of the discrete pricing grid as adding a component of noise to the empirically observed time series—which is, essentially, the latent price (ignoring other sources of market microstructure noise) rounded to the nearest tick.
- Fighting screen effect: In foreign exchange markets, during periods of market inactivity, quote contributors will frequently submit nominally different quotes to appear on the Reuters screen ([Zhou, 1996](#)).<sup>1</sup> This adds a small, non-autocorrelated noise to the series of quotes.

### 1.5 *Informational Effects*

In addition to the market microstructure noise from trading process frictions, a noise of a more theoretical sort comes from informational effects.

- Adverse selection: If there exists a set of informed traders, a transaction or price change may signal some of these traders' private information. Market makers produce a bid-ask spread in order to protect themselves from, among other things, the adverse selection of trading. These changing informational effects obscure the latent price and represent noise.
- Inventory control: Similar to the last point, another component of the bid-ask spread represents the costs market makers bear holding inventory and transacting.

---

<sup>1</sup>This is not prohibitively costly since the minimum quote denomination is significantly smaller than the average price movement.

- Order flow: the gradual response of prices to a block trade and the strategic component of order flow will also drive a wedge between the observed price and the latent price.

For more detail, see [Stoll \(2000\)](#); [Madhavan \(2000\)](#); [Biais et al. \(2005\)](#) for surveys and [O'Hara \(1998\)](#); [Hasbrouck \(2007\)](#) for textbook treatment.

### **1.6 Structural Market Microstructure Models**

With such data in hand, theories of market microstructure could be tested and new ones posited. [Biais et al. \(2005\)](#) survey the market microstructure literature, touching on topics such as asymmetric trader information, inventory holding costs, market maker power, call auctions, granularity of the pricing grid and their effects on price discovery, volume, duration, and other financial asset trading metrics. [O'Hara \(1998\)](#) gives an early textbook treatment of theoretical market microstructure models.

As an aside, deep understanding of such market microstructure models requires familiarity with how basic market features work, such as the limit order book. [Gourieroux and Jasiak \(2001\)](#), Chapter 14.1–14.3) give a short, and very useful, introduction to these features. [Hasbrouck \(2007\)](#) provides a deeper, more recent, textbook treatment.

### **1.7 High-Frequency Financial Econometric Models**

Among financial econometricians, there exists a tremendous amount of interest in modeling aspects of financial data at their finest level of detail, from transactions data. [Engle \(2000\)](#) dub this highest frequency *ultra-high-frequency*. A number of models exist for price data themselves. In early work, [Hausman et al. \(1992\)](#) use an ordered probit model to capture ultra-high-frequency price movements. Alternatively, [Rydberg and Shephard \(2003\)](#) model returns as multiplicative price-change, direction, and size variables to each be modeled conditionally. A more recent treatment by [Robert and Rosenbaum \(2011\)](#) models prices as a hybrid of discrete-time and continuous-time models.

At ultra-frequencies the autoregressive conditional duration (ACD) model of [Engle and Russell \(1997, 1998\)](#), which can be seen as a type of dependent Poisson process, provides a popular foundation for the modeling of durations. See [Hautsch \(2004\)](#); [Pacurar \(2008\)](#) for reviews of extensions to the ACD model. Of course, one may model price and durations in a bivariate sense. For a textbook treatment see [Tsay \(2005, Ch. 5.7\)](#).

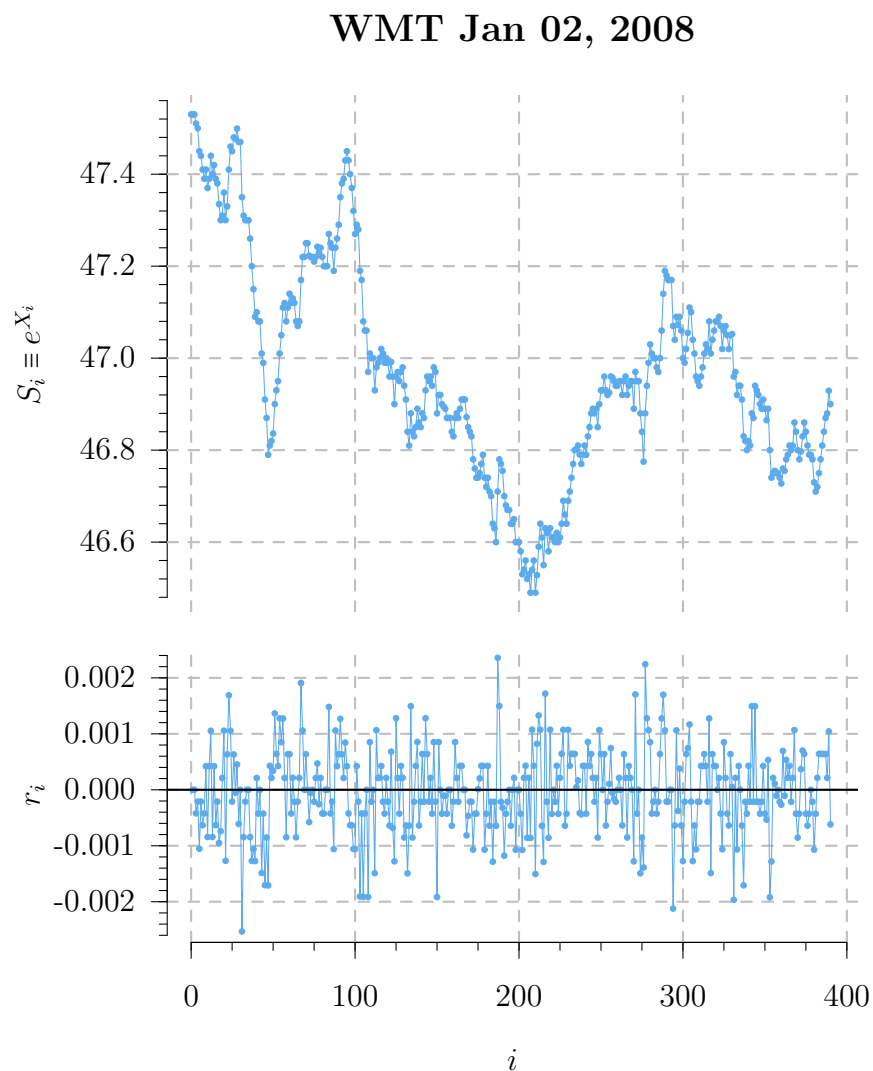
But our interest in following chapters is not in the modeling of high-frequency financial variables themselves but, rather, using high-frequency data to estimate lower-frequency quantities such as daily return variance, integrated variance, and jump variation. We give a brief introduction to this topic in the following section.

### ***1.8 Integrated Variance and Jump Variation from High-Frequency Financial Asset Returns***

Before we proceed with a short introduction to the subject of integrated variance, jump variation, and jump estimation from high-frequency realized quantities, we must mention a couple of surveys of the subject area. [McAleer and Medeiros \(2008\)](#) survey the topic of integrated variance estimation in a continuous model with different possible specifications of market microstructure noise. [Dumitru and Urga \(2012\)](#) provide a recent survey and simulation performance comparison of several procedures for jump detection in high-frequency asset prices. They find the estimator of [Lee and Mykland \(2008\)](#) gives the best performance.

To motivate the topic of integrated variance estimation, let us begin with an example question: “What is the return variance of Walmart (WMT), trading on the New York Stock Exchange (NYSE), on January 2, 2008?” We see the prices of Walmart in [Figure 1.1](#). A naïve, first-pass answer might be: “Zero! We observe the actual Walmart prices on January 2, 2008, and so the price series is no longer a random variable.” Though a technically reasonable answer, this misses the point of the question. The questioner most likely meant to ask, “What is the variance of

the underlying process generating Walmart returns on January 2, 2008, of which we observe one realization?”



**Figure 1.1:** 1-minute spot prices  $S_i$  and log-returns  $r_i$  for Walmart (trading on the New York Stock Exchange as *WMT*) on January 2, 2008. This implies a 391-element homogenous partition  $\mathcal{P}_{391}$  of  $[0, 390]$ , the parametrization of the day’s trading time interval of 9:30AM–4:00PM.

Another naïve answer may be, “Since we have one realization of Walmart’s underlying price process on January 2, 2008, our estimate of the day’s return variance should just be the sample-of-1 variance: the square of the entire day’s return.”<sup>2</sup> This is not an unreasonable answer. This estimator of day’s return variance is actually unbiased, albeit, extremely noisy. The problem with such an estimator is that it does not make use of all the information in the intraday returns.<sup>3</sup>

This might be understood analogously to estimating the probability weight of a weighted coin in a sequence of Bernoulli trials by the  $\{0,1\}$  value of the random variable, corresponding to  $\{\text{Tails, Heads}\}$  outcomes, on the first trial. This too gives an unbiased estimate of the probability weight of the coin but clearly misses all the information from subsequent trials which could potentially reduce the variance of the estimator. In fact, the sample mean of the random variable gives the minimum variance unbiased estimator of probability weight of the coin. In the case of estimating daily return variance, we’ll find that the sum of the squared intraday returns (realized variance) completes the analogy.

Of course, the vague nature of the motivating question produced the confused answers. To proceed carefully, we must reformulate the motivating question in terms of a clear mathematical model for the price processes for Walmart. The purpose of Section 2.1 is to build such a model from first principles of financial theory, arriving at the semimartingale plus model for log-prices  $X_t$ , where log-returns  $r_t$  may be written in the form of Theorem 2.1.27 as

$$r_t = X_t - X_0 = \int_0^t a_u du + \int_0^t \sigma_u dW_u + J_t.$$

where, for the log-price process  $X$ ,  $a$  represents the *spot drift* process,  $\sigma$  the *spot volatility* process,  $J$  a jump process, and  $W$  a standard Brownian innovations process.

---

<sup>2</sup>Embedded in this response is the assumption of zero mean drift in the price underlying price process.

<sup>3</sup>See [McAleer and Medeiros \(2008\)](#) for a discussion of this point.

This model appears in the literature most commonly with the name *jump-diffusion*. [Barndorff-Nielsen and Shephard \(2004\)](#) refer to this model as the *stochastic volatility semimartingale plus jumps* model or the *Brownian semimartingale plus jumps* model in [Barndorff-Nielsen and Shephard \(2006\)](#).

In this new context, we may reformulate the motivating question of the return variance for Walmart as finding.

$$\text{Var}\{r_t \mid \mathcal{F}_t\}, \quad \mathcal{F}_t \equiv \mathcal{F}\{a_u, \sigma_u\}_{u \in [0, t]}$$

where  $\mathcal{F}_t$  is the filtration generated by the sample paths of the spot drift  $a$  and spot volatility  $\sigma$  processes and we parameterize the day's trading time interval to  $[0, t]$ . We call  $\text{Var}\{r_t \mid \mathcal{F}_t\}$  the *conditional daily return variance* for appropriate parameterizations. From [Theorem 2.2.5](#) we find that when  $J \equiv 0$ , in addition to some technical conditions,

$$\text{Var}\{r_t \mid \mathcal{F}_t\} = \int_0^t \sigma_u^2 du \equiv IV_t$$

where we define the last quantity as *integrated variance*. Thus, in this special case, the problem of estimating conditional daily return variance reduces to one of estimating integrated variance. From the same theorem, we find that the sum of the intraday squared returns,

$$RV_t^{(n)} \equiv \sum_{i=1}^n r_i^2,$$

which is known as the *realized variance*, is a consistent estimator of integrated variance, a fact that we denote by

$$RV_t^{(n)} \rightarrow IV_t.$$

Use of sums of intraday squared returns can be found early in the work of [Merton \(1980\)](#). For early papers showing its unbiasedness and consistency in estimation of integrated variance see [Andersen and Bollerslev \(1998\)](#); [Andersen et al. \(2001, 2003\)](#); [Barndorff-Nielsen and Shephard \(2001, 2002\)](#); [Barndorff-Nielsen and Shephard \(2002\)](#). But of course, this is not the case when we allow  $J \neq 0$ , where from [Theorem 2.2.3](#) we

find

$$RV_t^{(n)} \rightarrow IV_t + JV_t,$$

where

$$JV_t \equiv \sum_{u=0}^t (dJ_u)^2,$$

and so realized variance is no longer a consistent estimator of integrated variance. [Barndorff-Nielsen and Shephard \(2004, 2006\)](#) emphasize these facts in their work. But because we find in [Theorem 2.2.6](#)

$$\text{Var}\{r_t \mid \mathcal{F}_t\} = IV_t + \text{E}\{JV_t \mid \mathcal{F}_t\}$$

we still have that realized variance is an asymptotically unbiased estimator of the conditional daily return variance. If our only purpose were to estimate conditional daily return variance, we would be in no worse of a position. But we have interest in estimating the components of conditional daily return variance, integrated variance and jump variation, separately as their values have implications for option pricing ([Nicolato and Venardos, 2003](#)), portfolio optimization ([Bäuerle and Rieder, 2007](#)), risk estimation ([Bollerslev et al., 2011](#)), and variance forecasting ([Andersen et al., 2007](#)). [Barndorff-Nielsen and Shephard \(2004, 2006\)](#) develop *realized bipower variation*

$$RBPV_t^{(n)} \equiv \sum_{i=2}^n |r_i| |r_{i-1}|$$

and show that it can be used as a consistent, jump-robust estimator of integrated variance by scaling it up as in [Theorem 2.2.10](#)

$$\frac{\pi}{2} RBPV_t^{(n)} \rightarrow IV_t.$$

Naturally, the residual of realized variance after removing the integrated variance

$$RV_t^{(n)} - \frac{\pi}{2} RBPV_t^{(n)} \rightarrow JV_t,$$

also as in [Theorem 2.2.10](#).

In Chapter 2 we build the semimartingale plus model from first principles of financial theory and derive these results more formally. We look for improvements upon the realized bipower variation-based estimators, which leads us to explore hard-thresholding and naïve shrinkage estimators of jump variation integrated variance. One may think of these estimators as employing a *jumps first* strategy of estimating jump sizes, then jump variation, and finally integrated variance as the residual of realized after removing jump variation.

## Chapter 2

## JV AND IV ESTIMATION IN A CONTINUOUS SEMIMARTINGALE MODEL

### 2.1 Toward the $\mathcal{SM}+$ Model

We begin with an abstract asset  $X$  which we associate with a logarithmic stochastic price process also denoted  $X \equiv (X_t)_{t \in [0, \infty)}$ , where  $X_t$  denotes the asset's log-price at time  $t$ . We ultimately wish to analyze the logarithmic returns of  $X$ ,  $r$ , on some finite trading time interval  $[0, T]$ , the latter representing one trading day in all simulation and empirical analyses to follow. To develop theoretical results, some theoretical mathematical model must govern the behavior of  $X$  and  $r$ . We ultimately use the semimartingale plus model ( $\mathcal{SM}+$ ) as the base model for analysis. This model often represents a starting point of analysis in the high-frequency financial econometrics literature and may be found under different names such as the *continuous stochastic volatility semimartingale plus jumps* model (Barndorff-Nielsen and Shephard, 2004), *Brownian semimartingale plus jumps* model (Barndorff-Nielsen and Shephard, 2006), *jump-diffusion* model (Fan and Wang, 2007), *special semimartingale* model (Andersen et al., 2003), or unnamed but using different notation (Lee and Mykland, 2008). In this section we start with a *general stochastic process* model ( $\mathcal{GSP}$ ) and carefully work toward the  $\mathcal{SM}+$  model so as to develop understanding of the latter's components. The uninterested reader may simply skip to Theorem 2.1.27 to find a statement of the  $\mathcal{SM}+$  model.

**Definition 2.1.1 (General stochastic process:  $\mathcal{GSP}$ )**

*If*<sup>1</sup>

---

<sup>1</sup>Where we emphasize words, precise definitions exist. Unless defined later, one can find fully

- $(\Omega, \mathcal{F}, \mathbb{F}, P)$  is a filtered complete probability space *satisfying the usual hypothesis of  $P$ -completeness and right continuity.*
- $X \equiv (X_t)_{t \in [0, \infty)}$  denotes an adapted, càdlàg stochastic process on  $(\Omega, \mathcal{F}, \mathbb{F}, P)$ , representing an asset with logarithmic price (log-price)  $X_t$  at time  $t$ .<sup>2</sup>
- $\Omega$  is a sample space, representing possible states in the asset's world.
- $\mathcal{F}$  is a  $\sigma$ -algebra, representing measurable events in  $\Omega$ .
- $\mathcal{F}_t \subseteq \mathcal{F}$  is a  $\sigma$ -algebra, representing measurable events at time  $t$ .
- $\mathbb{F} \equiv (\mathcal{F}_t)_{t \in [0, \infty)}$  is a filtration of the probability space  $(\Omega, \mathcal{F}, P)$ .
- $P : \mathcal{F} \rightarrow [0, 1]$  is a probability measure, representing true probabilities of events in  $\mathbb{F}$ .<sup>3</sup>

then we say the process  $X$  is a general stochastic process ( $X \in \mathcal{GSP}$ ).<sup>4</sup> □

Most stochastic analyses use this fully-specified probability model as a starting point, with it being implied if not explicitly stated. Additionally, in our financial context, we look to define the log-return process  $r$  as

**Definition 2.1.2 (Log-returns:  $r$ )**

For  $X \in \mathcal{GSP}$ , its logarithmic return (log-return) on a time interval  $[t_0, t_1]$

$$r_{X, [t_0, t_1]} \equiv X_{t_1} - X_{t_0}, \quad 0 \leq t_0 \leq t_1 < \infty.<sup>5</sup>$$

Where the stochastic process  $X$  is clearly implied on the time interval  $[0, t]$ ,

$$r_t \equiv r_{X, [0, t]}$$

precise definitions in [Protter \(2005, pg. 3\)](#) and [Rosenthal \(2006, pg. 7\)](#).

<sup>2</sup>We will loosely use  $X$  to refer to both the log-price process associated with the asset and the asset itself.

<sup>3</sup>In Chapter 3 we analyze returns from a Bayesian perspective and, so, relax this frequentist interpretation of a probability measure. In this case, we would appropriately replace *true probabilities* with *probability beliefs*

<sup>4</sup>Typically a general stochastic process, in its full generality, need not be an adapted, càdlàg stochastic process satisfying the usual hypothesis. Thus, our definition of a general stochastic process may be smaller than a probabilist may be accustomed.

<sup>5</sup>Even though we define the log-return mathematically for  $X \in \mathcal{GSP}$ , this quantity can meaningfully be thought of as a *return* only if  $X$  is a log-price process for an asset. We leave the statement as such in the interest of maintaining generality.

for shorthand use. □

Our interest is not in the log-return process  $r$  itself but, rather, the incremental log-returns associated with a sampling of the log-price process  $X$  on a trading time interval  $[0, t]$ . More specifically, in our simulation and empirical analyses we concern ourselves with the 1-minute log-returns of an asset on a daily trading time interval. We may represent a general, non-homogenous sampling of  $X$  on a trading time interval  $[0, t]$  as the set of values of the process at a set of times given by a *partition*, defined as follows:

**Definition 2.1.3 (Partition:  $\mathcal{P}$ )**

An  $(n + 1)$ -element partition of a time interval  $[0, t]$

$$\mathcal{P}_n \equiv \{\tau_0, \tau_1, \dots, \tau_n\}$$

is a set of times such that

$$0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_n = t < \infty.$$

We say such a partition is homogenous if

$$\tau_i = i \left( \frac{t}{n} \right), \quad i = 0, \dots, n.$$

□

and are now set to define log-returns on a partition.

**Definition 2.1.4 (Log-returns on a partition:  $r$ )**

For  $X \in \mathcal{GSP}$ , with a partition  $\mathcal{P}_n$  of some finite time interval is clearly implied,

$$r_{X,i} \equiv r_{X, [\tau_{i-1}, \tau_i]}, \quad i = 1, \dots, n,$$

for shorthand use. Where the stochastic process  $X$  is clearly implied,

$$r_{[t_0, t_1]} \equiv r_{X, [t_0, t_1]},$$

$$r_i \equiv r_{X,i}$$

for shorthand use. □

Figure 1.1 displays 1-minute spot prices  $S_i$  and log-returns  $r_i$  for Walmart (trading on the New York Stock Exchange as WMT) on January 2, 2008. This implies a 391-element homogenous partition  $\mathcal{P}_{391}$  of  $[0, 390]$ , the parametrization of the day's trading hours of 9:30AM–4:00PM.

Unsurprisingly, at this level of generality we cannot infer anything meaningful about the behavior of log-price and log-return processes  $X, r$ . For that, we must make additional assumptions informed by careful empirical observations of financial asset log-price processes.

**Definition 2.1.5 (Semimartingale:  $\mathcal{SM}$ )**

$X \in \mathcal{GSP}$  is a semimartingale ( $X \in \mathcal{SM}$ ) if there exist processes  $M, A$  with  $M_0 = A_0 = 0$  such that

$$X_t - X_0 = A_t + M_t$$

where  $M$  is a local martingale ( $M \in \mathcal{M}_{loc}$ ) and  $A$  is a finite variation process ( $A \in \mathcal{FV}$ ). □

In the financial context with a log-price process  $X$ , the finite variation process  $A$  may loosely be thought of as the drift and the local martingale component  $M$  the innovations process. Examples of semimartingales include Brownian motion, Poisson process, and Lévy processes. To understand why we might assume a log-price process is a semimartingale we have the following theorem:

**Theorem 2.1.6 (No arbitrage  $\rightarrow \mathcal{SM}$ )**

If  $X \in \mathcal{GSP}$  has the no arbitrage property, then  $X \in \mathcal{SM}$ .<sup>6</sup> □

*Proof.* See Schweizer (1992); Back (1991); Delbaen and Schachermayer (1994) for more rigorous statements and proofs. ■

---

<sup>6</sup>Even though the *no arbitrage property* is a mathematical condition on  $X \in \mathcal{GSP}$ , this property is meaningful only if  $X$  is an asset. We leave the statement as such in interest of maintaining generality.

A very recent, specialized literature has emerged arguing against the existence of certain specifications of the no arbitrage property in financial markets. Despite this, the no arbitrage assumption still represents a near-universal starting point across theoretical financial models. Well-known, classical, theoretical financial results such as the Black-Scholes pricing formula for a European call option or Arbitrage Pricing Theory rest on this assumption.

With our semimartingale log-price process  $X$ , we would like to introduce the concept of *quadratic variation*. Under certain assumptions and loosely speaking we will be able to regard the quadratic variation of  $X$  on a time interval as its variance on that interval. But quadratic variation is defined as a limit taken over increasingly fine partitions and so we will have to first define the *mesh* of a partition as follows:

**Definition 2.1.7 (Mesh:  $\|\mathcal{P}\|$ )**

*The mesh of a partition  $\mathcal{P}_n$*

$$\|\mathcal{P}_n\| \equiv \sup_{i \in \{1, \dots, n\}} |\tau_i - \tau_{i-1}|.$$

□

In the case of a simple, homogenously spaced partition of a time interval  $[0, t]$ , the mesh is simply  $\delta = \frac{t}{n}$ . Although our simulation and empirical analyses fall into this simple context, we maintain the generality of a non-homogenously spaced partition to stay consistent with elements of the financial econometrics literature and to emphasize that results do not require this simple context. With a mesh now defined, we are ready to generally define the *quadratic variation*

**Definition 2.1.8 (Quadratic Variation:  $[X]$ )**

*For  $X, Y \in \mathcal{SM}$ , the quadratic covariation on a time interval  $[0, t]$*

$$[X, Y]_t \equiv \lim_{n \rightarrow \infty} \sum_{i=1}^n r_{X,i} r_{Y,i}, \quad 0 \leq t < \infty,$$

where convergence is uniform on  $[0, t]$  in probability (ucp) and the limit is taken over all partitions  $\mathcal{P}_n$  of  $[0, t]$  with  $\lim_{n \rightarrow \infty} \|\mathcal{P}_n\| = 0$ .<sup>7</sup> The quadratic variation (QV)

$$[X]_t \equiv [X, X]_t$$

for shorthand use. □

For an intuitive understanding of why the limit in the previous theorem does not diverge to infinity, consider the semimartingale plus model of the previous chapter. There, returns over intervals, not containing any of the countable number of jumps, are of order  $O(\sqrt{\|\mathcal{P}_n\|})$  and so squared returns are of order  $O(\|\mathcal{P}_n\|)$ . But the number of terms in the summation is of order  $O\left(\frac{T}{\|\mathcal{P}_n\|}\right)$  and so the entire summation is of constant order  $O(T)$ . This is unaffected by the jumps as they are of constant order as well.

Of course, in a practical context we cannot observe quadratic variation. We can, however, observe finite approximations to quadratic variation—*realized variance*, defined as follows:

**Definition 2.1.9 (Realized Variance:  $RV$ )**

For  $X \in \mathcal{SM}$ , the realized variance on a time interval  $[0, t]$  with partition  $\mathcal{P}_n$

$$RV_{X,t}^{(n)} \equiv \sum_{i=1}^n r_{X,i}^2, \quad 0 \leq t < \infty.$$
<sup>8</sup>

Where the stochastic process  $X$  and partition  $\mathcal{P}_n$  are clearly implied

$$RV_t \equiv RV_{X,t}^{(n)}$$

for shorthand use. □

<sup>7</sup>Usually this is taken to be a result for semimartingale processes after one defines quadratic covariation more generally in terms of stochastic integration. See Protter (2005, pg. 66). To avoid unnecessary complication, we skip the typical set-up.

<sup>8</sup>Even though we define the realized variance mathematically for  $X \in \mathcal{SM}$ , this quantity can meaningfully be thought of as a *variance* (or, more accurately, an *approximation to variance*) only if  $X$  satisfies additional regularity conditions to be discussed later.

**Theorem 2.1.10 ( $RV \rightarrow [X]$ )**

For  $X \in \mathcal{SM}$ ,

$$\lim_{n \rightarrow \infty} RV_{X,t}^{(n)} = [X]_t,$$

where convergence is ucp on  $[0, t]$  and the limit is taken over all partitions  $\mathcal{P}_n$  of  $[0, t]$  with  $\lim_{n \rightarrow \infty} \|\mathcal{P}_n\| = 0$ . □

*Proof.* This follows trivially from previous definitions. ■

Still, the assumption of a semimartingale log-price processes is not enough to guarantee additional desirable theoretical properties of financial assets. For that, we define a *special semimartingale* process as follows:

**Definition 2.1.11 (Special semimartingale:  $\mathcal{SSM}$ )**

If  $X \in \mathcal{SM}$  has decomposition

$$X_t - X_0 = A_t + M_t$$

where  $A_0 = M_0 = 0$ ,  $M \in \mathcal{M}_{loc}$ ,  $A \in \mathcal{FV}$  and predictable ( $A \in \mathcal{PRE}$ ), then  $X$  is said to be a special semimartingale ( $X \in \mathcal{SSM}$ ).<sup>9</sup> □

One important result for special semimartingale processes is the uniqueness of their decomposition into finite variation and local martingale processes, stated as follows:

**Theorem 2.1.12 ( $\mathcal{SSM} : X = A + M$  canonical)**

If  $X \in \mathcal{SSM}$ , then its decomposition

$$X_t - X_0 = A_t + M_t$$

with  $A \in \mathcal{PRE}$  unique is called the canonical decomposition. □

*Proof.* See [Protter \(2005, Chpt. III.7, Thm. 31\)](#) ■

---

<sup>9</sup>Clearly,  $\mathcal{SSM} \subseteq \mathcal{SM}$ .

In the financial context this means the drift and innovations components of the log-price process can be uniquely defined, a necessary condition for what we would think of as coherent model for theoretical asset prices. Special semimartingales relate to other desirable theoretical properties of financial assets. For example:

**Theorem 2.1.13 ( $\mathcal{SSM} \rightarrow$  local risk premia exist)**

If  $X \in \mathcal{SSM}$  satisfies weak technical conditions, then local risk premia exist.<sup>10</sup>  $\square$

*Proof.* As it is written, the theorem omits weak necessary technical conditions. See [Back \(1991, Thm. 1\)](#) for a precise statement and proof.  $\blacksquare$

Without risk premia, pricing of assets by, for example, the Capital Asset Pricing Model or Black-Scholes option pricing model would be impossible. Of course, an asset log-price process being a special semimartingale is only a sufficient condition for the existence of local risk premia for that asset. [Schweizer \(1992\)](#) defines the concept of a *martingale density* and provides more technical conditions under which one can slightly relax the special semimartingale assumption. Nonetheless, this should help provide some intuition for why one would constrain the coherent set of asset log-price processes to special semimartingales. For additional intuition, the following theorem provides a necessary condition for a stochastic process to be a special semimartingale:

**Theorem 2.1.14 ( $\mathcal{SM} +$  bounded jumps  $\rightarrow \mathcal{SSM}$ )**

If  $X \in \mathcal{SM}$  and has bounded jumps, then  $X \in \mathcal{SSM}$ .  $\square$

*Proof.* From [Protter \(2005, Thm. III.26\)](#), a *classical semimartingale* is a semimartingale. The proof then follows the proof of [Protter \(2005, Thm. III.31\)](#).  $\blacksquare$

Thus, one could loosely think of special semimartingales as semimartingales with bounded jumps. Certainly, a continuous process has no jumps and therefore vacuously has bounded jumps, stated in the following theorem:

---

<sup>10</sup>Even though *local risk premia* are defined by a mathematical condition on a general stochastic processes  $X$ , they are meaningful only if  $X$  is a log-price process for an asset. We leave the statement as such in interest of maintaining generality.

**Theorem 2.1.15** ( $\mathcal{SM}^c \rightarrow \mathcal{SSM}^c$ )

If  $X$  is a semimartingale with almost sure continuous paths ( $X \in \mathcal{SM}^c$ ), then  $X \in \mathcal{SSM}$  and has almost sure continuous sample paths ( $X \in \mathcal{SSM}^c$ ) and both  $A$  and  $M$  in its canonical decomposition have continuous paths to be denoted by  $A \in \mathcal{FV}^c$  and  $M \in \mathcal{M}_{loc}^c$ .  $\square$

*Proof.* See proof of corollary to Protter (2005, Thm. III.31), again using Protter (2005, Thm. III.26). One can also look to the proof of Musiela and Rutkowski (2008, Prop. A.8.1) for an alternative exposition.  $\blacksquare$

Eventually, we rely on this theorem to imply the uniqueness of the decomposition of a semimartingale as  $X = A^c + M^c + J$  in Theorem 2.1.27. But, as we will see, unless the process  $X$  itself is assumed to be a special semimartingale (and not just the process  $A^c + M^c$  which is already a special semimartingale by continuity), we will not be able to uniquely decompose  $J$  into finite variation and local martingale jump components. Ultimately, we leave this further decomposition of  $J$  unspecified as it is custom in the literature where the starting point is the semimartingale plus model of Theorem 2.1.27. However, we draw attention to this now as it could represent an interesting starting point for the inquiry into the nature of jumps. But before even a decomposition  $X = A^c + M^c + J$  is possible, we must make an additional assumption about the log-price process:

**Assumption 2.1.16** ( $\sum |\Delta X| < \infty$ )

$$\sum_{0 \leq u \leq t} |\Delta X_u| < \infty$$

almost surely for each  $t > 0$ , where,

$$\Delta X_t \equiv X_t - X_{t-}$$

$$X_{t-} \equiv \lim_{u \nearrow t} X_u.$$

$\square$

One can look to Protter (2005, Hypothesis A) for the origin of this assumption. A log-price process satisfying this assumption has non-dense, or countable, jumps. With this assumption, we can guarantee the following decomposition:

**Theorem 2.1.17 ( $X = X^c + J$ )**

If  $X \in \mathcal{SM}$  satisfies Assumption 2.1.16, then  $X$  can be decomposed uniquely as

$$X_t = X_t^c + J_t,$$

where

$$J_t \equiv \sum_{0 \leq u \leq t} \Delta X_u,$$

$$X_t^c \equiv X_t - J_t$$

are the jump and continuous parts of  $X$ . □

*Proof.* The proof is trivial as  $X$  satisfying Assumption 2.1.16 guarantees the quantity  $\sum_{0 \leq u \leq t} \Delta X_u$  is well-defined almost surely for each  $t > 0$ . ■

$X$  can be decomposed further as:

**Theorem 2.1.18 ( $\mathcal{SM}: X = A^c + M^c + J$ )**

If  $X \in \mathcal{SM}$ , with its (canonical if  $X \in \mathcal{SSM}$ ) decomposition

$$X_t - X_0 = A_t + M_t,$$

satisfies Assumption 2.1.16, then in its unique decomposition

$$X_t = X_t^c + J_t,$$

$X_t^c$  can be decomposed uniquely as

$$X_t^c = A_t^c + M_t^c$$

where  $A^c \in \mathcal{FV}^c$  and  $M^c \in \mathcal{M}_{loc}^c$  are the continuous components of  $A$  and  $M$ . Furthermore,  $J_t$  can be decomposed (uniquely if  $X \in \mathcal{SSM}$ ) as

$$J_t = \sum_{0 \leq u \leq t} \Delta A_u + \sum_{0 \leq u \leq t} \Delta M_u. \text{<sup>11</sup>}$$

*Proof.* See proof for Andersen et al. (2003, Prop. 1) which appeals to Protter (2005, pgs. 70, 221). The uniqueness result follows immediately from the uniqueness of the canonical decomposition and the almost sure uniqueness of the continuous/jump decomposition. For validity of the further decomposition into purely continuous and discontinuous component process, see Lowther (2012). ■

At this point we are ready to introduce our primary variable of interest for this analysis, *jump variation*:

**Definition 2.1.19 (Jump variation:  $JV$ )**

For  $X \in \mathcal{SM}$  the jump variation ( $JV$ ) of  $X$  on a time interval  $[0, t]$

$$JV_{X,t} \equiv \sum_{t_0 \leq u \leq t_1} (\Delta X_u)^2.$$

with  $JV$  finite for  $X$  satisfying Assumption 2.1.16. Where an asset  $X$  is clearly implied,

$$JV_t \equiv JV_{X,t}$$

for shorthand use. □

It is worth noting that we introduce *jump variation* early in this exposition as it is so central to our analysis and technically sound to do so. However, for an accurate historical exposition, we would introduce *integrated variance* (Definition 2.2.2) first as the methods for estimating conditional daily return variance by realized quantities started from a purely diffusive model where jumps played no role (Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2001).

---

<sup>11</sup>To stay consistent with the exposition of Barndorff-Nielsen and Shephard (2004), we eventually reparameterize  $J_t = \sum_{i=1}^{N_t} C_i$ .

**Theorem 2.1.20** ( $\mathcal{SM} : [X] = [X^c] + \sum(\Delta X)^2$ )

If  $X \in \mathcal{SM}$  satisfies Assumption 2.1.16, then its quadratic variation can be decomposed into continuous and jump parts as

$$[X]_t = [X^c]_t + JV_t.$$

□

*Proof.* See remarks of Protter (2005, pg. 70) and proof in Dellacherie and Meyer (1982, pg. 355). ■

One can interpret jump variation as the contribution of jumps to the overall quadratic variation of a process or, in a financial context under certain assumptions, as the contribution of jumps to the log-return variance over a trading time interval.

**Theorem 2.1.21** ( $\mathcal{SM}^c : [X] = [M]$ )

If  $X \in \mathcal{SM}$  satisfies Assumption 2.1.16, with decomposition of its QV

$$[X]_t = [X^c]_t + JV_t,$$

then

$$[X^c]_t = [M^c]_t$$

and  $JV_t$  can be decomposed (uniquely if  $X \in \mathcal{SSM}$ ) as

$$JV_t = \sum_{0 \leq u \leq t} (\Delta M_u)^2 + \sum_{0 \leq u \leq t} (\Delta A_u)^2 + 2 \sum_{0 \leq u \leq t} \Delta A_u \Delta M_u.$$

Furthermore, if  $A \in \mathcal{FV}^c$ , then

$$JV_t = \sum_{0 \leq u \leq t} (\Delta M_u)^2$$

and so

$$[X]_t = [M^c]_t + \sum_{0 \leq u \leq t} (\Delta M_u)^2 = [M]_t.$$

*Proof.*

$$\begin{aligned}
[X]_t &= [M + A]_t \\
&= [M]_t + [A]_t + 2[M, A]_t, \\
&\quad (\text{by Polarization Identity, see Protter (2005, pg. 66)}) \\
&= [M^c]_t + \sum_{0 \leq u \leq t} (\Delta M_u)^2 + [A]_t + 2[M, A]_t, \\
&\quad (\text{by previous theorem}) \\
&= [M^c]_t + \sum_{0 \leq u \leq t} (\Delta M_u)^2 + [A]_t + 2 \sum_{0 \leq u \leq t} \Delta A_u \Delta M_u, \\
&\quad (\text{by Kallenberg (1997, Thm. 23.6(viii))}) \\
&= [M^c]_t + \sum_{0 \leq u \leq t} (\Delta M_u)^2 + \sum_{0 \leq u \leq t} (\Delta A_u)^2 + 2 \sum_{0 \leq u \leq t} \Delta A_u \Delta M_u, \\
&\quad (\text{by previous theorem and fact that } [A^c]_t = 0 \text{ for any } A^c \in \mathcal{FV}^c)
\end{aligned}$$

The final theorem statements follow immediately from arguments already made. See Andersen et al. (2003, Prop. 1) and Barndorff-Nielsen and Shephard (2002, pg. 463) for discussions and partial proofs. ■

We establish the following result to help with establishing the theorem following it

**Theorem 2.1.22** ( $\mathcal{M}_{loc}^2 : \mathbf{E}\{M^2\} = \mathbf{E}\{[M]\}$ )

*If  $M$  is a locally square integrable local martingale ( $M \in \mathcal{M}_{loc}^2$ ), then*

$$\mathbf{E}\{M_t^2 \mid \mathcal{F}_0\} = \mathbf{E}\{[M]_t \mid \mathcal{F}_0\}, \quad t \geq 0.$$

□

*Proof.* See Protter (2005, Cor. II.6.3). ■

Nonetheless, we can already obtain some insight into the nature of quadratic variation. With a local martingale  $M$  satisfying  $M_0 = 0$ , we can loosely interpret

the previous result as establishing that the expected quadratic variation of a local martingale over an interval is an unbiased forecast of its variance over that interval. The following theorem establishes this result for special semimartingales under certain assumptions:

**Theorem 2.1.23** ( $\mathcal{SSM}, \mathcal{DET} : \text{Var}\{r\} = \mathbf{E}\{[X]\}$ )

If  $X \in \mathcal{SSM}$  with canonical decomposition  $X = A + M$  where  $A \in \mathcal{FV}^c$  and  $M \in \mathcal{M}_{loc}^2$ , then

$$\text{Var}\{r_t \mid \mathcal{F}_0\} = \mathbf{E}\{[X]_t \mid \mathcal{F}_0\} + \text{Var}\{A_t \mid \mathcal{F}_0\} + \text{Cov}\{M_t, A_t \mid \mathcal{F}_0\}.$$

Furthermore, if  $M_t, A_t \mid \mathcal{F}_0$  are independent of each other ( $M_t \perp\!\!\!\perp A_t \mid \mathcal{F}_0$ ), then

$$\text{Var}\{r_t \mid \mathcal{F}_0\} = \mathbf{E}\{[X]_t \mid \mathcal{F}_0\} + \text{Var}\{A_t \mid \mathcal{F}_0\}.$$

Finally, if  $A$  is deterministic ( $A \in \mathcal{DET}$ ), then

$$\text{Var}\{r_t \mid \mathcal{F}_0\} = \mathbf{E}\{[X]_t \mid \mathcal{F}_0\}.$$

□

*Proof.* See [Andersen et al. \(2003, Thm. 1, Cor. 1\)](#). The main result is based on the previous theorem. ■

This theorem clearly points to the importance of quadratic variation in forecasting asset return variance. However, our interest lies in *now-casting* rather than forecasting. That is, we look to make inferences about different aspects of our log-price and return processes over a trading time interval given we observe one realization of those processes on that same interval. This is a signal processing problem, not a forecasting problem. So, this theorem is clearly expository and non-central to the rest of our analysis. Nonetheless, from this result, one can develop some sense for how quadratic variation relates to return variance.

We are now ready to formally introduce the semimartingale plus model, defined as follows:

**Definition 2.1.24** ( $\mathcal{SM}+$ ,  $\mathcal{SM}^{c+}$ )

An  $X \in \mathcal{SM}$  satisfying

- Assumption 2.1.16,
- $M^c \in \mathcal{M}_{loc}^2$  in its decomposition  $X_t = A_t^c + M_t^c + J_t$ ,

is a semimartingale plus ( $X \in \mathcal{SM}+$ ).<sup>12</sup> Additionally, if

- $X \in \mathcal{SM}^c$

so that  $J_t = 0$  for all  $t > 0$ , then  $X = A^c + M^c$  is a continuous semimartingale plus ( $X \in \mathcal{SM}^{c+}$ ). □

The significance of this model may be puzzling as all it requires over the semimartingale model are the satisfaction of Assumption 2.1.16 and square integrability of the continuous local martingale component. But it is the satisfaction of the latter which allows for it to be written as a stochastic integral, stated formally as follows (familiar readers should recognize this theorem as nothing more than the *Martingale Representation Theorem*):

**Theorem 2.1.25** ( $\mathcal{M}_{loc}^2 : M = \int \sigma dW$ )

Let  $W$  be a standard Brownian motion ( $W \in \mathcal{BM}$ ) with  $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq \infty}$  its completed natural filtration. Then every  $M \in \mathcal{M}_{loc}^2$  for  $\mathbb{F}$  has a representation<sup>13</sup>

$$M_t - M_0 = \int_0^t \sigma_u dW_u$$

where  $\sigma_u \in \mathcal{PRE}$ . □

*Proof.* Theorem statement and proof for the more general,  $n$ -dimensional case can be found in Protter (2005, Thm. IV.43). ■

Standard results allow us to write the continuous finite variation component as a Riemann integral, stated formally as follows:

<sup>12</sup>The *plus* represents the extra conditions satisfied. This is not a standard definition.

<sup>13</sup>Hereafter, we abbreviate as  $M \in \mathcal{M}_{loc}^2$  where “for  $\mathbb{F}$ ” is implied.

**Theorem 2.1.26** ( $\mathcal{FV}^c : A^c = \int a \, du$ )

For every  $A^c \in \mathcal{FV}^c$ ,

$$A_t^c - A_0^c = \int_0^t a_u \, du$$

where  $a_u \in \mathcal{PRE}$ . □

*Proof.* A standard result in analysis is that any finite variation function has derivative almost everywhere in the Lebesgue measure sense. Then by the second fundamental theorem of calculus we can write any path of  $A$  as a Riemann integral. ■

With these last two results, we can now write a semimartingale plus stochastic process in terms of stochastic integrals and a jump component, stated formally as follows:

**Theorem 2.1.27** ( $\mathcal{SM}+ : X = \int a \, du + \int \sigma \, dW + J$ )

If  $X \in \mathcal{SM}+$ , then

$$r_t = X_t - X_0 = \int_0^t a_u \, du + \int_0^t \sigma_u \, dW_u + J_t$$

where  $W \in \mathcal{BM}$ ,  $a, \sigma \in \mathcal{PRE}$  are the spot drift and spot volatility of returns,<sup>14,15</sup> and  $J_t$  is a jump process written as:

$$J_t = \sum_{i=1}^{N_t} C_i,$$

where  $N_t$  is a simple counting process denoting the number of jumps in the time interval  $[0, t]$  and  $C_i$  denotes the size of the  $i$ -th jump. □

*Proof.* This is an immediate consequence of the previous definition and two theorems. Also notice we reparameterized the jump process  $J_t$ . This is possible as Assumption 2.1.16 is satisfied, which guarantees a countable number of jumps. ■

<sup>14</sup>This terminology is specific to finance and should only be used when interpreting the general stochastic process  $X$  specifically as a log-price process.

<sup>15</sup>We should be careful in interpreting  $\sigma$  as a *volatility*. None of the assumptions have yet guaranteed  $\sigma > 0$ . This will, however, be assumed in the semimartingale plus plus model.

But, again, for an exposition that follows the historical development of this topic, the following theorem would precede the last as researchers considered variance in pure diffusion models before jump-diffusion models. The advantage of the exposition here is that we emphasize the continuous semimartingale plus model as a special case of the semimartingale plus model. This jibes with our interest in presenting jump variation as our primary variable of interest.

**Theorem 2.1.28** ( $\mathcal{SM}^{c+} : X = \int a du + \int \sigma dW$ )

If  $X \in \mathcal{SM}^+$  with  $X$  continuous, then  $X \in \mathcal{SM}^{c+}$  with

$$r_t = X_t - X_0 = \int_0^t a_u du + \int_0^t \sigma_u dW_u$$

□

*Proof.* This trivially follows from the previous theorem where, in the case that  $X$  is continuous,  $J_t = 0$  for all  $t > 0$ . ■

With the jump-diffusion and pure diffusion forms found in the previous two theorems, we arrive at the typical models found in financial econometric analyses. Notice that at this level of generality we can say little about the nature of the jump process  $J$ . [Barndorff-Nielsen and Shephard \(2004\)](#); [Lee and Mykland \(2008\)](#) provide commentary, noting that the semimartingale plus model is sufficiently general such that it allows for time-varying jump intensity and volatility. Alternatively, we may like to know whether the jumps occur in the finite variation component  $A$  or local martingale component  $M$  of the decomposition of the semimartingale log-price process  $X = A + M$ . This opens up an interesting line of questioning into the nature of jumps. [Back \(1991\)](#); [Andersen et al. \(2003\)](#) provide a discussion suggesting a reasonable model with special semimartingale log-price processes would only consider local martingale jumps. Satisfaction of the no arbitrage property would require the bizarre property that any finite variation jumps happen in simultaneity with counterbalancing local martingale

jumps and, thus, would wash out in their effect on the log-price process. This argument provides partial support for the simple model of jumps found in Chapter 3. Nonetheless, we do not push this topic further, choosing to focus on jumps in a most general scenario.

## 2.2 *IV and JV Estimation in Previous Work*

By this point we established our base model of interest: the semimartingale plus model. This model emerged from the mild assumptions that the log-price process  $X$  was a semimartingale, had absolutely summable jumps  $\Delta X$ , and had a square integrable continuous martingale component  $M^c$  in its decomposition in Theorem 2.1.18. We also introduced our primary variable of interest: jump variation  $JV$ . We have yet to say anything about its estimation. Also, we have yet to introduce the other side of the realized variance coin: integrated variance.<sup>16</sup> Before we can introduce it, though, we must state the following result for the quadratic variation of a square integrable local martingale process:

**Theorem 2.2.1** ( $\mathcal{M}_{loc}^2 : [M] = \int \sigma^2 dt$ )

If  $M \in \mathcal{M}_{loc}^2$  with  $M_0 = 0$ , then

$$\begin{aligned} [M]_t &= [\sigma \cdot W]_t \\ &= \int_0^t \sigma_u^2 dt \end{aligned}$$

where  $W \in \mathcal{BM}$ ,  $\sigma \in \mathcal{PRE}$ , and

$$(\sigma \cdot W)_t \equiv \int_0^t \sigma_u dW_u$$

denotes stochastic integration. □

---

<sup>16</sup>See Theorem 2.2.6 for the technical foundation behind this colorful statement.

*Proof.* From Protter (2005, Thm. I.28, Thm. II.22) where one finds the standard result that  $[W]_t = t$ . The theorem result then follows from the previous theorem and Protter (2005, Thm. II.29).<sup>17</sup> ■

Now we are ready to introduce our secondary variable of interest: integrated variance.

**Definition 2.2.2 (Integrated variance:  $IV$ )**

For  $X \in \mathcal{SM}^+$ ,<sup>18</sup> the integrated variance of  $X$  on a time interval  $[t_0, t_1]$

$$IV_{X,[t_0,t_1]} \equiv \int_{t_0}^{t_1} \sigma_u^2 du, \quad 0 \leq t_0 \leq t_1 < \infty. \text{<sup>19</sup>}$$

For shorthand use,

$$IV_{X,t} \equiv IV_{X,[0,t]}.$$

Where a stochastic process  $X$  is clearly implied,

$$IV_{[t_0,t_1]} \equiv IV_{X,[t_0,t_1]}$$

$$IV_t \equiv IV_{X,t}$$

for shorthand use. □

Thus, in conjunction with the previous theorem, we see that integrated variance is the quadratic variation of the diffusive innovations component of the semimartingale

<sup>17</sup>It is worth noting that by applying expectations across the equation in Protter (2005, Thm. II.29) one arrives at

$$\mathbb{E}\{[\sigma \cdot X]_t\} = \mathbb{E}\left\{\int_0^t \sigma_u^2 d[X]_t\right\},$$

a more general form of the well-known *Itô Isometry*

$$\mathbb{E}\left\{\left(\int_0^t \sigma_u dW_u\right)^2\right\} = \mathbb{E}\left\{\int_0^t \sigma_u^2 du\right\}.$$

Applying expectations to the equation in this theorem gives an Itô Isometry-like result.

<sup>18</sup>Of course,  $\mathcal{SM}^{c+} \subseteq \mathcal{SM}^+$ , so this definition holds whether  $X$  comes from the continuous model or not.

<sup>19</sup>In the notation, for simplicity, we suppressed the dependence of  $IV$  on the process  $X$ , which should not lead to confusion later in the paper.

plus process. But we can say more than that. Since the quadratic variation of a continuous finite variation process  $[A^c] = 0$ , the integrated variance is the quadratic variation of the entire diffusive component of the semimartingale plus process. Thus, integrated variance is invariant to changes in the finite variation drift process.

We introduce integrated variance here because it is technically suitable to do so. But thus far, we have not motivated why it is an important variable of interest. One can peek ahead to Theorem 2.2.5 where we show that integrated variance is the conditional daily return variance when log-prices evolve according to the continuous semimartingale plus model. Moreover, integrated variance is of interest in option pricing, portfolio optimization, risk measurement, and volatility forecasting (see references in Chapter 1).

In the following theorem we show that realized variance consistently estimates the sum of integrated variance and jump variation.

**Theorem 2.2.3 ( $\mathcal{SM}+ : RV \rightarrow IV + JV$ )**

If  $X \in \mathcal{SM}+$ , then

$$[X]_t = [M^c]_t + [J]_t = IV_{X,t} + JV_{X,t}.$$

Then,

$$RV_{X,t}^{(n)} \rightarrow IV_{X,t} + JV_{X,t}$$

where convergence is ucp. □

*Proof.* Results immediately follow from previous theorems. The  $RV$  convergence results follows from the definition of  $[X]$ . ■

From Theorem 2.2.6, we find that this gives a near-consistent estimator of the conditional daily return variance, an quantity of interest in its own right. However, we may wish to estimate integrated variance and jump variation separately. Of course, when we assume jumps  $J \equiv 0$ , the situation is much more straightforward, as in the following theorem:

**Theorem 2.2.4** ( $\mathcal{SM}^{c+} : RV \rightarrow IV$ )

If  $X \in \mathcal{SM}^{c+}$ ,

$$[X]_t = [M]_t = IV_{X,t}.$$

Then,

$$RV_{X,t}^{(n)} \rightarrow IV_{X,t}$$

where convergence is ucp. □

*Proof.* Results immediately follows from previous theorems. The  $RV$  convergence results follows from the definition of  $[X]$ . See [Andersen et al. \(2003, Prop. 3\)](#) for an alternative statement for the latter case in  $n$  dimensions. ■

Here, there is no jump variation with which to concern ourselves. Realized variance consistently estimates integrated variance. From the following theorem we find in the  $J \equiv 0$  case that integrated variance is equal to the conditional daily return variance and, thus, realized variance is also a consistent estimator of conditional daily return variance

**Theorem 2.2.5** ( $\mathcal{SM}^{c+} : RV \rightarrow \text{Var}\{r\}$ )

If  $X \in \mathcal{SM}^{c+}$  with  $(a, \sigma) \perp\!\!\!\perp W$ , then

$$r_t | \mathcal{F}_t \sim \mathcal{N}\left(\int_0^t a_u du, \int_0^t \sigma_u^2 dt\right)$$

where

$$\mathcal{F}_t \equiv \mathcal{F}\{a_u, \sigma_u\}_{u \in [0,t]}^{20}$$

denotes the  $\sigma$ -algebra generated from the sample paths of  $\{a_u, \sigma_u\}$  over  $[0, t]$  and  $\mathcal{N}(\mu, \sigma^2)$  denotes the normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Then

$$\mathbb{E}\{r_t | \mathcal{F}_t\} = \int_0^t a_u du$$

$$\text{Var}\{r_t | \mathcal{F}_t\} = IV_{X,t}.$$

---

<sup>20</sup>This may be confusing as we originally defined  $\mathcal{F}$  in Definition 2.1.1 as the  $\sigma$ -algebra in the definition of the probability space. But, this may be less confusing than introducing new notation.

and so,

$$RV_{X,t}^{(n)} \rightarrow \text{Var}\{r_t \mid \mathcal{F}_t\}.$$

□

*Proof.* See Andersen et al. (2003, Thm. 2), ■

Of course, as mentioned earlier, in the case where we allow jumps  $J \neq 0$ , realized variance will still asymptotically unbiasedly estimate the conditional daily return variance, which is now written as the sum of the integrated variance and jump variation as in the following theorem.

**Theorem 2.2.6 ( $\mathcal{SM}+$ :  $RV \rightarrow \text{Var}\{r\} + \text{error}$ )**

If  $X \in \mathcal{SM}+$  and  $W \perp\!\!\!\perp JV$ , then

$$\text{Var}\{r_t \mid \mathcal{F}_t\} = IV_t + \text{E}\{JV_t \mid \mathcal{F}_t\}$$

and so,

$$RV_{X,t}^{(n)} \rightarrow \text{Var}\{r_t \mid \mathcal{F}_t\} + (\text{E}\{JV_t \mid \mathcal{F}_t\} - JV_t),$$

so that

$$\text{E}\{RV_{X,t}^{(n)} \mid \mathcal{F}_t\} \rightarrow \text{Var}\{r_t \mid \mathcal{F}_t\}.$$

*Proof.* The final line follows from the facts that  $(\text{E}\{JV_t \mid \mathcal{F}_t\} - JV_t)$  is clearly a mean-zero error term and that  $RV \rightarrow QV$  ucp. The previous lines follow immediately from previous theorems. ■

Here it is clear we can interpret integrated variance as the contribution of the pure diffusive component of the log-price process to the conditional daily return variance and jump variation as the contribution of the pure jump component. We would like to estimate these quantities separately. But the estimators of Barndorff-Nielsen and Shephard (2004, 2006) found in Definition 2.2.9 will only have desired consistency properties if we assume the semimartingale plus log-price process satisfies additional

assumptions, in which case we refer to the process as a semimartingale plus plus process, stated formally as follows:

**Definition 2.2.7 (Semimartingale plus plus:  $\mathcal{SM}++$ )**

An  $X \in \mathcal{SM}+$ , written as in Theorem 2.2.3, satisfying

- $\sigma_t > 0$ , for all  $0 \leq t < \infty$
- $\sigma$  càdlàg and locally bounded away from zero
- $IV_t < \infty$ , for all  $0 \leq t < \infty$
- $a, \sigma \perp\!\!\!\perp W$

is a special semimartingale plus plus ( $X \in \mathcal{SM}++$ ) □

We now introduce *realized bipower variation*, which is an intermediate step along the path toward establishing a jump-robust estimator of integrated variance.

**Definition 2.2.8 (RBPV)**

For  $X \in \mathcal{SM}$ , the realized bipower variation on a time interval  $[0, t]$  with partition  $\mathcal{P}_n$

$$RBPV_{X,t}^{(n)} \equiv \sum_{i=2}^n |r_i| |r_{i-1}|, \quad 0 \leq t < \infty.$$

Whereas realized variance sums absolute intraday returns multiplied by the absolute value of themselves (i.e., the squared intraday returns), realized bipower variation sums the absolute intraday returns multiplied by the absolute value of their adjacent return. One should immediately see the significance of this subtle difference after recalling that jump returns are  $O(1)$  while diffusive returns are  $O(\sqrt{\|P\|})$ . Because jumps are assumed not dense, no term in the summation in  $RBPV$  will have two  $O(1)$  jump terms multiplying each other for sufficiently fine partitions  $\mathcal{P}_n$ . Thus, most terms will be  $O(\|P\|)$  while a countable set will be of  $O(\sqrt{\|P\|})$ . Since the number of terms in the partition  $n = O\left(\frac{t}{\|P\|}\right)$ ,  $RBPV = O(t)$  and, thus, will not exhibit any explosive behavior over partitions  $\|\mathcal{P}_n\| \rightarrow 0$ .

Still though, realized bipower variation will not consistently estimate the quadratic variation of the diffusive component of returns, integrated variance. Asymptotically it will be biased down, only capturing a fraction  $\frac{2}{\pi}$  of integrated variance. For an

interpretation of this result, consider analogously the classical calculus problem of maximizing the area of a rectangular region by choosing side lengths subject to the constraint of a constant perimeter. One solves the problem by choosing each side equal to  $\frac{1}{4}$  the fixed perimeter and, thus, making a square. With a fixed return *perimeter* then we should expect that the area  $r_i^2$  should be greater than the area  $|r_i||r_{i-1}|$ .

Of course, the previous only presents a colorful interpretation of the biasedness of realized bipower variation as an estimator of integrated variance. [Barndorff-Nielsen and Shephard \(2004\)](#) define their estimators of integrated variance and jump variation

**Definition 2.2.9** ( $\widehat{IV}, \widehat{JV}$ )

For  $X \in \mathcal{SM}$ , the [Barndorff-Nielsen and Shephard \(2004\)](#) estimators of integrated variance and jump variation on the time interval  $[0, t]$  with partition  $\mathcal{P}_n$

$$\begin{aligned}\widehat{IV}_{\text{BNS04}, X, t}^{(n)} &\equiv \frac{\pi}{2} RBPV_{X, t}^{(n)} \\ \widehat{JV}_{\text{BNS04}, X, t}^{(n)} &\equiv RV_{X, t}^{(n)} - \widehat{IV}_{\text{BNS04}, X, t}^{(n)}\end{aligned}$$

□

and rigorously show their consistency as follows:

**Theorem 2.2.10** ( $\mathcal{SM}++ : \widehat{IV} \rightarrow IV, \widehat{JV} \rightarrow JV$ )

If  $X \in \mathcal{SM}++$ , then

$$\begin{aligned}\widehat{IV}_{\text{BNS04}, X, t}^{(n)} &\rightarrow IV_{X, t} \\ \widehat{JV}_{\text{BNS04}, X, t}^{(n)} &\rightarrow JV_{X, t}\end{aligned}$$

□

*Proof.* See [Barndorff-Nielsen and Shephard \(2004, Theorem 2\)](#) ■

And so, in their work [Barndorff-Nielsen and Shephard](#) derive a mathematically elegant, consistent, jump-robust estimator of integrated variance and a consistent estimator of jump variation. But we suspect there is room for improvement.

In their formulation, [Barndorff-Nielsen and Shephard](#) estimate jump variation as the residual of realized variance after removing their realized bipower variation-based estimate of integrated variance. A problem with this approach is that it does not require the estimate of jump variation to be non-negative. The authors address this issue by shrinking all jump variation estimates below zero to zero. While this approach gives more comprehensible estimates of jump variation, the resulting estimator will end up being biased up in small samples.

Another concern with their approach is that it does little to leverage the sparse structure of jumps. Because they are sparse and more difficult to estimate, we suspect optimal finite sample estimators of jump variation and integrated variance would focus first on the problem of estimating jumps and jump variation and back out integrated variance as a residual rather than vice versa. This observation motivates the development of our naïve shrinkage estimators of jump variation and integrated variance in the following sections.

A final concern with their approach is whether convergence of their estimators to integrated variance and jump variation is fast enough so as to yield reasonable estimators in finite, practical contexts. In this chapter we demonstrate that even on practically fine grids of 1-minute intervals, there are serious deficits in the [Barndorff-Nielsen and Shephard](#) approach to estimation of integrated variance and jump variation. We show in Chapter 3 that optimal finite sample estimators of jump variation and integrated variance, derived from a simple empirical Bayesian model of sparse jumps, can greatly improve upon the estimators of [Barndorff-Nielsen and Shephard](#) in terms of both bias and variance.

### ***2.3 Jump Location Estimation in Previous Work***

As alluded to in the previous section, we look to develop an estimator of jump variation based on the estimation of jump sizes. We then look to back out an estimator of integrated variance as the residual of realized variance after removing the jump

variation estimate. [Fan and Wang \(2007\)](#) introduce an estimator of jump sizes, jump variation, and integrated variance based on hard-thresholding wavelet coefficients by the universal threshold. [Xue and Gençay \(2010\)](#) take this work further in the direction of jump detection by deriving the distribution for a test statistic for jumps based on scaling wavelet coefficients. In the case of the Haar wavelet, this test reduces that of [Lee and Mykland \(2008\)](#). More specifically, the jump statistic of [Lee and Mykland](#) scales each of the returns by a jump-robust estimate of each return interval's volatility, a fraction of the integrated variance over a window. We state it formally as follows:

**Definition 2.3.1 ( $LM_l$ )**

For  $X \in \mathcal{SM}$  on the time interval  $[0, t]$  with partition  $\mathcal{P}_n$ , the modified local [Lee and Mykland \(2008\)](#) jump statistic at time  $t_1 \in (\tau_{i-1}, \tau_i]$

$$LM_{l,X,t_1}^{(n)} \equiv \frac{r_i}{\widehat{\sigma}_i}$$

where

$$\widehat{\sigma}_i^2 \equiv \left( \frac{\tau_i - \tau_{i-1}}{t} \right) \widehat{IV}_{\text{BNS04},X,t}^{(n)}$$

□

As stated this statistic differs from the actual statistic of [Lee and Mykland \(2008\)](#) in several minor ways:

1. We allow partitions  $\mathcal{P}_n$  to be of non-homogenous spacing. This keeps with previous conventions and does not affect later asymptotic results.
2. We replace  $RBPV_{X,t}^{(n)}$  with  $\widehat{IV}_{\text{BNS04},X,t}^{(n)}$ . Effectively, we deflate the statistic of [Lee and Mykland](#) by the multiplicative factor  $\sqrt{\frac{2}{\pi}}$  to arrive at the modified statistic. Using  $\widehat{IV}$  helps with intuition and provides for a cleaner asymptotic convergence result.
3. We consider  $\widehat{IV}$  over  $[0, t]$  instead of  $[t_1 - K, t_1]$ , where  $K$  is a window length.

The specification of [Lee and Mykland](#) allows for a *real-time*<sup>21</sup> jump statistic and

---

<sup>21</sup>Their statistic at time  $t_1$  does not consider data at times  $t > t_1$ .

a desirable asymptotic convergence provided the window length  $K$  decreases at a suitable rate. Because our ultimate interest is to estimate jump variation over a full trading interval, we can take advantage of the full set of data on  $[0, t]$ . Also, because our interest is to compare finite sample properties of jump variation estimators, we can avoid the notational delicacies of an evolving window length  $K$ .

From Theorem 2.2.5 we know that conditional returns over intervals without jumps are normally distributed as

$$r_t | \mathcal{F}_t \sim \mathcal{N}\left(\int_0^t a_u du, \int_0^t \sigma_u^2 dt\right).$$

Thus, we would expect such returns scaled by a consistent estimator of the return interval's variance, say  $\widehat{IV}_{BNS04,t}$  as in Definition 2.3.1, would be normally distributed. Under weak technical assumptions, that is precisely the result Lee and Mykland find, stated formally as follows:

**Theorem 2.3.2 ( $\mathcal{SM}++$  : convergence of  $LM_l$ )**

If  $X \in \mathcal{SM}++$  with

- processes  $a, \sigma$  well-behaved<sup>22</sup>
- jump sizes and their mean and standard deviation processes predictable
- jump sizes independent and identically distributed (i.i.d.)
- jump sizes independent of jump intensity and the innovations process  $W$

and

- there is no jump at time  $t_1$ ,

then

$$LM_{l,X,t_1}^{(n)} \rightarrow \mathcal{N}(0, 1)$$

where convergence is uniform in distribution over all partitions  $\|\mathcal{P}_n\| \rightarrow 0$ .<sup>23</sup> If

<sup>22</sup>See Lee and Mykland (2008, Assumption 1) for a precise statement.

<sup>23</sup>We leave out the detail that the window length over which  $\widehat{IV}$  in  $LM_l$  is calculated must decrease at a suitable rate. The interested reader can see Lee and Mykland (2008, Theorem 1) for a precise statement.

- there is a jump at time  $t_1$ ,

then

$$LM_{l,X,t_1}^{(n)} \rightarrow \infty.$$

□

*Proof.* See Lee and Mykland (2008, Theorem 1, Theorem 2) for a precise statements and proofs. ■

Thus, based on this statistic Lee and Mykland construct a test for the existence of jumps in a trading time interval, stated formally as follows:

**Theorem 2.3.3 (Jump location estimation: local)**

For a partition  $\mathcal{P}_n$ , the tests which each reject the null hypothesis of no jumps in each of the intervals  $(\tau_{i-1}, \tau_i]$  for

$$\left| LM_{l,X,\tau_i}^{(n)} \right| > c_\alpha, \quad i = 1, \dots, n,$$

where

$$c_\alpha \equiv \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)$$

is the  $(1 - \frac{\alpha}{2})$ -th percentile of the normal distribution, each achieve asymptotic significance  $\alpha$  as  $\|\mathcal{P}_n\| \rightarrow 0$ .<sup>24</sup> Furthermore, provided  $\alpha \searrow 0$  (i.e.,  $c_\alpha \nearrow \infty$ ) at a suitable rate as  $\|\mathcal{P}_n\| \rightarrow 0$ , the test achieves asymptotic power equal to 1.<sup>25</sup> □

*Proof.* See Lee and Mykland (2008, Theorem 1–2). ■

But this test is local in nature and would be performed  $n$  times, yielding an uncomfortably large number of false positively detected jumps. For example, suppose an asset trading on the NYSE is known to have no jumps on a particular trading

<sup>24</sup>In other words, the probability of misclassifying a *single* interval as containing a jump tends toward  $\alpha$  as the mesh of the partition tends toward zero

<sup>25</sup>In other words, the probability of misclassifying a *single* interval as not containing a jump tends toward zero.

day. Performing this test at the 1% significance level on the 390 1-minute returns for that asset on that trading day would yield roughly 4 false positively detected jumps. Thus, a more functional test would be based on max order test statistics, taking into account the number of elements in the partition of the trading time interval. [Lee and Mykland \(2008\)](#) define such a statistic, stated formally as follows:

**Definition 2.3.4 ( $LM_g$ )**

For  $X \in \mathcal{SM}$  on the time interval  $[0, t]$  with partition  $\mathcal{P}_n$ , the modified global [Lee and Mykland \(2008\)](#) jump statistic at time  $t_1 \in (\tau_{i-1}, \tau_i]$

$$LM_{g,X,t_1}^{(n)} \equiv \frac{|LM_{l,X,t_1}^{(n)}| - C_n}{S_n}$$

where<sup>26</sup>

$$C_n = (2 \log n)^{1/2} - \frac{\log \pi + \log(\log n)}{2(2 \log n)^{1/2}}, \quad S_n = \frac{1}{(2 \log n)^{1/2}}.$$

□

They derive the following asymptotic results for their global test statistic, finding convergence to a generalized Pareto-like distribution, stated formally as follows:

**Theorem 2.3.5 ( $\mathcal{SM}++$ : convergence of  $LM_g$ )**

If  $X \in \mathcal{SM}++$  satisfies the conditions in [Theorem 2.3.2](#) and

- there is no jump in the time interval  $[0, t]$ ,

then

$$\max_{i \in \{1, \dots, n\}} LM_{g,X,\tau_i}^{(n)} \rightarrow \xi$$

where  $\xi$  is a random variable with cumulative distribution function

$$F_\xi(x) \equiv \Pr\{\xi \leq x\} = \exp(e^{-x})$$

and convergence is uniform in distribution over all partitions  $\|\mathcal{P}_n\| \rightarrow 0$ .<sup>27</sup> If

<sup>26</sup>In their definitions of  $C_n$  and  $S_n$ , [Lee and Mykland \(2008, Lemma 1\)](#) include a factor  $\frac{1}{c}$ , where  $c = \sqrt{\frac{2}{\pi}}$ . This factor does not appear in the definitions here because of the  $c$ -scaling which already occurred in our definition of  $LM_{l,X,t_1}^{(n)}$ .

<sup>27</sup>Again, we leave out details on the window length.

- there is a jump in the time interval  $[0, t]$ ,

then

$$\max_{i \in \{1, \dots, n\}} LM_{g, X, \tau_i}^{(n)} \rightarrow \infty.$$

□

*Proof.* See [Lee and Mykland \(2008, Lemma 1\)](#). ■

Thus, based on this test statistic [Lee and Mykland](#) construct a *globally-informed* test for the existence of jumps in a trading time interval. Before stating the test formally, we must introduce their critical value  $\beta_\alpha$  for the generalized Pareto-like distribution:

**Definition 2.3.6 ( $\beta_\alpha$ )**

Let  $\beta_\alpha$  be the  $(1 - \alpha)$ th percentile of the distribution for  $\xi$ , i.e.,  $\beta_\alpha$  is defined to solve

$$F_\xi(\beta_\alpha) \equiv 1 - \alpha$$

where  $0 < \alpha < 1$ . □

Now we are ready to state the *globally-informed* test for the existence of jumps in a trading time interval of [Lee and Mykland](#) as follows:

**Theorem 2.3.7 (Jump location estimation: global)**

For a partition  $\mathcal{P}_n$ , the set of tests which reject the null hypothesis of no jumps in the interval  $(\tau_{i-1}, \tau_i]$  for

$$LM_{g, X, \tau_i}^{(n)} > \beta_\alpha, \quad i = 1, \dots, n$$

achieves asymptotic global significance  $\alpha$  as  $\|\mathcal{P}_n\| \rightarrow 0$ .<sup>28</sup> Furthermore, provided  $\alpha \searrow 0$  (i.e.,  $\beta \nearrow \infty$ ) at a suitable rate as  $\|\mathcal{P}_n\| \rightarrow 0$ , the set of tests achieve asymptotic global significance of 0 and power equal to 1.<sup>29</sup>

<sup>28</sup>In other words, the probability of misclassifying *any* interval as containing a jump provided *none* of the intervals contains a jump tends toward  $\alpha$  as the mesh of the partition tends toward zero. We use the term *global significance* as the test considers all all intervals simultaneously and, thus, cannot be termed *significance* in the strictest sense.

<sup>29</sup>In other words, the probability of misclassifying *any* interval as containing or not containing a jump tends toward zero.

Alternatively, we may specify the test in terms of  $LM_{l,X,\tau_i}^{(n)}$  as

$$|LM_{l,X,\tau_i}^{(n)}| > C_n + S_n\beta_\alpha, \quad i = 1, \dots, n$$

□

*Proof.* See [Lee and Mykland \(2008, Theorem 3–5\)](#). ■

We write the alternative form of the theorem in terms of  $|LM_{l,X,\tau_i}^{(n)}|$  to unify this global approach with the previous local approach. In both cases, the null hypothesis of no jumps is rejected for  $|LM_{l,X,\tau_i}^{(n)}|$  exceeding sufficiently large thresholds  $\mathbf{t}$ :

**Definition 2.3.8** ( $\mathbf{t}_l, \mathbf{t}_g$ )

$$\begin{aligned} \mathbf{t}_l(\alpha) &\equiv \Phi\left(1 - \frac{\alpha}{2}\right) \\ \mathbf{t}_g(\alpha, n) &\equiv C_n + S_n\beta_\alpha \end{aligned}$$

where  $\Phi$  denotes the cumulative distribution function for a standard normal random variable. □

It is only the case that the thresholds in the global approach are larger in magnitude, as demonstrated by example values in [Table 2.1](#). Notice that thresholds for local tests for jump locations do not depend on the number of elements  $n$  in the partition, as we would expect given previous discussion.

We introduce the following definitions in order to set up an alternative statement about the asymptotic performance of the global test for jump locations.

**Definition 2.3.9** ( $\mathcal{J}, \mathcal{I}, N, \widehat{\mathcal{I}}, \widehat{N}$ )

Let

$$\mathcal{J}_t \equiv \{u \in [0, t] : dN_t \neq 0\}$$

**Table 2.1:** Representative local ( $t_l$ ) and global ( $t_g$ ) thresholds which absolute normalized intraday returns must exceed in order to reject the null hypothesis of jumps in an interval at the  $\alpha$  level of local and global significance.

	$\alpha$	$n$	$\mathbf{t}$
$t_l(\alpha)$	0.05		1.96
	0.01		2.58
$t_g(\alpha, n)$	0.05	390	3.89
	0.01	390	4.36

be the set of jump times in the interval  $[0, t]$ . For a partition  $\mathcal{P}_n$  of  $[0, t]$

$$\mathcal{I}_t^{(n)} \equiv \{i : \mathcal{J}_t \cap (\tau_{i-1}, \tau_i] \neq \emptyset\}$$

is the set of indices corresponding to intervals containing jumps and

$$N_t^{(n)} \equiv |\mathcal{I}_t^{(n)}|$$

is the number of elements in  $\mathcal{I}_t^{(n)}$ . In addition, for a given threshold  $\mathbf{t}$ ,

$$\widehat{\mathcal{I}}_{\mathbf{t}, t}^{(n)} \equiv \left\{ i \in \{1, \dots, n\} : \left| LM_{l, X, \tau_i}^{(n)} \right| > \mathbf{t} \right\}$$

is the set of indices corresponding to intervals classified as having jumps and

$$\widehat{N}_{\mathbf{t}, t}^{(n)} \equiv |\widehat{\mathcal{I}}_{\mathbf{t}, t}^{(n)}|$$

is the number of elements in  $\widehat{\mathcal{I}}_{\mathbf{t}, t}^{(n)}$ . □

Now we are ready to present the following theorem, which states that, asymptotically, we can perfectly identify jump locations using the global jump detection test.

**Theorem 2.3.10 (Jump location estimation: alternative)**

For any  $\epsilon > 0$ ,

$$\Pr \left\{ \widehat{N}_{t_g(\alpha,n),t}^{(n)} = N_t^{(n)}, \left\| \widehat{\mathcal{I}}_{t_g(\alpha,n),t}^{(n)} - \mathcal{I}_t^{(n)} \right\| < \epsilon \right\} \rightarrow 1$$

as  $\|\mathcal{P}_n\| \rightarrow 0, \alpha \rightarrow 0$  at suitable rates. □

*Proof.* This proof is nearly trivial given previous work. See [Fan and Wang \(2007\)](#) for an alternative proof but with wavelets. One should consider the Haar wavelet in order to reproduce a proof for the case here. ■

One can interpret this theorem as combining the size and power results of previous theorems. We mention it here as it unifies the presentation of [Lee and Mykland \(2008\)](#) with that of [Fan and Wang \(2007\)](#).

In the context of jump variation and integrated variance estimation and after reviewing the theory of jump detection estimation, one might naturally wonder how the latter can be used to estimate the former two. That is the subject of the next section where we use jump detection theory to estimate jump sizes and build hard-thresholding and naïve shrinkage estimators of jump variation and back out estimators of integrated variance as residuals of realized variance.

## 2.4 JV and IV Estimation via Jump Location Estimation

The methods presented in the previous section detected jumps in intervals of a partition where associated absolute normalized returns exceeded a given threshold. We can naturally extend this methodology to the estimation of jump sizes by a hard thresholding procedure. That is, for intervals where we detect jumps we estimate the jump size as the return itself and for intervals where we do not detect jumps we estimate the jump size to be zero. We can then build an hard-thresholding estimator of jump variation by taking the sum of the squared estimated jump sizes. We define our estimator formally as follows:

**Definition 2.4.1** ( $\widehat{JV}_{\text{HT}}, \widehat{IV}_{\text{HT}}$ )

For a given partition  $\mathcal{P}_n$  of a time interval  $[0, t]$  and a threshold  $\mathfrak{t}$ , the hard-thresholding estimator of jump variation and integrated variance

$$\begin{aligned}\widehat{JV}_{\text{HT},\mathfrak{t},t}^{(n)} &\equiv \sum_{j \in \widehat{\mathcal{I}}_{\mathfrak{t},t}^{(n)}} r_j^2 \\ \widehat{IV}_{\text{HT},\mathfrak{t},t}^{(n)} &\equiv RV_t^{(n)} - \widehat{JV}_{\text{HT},\mathfrak{t},t}^{(n)}.\end{aligned}$$

□

As we estimate squared jump sizes by squaring estimates of jump sizes, Jensen's inequality suggests that the hard-thresholding estimator of jump variation will be biased down in small samples. Nonetheless, the following theorem tells us that the hard-thresholding estimators are consistent.

**Theorem 2.4.2** ( $\widehat{JV}_{\text{HT}} \rightarrow JV$ )

$$\begin{aligned}\widehat{JV}_{\text{HT},\mathfrak{t}_g(\alpha,n),t}^{(n)} &\rightarrow JV_t \\ \widehat{IV}_{\text{HT},\mathfrak{t}_g(\alpha,n),t}^{(n)} &\rightarrow IV_t\end{aligned}$$

as  $\|\mathcal{P}_n\| \rightarrow 0, \alpha \rightarrow 0$  at suitable rates. □

*Proof.* One can see [Fan and Wang \(2007\)](#) for a proof in the context of wavelets. One would need to use the Haar wavelet in order to relate to the statement here. ■

These estimators represent a simple, first-pass approach to estimating jump variation and integrated variance from estimated jump sizes. Such estimators are even shown to be consistent. Still, we suspect we can do better. One concern with these estimators is the choice of threshold  $\mathfrak{t}$ . One can see from simulation results in [Section 2.7](#) that performance of these estimators varies tremendously across thresholds and that no one threshold categorically outperforms others across simulation models. Another concern is the implied hard-thresholding procedure in the estimation of jump

sizes embedded in the hard-thresholding estimators of jump variation and integrated variance. Such a procedure is too willing to ascribe movements in the log-return to the jump component process. It is a well-known result in thresholding theory that such procedures, though they may be unbiased, will have uncomfortably large variance. Thus, we expect estimators of jump variation based on soft-thresholding or shrinkage of the returns would outperform hard-thresholding estimators in terms of mean absolute error. We prefer shrinkage-based estimators as they do not require a choice of threshold as is the case with soft-thresholding estimators.

A shrinkage estimator of jump size would reduce the magnitude of the log-return to yield an estimate of the jump component. All non-zero returns would lead to non-zero estimates of jump sizes, although only large absolute normalized returns would produce large estimates of jump sizes. For a clear exposition of this topic see [Percival and Walden \(2000, Chapter 10\)](#). One such shrinkage estimators as being less dramatic than hard-thresholding estimators in their estimation of jump sizes. Still, we must decide on a rule for shrinking returns in order to get *good* estimates of jump sizes. [Percival and Walden \(2000, Chapter 10.4\)](#) provide example of a shrinkage rule derived from minimizing the mean squared error over a class of estimators of an underlying signal. But at the level of generality in the semimartingale plus model of log-prices no such model-based shrinkage rule is forthcoming. In [Chapter 3](#) we make additional assumptions about the log-price process which allow us to derive optimal shrinkage estimators of jumps sizes and jump variation. But here, we look to proceed maintaining the level of generality of the semimartingale plus model and derive a non-parametric, naïve shrinkage estimator of jump sizes and jump variation. We call this estimator *naïve* as it does not derive from any specific optimality considerations. However, as we will show, this estimator is not naïve at all, in the sense that it produces consistent estimates of jump variation, has more desirable theoretical properties than the realized bipower variation-based estimator of [Barndorff-Nielsen and Shephard](#), and can be shown to outperform the last in nearly every simulation scenario.

We define our naïve shrinkage estimators of jump variation and integrated variance formally as follows:

**Definition 2.4.3** ( $\widehat{\mathcal{J}}V_{\text{NS}}, \widehat{\mathcal{I}}V_{\text{NS}}$ )

For a given partition  $\mathcal{P}_n$  of a time interval  $[0, t]$ , the naïve-shrinkage estimator of jump variation and integrated variance

$$\begin{aligned}\widehat{\mathcal{J}}V_{\text{NS},t}^{(n)} &\equiv \sum_{i=1}^n [F_{\xi}(LM_{g,X,\tau_i}) r_i]^2 \\ \widehat{\mathcal{I}}V_{\text{NS},t}^{(n)} &\equiv RV_t^{(n)} - \widehat{\mathcal{J}}V_{\text{NS},t}.\end{aligned}$$

□

As  $F_{\xi}(\cdot)$  is a cumulative distribution function for the generalized Pareto-like distribution encountered in Theorem 2.3.5,  $0 < F_{\xi}(LM_{g,X,\tau_i}) < 1$  for all values of the global jump statistic. Thus, we may interpret  $F_{\xi}(LM_{g,X,\tau_i}) r_i$  as a shrunken return. We can construct an uncountably infinite number of shrinkage functions satisfying this property. What makes this particular shrinkage function special is that it tends toward 1 as  $LM_{g,X,\tau_i} \rightarrow \infty$  and 0 for  $LM_{g,X,\tau_i}$  small. So then, consider a return  $r_i$  whose normalization is large in absolute value, which leads to a large  $LM_{g,X,\tau_i}$  and  $F_{\xi}(LM_{g,X,\tau_i}) \approx 1$  and suggests that the return interval  $(\tau_{i-1}, \tau_i]$  contains a jump. The  $r_i^2$  then contributes nearly one-to-one to the naïve shrinkage estimator of jump variation. Alternatively, for  $r_i$  whose normalization is small in absolute value,  $LM_{g,X,\tau_i}$  is small and so  $F_{\xi}(LM_{g,X,\tau_i}) \approx 0$  and suggests that the return interval  $(\tau_{i-1}, \tau_i]$  does not contain a jump. The  $r_i^2$  contributes nearly nothing to the naïve shrinkage estimator of jump variation.

This discussion leads to a desirable property of the naïve shrinkage estimator:

**Theorem 2.4.4** ( $0 \leq \widehat{\mathcal{J}}V_{\text{NS}} \leq RV$ )

$$\begin{aligned}0 &\leq \widehat{\mathcal{J}}V_{\text{NS},t}^{(n)} \leq RV_t^{(n)} \\ 0 &\leq \widehat{\mathcal{I}}V_{\text{NS},t}^{(n)} \leq RV_t^{(n)}\end{aligned}$$

for all partitions  $\mathcal{P}_n$  of the time interval  $[0, t]$ .  $\square$

*Proof.* From the preceding discussion,

$$0 < F_\xi(LM_{g,X,\tau_i}) < 1$$

and so

$$0 < (F_\xi(LM_{g,X,\tau_i}))^2 < 1$$

for all values of the global jump statistic. Multiplying by  $r_i^2$  gives

$$0 < [F_\xi(LM_{g,X,\tau_i}) r_i]^2 < r_i^2$$

and so summation over all  $n$  gives the result for  $\widehat{JV}_{NS,t}^{(n)}$ . The result for  $\widehat{IV}_{NS,t}^{(n)}$  follows immediately as it is calculated as the residual of realized variance.  $\blacksquare$

Recall this is not the case with  $\widehat{JV}_{BNS04,t}^{(n)}$ ,  $\widehat{IV}_{BNS04,t}^{(n)}$  which, in finite samples, can lead to unintuitive, negative estimates of the jump variation and estimates of integrated variance greater than the realized variance.

Another desirable property of the naïve shrinkage estimators is their consistency, shown as follows:

**Theorem 2.4.5** ( $\widehat{JV}_{NS} \rightarrow JV$ )

$$\begin{aligned} \widehat{JV}_{NS,t}^{(n)} &\rightarrow JV_t \\ \widehat{IV}_{NS,t}^{(n)} &\rightarrow IV_t \end{aligned}$$

as  $\|\mathcal{P}_n\| \rightarrow 0$ .  $\square$

*Proof.* To sketch the proof we begin by splitting the summation in the expression for  $\widehat{JV}_{ST}$  into those indices that contain jumps and those that do not and take limits as

$$\lim_{\|\mathcal{P}\| \rightarrow 0} \widehat{JV}_{ST} = \lim_{\|\mathcal{P}\| \rightarrow 0} \sum_{i \in \{K, K+1, \dots, n\} \setminus J} F_\xi(LM_g(i)) r_i^2 + \lim_{\|\mathcal{P}\| \rightarrow 0} \sum_{i \in J} F_\xi(LM_g(i)) r_i^2$$

where  $J$  is a finite set of actual jump indices for any partition. The last term can be rewritten as

$$\begin{aligned} &= \lim_{\|P\| \rightarrow 0} \sum_{i \in J} F_\xi(LM_g(i)) r_i^2 \\ &= \lim_{\|P\| \rightarrow 0} \sum_{i \in J} F_\xi(LM_g(i)) (r_i^c + L_{l(i)})^2 \end{aligned}$$

where  $l(i)$  denotes the function mapping the indices in the partition to the counting numbers  $\{1, 2, \dots, N_T\}$  and  $r_i$  has been decomposed into the return from the continuous diffusion  $r_i^c$  and the return from the jump  $L_{l(i)}$ . We can decompose this expression further as

$$\begin{aligned} &= \lim_{\|P\| \rightarrow 0} \sum_{i \in J} F_\xi(LM_g(i)) (r_i^c)^2 + \lim_{\|P\| \rightarrow 0} \sum_{i \in J} F_\xi(LM_g(i)) 2r_i^c L_{l(i)} + \lim_{\|P\| \rightarrow 0} \sum_{i \in J} F_\xi(LM_g(i)) (L_{l(i)})^2 \\ &= 0 + 0 + \sum_{i \in J} (L_{l(i)})^2 \\ &= JV \end{aligned}$$

The first of these three terms goes to zero because (1)  $r_i^c = O(\sqrt{\|P\|})$  since it is only driven by a Brownian diffusion, (2) the number of elements in  $J$  is finite, and (3)  $0 \leq F_\xi(LM_g(i)) \leq 1$  is bounded. The second of these three terms goes to zero by a similar argument and also noting that each of the  $L_{l(i)}$  are constants unaffected by the partition taken. Finally, the third term converges to the true  $JV$  since  $LM_g(i) \rightarrow \infty$ , implying that  $F_\xi(LM_g(i)) \rightarrow 1$ , as noted earlier.

We complete the sketch of the proof if we can show

$$\lim_{\|P\| \rightarrow 0} \sum_{i \in \{K, K+1, \dots, n\} \setminus J} F_\xi(LM_g(i)) (r_i^c)^2 = 0$$

This is non-trivial. Though  $(r_i^c)^2 = O(\|P\|)$  the number of elements in the summation is  $O\left(\frac{T}{\|P\|}\right)$ . If we ignore the factor  $F_\xi(LM_g(i))$  in the expression then the summation would be of *constant* order when we in fact need it to be of *higher* order so that it

converges to zero. Thus we depend on the structure of  $F_\xi(LM_g(i))$  to produce the desired higher order. We *conjecture* that the *portion* of these elements  $i$  such that  $|F_\xi(LM_g(i))| < \epsilon$ , for some negligibly small  $\epsilon > 0$ , goes to one as  $\|P\| \rightarrow 0$ . If this is true the proof is complete. It remains to be shown. ■

In this section we showed two desirable theoretical properties of the naïve shrinkage estimators. They also agree with the heuristic of focusing efforts first on estimation of the sparse, jump component contribution to variance and diffusive component contribution second. Additionally, they benefit from a more immediately intuitive form, summing shrunken squared returns, than the form of [Barndorff-Nielsen and Shephard](#). One possible drawback of these estimators may be their bias in finite samples. Embedded in the naïve shrinkage estimators is an estimate of squared jump sizes by the square of estimated jumps sizes. Jensen’s inequality then suggests that our naïve shrinkage estimator of jump variation should be biased down in finite samples.<sup>30</sup> In the next few sections we turn to simulation testing to quantify this potential bias and more generally assess the performance of these estimators across a wide range of simulation models. We find excellent performance for the naïve shrinkage estimators across all models.

## 2.5 The Simulation Models

See Section 3.5 for precise definitions and explanations of the simulation models. We postpone presentation until that section as we use notation developed in Chapter 3 to define our simulation models.

## 2.6 Estimator Evaluation Criteria for Simulation Models

See Section 3.6 for precise definitions and explanations of the simulation models evaluation criteria mean percentage (of daily quadratic variation) error MPE (MPE, a

---

<sup>30</sup>In Chapter 3 we address this issue of estimating jump variation by summing the squares of estimated jump sizes and theoretically derive the form of the downward bias.

physically relevant measure of bias) and mean absolute percentage (of daily quadratic variation) error (MAPE, a physically relevant measure of variance).

## 2.7 Simulation Results

With our simulation models set up and estimator evaluation criteria ready we can now present simulation results for the performance of these various estimators of jump variation and integrated variance: hard-thresholding, naïve shrinkage, and realized bipower variation-based. We calculate results for a wide range of simulation models. Though we present a stochastic volatility simulation model in Section 3.5, we do not report results associated with that model here as they add little insight. Essentially, all methods of estimating jump variation and integrated variance perform slightly worse than in the constant volatility case.

In Tables 2.2–2.3 we present simulation results for the constant volatility simulation model with relative jump sizes (conditional on a non-zero jump being drawn) drawn from the  $\mathcal{U}(-4, 4)$  distribution.<sup>31</sup> Thus, conditional on a jump being drawn, the mean absolute relative jump size is 2. Clearly this is small as the relative return for the diffusive component of returns is distribution  $\mathcal{N}(0, 1)$ . Thus, one would easily mistake a mean absolute relative jump with a large relative return for the diffusive component. A quick look at empirical data suggests that conditional relative jumps may be more appropriately drawn from a  $\mathcal{U}(-10, 10)$  distribution.<sup>32</sup> Nonetheless, we still have interest in the robustness of our estimators to changes in jump size. We also hope to gain additional insight into the workings of these estimators by examining them in the case of very small conditional relative jump sizes.

For the case of zero jumps, we see that  $\widehat{\mathcal{J}}_{\text{HT}, \mathbf{t}_g(0.01, 390)}$  performs best in terms of bias and variance (MPE and MAPE). This is unsurprising as  $\mathbf{t}_g(0.01, 390)$  is the

---

<sup>31</sup>*relative* refers to the fact that these jumps are normalized by the volatility of the diffusive component over the relevant return interval. More details can be found in Chapter 3.

<sup>32</sup>For example, with WMT returns over the months January–March in 2008 one finds several 1-minute returns in excess of 10x standard deviation.

largest threshold among the hard-thresholding estimators. This means that a return would have to be considerably large for a jump to be detected in an interval and have that interval's squared return added to the hard-thresholding estimator of jump variation.

In fact, in the case of zero jumps, the best estimators of jump variation and integrated variance would be hard-thresholding estimators with a infinite thresholds, that is, they never detect jumps. Of course, we do not expect these sort of extreme estimators to perform well against alternatives as we add even a moderate number of jumps to the simulation model. Unsurprisingly, we see the performance of  $\widehat{\mathcal{J}}V_{\text{HT},t_g(0.01,390)}$  deteriorate as we add jumps to the simulation. This is a theme across all simulation models regardless of conditional relative jump size  $s$ . Even with only 3 jumps added to the simulation,  $\widehat{\mathcal{J}}V_{\text{NS}}$  outperforms  $\widehat{\mathcal{J}}V_{\text{HT},t_g(0.01,390)}$  (marginally) both in terms of bias and variance.

Also in the case of zero jumps we find that  $\widehat{\mathcal{J}}V_{\text{BNS04}}$  outperforms (marginally)  $\widehat{\mathcal{J}}V_{\text{NS}}$  in terms of bias and variance. This does not surprise us as we noted earlier the realized bipower variation-based estimator of jump variation should be biased down in small samples. That would certainly help the estimator when it is known a priori that jump variation is zero. The true question is how robust are errors in this estimator to jumps. We find that the estimator's downward bias increases as we increase the number of jumps in the model. To be fair, the downward bias of  $\widehat{\mathcal{J}}V_{\text{NS}}$  increases as well but not at so quick a rate. This effect is even more pronounced as the conditional relative jump size  $s$  is increased to more reasonable values.

One surprise which bears mention is the excellent performance of  $\widehat{\mathcal{J}}V_{\text{HT},t_l(0.01)}$  in terms of both bias and variance. Surprisingly, its performance is even robust to the number of jumps. Although, what may be happening is that  $t_l(0.01)$  produce is *just the right threshold*. To understand this last comment, consider extremely high or low thresholds. We would expect the resulting jump variation estimators to be biased down and up respectively. In the former case there are too many false negatives and

the latter case too many false positives. Because hard-thresholding jump variation estimators are monotonically decreasing in the size of the threshold, there must be some *optimal* threshold between the two extremes. What we observe here is that  $\mathfrak{t}_t(0.01)$  might just be that optimal threshold. Of course, this is likely to depend on the model assumed for jumps, a deficit from which  $\widehat{\mathcal{J}}_{\text{NS}}$  would not suffer as it does not depend on some choice of threshold. Another deficit of  $\widehat{\mathcal{J}}_{\text{HT},\mathfrak{t}_t(0.01)}$  can be seen by comparing its performance across the number of jumps starting with zero. At zero, it actually performs worse than almost all the other estimators, over-estimating the amount of jump variation. This is because it has a threshold that is *too low* and falsely detects too many jumps. As the number of jumps increase the bias and the variance of the estimator decrease. But this is because jump variation is increasing as the number of jumps increase and not because the number of jumps correctly being classified increases. Thus, on theoretical grounds,  $\widehat{\mathcal{J}}_{\text{HT},\mathfrak{t}_t(0.01)}$  does not appear as attractive as  $\widehat{\mathcal{J}}_{\text{NS}}$ .

Now consider the case where we increase conditional relative jump size from 4 to 7. In this case, the conditional mean absolute jump is 3.5, much more easily distinguished from a large diffusive component return. In fact, this serves as a more realistic model of jump sizes, as inferred from empirical data previously. Thus, we should take more seriously the results found in Tables 2.4–2.5.

Notice from Tables 2.4–2.5 that  $\widehat{\mathcal{J}}_{\text{BNS04}}$  performs worse than it did in the case of conditional relative jumps size  $s = 4$ . This suggests that the realized bipower variation-based estimator is not robust to increases in conditional relative jump size in finite samples. We can actually see this result extends to the cases of  $s = 10, 15$  by peeking ahead to results found in Tables 2.6–2.9. This lack of robustness is a result of the estimator of jump variation being computed as the residual of realized variance after removing scaled realized bipower variation. Conversely,  $\widehat{\mathcal{J}}_{\text{NS}}$  performs *better* as conditional relative jump sizes increase. This is because jumps should be more easily distinguishable by the Lee and Mykland jump detection statistic. As this is the

**Table 2.2:** MPE for  $s = 4$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.05)}$	27.7	25.4	21.5	12.2
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.01)}$	8.2	6.9	4.3	-2.8
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	-2.2	-7.0	-21.4
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	-3.0	-9.1	-25.3
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	-0.5	-4.0	-15.1
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.3	-2.2	-7.4	-19.0
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.05)}$	-27.7	-25.8	-21.5	-12.5
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.01)}$	-8.3	-7.2	-4.4	2.6
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	-0.3	1.9	7.0	21.1
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	-0.1	2.7	9.1	25.1
$\widehat{I}\mathcal{V}_{\text{NS}}$	-1.4	0.1	3.9	14.8
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	-0.4	1.9	7.3	18.8

basis for the shrinkage function in the naïve shrinkage estimator of jump variation, we would expect its performance in estimation jump variation to improve.

Results for the cases of conditional relative jump sizes of  $s = 10, 15$  in Tables 2.6–2.9 tell mostly similar stories. The realized bipower variation estimator continues to do underperform as jump sizes increase. The naïve shrinkage estimator and other hard thresholding estimators continue to outperform.

It is worth noting at this point the magnitudes of outperformance. For the extreme case of  $s = 15$  and 30 jumps, the mean absolute percentage error of  $\widehat{\mathcal{J}}_{\text{NS}}$  is 4.5 in comparison with 25.3 for  $\widehat{\mathcal{J}}_{\text{BNS04}}$ . That is a tremendous outperformance, made even

**Table 2.3:** MAPE for  $s = 4$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.05)}$	27.7	25.4	21.5	12.3
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.01)}$	8.2	7.0	5.3	5.4
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	2.9	7.4	21.4
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	3.4	9.3	25.3
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	2.1	4.8	15.1
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	3.8	7.6	19.0
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.05)}$	27.7	25.8	21.5	12.5
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.01)}$	9.4	8.6	6.8	5.5
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	5.8	6.1	8.3	21.1
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	5.8	6.3	9.8	25.1
$\widehat{I}\mathcal{V}_{\text{NS}}$	5.9	5.7	6.4	14.9
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	6.7	6.8	8.7	18.8

more tremendous by noting that the *percentage* in mean absolute percentage error is daily quadratic variation and not, the smaller, jump variation. However, this is an unrealistic scenario for 1-minute financial asset returns. More likely we find ourselves in the case of  $s = 10$  with 3 jumps. Here we find values 3.1 and 6.9, so that the naïve shrinkage estimator only modestly outperforms the realized bipower variation-based estimator.

**Table 2.4:** MPE for  $s = 7$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.05)}$	27.7	22.7	15.3	5.0
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.01)}$	8.2	6.2	3.4	-2.0
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	-1.5	-4.5	-16.3
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	-2.5	-7.1	-24.0
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	-0.0	-2.1	-10.3
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.3	-4.5	-12.7	-26.6
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.05)}$	-27.7	-23.1	-15.3	-5.2
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.01)}$	-8.3	-6.6	-3.4	1.7
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	-0.3	1.1	4.5	16.1
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	-0.1	2.1	7.1	23.7
$\widehat{I}\mathcal{V}_{\text{NS}}$	-1.4	-0.4	2.1	10.0
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	-0.4	4.1	12.7	26.4

## 2.8 Empirical Results

We analyze the 1-minute returns of Walmart (WMT) during the 60-day trading period of January 02, 2008 – March 31, 2008. In Figure 2.1 we plot the time series of daily annualized volatility measures in percentage terms. The first notable feature from this figure is that the percentage daily annualized realized volatility of Walmart is usually greater than the typical percentage daily annualized volatility on the S&P 500, 20%. We expect this for any one asset in comparison with a broad index such as the S&P 500. A second notable feature is that the realized bipower variation-based estimator of jump variation turns negative 9 of the 60 days. This suggests the theoretical issue of

**Table 2.5:** MAPE for  $s = 7$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.05)}$	27.7	22.7	15.3	6.0
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.01)}$	8.2	6.4	5.0	4.7
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	3.2	5.9	16.4
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	3.8	8.0	24.0
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	2.7	4.4	10.6
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	5.4	12.7	26.6
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.05)}$	27.7	23.1	15.3	5.5
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.01)}$	9.4	7.9	5.2	3.4
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	5.8	5.5	6.1	16.1
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	5.8	5.9	8.1	23.7
$\widehat{I}\mathcal{V}_{\text{NS}}$	5.9	5.3	4.7	10.1
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	6.7	7.3	12.9	26.4

$\widehat{\mathcal{J}}_{\text{BNS04}}$  possibly being negative is a practical issue as well. A third notable feature is that  $\widehat{\mathcal{J}}_{\text{NS}}$  is typically greater than  $\widehat{\mathcal{J}}_{\text{BNS04}}$ , with the two tracking each other mildly. Simulation results led us to expect this.

To get a better sense of the relationship between  $\widehat{\mathcal{J}}_{\text{NS}}$  and  $\widehat{\mathcal{J}}_{\text{BNS04}}$ , we draw a scatterplot of their values in Figure 2.2. With most values falling below/right of the 45° line, we see clearly that  $\widehat{\mathcal{J}}_{\text{NS}}$  typically exceeds  $\widehat{\mathcal{J}}_{\text{BNS04}}$ . As we would expect, a positive linear relationship between the two appears to exist. Although, there exists a large amount of variability. This suggests we cannot simply solve the problem of  $\widehat{\mathcal{J}}_{\text{BNS04}}$  being biased down in finite samples by scaling it up by a constant factor.

**Table 2.6:** MPE for  $s = 10$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.05)}$	27.7	19.6	10.4	1.9
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.01)}$	8.2	5.2	2.3	-1.4
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	-0.9	-2.4	-10.1
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	-1.6	-4.0	-15.1
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	0.2	-1.0	-6.4
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.3	-6.3	-15.2	-27.5
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.05)}$	-27.7	-20.0	-10.3	-2.2
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.01)}$	-8.3	-5.6	-2.3	1.2
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	-0.3	0.5	2.4	9.9
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	-0.1	1.2	4.1	14.9
$\widehat{I}\mathcal{V}_{\text{NS}}$	-1.4	-0.6	1.0	6.1
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	-0.4	5.9	15.2	27.3

Thus, we prefer the naïve shrinkage estimator.

We report summary statistics for  $\widehat{\mathcal{J}}_{\text{BNS04}}$  and  $\widehat{\mathcal{J}}_{\text{NS}}$  in Table 2.10. Consistent with the earlier story told, all listed quantiles for the empirical distribution of daily  $\widehat{\mathcal{J}}_{\text{NS}}$  as a percentage of daily  $RV$  are greater than the corresponding for  $\widehat{\mathcal{J}}_{\text{BNS04}}$ . Here, we get a sense for the physical significance of the magnitude of the differences between the two. The naïve shrinkage estimator suggests that jumps contribute 15.6% of daily variance while the realized bipower variation-based estimator suggests only 3.6%. With only the realized bipower variation-based estimator in hand, a risk-modeling practitioner may see the number 3.6% and forego the trouble of modeling jumps all-

**Table 2.7:** MAPE for  $s = 10$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_t(0.05)}$	27.7	19.6	10.5	4.1
$\widehat{\mathcal{J}}_{\text{HT},t_t(0.01)}$	8.2	5.8	4.5	4.0
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	3.3	4.8	10.4
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	3.7	5.7	15.2
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	3.1	4.2	6.9
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	6.9	15.2	27.5
$\widehat{I}\mathcal{V}_{\text{HT},t_t(0.05)}$	27.7	20.0	10.4	2.6
$\widehat{I}\mathcal{V}_{\text{HT},t_t(0.01)}$	9.4	6.9	3.8	2.2
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	5.8	4.9	4.2	9.9
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	5.8	5.1	5.2	14.9
$\widehat{I}\mathcal{V}_{\text{NS}}$	5.9	4.8	3.5	6.2
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	6.7	7.8	15.2	27.3

together. After all, they do not appear to contribute much to daily return variance. But the naïve shrinkage tells a different and striking story: jumps do not contribute an only negligible proportion of daily return variance. Jumps must be taken seriously, then, when modeling volatilities, whether it be for the purposes option pricing, risk-arbitrage, portfolio optimization, etc.

## 2.9 Conclusion

We began this chapter by motivating the semimartingale plus model from first principles of financial theory. Our derivation suggested that the model obscured the type

**Table 2.8:** MPE for  $s = 15$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.05)}$	27.7	14.9	5.4	0.4
$\widehat{\mathcal{J}}_{\text{HT},t_i(0.01)}$	8.2	3.8	1.2	-1.0
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	-0.5	-1.1	-5.7
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	-0.9	-1.9	-8.5
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	0.2	-0.4	-3.7
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.3	-8.1	-15.8	-25.3
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.05)}$	-27.7	-15.3	-5.3	-0.6
$\widehat{I}\mathcal{V}_{\text{HT},t_i(0.01)}$	-8.3	-4.2	-1.1	0.8
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.05,390)}$	-0.3	0.1	1.2	5.6
$\widehat{I}\mathcal{V}_{\text{HT},t_g(0.01,390)}$	-0.1	0.5	2.0	8.4
$\widehat{I}\mathcal{V}_{\text{NS}}$	-1.4	-0.6	0.4	3.5
$\widehat{I}\mathcal{V}_{\text{BNS04}}$	-0.4	7.7	15.9	25.1

of jump, whether finite variation or local martingale. Though it had no bearing on our theoretical developments thereafter, investigation into the type of jumps affecting financial log-price processes could prove a promising line of future research as it matters very much for the specific modeling of jumps.

We then presented the gold-standard of integrated variance and jump variation estimation: the realized bipower variation-based estimators of [Barndorff-Nielsen and Shephard](#). We sought finite sample improvements upon their estimators as the realized bipower variation-based estimators are motivated by asymptotic arguments and focus first on estimating *integrated variance*, the more easily modeled quantity in

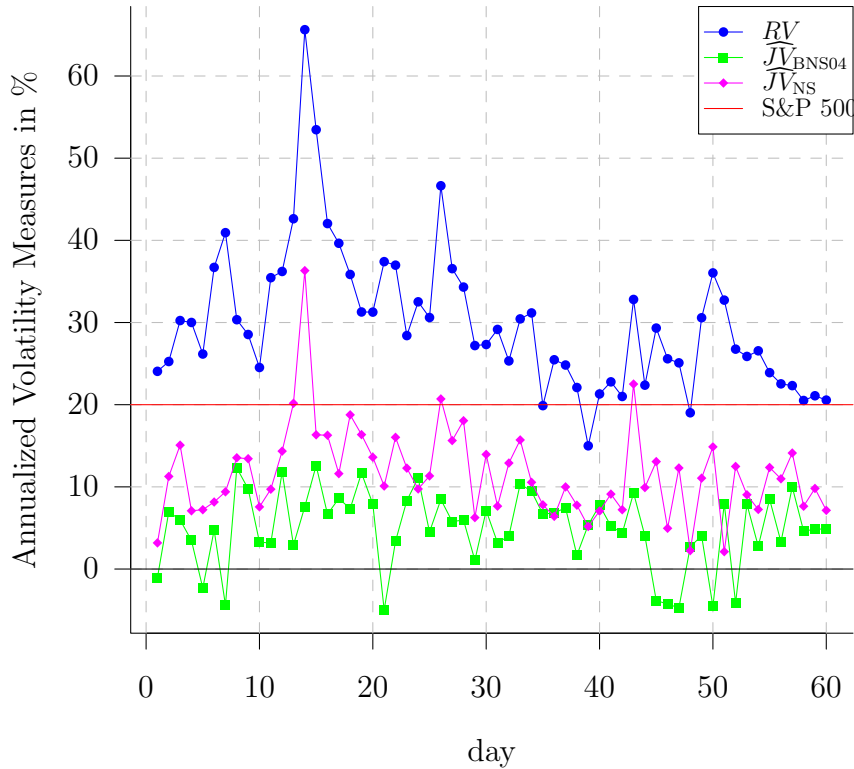
**Table 2.9:** MAPE for  $s = 15$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{HT},t_t(0.05)}$	27.7	15.0	6.1	3.0
$\widehat{\mathcal{J}}_{\text{HT},t_t(0.01)}$	8.2	5.0	3.8	3.1
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.05,390)}$	0.2	3.7	4.1	6.1
$\widehat{\mathcal{J}}_{\text{HT},t_g(0.01,390)}$	0.0	3.8	4.4	8.7
$\widehat{\mathcal{J}}_{\text{NS}}$	1.3	3.6	3.8	4.5
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	8.5	15.9	25.3
$\widehat{\mathcal{I}}_{\text{HT},t_t(0.05)}$	27.7	15.4	5.4	1.0
$\widehat{\mathcal{I}}_{\text{HT},t_t(0.01)}$	9.4	5.4	2.3	1.3
$\widehat{\mathcal{I}}_{\text{HT},t_g(0.05,390)}$	5.8	4.0	2.6	5.6
$\widehat{\mathcal{I}}_{\text{HT},t_g(0.01,390)}$	5.8	4.1	3.0	8.4
$\widehat{\mathcal{I}}_{\text{NS}}$	5.9	3.9	2.2	3.6
$\widehat{\mathcal{I}}_{\text{BNS04}}$	6.7	8.6	15.9	25.1

comparison with jump variation.

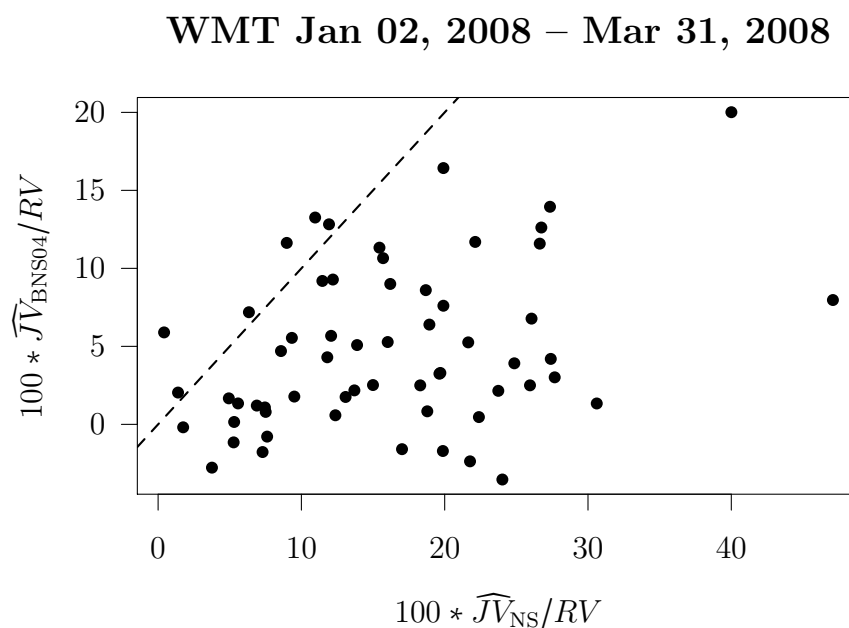
We motivated naïve shrinkage estimators as an intuitive, jump-first approach to estimating jump variation and integrated variance based on shrinking squared returns by the inverse of the asymptotic distribution of the global [Lee and Mykland \(2008\)](#) statistic. Across a wide range of finite sample simulations, we found naïve shrinkage estimators outperformed the realized bipower variation-based estimators both in terms of bias and variance. They were less likely to underestimate jump variation. Empirical studies told the same story: the contribution of jumps to daily return variance is greater when using naïve shrinkage estimators.

### WMT Jan 02, 2008 – Mar 31, 2008



**Figure 2.1:** Time series of daily annualized volatility measures in percentage terms from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008. Thus, daily  $RV$  actually refers to  $100\sqrt{252}\sqrt{RV}$ , the annualized daily realized *volatility*. We follow this with  $\widehat{JV}_{\text{BNS04}}$  and  $\widehat{JV}_{\text{NS}}$  similarly. The red S&P 500 line gives the rough annualized daily volatility of the S&P 500 over the last number of years, plotted for reference.

But we found outperformance of the naïve shrinkage estimators over realized bipower variation-based estimators in a signal processing sense. One may still wonder the physical relevance, in a financial sense, of using one estimator versus the other. We leave open a number of empirical questions for future research: (1) Do using these



**Figure 2.2:** Scatter plot of daily  $\widehat{J}V_{NS}$  vs.  $\widehat{J}V_{BNS04}$  as percentages of daily  $RV$  from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008.

different estimators of daily return variance and its components matter for the forecast of return variance? (2) Will theoretical option prices be substantially affected by using these different estimators? (3) Will using these different estimators impact measures of portfolio risk?

One question left open we answer in the following chapter: how can build a non-naïve shrinkage estimator using a model-based approach?

**Table 2.10:** Empirical summary of daily  $\widehat{\mathcal{J}}V_{\text{BNS04}}$  and  $\widehat{\mathcal{J}}V_{\text{NS}}$  from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008. We report as percentages of daily  $RV$ .

	$100 * \left( \frac{\widehat{\mathcal{J}}V_{\text{BNS04}}}{RV} \right)$	$100 * \left( \frac{\widehat{\mathcal{J}}V_{\text{NS}}}{RV} \right)$
Min.	-3.5	0.4
1st Qu.	1.2	8.9
Median	3.6	15.6
Mean	4.8	16.1
3rd Qu.	8.1	21.9
Max.	20.0	47.1

## Chapter 3

### *JV AND IV ESTIMATION IN A DISCRETE BAYESIAN MODEL*

In Chapter 2 we maintained the generality of the semimartingale plus model and derived the non-parametric naïve shrinkage and hard-thresholding estimators of jump variation and integrated variance which outperformed the estimators of [Barndorff-Nielsen and Shephard](#) in finite samples. We mentioned a drawback of our naïve shrinkage estimator: its ad hoc shrinkage approach. Typically one would derive a shrinkage function from optimality conditions. No such conditions were feasible as we maintained the generality of the semimartingale plus model. But here we borrow the fully-specified discrete empirical Bayesian model of a sparse signal plus noise of [Johnstone and Silverman \(2004, 2005\)](#) and apply this to asset returns, where the *noise* represents the *diffusive*<sup>1</sup> innovations component of asset returns and the *sparse signal* represents the jumps in the log asset prices. Within the context of this model we can derive the minimum mean-squared error estimator of jump variation in closed form and show outperformance over all estimators discussed in previous chapters.

#### **3.1 Discrete Model from Continuous**

Rather than stating outright the discrete empirical Bayesian model of a sparse signal plus noise of [Johnstone and Silverman \(2004, 2005\)](#), we build to it from the semimartingale plus model of Chapter 2. To do this, we must make a number of assumptions. We begin with the following:

---

<sup>1</sup>Of course, this being a discrete model, we can no longer call our innovations component a *diffusive component* as we did in Chapter 2. We include it here once to maintain continuity across chapters.

**Assumption 3.1.1**

(i) (*Zero Drift*).  $a_u = 0, \forall u \in [0, T]$ .

(ii) (*Homogenous Sampling*). An  $(n + 1)$ -element homogenous sampling of  $X_t$  over the time interval  $[0, T]$ ,  $\{X_0, X_\delta, X_{2\delta}, \dots, X_{(n-1)\delta}, X_T\}$ , where  $\delta = T/n$  is the width of the sampling interval.  $\square$

In the high-frequency setting, researchers frequently assume *Zero Drift. Homogenous Sampling* allows for neat exposition and a simple connection to the empirical Bayesian model of Section 3.2.

Using Assumption 3.1.1 and the following definitions:

**Definition 3.1.2 (Continuous  $\rightarrow$  Discrete)**

For  $i = 1, \dots, n$ ,

(i) (*Interval Volatility*).  $\sigma_i \equiv \sigma_{[\delta(i-1), \delta i]}$ .

(ii) (*Scaled Interval Jump*).  $\mu_i \equiv \frac{1}{\sigma_i}(J_{\delta i} - J_{\delta(i-1)})$ .

(iii) (*Log>Returns*).  $r_i \equiv X_{\delta i} - X_{\delta(i-1)}$ .

(iv) (*Scaled Log>Returns*).  $x_i \equiv \frac{1}{\sigma_i} r_i$ .  $\square$

The semimartingale plus model of Chapter 2 reduces to:

**Model 3.1.3 (Discrete, Stochastic Volatility)**

Observe  $r_i = \sigma_i x_i$ , where

$$x_i = \mu_i + \epsilon_i, \quad i = 1, \dots, n,^2$$

with

$$\epsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

where  $\mu_i$  and  $\sigma_i$  are unspecified jump and stochastic volatility processes.  $\square$

---

<sup>2</sup>We do not actually observe the scaled log-returns  $x_i$  but, rather, the log-returns  $r_i = X_{\delta i} - X_{\delta(i-1)} = \sigma_i x_i$ . Though inconsistent with the financial econometrics literature which typically uses  $x_i$  to denote log-returns, this non-standard notation makes simple the transition to the empirical Bayesian model of [Johnstone and Silverman \(2004, 2005\)](#); [Johnstone \(2011\)](#).

Of course, this model is still incompletely specified like the semimartingale plus model in the sense that we do not have enough information to simulate values from the model. All we have done is discretize and redefine variables. Complete specification comes in Section 3.2 where we assume jumps sparse and conditionally drawn from a double exponential distribution. But before, we must say something about the effects of the discretization on our primary variables of interest:

**Definition 3.1.4 (Discrete Approximations/Definitions for  $IV$ ,  $JV$ ,  $QV$ )**

*We may approximate the limiting quantities  $IV$ ,  $JV$ , and  $QV$ , defined in Chapter 2, by*

$$\begin{aligned} IV &\approx \sum_{i=1}^n \sigma_i^2 \\ JV &\approx \sum_{i=1}^n \sigma_i^2 \mu_i^2 \\ QV &\approx \sum_{i=1}^n \sigma_i^2 + \sum_{i=1}^n \sigma_i^2 \mu_i^2. \end{aligned}$$

*Alternatively, we may take these approximations to define  $IV$ ,  $JV$ , and  $QV$ , in Model 3.1.3*  $\square$

Thus, the approximation results from discretization error. In simulation and empirical analyses to follow we take these discretized approximations to  $IV$ ,  $JV$ , and  $QV$ , to define them.

We add another assumption to our model:

**Assumption 3.1.5 (Constant Volatility)**

$$\sigma_u = \bar{\sigma}, \quad \forall u \in [0, T]. \quad \square$$

With this assumption we restrict the price process tremendously. But the assumption of constant volatility represents a frequent starting point of inquiry. Additionally, it allows us to connect to the empirical Bayesian model of [Johnstone and Silverman](#). We re-derive major model results of [Johnstone and Silverman](#) under this assumption but have no problems relaxing it to stochastic volatility.

By assuming constant volatility, we may rewrite the interval volatility, written formally as:

**Result 3.1.6 (Simplified, Redefined Interval Volatility)**

In Model 3.1.3, under Assumptions 3.1.1 and 3.1.5, interval volatility over the  $i$ -th sampling interval reduces to

$$\begin{aligned}\sigma_{[\delta^{(i-1)}, \delta^i]} &= \sqrt{\int_{\delta^{(i-1)}}^{\delta^i} \bar{\sigma}^2 du} \\ &= \bar{\sigma} \sqrt{\delta} \\ &\equiv \sigma.\end{aligned}$$

for  $i = 1, \dots, n$ . □

Using this result, Model 3.1.3 reduces to

**Model 3.1.7 (Discrete, Constant Volatility)**

Model 3.1.3 with  $\sigma_i = \sigma$ ,  $\forall i = 1, \dots, n$ . □

In their form, Models 3.1.3 and 3.1.7 are too simple to yield empirically meaningful results. For example, maximum-likelihood methodology produces the unstable estimates  $\hat{\mu}_i = x_i$  and  $\hat{\sigma} \rightarrow 0$ . But we can use these models as a building blocks for the sufficiently-specified Models 3.2.1 and 3.2.2.

**3.2 The Bayesian Model**

In the following model, we augment Model 3.1.3 to account for our prior belief in sparse, fat-tailed jumps.

**Model 3.2.1 (Discrete, Stochastic Volatility, Empirical Bayesian)**

(Observations).

$$r_i = \sigma_i x_i \quad i = 1, \dots, n,$$

(Likelihood).

$$\mathcal{L}(x_i | \mu_i, \sigma_i) = \phi(x_i - \mu_i) \quad i = 1, \dots, n,$$

where  $\phi$  represents the standard normal density function.

(Prior Density).

$$\pi(\mu_i) = \begin{cases} 1 - w & \text{for } \mu_i = 0 \\ w\gamma(\mu_i) & \text{for } \mu_i \neq 0, \end{cases} \quad i = 1, \dots, n,$$

(with respect to the non-standard dominating measure  $\nu(d\mu_i) = \delta_0(d\mu_i) + d\mu_i$ )

where

$$w = \pi \{ \mu \neq 0 \},^3$$

is the prior jump probability and,

$$\gamma(\mu) = \frac{1}{2}a \exp(-a|\mu|),$$

is the prior density of  $\mu \mid \mu \neq 0$ , taken here as the Laplace( $a$ ) density.

(Hyperparameters:  $\sigma_i, w, a$ ).

Each has a degenerate prior distribution, possibly informed by the observed data  $\{r_i\}$ . □

Similarly, the following model represents the empirical Bayesian extension of Model 3.1.7

### Model 3.2.2 (Discrete, Constant Volatility, Empirical Bayesian)

Model 3.2.1 but with  $\sigma_i = \sigma, \forall i = 1, \dots, n$ . □

Essentially, this is the empirical Bayesian model of a sparse signal plus noise of [Johnstone and Silverman \(2004, 2005\)](#); [Johnstone \(2011\)](#) as applied to high-frequency asset returns. We say *empirical* Bayesian because of the possibility of hyperparameter estimation from observed data. In this way, the model blends frequentist and Bayesian techniques. We follow [Johnstone and Silverman](#) in estimating  $\sigma$  as the scaled median absolute deviation of observed returns  $\sigma x$ ,  $w$  by marginal maximum-likelihood on the scaled returns  $x$ , and setting  $a$  equal to 0.5 a priori.

---

<sup>3</sup>Because  $\mu_i$  and  $x_i$  are i.i.d. in the model, for simplicity, we can drop the  $i$  subscripts and take  $\mu$  and  $x$  to denote representative  $\mu_i$  and  $x_i$ .

### 3.3 Theoretical Results in the Bayesian Model

We have now built the fully-specified empirical Bayesian model of [Johnstone and Silverman](#) from the semimartingale plus model of Chapter 2. With this model, we can derive minimum mean-squared error and absolute error estimators of jump variation, which we do in Section 3.4. Before that, we must re-state some of the main results of [Johnstone and Silverman](#), building to their minimum mean-squared error and absolute error estimators of jump sizes.

We begin by writing the marginal density of  $x \mid \{\mu \neq 0\}$ .<sup>4</sup>

**Result 3.3.1 (Marginal Density of  $x \mid \{\mu \neq 0\}$ :  $g(x)$ )**

$$\begin{aligned} g(x) &\equiv (\phi * \gamma)(x) \\ &\equiv \int_{\mu \neq 0} \phi(x - \mu) \gamma(\mu) \, d\mu. \end{aligned}$$

□

We use the marginal density of  $x \mid \{\mu \neq 0\}$ ,  $g(x)$ , to write the marginal density  $x$  simply as follows:

**Result 3.3.2 (Marginal Density of  $x$ :  $p(x)$ )**

$$\begin{aligned} p(x) &= \int \phi(x - \mu) \pi(d\mu) \\ &= (1 - w)\phi(x) + wg(x). \end{aligned}$$

□

---

<sup>4</sup>We say *We* to maintain consistency. These results, though, are just those of [Johnstone and Silverman](#). We should more appropriately say *They*.

At this point, we leave  $w$  as an unspecified hyperparameter, interpreted as the prior probability of a jump occurring, i.e.,  $\mu \neq 0$ . Ultimately, we estimate  $w$  by marginal maximum likelihood of the posterior density of  $x$ ,  $p(x)$ . This blending of frequentist methods into an otherwise Bayesian model makes for an *empirical* Bayesian model.

We write the posterior density of  $\mu | x$  intuitively as follows:

**Result 3.3.3 (Posterior Density of  $\mu | x$ :  $\pi(\mu | x)$ )**

$$\begin{aligned} \pi(\mu | x) &= \frac{\pi(\mu) \mathcal{L}(x | \mu)}{p(x)} \\ &= \begin{cases} \frac{(1-w)\phi(x)}{p(x)} & \text{for } \mu = 0 \\ \frac{w\gamma(\mu)\phi(x-\mu)}{p(x)} & \text{for } \mu \neq 0. \end{cases} \end{aligned}$$

□

By integrating the posterior density of  $\mu | x$  over  $\mu \neq 0$  we get the posterior non-zero jump probability  $w(x)$ , derived as follows:

**Result 3.3.4 (Posterior Non-Zero Jump Probability:  $w(x)$ )**

$$\begin{aligned} w(x) &\equiv \pi(\mu \neq 0 | x) \\ &= \int_{\mu \neq 0} \frac{w\gamma(\mu)\phi(x-\mu)}{p(x)} d\mu \\ &= \frac{w}{p(x)} \int_{\mu \neq 0} \gamma(\mu)\phi(x-\mu) d\mu \\ &= \frac{wg(x)}{p(x)} \\ &= \frac{wg(x)}{(1-w)\phi(x) + wg(x)} \\ &\rightarrow 1 \quad \text{for large } |x|. \end{aligned}$$

□

which has the intuitive form of being the ratio of the marginal density of  $x \mid (\mu \neq 0)$ , weighted by the prior non-zero jump probability  $w(x)$ , to the marginal density of  $x$ .

By this point, we already defined the posterior density of  $\mu \mid x$ . But by deriving the posterior density of  $\mu \mid (x, \mu \neq 0)$  we can simplify calculations to follow.

**Result 3.3.5 (Posterior Density of  $\mu \mid (x, \mu \neq 0)$ :  $\gamma(\mu \mid x)$ )**

$$\begin{aligned}\gamma(\mu \mid x) &= \frac{\pi(\mu \mid x, \mu \neq 0)}{\pi(\mu \neq 0 \mid x)} \\ &= \frac{\pi(\mu \mid x, \mu \neq 0)}{w(x)} \\ &= \left( \frac{w\gamma(\mu)\phi(x - \mu)}{p(x)} \right) / \left( \frac{wg(x)}{p(x)} \right) \\ &= \frac{\gamma(\mu)\phi(x - \mu)}{g(x)}.\end{aligned}$$

□

With the posterior density of  $\mu \mid (x, \mu \neq 0)$  now derived, we may re-write the posterior density of  $\mu \mid x$  in the more intuitive form, as follows:

**Result 3.3.6 (Posterior Density of  $\mu \mid x$  as a Function of  $w(x)$ :  $\pi(\mu \mid x)$ )**

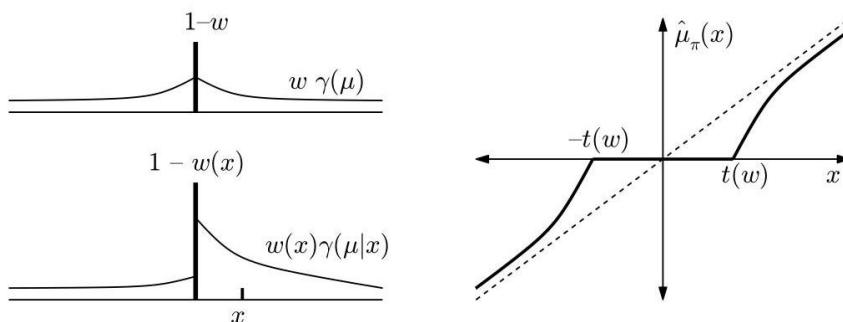
$$\pi(\mu \mid x) = \begin{cases} 1 - w(x) & \text{for } \mu = 0 \\ w(x)\gamma(\mu \mid x) & \text{for } \mu \neq 0, \end{cases}$$

□

One may now see clearly the analogy between the posterior density of  $\mu \mid x$  given here and the prior density of  $\mu \mid x$  found in Model 3.2.1. We simply replace prior parameters and densities  $w$  and  $\gamma(\mu)$  by posterior  $w(x)$  and  $\gamma(\mu \mid x)$ .

The top-left-hand-side subfigure of Figure 3.1 illustrates the prior distribution of jumps  $\mu$  in terms of the prior non-zero jump probability  $w$  and the prior conditional jump density of  $\mu \mid (\mu \neq 0)$ ,  $\gamma(\mu)$ . The figure indicates a  $1 - w$  mass of probability

at  $\mu = 0$ . The bottom-left-hand-side subfigure illustrates the posterior distribution of  $\mu \mid x$  in terms of the posterior non-zero jump probability  $w(x)$  and the posterior conditional jump density of  $\mu \mid (x, \mu \neq 0)$ ,  $\gamma(\mu \mid x)$ . Observing a positive scaled return  $x$  leads us to a posterior distribution with more density on positive scaled jumps  $\mu > 0$  than negative scaled jumps  $\mu < 0$ . The probability density over values  $\mu < 0$  represents our posterior belief that even though we observe a positive scaled return  $x$ , that return is composed of a negative jump  $\mu < 0$  and, necessarily, a positive innovation of magnitude greater than  $\mu$ .



**Figure 3.1:** Illustration from [Johnstone \(2011\)](#). Left: illustration of prior and posterior distributions of jumps  $\mu$ . Right: Illustration of the posterior median of  $\mu \mid x$

Up until this point we dealt with a general prior density for  $\mu$ . [Johnstone and Silverman \(2004, 2005\)](#) allow for two different prior specifications: the Laplace distribution and the quasi-Cauchy distribution. We have chosen to only investigate results for the Laplace prior and leave for future study the quasi-Cauchy distribution.

We now derive the posterior density of  $\mu \mid (x, \mu \neq 0)$  for the Laplace prior as follows:

**Result 3.3.7 (Posterior Density of  $\mu \mid (x, \mu \neq 0)$  for Laplace prior:  $\gamma(\mu \mid x)$ )**

$$\gamma(\mu \mid x) = \begin{cases} \frac{e^{-ax}}{D} \phi(\mu - x + a) & \text{for } \mu > 0 \\ \frac{e^{ax}}{D} \phi(\mu - x - a) & \text{for } \mu \leq 0, \end{cases}$$

where  $D = e^{-ax}\Phi(x - a) + e^{ax}\tilde{\Phi}(x + a)$  and  $\tilde{\Phi}(x) \equiv 1 - \Phi(x)$ . □

Jumps  $\mu$  only enter the posterior density of  $\mu \mid (x, \mu \neq 0)$  through the kernel of the normal distribution  $\phi(\cdot)$ . Noting this, one should see that the density then takes the form of two adjacent truncated normal distributions. This particular form allows for simple closed form calculation of the posterior mean of  $\mu \mid (x, \mu \neq 0)$  for the Laplace prior, written as follows:

**Result 3.3.8 (Posterior Mean of  $\mu \mid (x, \mu \neq 0)$  for Laplace Prior:  $\hat{\mu}_\gamma(x)$ )**

For  $x > 0$

$$\begin{aligned} \hat{\mu}_\gamma(x) &= x - a \frac{\{e^{-ax}\Phi(x - a) - e^{ax}\tilde{\Phi}(x + a)\}}{e^{-ax}\Phi(x - a) + e^{ax}\tilde{\Phi}(x + a)} \\ &\rightarrow x - a \quad \text{for large } x. \end{aligned}$$

□

We include the last line for interpretation's sake. This line says that, asymptotically, the observed scaled return increases one-for-one with the posterior mean of  $\mu \mid (x, \mu \neq 0)$ . Recall that  $a$  represents the scale parameter in the Laplace distribution which is inversely related to the dispersion of the distribution. Thus, increases in  $a$  may be interpreted as us believing that jumps are less dispersive, i.e., they are more tightly distributed around zero. If we believe this, it should then be unsurprising that increases in  $a$  lead to decreases in the absolute posterior mean of  $\mu \mid (x, \mu \neq 0)$  by  $a$ .

We can now use the posterior mean of  $\mu \mid (x, \mu \neq 0)$  in the posterior mean of  $\mu \mid x$ , written as follows:

**Result 3.3.9 (Posterior Mean of  $\mu \mid x$ :  $\hat{\mu}_\pi(x)$ )**

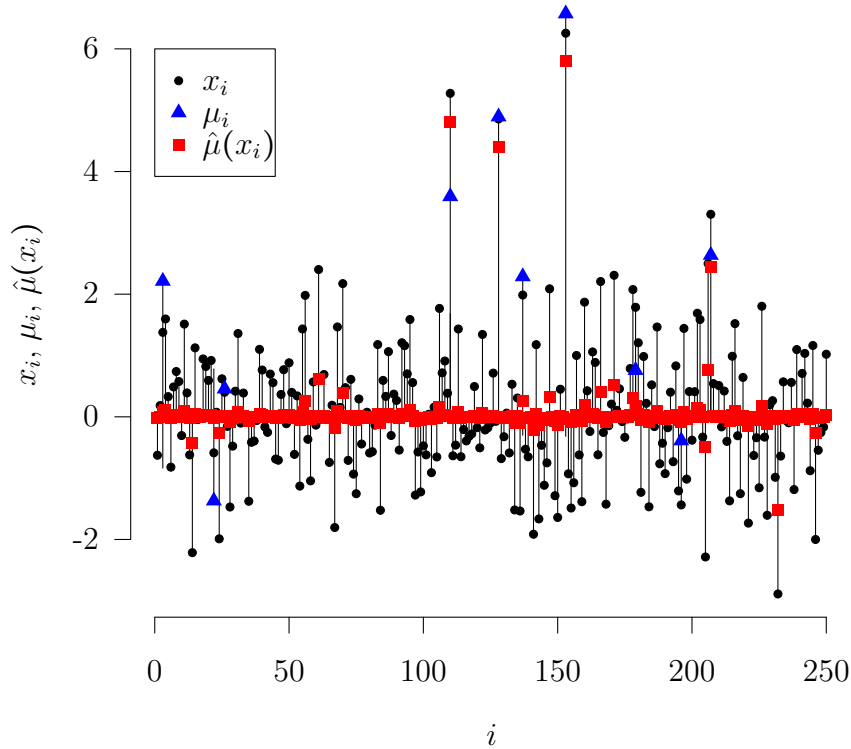
$$\begin{aligned}\hat{\mu}_\pi(x) &= w(x)\hat{\mu}_\gamma(x) \\ &\rightarrow x - a \quad \text{for large } x \text{ and Laplace}(a) \text{ prior.}\end{aligned}$$

□

Written in this way, the posterior mean of  $\mu \mid x$  has a particularly intuitive form. Because we associate the posterior zero-jump probability  $1 - w(x)$  with  $\mu = 0$ , this should have no effect on the posterior mean of  $\mu \mid x$  and hence does not enter the equation above. As a result, we just downweight the posterior mean of  $\mu \mid (x, \mu \neq 0)$  by the probability of non-zero jump occurring. Also, because  $w(x) \rightarrow 1$  at a suitable rate as  $x \rightarrow \infty$ , we find the same asymptotic result as in the posterior mean of  $\mu \mid (x, \mu \neq 0)$ .

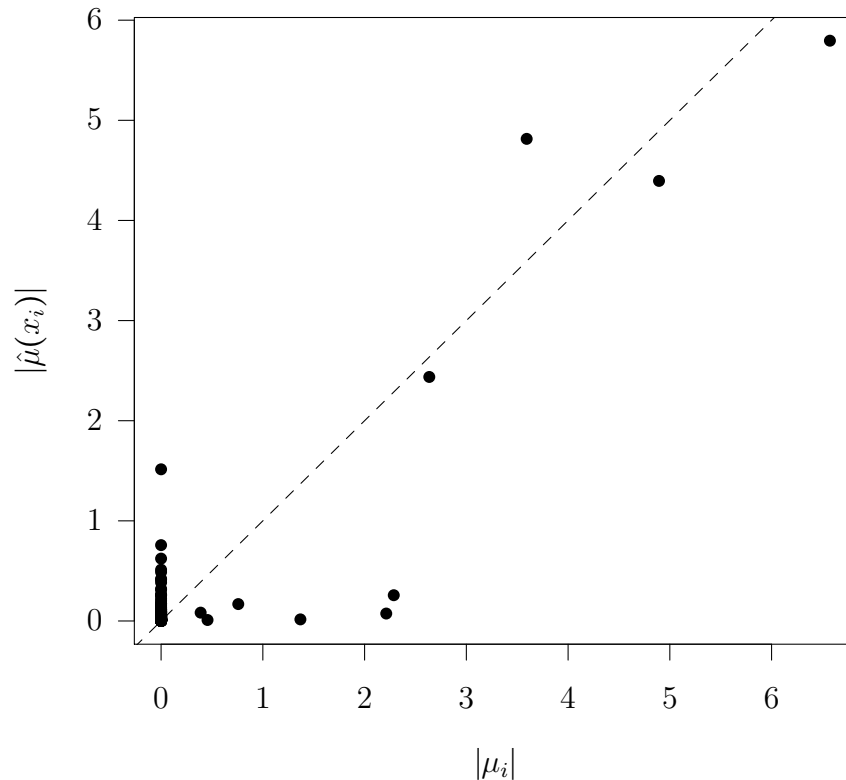
Figure 3.2 plots the time series of 250 simulated jumps  $\mu_i$ , returns  $x_i = \mu_i + \epsilon_i$ , and posterior means of jumps  $\hat{\mu}(x_i)$ , where  $\epsilon_i \sim \mathcal{N}(0, 1)$ , the number of jumps is set deterministically to 10, and the conditional jumps  $\mu_i \mid (\mu_i \neq 0) \sim \mathcal{U}(-7, 7)$ . We see that posterior mean of  $\mu \mid x$  accurately estimates the true jumps  $\mu$ . There are a couple exceptions, among them, a false positive at observation 235 and false negative at observation 5 and 140.

For another illustration of the accuracy of the posterior mean of  $\mu \mid x$  as an estimator of  $\mu$ , one may look to Figure 3.3, a scatter plot of 250 simulated absolute jumps  $|\mu_i|$  against absolute posterior means of jumps  $|\hat{\mu}(x_i)|$ , where innovation returns  $\epsilon_i \sim \mathcal{N}(0, 1)$ , the number of jumps is set deterministically to 10, and the conditional jumps  $\mu_i \mid (\mu_i \neq 0) \sim \mathcal{U}(-7, 7)$ . We see that absolute jumps  $|\mu_i|$  and the corresponding absolute posterior means  $|\hat{\mu}(x_i)|$  roughly line up along the 45° line as desired. A number of non-zero absolute posterior means line up along the line  $|\mu_i| = 0$ . A visualization of the posterior median would have a mass of points at the coordinate  $(0, 0)$ . We give more detail of this property toward the end of this section.



**Figure 3.2:** Time series plot of 250 simulated jumps  $\mu_i$ , returns  $x_i = \mu_i + \epsilon_i$ , and posterior means of jumps  $\hat{\mu}(x_i)$ , where  $\epsilon_i \sim \mathcal{N}(0, 1)$ , the number of jumps is set deterministically to 10, and the conditional jumps  $\mu_i \mid (\mu_i \neq 0) \sim \mathcal{U}(-7, 7)$ .

We omit the form for the posterior median of  $\mu \mid x$ , as in the Section 3.4 we focus energy on the derivation of the posterior mean of jump variation from the posterior means of  $\mu \mid x$ . Specific details on the posterior median can be found in [Johnstone and Silverman \(2004, 2005\)](#); [Johnstone \(2011\)](#). Still, one may gain insight into the form of the posterior median of  $\mu \mid x$  from the right-hand-side subfigure of Figure 3.1. In the figure, we see a thresholding region  $[-t(w), t(w)]$  in which the posterior median is zero. Increases in the non-zero jump probability  $w$  shrink the size of the thresholding



**Figure 3.3:** Scatter plot of 250 simulated absolute jumps  $|\mu_i|$  against absolute posterior means of jumps  $|\hat{\mu}(x_i)|$ , where innovation returns  $\epsilon_i \sim \mathcal{N}(0, 1)$ , the number of jumps is set deterministically to 10, and the conditional jumps  $\mu_i \mid (\mu_i \neq 0) \sim \mathcal{U}(-7, 7)$ . We draw a dashed  $45^\circ$  line for reference.

region, as a larger  $w$  weights more heavily our prior belief that a jump occurs.

The posterior mean of  $\mu \mid x$  has a similar form to the posterior median except it is an increasing function in  $x$  whereas the posterior median is a non-decreasing function. The posterior mean takes on the value of zero only for  $x = 0$ , But for values  $x \in [-t(w), t(w)]$  the posterior mean is only nominally different from zero.

So far, we have only spoken most generally about the estimation procedure. We

have said nothing about the estimation of the  $\sigma_i$  or  $w$  in practice. Before we move forward we present a concrete estimation procedure:

1. Set  $a = 0.5$  a priori
2. Assume  $\sigma_i = \sigma$  a constant and take  $\sigma = 1.48 \text{MAD}\{r_1, \dots, r_n\}$
3. Calculate  $x_i = \frac{r_i}{\sigma}$  for  $i = 1, \dots, n$
4. Calculate  $w$  by marginal maximum likelihood of  $\mathcal{L}(\tilde{w})$ . I.e., take

$$w = \operatorname{argmax}_{0 \leq \tilde{w} \leq 1} \sum_{i=1}^n \log\{(1 - \tilde{w})\phi(x_i) + \tilde{w}g(x_i)\}$$

where

$$g(x) \equiv (\phi * \gamma)(x) \equiv \int \phi(x - \mu)\gamma(\mu) d\mu$$

is interpreted as the *marginal density of  $x \mid \{\mu \neq 0\}$*

5. For  $i = 1, \dots, n$ , solve for the posterior density of  $\mu_i \mid x_i \dots$

[Johnstone and Silverman \(2004, 2005\)](#) use  $\sigma = 1.48 \text{MAD}\{r_1, \dots, r_n\}$  as it is robust to jumps and consistent in its estimation of the volatility of the innovations. [Johnstone and Silverman](#); [Johnstone and Silverman](#) also use  $a = 0.5$ , as it anchors the dispersion of the conditional jump distribution to be large relation to that of the innovations' distribution. This way, the marginal maximum likelihood procedure in choosing  $w$  in the next step may *distinguish* between jumps and normal innovations.

### 3.4 Jump Variation in the Bayesian Model

Now that we restated the basic results of the empirical Bayesian model of a sparse signal plus noise of [Johnstone and Silverman \(2004, 2005\)](#), we can use these results to derive results for our primary variable of interest: jump variation. Recall from Definition 3.1.4 that jump variation

$$JV \equiv \sum_{i=1}^n \sigma_i^2 \mu_i^2$$

and so is a function of constants  $\sigma_i$  (or thought of as random variables with degenerate priors) and random variables  $\mu_i$ , each with the Laplace( $a$ ) prior. Thus, we can derive a

prior distribution for jump variation. Similarly, we may derive a posterior distribution for jump variation from the posterior distributions of the  $\mu_i$ . But because our interest is in posterior point estimates of jump variation (posterior mean and posterior median) we may skip over these messy algebraic details.

A naïve first attempt at a posterior mean and median of jump variation might be the following:

**Definition 3.4.1 (Naïve Posterior Mean and Median of  $JV \mid \mathbf{x}$ )**

$$\widehat{JV}_{NEB,\ell_2}(\mathbf{x}) \equiv \sum_{i=1}^n \sigma_i^2 \hat{\mu}_{\pi,i}(x_i)^2$$

$$\widehat{JV}_{NEB,\ell_1}(\mathbf{x}) \equiv \sum_{i=1}^n \sigma_i^2 \tilde{\mu}_{\pi,i}(x_i)^2$$

□

We say *naïve* as these actually are not the posterior mean and median of jump variation but, rather, the sum of the scaled squared posterior means and medians of the individual jumps. Because the square function is concave, from Jensen's inequality we expect these estimators to underestimate the actual posterior mean and median of jump variation. Nonetheless, they will provide insight into the behavior of the actual posterior mean and median of jump variation. If it turns out (it will not) that a nice closed form for the posterior mean of jump variation does not come forward, we would have to resort to sampling procedures to estimate the posterior mean of jump variation. In this case the naïve estimators may look even more attractive as they provide a computationally cheap alternative to sampling procedures. So, the naïve estimators represent *cheap and dirty* benchmarks against which we should test the other more algebraically involved (but theoretically sound) estimators of jump variation.

As usual, we define our corresponding estimators of integrated variance as the residual of realized variance after removing jump variation.

**Definition 3.4.2 (Naïve Posterior Mean and Median of  $IV \mid \mathbf{x}$ )**

$$\widehat{IV}_{NEB,\ell_2}(\mathbf{x}) \equiv \sum_{i=1}^n r_i^2 - \widehat{JV}_{NEB,\ell_2}(\mathbf{x})$$

$$\widehat{IV}_{NEB,\ell_1}(\mathbf{x}) \equiv \sum_{i=1}^n r_i^2 - \widehat{JV}_{NEB,\ell_1}(\mathbf{x})$$

□

We define the actual posterior mean and median of jump variation as:

**Definition 3.4.3 (Posterior Mean and Median of  $JV \mid \mathbf{x}$ )**

$$\widehat{JV}_{EB,\ell_2}(\mathbf{x}) \equiv \mathbb{E}[JV \mid \mathbf{x}]$$

$$\widehat{JV}_{EB,\ell_1}(\mathbf{x}) \equiv \text{Median}[JV \mid \mathbf{x}].$$

□

Written in this general form, one gains little insight into the workings of the posterior mean and median of jump variation. The following result shows a relationship between the naïve posterior mean and the actual posterior mean of jump variation:

**Result 3.4.4 ( $\widehat{JV}_{NEB,\ell_2}(\mathbf{x})$  vs.  $\widehat{JV}_{EB,\ell_2}(\mathbf{x})$ )**

$$\widehat{JV}_{EB,\ell_2}(\mathbf{x}) = \widehat{JV}_{NEB,\ell_2}(\mathbf{x}) + \sum_{i=1}^n \sigma_i^2 \text{Var}[\mu_i \mid x_i]$$

□

*Proof.*

$$\begin{aligned}
\widehat{\mathcal{J}}_{\text{EB},\ell_2}(\mathbf{x}) &\equiv \text{E}[JV \mid \mathbf{x}] \\
&= \text{E}\left[\sum_{i=1}^n \sigma_i^2 \mu_i^2 \mid \mathbf{x}\right] \\
&= \sum_{i=1}^n \sigma_i^2 \text{E}[\mu_i^2 \mid x_i] \\
&= \sum_{i=1}^n \sigma_i^2 \hat{\mu}_{\pi,i}(x_i)^2 + \sigma_i^2 \text{Var}[\mu_i \mid x_i] \\
&= \widehat{\mathcal{J}}_{\text{NEB},\ell_2}(\mathbf{x}) + \sum_{i=1}^n \sigma_i^2 \text{Var}[\mu_i \mid x_i].
\end{aligned}$$

■

Thus, the naïve estimator of the posterior mean is actually a first order approximation to the posterior mean. It underestimates the actual posterior mean by the variance of the jumps. Thus, for very few or very tightly dispersed jumps, the naïve estimator of the posterior mean should perform adequately.

We now derive the main theoretical result of this chapter: a nice closed-form for the posterior mean of jump variation for the Laplace( $a$ ) prior.

**Result 3.4.5 (Posterior Mean of  $JV \mid \mathbf{x}$ :  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}(\mathbf{x})$ )**

*In Model 3.2.1 we derive the posterior mean of  $JV \mid \mathbf{x}$  as*

$$\widehat{\mathcal{J}}_{\text{EB},\ell_2}(\mathbf{x}) = \sum_{i=1}^n \sigma_i^2 \widehat{\mu}_{\pi}^2(x_i),$$

where  $\widehat{\mu}_{\pi}^2(x_i)$  is the posterior mean of  $\mu_i^2 \mid x_i$ . For Laplace( $a$ ) prior

$$\begin{aligned}
\widehat{\mu}_{\pi}^2(x) &= w(x) \left( x^2 + a^2 + 1 - 2ax \left\{ \frac{e^{-ax}\Phi(x-a) - e^{ax}\tilde{\Phi}(x+a)}{D} \right\} \right. \\
&\quad \left. - 2a \left\{ \frac{e^{-ax}\phi(x-a)}{D} \right\} \right) \tag{3.4.1}
\end{aligned}$$

$$\approx \hat{\mu}_{\pi}(x)^2 + 1 \quad \text{for large } x. \tag{3.4.2}$$

*Proof.*

$$\begin{aligned}
\widehat{JV}_{EB,\ell_2}(\mathbf{x}) &\equiv \mathbb{E}[JV \mid \mathbf{x}] \\
&= \mathbb{E}\left[\sum_{i=1}^n \sigma_i^2 \mu_i^2 \mid \mathbf{x}\right] \\
&= \sum_{i=1}^n \sigma_i^2 \mathbb{E}[\mu_i^2 \mid x_i], && \text{by linearity of expectations and i.i.d.,} \\
&= \sum_{i=1}^n \sigma_i^2 \widehat{\mu_\pi^2}(x_i), && \text{by definition,}
\end{aligned}$$

which gives the first part of the result. For the second part,

$$\begin{aligned}
\widehat{\mu_\pi^2}(x) &\equiv \mathbb{E}[\mu^2 \mid x] \\
&= w(x) \mathbb{E}[\mu^2 \mid x, \mu \neq 0] \\
&\equiv w(x) \widehat{\mu_\gamma^2}(x),
\end{aligned} \tag{3.4.3}$$

where  $\widehat{\mu_\gamma^2}(x)$  is the posterior mean of  $\mu^2 \mid (x, \mu \neq 0)$ .

We can rewrite the density of  $\mu \mid (x, \mu \neq 0)$  under the Laplace(a) prior, given in Result 3.3.7, as

$$\gamma(\mu \mid x) = \begin{cases} \alpha f_{TN}(\mu; x - a, 1, 0, \infty) & \text{for } \mu > 0 \\ (1 - \alpha) f_{TN}(\mu; x + a, 1, -\infty, 0) & \text{for } \mu \leq 0, \end{cases}$$

where  $\alpha = \frac{e^{-ax}\Phi(x-a)}{D}$  and  $f_{TN}(x; \mu, \sigma^2, x_{\text{lb}}, x_{\text{ub}})$  represents the density of a truncated normal random variable with mean  $\mu$ , variance  $\sigma^2$ , lower truncation  $x_{\text{lb}}$ , and upper truncation  $x_{\text{ub}}$ . Then, we can write

$$\begin{aligned}
\widehat{\mu_\gamma^2}(x) &= \alpha \mathbb{E}[TN_U^2] + (1 - \alpha) \mathbb{E}[TN_L^2] \\
&= \alpha \left\{ \mathbb{E}[TN_U]^2 + \text{Var}[TN_U] \right\} + (1 - \alpha) \left\{ \mathbb{E}[TN_L]^2 + \text{Var}[TN_L] \right\}
\end{aligned} \tag{3.4.4}$$

where  $TN_U$  is a truncated normal random variable with density function  $f_{TN}(\mu; x - a, 1, 0, \infty)$  and  $TN_L$  with  $f_{TN}(\mu; x + a, 1, -\infty, 0)$ .

From the formulas for the expectation and variance of a truncated normal random variable we get

$$\begin{aligned}
\mathbb{E}[TN_U] &= x - a + \frac{\phi(x - a)}{\Phi(x - a)} \\
\text{Var}[TN_U] &= 1 - \left(\frac{\phi(x - a)}{\Phi(x - a)}\right)^2 - \left(\frac{\phi(x - a)}{\Phi(x - a)}\right)(x - a) \\
\mathbb{E}[TN_L] &= x + a - \frac{\phi(x + a)}{\tilde{\Phi}(x + a)} \\
\text{Var}[TN_L] &= 1 - \left(\frac{\phi(x + a)}{\tilde{\Phi}(x + a)}\right)^2 + \left(\frac{\phi(x + a)}{\tilde{\Phi}(x + a)}\right)(x + a)
\end{aligned} \tag{3.4.5}$$

We substitute equations (3.4.5) into equation (3.4.4), reduce, substitute into equation (3.4.3), and arrive at the desired result in equation (3.4.1).

To show approximation (3.4.2) note that for large  $x$

$$\begin{aligned}
\frac{e^{-ax}\Phi(x - a) - e^{ax}\tilde{\Phi}(x + a)}{D} &\approx 1 \\
\frac{e^{-ax}\phi(x - a)}{D} &\approx 0 \\
w(x) &\approx 1.
\end{aligned}$$

By substituting these approximations into equation (3.4.1) we get for large  $x$

$$\begin{aligned}
\widehat{\mu}_\pi^2(x) &\approx x^2 + a^2 + 1 - 2ax \\
&= (x - a)^2 + 1 \\
&\approx \hat{\mu}_\pi(x)^2 + 1, \quad \text{by Result 3.3.9}
\end{aligned}$$

■

As alluded to earlier, the trick in showing this result comes from noticing that the posterior means of  $\mu|x$  are distributed as two adjacent truncated normal random variables. Thus, the  $\mu^2|x$  may be calculated in terms of the mean and variance of a truncated normal.

**Result 3.4.6 (Posterior Median of  $JV \mid \mathbf{x}$ :  $\widehat{JV}_{EB,\ell_1}(\mathbf{x})$ )**

Though *Johnstone and Silverman (2004, 2005)* find a closed form (up to an error function) for the posterior median of  $\mu \mid x$ , no such form is forthcoming for  $JV \mid \mathbf{x}$ . We can numerically determine the posterior median of  $JV \mid \mathbf{x}$ ,  $\widehat{JV}_{EB,\ell_1}(\mathbf{x})$ , by standard sampling procedures.  $\square$

As usual, we define our corresponding estimators of integrated variance as the residual of realized variance after removing jump variation.

**Definition 3.4.7 (Posterior Mean and Median of  $IV \mid \mathbf{x}$ )**

$$\widehat{IV}_{EB,\ell_2}(\mathbf{x}) \equiv \sum_{i=1}^n r_i^2 - \widehat{JV}_{EB,\ell_2}(\mathbf{x})$$

$$\widehat{IV}_{EB,\ell_1}(\mathbf{x}) \equiv \sum_{i=1}^n r_i^2 - \widehat{JV}_{EB,\ell_1}(\mathbf{x})$$

 $\square$ 

We must be careful when calling these estimators the posterior mean and median of integrated variance. They are not actually. In the context of the Bayesian problem, the  $\sigma_i$  are taken as known. So  $IV \equiv \sum_{i=1}^n \sigma_i^2$  is known as well. But we prefer to maintain consistency with Chapter 2 in focusing attention on the estimation of jump variation first and backing out an estimate of integrated variance as a residual.

Now that we derived our empirical Bayesian estimators of jump variation and integrated variance, we look to test their performance against other estimators across a number of simulation models.

**3.5 The Simulation Models**

The simulation models against which we test the performance of estimators of jump variation and integrated variance are given by all combinations of the following parameters:

### Model 3.5.1 (Simulation Models)

We make concrete Models 3.1.3 and 3.1.7 by setting

(Burn-in).  $n.\text{burn} = 200$ .

(Number of Observations after Burn-in).  $n = 390$ .

(Observations). For  $i = 1, \dots, n$ ,

$$r_i = \sigma_i x_i,$$

where

$$x_i = \mu_i + \epsilon_i, \quad \epsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

(Interval Volatility).

(stochastic). For  $i = 1, \dots, n$ ,

$$\sigma_i = \sigma_{\delta i} / \sqrt{252 * 390},$$

with  $\sigma_{\delta i}$  sampled from the continuous square root process

$$\sigma_t^2 = \sigma_0^2 + \int_0^t \theta_1 + \theta_2 \sigma_u^2 du + \int_0^t \theta_3 \sigma_u dW_u,$$

where  $\delta = 1/(252 * 390)$ ,  $(\theta_1, \theta_2, \theta_3) = (0.3215, -8.0369, 0.4324)$ , and  $\sigma_0^2 = -\theta_1/\theta_2 = 0.0400$ .

(constant). For  $i = 1, \dots, n$ ,

$$\sigma_i = \sigma = 0.000638 \approx 0.20 / \sqrt{252 * 390}.$$

(Scaled Jumps).  $\mu_i \mid \mu_i \neq 0 \sim \mathcal{U}(-s, s)$ , with  $s$  deterministically set equal to 4, 7, 10, and 15.

(Number of Jumps). Deterministically set equal to 0, 3, 10, and 30.

(Jump Locations). Each uniformly chosen from  $i = 1, \dots, n$

(Number of Simulations).  $m = 5000$

□

We interpret one of the 5000 simulations in Model 3.5.1 as one day’s worth of 1-minute returns of an asset trading on the New York Stock Exchange and, thus, the full set of simulations as  $5000/252 \approx 20$  years of 1-minute returns.

For 1-minute volatility, in the constant volatility case, we choose a value that annualizes to 20%, roughly the annualized volatility for the S&P 500 index. For 1-minute volatility, in the stochastic case, we use the square root process. We took the parameters estimated by Bakshi et al. (2006) from 5-minute US equity data and adjusted them to correspond to 1-minute samplings.

We draw scaled jumps from a uniform distribution. Some benefits of using this distribution are that it: (1) is fairly reasonable as in practice we do not think the distribution of jumps has infinite support, (2) stays consistent with the simulation methods of Johnstone and Silverman (2005), (3) allows us to understand more clearly the impact of jumps on estimator performance, and (4) avoids stacking the simulation deck in our favor by using a distribution other than the Laplace( $a$ ).

We choose supports for the uniform, conditional scaled jump distribution of  $(-4, 4)$ ,  $(-7, 7)$ ,  $(-10, 10)$ , and  $(-15, 15)$ .  $s = 7$  corresponds to the simulation scenario of Johnstone and Silverman (2005). From 1-minute returns on several large NYSE stocks,  $s = 10$  seems the best fit to the data. We choose  $s = 4$  and  $s = 15$  to check the robustness of the estimators to uncharacteristically small or large jumps. In the former case, a mean absolute relative jump would be of size 2—difficult to distinguish from a large innovations component.

We deterministically set the number of jumps equal to 0, 3, 10, and 30. We have clear interest in simulating the typical null hypothesis of a continuous process, i.e., 0 jumps. In other studies (see e.g., Xue and Gençay (2010); Fan and Wang (2007)) one typically finds the case of 3 jumps simulated in tested. But the residuals from stochastic volatility models with up to 3 jumps fit to 1-minute returns of large NYSE stocks suggest more jumps ought to be considered. In the case of 10 jumps in 1-minute returns for one day of trading, one can no longer explain jump behavior

solely by news events (Lee and Mykland, 2008). One must entertain liquidity events contributing jumps, i.e., low levels of liquidity leading to larger trading spreads and price return behavior. We consider the case of 30 jumps merely to check robustness of these procedures to large numbers of jumps. Though an unrealistic assumption for 1-minute financial return data, this may be more appropriate in fields where data may be modeled as a moderately sparse signal (e.g.,  $w = 0.10$ ) plus noise.

To test the performance of the various estimators of jump variation across these simulation models we develop our estimator evaluation criteria, mean percentage error (MPE) and mean absolute percentage error (MAPE), in the next section.

### **3.6 Estimator Evaluation Criteria for Simulation Models**

In this section we look to develop physically relevant criteria for evaluating the performance of estimators of jump variation and integrated variance. The most standard criteria researchers use for evaluating the finite sample performance of estimators are bias and variance. But closed form calculations of the bias and variance for any of these estimators in finite samples are not forthcoming. However, we may obtain approximations by calculating sample bias and variance for simulated models. Still, this is not quite what we want. The sample bias and variance will be unit-less, and, so, will not have any physically relevant interpretation. As follows, we define physically relevant estimator evaluation criteria alternatives to bias and variance (actually, volatility for the latter), mean percentage error (MPE) and mean absolute percentage error (MAPE):

#### **Definition 3.6.1 (MPE and MAPE)**

*We define the mean percentage error and mean absolute percentage error for an*

estimator  $\widehat{JV}$  of  $JV$  as

$$\begin{aligned} \text{MPE}(\widehat{JV}) &= 100 * \text{E} \left[ \frac{\widehat{JV} - JV}{QV} \right] \\ \text{MAPE}(\widehat{JV}) &= 100 * \text{E} \left[ \frac{|\widehat{JV} - JV|}{QV} \right] \end{aligned}$$

and similarly for an estimator  $\widehat{IV}$  of  $IV$

$$\begin{aligned} \text{MPE}(\widehat{IV}) &= 100 * \text{E} \left[ \frac{\widehat{IV} - IV}{QV} \right] \\ \text{MAPE}(\widehat{IV}) &= 100 * \text{E} \left[ \frac{|\widehat{IV} - IV|}{QV} \right] \end{aligned}$$

We define the bias criterion (MPE) so that positive values indicate the estimator is biased up.

Notice that the *percentage* in these criteria is of quadratic variation. Recall from Chapter 2 that the quadratic variation is the conditional return variance over the entire trading interval (taken to be one day in this analysis and, so, conditional *daily* return variance). We chose this as the percentage over a standard percentage for a couple reasons: (1)  $\frac{\widehat{JV}-JV}{JV}$  cannot be evaluated for the null case where  $JV = 0$  and (2) the standard percentage, though it provides a more physically relevant evaluation criterion than its unit-less alternatives, still does not quite aid intuition in the way we would like. We may find a large standard percentage bias in, say, jump variation. But if jump variation itself represents a small component of daily conditional return variance (5–10% according to some estimators) then we would typically not worry about the standard percentage bias. Using a percentage of quadratic variation avoids both of these problems.

Still, with these physically relevant alternatives to bias and variance, closed forms are not forthcoming. Instead, we approximate them from the sample of  $m = 5000$  simulations:

**Result 3.6.2 (Simulation Approximations to MPE and MAPE)**

*Using the output from simulation Model 3.5.1, we may approximate MPE and MAPE*

by

$$\begin{aligned} \text{MPE}(\widehat{JV}) &\approx \frac{100}{m} \sum_{j=1}^m \frac{\widehat{JV}_j - JV_j}{QV_j} \\ \text{MAPE}(\widehat{JV}) &\approx \frac{100}{m} \sum_{j=1}^m \frac{|\widehat{JV}_j - JV_j|}{QV_j} \\ \text{MPE}(\widehat{IV}) &\approx \frac{100}{m} \sum_{j=1}^m \frac{\widehat{IV}_j - IV_j}{QV_j} \\ \text{MAPE}(\widehat{IV}) &\approx \frac{100}{m} \sum_{j=1}^m \frac{|\widehat{IV}_j - IV_j|}{QV_j} \end{aligned}$$

where the subscript  $j$  on a quantity indicates that quantity's value for the  $j$ -th simulation. □

We report these approximations in the simulation results of the next section.

### 3.7 Simulation Results

Now that we defined and derived our estimators, detailed our simulation models, and explained our choice of estimator evaluation criteria, we are ready to report on the performance of estimators of jump variation and integrated variance in terms of MPE and MAPE across the simulation models.

Tables 3.1–3.2 report on results for the constant volatility simulation model with conditional scaled jumps drawn from the  $\mathcal{U}(-4, 4)$  distribution. For 0 jumps, we see that the naïve empirical Bayesian estimators actually perform better than their non-naïve empirical Bayesian counterparts. This does not surprise us as we already showed in Result 3.4.4 that  $\widehat{JV}_{\text{NEB}, \ell_2}$  is biased down by  $\sum_{i=1}^n \sigma_i^2 \text{Var}[\mu_i | x_i]$ . As the posterior  $\mu_i | x_i$  typically is not a near-degenerate distribution at  $\mu \equiv 0$ , even in the case where our simulation gives 0 *true* jumps. And since  $\widehat{JV}_{\text{NEB}, \ell_2} \geq 0$  necessarily, bias will be toward 0, the *true* jump variation from the simulation. However the difference is between MPE's of 0.2–0.3 and 0.8, less than 1% of the day's conditional return variance. So we do not consider these physically relevant differences.

We see that the non-naïve empirical Bayesian estimators of jump variation remain mostly robust to the number of jumps as we increase jumps to 3, 10, and 30 per trading day. We say *mostly* as there is a slight increase in the magnitude of bias. But this is merely an artifact of the conditional scaled jumps being drawn from a uniform rather than Laplace distribution. We see that the naïve estimators and realized bipower variation-based estimators are not robust, growing in both bias and variance (MPE and MAPE) with the number of jumps. We already saw this for the realized bipower variation-based estimator back in Chapter 2. This non-robustness for the naïve estimators is exactly what we would expect after deriving Result 3.4.4.

If we compare these results to those in Tables 2.2–2.3, we see that the naïve shrinkage estimator performs similarly to the empirical Bayesian estimators until the case of 30 jumps. It does not remain robust to the number of jumps as well as the empirical Bayesian estimators. The empirical Bayesian estimators do particularly well in this regard as the marginal maximum likelihood step in estimating  $w$  controls for readily we ascribe scaled return movements to scaled jumps. No feature like this exists in any of the naïve shrinkage, hard-thresholding, or realized bipower variation-based estimators. Of course, 30 jumps in one day’s worth of 1-minute financial returns is empirically unreasonable, but concerns of estimator robustness to the number of jumps remain relevant.

Of course, one should not base much of their judgement of these estimators on the  $s = 4$  scenario, which gives an unrealistically small support for the distribution of conditional scaled jumps. Tables 3.3–3.4 report on results for the more realistic scenario of conditional scaled jumps drawn from the  $\mathcal{U}(-7, 7)$  distribution. We find that the empirical Bayesian estimators perform extraordinarily well against all other estimators in terms of bias and variance, even as the number of jumps increase. With the conditional scaled jump size now larger, the empirical Bayes algorithm can distinguish jumps more easily. Both the naïve empirical Bayesian estimators and the realized bipower variation-based estimator perform worse than they did in the case of

**Table 3.1:** MPE for  $s = 4$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	-2.2	-6.6	-15.7
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	-2.1	-6.5	-15.5
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	-0.6	-2.3	-6.6
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	-0.4	-2.1	-6.3
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.2	-2.0	-7.3	-19.1
$\widehat{\mathcal{I}}_{\text{NEB},\ell_1}$	-0.1	2.1	7.0	15.8
$\widehat{\mathcal{I}}_{\text{NEB},\ell_2}$	-0.2	2.0	6.9	15.6
$\widehat{\mathcal{I}}_{\text{EB},\ell_1}$	-0.6	0.5	2.7	6.7
$\widehat{\mathcal{I}}_{\text{EB},\ell_2}$	-0.7	0.3	2.5	6.4
$\widehat{\mathcal{I}}_{\text{BNS04}}$	-0.1	1.9	7.7	19.1

$s = 4$ . This is a very unattractive feature of these estimators, as we would think larger jumps should be more easily distinguished and, hence, jump variation more accurately estimated. In comparison with Tables 2.4–2.5, the empirical Bayesian estimators still outperform the naïve shrinkage estimator in terms of bias and variance.

Tables 3.5–3.6 report on results for the still realistic scenario of conditional scaled jumps drawn from the  $\mathcal{U}(-10, 10)$  distribution. One interesting phenomenon we observe is that the naïve empirical Bayesian estimators seem to be robust to the size  $s$  of conditional scaled jumps. That is, for large numbers of jumps 10 and 30, the naïve empirical Bayesian estimators do slightly better with  $s = 10, 15$  than they did in the cases of  $s = 4, 7$ . To understand why, we must look back again to Result 3.4.4 where we derive the form of the bias as  $\sum_{i=1}^n \sigma_i^2 \text{Var}[\mu_i | x_i]$ . Conditional on a large

**Table 3.2:** MAPE for  $s = 4$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	2.7	6.9	15.8
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	2.7	6.8	15.5
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	3.3	5.7	8.7
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	3.3	5.6	8.5
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	3.7	7.6	19.1
$\widehat{I}_{\text{NEB},\ell_1}$	5.8	5.9	8.0	15.9
$\widehat{I}_{\text{NEB},\ell_2}$	5.8	5.9	7.9	15.7
$\widehat{I}_{\text{EB},\ell_1}$	6.0	6.1	6.7	8.6
$\widehat{I}_{\text{EB},\ell_2}$	6.0	6.2	6.7	8.5
$\widehat{I}_{\text{BNS04}}$	6.7	6.4	8.7	19.2

relative return  $x_i$ , we can more accurately estimate the jump  $\mu_i$  in relative terms. This manifests as a smaller  $\text{Var}[\mu_i | x_i]$  and, hence, smaller bias. Still, we prefer the non-naïve empirical Bayesian estimators which continue to perform excellently and outperform their naïve counterparts.

Tables 3.7–3.8 report on results for the still realistic scenario of conditional scaled jumps drawn from the  $\mathcal{U}(-15, 15)$  distribution. Here we see very much more of the same: excellent performance of the empirical Bayesian estimators in comparison with their naïve counterparts and the realized bipower variation-based estimator. A look back at Tables 2.8–2.9 show that the empirical Bayesian estimators still perform well against the naïve shrinkage and hard-thresholding alternatives. But the relative performance gain is small. This makes intuitive sense as all the methods based on

**Table 3.3:** MPE for  $s = 7$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	-2.2	-5.9	-11.8
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	-2.1	-5.6	-11.4
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	1.2	0.9	-1.8
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	1.4	1.2	-1.7
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.2	-4.3	-12.6	-26.1
$\widehat{\mathcal{I}}_{\text{NEB},\ell_1}$	-0.1	2.4	5.7	11.8
$\widehat{\mathcal{I}}_{\text{NEB},\ell_2}$	-0.2	2.2	5.4	11.5
$\widehat{\mathcal{I}}_{\text{EB},\ell_1}$	-0.6	-1.1	-1.1	1.9
$\widehat{\mathcal{I}}_{\text{EB},\ell_2}$	-0.7	-1.3	-1.4	1.8
$\widehat{\mathcal{I}}_{\text{BNS04}}$	-0.1	4.4	12.3	26.2

picking jumps, and estimating jump variation from those estimated jumps sizes, will more easily pick out jumps when they are incredibly large in relation to normal innovations. For  $s = 15$ , an individual could likely estimate jump variation accurately by summing the squares of jumps estimated manually from eye-balling a time series visualization of returns. In other words, fancy statistical methods only outperform *very* naïve jump variation estimation methods when the task of estimating jumps is a subtle one. That is the case with most empirical financial asset returns and not the case with the more unrealistic  $s = 15$  scenario. Thus, we still find value in the fancy statistical methods.

**Table 3.4:** MAPE for  $s = 7$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	3.3	6.5	11.8
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	3.3	6.3	11.4
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	3.9	4.8	4.7
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	3.9	4.8	4.7
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	5.3	12.6	26.1
$\widehat{I}_{\text{NEB},\ell_1}$	5.8	5.8	6.6	11.8
$\widehat{I}_{\text{NEB},\ell_2}$	5.8	5.8	6.5	11.5
$\widehat{I}_{\text{EB},\ell_1}$	6.0	6.1	5.1	3.8
$\widehat{I}_{\text{EB},\ell_2}$	6.0	6.1	5.1	3.7
$\widehat{I}_{\text{BNS04}}$	6.7	7.3	12.6	26.2

### 3.8 Empirical Results

We analyze the 1-minute returns of Walmart (WMT) during the 60-day trading period of January 02, 2008 – March 31, 2008. In Table 3.9 we find very different estimates of jump variation of Walmart on January 02, 2008 by  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$ . These estimates range from 23.8% of daily realized variance to 12.1% depending on the length of the volatility estimation window. This suggests that a constant volatility plus jumps model does not sufficiently capture the empirical qualities of financial asset returns. Stochastic volatility matters. Notice, though, that windows of 10 minutes to 30 minutes lead to a small range of jump variation estimates of 12.1% to 15.3% of daily realized variance. This feature remains consistent across trading days and other liquid, large-cap stocks. Thus,  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$  stabilizes around a 30-minute volatility estimation window, which we

**Table 3.5:** MPE for  $s = 10$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{JV}_{\text{NEB},\ell_1}$	0.2	-2.5	-6.3	-10.0
$\widehat{JV}_{\text{NEB},\ell_2}$	0.3	-2.3	-6.0	-9.6
$\widehat{JV}_{\text{EB},\ell_1}$	0.8	1.2	-0.2	-2.8
$\widehat{JV}_{\text{EB},\ell_2}$	0.8	1.4	-0.1	-2.7
$\widehat{JV}_{\text{BNS04}}$	0.2	-6.3	-15.6	-27.4
$\widehat{IV}_{\text{NEB},\ell_1}$	-0.1	2.5	6.3	10.2
$\widehat{IV}_{\text{NEB},\ell_2}$	-0.2	2.3	6.0	9.8
$\widehat{IV}_{\text{EB},\ell_1}$	-0.6	-1.2	0.2	3.0
$\widehat{IV}_{\text{EB},\ell_2}$	-0.7	-1.4	0.1	2.9
$\widehat{IV}_{\text{BNS04}}$	-0.1	6.3	15.5	27.7

use hereafter if not stated outright.

Table 3.10 compares  $\ell_1$  and  $\ell_2$  methods for  $JV$  and  $IV$  estimation for WMT on January, 02, 2008. We see little difference between  $\ell_1$  and  $\ell_2$  methods. This holds similarly across trading days and liquid, large-cap stocks. Thus, hereafter we focus attention on  $\ell_2$  methods as we have a closed form them.

In Figure 2.1 we plot the time series of daily annualized volatility measures in percentage terms. In this figure we revisit Figure 2.1 and add  $\widehat{JV}_{\text{EB},\ell_2}$ . We see that the empirical Bayesian estimator of jump variation is typically larger than that of the naïve shrinkage and realize bipower variation-based estimators. They do appear to be positively correlated. For example, one sees a sharp turn up in jump variation for both the empirical Bayesian and naïve shrinkage estimators on day 14.

**Table 3.6:** MAPE for  $s = 10$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	3.9	6.8	10.1
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	3.9	6.5	9.7
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	4.0	4.3	4.4
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	4.0	4.2	4.3
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	7.1	15.6	27.4
$\widehat{I}_{\text{NEB},\ell_1}$	5.8	5.0	6.6	10.2
$\widehat{I}_{\text{NEB},\ell_2}$	5.8	5.0	6.3	9.8
$\widehat{I}_{\text{EB},\ell_1}$	6.0	5.1	3.6	3.4
$\widehat{I}_{\text{EB},\ell_2}$	6.0	5.1	3.6	3.3
$\widehat{I}_{\text{BNS04}}$	6.7	7.8	15.6	27.7

To get a better sense of the relationship between  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$  and  $\widehat{\mathcal{J}}_{\text{BNS04}}$ , we draw a scatterplot of their values in Figure 3.5. With most values falling below/right of the 45° line, we see clearly that  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$  typically exceeds  $\widehat{\mathcal{J}}_{\text{BNS04}}$ . As we would expect, a positive linear relationship between the two appears to exist. Although, there exists a large amount of variability. This suggests we cannot simply solve the problem of  $\widehat{\mathcal{J}}_{\text{BNS04}}$  being biased down in finite samples by scaling it up by a constant factor. Thus, we prefer the naïve shrinkage estimator.

To get a better sense of the relationship between  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$  and  $\widehat{\mathcal{J}}_{\text{NS}}$ , we draw a scatterplot of their values in Figure 3.6. With most values falling below/right of the 45° line, we see that  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$  typically exceeds  $\widehat{\mathcal{J}}_{\text{NS}}$ . However, we do not see this effect as clearly as in the case of  $\widehat{\mathcal{J}}_{\text{BNS04}}$ . Again, we see a positive linear relationship in the

**Table 3.7:** MPE for  $s = 15$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	-3.0	-5.7	-8.2
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	-2.7	-5.4	-7.9
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	0.4	-1.4	-4.0
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	0.6	-1.3	-4.0
$\widehat{\mathcal{J}}_{\text{BNS04}}$	0.2	-7.8	-16.0	-25.0
$\widehat{\mathcal{I}}_{\text{NEB},\ell_1}$	-0.1	2.9	8.5	8.2
$\widehat{\mathcal{I}}_{\text{NEB},\ell_2}$	-0.2	2.7	8.1	7.9
$\widehat{\mathcal{I}}_{\text{EB},\ell_1}$	-0.6	-0.4	2.1	4.0
$\widehat{\mathcal{I}}_{\text{EB},\ell_2}$	-0.7	-0.6	1.9	4.0
$\widehat{\mathcal{I}}_{\text{BNS04}}$	-0.1	7.8	24.0	25.0

estimators with plenty of variability.

Table 3.11 adds to Table 2.10 a column for summary statistics for daily  $\widehat{\mathcal{J}}_{\text{EB},\ell_2}$ , as a percentage of daily realized variance from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008. The empirical Bayesian estimator of jump variation suggests that the contributions of jumps to daily return variance is even greater than that suggested by the naïve shrinkage estimator. We expected a greater contribution as we argued earlier that the naïve shrinkage estimators are still biased down in estimating jump variation. Table 3.11 tells an even more striking story than Table 2.10: jumps matter quite a lot in their contribution to daily return variance with the median estimate of the jump variation as a percentage of realized variance now up to 25.6% compared with the 3.6% suggested by realized

**Table 3.8:** MAPE for  $s = 15$ , constant volatility

Est.	# Jumps			
	0	3	10	30
$\widehat{\mathcal{J}}_{\text{NEB},\ell_1}$	0.2	4.4	6.2	8.2
$\widehat{\mathcal{J}}_{\text{NEB},\ell_2}$	0.3	4.3	6.0	7.9
$\widehat{\mathcal{J}}_{\text{EB},\ell_1}$	0.8	4.1	4.1	4.5
$\widehat{\mathcal{J}}_{\text{EB},\ell_2}$	0.8	4.1	4.0	4.5
$\widehat{\mathcal{J}}_{\text{BNS04}}$	3.3	8.4	16.0	25.0
$\widehat{\mathcal{I}}_{\text{NEB},\ell_1}$	5.8	4.6	5.7	8.2
$\widehat{\mathcal{I}}_{\text{NEB},\ell_2}$	5.8	4.5	5.5	7.9
$\widehat{\mathcal{I}}_{\text{EB},\ell_1}$	6.0	4.1	2.7	4.0
$\widehat{\mathcal{I}}_{\text{EB},\ell_2}$	6.0	4.1	2.6	4.0
$\widehat{\mathcal{I}}_{\text{BNS04}}$	6.7	8.8	16.0	25.0

bipower variation-based estimators.

### 3.9 Conclusion

In Chapter 2 we developed a naïve shrinkage approach to the estimation of jump variation. A detraction of that approach was its naïveté: the shrinkage function did not emerge from a model-based approach. We began this chapter by translating the incompletely specified, continuous time, semimartingale plus model of Chapter 2 to a fully-specified discrete time model, where the latter could be seen as the Bayesian sparse signal plus noise model of [Johnstone and Silverman \(2004, 2005\)](#). As in Chapter 2, we were motivated to find a jump-first approach to the estimation of jump variation and integrated variance. So, in the context of this model we derived a

**Table 3.9:** Jump *volatility* and integrated *volatility* for WMT on January 02, 2008 obtained from different minute volatility estimation windows. *full* corresponds to using the entire day to estimate each minute’s volatility. *60min* corresponds to using a rolling window of length 120 minutes, centered at the minute of interest, to estimate each minute’s volatility. And so on.

	Window	$100\sqrt{\cdot}$	$100\sqrt{252}\sqrt{\cdot}$	$\%RV$
$\widehat{JV}_{EB,\ell_2}$	full	0.8	12.1	25.4
	60min	0.7	11.7	23.8
	30min	0.6	8.7	13.2
	15min	0.5	8.4	12.1
	10min	0.6	9.4	15.3
	05min	0.7	11.6	23.3
$\widehat{IV}_{EB,\ell_2}$	full	1.3	20.8	74.6
	60min	1.3	21.0	76.2
	30min	1.4	22.4	86.8
	15min	1.4	22.6	87.9
	10min	1.4	22.1	84.7
	05min	1.3	21.1	76.7

closed-form solution for the posterior mean of jump variation and explained a sampling procedure for estimating the posterior median. Then from both we estimated integrated variance as the residual of realized variance after removing the estimate of jump variation. Across a wide range of finite sample simulations, we found the empirical Bayesian estimators outperformed both the naïve shrinkage estimators and the realized bipower variation-based estimators in terms of both bias and variance. From our empirical analysis, empirical Bayesian methods suggested that jumps contribute

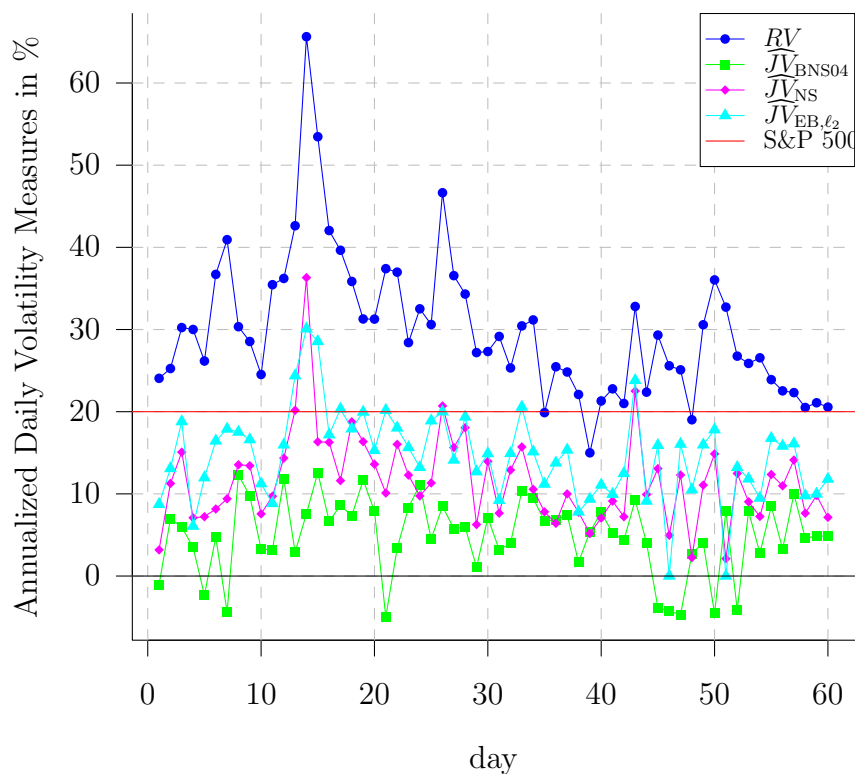
**Table 3.10:** Comparison of  $\ell_1$  and  $\ell_2$  methods for  $JV$  and  $IV$  estimation using 1-minute returns for WMT on January, 02, 2008.

	$100\sqrt{\cdot}$	$100\sqrt{252}\sqrt{\cdot}$	$\%RV$
$RV$	1.5	24.1	100.0
$\widehat{JV}_{EB,\ell_1}$	0.5	8.5	12.5
$\widehat{JV}_{EB,\ell_2}$	0.6	8.7	13.2
$\widehat{JV}_{NS}$	0.2	3.2	1.7
$\widehat{JV}_{BNS04}$	-0.1	-1.0	-0.2
$\widehat{IV}_{EB,\ell_1}$	1.4	22.5	87.5
$\widehat{IV}_{EB,\ell_2}$	1.4	22.4	86.8
$\widehat{IV}_{NS}$	1.5	23.8	98.3
$\widehat{IV}_{BNS04}$	1.5	24.1	100.2

**Table 3.11:** Empirical summary of daily  $\widehat{JV}_{BNS04}$ ,  $\widehat{JV}_{NS}$ , and  $\widehat{JV}_{EB,\ell_2}$ , from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008. We report as percentages of daily  $RV$ .

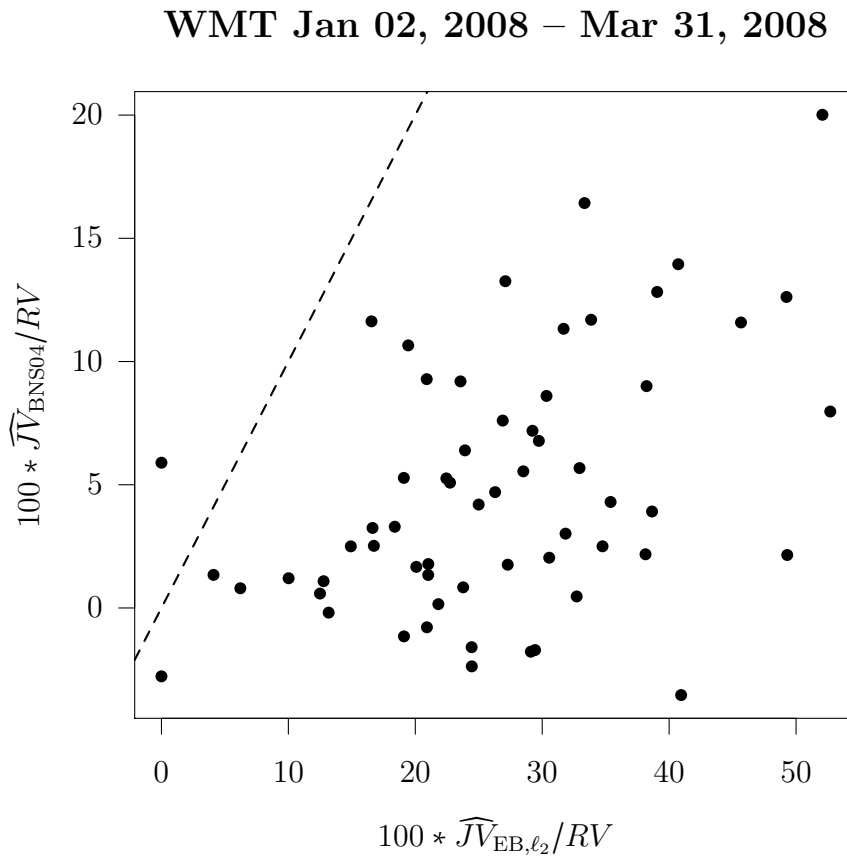
	$100 * \left( \frac{\widehat{JV}_{BNS04}}{RV} \right)$	$100 * \left( \frac{\widehat{JV}_{NS}}{RV} \right)$	$100 * \left( \frac{\widehat{JV}_{EB,\ell_2}}{RV} \right)$
Min.	-3.5	0.4	0.0
1st Qu.	1.2	8.9	19.4
Median	3.6	15.6	25.6
Mean	4.8	16.1	26.3
3rd Qu.	8.1	21.9	33.0
Max.	20.0	47.1	52.7

### WMT Jan 02, 2008 – Mar 31, 2008



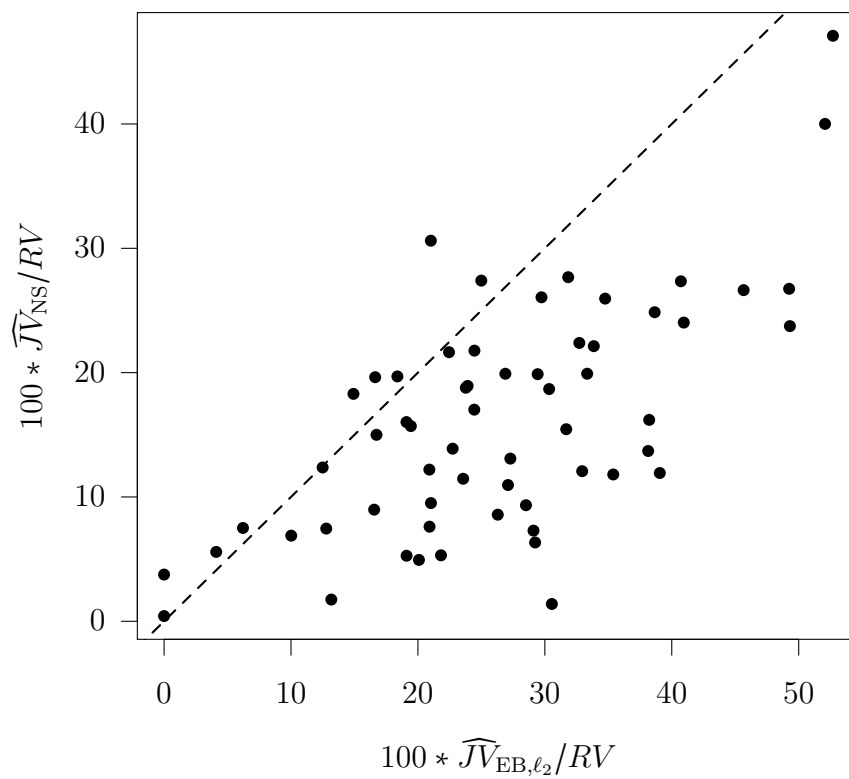
**Figure 3.4:** Time series of daily annualized volatility measures in percentage terms from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008. Thus, daily  $RV$  actually refers to  $100\sqrt{252}\sqrt{RV}$ , the annualized daily realized *volatility*. We follow this with  $\widehat{JV}_{BNS04}$ ,  $\widehat{JV}_{NS}$ , and  $\widehat{JV}_{EB,\ell_2}$  similarly. The red S&P 500 line gives the rough annualized daily volatility of the S&P 500 over the last number of years, plotted for reference.

about 26% of daily return variance. This stands in large contrast to the 3.6% from the realized bipower variation-based estimators.



**Figure 3.5:** Scatter plot of daily  $\widehat{J}V_{EB, \ell_2}$  vs.  $\widehat{J}V_{BNS04}$  as percentages of daily  $RV$  from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008.

## WMT Jan 02, 2008 – Mar 31, 2008



**Figure 3.6:** Scatter plot of daily  $\widehat{J}_{EB, \ell_2}$  vs.  $\widehat{J}_{NS}$  as percentages of daily  $RV$  from 1-minute returns of WMT over the 60 trading day period January 02, 2008 – March 31, 2008.

## Bibliography

- Aguiar-Conraria, L., N. Azevedo, and M. Soares (2008). Using wavelets to decompose the time-frequency effects of monetary policy. *Physica A: Statistical Mechanics and its Applications* 387(12), 2863–2878.
- Aguiar-Conraria, L. and M. Soares (2007). Using cross-wavelets to decompose the time-frequency relation between oil and the macroeconomy. *Working*.
- 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.
- Almasri, A. and G. Shukur (2003). An illustration of the causality relation between government spending and revenue using wavelet analysis on Finnish data. *Journal of Applied Statistics* 30(5), 571–584.
- Andersen, T. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.
- Andersen, T., T. Bollerslev, and F. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89(4), 701–720.
- Andersen, T., T. Bollerslev, and F. Diebold (2009). Parametric and Nonparametric Volatility Measurement. In Y. Aït-Sahalia and L. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and techniques*, pp. 67–138. North-Holland.
- Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71(2), 579–625.

- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001, March). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96(453), 42–55.
- Ariño, M. (1996). Forecasting time series via discrete wavelet transform. *Computing in Economics and Finance*.
- Arnéodo, A., J. Muzy, and D. Sornette (1998). “Direct” causal cascade in the stock market. *The European Physical Journal B-Condensed Matter and Complex Systems* 2(2), 277–282.
- Aussem, A., J. Campbell, and F. Murtagh (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Intelligence in Finance* 6(2), 5–12.
- Aussem, A. and F. Murtagh (1997). Combining neural network forecasts on wavelet-transformed time series. *Connection Science* 9(1), 113–122.
- Back, K. (1991). Asset pricing for general processes. *Journal of Mathematical Economics* 20(4), 371–395.
- Bakshi, G., N. Ju, and H. Ou-Yang (2006). Estimation of continuous-time models with an application to equity volatility dynamics. *Journal of Financial Economics* 82(1), 227–249.
- Barndorff-Nielsen, O. and N. Shephard (2001). Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2), 167–241.
- Barndorff-Nielsen, O. 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.

- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal* 12(3), C1–C32.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17(5), 457–477.
- Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2(1), 1–37.
- Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4(1), 1–30.
- Bäuerle, N. and U. Rieder (2007). Portfolio optimization with jumps and unobservable intensity process. *Mathematical Finance* 17(2), 205–224.
- Bayraktar, E., H. Poor, and R. Sircar (2007). Estimating the Fractal Dimension of the S&P 500 Index Using Wavelet Analysis. *Working*.
- Biais, B., L. Glosten, and C. Spatt (2005). Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets* 8(2), 217–264.
- Bollerslev, T., M. Gibson, and H. Zhou (2011). Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* 160(1), 235–245.
- Bowden, R. and J. Zhu (2008). The agribusiness cycle and its wavelets. *Empirical Economics* 34(3), 603–622.
- Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis* 51(4), 2232–2245.

- Campbell, J. Y., A. W. Lo, A. C. MacKinlay, and A. Y. Lo (1996). *The Econometrics of Financial Markets*. Princeton University Press.
- Capobianco, E. (2004). Multiscale Analysis of Stock Index Return Volatility. *Computational Economics* 23(3), 219–237.
- Chauhan, N., V. Ravi, and D. Karthik Chandra (2009). Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Systems with Applications* 36(4), 7659–7665.
- Chen, G., Y. Choi, and Y. Zhou (2008). Detections of changes in return by a wavelet smoother with conditional heteroscedastic volatility. *Journal of Econometrics* 143(2), 227–262.
- Conlon, T., M. Crane, and H. Ruskin (2008). Wavelet multiscale analysis for Hedge Funds: Scaling and strategies. *Physica A: Statistical Mechanics and its Applications* 387(21), 5197–5204.
- Connor, J. and R. Rossiter (2005). Wavelet Transforms and Commodity Prices. *Studies in Nonlinear Dynamics & Econometrics* 9(1), 433–465.
- Crowley, P., D. Maraun, and D. Mayes (2006). How hard is the euro area core? An evaluation of growth cycles using wavelet analysis. *Bank of Finland Discussion Paper* 18.
- Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys* 21(2), 207–267.
- Dacorogna, M., R. Gençay, U. A. Müller, R. Olsen, and O. Pictet (2001). *An Introduction to High-Frequency Finance* (1 ed.). Academic Press.
- Dalkir, M. (2004). A new approach to causality in the frequency domain. *Economics Bulletin* 3(44), 1–14.

- Daubechies, I. (1992). *Ten Lectures on Wavelets* (1 ed.). SIAM: Society for Industrial and Applied Mathematics.
- Davidson, R., W. Labys, and J. Lesourd (1998). Wavelet Analysis of Commodity Price Behavior. *Computational Economics* 11(1), 103–128.
- Delbaen, F. and W. Schachermayer (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann* 300, 463–520.
- Dellacherie, C. and P. Meyer (1982). *Probabilities and Potential B. Theory of Martingales*. Elsevier Science Ltd.
- Duchesne, P. (2006a). On testing for serial correlation with a wavelet-based spectral density estimator in multivariate time series. *Econometric Theory* 22(04), 633–676.
- Duchesne, P. (2006b). Testing for multivariate autoregressive conditional heteroskedasticity using wavelets. *Computational statistics & data analysis* 51(4), 2142–2163.
- Dumitru, A. and G. Urga (2012). Identifying jumps in financial assets: A comparison between nonparametric jump tests. *Journal of Business & Economic Statistics* 30(2), 242–255.
- Durai, S. and S. Bhaduri (2009). Stock prices, inflation and output: Evidence from wavelet analysis. *Economic Modelling* 26(5), 1089–1092.
- Elder, J. and H. Jin (2007). Long memory in commodity futures volatility: A wavelet perspective. *Journal of Futures Markets* 27(5), 411–437.
- Engle, R. (2000). The econometrics of ultra-high-frequency data. *Econometrica* 68(1), 1–22.

- Engle, R. and J. Russell (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance* 4(2), 187–212.
- Engle, R. and J. Russell (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.
- Fan, J. and Y. Wang (2007). Multi-Scale jump and volatility analysis for High-Frequency financial data. *Journal of the American Statistical Association* 102(480), 1349–1362.
- Fan, Y. and R. Gençay (2010). Unit root tests with wavelets. *Econometric Theory online*, 1–27.
- Fan, Y. and B. Whitcher (2003). A wavelet solution to the spurious regression of fractionally differenced processes. *Applied Stochastic Models in Business and Industry* 19(3), 171–183.
- Fernandez, V. (2005a). The international capm and a wavelet-based decomposition of value at risk. *Studies in Nonlinear Dynamics & Econometrics* 9(4), 1–35.
- Fernandez, V. (2005b). Time-scale decomposition of price transmission in international markets. *Emerging Markets Finance and Trade* 41(4), 57–90.
- Fernandez, V. (2006). The CAPM and value at risk at different time-scales. *International Review of Financial Analysis* 15(3), 203–219.
- Fernandez, V. (2008a). Multi-period hedge ratios for a multi-asset portfolio when accounting for returns co-movement. *Journal of Futures Markets* 28(2), 182–207.
- Fernandez, V. (2008b). The war on terror and its impact on the long-term volatility of financial markets. *International Review of Financial Analysis* 17(1), 1–26.

- Fernandez, V. and B. M. Lucey (2007). Portfolio management under sudden changes in volatility and heterogeneous investment horizons. *Physica A: Statistical Mechanics and its Applications* 375(2), 612–624.
- Fryzlewicz, P., S. van Bellegem, and R. von Sachs (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics* 55(4), 737–764.
- Galagedera, D. and E. Maharaj (2008). Wavelet timescales and conditional relationship between higher-order systematic co-moments and portfolio returns. *Quantitative Finance* 8(2), 201–215.
- Gallegati, M. (2008). Wavelet analysis of stock returns and aggregate economic activity. *Computational Statistics & Data Analysis* 52(6), 3061–3074.
- Gallegati, M. and M. Gallegati (2007). Wavelet Variance Analysis of Output in G-7 Countries. *Studies in Nonlinear Dynamics & Econometrics* 11(3).
- Gençay, R., N. Gradojevic, F. Selçuk, and B. Whitcher (2010). Asymmetry of information flow between volatilities across time scales. *Quantitative Finance*.
- Gençay, R., F. Selçuk, and B. Whitcher (2001a). Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A: Statistical Mechanics and its Applications* 289(3-4), 543–556.
- Gençay, R., F. Selçuk, and B. Whitcher (2001b). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.
- Gençay, R., F. Selçuk, and B. Whitcher (2001c). Scaling properties of foreign exchange volatility. *Physica A: Statistical Mechanics and its Applications* 289(1-2), 249–266.
- Gençay, R., F. Selçuk, and B. Whitcher (2003). Systematic risk and timescales. *Quantitative Finance* 3(2), 108–116.

- Gençay, R., F. Selçuk, and B. Whitcher (2005). Multiscale systematic risk. *Journal of International Money and Finance* 24(1), 55–70.
- Gourieroux, C. and J. Jasiak (2001). *Financial Econometrics: Problems, Models, and Methods*. Princeton University Press.
- Grané, A. and H. Veiga (2010). Wavelet-based detection of outliers in financial time series. *Computational Statistics & Data Analysis* 54(11), 2580–2593.
- Griffin, J. and R. Oomen (2010). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*.
- Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* 24(2), 127–161.
- Hasbrouck, J. (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, USA.
- Hausman, J., A. Lo, and A. MacKinlay (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* 31(3), 319–379.
- Hautsch, N. (2004). *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models* (1 ed.). Springer.
- Haven, E., X. Liu, C. Ma, and L. Shen (2009). Revealing the implied risk-neutral MGF from options: The wavelet method. *Journal of Economic Dynamics and Control* 33(3), 692–709.
- Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11(2), 359.
- Høg, E. and A. Lunde (2003). Wavelet estimation of integrated volatility. *Unpublished paper: Aarhus School of Business*.

- Hong, Y. and C. Kao (2004). Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica* 72(5), 1519–1563.
- Hong, Y. and J. Lee (2000). Wavelet-based estimation for heteroskedasticity and autocorrelation consistent variance-covariance matrices. In *Econometric Society World Congress 2000 Contributed Papers*. Econometric Society.
- Hong, Y. and J. Lee (2001). One-sided testing for ARCH effects using wavelets. *Econometric Theory* 17(06), 1051–1081.
- Huang, S. and T. Wu (2010). Integrating recurrent SOM with wavelet-based kernel partial least square regressions for financial forecasting. *Expert Systems with Applications* 37, 5698–5705.
- In, F. and S. Kim (2006a). The hedge ratio and the empirical relationship between the stock and futures markets: A new approach using wavelet analysis\*. *The Journal of Business* 79(2), 799–820.
- In, F. and S. Kim (2006b). Multiscale hedge ratio between the Australian stock and futures markets: Evidence from wavelet analysis. *Journal of Multinational Financial Management* 16(4), 411–423.
- In, F., S. Kim, V. Marisetty, and R. Faff (2008a). Analysing the performance of managed funds using the wavelet multiscaling method. *Review of Quantitative Finance and Accounting* 31(1), 55–70.
- In, F., S. Kim, V. Marisetty, and R. Faff (2008b). Analysing the performance of managed funds using the wavelet multiscaling method. *Review of Quantitative Finance and Accounting* 31(1), 55–70.
- Jagric, T., B. Podobnik, and M. Kolanovic (2005). Does the Efficient Market Hypothesis Hold?: Evidence from Six Transition Economies. *Eastern European Economics* 43(4), 79–103.

- Johnstone, I. M. (2011). *Gaussian estimation: Sequence and multiresolution models*.
- Johnstone, I. M. and B. W. Silverman (2004). Needles and straw in haystacks: Empirical bayes estimates of possibly sparse sequences. *The Annals of Statistics* 32(4), 1594–1649.
- Johnstone, I. M. and B. W. Silverman (2005). Ebayesthresh: R programs for empirical bayes thresholding. *Journal of Statistical Software* 12(8), 1–38.
- Kallenberg, O. (1997, January). *Foundations of Modern Probability* (1st ed.). Springer.
- Karuppiah, J. and C. Los (2005). Wavelet multiresolution analysis of high-frequency Asian FX rates, summer 1997. *International Review of Financial Analysis* 14(2), 211–246.
- Kim, S. and F. In (2003). The relationship between financial variables and real economic activity: Evidence from spectral and wavelet analyses. *Studies in Nonlinear Dynamics and Econometrics* 7(4), 22–45.
- Kim, S. and F. In (2005a). Multihorizon Sharpe Ratios. *The Journal of Portfolio Management* 31(2), 105–111.
- Kim, S. and F. In (2005b). The relationship between stock returns and inflation: new evidence from wavelet analysis. *Journal of Empirical Finance* 12(3), 435–444.
- Kim, S. and F. In (2007). On the relationship between changes in stock prices and bond yields in the G7 countries: Wavelet analysis. *Journal of International Financial Markets, Institutions and Money* 17(2), 167–179.
- Kyaw, N., C. Los, and S. Zong (2006). Persistence characteristics of Latin American financial markets. *Journal of Multinational Financial Management* 16(3), 269–290.

- Lee, H. S. (2001). Price and volatility spillovers in stock markets: A wavelet analysis. In *Manuscript presented at the 2001 Australasian Meeting of the Econometric Society*.
- Lee, H. S. (2004). International transmission of stock market movements: a wavelet analysis. *Applied Economics Letters* 11(3), 197–201.
- Lee, J. and Y. Hong (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory* 17(02), 386–423.
- Lee, S. S. and P. A. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial studies* 21(6), 2535–2563.
- Li, J., Z. Shi, and X. Li (2006). Genetic programming with wavelet-based indicators for financial forecasting. *Transactions of the Institute of Measurement & Control* 28(3), 285–297.
- Li, S. and S. Kuo (2008). Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks. *Expert Systems with Applications* 34(2), 935–951.
- Li, T. and M. Hinich (2002). A filter bank approach for modeling and forecasting seasonal patterns. *Technometrics* 44(1), 1–14.
- Lien, D. and K. Shrestha (2007). An empirical analysis of the relationship between hedge ratio and hedging horizon using wavelet analysis. *Journal of Futures Markets* 27(2), 127–150.
- Lin, S. and M. Stevenson (2001). Wavelet analysis of the cost-of-carry model. *Studies in Nonlinear Dynamics and Econometrics* 5(1), 87–102.
- Lowther, G. (2012). Decomposition of semimartingales. <http://math.stackexchange.com/a/139686/28463>.

- Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets* 3(3), 205–258.
- Malliavin, P. and M. E. Mancino (2002). Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics* 6(1), 49–61.
- McAleer, M. and M. C. Medeiros (2008). Realized volatility: A review. *Econometric Reviews* 27(1), 10–45.
- Merton, R. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4), 323–361.
- Mondal, D. and D. B. Percival (2008). Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics*, 1–24.
- Müller, U. A., M. M. Dacorogna, R. D. Dave, R. B. Olsen, O. V. Pictet, and J. E. von Weizsäcker (1997). Volatilities of different time resolutions—Analyzing the dynamics of market components. *Journal of Empirical Finance* 4(2-3), 213–239.
- Murtagh, F., J. L. Starck, and O. Renaud (2004). On Neuro-Wavelet modeling. *Decision Support Systems* 37(4), 475–484.
- Musiela, M. and M. Rutkowski (2008). *Martingale Methods in Financial Modelling* (2nd ed.). Springer.
- Nicolato, E. and E. Venardos (2003). Option pricing in stochastic volatility models of the ornstein-uhlenbeck type. *Mathematical finance* 13(4), 445–466.
- Nielsen, M. Ø. and P. Frederiksen (2008). Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. *Journal of Empirical Finance* 15(2), 265–286.

- Nikkinen, J., S. Pynnönen, M. Ranta, and S. Vähämaa (2010). Cross-dynamics of exchange rate expectations: a wavelet analysis. *International Journal of Finance & Economics*.
- Norsworthy, J., D. Li, and R. Gorener (2000). Wavelet-based analysis of time series: an export from engineering to finance. In *Proceedings of the 2000 IEEE Engineering Management Society. EMS - 2000 (Cat. No.00CH37139)*, pp. 126–132.
- O’Hara, M. (1998). *Market Microstructure Theory*. Wiley.
- Øksendal, B. K. (2002). *Stochastic Differential Equations: An Introduction with Applications* (5th ed.). Springer.
- Ozun, A. and A. Cifter (2008). Modelling long-term memory effect in stock prices. *Studies in Economics and Finance* 25(1), 38–48.
- Pacurar, M. (2008). Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys* 22(4), 711–751.
- Pan, Z. and X. Wang (1998). A Stochastic Nonlinear Regression Estimator Using Wavelets. *Computational Economics* 11(1), 89–103.
- Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.
- Power, G. and C. Turvey (2010). Long-range dependence in the volatility of commodity futures prices: Wavelet-based evidence. *Physica A: Statistical Mechanics and its Applications* 389(1), 79–90.
- Protter, P. E. (2005). *Stochastic Integration and Differential Equations* (2nd ed.). Springer.

- Raihan, S., Y. Wen, and B. Zeng (2005). Wavelet: A New Tool for Business Cycle Analysis. *Working*.
- Ramsey, J. (1999a). Regression over timescale decompositions: a sampling analysis of distributional properties. *Economic Systems Research* 11(2), 163–184.
- Ramsey, J. (1999b). The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 357(1760), 2593–2606.
- Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics & Econometrics* 6(3), 1–27.
- Ramsey, J. B. and C. Lampart (1998a). The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies in Nonlinear Dynamics and Econometrics* 3(1), 23–42.
- Ramsey, J. B. and C. Lampart (1998b). Decomposition of economic relationships by timescale using wavelets: Money and income. *Macroeconomic Dynamics* 2(01), 49–71.
- Ramsey, J. B. and Z. Zhang (1996). The application of waveform dictionaries to stock market index data. In Y. A. Kravtsov and J. Kadtke (Eds.), *Predictability of Complex Dynamical Systems*, pp. 189–208. Springer-Verlag.
- Ramsey, J. B. and Z. Zhang (1997). The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance* 4(4), 341–372.
- Renaud, O., J. L. Starck, and F. Murtagh (2003). Prediction based on a multiscale decomposition. *International Journal of Wavelets Multiresolution and Information Processing* 1, 217–232.

- Renaud, O., J. L. Starck, and F. Murtagh (2005). Wavelet-Based combined signal filtering and prediction. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 35(6), 1241–1251.
- Robert, C. and M. Rosenbaum (2011). A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics* 9(2), 344–366.
- Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance* 39(4), 1127–1139.
- Rosenthal, J. S. (2006). *A First Look at Rigorous Probability Theory* (2 ed.). World Scientific Publishing Company.
- Roşu, I. (2009). A Dynamic Model of the Limit Order Book. *Review of Financial Studies* 22(11), 4601–4641.
- Rua, A. and L. Nunes (2009). International comovement of stock market returns: A wavelet analysis. *Journal of Empirical Finance* 16(4), 632–639.
- Russell, J. and F. Engle (2009). Analysis of High-Frequency Data. In Y. Aït-Sahalia and L. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and techniques*, pp. 383–426. North-Holland.
- Rydberg, T. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* 1(1), 2–25.
- Samia, M., M. Dalenda, and A. Saoussen (2009). Accuracy and Conservatism of VaR Models: A Wavelet Decomposed VaR Approach Versus Standard ARMA-GARCH Method. *International Journal of Economics and Finance* 1(2), 174–184.
- Schweizer, M. (1992). Martingale densities for general asset prices. *Journal of Mathematical Economics* 21(4), 363–378.

- Shrestha, K. and K. Tan (2005). Real Interest Rate Parity: Long-Run and Short-Run Analysis Using Wavelets. *Review of Quantitative Finance and Accounting* 25(2), 139–157.
- Stoll, H. (2000). Friction. *Journal of Finance* 55(4), 1479–1514.
- Struzik, Z. (2001). Wavelet methods in (financial) time-series processing. *Physica A: Statistical Mechanics and its Applications* 296(1-2), 307–319.
- Subbotin, A. (2008). A multi-horizon scale for volatility. *Université Paris1 Panthéon-Sorbonne (Post-Print and Working Papers)*.
- Tang, L., H. Sheng, and L. Tang (2009). GARCH prediction using spline wavelet support vector machine. *Neural Computing & Applications* 18(8), 913–917.
- Tang, L., L. Tang, and H. Sheng (2009). Forecasting volatility based on wavelet support vector machine. *Expert Systems with Applications* 36(2), 2901–2909.
- Teyssiere, G. and P. Abry (2007). Wavelet analysis of nonlinear long-range dependent processes. Applications to financial time series. In *Long Memory in Economics*, pp. 173–238. Springer.
- Tkacz, G. (2001). Estimating the Fractional Order of Integration of Interest Rates Using a Wavelet OLS Estimator. *Studies in Nonlinear Dynamics and Econometrics* 5(1), 19–32.
- Tsay, R. S. (2005). *Analysis of Financial Time Series* (2nd ed.). Wiley-Interscience.
- Voev, V. and A. Lunde (2007). Integrated covariance estimation using High-Frequency data in the presence of noise. *Journal of Financial Econometrics* 5(1), 68–104.
- Vuorenmaa, T. (2004). A Multiresolution Analysis of Stock Market Volatility Using Wavelet Methodology. *Licentiate Thesis, University of Helsinki*.

- Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* 82(2), 385–397.
- Wong, H., W. Ip, and Y. Li (2001). Detection of jumps by wavelets in a heteroscedastic autoregressive model. *Statistics & Probability Letters* 52(4), 365–372.
- Wong, H., W. Ip, Z. Xie, and X. Lui (2003). Modelling and forecasting by wavelets, and the application to exchange rates. *Journal of Applied Statistics* 30(5), 537–553.
- Wood, R. (2000). Market Microstructure Research Databases: History and projections. *Journal of Business & Economic Statistics* 18(2), 140–145.
- Xue, Y. and R. Gençay (2010). Jump detection with wavelets for High-Frequency financial time series. *Working Paper*.
- Yamada, H. (2005). Wavelet-based beta estimation and Japanese industrial stock prices. *Applied Economics Letters* 12(2), 85–88.
- Yogo, M. (2008). Measuring business cycles: A wavelet analysis of economic time series. *Economics Letters* 100(2), 208–212.
- Yousefi, S., I. Weinreich, and D. Reinartz (2005). Wavelet-based prediction of oil prices. *Chaos, Solitons & Fractals* 25(2), 265–275.
- Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12(6), 1019–1043.
- Zhang, L. (2010). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics* 160(1), 33–47.
- Zhang, L., P. A. Mykland, and Y. Ait-Sahalia (2005). A tale of two time scales. *Journal of the American Statistical Association* 100(472), 1394–1411.

- Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 45–52.
- Zhou, J., T. Bai, and J. Tian (2008). The Integrated Methodology of KPCA and Wavelet Support Vector Machine for Predicting Financial Distress. *Advanced Data Mining and Applications 5139*, 508–515.
- Zivot, E. and J. Wang (2006). *Modeling Financial Time Series with S-PLUS* (2nd ed.). Springer.

## VITA

Brian Donhauser is a Ph.D. student at the University of Washington. He welcomes your comments to [brianjd@u.washington.edu](mailto:brianjd@u.washington.edu).