

Essays in Financial Econometrics

Daousis Emmanouil

Supervisor: Prof. T. Simos

The Department of Economics
University of Ioannina

This dissertation is submitted for the degree of
Doctor of Philosophy

May 2026

Above all, I am profoundly thankful to my parents for their unconditional love, constant encouragement, and steadfast support throughout my academic journey. They have stood beside me from the very beginning, believing in me even during moments of doubt. Their sacrifices, patience, and faith in my abilities have been the foundation upon which all my achievements rest. This thesis is as much theirs as it is mine.

Declaration

For the completion of my thesis i have received funding from the ELKE of the University of Ioannina, which i am thankful as their support was necessary for the successful finish of my research.

Daousis Emmanouil
May 2026

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Mr.Theodoros Simos, whose guidance, insight, and unwavering support were paramount throughout the course of this research and the writing of this thesis. His expertise, patience, and constructive feedback continuously challenged me to refine my ideas and strengthen my work. I am especially grateful for his encouragement during both the most demanding and the most rewarding stages of this journey. His mentorship has not only shaped this thesis but has also significantly contributed to my academic and professional development.

I would also like to extend my sincere thanks to Mr.Stylianos Arvanitis and Mr. Dimitrios Kainourgios for their valuable comments and thoughtful suggestions during the various writing phases of my PhD. Their critical insights and engaging discussions helped me better understand complex and demanding concepts, ultimately improving the clarity and rigor of this work. I am truly appreciative of the time and effort they dedicated to reviewing my research.

Abstract

From the early 1990's we observe the transition from the open floor trading of the stock exchange to a computer base electronic trading system. Steadily from early 2000's financial markets have moved towards a high frequency and later to algorithmic trading. Those advancements have fueled the creation and availability of high and ultra-high frequency financial data. Such data have shown an increase complexity of the markets dynamics that lower frequencies did not, such as microstructure noise, jumps and staleness. Although the literature has explore and identifies those stylized features of the markets, such as the stock to the cryptocurrency markets, little attention has been directed to addressing those issues regarding volatility estimation. The largest part of the literature is focusing on the effects and mitigation of the microstructure noise component and the disentangling of the discontinuous part of the underlying process from the continuous one. At those high frequencies is not uncommon to observe the price of a financial instrument staying stale for a short period of time before an agreement is sealed and the price updates. This results in a series of zeros in the return series of the underlying asset, which can heavily influence volatility estimation and jump inference. To that end the literature has explored ways to mitigate the effects of noise and create unbiased estimators of the volatility, recently this has also been done when it comes to staleness where robust estimators have been developed to avoid the either downward or upward bias classic estimators like the realized variance suffer from.

Empirical evidence have show that in tick by tick frequency microstructure noise is a prevalent feature of the underlying process along with staleness. Some evidence shows that up to 40% of the returns are zeros while around a similar percentage are the noise to signal ratio. This indicates that at high frequencies both stylized features are present in the data, volatility estimation and thus jump identification are robust to only one of those features. As far as we can see an estimator able to remain unbiased and consistent under the simultaneous presence of microstructure noise and staleness is yet to be developed. This observation also extends to jump inference where jump identification procedures fail to consistently identify discontinuities, leading to spurious jump counts. Our aim with this research was to explore what current estimators of the Integrated and quadratic variance have to offer, how the presence of noise and staleness affects the estimation of volatility

and jump inference and using those lessons to propose a correction for known estimators thus increasing their power under such conditions. Lastly by employing the corrected estimators to observe whether or not the jump inference could be improved.

To that end we have structured our study in three sections, Chapter 2 provides a comprehensive and systematic review of non-parametric jump identification methodologies developed in the literature. We conduct a meticulous comparison of the most widely used jump detection procedures, examining their statistical properties under a broad range of simulated environments. In particular, we evaluate their empirical size and power through an extensive Monte Carlo study that incorporates multiple stochastic price processes, alternative discretization schemes, and varying threshold parameter choices where applicable. Special attention is given to the impact of zero returns on jump detection performance, a feature often overlooked in previous studies. Based on these findings, we rank the competing methodologies and provide detailed guidance regarding their practical implementation, highlighting their relative strengths, weaknesses, and robustness across different data-generating scenarios.

Chapter 3 focuses on the estimation of integrated volatility in the simultaneous presence of market microstructure noise and price staleness, the latter inducing zero returns in the observed return series. After carefully outlining the econometric challenges posed by these dual sources of contamination and discussing the limitations of existing estimators, we propose a correction to the Multipower variation estimator, designed to remain robust under such distortions. We formally derive its theoretical and finite-sample distributional properties and establish its consistency under general conditions. An extensive Monte Carlo simulation study is conducted to evaluate its performance relative to conventional estimators, demonstrating that the proposed correction delivers stable and reliable volatility estimates even under the specific assumptions.

Finally, Chapter 4 builds directly on the methodological developments of the previous chapter. By employing the corrected integrated variance estimator, we revisit two established and widely used jump detection tests and propose appropriate modifications to account for the joint presence of noise and price staleness. Through a detailed simulation analysis, we reassess jump inference and update the empirical ranking of jump identification procedures. Our results indicate that, once contamination is properly addressed, jumps appear less frequent and less economically significant in explaining total price variation than previously documented in the empirical literature. These findings contribute to a more accurate understanding of volatility decomposition and the true role of discontinuities in high-frequency financial markets.

Table of contents

List of figures	10
List of tables	11
1 Literature Review	12
1.1 The Volatility Estimation Paradigm Shift	12
1.2 Theoretical Foundations: From Continuous-Time Semimartingales to Re- alized Measures	14
1.2.1 Quadratic Variation: The Theoretical Target	15
1.2.2 The Realized Variance Estimator: Theory and Convergence . . .	15
1.3 Early Explorations with High-Frequency Data	16
1.4 The Jump Component: Separating Continuous from Discontinuous Variation	16
1.4.1 Bipower and Multipower Variation: Theoretical Development . .	17
1.4.2 Threshold and Nearest-Neighbor Approaches	17
1.5 Jump Detection in High-Frequency Financial Data	17
1.6 Market Microstructure Noise	22
1.6.1 Market Microstructure Noise and Noise-Robust Estimators	23
1.6.2 Joint Presence of Jumps and Noise	25
1.7 Staleness	26
1.7.1 Consequences of Ignoring Zero Returns	28
1.8 Main Findings	30
2 Review of Jump tests	33
2.1 Introduction	33
2.2 Introspection of the jump tests	35
2.2.1 Tests based on squared variation	37
2.2.2 Tests based on P-power variation	40
2.2.3 Tests based on standardized returns	43
2.2.4 Test based on variance swap	44
2.2.5 Tests that account for microstructure noise	45

2.3	Monte Carlo Simulations	46
2.3.1	Appraisal of test performance	48
2.4	Empirical Evidence	59
2.4.1	Data Description	60
2.4.2	Empirical Findings	62
2.5	Conclusions	65
3	Volatility Estimation When Noisy Prices Go Silent: A Kernel Multipower Approach	67
3.1	Introduction	67
3.2	Model Setup	69
3.3	Existing IV Estimators	70
3.3.1	Multipower variation estimator	71
3.3.2	Two scales realized variance estimator	73
3.3.3	Realized kernel estimator	73
3.3.4	Truncated realized variance estimator	74
3.3.5	Truncated pre-averaging estimator	74
3.3.6	Truncated flat-top realized kernel estimator	75
3.4	Novel kernel estimator	76
3.5	Simulation Study	79
3.5.1	Staleness	82
3.5.2	Finite-Sample Performance	84
3.6	Conclusions	85
4	Jumps in Asset Prices: Illusion of Realized Inference under Noise at stand still	87
4.1	Introduction	87
4.2	Model Setup	90
4.3	Impact of Noise Distribution	94
4.4	Simulation Setup	95
4.4.1	Microstructure Noise Distortions	97
4.5	The Kernel-Smoothed Staleness-Corrected Multipower Variation Estimator	98
4.5.1	Setup and Definition	99
4.5.2	Asymptotic Properties	100
4.5.3	Tuning Parameters	100
4.6	Jump Inference	103
4.6.1	Correcting Existing Jump Tests	103
4.6.2	Alternative Jump tests	105
4.6.3	ABD jump test	106

4.7	Empirical Application	108
4.7.1	Estimation under Noise	110
4.8	Conclusions	112
	References	113

List of figures

3.1	Plot of log MSE vs noise for three estimators	80
3.2	Log MSE of QRV, MBRK and TTRPK estimators under MA(1) noise . .	81
3.3	Bias of staleness-corrected MPV and proposed MPV estimator vs Gaussian noise	83
3.4	Bias vs sample size N	84
4.1	Bias of Power Variation for the power r	93
4.2	Bias of Bipower Variation for the first power r_1	93
4.3	Bias of Bipower Variation under Gaussian i.i.d noise	93
4.4	Bias of Bipower Variation as a function of noise standard deviation	93
4.5	Rejection rate of corrected BNS test across staleness probabilities	95
4.6	Comparison of kernel plots: (a) Kernel example, (b) Kernel comparison. .	103
4.7	Rejection Rate of the BNS test as a function of the staleness probability .	104
4.8	BNS test of Kolokolov et al. (2020) and Corrected BNS test rejection rate as a function of the staleness probability	105
4.9	Distribution of test statistics from simulated daily samples	107
4.10	Distribution of ABD and corrected ABD	108
4.11	Rejection rate BNS tests as a function of staleness	110
4.12	BNS and corrected BNS tests for EUR/USD, 2021–2025	112

List of tables

2.1	Empirical size of the test with no jumps at 5% significance level	49
2.2	Empirical size of the tests with rare volatility jumps ($\lambda = 0.2$)	50
2.3	Empirical size of the tests with frequent volatility jumps ($\lambda = 5$)	52
2.4	Empirical size for rare and small jumps in volatility at 5% significance . .	53
2.5	Empirical size for volatility jumps at 5% level	54
2.6	Empirical power for rare and small jumps in price at 5% significance . . .	54
2.7	Empirical power for large and frequent jumps in price at 5% significance level	55
2.8	Power of tests at 5% significance level	56
2.9	Power of tests at 5% significance under Heston model	57
2.10	Infrequent trading without jumps at 5% level	59
2.11	Summary of Simulation Variables	59
2.12	Descriptive Statistics	60
2.13	Number of zeros in the data-set	61
2.14	Empirical Results	62
3.1	MSE comparisons for different values of the noise variance (i.i.d. noise). .	80
3.2	MSE comparisons for different values of the noise variance (MA(1) noise)	81
4.1	Rejection rate of the jump tests at 5% significance level under three sam- pling frequencies	109

Chapter 1

Literature Review

1.1 The Volatility Estimation Paradigm Shift

The birth of the well-known term volatility took place in 1952 by a young graduate student from the University of Chicago named Harry Markowitz in one of his papers published under the name Portfolio Selection [Markowitz \(1952\)](#). In his thesis, he argued that fund performance should be compared to the amount of risk that inherits, since there was no standardized and well accepted quantification of risk in the literature, he uses variance as a proxy. The term is used to describe a statistical concept that measures dispersion by measuring how much each observation diverges from the mean, expressed mathematically as the expected value of the squared deviations.

The importance of volatility can be seen in action in every market around the world as investors and fund managers utilize it to assess the market environment. High volatility usually associates with negative market performance meaning that sometimes can provide a window of opportunity for market participants to exploit the undervalued financial tools. Furthermore, the existence of different speculative financial securities, such as exchange-traded notes, exchanged traded funds, and VIX options and futures embodies its importance. It is not by accident that investors tend to use the following quote “When VIX is high time to buy when VIX is low look out below”, this rather famous piece of advice makes clear that volatility is vital for all market participants as its interpretation at the right moment can separate the winners from losers.

No matter how important volatility is a stand-out statistical measure may be, the real interest lays on volatility and its relation to price movements. Any event no matter its origin, macroeconomic, financial, or political can potentially trigger a volatile period for a given market or markets. It is common to associate volatility with risk, however, while related, the two concepts are distinct. Volatility refers to the magnitude of fluctuations in an asset’s price and is typically modeled as the standard deviation of returns. Formally, if

r_t denotes the return at time t , volatility is given by:

$$\sigma = \sqrt{\mathbb{E}[(r_t - \mu)^2]} \quad (1.1)$$

where $\mu = \mathbb{E}[r_t]$ is the expected return. This definition highlights that volatility captures both upward and downward deviations from the mean, and therefore reflects overall uncertainty rather than downside risk alone.

A closely related statistical measure is variance, which quantifies the dispersion of returns around their mean. It is defined as:

$$\text{Var}(r_t) = \mathbb{E}[(r_t - \mu)^2] \quad (1.2)$$

Variance provides a complete second-moment characterization of the distribution of returns, while volatility is simply its square root. Contrary to informal interpretations, variance is not the difference between volatility values at different time periods, but rather the fundamental quantity from which volatility is derived.

In financial applications, volatility is often treated as a time-dependent process σ_t , reflecting the evolving uncertainty in asset prices. Models such as GARCH or stochastic volatility frameworks explicitly describe the dynamics of σ_t^2 (i.e., conditional variance), making variance the central object in the modeling of volatility over time.

These statistical measures form the basis of estimation models used in financial economics, particularly in the valuation of derivative securities such as options. Volatility, defined as the degree of variation in an asset's price over time, plays a central role as a key input in option pricing models such as the Black-Scholes model and stochastic volatility models (Black and Scholes, 2019; Heston, 1993), portfolio optimization in the Markowitz framework, risk management metrics like Value-at-Risk, and as an indicator of market uncertainty. The right pricing of a financial instrument is crucial not only because it provides information about the objective value of an instrument and whether is over or under-evaluated, but also it helps in eliminating actions that can endanger the stability of the market.

Traditional approaches to volatility modeling, epitomized by the GARCH family (Bollerslev, 1986; Engle, 1982) and stochastic volatility (SV) models (Taylor, 1999), were inherently parametric and exclusively relied on daily or lower-frequency returns. These models conceptualized volatility as a latent, unobservable process requiring statistical filtration from historical returns. While profoundly influential, they encountered some well documented limitations such as, dependency on specific distributional assumptions, challenges in capturing rapid volatility dynamics, and the inherent difficulty in making forecasts using an observed bias version of the unobserved true process.

1.2 Theoretical Foundations: From Continuous-Time Semimartingales to Realized Measures

During the early 2000s the increased availability of high frequency data has fueled the growing body of literature surrounding the subject of disentangling the true variation from the discontinuous part. This transition moved the analytical focus from modeling and filtering latent volatility to directly measuring it using intraday price observations. The theoretical bedrock for this shift is the concept of Quadratic Variation (QV). For a broad class of asset price processes (semimartingales), the sum of squared intraday returns converges to the total QV over a fixed time interval (e.g., one trading day) as the sampling frequency approaches infinity. Crucially, this QV decomposes into the integrated variance (IV) from the continuous component of the price path and the cumulative sum of squared jumps. Realized Variance (RV), defined simply as the sum of squared intraday returns, thus provides a model-free, ex-post observable measure of total daily price variation ([Andersen and Bollerslev, 1998](#); [Barndorff-Nielsen and Shephard, 2002](#)). This innovation transformed volatility from an abstract statistical concept into an observable financial variable, enabling direct analysis and fostering the literature on high frequency econometrics.

The review is organized as follows, [section 1.2](#) establishes the theoretical foundations. [Section 1.3](#) reviews the literature regarding high frequency financial data dynamics, such as Microstructure noise. The sections also discuss the implications the empirical literature has demonstrated for classical integrated variance estimators. In [section 1.4](#) and [section 1.5](#) we discuss the early literature of disentangling the continuous from the discontinuous part of the underlying process and early constructions of non-parametric jump tests. [Section 1.6](#) explores in depth the effects of Noise on volatility estimation and reviews methodologies that have been developed to mitigate its effects. [Section 1.7](#) derives the definition of staleness and the theories behind its existence. The section also delves into implications of it when it concerns with volatility estimation and jump inference. We conclude our introduction to this study with some concluding remarks at [Section](#)

1.2 Theoretical Foundations: From Continuous-Time Semimartingales to Realized Measures

The standard modeling framework in continuous-time finance posits that the logarithm of an asset price, $X_t = \log(P_t)$, follows a semimartingale process. This class is sufficiently general to encompass both Brownian-motion-driven diffusions and discontinuous jumps, making it the workhorse for realistic asset pricing models. A general Itô semimartingale can be represented as:

$$dX_t = \mu_t dt + \sigma_t dW_t + J_t dN_t$$

where μ_t is a predictable, locally bounded drift process, σ_t is the càdlàg spot volatility process, W_t is a standard Brownian motion, J_t is the random jump size, and N_t is a counting

1.2 Theoretical Foundations: From Continuous-Time Semimartingales to Realized Measures

process governing jump arrivals. The semimartingale property is economically crucial as it is linked to the absence of arbitrage in complete markets under the "No Free Lunch with Vanishing Risk" condition [Delbaen \(2006\)](#).

1.2.1 Quadratic Variation: The Theoretical Target

For a semimartingale X_t , the quadratic variation (QV) process, $[X]_t$, is defined as the probability limit of sums of squared increments over increasingly fine partitions of the time interval $[0, t]$. It serves as the model-free, unconditional measure of total return variation. A key decomposition result states that for the process above, the QV over a fixed interval $[t-1, t]$ (e.g., one trading day) is:

$$[X]_t = \underbrace{\int_{t-1}^t \sigma_s^2 ds}_{\text{Integrated Variance (IV)}} + \underbrace{\sum_{t-1 \leq s \leq t} J_s^2}_{\text{Sum of Squared Jumps}} \quad (1.3)$$

This decomposition lies at the heart of high-frequency volatility measurement as the total realized variation equals continuous integrated variance plus the contribution from discontinuous price jumps.

1.2.2 The Realized Variance Estimator: Theory and Convergence

The natural estimator for QV, given discrete observations, is the **Realized Variance (RV)**. Consider partitioning the interval $[t-1, t]$ into n subintervals of length $\Delta = 1/n$. Let the i -th intraday return be $r_{t,i} = X_{t-1+i\Delta} - X_{t-1+(i-1)\Delta}$, then RV could be defined as:

$$RV_t^{(n)} = \sum_{i=1}^n r_{t,i}^2. \quad (1.4)$$

Under mild regularity conditions, as $n \rightarrow \infty$ ($\Delta \rightarrow 0$), the estimator converges in probability to the QV:

$$RV_t^{(n)} \xrightarrow{P} [X]_t = IV_t + \sum_{s \in [t-1, t]} J_s^2.$$

[Barndorff-Nielsen and Shephard \(2002\)](#) provided the asymptotic theory, establishing a central limit theorem (CLT) for RV in a pure continuous stochastic volatility setting (no jumps, σ_t independent of W_t):

$$\frac{RV_t^{(n)} - IV_t}{\sqrt{\frac{2}{3} \int_{t-1}^t \sigma_s^4 ds}} \xrightarrow{d} N(0, 1),$$

where the integrated quarticity $\int \sigma_s^4 ds$ can be consistently estimated by realized quarticity measures. This CLT provided the statistical foundation for using RV as a highly efficient, asymptotically normal estimator of IV in the absence of jumps.

1.3 Early Explorations with High-Frequency Data

Initial investigations using high-frequency data primarily focused on characterizing intraday periodic patterns and estimating parameters of specific continuous-time models (Bergstrom, 1984; Nowman, 1997). Zhou (1996) represents an important precursor, utilizing high-frequency foreign exchange data to study volatility dynamics and empirically highlighting challenges like autocorrelation in returns that was used as an early indication of microstructure effects. This work set the stage for a more systematic theoretical treatment. The influential works of Andersen and Bollerslev (1998) and, especially, Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003) constituted the definitive breakthrough. Andersen et al. (2003) comprehensively articulated the case for RV. They demonstrated empirically that RV, constructed from moderately sampled intraday data (e.g., 5-minute returns), provided an unbiased estimate of daily volatility compared to the traditional, extremely noisy proxy of daily squared returns. Theoretically, they formalized the convergence of RV to the underlying integrated variance under a general continuous stochastic volatility framework. This transformed volatility from a latent statistical variable into an observable economic variable, thereby providing the bases for the field of high-frequency financial econometrics.

1.4 The Jump Component: Separating Continuous from Discontinuous Variation

Jumps capture sudden, large price movements typically associated with the arrival of significant news, liquidity shocks, or macroeconomic announcements (Merton, 1976). From an economic perspective, distinguishing between the smooth, "normal" variation induced by continuous information flow and "abnormal," discontinuous jumps is paramount. Models incorporating jumps, such as affine jump-diffusions (Duffie et al., 2000), provide a better fit to key empirical features like the volatility smirk in option prices (Christoffersen et al., 2009) and the heavy-tailed distribution of returns (Chernov et al., 2003).

1.4.1 Bipower and Multipower Variation: Theoretical Development

[Barndorff-Nielsen and Shephard \(2004, 2006\)](#) introduced **Bipower Variation (BPV)** as a jump-robust estimator of the integrated variance. Their key insight was that for a continuous semimartingale, while RV converges to QV, the sum of products of adjacent absolute returns, appropriately scaled, converges solely to IV:

$$BPV_t^{(n)} = \mu_1^{-2} \sum_{i=2}^n |r_{t,i}| |r_{t,i-1}| \xrightarrow{P} IV_t, \quad \text{where } \mu_1 = E(|Z|) = \sqrt{2/\pi}, \quad Z \sim N(0, 1). \quad (1.5)$$

Consequently, the difference $RV_t - BPV_t$ provides a consistent estimator for the sum of squared jumps on day t . **Tripower** and higher-order **Multipower Variations** were subsequently developed to improve robustness to multiple consecutive jumps and to estimate integrated quarticity for studentization purposes. [Barndorff-Nielsen et al. \(2006b\)](#) provided the asymptotic theory for these statistics in the presence of jumps.

1.4.2 Threshold and Nearest-Neighbor Approaches

Multipower variation estimators can be sensitive to the "smearing" effect of several small, consecutive jumps. **Threshold-based methods** offered a more localized alternative. [Mancini \(2008, 2009\)](#) pioneered a thresholding principle: if a return's magnitude is below a vanishing threshold (typically a function of local volatility), it is likely from the continuous component, if above, it is classified as a jump contribution. The continuous integrated variance is then estimated by summing the squares of returns truncated at this threshold. [Corsi et al. \(2010\)](#) developed **Threshold Bipower Variation (TBPV)**, which modifies the BPV estimator by replacing raw returns with their threshold-truncated versions, significantly improving robustness to consecutive jumps.

1.5 Jump Detection in High-Frequency Financial Data

The question of whether asset prices evolve continuously or exhibit discontinuities is fundamental to modern financial econometrics. The presence of jumps has profound implications for asset pricing, risk management, volatility forecasting, and derivative valuation. Models that ignore jumps may severely underestimate tail risk, misprice options, and yield misleading inference regarding volatility dynamics. At the same time, detecting jumps is intrinsically difficult because financial prices are observed at discrete times, whereas the underlying price process evolves in continuous time. As a result, even a purely continuous diffusion process will almost surely appear as a sequence of abrupt movements when sampled at sufficiently high frequency.

This tension between continuous-time theory and discrete-time observation motivates a vast literature devoted to jump detection. The central challenge is to distinguish true discontinuities in the sample path from large realizations of the continuous Brownian component. Early contributions, such as [Aït-Sahalia \(2002\)](#), emphasized that certain forms of apparent discontinuity are compatible with a continuous sample path, while others are not. By exploiting the probabilistic structure of transition densities and coincidence probabilities, this study laid the foundations for a rigorous statistical treatment of jumps.

Throughout the literature, asset prices are typically modeled as Itô semimartingales, a highly flexible class of stochastic processes that encompasses continuous diffusions, jump-diffusion models, and more general Lévy-driven dynamics. Let X_t denote the logarithmic price process. A integral representation is given by

$$X_t = X_0 + \int_0^t \alpha_u du + \int_0^t \sigma_u dW_u + Z_t, \quad (1.6)$$

where α_u is a locally bounded drift process, σ_u is a càdlàg spot volatility process, W_u is a standard Brownian motion, and Z_t captures the jump component. In much of the empirical literature, jumps are assumed to follow a compound Poisson process,

$$Z_t = \sum_{j=1}^{N_t} k_j, \quad (1.7)$$

where N_t is a Poisson counting process with intensity λ_t , and $k_j \neq 0$ denotes the jump sizes.

This specification implies that the quadratic variation of the process over a fixed interval decomposes naturally into a continuous and a discontinuous component:

$$QV_t = \int_{t-1}^t \sigma_s^2 ds + \sum_{j=1}^{N_t} k_j^2. \quad (1.8)$$

The first term corresponds to the integrated variance (IV) generated by the continuous diffusion component, while the second term captures the cumulative contribution of jumps. Because these two sources of variation have distinct statistical and economic interpretations, disentangling them is of primary importance.

The availability of high-frequency data has revolutionized empirical volatility measurement. Following the seminal work of [Andersen and Bollerslev \(1998\)](#), the quadratic variation of a semimartingale can be consistently estimated using realized variance.

However, realized variance alone cannot distinguish between continuous and jump-induced variation. This limitation led to the development of alternative realized measures that are robust to jumps. A key insight, formalized by [Barndorff-Nielsen and Shephard](#)

(2004), is that certain functionals of high-frequency returns converge to integrated variance even in the presence of a finite number of jumps. This observation underpins a large class of jump detection procedures.

Most jump detection tests are built around a simple but powerful idea: compare a jump-sensitive estimator of quadratic variation with a jump-robust estimator of integrated variance. If the price path is continuous over the interval of interest, both estimators converge to the same limit. If jumps are present, realized variance captures both continuous and discontinuous variation, while the jump-robust estimator does not, leading to a statistically significant discrepancy.

It is important to emphasize that the majority of jump tests are not designed to estimate the exact number, size, or timing of jumps. Instead, they test the null hypothesis of path continuity over a given interval against the alternative of at least one discontinuity. Rejection of the null therefore indicates that the process is not continuous on that interval, but does not necessarily imply frequent or large jump activity.

The literature on jump detection can be broadly classified into several categories based on the designed of the test statistic:

1. Tests based on squared variation and multipower variation
2. Tests based on higher-order p -power variation
3. Tests based on standardized returns and extreme value theory
4. Tests derived from variance swap replication arguments
5. Tests explicitly accounting for microstructure noise

Each class of tests embodies different assumptions about the underlying price process and exhibits distinct strengths and weaknesses.

The earliest and most influential jump tests rely on squared variation measures. The bipower variation estimator introduced by [Barndorff-Nielsen and Shephard \(2004\)](#) provides a consistent estimate of integrated variance in the presence of finite activity jumps. Building on this result, [Barndorff-Nielsen and Shephard \(2006\)](#) propose a test that exploits the difference between realized variance and realized bipower variation. Under the null hypothesis of no jumps, the difference converges to zero after appropriate standardization, while under the alternative it diverges.

The economic intuition behind this approach is straightforward. When jumps are rare, the probability of observing two consecutive returns contaminated by jumps vanishes as the sampling frequency increases. Consequently, products of adjacent absolute returns are driven primarily by the diffusion component. This insight has led to a wide family of multipower variation estimators and associated test statistics.

Despite their popularity, squared-variation-based tests face several challenges. Finite sample bias can be substantial, particularly when jumps cluster or when volatility varies rapidly. Moreover, multipower variation estimators may perform poorly in the presence of market microstructure noise or zero returns, motivating subsequent refinements.

To address the finite-sample bias of multipower variation estimators, [Corsi et al. \(2010\)](#) introduce threshold-based methods that combine multipower variation with truncation rules inspired by [Mancini \(2009\)](#). The key idea is to truncate large returns that are likely to be associated with jumps when estimating integrated variance.

These threshold-based estimators yield improved finite-sample performance and form the basis for the CPR test, which has been shown to outperform traditional bipower-based tests in environments with clustered jumps or high volatility. The broader lesson from this line of research is that jump detection accuracy depends critically on the quality of the integrated variance estimator.

An alternative approach exploits the asymptotic behavior of realized power variations of order p . As shown by [Aït-Sahalia and Jacod \(2009\)](#) and [Podolskij and Ziggel \(2010\)](#), the limiting behavior of power variation depends sharply on whether the underlying process is continuous and on the value of p . For $p > 2$, large increments dominate, making the statistic sensitive to jumps; for $p < 2$, small increments dominate, emphasizing the diffusion component.

By comparing realized power variations computed at different sampling frequencies, these tests are able to distinguish between continuous paths, finite-activity jumps, and infinite-activity jump processes. A major advantage of this framework is its robustness to both finite and infinite jump activity, characterized by the Blumenthal–Gettoor index.

Another prominent class of jump tests relies on standardized returns. [Andersen et al. \(2007a\)](#) and [Lee and Mykland \(2008\)](#) propose tests based on the idea that, in the absence of jumps, returns scaled by a consistent estimator of local volatility should follow a standard normal distribution. Extreme realizations of standardized returns are therefore indicative of jumps.

These tests employ tools from extreme value theory to control for multiple testing across intraday observations. They are particularly useful for identifying the timing of jumps, rather than merely their existence. However, their performance depends critically on the choice of window length used to estimate local volatility, and they may struggle to distinguish jumps from large but continuous volatility shocks.

A conceptually distinct approach is proposed by [Jiang and Oomen \(2008\)](#), who derive jump tests from variance swap replication arguments. In the absence of jumps, the difference between arithmetic and logarithmic returns aggregates to half of the integrated variance. Deviations from this equality therefore signal the presence of jumps.

1.5 Jump Detection in High-Frequency Financial Data

This approach has several appealing features. It is naturally linked to financial theory, remains robust to certain forms of microstructure noise, and does not rely directly on multipower variation estimators. However, its implementation requires careful estimation of higher-order variation measures and relies on high-frequency data.

As sampling frequencies increase, market microstructure noise becomes a dominant concern. Bid–ask bounce, price discreteness, and asynchronous trading can severely distort realized measures. Tests that ignore noise may misclassify noise-induced fluctuations as jumps.

To address this issue, [Lee and Mykland \(2012\)](#) and [Aït-Sahalia et al. \(2012\)](#) develop noise-robust jump tests using pre-averaging and related techniques. These methods explicitly model the observed price as the sum of an efficient price and a noise component, and adjust the test statistics accordingly. Simulation and empirical evidence suggest that accounting for noise substantially improves size control without sacrificing power.

An important distinction in the literature concerns the nature of jump activity. Early models focus on finite activity jumps driven by Poisson processes. More recent work recognizes that asset prices may exhibit infinite activity jumps, characterized by a large number of small discontinuities. Lévy-driven models capture this behavior and motivate alternative testing strategies based on truncated power variations and distributional diagnostics.

Tests designed for finite activity jumps may lose power or become inconsistent in the presence of infinite activity. Consequently, the choice of jump test must be aligned with the assumed jump structure of the data.

Empirical studies consistently find that jumps account for a relatively small fraction of total quadratic variation, often on the order of one to a few percent. Nevertheless, jumps play a disproportionate role in driving extreme returns and volatility spikes, and are closely linked to the arrival of macroeconomic and firm-specific news.

From a modeling perspective, separating continuous and jump components improves volatility forecasting and option pricing performance. From a risk management perspective, accurate jump detection is essential for stress testing and tail risk assessment.

The jump detection literature has evolved from simple squared-variation-based tests to a rich set of methodologies capable of handling finite and infinite activity jumps, stochastic volatility, and market microstructure noise. While no single test dominates in all settings, the collective body of work provides a powerful toolkit for diagnosing discontinuities in asset prices.

The choice of jump test should be guided by the sampling frequency, the expected nature of jump activity, and the presence of noise. In practice, empirical researchers often apply multiple tests in parallel to ensure robustness. Ultimately, jump detection remains a cornerstone of modern financial econometrics, bridging theoretical modeling and empirical reality.

1.6 Market Microstructure Noise

The literature on realized variance has moved from simpler estimators to more advanced methods that can better handle the complex nature of high-frequency financial data. While early estimators focused on consistency under idealized assumptions, subsequent research has systematically addressed jumps, market microstructure noise, and their joint presence. Truncation, pre-averaging, kernel smoothing, and wavelet techniques represent complementary strategies in this endeavor. The ongoing challenge lies in balancing theoretical robustness, computational feasibility, and empirical performance, a theme that continues to motivate active research in volatility estimation.

Volatility plays a central role in modern financial economics, underpinning asset pricing, derivative valuation, risk management, and portfolio allocation. Accurate measurement of volatility is therefore of both theoretical and practical importance. Traditional approaches to volatility estimation relied on low-frequency data, such as daily or weekly returns, and parametric models including ARCH and GARCH-type specifications. While useful, these approaches are limited by their dependence on restrictive model assumptions and their inability to fully exploit the information contained in high-frequency financial data. The availability of intraday transaction data over the last few decades has fundamentally transformed the volatility estimation literature, giving rise to realized variance (RV) and its numerous extensions.

The seminal contribution of [Andersen and Bollerslev \(1998\)](#) introduced the realized variance estimator, defined as the sum of squared intraday returns over a fixed interval, typically one trading day. Subsequent work by [Andersen et al. \(2003\)](#) and [Barndorff-Nielsen and Shephard \(2002\)](#) established the theoretical foundations of RV within a continuous-time semimartingale framework. Under idealized conditions (namely, frictionless markets, continuous sample paths, and noiseless price observations) realized variance converges in probability to the quadratic variation of the price process as the sampling frequency increases.

In continuous diffusion models, quadratic variation coincides with integrated volatility (IV), defined as the integral of the spot variance process over time. This result provides strong justification for the use of realized variance as a nonparametric estimator of volatility. Moreover, realized variance has been shown to outperform traditional model-based estimators in forecasting future volatility and risk measures, further contributing to its widespread adoption in empirical finance.

Despite the elegance of the RV framework, empirical evidence overwhelmingly suggests that asset prices exhibit discontinuities or jumps. Such jumps may arise from macroeconomic announcements, earnings releases, or sudden changes in market sentiment.

[Barndorff-Nielsen and Shephard \(2004\)](#) and [Hansen and Lunde \(2006\)](#), among others, documented that ignoring jumps leads to biased volatility estimates and distorted inference.

Allowing for jumps complicates volatility measurement because the quadratic variation of a jump-diffusion process consists of two distinct components: the integrated volatility of the continuous part and the jump variation associated with discontinuities. While standard RV consistently estimates total quadratic variation, it does not disentangle these components. This distinction is economically meaningful. [Andersen et al. \(2007a\)](#) and [Corsi et al. \(2010\)](#) demonstrated that separating continuous and jump components improves volatility forecasting performance. Similarly, [Zhang et al. \(2009\)](#) showed that both integrated volatility and jump variation have distinct and significant effects on credit default swap spreads, highlighting the importance of jump-sensitive volatility measures.

An alternative strategy for dealing with jumps relies on truncation. [Mancini \(2008, 2009\)](#) proposed excluding intraday returns whose absolute values exceed a threshold that depends on the sampling frequency. The intuition is that large returns are more likely to be generated by jumps rather than the continuous component. The resulting truncated realized variance estimator consistently estimates integrated volatility in the presence of both finite and infinite activity jumps.

[Andersen et al. \(2012\)](#) introduced a related approach based on nearest-neighbor truncation, where squared returns are truncated if they are significantly larger than adjacent returns. This method avoids the need for explicit threshold selection and performs well in jump-dominated environments. Despite their effectiveness against jumps, truncation-based RV estimators inherit the noise sensitivity of the classical RV estimator. As a result, they become inconsistent when prices are contaminated by market microstructure noise.

1.6.1 Market Microstructure Noise and Noise-Robust Estimators

Market microstructure noise refers to the deviation of observed transaction prices from the latent efficient price that reflects the fundamental value of an asset. A standard econometric representation models the observed log-price Y_t as

$$Y_t = X_t + \varepsilon_t, \tag{1.9}$$

where X_t denotes the latent efficient log-price process, typically assumed to be a continuous Itô semimartingale, and ε_t represents microstructure noise. The noise component captures bid–ask bounce, price discreteness due to tick size constraints, order processing costs, inventory effects, and other trading frictions. It is commonly assumed that ε_t is i.i.d. with mean zero and variance η^2 , and independent of X_t .

In frictionless markets, the integrated volatility over the interval $[0, T]$,

$$IV_T = \int_0^T \sigma_t^2 dt, \quad (1.10)$$

can be consistently estimated using realized variance,

$$RV_n = \sum_{i=1}^n (\Delta_i^n X)^2, \quad \Delta_i^n X = X_{t_i} - X_{t_{i-1}}, \quad (1.11)$$

as the sampling frequency $n \rightarrow \infty$. However, when prices are observed with noise, returns satisfy

$$\Delta_i^n Y = \Delta_i^n X + \Delta_i^n \varepsilon, \quad (1.12)$$

and the realized variance estimator becomes

$$\begin{aligned} RV_n &= \sum_{i=1}^n (\Delta_i^n Y)^2 \\ &= \sum_{i=1}^n (\Delta_i^n X)^2 + \sum_{i=1}^n (\Delta_i^n \varepsilon)^2 + 2 \sum_{i=1}^n \Delta_i^n X \Delta_i^n \varepsilon. \end{aligned} \quad (1.13)$$

Since $\sum_{i=1}^n (\Delta_i^n \varepsilon)^2 \approx 2n\eta^2$, the noise-induced component grows linearly with the sampling frequency. As shown by [Hansen and Lunde \(2006\)](#) and [Bandi and Russell \(2008\)](#), this causes realized variance to diverge as $n \rightarrow \infty$, implying that high frequency sampling (higher than 5 minutes) exacerbates estimation error, as it will be demonstrated later in the study.

To address this issue, [Zhang et al. \(2005\)](#) propose the two-scales realized variance (TSRV) estimator. Let $RV^{(K)}$ denote realized variance computed using sparse sampling every K observations, and let $RV^{(1)}$ denote the full-sample realized variance. The TSRV estimator is defined as

$$\widehat{IV}_{\text{TSRV}} = RV^{(K)} - \frac{K}{n} RV^{(1)}. \quad (1.14)$$

By combining information from coarse and fine sampling frequencies, TSRV corrects for noise-induced bias and yields a consistent estimator of integrated volatility. [Zhang \(2006\)](#) extends this idea to the multi-scales realized variance estimator, which optimally aggregates realized variance measures across multiple sampling scales and improves convergence rates.

An alternative and influential class of estimators is realized kernel methods, introduced by [Barndorff-Nielsen et al. \(2008\)](#) and further developed in [Barndorff-Nielsen et al. \(2011\)](#). These estimators exploit the serial dependence induced by microstructure noise in high-frequency returns by weighting autocovariances using a kernel function. The realized

kernel estimator takes the form

$$\widehat{IV}_{\text{RK}} = \sum_{h=-H}^H k\left(\frac{h}{H}\right) \sum_{i=|h|+1}^n \Delta_i^n Y \Delta_{i-|h|}^n Y, \quad (1.15)$$

where $k(\cdot)$ is a kernel function and H is a bandwidth parameter. Appropriate kernel choices filter out noise-induced autocorrelation while preserving information about the latent price variation.

Closely related are pre-averaging estimators developed by [Jacod et al. \(2009\)](#) and [Jacod et al. \(2010\)](#). These methods construct locally averaged returns,

$$\bar{\Delta}_i^n Y = \sum_{j=1}^{k_n} g\left(\frac{j}{k_n}\right) (Y_{t_{i+j}} - Y_{t_{i+j-1}}), \quad (1.16)$$

where $g(\cdot)$ is a weight function and k_n is a window size. By smoothing high-frequency returns prior to forming quadratic variation, pre-averaging reduces noise contamination while retaining consistency and asymptotic efficiency.

Overall, this literature shows that microstructure noise fundamentally alters the asymptotic behavior of high-frequency volatility estimators. Reliable inference therefore requires econometric techniques that explicitly account for trading frictions rather than relying on increasingly fine sampling of observed prices.

1.6.2 Joint Presence of Jumps and Noise

While the aforementioned estimators address either jumps or noise, real financial data typically exhibit both features simultaneously. This has motivated a growing literature on estimators that are robust to both jumps and market microstructure noise. [Podolskij and Vetter \(2009a,b\)](#) and [Christensen et al. \(2014\)](#) combined pre-averaging techniques with bipower variation, yielding estimators that are consistent in the presence of finite activity jumps and noise. [Christensen et al. \(2018\)](#) further enhanced this framework by incorporating truncation indicators into pre-averaged bipower variations.

Other approaches adopt signal-processing techniques. [Fan and Wang \(2007\)](#) and [Barunik and Vacha \(2015\)](#) used wavelet methods to detect and remove jumps before applying noise-robust estimators such as TSRV or MSRV. Although effective in practice, wavelet-based methods are computationally intensive and their theoretical properties under infinite activity jumps have received little attention and have yet to be established.

[Christensen et al. \(2010\)](#) proposed estimators based on intraday quantile ranges, which are naturally robust to extreme observations caused by jumps. [Li et al. \(2014\)](#) introduced a truncated pre-averaging estimator by excluding large pre-averaged returns, while [Barndorff-](#)

Nielsen et al. (2008) and Barndorff-Nielsen et al. (2011) developed related kernel-based approaches that address both noise and jumps. Varneskov (2017) combined realized kernels with nearest-neighbor truncation, resulting in estimators that jointly address jumps and noise.

Recent research has continued to refine truncation-based methodologies, demonstrating that truncating noise-robust estimators often yields improved finite-sample performance. The truncated two-scales and truncated flat-top realized kernel estimators introduced in this line of work are consistent under both noise and jump contamination and are relatively easy to implement. Importantly, truncated estimators are closely related to wavelet-based approaches but offer substantial computational advantages.

Beyond volatility estimation, truncation methods have found applications in studying the interaction between jumps and volatility (Jacod and Todorov, 2010), estimating covariation between returns and volatility (Aït-Sahalia, 2002), and testing for jumps and cojumps in multivariate settings (Barndorff-Nielsen et al., 2011; Jacod et al., 2009; Lee and Mykland, 2012). This broader literature underscores the central role of jump detection and noise-robust inference in high-frequency financial econometrics.

1.7 Staleness

In empirical finance, the literature seems to suggest to sample asset prices on evenly spaced time grids, such as one- or five-minute intervals. This practice is motivated by convenience, comparability across assets, and the requirements of many econometric techniques that assume regularly spaced observations. However, when transaction prices or quotes are sampled in this way, a striking empirical regularity emerges: a large fraction of observed returns are zero. For the stocks belonging to the S&P 500 the percentage of zero returns is 33.96% at the one-minute frequency, 16.08% at five minutes, and still as high as 11.52% at the relatively low intraday frequency of ten minutes (Kolokolov et al. (2020)). Zeros are therefore a non-negligible feature of high-frequency financial data, even for the most liquid equities in the market.

The omnipresence of zero returns reflects fundamental features of the trading process. Two main mechanisms are typically highlighted in the literature. The first is price staleness, an economically meaningful phenomenon whereby the efficient price evolves continuously, but the observed transaction price remains unchanged because no trade occurs or because incoming orders do not suffice to move the price. The second is price discreteness, arising from the fact that prices are quoted on a discrete grid determined by the tick size. While both mechanisms can generate zero returns, they have different economic interpretations and econometric implications. Bandi et al. (2017) emphasize the role of price staleness as

a structural feature of high-frequency markets, while [Bandi et al. \(2020\)](#) discuss methods to disentangle staleness from discreteness.

Despite the empirical relevance of zero returns, the vast majority of continuous-time and discrete-time models used in finance postulate the mere impossibility of observing a zero return over any positive time interval. Classical diffusion models, jump–diffusion processes, and their numerous extensions imply that returns have continuous distributions with zero probability mass at exactly zero. This modeling choice is often made for analytical tractability, but it stands in stark contrast with the data. As a result, zero returns are typically treated as an inconvenience to be filtered out, ignored, or smoothed away, rather than as informative observations generated by the data-generating process.

Price staleness refers to the phenomenon whereby the observed price remains unchanged over a given sampling interval even though the underlying efficient price may have moved. At very high frequencies, this is primarily due to infrequent trading, strategic order submission, and market frictions such as order-processing costs and inventory considerations of liquidity providers. When no transaction occurs within a given interval, the last observed price is mechanically carried forward, producing a zero return. Even when trades do occur, they may take place at the same price level, particularly in the presence of a relatively large tick size or clustered order flow.

From an economic perspective, staleness is closely related to market liquidity. Assets with lower trading intensity or more episodic order flow tend to exhibit higher proportions of zero returns. However, as the evidence from highly liquid S&P 500 stocks shows, staleness is not confined to illiquid assets. Instead, it is a pervasive feature of modern electronic markets, reflecting the interaction between high-frequency trading, order book dynamics, and discrete trading opportunities. [Bandi et al. \(2017\)](#) formalize this intuition by modeling the observed price as a discretely sampled version of a latent efficient price that is only revealed when trading activity occurs. In their framework, zero returns arise naturally as periods of non-revelation.

Importantly, staleness implies that zero returns are informative. They convey information about the absence of trades, the state of liquidity, and the intensity of market activity. Ignoring zeros therefore amounts to discarding economically meaningful signals. Moreover, staleness induces serial dependence in returns, as consecutive zero returns are more likely to occur during low-activity periods. This feature violates the assumptions of many standard models and estimators that rely on independence or weak dependence.

Price discreteness provides a complementary explanation for zero returns. Because prices can only change in multiples of the minimum tick size, small movements in the efficient price may not be reflected in observed transaction prices. When the efficient price fluctuates within a tick interval, the observed price remains constant, generating zero

returns upon sampling. This mechanism is particularly relevant at high frequencies and for assets with relatively large ticks compared to their volatility.

The literature on rounding and discreteness has a long tradition in financial econometrics. More recently, [Aït-Sahalia and Yu \(2008\)](#) study the impact of market microstructure noise, including rounding effects, on volatility estimation. While these models can generate zero returns mechanically, they typically abstract from the strategic and institutional features that give rise to staleness. [Bandi et al. \(2020\)](#) emphasize that, although discreteness and staleness may be observationally similar, conflating the two can lead to misleading conclusions about market quality and price discovery.

1.7.1 Consequences of Ignoring Zero Returns

Ignoring the presence of zero returns can have several detrimental consequences. First, standard estimators of volatility, such as realized variance, may be biased when applied to data with a high fraction of zeros. Staleness effectively induces irregular sampling, violating the assumptions underlying many asymptotic results. Second, inference procedures that rely on continuous return distributions may suffer from size distortions and loss of power. Third, model misspecification can lead to incorrect economic interpretations, for instance by overstating the degree of market efficiency or understating the role of liquidity frictions.

More broadly, treating zero returns as impossible or irrelevant obscures an important dimension of market behavior. Zeros are not merely artifacts of data handling but are manifestations of the trading process itself. A comprehensive modeling approach to high-frequency returns must therefore account explicitly for price staleness and discreteness. This recognition motivates a re-examination of standard econometric tools and calls for models that align more closely with the empirical realities of high-frequency financial markets.

In light of these considerations, a central question arises: what are the dangers of ignoring the presence of zeros in the data-generating process, and how can models be adapted to incorporate them in a theoretically coherent and empirically relevant manner?

This was the question that motivate a recent work from [Kolokolov et al. \(2020\)](#) where they set to investigate how zeros in the returns impacts jump inference and how many of the identified discontinuities that modern jump tests have identified are actually jumps and not a product of staleness. To illustrate the effects this stylized fact along with the correction that the authors proposed we begun with the usual assumption that the log price process could be considered as:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t, \quad (1.17)$$

where W_t is a standard Brownian motion, σ_t is a stochastic volatility process, and J_t is a finite-activity jump process. We assume the the discontinuous part could be modeled as :

$$Z_t = \sum_{j=1}^{N_t} k_j \quad (1.18)$$

where N_t is the Poisson random variable. Rather than observing X_t directly, the econometrician observes transaction prices on a discrete grid $\{t_i\}_{i=1}^n$. Let

$$Y_{t_i} = X_{\tau(i)}, \quad \tau(i) \leq t_i, \quad (1.19)$$

where $\tau(i)$ denotes the most recent time at which the price was updated. Observed returns are therefore given by

$$\Delta Y_i = Y_{t_i} - Y_{t_{i-1}}. \quad (1.20)$$

When prices are stale, $\tau(i) = \tau(i-1)$ and $\Delta Y_i = 0$. When an update finally occurs, the observed return aggregates multiple latent price increments. Standard jump tests rely on comparing realized variance,

$$RV_n = \sum_{i=1}^n (\Delta Y_i)^2, \quad (1.21)$$

to jump-robust measures of integrated volatility such as bipower variation,

$$BV_n = \sum_{i=2}^n |\Delta Y_i| |\Delta Y_{i-1}|. \quad (1.22)$$

Under synchronous sampling and no jumps, RV_n and BV_n converge to the same limit. However, suppose prices remain stale for k consecutive intervals and then update at time t_i . Even if the latent price follows a purely continuous process, the resulting return satisfies

$$\Delta Y_i = X_{t_i} - X_{t_{i-k}}, \quad (1.23)$$

with

$$\mathbb{E}[(\Delta Y_i)^2] = \int_{t_{i-k}}^{t_i} \sigma_s^2 ds. \quad (1.24)$$

Thus, staleness produces large, isolated returns mimicking the discontinuous part and resolving in considerably higher identification of jumps. To disentangle genuine jumps from stale-price effects, [Kolokolov et al. \(2020\)](#) introduce a latent updating indicator:

$$S_i = \begin{cases} 1, & \text{if the price updates at } t_i, \\ 0, & \text{if the price is stale.} \end{cases} \quad (1.25)$$

Observed returns can then be written as

$$\Delta Y_i = S_i \left(X_{t_i} - X_{t_{\tau(i-1)}} \right). \quad (1.26)$$

The updating process $\{S_i\}$ is modeled probabilistically, typically as a Bernoulli process with intensity λ_S . This assumption reflects the discrete nature of transaction prices and allows staleness to be treated as a structural component of the observation process rather than as noise to be filtered out.

The key identification insight of the paper is that jumps and staleness generate distinct temporal patterns. Jumps produce isolated large returns, whereas staleness generates clusters of zero returns followed by a large observation.

Formally, the density of ΔY_i can be expressed as a mixture:

$$f(\Delta Y_i) = P(S_i = 0) \mathbb{I}\{\Delta Y_i = 0\} + P(S_i = 1) f_X(\Delta Y_i), \quad (1.27)$$

where f_X is the conditional density implied by the latent price process, including possible jumps.

By embedding this mixture structure into a full likelihood and working under in-fill asymptotics, the authors show that jump intensity λ_J and staleness intensity λ_S are separately identifiable. Ignoring staleness leads to asymptotic bias:

$$\hat{\lambda}_J^{\text{naive}} \xrightarrow{P} \lambda_J + \lambda_S. \quad (1.28)$$

Empirical applications reveal that a substantial fraction of detected jumps disappears once staleness is explicitly modeled. Assets with lower liquidity exhibit higher estimated staleness intensities, and periods of market stress amplify the confounding effects of stale prices.

These findings suggest that part of the jump evidence documented in the literature may reflect liquidity frictions rather than true discontinuities in the efficient price.

[Kolokolov et al. \(2020\)](#) provide a compelling re-examination of jump detection in high-frequency data. By explicitly modeling the observation process, the paper highlights an identification problem that has been largely overlooked.

1.8 Main Findings

This thesis was devoted to the study of discontinuities present in financial assets, we focus on the identification, estimation of volatility and jump inference while accounting for stylized features of high frequency financial data. This section outlines the methodologies, employed in this study along with the main findings each chapter provided us.

In our second chapter we provide a review of jump tests, specifically we focused on the non-parametric one-sided upper-tailed tests that have been proposed in the literature. We have conducted a meticulous comparison of the most widely used jump detection procedures, examining their statistical properties using various simulated conditions. In particular, we evaluate their empirical size and power through an extensive Monte Carlo study that incorporates multiple stochastic price processes, alternative discretization schemes, and varying threshold parameter choices where applicable. Furthermore we examine how zeros in the return series affect jump inference for those procedures, namely if and when their asymptotic theory fails, a feature often overlooked in previous studies. Based on these findings, we rank the competing methodologies and provide detailed guidance regarding their practical implementation, highlighting their relative strengths, weaknesses, and robustness across different data-generating scenarios. Specifically our simulations and empirical results indicate that no single jump detection test performs optimally across all market conditions, making the use of multiple procedures essential for reliable inference. Test performance drops notably in the presence of additive noise or zero returns. Among individual tests, LM12 and ASJL exhibit increased size relative to nominal level and to scenarios with frequent volatility jumps, indicating heightened sensitivity and a tendency to over identify jumps. The ASJ, CPR, and PZ tests are highly dependent on the threshold parameter, we choose to use the values recommended by the authors leads to substantial size distortions. For the PZ test, the chapter considers two main variants, the PZ2 and PZ4 where their difference is the choice of the threshold value. PZ4 performs best at the highest sampling frequencies, while PZ2 becomes more effective as frequency decreases. Across tests, convergence is more evident at lower frequencies. Simulations and empirical results show that the BNS test maintains moderate size and power across most settings, whereas the ASJ test is consistently the least powerful. All procedures are sensitive to sampling frequency, if the frequency is high enough the effects of microstructure noise take effect and lead to substantial bias and spurious jump detection. On the other hand if the frequency is too low the, much of the information about the discontinuous part has been lost and the procedures then tend to underestimate the jump activity. Based on the Monte Carlo simulations we find that the window between 20 seconds for high frequency observations and 5 minutes for a relative low frequency sampling tend to produce the best result. The simulation framework incorporates three stochastic volatility models and two discretization schemes, with the latter having no statistically significant effect. Among models, Heston is the least volatile and Double Heston the most, a pattern also seen in their sizes. Empirical test sizes at the 1 minute frequency exceed simulated ones, suggesting either higher noise and zero prevalence in real data. Interestingly, when no jumps or noise are present, test disagreement is substantial, but the introduction of jumps, noise, and zeros leads to greater convergence. Volatility only jumps are not detected by any test, even at high frequencies,

and jump size matters more than jump frequency. Finally, LM12 and ASJL perform better, primarily at higher frequencies, especially when additive noise is introduced.

In the third chapter, under the assumption that the price process of high frequency financial assets is contaminated by the simultaneous presence of microstructure noise and staleness, we provide a correction to the multipower variation estimator. We formally derive its theoretical and finite-sample distributional properties and establish its consistency under general conditions. A extensive simulation study is conducted to evaluate its performance relative to conventional estimators. Our findings align with the recent literature indicating that stochastic volatility models driven by Brownian motion have significantly less share in the overall observed price behavior. Our results suggest that both staleness and noise must be embedded in the assumptions underlying high-frequency data generation. Ignoring these frictions can bias multipower variation estimators and undermine broader econometric inference, leading to bias estimates of the volatility process.

Finally, in chapter 4 and by utilizing the proposed correction for the multipower variation estimator from the previous chapter, we apply it to two well studied jump test, in an attempt to improve jump inference under the assume underlying process. Through a detailed simulation analysis, we reassess jump inference and update the empirical ranking of jump identification procedures.

Our work done throughout those chapters have lead to some interesting conclusions that we believe can expand the current literature on the field of high-frequency jump identification. In addition we fill that our results provided us with a deeper understanding of the econometrics involved around jump inference and the challenges faced by it.

Chapter 2

Review of Jump tests

This chapter consists a meticulous comparison of non-parametric jump detection procedures for discontinuities in the price process of financial instruments. We provide an in depth analyses of all the jump identification procedures that have been developed. Through an in depth Monte Carlo exercise we give focus to the robustness of the tests by displaying their size and power under different simulated conditions. We differentiate from the literature firstly as we have consider multiple stochastic models and discretization procedures for the simulations, secondly we document the impact of different values of the threshold parameter for the test that depend on it. Thirdly the impact of zeros under all stochastic models and jump tests was also studied. Finally we provide detailed recommendations for their use.

2.1 Introduction

In the early days of the 20th century the use of advanced mathematics in the finance domain was to say at least limited, this status quo changed when Louis Bachelier, a french mathematician in the 1900s defended and successfully published his Ph.D. thesis under the title "The theory of speculation" [Samuelson et al. \(2006\)](#). He is considered today the father of stochastic finance and financial engineering since he was the first to use the idea of a purely stochastic process to model stock dynamics. To achieve this he use the notion of Brownian motion to describe the t-forward price of a stock and derive a successful pricing formula for European options. His work was as influential as it was pioneering because it set the basis for the work of Albert Einstein and in the finance discipline since it was the predecessor for the Black and Scholes model, also his research predates what is known today as the efficient market hypothesis. From that moment onward the aeolian bag opens and to this day numerous stochastic models have been developed becoming better in capturing the stock price and volatility dynamics, but in a strange twist of faith

Bachelier's model was used again two years before, the model assumed that the stock price can take negative values with a positive probability, this first was consider the greatest limitation of the model, but in April of 2020 the oil futures contracts for the first time display negative prices, furthermore the global financial crisis of 2008 and the recent Pandemic have triggered negative interest rates, spiking once again after over a century the interest for his model. Since empirical as well as theoretical research has backed the use of the stochastic volatility model to capture the return dynamics. Those can be decomposed into changes due to the existence of volatility clustering and a purely stochastic part, the jump component. The ability to observe the time a jump takes place and its size is crucial in several fields, most notably in option pricing where the ability to separate the noise of Brownian motion from the jump part is important as the two display different hedging requirements and possibilities. Portfolio allocation remains yet another significant area in which the distinction between Brownian and jump risks can optimize the demand for assets subject to both kinds of risk. Risk management consists of another essential task where the separation of Brownian noise and jump come into hand, as the Brownian part can be managed by employing Gaussian tools while the jumps part can be controlled through tail statistics for example the VaR. Jumps and their existence have the ability to influence the outcome of studies due to the fact that can influence financial variables such as asset prices, interest rates, and currencies [Ait-Sahalia \(2004\)](#).

The significance of the jump part had led to the creation of a different class of models attempting to capture it, to better explain the dynamics of the assets. First in this race was the jump-diffusion models such as the Merton option pricing formula [Merton \(1976\)](#). But this approach has certain drawbacks, firstly there is no closed form expression for the Likelihood function and secondly, the number of parameters to be estimated is large. Affine Jump Diffusion models (AJD), consist of yet another family of mathematical expressions attempting to model the asset price dynamics, in this class belong to the models of Vasicek 1977 and Cox Ingersoll and Ross 1985. Even though they work well, they tend to demand a high number of restrictive assumptions. Finally, there are nonlinear volatility models such as the Gallant and Tauchen 2002, but they are computationally intensive and they heavily rely on a supplementary model. From the firsts attempting to point the discontinuities of the price process in a discrete-time is the [Ait-Sahalia \(2002\)](#) by exploiting the properties of transition densities for option data, was able to separate the jumps from the diffusion part, indicating that the assumption of the price process been driven purely by a diffusion model does not hold. With the surface of high-frequency data the jump testing procedures started to drift towards a non-parametric model-free methodology, [Mancini et al. \(2001\)](#) was the first non-parametric jumps diffusion model that forms the bases for the rest of the testing procedures. This chapter will explore all the non parametric jump detection procedures that have been proposed to date, Except for the initial contribution of [Barndorff-Nielsen](#)

and Shephard (2004) in estimating the quadratic variation in discrete returns we also include the work of Huang and Tauchen (2005), Lee and Mykland (2008), Jiang and Oomen (2008), Corsi et al. (2010), Podolskij and Vetter (2009b), Aït-Sahalia and Jacod (2009), Andersen et al. (2007b) and Fan and Fan (2011). With the mounting plethora of jump detecting procedures the question of what test performs the best in which data and under what condition still remains, such the main aim of the chapter is to provide a comprehensive comparison of all in a thorough Monte Carlo simulation as well as applying them to empirical data. To our knowledge, three previous papers address the same issues, in particular, we refer to the work of Theodosiou and Zikes (2011) investigating the performance of the test as well as the deviations from each other focusing on the effects of microstructure noise, the existence of zero returns and the impact of diurnal volatility pattern. Dumitru and Urga (2012) examine the influence of different sampling frequencies, different levels, and persistence of volatility, microstructure noise, and of jump size and intensities in the efficiency of the testing procedures. Maneesoonthorn et al. (2020) similarly examines the robustness of the test, under the impact of volatility jumps, microstructure noise, and of sampling frequency. As such the chapter aims to conduct a detailed examination of the jump test in a Monte Carlo simulation exercise with multiple stochastic volatility models and employing two different discretization procedures under different environments and conditions. The contribution of our paper is threefold, first and foremost is a meticulous comparison of major non-parametric jump testing procedures under different stochastic models and discretization methods, jump sizes and jump arrival times, different distributions of additive noise, and in three different sampling frequencies. We also obtain results for the impact of zero returns in the jump test, since the financial data are likely not updating every second. Secondly, we try to determine the most effective threshold choice for the PZ test in order to balance the trade between size distortion and jump detection. Lastly, we conduct an empirical study of the jump tests in stock, Foreign Exchange, and cryptocurrency markets. We hope for our research to provide helpful insights to both researchers and practitioners.

2.2 Introspection of the jump tests

This section reviews the main documented Jump identification procedures that have been developed and reviews the mechanism of each one, responsible to identify discontinuities in the price process. Let us assume that the logarithmic price process belongs to the class of Brownian semimartingales, it can be written as

$$d \log X_t = (\mu_t - \frac{1}{2} \sigma_t^2) dt + \sigma_t dW_t + dZ_t \quad (2.1)$$

$$\log X_t = \log X_0 + \int_0^t (\mu_s - \frac{1}{2} \sigma_s^2) ds + \int_0^t \sigma_s dW_s + Z_t \quad (2.2)$$

μ_t is a drift parameter, W_t is a Brownian motion at time t , σ_t denotes the spot volatility and Z represents the jump process at time t , defined as

$$Z_t = \sum_{j=1}^{N_t} k_j \quad (2.3)$$

N_t consists of the increment of a discrete counting process and k_j is a non-zero random variable.

$$P(N_t = n) = e^{-\lambda} \frac{\lambda^n}{n!}. \quad (2.4)$$

Taking the baton from the influential work of [Andersen and Bollerslev \(1998\)](#), dividing the process $[t-1, t]$ into M equally spaced squared intraday returns $\Delta_n = 1/M$ and summing them it is possible to obtain the realized variance (RV_t) to use as an estimator of the quadratic variation.

$$RV_{t,M} = \sum_{i=1}^M r_{ti}^2 \quad (2.5)$$

where r is the i -th log-intraday return computed as

$$r_{ti} = \log X_{t-1+i\Delta_n} - \log X_{t-1+(i-1)\Delta_n}, \quad i = 1, \dots, M. \quad (2.6)$$

The continuous and discontinuous components of the price process can be approximated by the realized variance. The total quadratic variation can be defined as

$$QV_t = \lim_{M \rightarrow \infty} RV_{t,M} = \int_{t-1}^t \sigma_s^2 ds + \sum_{j=1}^{N_t} k_j^2 \quad (2.7)$$

The integrated variance is the first part of the equation whereas the discontinuous path is represented by the second part of it. It is essential to decompose the two different sources of variability as they have different natures and should therefore be modeled separately. The first part of the [Equation 2.7](#) is the integrated variance (IV) while the second part is the jump part. In a continuous time process, the early tests drew power from the difference between a jump robust variation measure and the variation from the continuous part. The first is been achieved through the quadratic variation while the last from the integrated variance. The tests that exist and are reviewed in this section are

separated into 5 categories according to what properties or financial tools they use, 1) Squared variation, 2) Higher order p-variation, 3) Standardized returns, 4) Variance swaps and 5) Microstructure noise. An important disclaimer here is that even though their name suggests that those are jump detection procedures, there were not designed to provide information regarding the presence or the absence of jumps in a process. Their purpose is to be able to distinguish in a specific time interval $(0, t]$ if the process is continuous or not.

All procedures that will be reviewed are concerned whether the realized process is continuous either at a specific time or over a given interval. The tests of [Andersen et al. \(2007b\)](#) and, [Lee and Mykland \(2008\)](#) use the null of continuity at a specific time t_i , whereas the rest of the procedures test continuity over a fixed interval, such as a trading day. In all cases, the alternative implies at least one jump. The majority of the testing procedures use the realized multi-power variation estimators which remain robust to a finite number of jumps in a certain time interval, for that reason all the testing procedures work well when a finite number of jumps can occur in a time interval. Exception to this are the test of [Aït-Sahalia and Jacod \(2009\)](#) [Podolskij and Ziggel \(2010\)](#).

2.2.1 Tests based on squared variation

Barndorff-Nielsen and Shephard 2006

As it has been mentioned the difference between the quadratic variation (QV) and an estimator of the integrated variance (IV) can provide the contribution of the jump component to the price evolution. [Barndorff-Nielsen and Shephard \(2004\)](#) provide a consistent estimator for the IV named realized bipower variation which is robust to the presence of jumps. Based on their previous work they construct a test statistic that exploits the difference between the realized variance (RV) and the realized bipower variation. All the tests that belong to this category utilize the realized variance RV and different estimators for the integrated variance to construct the test statistic

$$BV_{t,M} = \mu_1^{-2} \frac{M}{M-1} \sum_{i=2}^M |r_{t,i-1}| |r_{ti}| \quad \mu_1 = \sqrt{2/\pi} \quad (2.8)$$

The premise of the test is that under the assumption of finite activity jumps the probability to observe two successive jumps approaches zero as the sampling frequency increases. As such the product of two consecutive returns should theoretically be driven solely by the

diffusion part alone. The test statistic can be obtained by

$$BNS_{t,M} = \frac{1 - \frac{BV_{t,M}}{RV_{t,M}}}{\sqrt{\left(\frac{\pi^2}{4} + \pi - 5\right) \frac{1}{M} \max\left(1, \frac{TPQ_{t,M}}{BV_{t,M}^2}\right)}} \xrightarrow{L} N(0, 1). \quad (2.9)$$

Where $TPQ_{t,M}$ stands for the realized tripower quaticity and is defined as

$$TPQ_{t,M} = \mu_{4/3}^{-3} \frac{M^2}{M-2} \sum_{i=3}^M |r_{t,i-2}|^{4/3} |r_{t,i-1}|^{4/3} |r_{ti}|^{4/3}, \quad (2.10)$$

and $\mu_{4/3} = E(|U|)^{4/3}$ where U is a standard normal random variable. The adjusted jump ratio statistic is

$$BNS = \frac{1 - \frac{MPV_{t,M}}{RV_{t,M}}}{\sqrt{\theta_p M^{-1} \max\left(1, \frac{MPQ_{t,M}}{MPV_{t,M}^2}\right)}} \quad (2.11)$$

MPQ and MPV denote the multipower quaticity and multipower variation given in equations Equation 2.12 and Equation 2.13 respectively and θ_p is provided by Huang and Tauchen (2005).

$$MPQ_{t,M}(p) = \mu_{4/p}^{-p} \frac{M^2}{M-p+1} \sum_{i=p}^M \prod_{j=0}^{p-1} |r_{t(i-j)}|^{4/p} \quad (2.12)$$

$$MPV_{t,M}(p) = \mu_{2/p}^{-p} \frac{M^p}{M-p+1} \sum_{i=p}^M \prod_{j=0}^{p-1} |r_{t(i-j)}|^{2/p} \quad (2.13)$$

Corsi, Pirino, Reno 2010 CPR test

Their paper is based on a conflicting notion reported in the work of Andersen et al. (2007b), as well as Forsberg and Ghysels (2007), namely those papers, provide evidence of a negative impact of jumps on future volatility. Corsi et al. (2010) show that the results mentioned in those papers are due to the limitations of the realized multipower variations estimators, and specifically, the culprit appear to be the small sample bias. By combining the realized multipower variations and threshold realized variance Mancini (2009) they introduce a new estimator for the integrated variance, the threshold multipower variation.

By exploiting their new estimator they construct a new test statistic for detecting jumps ¹.

$$CPR = \frac{1 - \frac{C-TBV_t}{RV_t}}{(\sqrt{\frac{\pi^2}{4} + \pi - 5})(1/M) \max(1, \frac{C-TT_{ri}PV_t}{C-TBV_t^2})} \xrightarrow{L} N(0, 1) \quad (2.14)$$

the realized threshold bipower and tripower variation are defined as

$$C-TBV_t = \frac{\pi}{2} \sum_{i=2}^M Z_1(r_{ti}, \theta_i) Z_1(r_{t,i-1}, \theta_{i-1}), \quad (2.15)$$

$$C-TTriPV_t = \mu_{4/3}^{-3} \sum_{i=3}^M Z_1(r_{ti}, \theta_i) Z_1(r_{t,i-1}, \theta_{i-1}) Z_1(r_{t,i-2}, \theta_{i-2}). \quad (2.16)$$

$Z(r_j, \theta_j)$ is a returns function with a scale constant (θ_j) that takes into account the local volatility estimator ($\hat{V}_j$)

$$Z(r_i, \theta_i) = \begin{cases} |r_i|, & r_i^2 < \theta_i \\ \frac{1}{2M(-C_\theta)\sqrt{\pi}} \left(\frac{2\theta_i}{c_\theta^2}\right)^{\gamma/2} \Gamma\left(\frac{\gamma+1}{2}, \frac{c_\theta^2}{2}\right), & r_i^2 > \theta_i \end{cases} \quad (2.17)$$

For the $Z(r_j, \theta_j)$ we set $\gamma = 1$ in the first case and $\gamma = 4/3$ in the second case. To compute Equation 2.16 we need $Z_1(r_j, \theta_j)$ which can be computed as:

$$Z_1(r_{ti}, \theta_i) = \begin{cases} |r_{ti}|, & r_{ti}^2 < \theta_i \\ 1.094\sqrt{\theta_i}, & r_{ti}^2 \geq \theta_i. \end{cases} \quad (2.18)$$

The threshold θ_j is computed as

$$\theta_j = c_\theta^2 \hat{V}_j \quad (2.19)$$

Where c_θ is a scale-free constant and $\hat{V}_j$ denotes the local volatility estimator given by

$$\hat{V}_t^Z = \frac{BV_t}{K-2} \quad (2.20)$$

For the local volatility estimate Equation 2.20 the authors suggest using a non-parametric filter (such as the Gaussian kernel) that with several iterations managed to remove the jumps, for that purpose practitioners can use either the method of Lee and Mykland (2008) or the GARCH(1,1) as in Mancini and Renò (2011). In line with the recommendations Corsi et al. (2010) we use $c_\theta = 3$ and $L = 25$.

¹The test statistic, will tend to converge to its asymptotic distribution, shown by the authors to be the standard normal as the sample size $M \rightarrow \infty$

Andersen, Dobrev, Schaumburg 2009

The shortcomings of the bipower variation has been documented and for that reason it has been generalized to the tripower variation (see [Barndorff-Nielsen et al. \(2006a\)](#)) and although the finite sample performance has improved, it remains to be sensitive to the effects of microstructure noise in high-frequency data and the class of multipower variation estimators in general face issues when the returns series displays zeros. As their predecessors they produce new estimators, minimum realized variance (MinRV) and median realized variance (MedRV) that rely on the nearest neighbor truncation, in an effort to overcome the issues of existent estimators of the integrated volatility in the presence of jumps.

$$MinRV_t = \frac{\pi}{\pi - 2} \frac{M}{M - 1} \sum_{j=2}^M \min(|r_{ti}|, |r_{ti-1}|)^2 \quad (2.21)$$

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \frac{M}{M - 2} \sum_{i=3}^M \text{med}(|r_{ti}|, |r_{ti-1}|, |r_{ti-2}|)^2 \quad (2.22)$$

$$MinRQ_t = \frac{\pi M}{3\pi - 8} \frac{M}{M - 1} \sum_{i=2}^M \min(|r_{ti}|, |r_{ti-1}|)^4 \quad (2.23)$$

$$MedRQ_t = \frac{3\pi M}{9\pi + 72 - 52\sqrt{3}} \frac{M}{M - 2} \sum_{i=3}^M \text{med}(|r_{ti}|, |r_{ti-1}|, |r_{ti-2}|)^4 \quad (2.24)$$

To detect jumps a method of comparing those estimators with the realized variance (RV_t) is important, for that reason, the authors propose two new test statistics. $MinRQ_t$ and $MedRQ_t$ designate the minimum realized and median realized quaticity as estimates for the integrated quaticity.

$$MinRQ_{t,M} = \frac{1 - \frac{MinRV_t}{RV_t}}{\sqrt{1.81M^{-1} \max(1, \frac{MinRQ_t}{MinRV_t^2})}} \xrightarrow{L} N(0, 1) \quad (2.25)$$

$$MedRQ_{t,M} = \frac{1 - \frac{MedRV_t}{RV_t}}{\sqrt{0.96M^{-1} \max(1, \frac{MedRQ_t}{MedRV_t^2})}} \xrightarrow{L} N(0, 1) \quad (2.26)$$

2.2.2 Tests based on P-power variation

Podolskij and Ziggel 2010 PZ test

This category envelops the tests that depend on higher-order p-power variation dynamics to detect discontinuities. For a price process, p_t the estimator of the p-power integrated variance (IV) can be defined with the help of the realized bipower variation of [Barndorff-](#)

Nielsen and Shephard (2004). $B(p, \hat{\Delta}_n)_t = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} |r_{ti}|^p$. The different components of the variation can be decomposed through the limiting behavior of the estimator for different values of p . When $p = 2$ then it converges to the realized quadratic variation as $n \rightarrow \infty$. For $0 < p < 2$ the estimator converges to

$$\frac{\Delta_n^{1-p/2}}{m_p} B(p, \hat{\Delta}_n)_t \xrightarrow{P} A(p_t) \quad (2.27)$$

$$A(p_t) = \int_0^t |\sigma_s|^p \quad (2.28)$$

Podolskij and Ziggel (2010) based their procedure on the truncated power variation of Mancini and Renò (2011) threshold estimator. The test statistic is derived from the difference between realized power variation and the threshold estimator perturbed by some external positive i.i.d random variable $(\eta_i)_{1 \leq i \leq \lfloor t/\delta \rfloor}$ given that $E[\eta_i] = 1$

$$T(p, \Delta_n)_t = M^{p-1/2} \sum_{i=1}^{t\delta} |r_i|^p (1 - \eta_i \mathbf{1}_{|r_i| \leq c\delta^w}), p \geq 2 \quad (2.29)$$

where $c = 2.3\sqrt{BV_t}$ and $w = 0.4$. $\mathbf{1}\{.\}$ is a condition for the difference inside the parenthesis, if the inequality holds then the difference inside the parenthesis is executed else the threshold estimator is derived only by the p -power of the returns.

$$PZ_{t,M}(p) = \frac{T(p, \Delta_n)_t}{\sqrt{\text{Var}[\eta_i] M^{p-1} \sum_{i=1}^M |r_{ti}|^{2p} \mathbf{1}_{\{|r_{ti}| \leq c\Delta_n^w\}}}} \xrightarrow{L} N(0, 1). \quad (2.30)$$

The authors suggest to sample the random i.i.d variable $((\eta_i)_{1 \leq i \leq \lfloor t/\delta \rfloor})$ from the following distribution

$$P^\eta = \frac{1}{2}(\delta_{1-t} + \delta_{1+t}) \quad (2.31)$$

and also recommend for the value of t to be a small number ($t = 0.1, t = 0.05$)

Ait-Sahalia and Jacob 2009

A similar approach for detecting discontinuities is that of Ait-Sahalia and Jacod (2009). Following the authors we define

$$A(p)_t = \int_0^t |\sigma_s|^p \quad B(p)_t = \sum_{s \leq t} |\Delta X_s|^p \quad (2.32)$$

Where p is a positive real number.² The realized power variation is given by

$$\hat{B}(p, K)_t = \sum_{j=1}^{[M/K]} |r_i|^p \quad (2.33)$$

The limiting behavior of the Equation 2.33 converge locally in t in probability.

$$\left\{ \begin{array}{l} p > 2 \Rightarrow B(\hat{p}, K)_t \xrightarrow{\mathbb{P}} B(p) \quad \text{Lépingle 1976} \\ p = 2 \Rightarrow B(\hat{p}, K)_t \xrightarrow{\mathbb{P}} A(2)_t + B(2)_t \quad \text{Realized quadratic variation} \\ p < 2 \Rightarrow \frac{\Delta_n^{1-p/2}}{m_p} \hat{B}(p, K)_t \xrightarrow{\mathbb{P}} A(p_t) \\ X \text{ is continuous} \Rightarrow \frac{\Delta_n^{1-p/2}}{m_p} \hat{B}(p, \Delta_n)_t \xrightarrow{\mathbb{P}} A(p_t) \end{array} \right. \quad (2.34)$$

By examining the above expression, the authors observe that in the case that $p > 2$ the sampling K behavior as it converges to 0, has no effect on the limit of the $\hat{B}(p, K)_t$ provided that jumps are present in $[0, t]$. Even in the case that X is continuous on $[0, t]$ then similarly $\hat{B}(p, K)_t$ will converge to a limit that is indifferent to the sampling changes. The test statistic is constructed by comparing the realized power variation at two different scales $1/M$ and K/M with $K \in \mathbb{N}$.

$$\hat{S}(p, K)_t = \frac{B(p, K)_t}{B(p, 1)_t} \rightarrow K^{p/2-1} \quad (2.35)$$

If the null hypothesis is that jumps are not present (X is continuous) the test statistic is

$$ASJ_{t,M}(p, K) = \frac{\hat{S}(p, K)_t - K^{p/2-1}}{\sqrt{\hat{V}_{t,M}}} \xrightarrow{L} N(0, 1) \quad (2.36)$$

$\hat{V}_{t,M}$ denotes the asymptotic variance of the test statistic given as

$$\hat{V}_{t,M} = M^{-1} \frac{N(p, K) \hat{A}_{t,M}(2p)}{A_{t,M}(p)^2} \quad (2.37)$$

$$\hat{A}_{t,M}(p) = \frac{(1/M)^{1-p/2}}{\mu_p} \sum_{i=1}^M |r_{ti}|^p I_{\{|r_{ti}| \leq a(1/M)^\omega\}} \quad (2.38)$$

²Note: $A(p)$ is finite for all $p > 0$ (under Assumption 1 Aït-Sahalia and Jacod (2009)), where- as $B(p)$ is finite-valued if $p \geq 2$ but often not when $p < 2$.

$$N(p, K) = \frac{1}{\mu_p^2} (K^{p-2}(1+K)\mu_{2p} + K^{p-2}(k-1)\mu_p^2 - 2k^{p/2-1}\mu_{K,p}) \quad (2.39)$$

$$\mu_{K,p} = E(|U|^p |U + \sqrt{K-1}V|^p) \quad (2.40)$$

Alternative one can use the estimator form [Mancini and Renò \(2011\)](#) to produce an estimate of the local variance. U and V are standard random normal variables. To contact the test prerequisite are the values of p, k, a, and ω , following the recommendations of the authors we set the values as follows $p = 4$, $K = 2$ and $\omega = 0.47$

2.2.3 Tests based on standardized returns

Andersen, Bollerslev, Dobrev 2007

The tests in this category utilize the behavior of the return to distinguish between the jump and the continuous part. The authors build their test upon the assumption that the asset price process follows a Gaussian distribution, by generalizing this, they assumed that if the returns are computed over a pre-determined fixed window of time and scaled by the square root of a consistent estimator of the corresponded integrated variance, then those returns should theoretically follow a Gaussian distribution as well, provided that jumps are not present. Because the ABD test and the test of [Lee and Mykland \(2008\)](#) (LM test) are very similar we choose to present them both at the same time to save space. For both tests a local volatility estimator that remains robust to the presence of jumps is crucial. Provided the intraday return r_{ti} a local volatility estimate $\hat{V}_{t,M}$ is necessary the authors as well as much of the literature uses the bipower variation estimator

$$Z_{ti} = \frac{r_{ti}}{\hat{V}_{ti}} \quad (2.41)$$

$$\hat{V}_{t,i} = \frac{1}{K-2} \sum_{k=i-K+2}^{i-1} |r_{ti}| |r_{t(i-1)}| \quad (2.42)$$

K denotes the size of the time interval used for estimating the instantaneous volatility. Provided that a jump occurs in the interval $[o, t]$ this ratio should be large, while the first attempt would be to compare it with a normal threshold this may yield misleading results as the issue of the false discovery rate arise since the test is applied to intraday time. The basic intuition for the tests is

$$ABD_{ti} = \frac{|Z_{ti}| - C_M}{S_M} \xrightarrow{L} \xi \quad P(\xi) = \exp(-e^{-x}) \quad (2.43)$$

$$C_M = \frac{(2\log(M))^{1/2}}{\mu_1} - \frac{\log \pi + \log(\log(M))}{2\mu_1(2\log(M))^{1/2}} \quad (2.44)$$

$$S_M = \frac{1}{\mu_1(2\log(M))^{1/2}} \quad (2.45)$$

Although the two tests are similar, there is one important difference, [Andersen et al. \(2007b\)](#) provide no suggestion regarding the size of the local volatility estimate. On the contrary LM test provides evidence that K depends on the sampling frequency and proposes a

$$K = \sqrt{252 * M} \quad (2.46)$$

M denotes the daily number of observations. The null hypothesis is the continuity of the sample path, we reject the null if $\frac{|Z_t - C_M|}{S_M} > \beta$ where β is $\beta = -\log(-\log(1 - a))$

2.2.4 Test based on variance swap

Jiang and Oomen 2008

Another approach for the identification of jumps is proposed by [Jiang and Oomen \(2008\)](#), which originates from the bipower variation test of [Barndorff-Nielsen and Shephard \(2006\)](#) but rather than comparing the RV to a jump robust variance measure, it depends on the comparison of RV to a jump sensitive variance measure that draws power from the higher order moments of the returns. The authors observe that the difference between arithmetic returns and logarithmic returns equates to half of the integrated variance provided that jumps are absent. As such the test statistic is constructed as follows

$$SwV_{t,M} = 2 \sum_{i=1}^M (R_{ti} - r_{ti}) \xrightarrow{p} \int_{t-1}^t \sigma_u^2 du \quad (2.47)$$

with R_{ti} denoting the arithmetic return and r_j the logarithmic return. If there is no jump in the time frame the discrepancy between the $SwV_{t,M}$ and the $RV_{t,M}$ will converge to zero. As such the test statistic can be defined as

$$JO_{t,M} = \frac{BV_{t,M}}{M\sqrt{\Omega_{SwV}}} \left(1 - \frac{RV_{t,M}}{SwV_{t,M}} \right) \xrightarrow{L} N(0, 1). \quad (2.48)$$

and Ω_{SwV} can be found using the multipower variation of [Barndorff-Nielsen et al. \(2006a\)](#)

$$\Omega_{SwV} = 3.05 \frac{M^3}{M-3} \sum_{i=1}^M \prod_{K=0}^3 |r_{ti-k}|^{3/2} \quad (2.49)$$

[Equation 2.49](#) is an estimator of the integrated sixticity $\int \sigma_u^6 du$ with the value for $p = 4$ or $p = 6$ as suggested by the authors.

2.2.5 Tests that account for microstructure noise

Lee and Mykland 2012

This family of tests is designed to accommodate the effects of microstructure noise. According to [Lee and Mykland \(2012\)](#) tests that do not take into consideration the effects of noise can discard small jumps as noise. For that reason, they suggest a process to remove the noise from the returns, and by using the method of [Podolskij and Vetter \(2009b\)](#) they are able to find the noise variance. The test statistic can be computed as follows ³

$$T_{LM12} = \frac{\max |X(t_i)| - A_M}{B_M} \xrightarrow{L} f(x) = e^{-e^{-x}} \quad (2.50)$$

$$A_M = 2 \log(\lfloor \frac{M}{KM} \rfloor)^{1/2} - \frac{\log \pi + \log(\log \lfloor \frac{M}{KM} \rfloor)}{2(2 \log \lfloor \frac{M}{KM} \rfloor)^{1/2}} \quad (2.51)$$

$$B_n = 2 \log(\lfloor \frac{n}{KM} \rfloor)^{1/2} \quad (2.52)$$

$$X(t_i) = \sqrt{\frac{M}{V_{t,M}}} (\hat{p}_{ti+KM} - \hat{p}_{ti}) \quad (2.53)$$

Ait-Sahalia and Jacob 2012

[Ait-Sahalia et al. \(2012\)](#) attempt to make a robustification of their previously developed test namely the ASJ test. This along with the majority of the rest of the testing procedures regarded in this chapter fail to accommodate the existence of noise and its implications for the test statistic. For that reason and with the recent availability of high-frequency data the authors attempt to fortify their test statistic against the impact of microstructure noise by utilizing the method of [Jacod and Todorov \(2010\)](#). They consider three alternative distributions of noise, Gaussian, Student, and a mixture of both.

$$\text{Gaussian noise: } \varepsilon_t = 2\sigma_t \left(\frac{1}{M}\right)^{1/2} \varepsilon_t^A \quad (2.54)$$

$$\text{t-distributed noise: } \varepsilon_t = 2\sigma_t \left(\frac{1}{M}\right)^{1/2} \frac{\varepsilon_t^\beta}{\sqrt{\frac{df}{df-2}}} \quad (2.55)$$

$$\text{Mixture of both: } \varepsilon_t = 2\sigma_t \left(\frac{1}{M}\right)^{1/2} \left(\frac{\varepsilon_t^A + \varepsilon_t^\beta}{\sqrt{\frac{df}{df-2}}} \right) \quad (2.56)$$

³K denotes the window size where the test statistic is applied and is estimated using the [Equation 2.46](#)

Where ε_t^A and ε_t^β are assumed to be independent i.i.d following a normal distribution with $N(0, 1)$ and a t-distribution with degree of freedom $df = 2.5$. Their results suggest that under the null hypothesis of a continuous path their robustified test statistic performs better than the 2009 test. The test statistic in form remains the same as in the ASJ test, but the power variation is computed by replacing Equation 2.33 with the realizing p-power variation of Podolskij and Ziggel (2010).

2.3 Monte Carlo Simulations

In this section, we attempt to assess the finite behavior of the tests and determine their size and power. In contrast to previous studies for the test behavior, this simulation exercise is set in order to surface different aspects of the test behavior by exploiting two different stochastic models and two different discretization procedures. In all of the simulation exercises conducted in this section, we eliminate the first 500 data points to eliminate contamination from the initial conditions. At first, we consider a Heston 1993 model with the innovations being driven by Brownian increments alone and simulate 23400 data points analogous to 6.5 hour trading day for 10000 days, with 1-second increments.

$$dS_t = \mu_t S_t dt + \sqrt{V_t} S_t dW_t^S \quad (2.57)$$

$$dV_t = k(\theta - V_t)dt + \sigma \sqrt{V_t} dW_t^V \quad (2.58)$$

We assume that the Brownian increments dW_t^S and dW_t^V are correlated, such that $\text{corr}(dW_t^S, dW_t^V) = \rho$. Here, S_t denotes the asset price process and V_t denotes the volatility process. The parameter σ represents the volatility of volatility, k is the speed of mean reversion in the volatility process, and θ is the long-run mean level to which volatility reverts. Adhering to the theoretical restrictions $2k\theta \geq \sigma^2$ we set the parameters as $k = 5$, $\theta = 0.16$, $\sigma = 0.05$ and $\rho = -0.5$. The tests are fit to the data in three different sampling frequencies, of 5 and 20 seconds as well as 1 minute. To model the alternative hypothesis, that of discontinuities in the price and volatility process we consider adding to the aforementioned model a jump part.

$$dS_t = \mu_t S_t dt + \sqrt{V_t} S_t dW_t^S + dJ_t \quad (2.59)$$

$$dV_t = k(\theta - V_t)dt + \sigma \sqrt{V_t} dW_t^V \quad (2.60)$$

The jumps follow a compound Poisson process with the parameter λ taking values from 0.2 up to 5. With mounting empirical evidence that jumps are present in the price as well as in the volatility process, we include also jumps in the volatility dynamics. With the

Euler discretization method, we consider also a two-factor stochastic volatility model, as according to mounting literature the one-factor model described above fails to capture the higher moments of the asset dynamics. [Todorov and Tauchen \(2011\)](#), [Chernov et al. \(2003\)](#) and [Bollerslev and Zhou \(2002\)](#) all fostered in their studies evidence regarding this shortcoming of the classic Heston model. The TFSV (Two Factor Stochastic Volatility Model) is described as

$$dS_t = \mu_t dt + \exp[\beta_0 + \beta_1 V_{1t} + \beta_2 V_{2t}] dW_s(t) + dJ_s(t) \quad (2.61)$$

$$dV_{1t} = \alpha_{v1} V_{1t} dt + dW_{v1} \quad (2.62)$$

$$dV_{2t} = \alpha_{v2} V_{2t} dt + [1 + \beta_{v2} V_{2t}] dW_{v2t} \quad (2.63)$$

$\text{corr}(dW_{st}, dW_{v1}) = \rho_1$ and $\text{corr}(dW_{st}, dW_{v2}) = \rho_2$ and dW_{st} , dW_{v1} and dW_{v2} are standard Gaussian random variables. This model was used due to its feedback term in the dV_{2t} function capable of creating larger price movements in contrast to a simple diffusion model, this was used to emulate a highly volatile asset price process, which under those conditions and in discrete time favors the miss-identification of a large price movement as a jump. The parameters for the TFSV model are in line with [Huang and Tauchen \(2005\)](#).

The two previous models are based on the Euler clock set to 1-second increments to create discrete data points, the next approach relies on the exact discrete model procedure developed by [Nowman \(1997\)](#) based on the exact solutions of [Bergstrom \(1984\)](#) (theorem 2) regarding short interest rate process. We start with a stochastic differential equation of the form

$$dV_t = (a + \beta V_t) dt + \sigma V_{t-1}^\gamma dZ + dJ_t \quad (2.64)$$

The above equation can be written in integral form as:

$$V(t) - V(0) = \int_0^t [a + \beta V_s] ds + \sigma(V_{t-1})^\gamma \int_0^t dZ + Y_t \int_0^t dJ_t \quad (2.65)$$

Where the Z is the random measure, Y_t is the jump size and dJ_t is the Poisson. By applying the Gaussian estimation methods we obtain the following expression

$$V_t = e^\beta V_{t-1} + \frac{a}{\beta} (e^\beta - 1) + \varepsilon_t \quad (2.66)$$

The price defined as [Equation 2.58](#) and the volatility process described above, comprise our finalized model.

2.3.1 Appraisal of test performance

Size of the tests under no jumps

Table 2.1 reports the size for the Heston model with no jumps in price or volatility at a 5% significance level in 5 and 20 seconds and 1-minute sampling frequency, assuming noise is absent. For the Heston model we observe that the BNS test attains reasonable performance across all of the sampling frequencies but it tends to be slightly undersized in the higher frequencies. Using the threshold multipower variation in the BNS test we observe an increase in the size in most cases but this result is not consistent. This was proposed in the paper of Corsi et al. (2010) but it does not seem to provide any statistical advantage. The ASJ test is undersized in low frequencies but swiftly becomes oversized as the sampling frequency decreases. Also, it heavily depends on the choice of the threshold parameter α as it determines whether a return is a jump or part of the Brownian increments. We observe that increasing the parameter leads to a lower size but at risk of miss identifying small jumps as noise. The JO test works well but it becomes slowly oversized in lower frequencies as does the LM test as well. MinRV and MedRV tests perform reasonably but they tend to display higher size in low frequencies while we observe a marginally better size performance by the MinRV test. CPR test behaves poorly in low frequencies but reaches a value close to the nominal level in lower frequencies. For the PZ test is important to choose the threshold parameter level as in the ASJ test, following the recommendations of the authors Podolskij and Vetter (2009b) we choose an initial value of $c = 2.3$ and later experiment with $c = 3$ which mitigates the issue of the test at low frequencies, more we do not encourage the use of a value for the threshold greater than 3.5. Furthermore, for the power of the test, we use $P = 2$ and $P = 4$ for the PZ2 and PZ4 accordingly. Finally, the tests designed to accommodate for the noise LM12 and ASJL display size close to the nominal level but they become undersized rapidly.

Moving to the two-factor stochastic volatility model the trend in most cases remains but the size of the majority of the procedures increases as expected since the model simulates a highly volatile asset process. This indicates that our initial hypothesis of a high likelihood for the tests to misidentify noise as jumps are to some extent true, this was after all the paradigm of using this model as we aimed at identifying what jump tests were able to distinguish between Brownian increments and discontinuities in a highly volatile environment. It should be mentioned that we did not alter the parameters of the tests, that is to say, that the power as well as the threshold levels have not been changed from the Heston model for any of the testing procedures. The reason is that firstly we wanted to observe their behavior in the setting proposed by their corresponding authors and secondly we wanted to be able to compare their performance, by calibrating each one it is possible to have obtained more conservative sizes but in that way we feel that we will have lost any

comparative value. The test that displays sizes close to the nominal level and also remains consistent throughout the sampling frequencies are from best to worst the BNS, MED and MIN, and the LM test, while the biggest size distortions have been observed in the lower frequencies from the JO, PZ2, and PZ4, LM12, CPR, and the ASJL. The last model to investigate the performance of the jump tests is the [Nowman \(1997\)](#). The reason we use the model was to document the impact of a different discretization procedure on the tests themselves. While there are some discrepancies in their power of them but most notably in their size, it is difficult from our advantage point to determine to what extent discretization plays a role if any. From a power point of view, most of them seem to increase in power not only in comparison to the TFSV model but also when compared to the Heston model. Following the theme of the previous paragraphs, the best performing models in terms of size are BNS, MIN, and MED, CPR is slightly oversized in the 5 seconds. The JO test has a reasonable size performance in the 20-second sampling and LM which again is oversized in high frequencies.

Table 2.1 Empirical size of the test with no jumps at 5% significance level

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.057	0.038	0.039	0.051	0.072	0.075	0.076	0.11	0.17	0.037	0.058
20s	0.042	0.055	0.063	0.068	0.051	0.048	0.051	0.27	0.28	0.078	0.015
1m	0.045	0.042	0.078	0.086	0.047	0.044	0.048	0.33	0.30	0.012	0.020
Double Heston											
5s	0.069	0.050	0.063	0.058	0.057	0.051	0.071	0.074	0.073	0.030	0.084
20s	0.073	0.055	0.060	0.061	0.052	0.052	0.066	0.13	0.14	0.098	0.009
1m	0.038	0.048	0.072	0.069	0.046	0.048	0.035	0.27	0.31	0.099	0.006
Nowman											
5s	0.039	0.040	0.032	0.076	0.069	0.053	0.085	0.098	0.12	0.049	0.054
20s	0.064	0.059	0.043	0.055	0.053	0.056	0.058	0.28	0.26	0.100	0.013
1m	0.081	0.041	0.069	0.037	0.048	0.055	0.048	0.35	0.38	0.180	0.002

Empirical Size with volatility jumps

To appraise the test behavior in an empirical relevant scenario we incorporate into the stochastic models volatility jumps with varying intensities and jump sizes. To examine the variation of the tests we consider a compound Poisson type jumps from low infrequent jumps ($\lambda = 0.2$ $\sigma^2 = 0.09$) up to large and frequent ones ($\lambda = 5$ $\sigma^2 = 0.0025$) for the volatility process. In terms of size the best performing tests as shown in [Table 2.2](#) are the BNS, MIN and MED, ASJ, and the JO being slightly oversized across all frequencies. The surprise to this point was the ASJ test achieved a size close to the nominal level while reducing its size in low frequencies. CPR is oversized in high frequencies, but the

PZ2, PZ4, LM12, and ASJL tests all are greatly oversized. The trend our exercise have surfaced is that with the increase in the jump size and the frequency the size of the test is increasing as does their power, this phenomenon indicates that the testing procedures are able to distinguish between jumps and a continuous path when the discontinuities are large and frequent. Table 2.2 shows little differences concerning the no jump case which is to be expected when considering that with the parameters we have set a jump occur approximately every 20 days with a relatively small size. As it has been documented in the work of Huang and Tauchen (2005) since there is a positive and significant probability that there are no jumps in a finite time interval when the test is applied on a day-to-day basis there not able to consistently discriminate between a continuous and a discontinuous process.

Table 2.2 Empirical size of the tests with rare volatility jumps ($\lambda = 0.2$)

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.040	0.034	0.033	0.056	0.062	0.078	0.052	0.110	0.170	0.030	0.048
20s	0.049	0.055	0.041	0.072	0.050	0.068	0.044	0.260	0.280	0.039	0.013
1m	0.042	0.047	0.052	0.070	0.058	0.035	0.032	0.300	0.280	0.010	0.035
Double Heston											
5s	0.065	0.049	0.065	0.062	0.059	0.052	0.080	0.074	0.073	0.032	0.077
20s	0.073	0.052	0.066	0.065	0.072	0.064	0.068	0.130	0.140	0.081	0.085
1m	0.033	0.049	0.081	0.069	0.063	0.077	0.044	0.250	0.300	0.096	0.009
Nowman											
5s	0.035	0.041	0.032	0.063	0.076	0.064	0.078	0.099	0.120	0.054	0.065
20s	0.052	0.060	0.038	0.055	0.043	0.052	0.046	0.320	0.300	0.120	0.010
1m	0.072	0.046	0.057	0.047	0.052	0.055	0.045	0.410	0.320	0.220	0.0027

Table 2.3 reports empirical rejection frequencies for frequent and large volatility jumps ($\lambda = 5$) at 5% significance level. In line with the work of Kolokolov and Renò (2024), the discussion focuses on size control, robustness across sampling frequencies, and the interaction between jump intensity and discretization.

A first salient feature of the table is that several tests that performed adequately under rare-jump scenarios continue to exhibit satisfactory size control when volatility jumps occur frequently. In particular, ASJ, BNS, JO, MED, MIN, and CPR deliver empirical sizes that remain close to the nominal level across all models and sampling frequencies. Rejection rates for these tests are generally contained within the 3–7% range, with only mild over-rejection for JO and LM at the one-minute frequency under the Double Heston specification. Importantly, the stability of these tests across the Heston, Double Heston, and Nowman models indicates that frequent volatility jumps per se do not necessarily undermine size control, provided the test construction appropriately accounts for high jump activity.

The LM test again shows a modest tendency toward over-rejection, with empirical sizes often in the 6–9% range, particularly under the Double Heston model at the highest sampling frequency. While this distortion is limited relative to some alternatives, it suggests that the asymptotic approximations underlying LM may be less accurate when jump arrivals are dense, even if jump magnitudes remain small.

A distinct pattern emerges for the PZ-type tests. Both PZ2 and PZ4 exhibit persistent and non-negligible over-rejection across all DGPs, with empirical sizes ranging from roughly 10% to 22%. Unlike the rare-jump case, however, the dependence on sampling frequency is weaker and less monotonic. This attenuation suggests that when volatility jumps are frequent, the dominant source of size distortion for PZ-type statistics is not discretization alone, but rather a structural sensitivity to high jump intensity that is not adequately filtered under the null.

The behavior of LM12 is particularly noteworthy. While LM12 performs reasonably well under the Heston and Double Heston specifications, it breaks down dramatically under the Nowman model at 20-second and one-minute sampling frequencies, with rejection rates exceeding 50%. This sharp deterioration stands in contrast to its otherwise acceptable performance and points to a strong interaction between model-induced persistence in volatility dynamics and the test’s finite-sample properties. Such sensitivity has also been showcased in [Kolokolov and Renò \(2024\)](#) regarding procedures that conflate genuine jumps with features induced by temporal aggregation or latent staleness-like effects.

Finally, ASJL maintains comparatively good size control across most configurations, with rejection frequencies generally close to 5% and only mild deviations under the Double Heston model. Relative to the rare-jump scenario, ASJL appears to benefit from higher jump intensity, suggesting that its finite-sample behavior improves when the underlying jump process is more active.

Taken together, the results indicate that frequent volatility jumps do not uniformly worsen size distortions across tests. Instead, the impact is highly test-specific. Procedures designed to be robust to jump contamination such as ASJ, BNS, MED, MIN, and CPR, retain reliable size properties even in high-intensity jump environments, whereas PZ-type tests and, in some cases, LM12 remain prone to substantial over-rejection. Consistent with the message of [Kolokolov and Renò \(2024\)](#), these findings underscore that size distortions in jump tests are often driven less by the presence of jumps per se than by how discretizations, persistence, and jump filtering interact in finite samples.

So far, the analysis has focused on the impact of jump frequency, in the tables below, out attention shifts to explore how the size and frequency can impact the ability of the procedures to disentangle the jump part from the Brownian increment. A growing body of empirical research highlights that even infrequent and small movements in the volatility process can carry meaningful information. These subtle volatility jumps, though less

Table 2.3 Empirical size of the tests with frequent volatility jumps ($\lambda = 5$)

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.032	0.048	0.037	0.063	0.058	0.058	0.066	0.140	0.190	0.039	0.041
20s	0.045	0.050	0.051	0.077	0.052	0.071	0.044	0.160	0.100	0.048	0.063
1m	0.040	0.053	0.054	0.063	0.061	0.059	0.061	0.220	0.190	0.025	0.033
Double Heston											
5s	0.056	0.045	0.055	0.089	0.031	0.049	0.053	0.220	0.180	0.084	0.067
20s	0.043	0.054	0.061	0.063	0.037	0.035	0.048	0.200	0.120	0.076	0.052
1m	0.044	0.059	0.088	0.068	0.063	0.072	0.069	0.150	0.100	0.097	0.046
Nowman											
5s	0.042	0.037	0.052	0.065	0.046	0.055	0.064	0.130	0.120	0.063	0.059
20s	0.049	0.041	0.061	0.065	0.047	0.054	0.056	0.099	0.087	0.670	0.048
1m	0.037	0.059	0.072	0.063	0.038	0.044	0.059	0.110	0.089	0.560	0.028

conspicuous, play a crucial role in shaping high-frequency dynamics and have been documented in the literature following the insights of [Jacod and Todorov \(2010\)](#). For that reason in this section we expose the tests to rare and frequent volatility Poisson distributed jumps. In [Table 2.4](#) (small volatility jumps) it can be seen that the tests display values close to the no price jump case, the best performing is the BNS, MED, ASJ, and JO indicating that in terms of size the trend continuous, CPR becomes undersized as the frequency increases. The PZ family of tests (PZ2 and PZ4) are greatly oversized in all frequencies while the tests designed to accommodate the noise are severely undersized. Moving to the Double Heston model the story changes, the BNS test seems to be the most consistent and close to the nominal level. The ASJ test is oversized and becomes quickly undersized in low frequencies, similar to the CPR test. The MED and MIN test increase in size as the frequency increases along with the LM test. In the PZ family, there is an improvement in the 5-second frequency which quickly diminishes. The LM12 test becomes oversized rapidly while the opposite course follows the ASJ test. The results are not confounding since the model (Double Heston) was used to simulate a more volatile market as described in [section 2.3](#) (Monte Carlo Simulation), accompanied by the rare volatility jumps that can make it difficult for the test to distinguish between Brownian variations and jump components. In the last model (Nowman) the results seem to be slightly better in terms of the empirical size, the BNS test holds its consistency followed by the MIN test. The ASJ test is undersized and becomes oversized as the frequency increases, the JO test provides a size close to the nominal level only in the 20-second sampling frequency as does the LM test as well. The MED test becomes undersized with the increase in frequency, while the PZ family of tests is grossly oversized in all frequencies. On the one hand, the LM12 test displays a good size in the lowest frequency but quickly becomes oversized and on the other, the ASJL test follows exactly the opposite course. Finally, the CPR test is oversized

in the 5 seconds, but as the frequency increases its size is getting closer to the nominal level. From this table it is clear that the tests perform the best in the Heston model, in the rest (Double Heston and Nowman) their size are greater in nearly all tests and all frequencies, indicating that most of the procedures are facing a challenge to distinguish between the continuous and the jump part in volatile circumstances. Furthermore, the BNS test is the most consistent in terms of the empirical size in all of the simulation exercises so far.

Table 2.4 Empirical size for rare and small jumps in volatility at 5% significance

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.045	0.035	0.058	0.076	0.055	0.054	0.077	0.190	0.180	0.110	0.180
20s	0.031	0.050	0.062	0.071	0.050	0.059	0.048	0.270	0.260	0.150	0.098
1m	0.040	0.051	0.081	0.064	0.062	0.071	0.061	0.470	0.470	0.210	0.170
Double Heston											
5s	0.051	0.042	0.054	0.081	0.036	0.048	0.052	0.210	0.190	0.100	0.170
20s	0.046	0.055	0.066	0.069	0.031	0.039	0.043	0.200	0.180	0.088	0.100
1m	0.033	0.044	0.071	0.063	0.057	0.055	0.048	0.230	0.170	0.120	0.220
Nowman											
5s	0.047	0.039	0.057	0.068	0.048	0.052	0.063	0.200	0.190	0.097	0.120
20s	0.042	0.047	0.062	0.065	0.046	0.050	0.054	0.250	0.230	0.130	0.099
1m	0.039	0.051	0.081	0.072	0.037	0.044	0.056	0.460	0.460	0.190	0.130

By implementing frequent volatility jumps the tests seem to perform better under all simulation models, in the Heston model the best performing test in terms of size is the BNS, ASJ although it is slightly undersized in the 5-second sampling and the JO with the same pattern. The worst remains the PZ family which is quite oversized in all frequencies. In the Double Heston model in the top category, we have the addition of the CPR test and in the last simulation model, CPR holds its good performance and the MIN tests also display size close to the nominal level. The empirical size of the tests is close to the no jump in price case indicating that volatility jumps alone are difficult to be detected even when significantly increasing their jump arrival frequency (λ), the only substantial difference able to observe where only in the case the jump size was increased considerably.

Power under price jumps

Switching to power evaluation in the case of rare and small jumps (Table 2.6) the best performing are the PZ tests with the highest power score achieved in the 5-second frequency, the CPR, JO, and BNS tests are very close, and maintain a high score in the 1 minute as well. Next, comes the LM and LM12 test but the latter is sensitive to the sampling frequency reducing to an almost 80% power score in 1 minute from 90% score in 5 seconds. Second, from last are the MED and MIN tests maintaining a respectable power

2.3 Monte Carlo Simulations

Table 2.5 Empirical size for large and frequent jumps in volatility at 5% significance level

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.052	0.040	0.063	0.075	0.059	0.054	0.081	0.200	0.190	0.110	0.130
20s	0.052	0.052	0.061	0.071	0.056	0.048	0.056	0.260	0.240	0.130	0.100
1m	0.043	0.069	0.058	0.083	0.042	0.044	0.066	0.430	0.430	0.210	0.160
Double Heston											
5s	0.058	0.052	0.063	0.085	0.045	0.051	0.067	0.210	0.200	0.140	0.170
20s	0.051	0.061	0.067	0.078	0.049	0.054	0.058	0.270	0.280	0.095	0.099
1m	0.049	0.052	0.073	0.076	0.060	0.063	0.059	0.240	0.220	0.120	0.230
Nowman											
5s	0.055	0.038	0.061	0.085	0.041	0.051	0.082	0.190	0.180	0.140	0.170
20s	0.043	0.047	0.064	0.073	0.051	0.055	0.055	0.260	0.240	0.087	0.091
1m	0.047	0.048	0.068	0.071	0.062	0.058	0.046	0.450	0.450	0.096	0.230

score ranging from high 80% in 5 seconds to low 90% in the 1 minute. In the same space is also the ASJL test but only in the high frequency as it suffers from the same thing as the LM12 test. Last, finishes the ASJ test maintaining the power of around 60%. While there are no significant variations among the three distinct simulation models, it is noteworthy that the TFSV model has decreased the power in most of the tests, other than that the tests appear to be consistent.

Table 2.6 Empirical power for rare and small jumps in price at 5% significance

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.684	0.999	0.998	0.978	0.863	0.892	0.999	1.000	1.000	0.901	0.885
20s	0.679	0.989	0.992	0.978	0.880	0.903	0.989	0.999	0.999	0.821	0.776
1m	0.643	0.981	0.985	0.971	0.855	0.988	0.921	0.963	0.965	0.796	0.749
Double Heston											
5s	0.660	0.999	0.992	0.968	0.883	0.886	0.992	0.998	0.999	0.891	0.898
20s	0.652	0.971	0.990	0.964	0.877	0.893	0.992	0.999	0.999	0.887	0.888
1m	0.638	0.957	0.986	0.961	0.849	0.878	0.955	0.985	0.982	0.799	0.846
Nowman											
5s	0.789	0.999	1.000	0.998	0.988	0.994	0.999	0.999	0.999	0.928	0.915
20s	0.775	0.997	0.998	0.946	0.887	0.886	0.997	0.999	0.999	0.899	0.875
1m	0.699	0.995	0.991	0.935	0.892	0.888	0.994	0.995	0.996	0.854	0.867

The data are not so far contaminated with additive noise, only with small and rare price jumps, this is about to change, in [Table 2.7](#) we report the results for frequent large price jumps and as in the case of size, the power as well increases. In this scenario, while the jumps are easier to detect (due to their size) the high frequency of appearances may confuse some of the tests, for reference the frequency is 25 times higher than in the previous case. The PZ2, PZ4, CPR, JO, and LM tests achieving a power ranging from 97% to 99% with

the lowest being the LM test, and in the highest end are the PZ2, PZ4, and CPR tests in 1-second sampling and decreasing to a range between 90% and 96% in the 1-minute frequency for the Heston model. It is worth noting that in the case of the PZ test it is possible to increase its power by altering the threshold parameter (increasing it) which will lead to a higher jump detection rate but with the trade of larger size distortions. This applies also to the CPR test but it is less sensitive to the choice of the threshold, this is in line with what [Theodosiou and Zikes \(2011\)](#) had documented in their work for the two-factor stochastic volatility model.

Table 2.7 Empirical power for large and frequent jumps in price at 5% significance level

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.756	0.999	0.999	0.982	0.965	0.958	1.000	1.000	1.000	0.943	0.904
20s	0.869	0.985	0.997	0.975	0.984	0.981	0.999	1.000	1.000	0.977	0.923
1m	0.748	0.981	0.988	0.973	0.981	0.976	0.989	0.999	0.998	0.948	0.912
Double Heston											
5s	0.771	0.988	0.999	0.975	0.966	0.961	0.999	0.999	0.999	0.925	0.985
20s	0.816	0.999	0.984	0.979	0.979	0.982	0.997	0.999	0.999	0.957	0.976
1m	0.768	0.981	0.980	0.974	0.972	0.982	0.991	0.995	0.994	0.942	0.966
Nowman											
5s	0.799	0.999	0.999	0.992	0.978	0.975	0.999	1.000	1.000	0.964	0.972
20s	0.786	0.999	0.997	0.989	0.969	0.908	0.999	0.999	0.999	0.959	0.951
1m	0.782	0.986	0.988	0.981	0.964	0.888	0.997	0.999	0.999	0.945	0.864

Noise

In our simulation exercise so far we have employed a simplified theoretical model for the evolution of the price process by hypothesizing that it is generated by the jump-diffusion model. Unfortunately in the real world financial markets do not always assume a random walk instead, it deviates (sometimes a lot) creating some so-called inefficiencies. Traders and market practitioners in general, create strategies to exploit those anomalies forcing the market back to its original walk. Despite that, there has been extensive theoretical as well as empirical work to explain those anomalies most of the literature and especially the one regarding jump identification tends to regard them as noise. In this section, we consider for all the models (Heston, Two-factor stochastic volatility, and Nowman) the existence of three different distributions of noise, Gaussian, Student, and a mixture of both. [Table 2.8](#) presents the power of the tests at a 5% significance level for the Heston model with no jumps in price or volatility, the hierarchy remains somewhat similar to the previous instances without the noise, with the PZ test displaying the best power followed closely by the ASJL, CPR, MED, JO, and BNS test. The LM12 test has high power in 5 seconds

which drops quickly when the frequency reduces to 1 minute. The MIN test has a hard time with a power lower than his brother (MED) and it tends to work best in the 20-second region. The ASJ test seems to be lacking in terms of power and it works best in the high frequency of 5 seconds. As a general rule across all of the testing procedures is that the ones that had incorporated in their design the existence of noise and were made to remain robust under its influence, seem to have an edge. Finally, between the three different types of noise we observe little differences in terms of power, the student-type noise appears to reduce the power of the test slightly but increasing the frequency to less than 5 seconds alleviates the issue.

Table 2.8 Power of the tests at 5% significance level under three different distributions of noise for the Heston 1993 model

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Gaussian											
5s	0.519	0.996	1.000	0.869	1.000	0.857	0.999	1.000	1.000	1.000	0.999
20s	0.496	0.999	1.000	0.752	1.000	0.992	0.999	1.000	1.000	0.998	0.888
1m	0.511	0.997	1.000	0.721	0.997	0.972	0.999	1.000	1.000	0.918	0.512
Student											
5s	0.554	0.999	0.997	0.999	0.820	0.640	1.000	1.000	1.000	0.998	0.999
20s	0.400	1.000	0.995	0.890	0.917	0.858	0.999	1.000	1.000	0.989	0.784
1m	0.580	0.999	0.999	0.887	0.965	0.980	1.000	1.000	1.000	0.930	0.413
Mixture											
5s	0.540	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.994	0.999
20s	0.504	0.996	1.000	0.898	1.000	1.000	1.000	0.999	1.000	0.998	0.800
1m	0.518	1.000	1.000	0.865	1.000	1.000	0.998	1.000	1.000	0.906	0.819

Simultaneous price and volatility jumps with additive noise

Up to this point we try to simulate price and volatility data with noise or price jumps or volatility jumps of different sizes and frequencies, nonetheless, our stochastic models remain a simplified approach to the empirical data, and thus it is not possible to determine the realistic behavior of the tests from what we have observed so far alone. To push the limits of their limiting theories and to truly put their name to the test it is apparent that we need to have a model that incorporates volatility jumps, price jumps and noise simultaneously, under this conditions the tests are thought to display similar properties in terms of size and power as in an empirical set of data. Table 2.8 shows the power of the tests under the Heston model with small volatility jumps, large price jumps, and noise on the left side and large price and volatility jump as well as noise on the right side. The noise is assumed to be Gaussian and similarly to the beginning of this section, the jumps assume a compound Poisson process with the parameter $\lambda = 0.5$ for both circumstances. The jump size for the first case (small volatility and large price jumps) for the volatility

Table 2.9 Power of the tests at 5% significance level with jumps in price and volatility and Gaussian noise under the Heston (1993) model

Large Price and small Volatility jumps				Large Price and large Volatility jumps			
	5s	20s	1m	5s	20s	1m	
ASJ	0.445	0.387	0.111	0.697	0.625	0.299	
BNS	0.642	0.511	0.473	0.999	0.893	0.658	
JO	0.289	0.215	0.176	0.885	0.822	0.713	
LM	0.989	0.916	0.847	1	0.987	0.846	
MED	0.719	0.618	0.56	0.984	0.945	0.855	
MIN	0.644	0.535	0.444	0.997	0.998	0.889	
CPR	0.778	0.741	0.657	1	0.999	0.914	
PZ2	0.985	0.878	0.786	0.999	0.998	0.903	
PZ4	0.988	0.844	0.775	1	0.999	0.892	
LM12	0.638	0.542	0.489	0.755	0.613	0.456	
ASJL	0.37	0.369	0.285	0.868	0.767	0.557	

is normally distributed with $N(0, 1.5\%)$ and for the price $N(0, 5\%)$. In the second case, the volatility jumps are $N(0, 5\%)$ similarly to the price jumps, the noise is Gaussian with $\sigma = 0.003$. The table below shows what we have described in a laconic form.

As can be seen the tests face their greeter challenge so far validating our initial conjecture, in the left-hand side all of the tests except for the PZ have not reached a power close to 90% in the 5-second sampling frequency. Behind it comes the CPR test with the MED test close by, the rest of them fall behind with the BNS and LM12 leading. The least power score is recorded in the JO and ASJ test. As the frequency decreases so does the power as in the previous cases but at this instant, the rate is considerably greater reaching the 1-minute frequency none of the tests are able to touch even the 80% mark. In terms of size the JO, CPR, and PZ tests are considerably oversized ranging from 25% (JO) up to 58% (PZ) while the BNS and ASJ test at 5-second sampling presents a size of zero but reach a value close to but smaller than the nominal level at the 1-minute frequency. On the right-hand side in terms of power the test performs better, leading are the CPR, PZ and LM followed by the BNS, MED, and MIN at the 5 seconds, moving toward the 1 minute the ranking remains the same but the power of the tests quickly drops. To synopsise under this scenario all of the testing procedures face a challenge when trying to discriminate between a continuous and a jump part, likewise, the jump tests were more vulnerable in the case of small volatility and large price jumps.

Zero returns

With the recent advancements in computer processing along with the ever higher transmission networks, high-frequency financial data are becoming more and more available, we have reached the point where it is possible to obtain all the transactions that took place in a given time frame for a financial instrument. By observing a set of such data, it is noticeable that the price does not change at specific time intervals of equal length. It is typical for financial instruments and especially for stocks that the time between price updates is not constant but a variable, this is due to investment behavior, for example, the market tends to have a higher activity during the open and closing hours and also during the noon. Furthermore, events for example announcements from the state or a company may trigger a high activity in a relatively low volume instrument. Since the modern portfolio theory suggest that the available information of a market is been fused to the next price update is of paramount importance to model the time interval between successive observation of price change. In that field, several papers attempt to predict this time interval by considering it as a stochastic variable (see the work of [Engle and Russell \(1998\)](#) and [Bauwens and Veredas \(2004\)](#)), but except for the work of [Theodosiou and Zikes \(2011\)](#) the effects of irregular time intervals has been incorporated neither in the construction (or robustification for that matter) nor in the testing of the jump test. In this section we attempt to observe the behavior of the testing procedures by considering that time between successive prices are not constant, presume that X_i is the length of time between a cumulative price update from the time t_{i-1} to t_i , then the duration is formulated as $X_i = t_i - t_{i-1}$. From the work of [Engle and Russell \(1998\)](#), it is documented that financial data tend to exhibit serial dependence, we can rewrite the equation for the duration as the product of a latent variable (accounting for the serial correlation) and an i.i.d random variable (error term) as

$$X_i = \Psi_i \varepsilon_i \quad (2.67)$$

where $\psi_i = E[X_i | \Omega_{i-1}]$ and Ω_{i-1} encompasses all information available at time t_{i-1} . Equation 69 shows how the latent variable is computed.

$$\Psi_i = e^{\psi_i} \quad (2.68)$$

where for $|\beta| < 1$

$$\psi_i = \omega + \beta \psi_{i-1} + u_i \quad (2.69)$$

[Table 2.11](#) summarizes the results for all the stochastic models with the duration between successive prices driven by a diffusion process without the existence of jumps in price or volatility. We expect that the test based on multipower variation will suffer in contrast to all others due to the inherent limitation of the bipower variation displaying

Table 2.10 Infrequent trading for all the stochastic models with no jumps at 5% significant level

	ASJ	BNS	JO	LM	MED	MIN	CPR	PZ2	PZ4	LM12	ASJL
Heston											
5s	0.056	0.055	0.058	0.062	0.045	0.041	0.096	0.110	0.099	0.088	0.140
20s	0.052	0.050	0.057	0.066	0.051	0.045	0.140	0.180	0.170	0.130	0.104
1m	0.039	0.048	0.067	0.041	0.048	0.051	0.056	0.150	0.160	0.210	0.290
Double Heston											
5s	0.069	0.065	0.070	0.073	0.058	0.055	0.120	0.230	0.230	0.150	0.170
20s	0.072	0.060	0.073	0.085	0.063	0.063	0.220	0.250	0.240	0.110	0.130
1m	0.061	0.075	0.082	0.079	0.069	0.072	0.310	0.380	0.420	0.180	0.260
Nowman											
5s	0.060	0.062	0.066	0.069	0.054	0.051	0.108	0.210	0.180	0.140	0.100
20s	0.067	0.054	0.069	0.077	0.056	0.058	0.206	0.230	0.210	0.110	0.101
1m	0.058	0.065	0.071	0.074	0.065	0.066	0.230	0.360	0.390	0.160	0.240

two successive sums equal to zero if a zero is present. Considering that in the case of the realized variance only one sum will be equal to zero, any test that relies on the difference between the two will be oversized. The simulation exercise confirms our intuition regarding

Table 2.11 Summary of Simulation Variables

Small Volatility and Large Price jumps		Large Price and large Volatility jumps	
Volatility	$N(0, 1.5\%)$	Volatility	$N(0, 5\%)$
Price	$N(0, 5\%)$	Price	$N(0, 5\%)$
Noise	$\sigma = 0.03$	Noise	$\sigma = 0.03$

the BNS and CPR tests, both are oversized in all frequencies, and the MED and MIN test provides the best performance in terms of size. Also, the JO and ASJ test while affected and slightly oversized, display good behavior, and finally the PZ tests while oversized (as in the previous occasions) appears that they do not have been seriously affected by the infrequent trading.

2.4 Empirical Evidence

In this section we apply the jump detection procedures to empirical data to observe whether or not our results from the simulation exercise carry through. This enables us to answer two important questions, firstly are the tests capable of detecting discontinuities, not in a controlled environment but in an empirical set of data? Secondly, it provides information regarding the effectiveness of the simulation models in capturing the dynamics of the different markets. The data are collected from three markets, the stock, foreign

exchange, and cryptocurrency markets. The data were selected in a manner that ensures a representative sample of all markets and financial instruments.

2.4.1 Data Description

Individual Stocks

The equity data were collected from the FirstRate Data in 1-minute frequency, we consider the stocks of S&P500 and Amazon from 1/04/2022 to 14/04/2022 for both, containing 3955 and 4836 observations respectively.

Foreign Exchange

For the Foreign Exchange market we choose the EURO-USD and EUR-GBP exchange, the data are collected in 1-minute frequency from 1/04/2022 to 15/04/2022 aggregating to 25518 observations each.

Cryptocurrencies

For the digital market we collect data from the CRYPTO database which is a platform containing all the crypto exchanges worldwide and provides high-frequency data. The data are from 30/08/2022 to 6/09/2022 containing 10000 observations. The simulation exercises in [section 2.3](#) revealed that the testing procedures work the best in the 20-second and 1-minute frequencies regarding their power score. The highest frequency of available data where at 1-minute intervals, in this frequency the power is high but the effects of microstructure noise are negligible.

Table 2.12 Descriptive Statistics

Category	Asset	Mean	Max	Min	S.Dev.	Skewness	Kurtosis
Individual Equities							
	Amazon	28.64	30.65	27.00	0.83	0.46	2.14
	S&P500	4481.40	4592.10	4382.50	53.29	0.27	1.80
Foreign Exchange							
	EUR/USD	1.00	1.10	1.00	0.0063	0.84	3.11
	EUR/GBP	0.83	0.84	0.82	0.0030	0.64	3.61
Crypto-Currencies							
	BTC	19954.60	20502.60	19573.30	199.52	0.72	2.54
	ETH	1572.10	1674.50	1476.00	28.00	0.73	4.94

In [Table 2.12](#) the basic statistics of the data set are provided, from first glance it is apparent (as expected from the financial theory) that the foreign exchange market is the

one closest to the normal distribution in terms of skewness and kurtosis while also have the smallest standard deviation. In the opposite spectrum lay equity and the crypto market. We can see that the S&P500 index and the BTC (Bitcoin) display the highest dispersion and are extremely volatile, while ETH (Ethereum) displays moderate tails and is skewed to the left. As a compliment to the aforementioned table we provide in table 11 the number of zeros contained in each data set and their percentage, the simulation exercise in the previous section revealed that there is a correlation between the number of zeros (infrequent trading) and the number of jumps detected by the testing procedures. We aim at observing whether or not this result carries in the empirical data.

Table 2.13 Number of zeros in the data-set

Data	1m	5m	30m
Amazon	92 (0.02)	42 (0.037)	7 (0.028)
S&P500	28 (0.007)	7 (0.008)	0 (0)
EUR\USD	12262 (0.48)	51 (0.018)	4 (0.008)
EUR\GBP	12491 (0.48)	100 (0.035)	11 (0.023)
BTC	670 (0.067)	6 (0.003)	0 (0)
ETH	1888 (0.18)	16 (0.008)	0 (0)

The highest percentage of zeros is seen in the Foreign Exchange followed by the Cryptocurrency market, similarly to section 3.1.5 where we discuss the effects of infrequent trading in the detection of jumps, the test we expect to be oversized are the ones depending on the multipower variation measures. The BNS and CPR are expected to be oversized while in the ASJ and JO test the effects should be smaller. The PZ test can be oversized and affected by the zeros but it heavily depends on the choice of the c parameter. With the reduction in frequency, the number of zeros also reduces and some assets in the 30-minute frequency have no zeros. Finally, by extending the results of the simulation exercises we expect the size of the tests to get smaller as one move to lower frequencies.

2.4.2 Empirical Findings

In this section we apply the jump detecting procedures in a set of empirical data retrieved from three distinct markets. For the values of the threshold parameters for the CPR, ASJ, and PZ test we use the same values as in the simulation exercise (small and medium) following the recommendations of the authors. We choose to sample at 1, 5, and 30 minutes with a significance level of 5%, the results are summarised in tables 12. As expected the BNS, LM, CPR, and PZ tests are generally oversized with the highest proportion of identified jumps, while the ASJL, LM12, ASJ, and JO display lower values. By moving to lower frequencies the number of jumps identified by the tests is decreasing but in a nonlinear way. From 1 to 5 minutes we observe an average decrease of 8% in jump identification while from 5 to 30 minutes of sampling, on average we observe a decrease of 17%. This phenomenon indicates that there is an accelerated decrease in jump identification up to the 10-minute sampling frequency, from that point onward the microstructure noise gets negligible and the number of zeros in the series is considerably less, as a result, we advocate that those two statistics seem to play a crucial role in the performance of the jump tests, this observation is in line to what [Theodosiou and Zikes \(2011\)](#) have documented. We differentiate from the literature in that we have applied the infrequent trading to multiple stochastic models to gain a better perspective and observe which of the three stochastic models used in the simulation is closer to the empirical data at least in the case of infrequent trading.

Table 2.14 Empirical Results

Category	Procedure	1 minute	5 minutes	30 minutes
Amazon	ASJ	0.254	0.153	0.075
	BNS	0.376	0.253	0.094
	JO	0.223	0.237	0.062
	LM	0.394	0.269	0.094
	MED	0.343	0.182	0.066
	MIN	0.331	0.178	0.064
	CPR	0.389	0.235	0.097
	PZ2	0.399	0.196	0.099
	PZ4	0.396	0.214	0.110
	LM12	0.215	0.126	0.078
	ASJL	0.172	0.110	0.083
	ASJ	0.187	0.124	0.074

Continued on next page

Category	Procedure	1 minute	5 minutes	30 minutes
	BNS	0.321	0.223	0.085
	JO	0.164	0.112	0.053
	LM	0.333	0.235	0.089
	MED	0.256	0.156	0.058
	MIN	0.245	0.143	0.052
	CPR	0.352	0.200	0.092
	PZ2	0.289	0.172	0.087
	PZ4	0.294	0.169	0.075
	LM12	0.121	0.093	0.048
	ASJL	0.098	0.082	0.051
EUR/USD	ASJ	0.448	0.325	0.234
	BNS	0.640	0.472	0.353
	JO	0.352	0.283	0.216
	LM	0.652	0.451	0.324
	MED	0.556	0.420	0.322
	MIN	0.567	0.380	0.233
	CPR	0.651	0.398	0.286
	PZ2	0.643	0.327	0.257
	PZ4	0.638	0.383	0.255
	LM12	0.527	0.334	0.199
	ASJL	0.581	0.432	0.200
EUR/GBP	ASJ	0.487	0.330	0.210
	BNS	0.641	0.554	0.353
	JO	0.372	0.338	0.218
	LM	0.651	0.556	0.378
	MED	0.623	0.532	0.326
	MIN	0.617	0.510	0.311
	CPR	0.605	0.544	0.387
	PZ2	0.653	0.537	0.415
	PZ4	0.648	0.543	0.389
	LM12	0.572	0.439	0.202
	ASJL	0.597	0.486	0.300
	ASJ	0.724	0.642	0.588
	BNS	0.843	0.689	0.624

Continued on next page

Category	Procedure	1 minute	5 minutes	30 minutes
	JO	0.703	0.601	0.593
	LM	0.951	0.800	0.712
	MED	0.760	0.699	0.582
	MIN	0.772	0.687	0.576
	CPR	0.982	0.723	0.602
	PZ2	0.976	0.718	0.638
	PZ4	0.954	0.762	0.633
	LM12	0.732	0.654	0.554
	ASJL	0.727	0.637	0.561
ETH	ASJ	0.732	0.636	0.593
	BNS	0.846	0.698	0.646
	JO	0.721	0.618	0.610
	LM	0.963	0.836	0.724
	MED	0.774	0.715	0.597
	MIN	0.778	0.710	0.583
	CPR	0.999	0.736	0.616
	PZ2	0.983	0.722	0.639
	PZ4	0.976	0.755	0.642
	LM12	0.821	0.666	0.589
	ASJL	0.845	0.643	0.567

The highest portion of jumps in a day was identified in the digital market and especially in the ETH (Ethereum) which displayed 99.9% of days as having jumps (CPR test). The size swiftly drops to 61.6% in 30 minutes. The results indicate that the highest number of jumps appear in the digital market ranging from 56% up to the 99.9%, the Forex market comes second ranging from 20% up to 60% and the stock market has the least number of jump identification ranging from 10% up to 35%. The results can be explained with the descriptive statistics provided in [Table 2.12](#), as long as there is a sufficient number of zeros in the series and the frequency is high enough the test will be in the majority oversized. The way the data set was collected created a possibility to further investigate the impact of zeros in the empirical data, the BTC (Bitcoin), ETH (Ethereum), and the S&P500 all have no zeros in the 30-minute sampling frequency. With such a low frequency and without zeros we should observe a size closer to the simulations, in fact, the S&P500 displays in the 30-minute frequency a 7% jump detection which is close to the size in the simulations. But contrary in the digital market the proportion of days having jumps remains, relatively to the simulations, high.

2.5 Conclusions

The contribution of our chapter is threefold, firstly we extend the existing literature by robust and meticulous analysis and comparison of all the available jump tests under different distributions of noise and a variety of jump sizes and jump arrival times in both the volatility and in the price process. We also enhance in the simulations the use of three alternative stochastic models and two discretization procedures. Secondly, we study the effects of irregular trading for all of the jump testing procedures. The final contribution of our work is the application of the jump tests to a set of empirical data from three distinct markets, the stock, Forex, and cryptocurrency providing a broad spectrum for their behavior. The findings of the paper indicate that there is not a single test that performs optimally under all conditions and markets, we advise the use of multiple jump tests in different sampling frequencies to determine the relatively better performance. The incorporation of either additive noise and/or zeros (due to the infrequent trading) has a significant negative impact on the size and power of the test. We observe that the LM12 and ASJL tests display a higher size when compared not only to the nominal level but also to the case of frequent volatility jumps. This indicates that those are becoming severely oversized and are not able to distinguish between Brownian and jump parts. In addition, the ASJ, CPR, and PZ tests are sensitive to the choice of the threshold parameter, in our work we mostly use the values recommended by the authors leading to considerable size distortions, nonetheless since a study for the optimal threshold choice for either of the aforementioned test is yet to exist we recommend the implementation of different values to obtain the best result. For the PZ test alone we use two variations of it the PZ2 and PZ4, their difference lies in the choice of the threshold parameter, for the first one we use the recommended by the author's value of 2, and for the second we use a value of 4. We conclude that in both the simulation and in empirical data the PZ4 works best in the highest frequency while the PZ2 gains ground as the frequency decreases. In addition, we have observed that the tests tend to converge in the lower frequencies. From the simulations and the empirical data, the BNS test seems to provide a moderate size and power in most conditions and sampling frequencies, the opposite stands for the ASJ test which was the least powerful. Furthermore, we detect that all of the tests are also sensitive to the choice of the sampling frequency, if it is too high the data are contaminated to a great extent with microstructure noise and a lot of zeros if the frequency is too low then most of the information is lost, both of which header the performance of the tests, from what we discover so far the optimal frequency for high-frequency data was 20 seconds while for lower frequency data the 5 minutes. In the simulation exercise, we use multiple stochastic models and two discretization procedures, the second does not seem to have an impact on the test's behavior, but the stochastic model in use is quite the opposite. The Heston model is the least volatile while the Double Heston

is the most volatile with the Nowman model in the middle. This was also reflected in the size of the tests, for the Double Heston the tests displays the highest size, and in the Heston model the lowest size. The empirical size of the tests is considerably larger when compared to the simulations in the 1-minute frequency which is common for both the simulation and the empirical application, this leads to two assumptions, first either our data have a greater amount of zeros and noise (which is probable) or the simulation models have failed to capture the full spectrum of the asset dynamics. The procedures while they vary when compared to one another, their behavior remains consistent. An interesting result is that when no jumps or noise are present in the time series then the disagreement of the procedures is great but when we introduce jumps, noise and zeros then the tests tend to converge. When we introduce jumps in the volatility process alone all of the tests are not able to capture them even in the case of considerably increasing the jump frequency. The jump size rather than the frequency appears to be significant in terms of the volatility jumps. In the simulations we allow the jump size to follow the normal distribution in some cases while in others has a fixed value, from what we have observed this had little to no impact either on the size or the power of the tests. Finally as expected the ASJL and LM12 tests are robust only in the higher frequencies and have a competitive edge when additive noise is introduced. Future research could compare the performance of the ASJ, CPR, and PZ tests across different threshold choices to identify an optimal value. It would also be useful to examine their behavior under random sampling, since their limit theories assume equidistant sampling, to see whether the results remain valid.

Chapter 3

Volatility Estimation When Noisy Prices Go Silent: A Kernel Multipower Approach

This chapter is concerned with the estimation of integrated volatility from high-frequency log-price data when both market microstructure noise and price staleness are present. We begin by outlining the challenges posed by these two forms of contamination and reviewing the limitations of existing estimators in this setting. We then introduce a novel estimator designed to remain robust to both noise and staleness and establish its finite-sample properties. Finally, we conduct a comprehensive simulation study to assess its performance, demonstrating that the proposed estimator delivers stable and reliable volatility estimates even under severe data distortions.

3.1 Introduction

The modeling and quantification of financial market volatility have long stood at the forefront of research in financial economics and econometrics. Volatility is central to numerous applications: it informs portfolio construction, risk management practices, and derivative valuation; it supports regulatory stress testing and systemic risk assessment; and it shapes our understanding of market efficiency and price formation. Because volatility dynamics are tightly linked to risk premia and asset-pricing frameworks, obtaining reliable estimates of volatility remains a foundational task in modern finance. The past twenty years have seen a profound shift in volatility research due to the widespread availability of high-frequency financial data. Traditional methods relying on daily or lower-frequency observations such as GARCH-type models are now frequently supplemented or replaced by realized measures constructed from intraday returns. The influential realized variance

(RV) estimator of Andersen and Bollerslev (1998), Andersen et al. (2003), and Barndorff-Nielsen and Shephard (2002) computes volatility as the sum of squared high-frequency returns. Under continuous semimartingale dynamics and frictionless sampling, RV consistently estimates the quadratic variation (QV) of the price process. This development has enabled a “model-free” approach to volatility measurement and forecasting, grounded in the statistical structure of high-frequency data rather than parametric assumptions. Yet empirical work soon made clear that these ideal conditions rarely hold in practice. At very high sampling frequencies such as one-minute or tick-by-tick data various frictions and empirical regularities distort realized measures. Two features that have become especially prominent in the jump-related literature are market microstructure noise and price staleness. Both, if ignored, introduce substantial biases and complicate inference about discontinuities in asset prices. Microstructure noise arises from bid–ask bounce, discrete pricing, order-processing costs, and asynchronous trading. As shown by Zhou (1996) and Hansen and Lunde (2006), noise causes the RV estimator to become inconsistent as sampling frequency increases, since squared returns increasingly capture market frictions rather than genuine price variation. This issue is particularly severe for highly liquid assets, where near-continuous trading makes it impossible to sample “true” prices at the highest frequencies. To mitigate noise, the literature has proposed several robust estimators. The two-scales realized variance (TSRV) estimator of Zhang et al. (2005) averages RV computed at different sampling frequencies to offset noise effects, with further improvements introduced in the multi-scales estimator of Zhang (2006). Kernel-based estimators Barndorff-Nielsen et al. (2008) and pre-averaging methods Jacod et al. (2009, 2010) smooth high-frequency returns to reduce noise contamination. Wavelet-based approaches Barunik and Vacha (2015); Fan and Wang (2007) filter noise in the frequency domain. Collectively, these methods yield consistent estimators of integrated volatility (IV) even when additive microstructure noise is present. A second key empirical feature of asset returns is the presence of jumps. For example, Andersen et al. (2007a) and Corsi et al. (2010) show that separating QV into its continuous (IV) and jump components enhances volatility forecasting. Jumps also influence derivative pricing by generating heavy tails and altering risk-neutral dynamics. In credit markets, Zhang et al. (2009) document that both IV and jump variation significantly affect credit default swap spreads. Consequently, detecting and quantifying jumps has become a major research area. Seminal nonparametric jump tests include Barndorff-Nielsen and Shephard (2006), Lee and Mykland (2008), and Aït-Sahalia and Jacod (2009), who propose statistics based on multipower variation, thresholding, or comparisons of truncated realized measures. Later studies such as Theodosiou and Zikes (2011), Dumitru and Urga (2012), and Maneesoonthorn et al. (2020) provide extensive Monte Carlo evaluations, highlighting how jump inference is affected by volatility clustering, noise, and sampling frequency. These findings reveal that different

tests often produce conflicting conclusions, underscoring the need for robust inference tools.

In contrast to noise and jumps, price staleness has only recently been recognized as a significant distortion in high-frequency data. Following [Bandi et al. \(2020\)](#), we use the term staleness to describe intervals in which transaction prices do not update, producing consecutive zero returns. This phenomenon is widespread even for highly liquid instruments such as S&P 500 futures, where zero returns account for roughly one-third of observations at the one-minute frequency, 16% at five minutes, and over 10% at ten minutes [Bandi et al. \(2020, 2017\)](#). Staleness arises when fundamental shocks are smaller than the tick size or when trading activity temporarily slows. Distinguishing between discreteness and inactivity is itself challenging ([Bandi et al. \(2020\)](#)). Ignoring staleness has two major implications for volatility and jump estimation. First, it increases the mass of the return distribution at zero. Second, it affects the conditional distribution of subsequent returns, which tend to be larger following sequences of zeros, thereby fattening the tails. These features mimic the statistical signatures of jumps, making the return distribution appear more leptokurtic. As [Kolokolov and Renò \(2024\)](#) argue, jump-diffusion models that neglect staleness will systematically overstate jump activity. Recent contributions have begun to address staleness directly. [Kolokolov and Renò \(2024\)](#) propose staleness-robust multipower variation estimators and show that conventional estimators substantially overdetect jumps when zeros are frequent. They also demonstrate that jump activity index (JAI) estimators become biased downward, creating the false impression that volatility is driven by infinite-activity jumps rather than continuous Brownian motion. Their corrected estimators remove these distortions and yield activity index estimates consistent with a Brownian component. Despite these developments, an important gap remains: the joint impact of microstructure noise and staleness and the resulting biases when both are present simultaneously has not been systematically examined. Most noise-robust estimators (e.g., realized kernels, pre-averaging) do not account for staleness, while staleness-robust procedures based on multipower variation remain vulnerable to noise. Consequently, empirical researchers face a trade-off: they must choose between estimators that are valid under noise but not staleness, or those that handle staleness but break down in the presence of noise.

3.2 Model Setup

Our analysis begins under the premise that the efficient log-price process of a financial asset follows a Brownian semimartingale process. Specifically, we denote by $(X_t, t \in [0, 1])$ of an asset. We represent by $t = 0$ the opening of the trading day and $t = 1$ the closing.

The process is initialized at an initial value $X_0 \in \mathbb{R}$, and its dynamics is described as:

$$dX_t = \mu_t dt + \sigma_t dW_t + J_t dN_t \quad (3.1)$$

where W_t is a standard Brownian motion, and J_t is a pure jump process, both of which are defined in a probability space $(\Omega, (F_t)_{t>0}, P)$. The coefficients μ_t and σ_t are adapted and locally bounded. We assume the volatility follows a stochastic process defined as:

$$\sigma_t = \sigma_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \int_0^t Z_s dZ_s \quad (3.2)$$

where σ_0 denote the initial value for the volatility processes. W_t and Z_s are independent Brownian motions. Equation 4.1, Equation 3.2 represent the general form of a stochastic volatility model. Further we assume that the jump part has finite activity and is expressed as $J_t = \sum_{i=1}^{M_t} Y_i$ where $M_t < \infty$ for finite t .

Much of the existing literature (e.g., Barndorff-Nielsen et al. (2008), Varneskov (2017), Christensen et al. (2010)) assumes that price observations within the interval $(0, 1]$ are uniformly spaced over time. This simplification facilitates technical analysis by treating observation times as equidistant fractions of the total sample size, N . Following these studies, we adopt the convention that $t_i = i/N$ for $i = 1, \dots, N$. A more general approach relaxes this assumption, permitting sampling times to be independent of both the price process and market noise, while requiring that the maximum interval between consecutive observations remains sufficiently small. Formally, this can be expressed as $\max(t_i - t_{i-1}) \leq C \frac{1}{N}$, where C is a constant ensuring that the largest time gap is at most proportional to $1/N$.

We include market microstructure noise in the efficient log price process Y_t , thus we observe X_t which is assumed to be generated as:

$$Y_{ti} = X_{ti} + u_{ti} \quad \text{for } i = 1, \dots, n, \quad (3.3)$$

where u_{ti} denotes the noise for the i -th observation with mean zero and variance h .

3.3 Existing IV Estimators

Numerous methods have been proposed in the literature to estimate an asset's integrated volatility (IV), defined as the integral of its squared volatility over time:

$$IV = \int_0^1 \sigma_t^2 dt. \quad (3.4)$$

A foundational approach, introduced by [Andersen and Bollerslev \(1998\)](#), approximates IV under, the assumption that the process is not been contaminated with noise or jumps, by summing the squares of intraday returns, yielding the well-known realized variance (RV) estimator:

$$RV = \sum_{i=1}^n (\Delta_i X)^2 \quad (3.5)$$

where $\Delta_i X = Y_i - Y_{i-1}$. [Barndorff-Nielsen and Shephard \(2002\)](#) established the asymptotic mixed normality of this RV estimator under these same conditions, showing that $M^{1/2}(RV - IV)$ is asymptotically mixed Gaussian given that it is not contaminated with jumps or noise:

$$M^{1/2}(RV - IV) \xrightarrow{L} MN(0, 2 \int_0^1 \sigma_s^4 ds), \quad (3.6)$$

as $M \rightarrow \infty$ where it becomes stable due to convergence law. Next we present some of the well known IV estimators in the literature and discuss their theoretical properties.

3.3.1 Multipower variation estimator

However, as noted, this classical RV estimator is not robust to the presence of jumps or noise in the underlying process. Consequently, a wide range of modified estimators have been developed to address these practical challenges. The following section reviews the most significant of these robust estimators.

Under this broader condition, standard estimators including realized variance (RV), realized kernels, pre-averaging, and multi-scale volatility measures retain their convergence properties toward integrated variance (IV) (see, e.g., [Barndorff-Nielsen et al. \(2011\)](#), [Fan et al. \(2012\)](#), [Kim et al. \(2016\)](#), [Wang and Zou \(2010\)](#)). When sampling shifts to tick-time, endogenous sampling may arise ([Robert and Rosenbaum \(2010\)](#), [Robert and Rosenbaum \(2012\)](#)). Notably, [Li et al. \(2014\)](#) demonstrated that even under endogenous sampling, the RV estimator maintains low error in the absence of noise and jumps. The general form of the Multipower variation estimator could be defined for a vector of m positive real numbers $r = [r_1, \dots, r_m]$ as:

$$MV(Y; r) = \frac{1}{n} \sum_{i=1}^{n-m+1} |\sqrt{n} \Delta_i X|^{r_1} \cdots |\sqrt{n} \Delta_{i+m-1} X|^{r_m} \quad (3.7)$$

Since this consist a general from we can obtain the realized variance by setting the powers as $r_1 = 2$ and $r_2 = 1$. In similar fashion we can select the powers as $r_1 = r_2 = 1$ to derive the Bipower variation estimator of [Barndorff-Nielsen et al. \(2008\)](#). The general multipower variation estimator could provide accurate estimates of the integrated variance as long as

the sampling frequency remains below the 5-minute mark. This is due to both staleness as demonstrated by [Kolokolov and Renò \(2024\)](#) but also due to microstructure noise where in higher frequencies becomes a prevalent stylized feature of the data. When the underlying series is contaminated with staleness, multipower based estimators yield biased estimates of the integrated variance. As shown by [Kolokolov and Renò \(2024\)](#), for some Bernoulli i.i.d variates we could explicitly show the bias as ¹:

$$MV(Y; r) \xrightarrow{p} \left(\prod_{j=1}^m \mu_{r_j} \right) \frac{(1 - p^\emptyset)^{m+1}}{p^\emptyset} \text{Li}_{-\frac{r_1}{2}}(p^\emptyset) \int_0^1 |\sigma_s|^{r^+} ds, \quad (3.8)$$

To account for a more realistic dynamics that includes staleness, the authors have introduced the staleness robust multipower variation estimator, that is designed to provide accurate estimates of the integrated variance defined as:

$$MV^c = \Delta_n \sum_{i=1}^{N_n - m + 1} |\Delta_{n,i}^{-1/2} \Delta_{ti} X|^{r_1} \dots |\Delta_{n,i+m-1}^{-1/2} \Delta_{ti+m-1} X|^{r_m} \quad (3.9)$$

where N_n is the stochastic grid of non-zero returns. [Equation 4.4](#), indicates that the bias of the estimator depends on three factors: the first power r_1 , the zero probability $p^\emptyset$, and the number of powers m . This bias results from two competing effects. The first effect is an attenuation caused when the product of consecutive returns equals zero due to either returns being zero. This creates a downward multiplicative bias. The second effect is volatility inflation that occurs when multiplying nonzero returns. In this case, the first return in the multiplication is computed from price differences where the time interval depends on the preceding sequence of zero returns. This effect can produce either upward or downward bias depending on the power r_1 . Under i.i.d. conditions, this bias takes a specific form. Because only the first return is influenced by prior staleness, the bias depends exclusively on the first power r_1 , while the other powers have no effect. Notably, realized variance ($r_1 = 2$ with $m = 1$) exhibits no bias for fundamental reasons.

¹where $\text{Li}_s(z)$ denotes the polylogarithm function of order s and argument z . So in this case the $\text{Li}_{-r_1/2}(p^\emptyset)$. It enters the expression because the number of consecutive stale observations is assumed to follow a geometric distribution. Its role here is to average the impact of all possible zero-return lengths on the estimator, $\text{Li}_{-r_1/2}(p^\emptyset) = \sum_{k=1}^{\infty} z^k / k^s$

3.3.2 Two scales realized variance estimator

The estimator proposed by [Zhang et al. \(2005\)](#), named as two-scales realized variance (TSRV) and is defined as:

$$TSRV = \frac{1}{K} \sum_{i=K}^M (Y_i - Y_{i-K})^2 - \frac{M-K+1}{MK} \sum_{i=1}^M (\Delta_x)^2 \quad (3.10)$$

The estimator is shown to consistently estimate the IV under the assumption that the underlying process is contaminated with microstructure noise.

The first term on the right-hand side of equation can be interpreted as the average of K realized variance estimators. These estimators have the same structure as the original realized variance RV , but use returns over K periods (i.e., $Y_i - Y_{i-K}$) instead of single-period returns. In the absence of noise, this term converges in probability to the integrated volatility (IV), while the second term serves to offset the bias introduced by market microstructure noise.

Assuming the underlying log-price process r_t is continuous and the noise terms u_i are independently and identically distributed (i.i.d.) and independent of the efficient price process y , [Zhang et al. \(2005\)](#) proved that $m^{1/6}(TSRV - IV)$ converges stably in law to a mixed Gaussian distribution. In cases where r_t is not continuous, it can be verified that the estimation error $(TSRV - IV)$ remains of order $O_p(m^{-1/6})$.

3.3.3 Realized kernel estimator

In their seminal work [Barndorff-Nielsen et al. \(2011\)](#) define the flat-top realized kernel as

$$RK = \gamma_0(X_\delta) + \sum_{h=1}^H k\left(\frac{h-1}{H}\right) (\gamma_h(X_\delta) + \gamma_{-h}(-X_\delta)) \quad (3.11)$$

The term $\gamma_0(X_\delta)$ is the Realized Variance (RV), the $k(\cdot)$ term refers to non-stochastic kernel weight which satisfies the conditions $k(0) = 1$ and $k(1) = 0$. This estimator was introduced in order to smooth away the impact of noise with the use of kernel functions. A difficulty in empirically applying the formula is that the estimator $K(X_\delta)$ requires some data outside the time frame $[0,1]$. Thankfully [Barndorff-Nielsen et al. \(2008\)](#) provides evidence that we can change the start and end points without any effect on the estimator. In the absence of jumps, it should converge stably in law to a Mixed Gaussian distribution as has also been proved by the authors. We employ the Parzen kernel for kernel weight functions as it satisfies the required conditions and is unlikely to produce negative estimates.

3.3.4 Truncated realized variance estimator

The truncation method has been applied to realized variance by [Mancini \(2008, 2009\)](#) and [Mancini and Renò \(2011\)](#), resulting in the truncated realized variance (TRV) estimator, which consistently estimates integrated volatility (IV) even in the presence of jumps.

The core idea of truncation is to identify and remove returns likely contaminated by jumps. For a fine sampling interval, a large absolute price change is a strong indicator of a jump. The method sets a threshold, $T(m)$, that converges to zero slower than the typical magnitude of a continuous price move as the sampling frequency increases ($m \rightarrow \infty$). Returns exceeding this threshold are excluded from the volatility sum. Formally, the TRV estimator is defined as:

$$TRV = \sum_{i=1}^m (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq T(m)\}}. \quad (3.12)$$

[Mancini \(2009\)](#) showed that in the presence of finite-activity jumps, TRV possesses the same asymptotic distribution as the standard RV estimator would in a jump-free setting, as the truncation effectively removes jump effects with an information loss of order $o_P(m^{-1/2})$. This consistency result was also extended to infinite-activity Lévy jumps. Furthermore, [Mancini and Renò \(2011\)](#) demonstrated that for finite-variation jump processes, the asymptotic distribution of TRV remains unchanged from that of the jump-free RV estimator.

In summary, while estimators like Bipower Variation (BPV) and TRV are designed to handle jumps, other estimators like Two-Scale realized variance (TSRV), Realized Kernel (RK), and Pre-Averaging (PA) focus on robustness to market microstructure noise. A natural progression in the literature has been to combine these approaches to create estimators robust to both issues. This has led to the development of hybrid methods, such as applying pre-averaging to bipower variations [Podolskij and Vetter \(2009a\)](#), [Podolskij and Vetter \(2009b\)](#), [Christensen et al. \(2014\)](#) and creating a truncated pre-averaging estimator ([Jing et al., 2014](#)).

3.3.5 Truncated pre-averaging estimator

Estimators based on truncated pre-averaging, which provide consistent integrated volatility (IV) estimation in the presence of both infinite-activity jumps and market microstructure noise, have been developed by several authors, including [Wang et al. \(2013\)](#), [Jing et al. \(2014\)](#), and [Koike \(2016\)](#). While [Wang et al. \(2013\)](#) established consistency without specifying a convergence rate, the latter two papers derive a central limit theorem (CLT) for their respective estimators.

Jing et al. (2014) proposed a detection method for jumps using the absolute value of the pre-averaged return, $(\Delta_{i,k_0}X(g))$, as defined in their pre-averaging estimator ². An unusually large value of this statistic is likely indicative of a jump within the corresponding time window $]t_i, t_{i+K_0-1}]$, which would otherwise distort the asymptotic distribution of the standard pre-averaging (PA) estimator. To counteract this, they introduced the truncated pre-averaging (TPA) estimator, which modifies the PA estimator by removing squared terms $(\Delta_{i,k_0}X(g))^2$ whenever $|\Delta_{i,k_0}X(g)|$ exceeds a predetermined threshold u_m . The threshold is chosen such that $u_m m^{\varpi_1} \rightarrow 0$ and $u_m m^{\varpi_2} \rightarrow \infty$ for some $0 < \varpi_1 < \varpi_2 < 1/4$. The TPA estimator is defined as:

$$TPA = \frac{1}{K_0 \phi_2} \sum_{i=0}^{m-K_0+1} (\Delta_{i,K_0}X(g))^2 \mathbf{1}_{\{|\Delta_{i,K_0}X(g)| \leq u_m\}} - \frac{\phi_1}{2K_0^2 \phi_2} \sum_{i=1}^m (\Delta_i X)^2, \quad (3.13)$$

where $\phi_1 = \int_0^1 g'(u)^2 du$ and $\phi_2 = \int_0^1 g(u)^2 du$ similar to the pre-averaging estimator of Jacod et al. (2009). Notably, truncation is not applied to the second, bias-correction term, as the impact of jumps on this component is of order $O_P(m^{-1/4})$ and the downward bias from truncating the first component is negligible.

Under the assumption that the log-price follows the dynamics of the efficient log-price process ³ the noise process shares properties with that in Jacod et al. (2009), and the infinite-activity jump process satisfies certain conditions, Jing et al. (2014) proved that the TPA estimator shares the same asymptotic distribution as the pre-averaging estimator of Jacod et al. (2009) derived in a jump-free setting by Jacod et al. (2009).

In a related development, Koike (2016) incorporated the Hayashi-Yoshida technique into the truncated pre-averaging framework. Originally designed by Hayashi and Yoshida (2005) to handle asynchronous data in multivariate settings, this adaptation yields an estimator called the truncated pre-averaging Hayashi-Yoshida estimator, that is designed to remain robust not only to jumps and noise but also when the assumption regarding equidistant sampling is violated.

3.3.6 Truncated flat-top realized kernel estimator

Another novel truncated estimator proposed by Brownlees et al. (2020) is the truncated flat-top realized kernel (TFTRK) estimator. This estimator consistently estimates the integrated volatility (IV) in the presence of jumps and M -dependent noise, achieving the optimal convergence rate of $m^{1/4}$. Here, M -dependence means that the noise terms u_i and

²Jacod et al. (2009) use a pre-averaging to construct an estimator of IV robust against microstructure noise as described in equation 10 of their study.

³The dynamics of the efficient process are described in equations 1 and 2 of the Brownlees et al. (2020)

u_j are independent whenever $|i - j| > M$. Under those assumption the authors define the estimator as:

$$TFTRK = \sum_{i=1}^m \left(\sum_{h=1}^{H(1+C)} k\left(\frac{h}{H}\right) (\Delta_{i-h}X + \Delta_{i+h}X) + \Delta_iX \right) \Delta_iX 1_{E_i} \quad (3.14)$$

1_{E_i} consist of a condition that states $E_i = \{\beta_j < r_m\}$ for $j = i - H(1 + C), \dots, i + (1 + C)$. If we remove the conditional term at the end of equation, then we are left with the same formula as the realized kernel. In order to smooth away the impact of boundary effects, the authors choose a similar jittering technique to (Barndorff-Nielsen et al., 2008), (Barndorff-Nielsen et al., 2011) and to (Varneskov, 2017). β_j is a jump indicator function, its purpose is to be able to eliminate jumps for the series. The formula is

$$\beta_i(m) = \frac{1}{K_3} \sum_{j=1}^{i+K_3-1} (x_j - x_{j-K_3}) \quad (3.15)$$

If the value of $\beta_i(m)$ is larger than a threshold value then returns that produce this result will be truncated. For that reason, the TFTRK estimator is sensitive to the choice of this threshold value.

3.4 Novel kernel estimator

A novel estimator introduced in this paper is the kernel-weighted staleness corrected multipower variation estimator (KS-MPV) which is designed to consistently estimate IV in the presence of jumps, gaussian i.i.d noise and staleness. Due to the use of kernel the estimator exhibits the same convergence as the realized kernel estimator, namely $m^{-1/4}$ under the assumption that the underlying process is contaminated with microstructure noise. In the absense of noise, the convergence returns to the classic realized variance $m^{-1/2}$.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,1]}, \mathbb{P})$ be a filtered probability space and $\{\tau_0, \dots, \tau_{N_n}\}$ denote the stochastic grid of non-zero returns, constructed from the original observation times $\{t_0, \dots, t_n\}$ by skipping intervals where the price fails to update (staleness). The process is observed on a regular grid $t_i = i\Delta_n$, $i = 0, \dots, n$, with $\Delta_n = 1/n$.

Staleness is introduced via an i.i.d. Bernoulli sequence $\{S_i\}_{i \geq 1}$ such that

$$\Delta_{t_i}X^{\text{obs}} = S_i \Delta_{t_i}X, \quad \mathbb{P}(S_i = 1) = 1 - p^0, \quad (3.16)$$

independently of X . When $S_i = 0$, the observed return is zero. The stochastic grid of nonzero returns could be expressed as:

$$\tau_0 = 0, \quad \tau_i = \inf\{t_j > \tau_{i-1} : S_j = 1\}, \quad (3.17)$$

and let N_n denote the number of nonzero returns. The durations between successive nonzero returns are

$$\Delta^{(n,k)} = \tau_k - \tau_{k-1}. \quad (3.18)$$

Under i.i.d. additive noise and staleness,

$$\Delta^{(n,i)} = K_i \Delta_n, \quad K_i \sim \text{Geometric}(1 - p^0). \quad (3.19)$$

We follow [Kolokolov and Renò \(2024\)](#) in defining the duration normalized returns accounting for the existence of staleness as:

$$Y_k = \Delta^{(n,k)-1/2} \left(X_{\tau_k}^{\text{obs}} - X_{\tau_{k-1}}^{\text{obs}} \right), \quad k = 1, \dots, N_n. \quad (3.20)$$

To define the kernel-weighted staleness corrected multipower variation estimator we let $m \geq 1$ be fixed and let $r_1, \dots, r_m \in (0, 2]$. Defining $r^+ := \sum_{j=1}^m r_j$, thus $k : [0, 1] \rightarrow \mathbb{R}$ is a bounded kernel satisfying

$$\int_0^1 k(u) du = 1, \quad \int_0^1 k(u)^2 du < \infty.$$

where by $k(u)$ we denote the Parzen kernel:

$$k(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & |u| \leq 1/2, \\ 2(1 - |u|)^3, & 1/2 < |u| \leq 1, \\ 0, & |u| > 1. \end{cases} \quad (3.21)$$

We denote as H_n the kernel bandwidth such that $H_n \rightarrow \infty$ and $H_n/N_n \rightarrow 0$, under those conditions the kernel smoothed returns are defined as:

$$Y_k = \frac{1}{\sqrt{H_n}} \sum_{h=0}^{H_n} k\left(\frac{h}{H_n}\right) Y_{k-h}, \quad k = H_n + 1, \dots, N_n. \quad (3.22)$$

The normalization $1/\sqrt{H_n}$ ensures that $\text{Var}(\tilde{Y}_k) = O(1)$ while reducing the contribution of microstructure noise. This kernel tapers the contributions of blocks near the sample boundaries, analogous to spectral windowing in long-run variance estimation. Unlike a simple multiplicative constant, it smoothly reduces the influence of early and late

observations, which improves finite-sample properties and reduces edge effects. By using the stochastic grid $\{\tau_i\}$, the estimator accounts for missing returns due to price staleness. Blocks are automatically adjusted by the durations $\Delta^{(n,i)}$, preserving the correct scaling. The Parzen kernel reduces bias from incomplete multipower blocks at the boundaries. Since the asymptotic bias arises from the interaction between the stochastic durations and the first power, the kernel weighting scheme does not affect the probability limit. Let $m \geq 1$ and $r_1, \dots, r_m \in (0, 2]$, and define $r^+ = \sum_{j=1}^m r_j$. The kernel smoothed staleness corrected multipower variation estimator is defined as:

$$\text{MPV}_n^{KS}(r) = \frac{1}{N_n} \sum_{k=H_n+1}^{N_n-m+1} \prod_{j=1}^m \left| \tilde{Y}_{k+j-1} \right|^{r_j}. \quad (3.23)$$

When the underlying process is assumed to be absent of microstructure noise, then it can be shown that the convergence rate of the estimator converge to the optimal rate of $m^{-1/2}$

$$\text{MPV}_n^{KS}(r) \xrightarrow{P} \left(\prod_{j=1}^m \mu_{r_j} \right) \int_0^1 \sigma_s^{r^+} ds^4 \quad (3.24)$$

Under the assumption that the underlying process is contaminated with i.i.d microstructure noise, given that $H_n \sim cN_n^{1/2}$, then

$$\text{MPV}_n^{KS}(r) - \mathbb{E}[\text{MPV}_n^{KS}(r)] = O_p(N_n^{-1/4}). \quad (3.25)$$

The bandwidth balances the smoothing bias of order H_n/N_n and the variance reduction of order $1/H_n$, yielding the optimal noise-robust rate.

The proposed KS-MPV estimator directly addresses these distortions by modifying the construction of the multipower variation in three ways. First, the summation is restricted to nonzero returns, thereby eliminating artificial zero-return blocks that arise from stale prices. Second, each increment is normalized by its sampling duration $\Delta^{(n,i)-1/2}$, ensuring that longer inactive intervals contribute proportionally less to the overall variation. Third, the inclusion of a smooth Parzen kernel $k(u)$ introduces a continuous weighting scheme that downweights boundary terms and smooths local irregularities in the stochastic sampling grid. Together, these adjustments preserve the estimator's consistency while reducing its small-sample variance and mitigating the finite-sample bias caused by irregular observation times.

⁴The associated estimation error is of stochastic order ($O_p(N_n^{-1/2})$)

3.5 Simulation Study

In this section we investigate through simulations the finite distributional properties of the review estimators along with the proposed KS-MPV estimator. To that end we perform a Monte Carlo exercise using the stochastic volatility model described in [Equation 4.1](#), and [Equation 3.2](#). We set the parameters of the data generating process as $\mu_t = 0, k = 5, \theta = 0.16, \sigma_v = 0.5$ and $\rho = -0.5$ adhering to the theoretical restriction $2k\theta \geq \sigma_v$ ⁵. We employ an Euler clock set to simulate the high-frequency observations, at 1 second increment. We generate 21600 observations per simulated day which is equivalent of a typical 6.5 hour trading day. For all the kernel-based estimators we use the parzen kernel and following [Brownlees et al. \(2020\)](#) and [Varneskov \(2017\)](#) and choose $C = H^{-3/5}$ and $H = \lfloor 0.5m^{1/2} \rfloor = 77$.

[Table 3.1](#) reports the mean square error (MSE) of the QRV, MBRK, and TFTRK, TTSRV, TPV, MBV and KS-MPV estimators as a function of the standard deviation U of the i.i.d. noise. The estimators exhibit markedly different sensitivities to noise. The QRV estimator shows a particularly high sensitivity to the noise level u , its MSE is relatively large at low noise levels, decreases sharply as U increases and reaches a minimum around $U = 0.01$, and then increases again for larger noise values, indicating unstable performance and high sensitivity to noise. In contrast, the MBRK estimator is more robust, with MSE values remaining within a relatively narrow range across all noise levels, reflecting stable but moderate performance. The TFTRK estimator consistently achieves the lowest MSE for all values of U and shows very limited variation as the noise level increases, demonstrating superior accuracy and robustness to i.i.d. noise compared to the other two estimators. The TTSRV, TPV and MBV show a consistent low MSE across the noise levels with the lowest values being seen by the first. The KS-MPV proposed in this study displays an initially high MSE, which is attributed to the use of the parzen kernel for mitigating the effects of microstructure noise. When the underlying series has either low or no noise present the parzen kernel will tend to over-smooth the data, thus introducing a bias in the estimation. As such, one limitation of the our proposed multipower variation estimator is its sensitivity to the kernel bandwidth.

We now proceed to explore how alternative distribution for the noise affects the MSE of the estimator. To that end we assume that the underlying process X is contaminated with additive microstructure noise following an MA(1) process. Similarly to [Varneskov \(2017\)](#) [Kim et al. \(2016\)](#) and [Brownlees et al. \(2020\)](#) we assume that the additive microstructure noise could be modeled as:

$$u_i = \theta_{i+1} - 0.5\theta_i \quad (3.26)$$

⁵The selection of the parameters satisfy the theoretical restrictions and there in line to [Maneesoonthorn et al. \(2020\)](#)

Table 3.1 MSE comparisons for different values of the noise variance (i.i.d. noise).

U	0	0.002	0.004	0.006	0.008	0.01	0.012	0.014
QRV	0.7461	0.6892	0.5296	0.3125	0.1052	0.0021	0.1240	0.6240
MBRK	0.0693	0.0586	0.0615	0.0641	0.0682	0.0634	0.0756	0.0598
TFTRK	0.0213	0.0206	0.0213	0.0218	0.0226	0.0215	0.0226	0.0199
TTSRV	0.0395	0.0400	0.0376	0.0389	0.0367	0.0399	0.0381	0.0393
TPV	0.1320	0.1400	0.0950	0.1340	0.0990	0.1380	0.1020	0.1350
MBV	0.3740	0.4520	0.5680	0.3490	0.3990	0.4470	0.3720	0.4400
KS-MPV	0.2461	0.2030	0.0725	0.0538	0.0145	0.0160	0.0110	0.0120

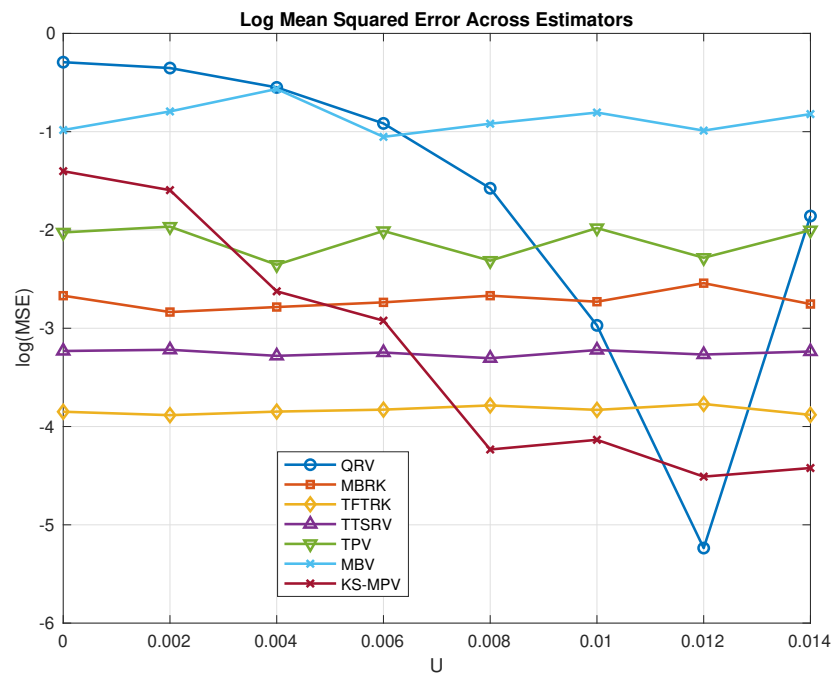


Fig. 3.1 The figure shows the plot of log MSE of the QRV, MBRK and TFTRK estimators as a function of additive Gaussian i.i.d. noise.

where θ_i is considered a discrete i.i.d and we assume that the standard deviation of the noise is $N(0, U^2/1.25)$ to normalize the noise and achieve similar long-run variance .

Table 3.2 MSE comparisons for different values of the noise variance (MA(1) noise)

U	0	0.002	0.004	0.006	0.008	0.01	0.012	0.014
QRV	0.8050	0.6460	0.5630	0.2770	0.1960	0.1210	0.1830	0.6930
MBRK	0.0970	0.1320	0.1010	0.1110	0.0920	0.1110	0.0840	0.0950
TFTRK	0.0570	0.0630	0.0460	0.0730	0.0340	0.0590	0.0460	0.0550
TTSRV	0.0640	0.0770	0.0550	0.0820	0.0400	0.0800	0.0570	0.0650
TPV	0.1480	0.1330	0.1000	0.1280	0.0970	0.1440	0.1060	0.1110
MBV	0.3880	0.6030	0.4700	0.4230	0.4470	0.3720	0.3760	0.4260
KS-MPV	0.3261	0.2501	0.0725	0.0538	0.0450	0.0360	0.0410	0.0421

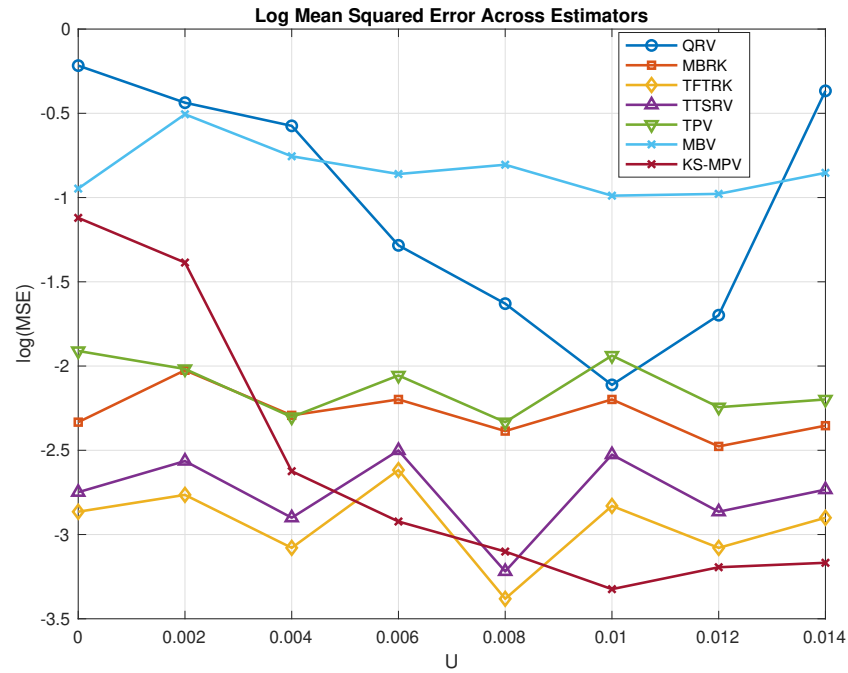


Fig. 3.2 The figure shows the plot of log MSE of the QRV, MBRK and TTRPK estimators as a function of additive MA(1) noise.

Table 3.2 reports the mean square error (MSE) of the QRV, MBRK, and TFTRK, TTSRV, TPV, MBV and KS-MPV estimators as a function of the standard deviation U of MA(1) noise. Compared to the i.i.d. noise case, the presence of temporal correlation affects the estimators differently. The QRV estimator remains highly sensitive to the noise level, its MSE is large at low noise levels, decreases as U increases and reaches a minimum around $U = 0.012$, before increasing again for larger noise values. This behavior indicates that QRV is particularly vulnerable to correlated noise and exhibits unstable

performance across noise regimes. The MBRK estimator shows improved robustness to MA(1) noise, with MSE values remaining within a relatively narrow range across all noise levels, suggesting stable performance despite the presence of noise correlation. The TFTRK estimator consistently achieves the lowest MSE for all values of U and displays minimal variation as the noise level increases, demonstrating strong robustness and superior accuracy in the presence of MA(1) noise. The MBV, TTSRV and MBRK estimators display stable MSE across the noise levels with the TTSRV having the lowest values of them all. As a general observation all estimators show a greater variability in their estimation performance.

3.5.1 Staleness

To evaluate the performance of our proposed correction together with the benchmark IV estimators, we introduce price staleness into the model following the framework of [Kolokolov and Renò \(2024\)](#). Under this setting, the data generating process is specified as

$$X_{t_i}^s = (1 - S_{t_i})X_{t_i}^s + S_{t_i}X_{t-1}, \quad (3.27)$$

where observed prices may reflect either the current or the previous efficient price, depending on the staleness indicator as described in [Equation 3.16](#).

As a consequence, the occurrence of zero returns is no longer governed by a constant probability but instead displays systematic intraday variation. Specifically, this probability follows an inverted U-shaped pattern over the trading day, attaining its highest values around mid-day and remaining lower during the opening and closing intervals. To reproduce this empirically documented feature, we model the time-dependent probability of zero returns in line with the authors' approach as

$$p_t = p^U \cdot \gamma_t^{(p)}. \quad (3.28)$$

Here, p^U represents the baseline level of staleness in the absence of intraday effects and can be interpreted as the unconditional probability of observing a zero return. The factor $\gamma_t^{(p)}$ introduces intraday variation⁶ scaling this baseline probability over time, thereby generating the characteristic inverted U-shaped profile. This scaling factor is obtained by inverting the intraday volatility pattern γ_t defined in [Equation 4.7](#) and subsequently normalizing it using a constant γ_0 :

$$\gamma_t^{(p)} = \frac{\gamma_0}{\gamma_t}. \quad (3.29)$$

⁶For the $\gamma_t^{(p)}$ value we follow [Kolokolov et al. \(2020\)](#) and model it as: $\gamma_t = \frac{1}{1033}(0.1271t^2 - 0.1260t + 0.1239)$

The constant γ_0 is selected such that the average value of $\gamma_t^{(p)}$ over a full trading day equals one. This normalization guarantees that, despite intraday fluctuations in the probability of zero returns, the expected daily frequency of zero returns remains anchored to the baseline staleness level p^U .

Figure 3.3 illustrates the bias of the staleness corrected mutlipower variation estimator and the one we proposed, The underlying process is constructed according to Equation 4.1, Equation 3.2, we further assume the the process is contaminated with additive noise ⁷ and a 30% probability of staleness. Similarly to the previous instances with the MSE our estimator is underestimating the true variance while becoming unbiased when the noise standard deviation is 0.013 and it becomes slightly over biased at the highest noise level. On the opposite the staleness corrected MPV estimator tends to be unbiased estimator when noise is absent but it becomes over biased as the noise variance increases. This figure demonstrates that even moderate amounts of microstructure noise when present along side staleness can have a significant impact when attempting to estimate the variance of the underlying process.

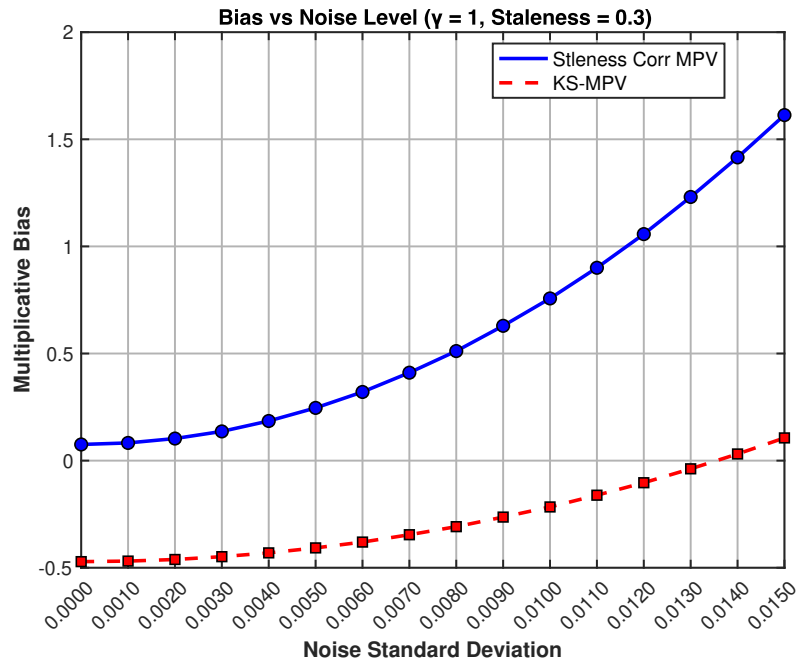


Fig. 3.3 Bias of the staleness-corrected MPV estimator and the MPV estimator proposed in this study as a function of the Gaussian i.i.d. additive noise.

⁷The noise is assumed to be Gaussian i.i.d with zero mean an a standard deviation of 0.01

3.5.2 Finite-Sample Performance

Having explored the behavior of the estimators under different assumed distribution of additive noise, and taking into account their behavior under staleness which has received little academic interest, now we shift our attention to explore the finite-sample behavior of the proposed estimator through a series of Monte Carlo simulations. In particular, we examine the limiting behavior of our proposed estimator captured through bias as a function of the sample size. By gradually increasing the sample size, we assess the convergence of the estimator toward its asymptotic properties. The results demonstrate that the estimator exhibits non-negligible bias in small samples, which steadily decreases as the number of observations grows, ultimately vanishing in large samples. This provides empirical evidence of the estimator's consistency and supports its asymptotic behavior.

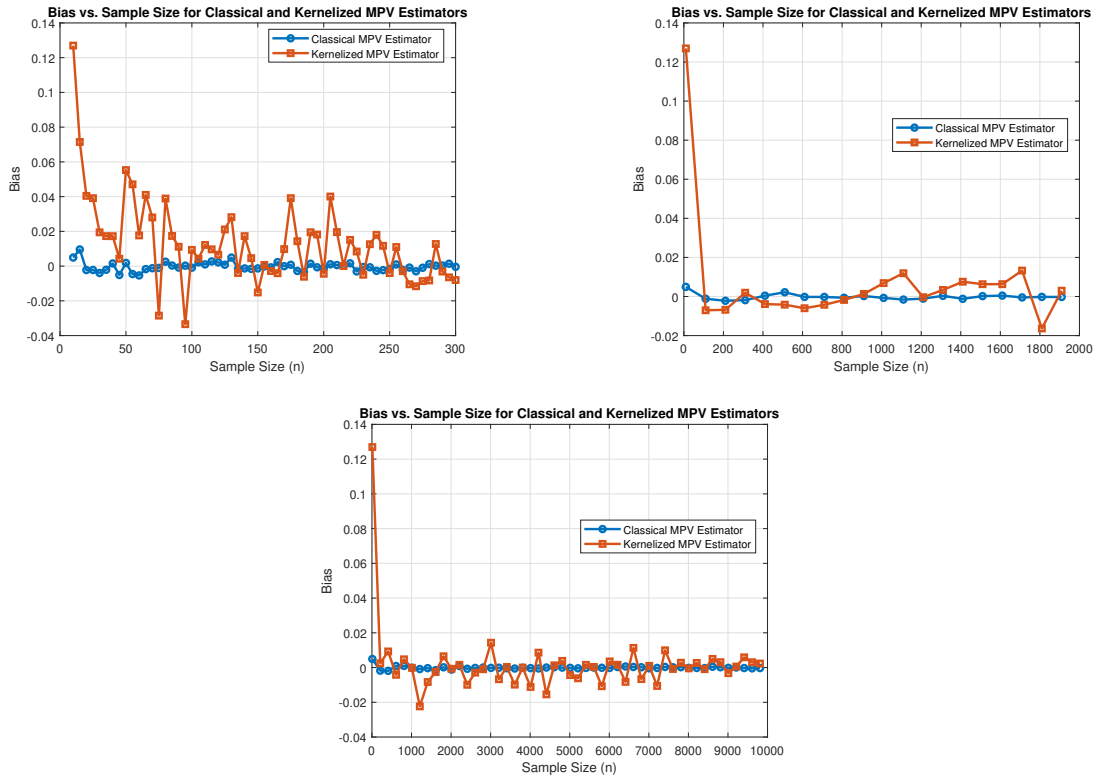


Fig. 3.4 Bias of the two estimators as a function of the sample size N .

Figure 3.4 show the bias of our corrected estimator as function of the sample size. The blue line (Classic MPV estimator) represents the realized variance of (Andersen and Bollerslev (1998), Andersen et al. (2003), and Barndorff-Nielsen and Shephard (2002)) while the orange line (kernalized MPV estimator) represents the corrected KS-MPV estimator. The underlying process is assumed to be following a pure Brownian motion with microstructure noise and staleness to be considered absent.

The simulation results appear broadly consistent with the asymptotic properties established in [section 3.4](#). In particular, the kernel-staleness corrected multipower variation estimator exhibits a noticeably larger finite-sample variance relative to the standard realized variance estimator. This behavior is expected from the asymptotic variance properties, since kernel smoothing introduces overlapping windows and therefore additional dependence across adjacent observations. As shown in the variance calculations, this overlap increases the stochastic fluctuation of the estimator through the $(O(H_n/n))$ variance term. At the same time, the simulations indicate that the proposed estimator becomes approximately unbiased as the sample size increases, though convergence occurs at substantially larger sampling frequencies relative to the standard realized variance estimator. This slower convergence is consistent with the asymptotic theory. The correction mechanism compensates for stale-price contamination through duration normalization and kernel smoothing, but these adjustments increase finite-sample variability and slow convergence toward the asymptotic limit. By contrast, the standard realized variance estimator becomes approximately unbiased much earlier in the simulations because, in the absence of staleness correction, it does not experience the additional variability induced by kernel overlap and random duration adjustments. However, under staleness, the asymptotic theory suggests that it converges to a biased limit, whereas the proposed correction for the estimator remains asymptotically centered around the corrected integrated volatility functional.

Overall, the simulation evidence supports the theoretical properties and the resulting tradeoff implied by the asymptotic analysis: the proposed estimator sacrifices finite-sample efficiency in exchange for robustness to staleness contamination and asymptotic consistency under irregular observation durations.

3.6 Conclusions

Failing to account for zero returns and market microstructure noise in high-frequency financial data can distort conclusions about how often price jumps occur, how large they are, and what role they play in asset dynamics. By introducing a correction that explicitly incorporates both noise and price staleness, we attempted to restore the statistical validity of the multipower variation estimator. Our findings align with standard stochastic volatility models driven by Brownian motion and suggest that pure jump components may be less central to explaining observed price behavior. More broadly, we argue that both staleness and noise must be embedded in the assumptions underlying high-frequency data generation. Ignoring these frictions can bias multipower variation estimators and undermine broader econometric inference. Further theoretical work is needed on modeling price rounding, as emphasized by [Li and Mykland \(2015\)](#). Although the estimators we examine perform

reasonably well unless rounding is severe, developing a more robust correction remains an open challenge. Our heuristic adjustment reduces the impact of aggressive rounding, but accurately estimating multipower variation in the joint presence of rounding, noise, and staleness is still unresolved. Even so, our results demonstrate that properly accounting for these frictions leads to more accurate measurements and more powerful tests of price dynamics. Specifically the Monte Carlo simulations performed indicate that the correction proposed reduces the bias of the multipower variation estimator when the underlying process is contaminated by both microstructure noise and staleness thus providing a more accurate estimation of the true volatility. Our proposed estimator is nonetheless without its limitations, we observe that when microstructure noise is low and/or staleness is absent (as is when sampling frequency is lower than 5-minutes) for the underlying series then the bias of it is a higher than some alternative competing estimators. In addition our results also demonstrated that our correction requires significantly more data to provide accurate estimates of the volatility.

Chapter 4

Jumps in Asset Prices: Illusion of Realized Inference under Noise at stand still

This chapter investigates the estimation of integrated volatility from high-frequency financial data in the presence of market microstructure noise and price staleness, which induces zero returns. Expanding on existing econometric techniques, we propose a novel estimator that is robust to both sources of contamination and derive its finite-sample distributional properties. Through an extensive Monte Carlo simulation study, we compare its performance to standard estimators, highlighting its robustness in disentangling the true volatility dynamics from spurious jump components. Our findings show that jumps are less frequent and less significant in contributing to price variation than the empirical literature indicate.

4.1 Introduction

The modeling and measurement of financial market volatility has long been a central topic in both financial economics and econometrics. Volatility plays a fundamental role in a wide range of applications namely, it is a crucial input in portfolio allocation, risk management, and derivative pricing as it informs regulatory stress testing and systemic risk monitoring while also underlies our understanding of market efficiency and price formation. In particular, volatility dynamics are deeply connected to notions of risk premia and asset pricing models, making reliable estimation of volatility a cornerstone of modern financial research. Over the past two decades, the increasing availability of high-frequency financial data has transformed the empirical study of volatility. Traditional approaches based on daily or lower-frequency data, such as GARCH-type models, are increasingly complemented or replaced by realized measures of volatility derived from intraday data.

The seminal realized variance (RV) estimator, introduced by [Andersen and Bollerslev \(1998\)](#), [Andersen et al. \(2003\)](#), and [Barndorff-Nielsen and Shephard \(2002\)](#), calculates volatility as the sum of squared intraday returns. Under the assumption of continuous semimartingale dynamics and frictionless observation, the RV estimator is a consistent measure of the quadratic variation (QV) of the price process. This line of work has opened the door to a “model-free” approach to volatility measurement and forecasting, one that relies on the statistical properties of high-frequency data rather than on parametric model assumptions. However, empirical research quickly revealed that idealized conditions rarely hold in practice. When volatility is studied at ultra-high frequencies (e.g., one-minute returns or tick), a number of frictions and stylized facts distort the behavior of realized measures. Two frictions that have become prominent in the jump related literature are market microstructure noise and price staleness. Both introduce severe biases if ignored, and both complicate inference about jumps in asset prices.

Sources of noise include bid–ask bounce, price discreteness, order processing costs, and asynchronous trading. As [Zhou \(1996\)](#) and [Hansen and Lunde \(2006\)](#) showed, the presence of noise renders the traditional RV estimator inconsistent as the sampling frequency increases, because squared returns increasingly reflect microstructure artifacts rather than true economic variation. This problem is especially acute in highly liquid assets where trading is nearly continuous, making it impossible to accurately sample prices at the highest available frequency. To address noise, researchers have developed a variety of robust estimators. The two-scales realized variance (TSRV) estimator of [Zhang et al. \(2005\)](#) averages multiple subsampled RV estimates to cancel out noise-induced variation, with further refinements in the multi-scales estimator of [Zhang \(2006\)](#). Kernel-based methods [Barndorff-Nielsen et al. \(2008\)](#) and pre-averaging approaches [Jacod et al. \(2009, 2010\)](#) smooth away high-frequency noise using weighted combinations of returns. Other contributions, such as wavelet methods [Barunik and Vacha \(2015\)](#); [Fan and Wang \(2007\)](#), exploit the frequency domain to filter noise. Collectively, these techniques have produced estimators of integrated volatility (IV) that remain consistent under the influence of additive microstructure noise.

Another important stylized fact of asset returns is the presence of discontinuities, or jumps. Those can be generated by news arrivals, liquidity shocks, or structural breaks, and they carry distinct economic implications. For instance, [Andersen et al. \(2007a\)](#) and [Corsi et al. \(2010\)](#) show that decomposing QV into the continuous (IV) and the jump components (dJ_t) improves volatility forecasting. Moreover, jumps affect derivative pricing, as they generate non-normal tails and alter risk-neutral dynamics. On the credit side, [Zhang et al. \(2009\)](#) demonstrate that both integrated volatility and jump variation significantly influence credit default swap spreads.

Testing for jumps and estimating their intensity has therefore become an active area of research. Pioneering nonparametric jump tests include [Barndorff-Nielsen and Shephard \(2006\)](#), [Lee and Mykland \(2008\)](#), and [Aït-Sahalia and Jacod \(2009\)](#), who developed statistics based on multipower variation, thresholding, or comparisons of truncated realized measures. Subsequent contributions provide comprehensive Monte Carlo evaluations of jump tests, such as [Theodosiou and Zikes \(2011\)](#), [Dumitru and Urga \(2012\)](#), and [Maneesoonthorn et al. \(2020\)](#). These studies highlight the sensitivity of jump inference to factors such as volatility clustering, noise contamination, and sampling frequency. They also show that different tests can yield conflicting results in empirical settings, motivating the development of robust and reliable inference procedures.

While noise and jumps have received considerable attention, price staleness has only recently been recognized as a key distortion in high-frequency data. we borrow the term Staleness from the [Kolokolov et al. \(2020\)](#), referring to it as the intervals where transaction prices have not changed, resulting in consecutive zeros in the return series. This phenomenon is pervasive even in highly liquid assets such as the S&P 500 financial instruments, where the share of zero returns is roughly one-third at the one minute horizon, 16% at five minutes, and still over 10% at ten minutes [Bandi et al. \(2020, 2017\)](#). When a fundamental shock is smaller than the tick size, it may not result in a visible price change, even though the asset's true value has shifted. Another source is trading, since it is irregular, there will be periods of low market activity, resulting in prices remaining stale. Distinguishing between discreteness and inactivity is by itself a challenge [Bandi et al. \(2020\)](#). Ignoring staleness has two key consequences for volatility and jump estimation. First, it inflates the return distribution at zero, increasing the empirical mass at the origin. Second, it alters the conditional variance of subsequent returns, as returns following a sequence of zeros tend to be larger, fattening the tails of the distribution. Together, these effects mimic the statistical properties of jumps, making the distribution appear more leptokurtic. As argued by [Kolokolov et al. \(2020\)](#), jump-diffusion models that ignore staleness will spuriously infer more jumps than actually exist.

Recent work has begun to account for staleness explicitly. [Kolokolov et al. \(2020\)](#) proposed a staleness-robust multipower variation estimators, showing theoretically and empirically that traditional estimators systematically tend to detect a considerably higher amount of jumps in the presence of zero returns. They also demonstrate that jump activity index (JAI) estimators become biased downward as zeros become more frequent, leading to the false impression that volatility is driven by infinite activity jump processes rather than continuous Brownian motion. Their corrected estimators eliminate these biases, yielding activity index estimates consistent with a Brownian driven component.

Despite these advances, the literature still faces an important gap, since the joint treatment of noise and staleness and their resulting inference bias of the simultaneous

presences of both stylized facts has yet to be studied. Most noise-robust estimators (e.g., realized kernels, pre-averaging) do not account for staleness, while the staleness robust procedure build on multipower variation estimators, remain sensitive to microstructure noise. As a result, applied researchers must either choose estimators that will provide accurate estimation of the volatility under the assumption that the underlying process is being contaminated with noise, or with estimators that will remain consistent under the assumption of staleness.

This paper builds on and extends the literature in three main ways. First, we introduce kernel-based estimators of integrated variance that are corrected for staleness, thereby delivering robustness under the simultaneous presence of additive and zero returns. Second, we adapt jump inference procedures to incorporate these kernel-corrected volatility measures, ensuring that jump detection remains valid in realistic high-frequency settings. Third, we evaluate the finite-sample performance of these procedures through Monte Carlo simulations, embedding staleness explicitly in the data-generating process. Our simulations confirm that ignoring both stylized facts leads to severe upward bias in jump detection whereas our corrected methods substantially mitigate these distortions. Finally, we apply our framework to high-frequency data on foreign exchange and cryptocurrency data. Empirically, we find that after correcting for both noise and staleness, (i) the number of detected jumps declines sharply, (ii) the contribution of jumps to quadratic variation is much smaller than previously documented, and (iii) activity index estimates align closely with the presence of a Brownian component in price dynamics. These findings help reconcile the discrepancies in the literature regarding the prevalence of jumps and the nature of volatility dynamics, showing that much of the prior evidence for jump activity is likely an artifact of the resulting bias.

The chapter proceeds as follows, [section 4.2](#) outlines the econometric framework and staleness-robust kernel estimators. [section 4.3](#) discusses the role of additive noise. [section 4.4](#) presents simulation evidence. [section 4.5](#) revisits the contribution of jumps to quadratic variation. [section 4.6](#) applies the corrected jump test to FX and equity data. [section 4.7](#) concludes.

4.2 Model Setup

Our analysis begins under the premise that the efficient log-price process of a financial asset follows a Brownian semimartingale process. Specifically, let X_t denote the efficient log-price process, where $t = 0$ corresponds to the opening of the trading day and $t = 1$ represents the closing. The process is initialized at an initial value X_0 , and its dynamics is

described as:

$$X_t^e = X_0 + A_t + M_t, \quad (4.1)$$

We define A_t and M_t as:

$$A_t := \int_0^t \mu_s ds, \quad M_t := \int_0^t \sigma_s dW_s. \quad (4.2)$$

where W_s is a standard Brownian motion, μ_s represents the drift process, and σ_s is the volatility process. For volatility, we assume that it follows a stochastic process similar to Equation 3.2 described in the previous chapter, where μ_s , μ_s^v , σ_s^v and V_s could be considered as adapted càdlàg bounded process. Equation 4.1, Equation 3.2 represent the general form of a stochastic price and volatility process.

Much of the existing literature (e.g., Barndorff-Nielsen et al. (2008), Varneskov (2017), Christensen et al. (2010)) assumes that price observations within the interval $(0, 1]$ are uniformly spaced over time. This simplification facilitates technical analysis by treating observation times as equidistant fractions of the total sample size, N . Following these studies, we adopt the convention that $t_i = i/N$ for $i = 1, \dots, N$. A more general approach relaxes this assumption, permitting sampling times to be independent of both the price process and market noise, while requiring that the maximum interval between consecutive observations remains sufficiently small. Formally, this can be expressed as $\max(t_i - t_{i-1}) \leq C \frac{1}{N}$, where C is a constant ensuring that the largest time gap is at most proportional to $1/N$. Under this broader condition, standard estimators including realized variance (RV), realized kernels, pre-averaging, and multi-scale volatility measures retain their convergence properties toward integrated variance (IV) (see, e.g., Barndorff-Nielsen et al. (2011), Fan et al. (2012), Kim et al. (2016), Wang and Zou (2010)). When sampling shifts to tick-time, endogenous sampling may arise (Robert and Rosenbaum (2010), Robert and Rosenbaum (2012)). Notably, Li et al. (2014) demonstrated that even under endogenous sampling, the RV estimator maintains low error in the absence of noise and jumps. The general form of the Multipower variation estimator could be defined as:

$$MV(X; r) = \frac{1}{n} \sum_{i=1}^{n-m+1} |\sqrt{n}\Delta_i X|^{r_1} \dots |\sqrt{n}\Delta_{i+m-1} X|^{r_m} \quad (4.3)$$

Since this consist a general from we can obtain the realized variance by setting the powers as $r_1 = 2$ and $r_2 = 1$. In similar fashion we can select the powers as $r_1 = r_2 = 1$ to derive the Bipower variation estimator of Barndorff-Nielsen et al. (2008). The general multipower variation estimator could provide accurate estimates of the integrated variance as long as the sampling frequency remains bellow the 5-minute mark. This is due to both staleness as demonstrated by Kolokolov et al. (2020) but also due to microstructure noise where in

higher frequencies becomes a prevalent stylized feature of the data. When the underlying series is contaminated with staleness, multipower based estimators yield biased estimates of the integrated variance. As shown by [Kolokolov et al. \(2020\)](#), for some Bernoulli i.i.d variates we could explicitly show the bias as shown in [Equation 3.8](#).

To account for a more realistic dynamics that includes staleness, the authors have introduced the staleness robust multipower variation estimator, that is designed to provide accurate estimates of the integrated variance defined as:

$$MV^c = \Delta_n \sum_{i=1}^{N-m+1} |\Delta_{n,i}^{-1/2} \Delta_{ti} X|^{r_1} \dots |\Delta_{n,i+m-1}^{-1/2} \Delta_{ti+m-1} X|^{r_m} \quad (4.4)$$

[Equation 4.4](#), indicates that the bias of the estimator depends on three factors: the first power r_1 , the zero probability p_0 , and the number of powers m . This bias results from two competing effects. The first effect is an attenuation caused when the product of consecutive returns equals zero due to either returns being zero. This creates a downward multiplicative bias. The second effect is volatility inflation that occurs when multiplying nonzero returns. In this case, the first return in the multiplication is computed from price differences where the time interval depends on the preceding sequence of zero returns. This effect can produce either upward or downward bias depending on the power r_1 . Under i.i.d. conditions, this bias takes a specific form. Because only the first return is influenced by prior staleness, the bias depends exclusively on the first power r_1 , while the other powers have no effect. Notably, realized variance ($r_1 = 2$ with $m = 1$) exhibits no bias for fundamental reasons. [Figure 4.1](#) illustrates how the multiplicative bias for the power variation (left panel) and the staleness corrected bipower variation (right panel) changes with the first power parameter, across three staleness probabilities.¹ For power variation, the estimator tends to be asymptotically lower than the true value when $r < 1.8$, and higher when $r > 2$. In the case of bipower variation, the estimator is effectively unbiased when the first power r_1 is close to 3, but shows a downward bias when r_1 is smaller than that. [Figure 4.3](#) shows the bias of the staleness corrected multipower variation estimator under the simultaneous presence of gaussian i.i.d noise and staleness. Similarly to [Figure 4.1](#) and [Figure 4.2](#) we examine the behavior of the estimator under three staleness probabilities (10%, 20%, 30%) and fixed the standard deviation of noise to be equal to 0.015. The results of the simulations indicate that even gaussian i.i.d noise introduces significant bias in the staleness corrected multipower variation estimator. Specifically the estimator is no longer unbiased under any power or staleness probability and the bias has increased significantly.

¹[Equation 4.3](#) is a generalized form of a Multipower variation estimator while [Equation 4.4](#) is the staleness corrected version of [Kolokolov et al. \(2020\)](#). By setting the power r it is possible to obtain the realized variance, bipower variation and tripower variation estimators. The plots have been generated based on a Brownian motion with additive microstructure noise and staleness

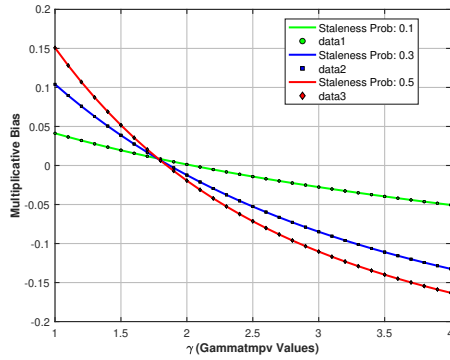


Fig. 4.1 Bias of Power Variation for the power r

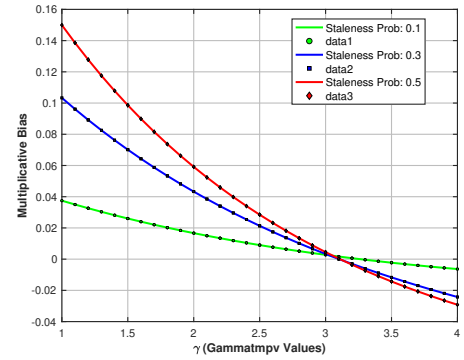


Fig. 4.2 Bias of Bipower Variation for the first power r_1

Furthermore, under the smallest percentage of zeros (10% staleness probability) the bias increases with a considerably higher rate than the rest indicating the sensitivity of the estimator under those conditions. At higher number of zeros the estimator truncates more observations and by doing so, it also smooths some of the noise component leading to a reduced bias under such conditions. Figure 4.4 displays the bias of the staleness corrected

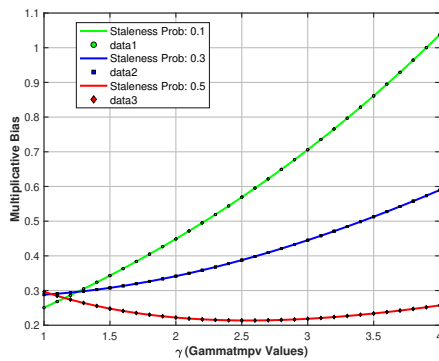


Fig. 4.3 Bias of Bipower Variation under Gaussian i.i.d noise

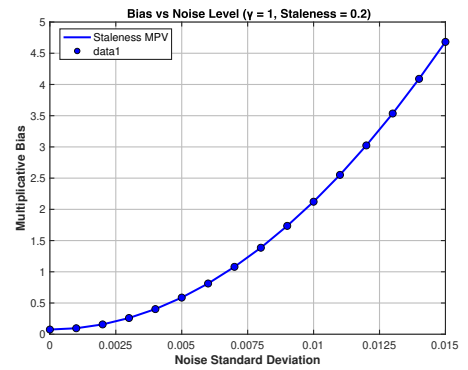


Fig. 4.4 Bias of Bipower Variation as a function of noise standard deviation

bipower variation estimator with the probability of zeros set as constant and equal to 20% , as a function of the i.i.d noise standard deviation. The figure illustrates that as the noise standard deviation increases the bias of the estimator increases exponentially, thus even moderate amounts of microstructure noise could lead to higher than true volatility estimates resulting in spurious jump detection.

This short simulation exercise indicates some useful results about existing estimators of the quadratic variance, especially when it is attempting to make it robust to the stylized features of the underlying process, namely staleness and microstructure noise. We observed that the innovative estimator of [Kolokolov et al. \(2020\)](#) works very well, particularly under

conditions of heavy staleness, without adding methodological complexity or significantly increasing computation time. As shown the estimator is specifically designed to disentangle price jumps from spurious discontinuities caused by stale quotes, a challenge that often undermines standard jump-detection procedures. In contrast to traditional methods, which may either over-reject genuine jumps or mistakenly classify staleness as a jump, their approach provides a more robust discrimination mechanism. This property becomes especially relevant in high-frequency financial data, where staleness is pervasive and can bias volatility and jump inference. By relying on a cleverly constructed approach that remains computationally light, their estimator achieves both accuracy and efficiency, making it superior in practice when dealing with microstructure frictions.

But we have also observed some limitations, namely as noise is introduced in the process the estimator provides biased estimates of the variance. Specifically, we observe that the smaller the staleness probability, the highest the bias the estimator produces under noise.² Furthermore, the sampling frequency also heavily influences the results, higher frequencies lead to higher bias even at low levels of noise and staleness. This behavior along with the lack of alternative estimators that will provide unbiased estimates of the volatility under the simultaneous presence of noise and staleness motivate our work. In the next sections of the study we provide a correction to the staleness multipower variation estimator to improve its distributional characteristics under both sources of contamination.

4.3 Impact of Noise Distribution

In [Figure 4.7](#) we demonstrate the impact of the additive noise distribution on the rejection rate of the BNS test as a function of the staleness probability. For Gaussian noise, the rejection rate remains very low when staleness is minimal. At a staleness probability of 0, the rejection rate is nearly zero, and it only rises slightly as staleness increases to 0.1. However, beyond that point, the rejection rate begins to climb more rapidly, reaching over 80% when the probability of staleness hits 0.5. This shows that even with relatively mild, normally distributed noise, stale data can significantly increase the likelihood of rejection. The behavior under student-t distributed noise is much more severe. Even with no staleness,

²Although the estimator is designed to be robust to price staleness, the presence of additive microstructure noise can generate a different bias pattern. When observed prices are contaminated by i.i.d. noise, the resulting returns include the noise difference. In this setting, lower staleness probabilities imply that a larger fraction of observations contains non-zero underlying returns, so the estimator is applied more frequently to returns that combine signal and noise. Because power or multipower transformations are nonlinear, they amplify the contribution of the noise component, especially at higher powers. By contrast, higher staleness probabilities generate more zero returns, which mechanically dampens the contribution of noise in power-type statistics. As a result, while lower staleness initially produces smaller bias, the noise-induced bias can increase more rapidly as the power of the statistic rises, eventually exceeding the bias observed under higher staleness probabilities.

the rejection rate is already above 40%, which suggests that the heavy-tailed nature of this noise introduces substantial bias. As the probability of staleness rises, the rejection rate increases dramatically, surpassing 70% by 0.15 and reaching almost 100% at 0.5. This pattern indicates that when data is influenced by heavy-tailed noise, even a small amount of staleness can cause the system to reject almost all updates or inputs. Overall, the figure highlights two important insights. First, rejection rates increase with the probability of staleness regardless of noise type, but the rate of increase depends heavily on the noise distribution. Second, Gaussian noise leads to a gradual and predictable rise in rejection, while student-t causes much higher rejection levels and greater sensitivity to staleness. For practitioners this behavior could create issues since there is not always possible to know the empirical distribution of noise in the data, leading to a spurious detection of jumps.

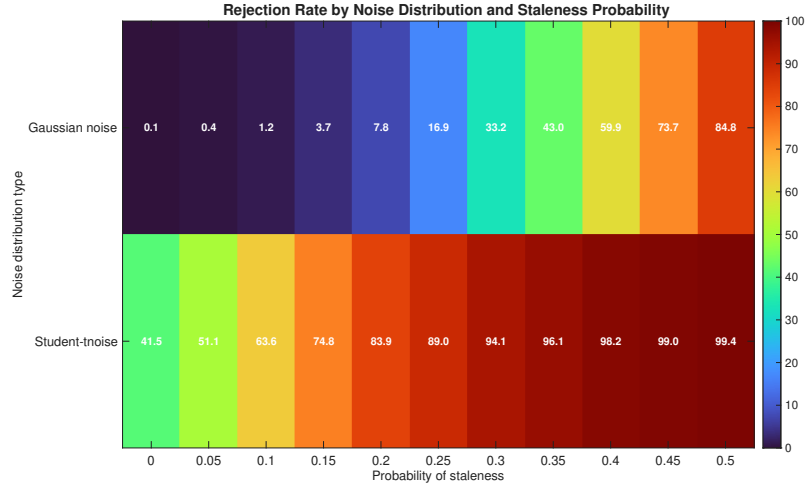


Fig. 4.5 Heatmap displaying the rejection rate of the corrected BNS test across 10 staleness probabilities using the data-generating process of Equation 4.5 under two different assumed distributions for the additive noise.

4.4 Simulation Setup

To evaluate the impact of staleness and noise on jump inference in finite samples we perform a simulation of high-frequency observations of a log-price process X_{t_i} over a trading day of length $T = 6.5$ hours. Our setup extends the classical framework of Andersen et al. (2002) by incorporating staleness and noise, as in modern market microstructure literature. The efficient price Y_t follows a jump-diffusion process with stochastic drift α_t , volatility σ_t , and finite-activity jumps J_t , while the observed price X_{t_i} is contaminated by additive Gaussian noise ε_{t_i} and Bernoulli-driven staleness B_{t_i} , inducing delays with probability p^θ . Observations are sampled on an equispaced grid $\{t_i = i\Delta_n\}_{i=0}^n$ with $\Delta_n = 5$ minutes, reflecting typical high-frequency data settings. This design allows us to experiment with

the effects of staleness and noise not only individually but also under the assumption of the simultaneous existence of both. We assume that the efficient log-price process Y_t follow an Itô semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$:

$$\begin{aligned} dY_t &= \mu_t dt + \sqrt{V_t} dB_t + Z_t dN_t, \\ dV_t &= \kappa(\theta - V_t) dt + \xi \sqrt{V_t} dW_t^V, \end{aligned} \tag{4.5}$$

The process Y_t is governed by four key components, first, the drift term μ_t which is a locally bounded stochastic process that captures the conditional trend in prices, reflecting predictable momentum or mean-reversion effects while ensuring integrability over finite horizons. Second, the volatility process σ_t is *càdlàg* and adapted to the filtration $\mathcal{F}_t$, representing stochastic volatility with possible discontinuities that mimic real-world abrupt changes in market uncertainty. The process σ_t modulates the intensity of the third component B_t , a standard Brownian motion that generates the continuous, erratic fluctuations inherent in financial markets, with its increments providing the Gaussian noise backbone of the diffusion. Finally, the jump process $J_t = \sum_{k=1}^{N_t} \kappa_k$ introduces finite-activity discontinuities, where N_t (a finite counting process for $t < \infty$) governs the arrival of jumps, and κ_k denotes their random sizes. This structure allows Y_t to replicate the abrupt price movements observed in asset prices during news events or liquidity shocks, while maintaining tractability by excluding infinite-activity jump phenomena. To test our proposed correction along with the competing estimators of IV, we incorporate staleness in the model above according to the definition of [Kolokolov et al. \(2020\)](#). Thus the data generating process assumes the form:

$$X_{t_i} = (1 - B_{t_i})(Y_{t_i} + \varepsilon_{t_i}) + B_{t_i}(Y_{t_{i-1}} + \varepsilon_{t_{i-1}}), \tag{4.6}$$

In this manner the likelihood of encountering a zero return is not fixed but exhibits systematic intraday variation. This temporal pattern follows a distinct inverted U-shaped trajectory, peaking during the middle of the simulated trading day while being comparatively lower at both the opening and closing periods. To capture this important empirical regularity, we specify the time-varying probability of zero returns following the authors as:

$$p_t = p^U \cdot \gamma_t^{(p)}, \tag{4.7}$$

In this model specification, p^U denotes the fundamental staleness level that would occur without intraday variations, what can be considered as the unconditional staleness probability. The component $\gamma_t^{(p)}$ acts as a time-varying factor that adjusts this baseline probability throughout the trading day, generating the distinctive inverted U-shaped profile.

The adjustment factor $\gamma_t^{(p)}$ itself is constructed by taking the intraday volatility pattern γ_t from Equation 4.7 and inverting it, then normalizing it by a constant γ_0 :

$$\gamma_t^{(p)} = \frac{\gamma_0}{\gamma_t}. \quad (4.8)$$

The normalizing constant γ_0 plays a crucial role in this specification. It is carefully chosen so that when we average $\gamma_t^{(p)}$ over an entire trading day, this average equals exactly one. This normalization ensures that while the probability of zero returns fluctuates throughout the day, the total daily number of zero returns remains centered around the baseline level p^U .

4.4.1 Microstructure Noise Distortions

The BNS jump test's sensitivity to microstructure noise has been extensively studied in the financial econometrics literature. When prices are contaminated with additive noise, the test statistic becomes systematically biased due to how noise propagates differently through its components. Research by Zhou (1996) and Barndorff-Nielsen et al. (2008) shows that while realized variance (RV) overestimates true volatility by approximately $2n\omega^2$ in the presence of i.i.d. noise $\varepsilon_t \sim (0, \omega^2)$, bipower variation (BV) suffers from a more complex downward bias scaling as $-\frac{2}{\pi}n\omega^2$. This creates an artificial inflation of the jump test's numerator, where the bias magnitude depends on the noise-to-signal ratio $2\omega^2(1 + \frac{1}{\pi}) / \int_0^t \sigma_s^2 ds$.

Staleness introduces additional distortions through return autocorrelation. As demonstrated by Hansen and Lunde (2006), when prices fail to update with probability $p^\emptyset$, the expected bipower variation drops to $(1 - p^\emptyset)^2 \int_0^t \sigma_s^2 ds$ while realized variance remains consistent. This explains why markets with frequent stale pricing, particularly during low liquidity periods exhibit exaggerated jump test statistics. The effect is compounded in the test's denominator, where tripower quarticity becomes severely attenuated. Jacod et al. (2009) establishes that Q_t/BV_t^2 converges to a value below unity under staleness, effectively disabling the variance adjustment in the denominator and further inflating the test statistic.

Empirical studies quantify these theoretical biases in practical settings. Dovonon et al. (2019) finds that for typical equity data sampled at 5-minute intervals with noise magnitude $\omega = 0.001$, the BNS test's actual size can exceed 50% when testing at nominal 5% levels. The distortion follows an approximate relationship $\Phi(-1.96 + 0.28 \cdot n^{1/2} \omega / \int \sigma_s^2 ds)$, where Φ is the standard normal CDF. This formula reveals how the over-rejection probability grows with both sampling frequency (n) and noise magnitude (ω), while being mitigated by integrated volatility. These findings underscore why raw application of the BNS test in

4.5 The Kernel-Smoothed Staleness-Corrected Multipower Variation Estimator

noisy high-frequency environments often produces misleading jump detection results. To account for both sources of bias and improve the finite distributional properties of existing jump tests we proceed to the main contribution of the paper, presenting the kernel-weighted staleness robust multipower variation estimator hereafter called (KS-MPV). The goal is to maintain consistency for integrated volatility powers while mitigating distortions due to microstructure noise and zero returns.

Assuming that the underlying process is contaminated with staleness with probability $p^\theta \in [0, 1)$ and Gaussian i.i.d noise $\varepsilon_{t_i} \sim (0, \omega^2)$ the test is subject to two separate forms of bias as discussed in the beginning of the section. This hypothesis implies an upward bias in the RV estimator due to the presence of noise, which can be viewed as :

$$E[RV_t] \approx \int_0^t \sigma_s^2 ds + 2n\omega^2 \quad (4.9)$$

thus the additive noise contributes a positive bias proportional to the sampling frequency and the noise variance. In contrast the bipower variation estimator experience a downward bias due to two dynamics, first when prices remain stale with probability p^θ , the expected number of nonzero returns shrinks by a factor of $(1 - p^\theta)^2$, attenuating the volatility measure even in the absence of noise. Second, additive noise induces an additional negative bias of approximately $-\frac{2}{\pi}n\omega^2$. Combining both effects yields

$$E[BV_t] \approx (1 - p^\theta)^2 \int_0^t \sigma_s^2 ds - \frac{2}{\pi}n\omega^2 \quad (4.10)$$

Because the realized variance is upward biased and the bipower variation downward biased, the BNS jump test statistic, driven by the ratio $(1 - BV_t/RV_t)$ tends to be systematically inflated, leading to spurious jump detections in the presence of either noise or staleness. To mitigate the bias from both stylized features of the underlying process we introduce the Kernel weighted staleness corrected Multipower variation estimator (KS-MPV).

4.5 The Kernel-Smoothed Staleness-Corrected Multipower Variation Estimator

In this chapter, we employ the kernel-smoothed staleness-corrected multipower variation (KS-MPV) estimator developed and analyzed in Chapter 3. For completeness and to fix notation, we briefly recall its construction and main asymptotic properties. Full technical derivations and proofs are provided in the previous chapter.

4.5.1 Setup and Definition

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,1]}, \mathbb{P})$ be a filtered probability space. Observations are recorded on a regular grid $t_i = i\Delta_n$, $i = 0, \dots, n$, with $\Delta_n = 1/n$. Due to price staleness, only nonzero returns are retained, generating a stochastic grid $\{\tau_0, \dots, \tau_{N_n}\}$ of effective observation times.

Denote by

$$\Delta^{(n,k)} = \tau_k - \tau_{k-1}$$

the durations between successive nonzero returns. The duration-normalized returns are defined as

$$Y_k = \Delta^{(n,k)-1/2} \left(X_{\tau_k}^{\text{obs}} - X_{\tau_{k-1}}^{\text{obs}} \right), \quad k = 1, \dots, N_n. \quad (4.11)$$

To attenuate microstructure noise and mitigate boundary effects, we introduce kernel smoothing. Let $k : [0, 1] \rightarrow \mathbb{R}$ be a bounded kernel satisfying

$$\int_0^1 k(u) du = 1, \quad \int_0^1 k(u)^2 du < \infty.$$

Throughout, we employ the Parzen kernel defined by

$$k(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & |u| \leq 1/2, \\ 2(1 - |u|)^3, & 1/2 < |u| \leq 1, \\ 0, & |u| > 1. \end{cases} \quad (4.12)$$

Let H_n be a bandwidth sequence satisfying

$$H_n \rightarrow \infty, \quad \frac{H_n}{N_n} \rightarrow 0.$$

The kernel-smoothed returns are given by

$$\tilde{Y}_k = \frac{1}{\sqrt{H_n}} \sum_{h=0}^{H_n} k\left(\frac{h}{H_n}\right) Y_{k-h}, \quad k = H_n + 1, \dots, N_n. \quad (4.13)$$

The normalization $1/\sqrt{H_n}$ ensures that $\text{Var}(\tilde{Y}_k) = O(1)$ while reducing the contribution of microstructure noise.

For $m \geq 1$ and powers $r_1, \dots, r_m \in (0, 2]$, define $r^+ = \sum_{j=1}^m r_j$. The KS-MPV estimator is

$$\text{MPV}_n^{KS}(r) = \frac{1}{N_n} \sum_{k=H_n+1}^{N_n-m+1} \prod_{j=1}^m \left| \tilde{Y}_{k+j-1} \right|^{r_j}. \quad (4.14)$$

4.5.2 Asymptotic Properties

Under the assumptions stated in the [section 3.4](#), if the underlying process is free of microstructure noise, the estimator is consistent and satisfies:

$$\text{MPV}_n^{KS}(r) \xrightarrow{p} \left(\prod_{j=1}^m \mu_{r_j} \right) \int_0^1 \sigma_s^{r^+} ds, \quad \text{MPV}_n^{KS}(r) = O_p(N_n^{-1/2}). \quad (4.15)$$

In the presence of i.i.d. microstructure noise and with bandwidth choice

$$H_n \sim cN_n^{1/2},$$

the centered estimator satisfies

$$\text{MPV}_n^{KS}(r) - \mathbb{E}[\text{MPV}_n^{KS}(r)] = O_p(N_n^{-1/4}), \quad (4.16)$$

which reflects the optimal trade-off between smoothing bias of order H_n/N_n and variance reduction of order $1/H_n$.

These results establish the consistency and noise-robust convergence rate of the KS-MPV estimator. In the remainder of this chapter, we build on these properties to develop the new results presented below. In previous chapter ([section 3.4](#)) you can find more details about the construction of the estimator.

4.5.3 Tuning Parameters

As is evident from the outline of the test procedures, each approach requires decisions to be made regarding the distinction between a continuous increment ($\sqrt{V_t}dW_t$) from a discontinuous increment (dJ_t). These ‘tuning’ parameters fundamentally affect the test outcome, influencing both the number and magnitude of identified jumps, as well as any inferential statistics that rely upon them. In this section, we examine the implications of alternative selections for the significance level, kernel specification, and other key tuning values and provide justification for the choices made.

Significance level

All jump detection tests necessarily depend on the choice of a significance level, which determines the critical value beyond which the null hypothesis of “no jump” is rejected. Although various studies have suggested specific, typically conservative, significance levels ([Andersen et al. \(2007b\)](#), [Dumitru and Urga \(2012\)](#)), there is no broad consensus in the literature on an optimal choice. It is important to note that a higher significance level

4.5 The Kernel-Smoothed Staleness-Corrected Multipower Variation Estimator

naturally increases the probability of identifying apparent “jumps,” including those of small magnitude. Conversely, if the analysis aims to capture only substantial discontinuities, a smaller significance level should be used. In our empirical analysis, we report results for a significance level of 1% since our aim was to illustrate the robustness of the corrected estimator and the resulting jump test. The outcomes for an alternative significance level of 5% are consistent with those obtained with the 1% but where not mentioning them here to conserve space, but there are available upon request.

Sampling Interval

The frequency at which intraday returns are sampled substantially affects jump test performance. There is general agreement in the literature that the estimation of quadratic variation (QV) and integrated variance (IV) is most reliable at approximately five-minute intervals, which balance the trade-off between temporal precision and the contamination of estimates by market microstructure noise ([Bandi and Russell \(2008\)](#)). Whether this optimality extends to jump detection, however, remains less clear. The sampling frequency impacts the ASJ and ASJL tests both through the choice of Δ_M , the interval length, and through the tuning constant k , which defines the centering of the statistic under the null. Consistent with prior literature, we adopt $k = 2$ and evaluate test performance over three alternative intraday sampling frequencies to illustrate the influence of this parameter.

Kernel Choice

In addition to the preceding parameters, kernel selection represents another important tuning dimension, particularly for kernel-based estimators of the IV. The kernel function determines how autocovariances of high-frequency returns are weighted, directly influencing the degree of bias correction for microstructure noise, the convergence rate and variance of the estimator. Common kernels include the Bartlett, Parzen, and Tukey–Hanning, each offering distinct bias–variance trade-offs. A smoother kernel (e.g., Parzen) typically yields more stable variance estimates but may attenuate sharp jumps, while more localized kernels (e.g., Bartlett) provide better sensitivity to discontinuities at the cost of increased small-sample variability.(see [Barndorff-Nielsen et al. \(2009\)](#), [Barndorff-Nielsen et al. \(2008\)](#)) In our analysis, we use the Parzen method since it provides better consistency across the different assumed distributions of additive noise used in this study and manages to provide lower type I error across most simulated conditions. We report below the impact of the choice of the kernel on the estimators performance and as a result to the jump identification capability of the jump test. We employ the epanechnikov kernel as:

$$k(u) = \frac{3}{4}(1 - u^2) \quad |u| \leq 1 \quad (4.17)$$

4.5 The Kernel-Smoothed Staleness-Corrected Multipower Variation Estimator

One key consideration about the correction the paper proposes to the multipower variation estimator is the choice of a kernel weighting scheme. Just as the realized variance estimator is sensitive to the choice of the threshold, our proposed estimator is similarly affected by the selection of the kernel. Different kernels apply distinct weighting methodologies, which can significantly influence the estimator's behavior in the presence of noise. [Figure 4.6a](#) illustrates how different kernel functions assign weights to observations in a time series, with the newest data receiving the highest weight and older observations gradually discounted. This weighting mechanism is central to kernel smoothing techniques used in high-frequency financial econometrics, where microstructure noise, such as bid-ask bounce, discrete price movements, and asynchronous trading, can severely distort volatility estimates. Each of the kernels shown in the figure, Triangular, Exponential, Gaussian, Epanechnikov, and Parzen, has a distinct philosophy of smoothing. The Triangular kernel applies a linear decay, offering simplicity but risking abrupt cutoffs that may overlook subtle patterns. The Exponential kernel decays rapidly, emphasizing recent data and capturing short-term shifts, yet its aggressiveness can amplify transient noise. The Gaussian kernel, with its smooth bell-shaped curve, provides excellent noise suppression but may lag in responsiveness, especially during market turbulence. The Epanechnikov kernel, optimal in a mean-squared error sense, balances efficiency and bias but still imposes a sharp boundary. In contrast, the Parzen kernel stands out for its gradual, symmetric decay and compact support, offering a more refined balance between responsiveness and stability. Its smooth tapering of weights helps preserve underlying signal structure while effectively dampening noise, making it particularly well-suited for volatility estimation in noisy environments. As highlighted in the literature by [Barndorff-Nielsen et al. \(2008\)](#) and [Kristensen \(2010\)](#), kernel-based estimators, especially when paired with pre-averaging can mitigate microstructure distortions, and the Parzen kernel's shape lends itself to lower estimation error and greater robustness. Ultimately, the choice of kernel has a direct impact on the quality of inference, and the Parzen kernel offers a compelling advantage for practitioners seeking reliable, that can provide unbiased estimates of the volatility in noisy high frequency data.

[Figure 4.6b](#) illustrates the impact of kernel selection on the performance of the jump test in the presence of both noise and staleness. The figure shows the rejection rate of the test statistic under the assumption that jumps are absent. With this assumption the rejection rate of the Null should remain close to the significance level (i.e in this case close to 1%). Any higher rate indicates a false rejection of the null (type 1 error) and indicates a difficulty of test in disentangling the jump part from the continuous one. Specifically, we apply the BNS tests using the data-generating process outlined in [Equation 4.5](#), incorporating additive noise with a constant variance of 0.015 across ten staleness probability levels. At low staleness levels, the difference between the two kernel variants is minimal. However, as the

proportion of zeros in the data increases, this behavior shifts. Beginning at approximately 30% staleness, the divergence between the kernels becomes more pronounced, reaching a maximum difference at 40%, with the Parzen version of the test showing a 20% lower reduction rate.

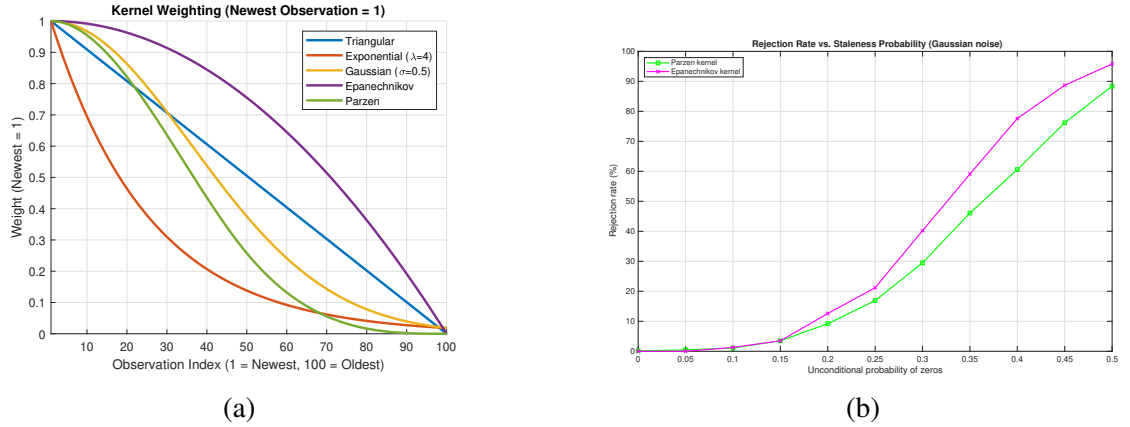


Fig. 4.6 Comparison of kernel plots: (a) Kernel example, (b) Kernel comparison.

4.6 Jump Inference

4.6.1 Correcting Existing Jump Tests

This chapter examines how the simultaneous presence of market microstructure noise and price staleness affects jump inference. To that end, it takes advantage of the kernel-weighted multipower variation estimator specified in Equation 4.14, designed to remain robust to both stylized features and applying it to the jump testing methodology of Barndorff-Nielsen and Shephard (2006) (BNS test). The analysis considers three specifications, the original BNS jump test, its staleness-corrected variant by Kolokolov et al. (2020), and the proposed correction of it. By evaluating these methods under the data generating process described in Equation 4.5 introducing the time-varying staleness (see Equation 4.7 and Equation 4.8) and by considering additive Gaussian i.i.d noise, the chapter aims to clarify how these combined distortions influence rejection behavior and whether the proposed correction improves the performance and finite distribution properties of the test under the influence of additive noise and zeros ³.

³Equation 4.18 shows the corrected test statistic for the BNS test of Barndorff-Nielsen and Shephard (2004) using the equation 4.10

$$\text{BNS} = \frac{1 - \frac{\text{MV}(X;[1|1])}{\text{MV}(X;[2])}}{\sqrt{\frac{\theta}{n} \max\left(1, \frac{\text{MV}(X;[4/3|4/3|4/3])}{\text{MV}(X;[1|1])^2}\right)}} \quad (4.18)$$

Figure 4.7 demonstrates the rejection rate of the three versions of the BNS test as a function of the staleness probability. In Figure 4.7 the original BNS test starts at 4% and rapidly increases to a 100% rejection rate at a 5% staleness level, while the alternative versions provide reduced rates. At 20% staleness level the the second version of the BNS rejects the null around 38% of the time, while the correction proposed in here reduces that to 10%. This difference extends up to 50% staleness. By assuming a student-t distribution for the additive noise in Figure 4.7, the corrected BNS test seems to better illustrate the potential in reducing the false rejection of the null.

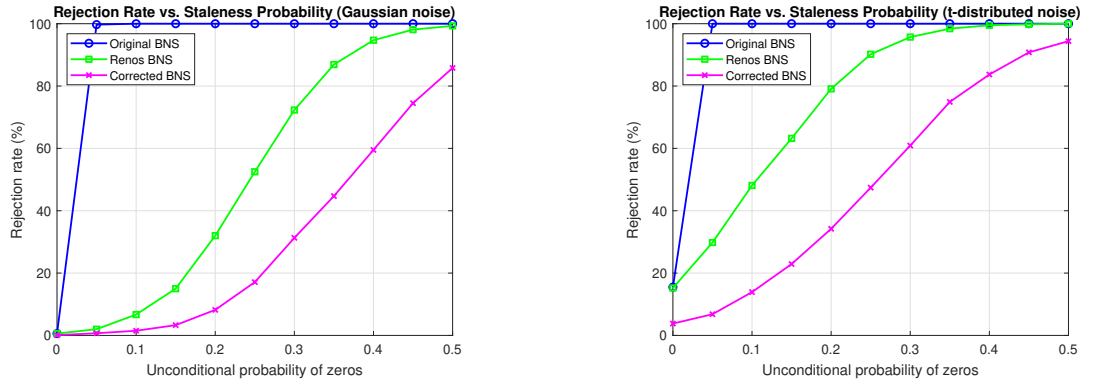


Fig. 4.7 Rejection Rate of the BNS test as a function of the staleness probability

Finally, the Figure 4.8 presents the rejection rates of two jump detection procedures, the staleness robust BNS test shown in Kolokolov et al. (2020) and the corrected BNS test this study proposes with the significance level set to 1% across varying probabilities of staleness and three different sampling frequencies, $N = 78$ which corresponds to a comparable 5-minute sampling, $N = 1000$ which corresponds to sampling every 25 seconds, and finally $N = 23400$ which corresponds to sampling every second. Since the nominal rejection rate is 1%, any systematic increase reflects distortions caused by stale observations and noise. A clear and consistent pattern emerges across all sampling frequencies, the corrected BNS test exhibits greater robustness to competing methodologies in this setting. Across all sampling frequencies the corrected version displays significant lower rejection rate, indicating fewer false positives and a more stable performance under the assumption that the logarithmic returns series is contaminated with additive noise and zeros. At the lowest sampling frequency ($N = 78$), both tests behave relatively well, but corrected BNS remains the more conservative of the two. As the sampling frequency increases to $N = 1000$, sensitivity to staleness becomes more pronounced, with the staleness BNS test exhibiting a markedly

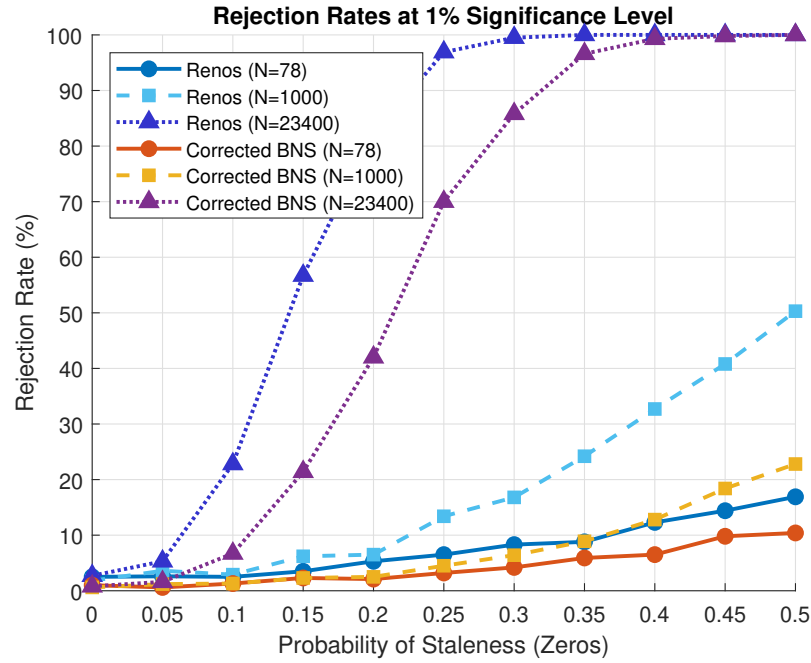


Fig. 4.8 BNS test of [Kolokolov et al. \(2020\)](#) and Corrected BNS test rejection rate as a function of the staleness probability

faster rise in rejection rates than Corrected version. The disparity becomes most striking at the highest sampling frequency ($N = 23400$), where the staleness test deteriorates rapidly, even moderate levels of staleness propel its rejection rate above 50%, and it reaches rejection levels close to that of the classical BNS procedure soon after. Although the corrected BNS also becomes more sensitive at this dense sampling frequency, it rises more gradually, maintaining significantly lower rejection rates relative to the staleness methodology until very high number of zeros levels are reached. Overall, the figure highlights that the simultaneous presence of additive noise and staleness compromises both tests, but the correction proposed by this study seem to offer some advantages in this regard, specifically in reducing on average the type I error of the BNS Jump test by 20%.

4.6.2 Alternative Jump tests

To complement the results show in the previous sections of the study we perform a further Monte Carlo experiment to visualize the distribution of several recognized jump tests against the corrected version of the BNS test proposed ([Equation 4.18](#)). The data generating process is the one described in [Equation 4.5](#) with a Gaussian i.i.d additive noise and a constant 10% staleness. Along with the two version of the BNS test show during the previous sections of the study, we report in [Figure 4.9](#) the distribution of the test statistic for some competing methodologies of jump identification, namely we have considered the

contributions of [Andersen et al. \(2007b\)](#) (ADS test)⁴, [Aït-Sahalia and Jacod \(2009\)](#) (ASJ test), and the [Jiang and Oomen \(2008\)](#) (JO test).

The bipower variation in the BNS test compares realized variance, which captures both diffusive and discontinuous variation, with bipower variation, which is consistent for the continuous component. Its statistic is constructed as a ratio of these measures. In the simulated series, the presence of zero returns depresses bipower variation (the denominator), while Gaussian noise inflates realized variance (the numerator). This combination produces a statistic with inflated variance and heavier empirical tails, explaining the size distortions observed in the plot. The [Jiang and Oomen \(2008\)](#) (JO) test exploits the variance swap replication identity, where the payoff equals half the integrated variance under continuity. Its statistic is based on the discrepancy between proportional and log returns. In the simulation, Gaussian noise perturbs realized variance upward, while staleness reduces the informativeness of log returns. The resulting replication error is systematically biased, yielding a distribution that diverges from the standard normal and exhibits poor size control. The [Aït-Sahalia and Jacod \(2009\)](#) (ASJ) test uses ratios of multipower variations across different sampling frequencies to distinguish continuous variation from jumps. By construction, it is more robust to noise than BNS or JO, but the presence of zero returns at fine scales dilutes the multipower variation, skewing the ratio. This explains why the ASJ statistic in the simulation is closer to normality but still exhibits finite-sample miscalibration. The [Andersen et al. \(2012\)](#) (ADS) test incorporates filtering and smoothing techniques that dampen the impact of noise and staleness. Its construction makes it less sensitive to the distortions present in the simulated data, which is reflected in the plot by a distribution that aligns more closely with the standard normal, with improved control of test size. Finally, the corrected version of the BNS test explicitly adjusts for the biases introduced by noise and zero returns. By stabilizing the denominator against staleness and mitigating noise-induced inflation in the numerator, the corrected statistic restores proper standardization. This is evident in the simulation, where its distribution overlaps the standard normal almost perfectly, with mean near zero, variance close to one, and quantiles matching theoretical expectations.

4.6.3 ABD jump test

Along with the BNS jump test of [Barndorff-Nielsen and Shephard \(2004\)](#) we employ our proposed methodology for handling the simultaneous existence of noise and staleness to

⁴We implement the ADS test following the methodology outlined by [Andersen et al. \(2012\)](#), specifically using their sixth equation to estimate integrated quarticity (IQ) via the MedRQ estimator. For the ASJ test, we adopt the authors' recommended configuration by setting the power variation parameters to $p = 4$ and $k = 2$. Since our null simulations exclude jump components, we compute the variance of the ASJ test statistic without trimming extreme returns. The PZ test is conducted with $p = 2$

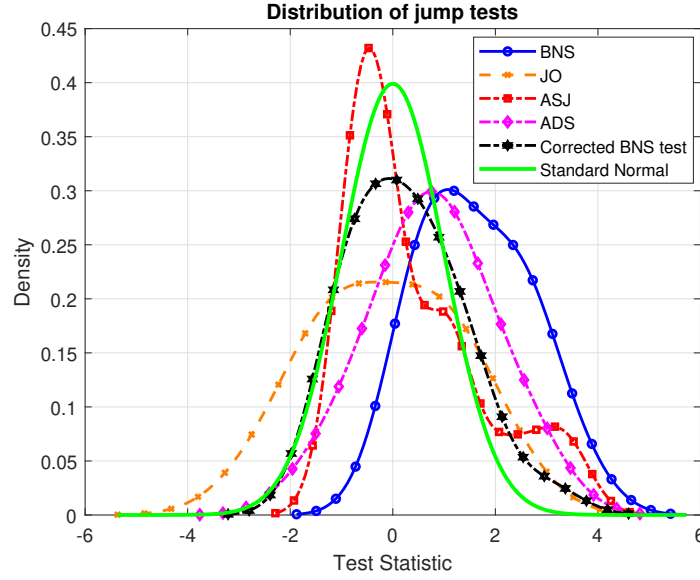


Fig. 4.9 Distribution of test statistics from simulated daily samples at 5-minute intervals ($n = 78$ returns), based on the ADS, ASJ, BNS, corrected BNS, and JO tests.

the ABD test of [Andersen et al. \(2012\)](#). Testing for the presence of a jump in the interval (t_{i-1}, t_i) the authors defined the test statistic as:

$$ABD_i = \frac{\Delta_i X_t}{\sqrt{M^{-1}BV_t}}, \quad i = 1, \dots, M, \quad (4.19)$$

where σ_t denotes the bipower variation, used for the estimation of the instantaneous volatility.

Instead of deriving price jump test statistics from various volatility measures, an alternative approach is to examine the behavior of appropriately standardized returns directly. Under the assumption that asset prices follow a Brownian motion, returns calculated over a selected time interval and scaled by the square root of a consistent estimator of the integrated variance should follow a standard normal distribution asymptotically provided there are no price jumps. Two tests leveraging this principle were introduced by [Andersen et al. \(2007b\)](#) and [Lee and Mykland \(2008\)](#), referred to as ABD and LM, respectively. ABD apply multiple two-sided tests to standardized returns throughout the trading day, whereas LM propose a one-sided test based on the largest absolute standardized return. In this section, we apply the staleness-noise corrected multipower variation estimator to the ABD procedure to observe how the noise corrected estimator impacts the finite distribution of the tests and whether or not it can improve jump inference for procedures other than the BNS test that has previously shown. [Figure 4.10](#) shows the distribution of ABD and corrected ABD under the assumption that the underlying process is contaminated with Gaussian i.i.d noise with 0.015 standard deviation. The figure on the left shows the finite

distribution of the tests under Gaussian i.i.d noise and a constant staleness of 10%. On the right the simulation setup remains the same but the assumed distribution for the noise changes to a student-t distribution. The figures indicate that the correction done to the jump tests using the KS-MPV estimator seems to provide some benefits in this scenario by correcting the distribution of the test and bringing it closer to the asymptotic one.

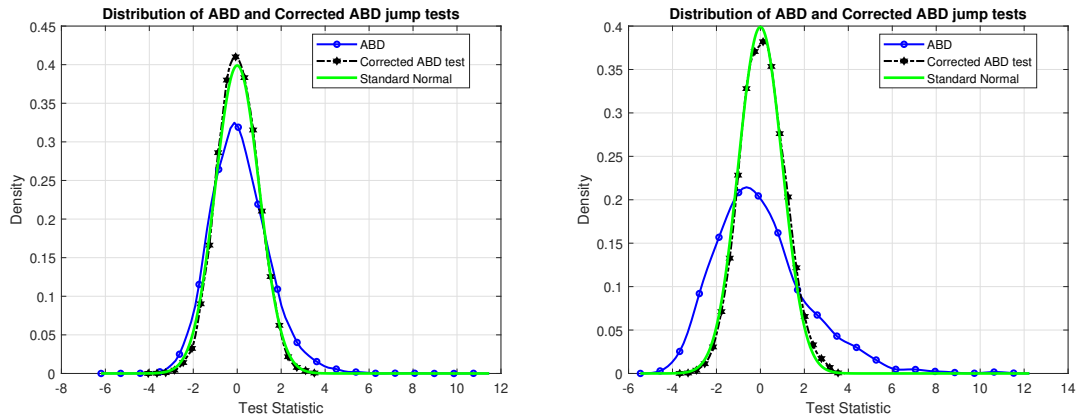


Fig. 4.10 Distribution of ABD and corrected ABD for (a) Gaussian i.i.d. additive noise and (b) t-Student additive noise.

4.7 Empirical Application

We conduct an empirical analyses to assess the performance of the corrected test in equity, foreign exchange and cryptocurrency markets using high-frequency transaction data. We focus on three forex pairs, EUR/USD, GBP/USD and JPY/USD over the period 2021 to 2025. Prices are sampled at 1-minute frequency, yielding 1440 observations per trading day for a total of 900 days. For the equity market we use the IBM stock from November of 2015 to November of 2025 with 390 observations per day. Transaction prices are recorded at the 1-minute frequency and aggregated to 5 and 10-minute grids following standard filtering procedures to remove outliers, bad ticks, and periods of zero volume. After truncating days with insufficient trading activity, the dataset comprises more than two million daily time series, ensuring robustness in inference. To retrieve the data we use the HistData for foreign exchange at minute sampling frequency, for the equity we use the Kibot database and lastly for the cryptocurrency market we use the Cryptodatadownload platform. Unfortunately due to data availability we are restricted and could not expand further our dataset. Nonetheless we believe that the existing dataset is sufficient to derive conclusions regarding the empirical distribution of the tests and to confirm the simulation results provided in the previous sections. Our dataset is sufficient

to provide empirical evidence of the bias induced in jump tests due to the simultaneous presence of microstructure noise and staleness.

Table 4.1 Rejection rate of the jump tests at 5% significance level under three sampling frequencies

	BNS	JO	ASJ	ADS	Corrected BNS
EUR/USD					
1 min	0.66848	0.34891	0.22065	0.48913	0.12283
5 min	0.24674	0.27065	0.19674	0.19239	0.052174
10 min	0.2087	0.25978	0.18696	0.15652	0.051087
GBP/USD					
1 min	0.65188	0.32594	0.16962	0.41131	0.10976
5 min	0.2439	0.24612	0.21951	0.16851	0.080931
10 min	0.1929	0.22395	0.2439	0.12528	0.052106
JPY/USD					
1 min	0.7617	0.42873	0.24266	0.62024	0.19478
5 min	0.31121	0.30903	0.2111	0.25245	0.062024
10 min	0.21545	0.26986	0.16975	0.16757	0.055495
IBM					
1 min	0.31683	0.18203	0.14397	0.19486	0.093994
5 min	0.20128	0.24897	0.19074	0.16323	0.11004
10 min	0.17928	0.30491	0.24484	0.15314	0.13159

Table 4.1 shows the rejection rate of the assumed jump tests under three sampling frequencies, of 1 minute, 5 minutes and 10 minutes for all of the empirical data used in this section. As has been shown in previous literature ([Theodosiou and Zikes \(2011\)](#), [Dumitru and Urga \(2012\)](#)) the tests size is inconsistent among the different procedures and across the three sampling frequencies. Furthermore the size of the tests reduces as the frequency reduces, for example the BNS test goes from a 66.8% rejection rate to 20.8% from 1 minute to the 10-minute for the EUR/USD pair. There are however some exceptions to the rule, namely the ASJ in the GBP/USD pair increases its size from 16.9% at 1-minute to 24.3% at the 10-minute. In terms of jump inference at the 1-minute frequency, the BNS tests seem to provide the largest size with an average rejection rate of 69% for the forex and 31.6% for the IBM stock. The ADS test comes second with an average rejection rate of 50.6% on the forex and 19.4% on the IBM stock. The best performance in terms of size is observed in the ASJ test with an average size of 21% for the forex and 14.3% for the IBM stock. The lowest size is been documented by the corrected BNS test, showing a 14.1% rejection rate for the forex pairs on average and a 9.3% rejection rate on the IBM stock at the same sampling frequency. The tests behavior showcase that jumps as it has been mentioned in the [Kolokolov et al. \(2020\)](#) are rare events and the increase detection most procedures

indicate is an artifact of the bias induced by both the existence of microstructure noise and zeros present in financial data when observed in high frequency.

To compliment our empirical analysis, we examine how staleness impacts the rejection rate of the BNS test. Figure 4.11 presents a scatter plot showing the frequency with which the BNS test rejects the null hypothesis of no jumps, plotted against the percentage of zeros in the return series. The results reveal a clear positive relationship: as staleness increases, so does the likelihood of spurious jump detection. This pattern holds consistently across all forex pairs, reinforcing our earlier findings and highlighting the sensitivity of jump tests to stale data.

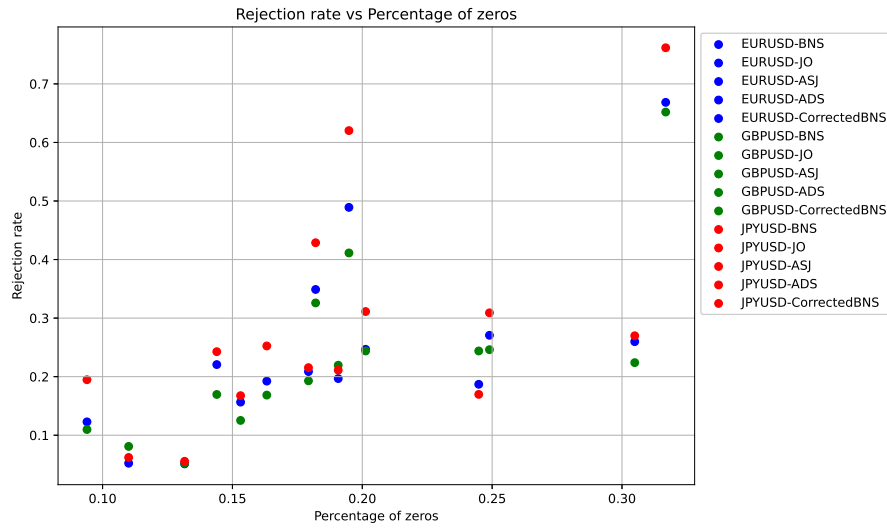


Fig. 4.11 Rejection rate BNS tests as a function of staleness

4.7.1 Estimation under Noise

In this section we seek to uncover the relationship between the microstructure noise present in our empirical dataset and the rejection rate of the jump tests considered in this study along with our proposed correction for the BNS test. To that end we use the Aït-Sahalia and Yu (2008) to estimate the Daily equal-weighted average of microstructure noise α as denoted in their paper. Under these assumptions, the variance of the microstructure noise can be consistently estimated by the equally-weighted intraday return variance,

$$\hat{a}_t^2 = \frac{1}{2(N_t - 1)} \sum_{i=2}^{N_t} (\Delta \log P_{t_i})^2, \quad \hat{a}_t = \sqrt{\hat{a}_t^2}, \quad (4.20)$$

where $\Delta \log P_{t_i} = \log P_{t_i} - \log P_{t_{i-1}}$ denotes the intraday log-return. This estimator is unbiased under the standard assumptions of no serial correlation in the noise and independent

increments of the efficient price at the intraday frequency. To demonstrate the magnitude of microstructure noise relative to total observed price variation, we compute the noise-to-signal ratio (NSR):

$$\text{NSR}_t = \frac{\hat{a}_t^2}{\sum_{i=2}^{N_t} (\Delta \log P_{t_i})^2}, \quad (4.21)$$

which represents the fraction of observed variation attributable to microstructure effects rather than fundamental price movements.

We evaluate the presence of discrete jumps in the efficient price using the BNS jump test, computed daily from intraday returns sampled at minute frequency. For each day, we register a return as jump is registered if

$$|\text{BNS statistic}_t| > 2.57, \quad (4.22)$$

corresponding to a 1% significance level under standard asymptotic theory.

To examine the temporal interplay between microstructure noise and jump occurrences, we construct a combined figure for each dataset, where the daily noise level $\hat{a}_t$ is plotted on the left axis (black line), the daily BNS test statistic is plotted on the right axis (blue line). Horizontal dashed lines mark the ± 2.57 critical values used for jump detection. This representation allows a direct comparison of days with elevated microstructure noise against days exhibiting statistically significant jumps. It facilitates an assessment of whether spikes in observed noise coincide with increased detection of jumps, which may indicate that microstructure frictions contribute to the spurious identification of jumps in high-frequency data.

Figure 4.12 shows the noise and tests statistic for the BNS and corrected BNS tests for the sample period from 2021 to 2025. The times that the tests statistic is above or below the red dashed line signifies the rejection of the null and the identification of a jump in that day. The figure illustrate that increased levels of noise will lead to a higher jump count in the original BNS tests. Specifically during the period January to July in 2023 most of the values of the test statistic are above the 1% critical value and have identified those days as jumps. When shifting to the corrected approach accounting for both the microstructure noise and staleness, we observe a similar behavior but with considerably lower percentage of the series outside the critical values signaling the considerably lower rejection rate of the test compared to its previous version. This behavior has also been demonstrated in Table 4.1 but we believe that the figure paint a clear representation of the effects on jump inference of ingoing the stylized features of noise and staleness.

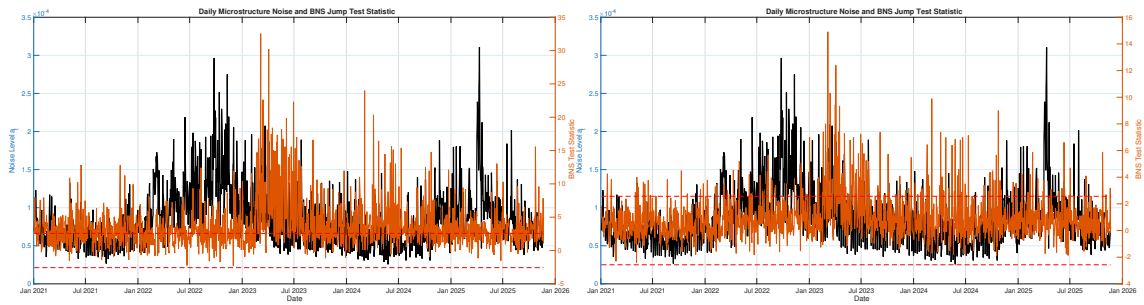


Fig. 4.12 BNS tests (left) and corrected BNS test (right) for the EUR/USD pair from 2021 to 2025. The black line shows the noise estimates, and the orange line shows the test statistic values, with a red dashed line marking the 1% significance critical value. Noise estimation and application of the jump tests are performed at a daily frequency.

4.8 Conclusions

By ignoring zero returns and market microstructure noise in financial data can lead to misleading conclusions about the frequency, magnitude, and role of jumps in price dynamics. By introducing a correction that explicitly accounts for the joint presence of market microstructure noise and staleness, we restore the statistical reliability of two jump tests. Applied to foreign exchange data and IBM stock, the corrected estimator produces results that differ considerably from prior studies about jump inference showcasing that jumps are more rare events than previously thought.

Our results are consistent with standard stochastic volatility models driven by Brownian motion, and indicate that pure jump components may play a less prominent role in explaining the observed dynamics. More broadly, we argue that staleness combined with noise must be incorporated into the assumptions of high-frequency data generation, since ignoring these frictions can distort not only multipower estimators but also general econometric inference.

Additional theoretical work is also needed on modeling rounding, as highlighted by Li and Mykland (2015). While the estimators we study perform adequately unless rounding is excessive, developing a more effective correction remains desirable. Our heuristic adjustment helps mitigate the impact of aggressive rounding, but the challenge of estimating multipower variation under the simultaneous presence of rounding, noise, and staleness remains unresolved. Nevertheless, our results show that properly understanding these frictions in high-frequency prices can yield more precise measurements and stronger tests of price dynamics.

References

- Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. *The Journal of Finance* 57, 2075–2112.
- Aït-Sahalia, Y., 2004. Disentangling diffusion from jumps. *Journal of financial economics* 74, 487–528.
- Aït-Sahalia, Y., Jacod, J., 2009. Testing for jumps in a discretely observed process. *The Annals of Statistics* , 184–222.
- Aït-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. *Journal of Econometrics* 168, 207–222.
- Aït-Sahalia, Y., Yu, J., 2008. High frequency market microstructure noise estimates and liquidity measures. Technical Report. National Bureau of Economic Research.
- Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. *The Journal of Finance* 57, 1239–1284.
- Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review* , 885–905.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007a. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics* 89, 701–720.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andersen, T.G., Bollerslev, T., Dobrev, D., 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics* 138, 125–180.
- Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75–93.
- Bandi, F.M., Kolokolov, A., Pirino, D., Renò, R., 2020. Zeros. *Management Science* 66, 3466–3479.
- Bandi, F.M., Pirino, D., Reno, R., 2017. Excess idle time. *Econometrica* 85, 1793–1846.
- Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies* 75, 339–369.

- Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Podolskij, M., Shephard, N., 2006a. A central limit theorem for realised power and bipower variations of continuous semi-martingales, in: *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*. Springer, pp. 33–68.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–169.
- Barndorff-Nielsen, O.E., Shephard, N., 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, 253–280.
- Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics* 2, 1–37.
- Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics* 4, 1–30.
- Barndorff-Nielsen, O.E., Shephard, N., Winkel, M., 2006b. Limit theorems for multipower variation in the presence of jumps. *Stochastic processes and their applications* 116, 796–806.
- Barunik, J., Vacha, L., 2015. Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *Quantitative Finance* 15, 1347–1364.
- Bauwens, L., Veredas, D., 2004. The stochastic conditional duration model: a latent variable model for the analysis of financial durations. *Journal of econometrics* 119, 381–412.
- Bergstrom, A.R., 1984. Continuous time stochastic models and issues of aggregation over time. *Handbook of econometrics* 2, 1145–1212.
- Black, F., Scholes, M., 2019. The pricing of options and corporate liabilities, in: *World Scientific Reference on Contingent Claims Analysis in Corporate Finance: Volume 1: Foundations of CCA and Equity Valuation*. World Scientific, pp. 3–21.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31, 307–327.
- Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33–65.
- Brownlees, C., Nualart, E., Sun, Y., 2020. On the estimation of integrated volatility in the presence of jumps and microstructure noise. *Econometric Reviews* 39, 991–1013.
- Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G., 2003. Alternative models for stock price dynamics. *Journal of Econometrics* 116, 225–257.

- Christensen, K., Hounyo, U., Podolskij, M., 2018. Is the diurnal pattern sufficient to explain intraday variation in volatility? a nonparametric assessment. *Journal of Econometrics* 205, 336–362.
- Christensen, K., Oomen, R., Podolskij, M., 2010. Realised quantile-based estimation of the integrated variance. *Journal of Econometrics* 159, 74–98.
- Christensen, K., Oomen, R.C., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* 114, 576–599.
- Christoffersen, P., Heston, S., Jacobs, K., 2009. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science* 55, 1914–1932.
- Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159, 276–288.
- Delbaen, F., 2006. *The mathematics of arbitrage*. Springer.
- Dovonon, P., Gonçalves, S., Hounyo, U., Meddahi, N., 2019. Bootstrapping high-frequency jump tests. *Journal of the American Statistical Association* 114, 793–803.
- Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.
- Dumitru, A.M., Urga, G., 2012. Identifying jumps in financial assets: a comparison between nonparametric jump tests. *Journal of Business & Economic Statistics* 30, 242–255.
- Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.
- Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.
- Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association* 107, 412–428.
- Fan, J., Wang, Y., 2007. Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association* 102, 1349–1362.
- Fan, Y., Fan, J., 2011. Testing and detecting jumps based on a discretely observed process. *Journal of Econometrics* 164, 331–344.
- Forsberg, L., Ghysels, E., 2007. Why do absolute returns predict volatility so well? *Journal of Financial Econometrics* 5, 31–67.
- Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24, 127–161.
- Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.

- Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies* 6, 327–343.
- Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. *Journal of financial econometrics* 3, 456–499.
- Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications* 119, 2249–2276.
- Jacod, J., Podolskij, M., Vetter, M., 2010. Limit theorems for moving averages of discretized processes plus noise .
- Jacod, J., Todorov, V., 2010. Do price and volatility jump together? *The Annals of Applied Probability* 20, 1425–1469.
- Jiang, G.J., Oomen, R.C., 2008. Testing for jumps when asset prices are observed with noise—a “swap variance” approach. *Journal of Econometrics* 144, 352–370.
- Jing, B.Y., Liu, Z., Kong, X.B., 2014. On the estimation of integrated volatility with jumps and microstructure noise. *Journal of Business & Economic Statistics* 32, 457–467.
- Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and their Applications* 126, 3527–3577.
- Koike, Y., 2016. Estimation of integrated covariances in the simultaneous presence of nonsynchronicity, microstructure noise and jumps. *Econometric Theory* 32, 533–611.
- Kolokolov, A., Livieri, G., Pirino, D., 2020. Statistical inferences for price staleness. *Journal of Econometrics* 218, 32–81.
- Kolokolov, A., Renò, R., 2024. Jumps or staleness? *Journal of Business & Economic Statistics* 42, 516–532.
- Kristensen, D., 2010. Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* 26, 60–93.
- Lee, S.S., Mykland, P.A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies* 21, 2535–2563.
- Lee, S.S., Mykland, P.A., 2012. Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics* 168, 396–406.
- Li, Y., Mykland, P.A., 2015. Rounding errors and volatility estimation. *Journal of Financial Econometrics* 13, 478–504.
- Li, Y., Mykland, P.A., Renault, E., Zhang, L., Zheng, X., 2014. Realized volatility when sampling times are possibly endogenous. *Econometric theory* 30, 580–605.
- Mancini, C., 2008. Large deviation principle for an estimator of the diffusion coefficient in a jump-diffusion process. *Statistics & probability letters* 78, 869–879.
- Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36, 270–296.

- Mancini, C., Renò, R., 2011. Threshold estimation of markov models with jumps and interest rate modeling. *Journal of Econometrics* 160, 77–92.
- Mancini, C., et al., 2001. Disentangling the jumps of the diffusion in a geometric jumping brownian motion. *Giornale dell’Istituto Italiano degli Attuari* 64, 19–47.
- Maneesoonthorn, W., Martin, G.M., Forbes, C.S., 2020. High-frequency jump tests: Which test should we use? *Journal of econometrics* 219, 478–487.
- Markowitz, H., 1952. Modern portfolio theory. *Journal of Finance* 7, 77–91.
- Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics* 3, 125–144.
- Nowman, K.B., 1997. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *The journal of Finance* 52, 1695–1706.
- Podolskij, M., Vetter, M., 2009a. Bipower-type estimation in a noisy diffusion setting. *Stochastic processes and their applications* 119, 2803–2831.
- Podolskij, M., Vetter, M., 2009b. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps .
- Podolskij, M., Ziggel, D., 2010. New tests for jumps in semimartingale models. *Statistical inference for stochastic processes* 13, 15–41.
- Robert, C.Y., Rosenbaum, M., 2010. On the microstructural hedging error. *SIAM Journal on Financial Mathematics* 1, 427–453.
- Robert, C.Y., Rosenbaum, M., 2012. Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 22, 133–164.
- Samuelson, P.A., Davis, M., Etheridge, A., 2006. Louis Bachelier’s theory of speculation: the origins of modern finance. *JSTOR*.
- Taylor, S.J., 1999. Markov processes and the distribution of volatility: a comparison of discrete and continuous specifications. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 357, 2059–2070.
- Theodosiou, M.G., Zikes, F., 2011. A comprehensive comparison of nonparametric tests for jumps in asset prices. Available at SSRN 1895364 .
- Todorov, V., Tauchen, G., 2011. Volatility jumps. *Journal of Business & Economic Statistics* 29, 356–371.
- Varneskov, R.T., 2017. Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels. *Econometric Theory* 33, 1457–1501.
- Wang, K., Liu, J., Liu, Z., 2013. Disentangling the effect of jumps on systematic risk using a new estimator of integrated co-volatility. *Journal of Banking & Finance* 37, 1777–1786.
- Wang, Y., Zou, J., 2010. Vast volatility matrix estimation for high-frequency financial data .

- Zhang, B.Y., Zhou, H., Zhu, H., 2009. Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *The Review of Financial Studies* 22, 5099–5131.
- Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019–1043.
- Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.
- Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics* 14, 45–52.