Aarhus University

Department of Economics and Business Economics

Master's thesis

Robustness of Rough Volatility (Robusthed af ru volatilitet)

Author:
Sebastian Schmidt-Larsen
(201809155)

Supervisor: Bezirgen Veliyev
Hand in date: 16.09.2024
Assignment can be made public
Number of characters, including tables and figures: 163.800

Contents

1 Introduction **4**

2 Introductory theory **6**
2.1 Spaces and measures . . . 7
2.2 Stochastic processes . . . 9
2.3 Stochastic integrals . . . 11
2.4 Semimartingales . . . 12
2.5 Fractional Brownian motion . . . 14

3 Jumps in financial econometrics **15**
3.1 Lévy processes - building blocks for jumps . . . 15
3.2 Effects of adding jumps . . . 17

4 Introduction to volatility estimation **18**
4.1 IV estimation with no noise . . . 20

5 Microstructure noise **22**
5.1 Microstructure noise models . . . 23
5.2 Noise assumption . . . 24
5.3 Pre-averaging . . . 24

6 Testing for jumps **25**
6.1 Review of the method in Bibinger & Sonntag (2023) . . . 26

7 The fSV model setting **28**
7.1 General jump setup . . . 29
7.2 Moments of the $(IV_t)_{t\in\mathbb{N}}$ process . . . 29
7.3 Jumps in log-price . . . 30
7.4 Jumps in volatility . . . 31
7.5 Note regarding a fSV model with co-jumps . . . 33

8 GMM estimation procedure **33**
8.1 GMM estimation presented in Bolko et al. (2023) . . . 33
8.1.1 Setting up the GMM estimation procedure for the fSV model . . . 34
8.1.2 Asymptotic normality . . . 36
8.2 GMM estimation procedure with jumps in X . . . 39

9 Simulation **41**
9.1 Fractional Brownian Motion . . . 42
9.2 Simulation of the fSV model as defined in Section 2 and in Bolko et al. (2023) . . . 43
9.3 Simulation of the fSV model in the general setup . . . 44
9.4 GMM estimation of replication study . . . 45
9.4.1 Results of replication study . . . 45
9.5 GMM estimation with fewer days observed . . . 45
9.6 Kernel specification in the weight function . . . 46
9.7 Lower number of moments . . . 46
9.8 The sampling frequency of RV . . . 46
9.9 Fractional stochastic volatility model with microstructure noise . . . 47
9.9.1 Practical concerns in calculating initial values . . . 48
9.9.2 Comment on the results . . . 48
9.10 Jump model . . . 49
9.10.1 Truncation approach . . . 49
9.10.2 When using Realized variance on jump model . . . 49
9.10.3 Initial values . . . 49
9.10.4 Results . . . 51

9.11 A note on convergence . . . 51
9.12 Kernel density estimate of standardized H under the different model settings . . . 53
9.13 Jump testing . . . 53

10 Empirical **54**
10.1 Data . . . 54
10.1.1 Data cleaning . . . 54
10.2 Empirical evidence for roughness . . . 55
10.3 Empirical evidence for jumps . . . 55
10.3.1 Jumps in prices . . . 55
10.3.2 Jumps in volatility . . . 57
10.4 Empirical analysis of the fSV model in the presence of price jumps . . . 57

11 Conclusion **59**

A Appendix **63**
A.1 Properties of a Poisson and compound Poisson process . . . 63
A.2 Plots of $\kappa(\ell)$ and $\mathbb{E}[IV_t IV_{t+\ell}]$ . . . 64
A.3 Algorithms . . . 65
A.4 Replication of the two-stage procedure as presented in Gatheral et al. (2022) . . . 67
A.5 GMM with no jumps in X . . . 67
A.6 Figures . . . 69
A.7 Tables . . . 73

B Generalized method of moments **76**
B.1 Identification . . . 77
B.2 Weighting matrix . . . 79
B.3 Consistency and asymptotic normality . . . 80
B.4 Testing over-identifying restrictions . . . 81
B.5 Numerical optimization . . . 81

Abstract

This thesis investigates the robustness of the generalized method of moment (GMM) approach presented in Bolko et al. (2023) for a fractional stochastic volatility (fSV) model. A comprehensive simulation study is performed, and we provide clear evidence that characteristic should be considered when the method is applied, as both microstructure noise and jumps significantly deteriorate the estimation of underlying model parameters. A jump test under rough volatility is presented, and although rare, we show the presence of volatility jumps across the 30 constituents of the Dow Jones industrial average. Introductory theory on probability spaces, measures, random variables, and stochastic processes is presented. We delve into the fundamentals of jump processes, techniques for volatility estimation, and the impact of microstructure noise. A general specification of the fSV model that includes both co-jumps and independent jumps in the fractional Ornstein-Uhlenbeck process and log prices is introduced. We provide a moment structure for the fSV model with the presence of price or volatility jumps and show the technical challenges apparent when GMM is performed on these moments. In the appendix, an algorithm for the simulation of the general fSV with jumps together with algorithms for different aspects of the GMM estimation helps clarify the approach for the numerical implementation.

1 Introduction

Volatility is of substantial significance in the assessment and prediction of financial risk, with applications in asset and derivative pricing, risk management, hedging strategies, and portfolio selection. Volatility is a latent variable, but using a time series of observed prices, it is possible to estimate the variation in asset prices over time. These estimates are mainly based on sums of quadratic variation. In time series observed at low frequencies (i.e., daily or lower), this variation is typically modeled by a time-series model, e.g. ARIMA, GARCH. In this thesis, we examine volatility through the application of high-frequency financial data. High-frequency financial data can be observed at tick-by-tick resolution (which is obtained in milliseconds or even nanoseconds) and provides access to a wealth of information such as price, order size, timestamps, and much more.

At high frequencies, econometricians employ forms of realized variance, i.e. forms of squared returns, as a proxy for volatility. We generally refer to this as realized volatility (RV). In order to model high-frequency financial data, the focus is on the application of continuous-time stochastic models. It is important to note that continuous-time stochastic models do not advocate the need to assume models that have continuous paths.

In the field of mathematical finance, there is a well-established consensus that dynamic financial models are more realistic when jump processes are incorporated. This idea originates from the work of Press (1967). Back (1991) presented the general case of models, which belongs to the class of semimartingale processes, where asset prices can jump, creating a discontinuous path. The early work of Bakshi et al. (1997), Bates (2000) and Pan (2002) incorporates jumps in the volatility process instead. Duffie et al. (2000) were the first (to our knowledge) to consider a double jump specification for jumps in both asset prices and the corresponding volatility. Large movements in asset prices, such as a crash, can be modeled by a jump component in the price process but this have no impact on the future distribution of returns, while jumps in volatility are a persistent (and rapidly moving) factor that drives the conditional volatility of returns (Eraker et al. 2003).

In high-frequency econometrics Barndorff-Nielsen & Shephard (2004) introduced a volatility estimator based on bipower variation, which was shown to be robust to rare jump in the log-price process, and the development of a pre-averaging approach by Jacod et al. (2009), Podolskij & Vetter (2009*a*), Podolskij & Vetter (2009*b*), has made estimators robust to both microstructure noise and jumps. At high frequency, microstructure noise might be due to liquidity effects, bid/ask bounces, misrecording and rounding errors (Barndorff-Nielsen et al. 2009).

The degree of jump variation in asset price data has been documented to be of various sizes within the existing literature. However, Christensen et al. (2014) demonstrated, based on high-frequency data, that the magnitude of the jump component in asset prices was significantly smaller than initially believed. Although the jump component in asset prices empirically is small, it is important from a theoretical point of view. Given this perspective, various cases of processes can be examined: both the price and volatility processes are devoid of jumps, either the price or the volatility contains jumps, independent jumps that occur in the price and volatility, or contemporaneous jumps (co-jumps) occur in both.

Statistical tests for significant co-jumps and disjoint jumps in both prices and volatility were presented in Jacod & Todorov (2010). The occurrence of co-jumps is commonly linked to simultaneous price and volatility changes of opposite signs (Bandi & Reno 2016). This negative correlation between shocks to the variance and shocks to the asset returns can be seen as a leverage effect (Bandi & Renò 2012). For an overview of the important jump literature up until 2014, see Christensen et al. (2014).

The volatility of an asset price is not fixed but stochastic. The early arguments for this include Mandelbrot (1963) and Fama (1963). Stochastic volatility (SV) modeling, as we know it, is based on the early discrete-time model proposed by Clark (1973). Discrete-time models remain prevalent in financial data analysis. However, to handle the growing amount of high-frequency data, such as on asset prices, continuous-time SV models are increasingly employed. In continuous time, the driving distribution for the model describing the log-process is a standard Brownian motion. The standard Brownian motion together with a process for spot volatility constitutes the model. The squared spot volatility process is usually assumed to be log-normal such that it takes a non-negative value and the spot volatility is assumed to have a càdlàg sample path. The other part of the model

for the log-price is the so-called drift process, which describes the trend in the price over time. As mentioned above, the log-price and volatility process can include a third component, namely jumps.

A field of continuous-time SV modeling is fractional stochastic volatility (fSV), where the stochastic process of the volatility is modeled with a fractional Brownian motion (fBm) with Hurst exponent $H \in (0, 1)$ instead of a standard Brownian motion which has $H = 0.5$. The fBm has two attractive properties: Its ability to generate paths that have varying levels of Hölder regularity ('roughness') and its ability to model long-range dependence. These two properties are controlled by the Hurst exponent and are linked through self-similarity (Cont & Das 2022). It should be noted that the fBm is neither a semimartingale nor has independent increments when $H \neq 0.5$, which in turn complicates modeling and estimation. Despite these mathematical complications, there has been a significant increase in interest, in the literature on finance, to incorporate an fBm into models.

Comte & Renault (1996) introduced the fractional stochastic volatility (fSV) model based on a fBm and Comte & Renault (1998) introduced a SV model where the log volatility follows a fractional Ohrstein-Uhlenbeck (fOU) process. Both models assumed $H \in \left(\frac{1}{2}, 1\right)$ and did not have roughness in the sample-path, but instead exhibited the long-memory property. Over the last two decades, the consensus has been that volatility displays this long-range dependence (Wang et al. 2021).

In the same period, theoretical developments in implied volatility modeling considered the case where $H \in \left(0, \frac{1}{2}\right)$, thereby introducing roughness in volatility (see, e.g. Alos et al. (2007) and Fukasawa (2017)). Omar et al. (2016) provided a microstructural foundation for rough volatility. Gatheral et al. (2022) (first published in 2014 as a working paper) subsequently showed empirical evidence of rough volatility in both the DAX and Bund futures contracts and S&P and NASDAQ indices. The empirical evidence was then extensively extended by Bennedsen et al. (2022), who studied data on the E-mini S&P 500 futures contracts and a large number of US equities (almost two thousand). Bennedsen et al. (2022) also introduced a class of continuous-time models using Brownian semistationary processes for the log-normal SV model with a general fractal index. This can simultaneously combine roughness and long-memory characteristics.

Different methods have been proposed in relation to rough volatility estimation. In the case of a log-normal fOU SV model, Wang et al. (2021) propose a two-stage method that involves estimating H using the change-of-frequency estimator and estimating the other parameters using the method-of-moments in the second stage. It should be noted that the two-stage method is not efficient. Cont & Das (2022) proposes a non-parametric method using the concept of normalized p-th variation along a sequence of partitions of data. Bennedsen et al. (2022) proposes an estimator for the roughness index based on nonlinear least squares regression. Fukasawa et al. (2022) proposes a quasi-likelihood estimator based on a Whittle-type approximation to the autocovariance of the fOU process and shows that the adapted Whittle estimator is consistent under high-frequency asymptotics. Bolko et al. (2023) expanded the standard generalized method of moments (GMM) technique in the case of a fractional log-normal SV model to jointly estimate parameters that feature a convenient integral form for moment expressions. Furthermore, they introduce a bias correction that accounts for measurement error, without introducing additional nuisance parameters. Compared to Fukasawa et al. (2022), the GMM estimator was shown to be asymptotic normal.

From the initial stages of SV research, GMM has been widely employed as an estimation technique. Renault (2009) formulate two possible explanations for this. Firstly, moments of a financial time series are used in volatility forecasting. Secondly, its simplicity, since evaluating the exact likelihood function in the context of parametric volatility models is difficult. An idea introduced by Bollerslev & Zhou (2002) is to match the sample moments of realized volatility to the population moments of the integrated volatility, by which we can formulate a GMM-type volatility estimator. In this thesis, we adopt this approach, grounding our estimator formulation on the time series properties of integrated volatility.

Estimation of volatility aside, a fundamental question arises: Do we encounter both roughness in the path of volatility and jumps in both asset prices and volatility? It has been proposed that in some cases there is no need for jumps under roughness (Bayer et al. 2016). Meanwhile, Wang et al. (2021) suggested that a better model for log-RV could be a fOU model with jumps. So far, jump

diffusion models and rough volatility models have been treated mainly in separate applications, and we were only able to find two papers that connect roughness and jumps in a single model (Jaber & Carvalho (2023) and Bolko et al. (2023)). It should be noted that Bolko et al. (2023) treats the case with both roughness and jumps in their supplementary online-only appendix.

Although rough volatility is an active and promising research area, different conclusions and recommendations have been drawn regarding whether volatility can be considered rough or not. An argument against rough volatility is that simpler models can fit observations just as well (Rogers 2023). Another argument follows from the premise that the observed roughness in volatility is the result of estimation errors arising from the microstructure noise characteristic in the process of estimating the realized volatility (Cont & Das 2022). Subsequently, several efforts have been made to address the aforementioned critique (Fukasawa et al. (2022) and Bolko et al. (2023)).

A possible approach to answering the question whether the volatility is rough or not is to include jump components in the model. Shown by Bolko et al. (2023), when including a jump component in the fOU process and performing their GMM estimation procedure (thus using wrongly matched moment conditions), the estimate for H is slightly lower than when the model is correctly specified. We examine a modification of the GMM approach to incorporate moment conditions for models that include jump processes.

This thesis will study the estimation of the fSV model presented in Bolko et al. (2023). First, we will replicate the results in a simulation study. Secondly, we apply the GMM procedure to empirical data, which in our case are the 30 constituents of the Dow Jones Industrial Average, in contrast to Bolko et al. (2023), which focuses on broad stock indices. Thirdly, we perform an extensive number of robustness checks. These include sampling frequency, number of observations, kernel and moment specification, microstructure noise, and jumps in both price and volatility. Building on a framework with jumps and co-jumps has been shown to have empirical evidence (Bandi & Reno 2016). The presence of jumps can be tested by non-parametric methods published in e.g. Barndorff-Nielsen & Shephard (2006), Lee & Mykland (2008) and Aït-Sahalia & Jacod (2009), under the assumption of a semimartingale volatility process. We review a new jump test and jump location method in the fractional model setting, where either the price the volatility process is not a semimartingale; the method is published in Bibinger & Sonntag (2023). We test the method on both simulated and empirical data.

Our results reveal several interesting features of the empirical data. When jumps are present, robust methods must be applied when the main focus is on estimation of volatility. This can be done in several ways. If, on the other hand, one has an interest in estimating the parameters in the jump size distribution, estimation procedures including these parameters must be developed. This thesis will examine two modified GMM procedures in the presence of a jump process. Both modified methods are in an early stage, and we will not formally derive the consistency and asymptotic results, but merely test them on simulated and empirical data. The notation and most of the theorems and assumptions throughout this thesis will follow the same form as in Bakshi et al. (1997) and Aït-Sahalia & Jacod (2014).

The remainder of the thesis is organized as follows. Section 2 contains introductory theory and the model setup, the next section is dealing with jump processes in general, Section 4 is on volatility estimation, then a theoretical section on microstructure noise, and Section 6 gives a review of a jump test in the fractional model setup. The fractional SV model is explained more in Section 7 and the GMM procedure in Section 8. The simulation results and empirical analysis are presented in Sections 9 and 10. The thesis finishes with a conclusion, and two Appendices contain multiple tables, figures, and a general description of GMM.

2 Introductory theory

As stated in the opening section, this thesis focuses mainly on the modeling and estimation of volatility in financial data. Volatility evolves in a random matter over time.

In mathematical finance, the concept of stochastic volatility models is when the variance in the stochastic log-price model is a stochastic process itself, and is not e.g. a constant as in the classic Black-Scholes model. So, volatility is assumed to be a positive stochastic process, often with a mean reverting property.

In order to define, simulate and estimate volatility we need to introduce concepts regarding stochastic processes in time, which are observed via high frequency data at discrete time points. Although volatility is regarded as a latent (unobseved) process in time, it is possible via returns in (log) price to do inference regarding the features that drive the volatility. The model structure studied in this thesis consists of a process for the log-asset price and a process for the spot variance.

We denote the stochastic process for the efficient log-price process $X = (X_t)_{t\geq 0}$ and assume it is an Itô process given by (Bolko et al. 2023):

$$X_t = X_0 + \int_0^t \mu_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s, \quad t \geq 0.$$

X is an adapted continuous time process defined on the filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. Here, the adapted filtration $\mathcal{F}_t$ is the information available at time $t, t \geq 0$. The asset price process is of a semimartingale form as we assume the market to be arbitrage free. The starting value X_0 is $\mathcal{F}_0$-measurable, and the drift $\mu = (\mu_t)_{t\geq 0}$ is predictable. The volatility process $\sigma = (\sigma_t)_{t\geq 0}$ is càdlàg. As usual, $W = (W_t)_{t\geq 0}$ is a standard Brownian motion.

The spot variance process $\sigma^2 = \left(\sigma_t^2\right)_{t\geq 0}$ is assumed to be log-normal and defined as

$$\sigma_t^2 = \xi \exp\left(Y_t - \frac{1}{2}\kappa(0)\right), \quad t \geq 0.$$

The parameter $\xi \in \Xi \subset (0, \infty)$ is the unconditional mean of σ_t^2. The process $Y = (Y_t)_{t\geq 0}$ is assumed to be a zero mean and stationary Gaussian process, the autocovariance function is

$$\kappa(u) = Cov(Y_0, Y_u), \quad u \geq 0.$$

The function κ is parameterized by $\phi \in \Phi \subset \mathbb{R}^p$. Both Ξ and Φ are assumed to be compact sets. We denote $\theta = (\xi, \phi) \in \Theta = \Xi \times \Phi$.

It is not assumed that the volatility is a semimartingale, but is the exponential of Y, which is a fractional Ornstein-Uhlenbeck (fOU) process:

$$Y_t = v \int_0^t e^{-\lambda(t-s)} \mathrm{d}B_s^H, \quad t \geq 0 \tag{1}$$

where $v, \lambda > 0$, and $B^H = \left(B_t^H\right)_{t\geq 0}$ is a fractional Brownian motion with the Hurst parameter $H \in (0, 1)$. For $H = 0.5$, $B^H = W$, and a semimartingale. Arbitrage opportunities is possible if the log-price process is a fractional Brownian motion (Rogers 1997), but we disregard this.

The model setting is completely continuous, but, as mentioned in the introduction, empirical evidence of jumps in both log-price and volatility suggest addition of jump processes. The notion of jump processes will be introduced and treated in a later section.

The volatility model setting includes many concepts from probability theory. The following sections are devoted to presenting definitions and theorems regarding probability spaces, measures, random variables, and especially regarding the field of stochastic processes.

Regarding the presented theorems; these are given without proofs, and the reader is referred to the given references for proofs.

2.1 Spaces and measures

To begin this subsection, we will present a number of definitions explaining the concepts of probability spaces and measures. The main references for this subsection are Billingsley (2017), Resnick (2019) and Tankov (2003).

Definition 1. *A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the sample space, $\mathcal{F}$ is a σ-field and $\mathbb{P}$ is a probability measure.*

Definition 2. *Ω is an arbitrary space that consists of all possible outcomes. An individual element of Ω is denoted by ω.*

The space Ω will represent all possible scenarios of the market, and $\omega \in \Omega$ can be thought of as the evolution of the log-price.

Definition 3. *A non-empty class $\mathcal{F}$ of subsets of Ω is called a field if it contains Ω and is closed under the set operations: finite union, finite intersection and complements. That is,*

$$\begin{aligned}&\Omega \in \mathcal{F}\\ &A \in \mathcal{F} \text{ implies } A^c \in \mathcal{F}\\ &A, B \in \mathcal{F} \text{ implies } A \cup B \in \mathcal{F}.\end{aligned}$$

Definition 4. *(Resnick 2019) A σ-field $\mathcal{F}$ is a non-empty class of subsets of Ω closed under countable union, countable intersection and complements. A synonym for σ-field is σ-algebra.*

The last component of the triplet is the probability measure $\mathbb{P}$, which is a special form of the more general concept of measures.

Definition 5. *A probability measure $\mathbb{P}$ satisfy the following conditions:*

- $0 \leq P(A) \leq 1$ *for* $A \in \mathcal{F}$.
- $P(\varnothing) = 0, P(\Omega) = 1$.
- $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P\left(A_n\right)$, *for disjoint events in* $\mathcal{F}$ $\{A_n, n \geq 1\}$. *This condition is called countable additivity.*

An event is a set of outcomes. $\mathcal{F}$ contains all the events for which it is possible to get information. $\mathcal{F}$ is called the event space, where an event is a set of outcomes. The probability function $\mathbb{P}$ assigns a probability to each event in $\mathcal{F}$.

One way to construct a σ-field is to start with the sets that one would like to be measurable with respect to $\mathbb{P}$ and then continue adding new sets by taking unions and complements until a σ-field is formed (Tankov 2003).

A set $A \in \mathcal{F}$ with $P(A) = 1$ is said to occur almost surely and if $P(A) = 0$ the event A is impossible.

The next definitions and a theorem are about random variables.

Definition 6. *A random variable is a real-valued function $X(\omega), \omega \in \Omega$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. X is a measurable function with respect to $\mathcal{F}$.*

In other words:

$$X : \Omega \mapsto \mathbb{R}.$$

Generally, random variables can be functions with values in $\mathbb{R}^d, d \geq 1$. However, in this thesis, we only deal with one-dimensional random variables, so all definitions and theorems are formulated for one-dimensional entities.

Definition 7. *For the random variable X the law is the probability measure μ on $\mathbb{R}$, defined by*

$$\mu(A) = P[X \in A], \quad A \in \mathbb{R}$$

Another way of expressing that a random variable is measurable is the following theorem.

Theorem 1. *A random variable Y is measurable iff for the random variables $X_1, \ldots, X_n$:*

$$Y = f\left(X_1, \ldots, X_n\right)$$

for some $f : \mathbb{R}^n \to \mathbb{R}$.

The remaining definitions in this section deals with a sequence of random variables $X_1, X_2, \ldots$. They define convergence rules that will be important for estimating volatility. They are convergence in probability, distribution, and in law.

Definition 8. *(Billingsley 2017) If* $\lim_{n\to\infty} P\left[|X_n - X| \geq \epsilon\right] = 0$ *holds for each positive* ϵ*, then* X_n *is said to converge to* X *in probability, written* $X_n \xrightarrow{P} X$*.*

Definition 9. *(Billingsley 2017) The distribution functions* F_n *are said to weakly converge to the distribution function* F *if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every continuity point x *of* F*, which we express as* $F_n \xrightarrow{D} F$*.*

Theorem 2. *If* $F_n \xrightarrow{D} F$ *then* $X_n \xrightarrow{\mathcal{L}} X$ *in which we say that* X_n *converges in law to* X *for* $n \to \infty$*.*

The distribution function $F : \mathbb{R} \to [0, 1]$ of a random variable $X : \Omega \to \mathbb{R}$ is defined by:

$$F(x) = P(\omega \in \Omega : X(\omega) \leq x)$$

or in the short-hand notation $F(x) = P(X \leq x)$.

2.2 Stochastic processes

The next subsection deals with the concept of stochastic processes. It will be the foundation for most of the subjects in this thesis. Inspiration comes from Aït-Sahalia & Jacod (2014), Lindgren (2006), Tankov (2003) and Billingsley (2017).

The term *process* is defined by:

Definition 10. *A process* $(X_t)_{t\in T}$ *is a collection of random variables on the time set* T*, defined on the same underlying probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, taking values in the Euclidean space* $\mathbb{R}$*, and indexed on the positive half line* $\mathbb{R}_+ = [0, \infty)$*.*

Any outcome $\omega \in \Omega$ can be considered as the path of the process. This implies that we can consider $(X_t)_{t\in T}$ as a single random variable with values in an appropriate space. The process is a discrete time process if $T = \mathbb{N}$ and a continuous time process if $T \subset \mathbb{R}_+$. We interpret t as time and we observe the path up to the current time $s \mapsto X_s(\omega)$ for all $s \in [0, t]$ where we refer to this "history" of the process as the filtration $\mathcal{F}_t$.

Definition 11. *(Tankov 2003) A filtration on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is an increasing family of* σ*-fields*

$$(\mathcal{F}_t)_{t\in T} : \forall t \geq s \geq 0, \mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}.$$

In other words, $\mathcal{F}_t$ is, for each t, the σ-field generated by variables X_s for $s \in [0, t]$. We write $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t)$ and by this we indicate that $\mathcal{F}_t$ is generated by all subsets and $\mathcal{F}_t$ is the information known at time t. A filtered probability space is the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration and will be written $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, \mathbb{P})$.

Definition 12. *(Tankov 2003) The stochastic process* $(X_t)_{t\in T}$ *is said to be adapted to filtration* $(\mathcal{F}_t)_{t\in T}$ *if each random variable* X_t *is* $\mathcal{F}_t$*-measurable,*

$$X_t^{-1}(B) \in \mathcal{F}_t, \forall B \in \mathbb{R},$$

where B *is a Borel set.*

A way of expressing this is to say that for all $t \in T$ the value of X_t is known at time t. Often, this is called a non-anticipating process.

Definition 13. *(Billingsley (2017), Tankov (2003))*

A càdlàg process $(X_t)_{t\in T}$ *on the filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, \mathbb{P})$ *is said to be a martingale if* X *is adapted to* $(\mathcal{F}_t)_{t\in T}$ *and*

- $E\left[|X_t|\right] < \infty, \forall t \in T$
- $E\left[X_s \mid \mathcal{F}_t\right] = X_t, \forall s > t$ *with probability 1.*

$\mathcal{F}_t$ is the collection of events whose occurrence can be determined from observations of the process X up to time t. $\mathcal{F}_t$ can be interpreted as the evolution of knowledge over time.

Definition 14. *(Tankov (2003)) The smallest σ-field on $\mathbb{R}$, which contains all open subsets of $\mathbb{R}$, is called the collection of Borel subsets of $\mathbb{R}$, and an element of this σ-field is called a Borel set.*

So, the present value of a martingale is the best prediction of the future value. The term càdlàg is defined as:

Definition 15. *(Tankov 2003) A function $f : T \to \mathbb{R}$ is said to be càdlàg if it is right continuous with left limit. For $\forall t \in T$ the limits*

$$f(t-) = \lim_{s \to t, s<t} f(s)$$

$$f(t+) = \lim_{s \to t, s>t} f(s)$$

exist and $f(t) = f(t+)$.

By definition every continuous function is càdlàg. But as we will explore later in this thesis, càdlàg functions permit a countable number of discontinuity points $\{t \in T, f(t) \neq f(t-)\}$.

The dynamics of a stochastic process $(X)_{t\geq 0}$ is generally specified as a stochastic differential equation.

Definition 16. *A stochastic differential equation (SDE) governing a stochastic process $(X)_{t\geq 0}$ can be expressed as follows:*

$$\mathrm{d}X_t = \mu_t\left(X_t\right)\mathrm{d}t + \sigma_t\left(X_t\right)\mathrm{d}W_t, \quad X_0 = x.$$

Here, $(W)_{t\geq 0}$ denotes a Wiener process, while $\mu_t(X_t)$ and $\sigma_t(X_t)$ are deterministic functions. The functions μ and σ are, respectively, termed the drift and diffusion function. The resulting process $(X)_{t\geq 0}$ is commonly referred to as a diffusion process.

Definition 17. *The process $(W_t)_{t\in T}$ is adapted to $(\mathcal{F}_t)_{t\in T}$ and*

- *$W_0 = 0$*
- *$t \mapsto W_t$ is a continuous function of t, i.e. W has continuous paths*
- *W has independent increments relative to the filtration $(\mathcal{F}_t)_{t\in T}$: If $t \geq s$, then $W_t - W_s$ is independent of $\mathcal{F}_s$*
- *For $t \geq s$, $W_t - W_s \sim N(0, t-s)$*

The only continuous martingale with stationary and independent increments is a Brownian motion.

In a Brownian motion the random "movements" increase linearly in size over time, and a Brownian motion can be shown to have a random fractal structure since it is a sum of arbitrary many increments in any time interval. So, any part of the motion has the same distribution as the original motion. If $C > 0$ is a constant, then

$$\tilde{W}_t = \frac{1}{\sqrt{C}} W_{C\cdot t}$$

has the same distribution as W_t, proving that $\tilde{W}$ is a Brownian motion. The covariance function for a Brownian motion is

$$E[W_t \cdot W_s] = min(t,s) = E[W_t \cdot W_s] = \frac{1}{2}\left(|t| + |s| - |t-s|\right).$$

Stationarity and ergodicity are two important properties of stochastic processes.

Definition 18. *Let τ be a constant and let t be the index time, then the stochastic process $(X_t)_{t\in T}$ is strictly stationary if the n-dimensional distributions of*

$$X_{t_1+\tau}, \ldots, X_{t_n+\tau}$$

are independent of τ.

A less strict form of stationarity is weak stationarity.

Definition 19. *$(X_t)_{t\in T}$ is weakly stationary, if it has a constant mean and has a covariance function that only depends on its time lag,*

$$\mathrm{Cov}\,(X_t, X_{t+s}) = f(s) \quad \forall t, s \geq 0.$$

The Wiener process is neither strict nor weak stationary, but the increments $W_t - W_s, s < t$ are stationary. The simple diffusion process is also not stationary. A process is Gaussian if all of its finite-dimensional distributions are multivariate Gaussian distributions.

A stationary process is ergodic if the following theorem holds.

Theorem 3. *If the covariance function is continuous for all $s, t \geq 0$ and*

$$\frac{1}{T}\int_0^T \mathrm{Cov}\,[X_s, X_{t+s}]\, ds \to 0, \quad T \to \infty$$

then

$$\lim_{T\to\infty} E\left[\left|\frac{1}{T}\int_0^T X_s ds - E\,[X_t]\right|^2\right] \to 0.$$

2.3 Stochastic integrals

Stochastic integration will, for convenience, be introduced using the standard Brownian motion W.

Consider the simple example from before

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 = 0.$$

We can integrate over the interval $[0, t]$ on both sides to get

$$\int_0^t \mathrm{d}X_s = \int_0^t \mu \mathrm{d}s + \int_0^t \sigma \mathrm{d}W_s, \quad X_0 = 0.$$

The first integral on the right-hand side of the equation is a standard integral, while the second is a stochastic integral. The integral on the left-hand side is also a stochastic integral.

In general we consider the expression $\int_0^t C_s dW_s$, where C is a piecewise constant process of the form

$$C_t = \sum_{i\geq 1} C_{t_{i-1}} 1_{[t_{i-1}, t_i)}(t) \tag{2}$$

where $0 = t_0 < t_1 < \cdots$ and $t_n \to \infty$ as $n \to \infty$. In this context, one might define

$$\int_0^t C_s dW_s = \sum_{i\geq 1} C_{t_{i-1}} \left(W_{t\wedge t_i} - W_{t\wedge t_{i-1}}\right) \tag{3}$$

where $t \wedge t_i = min(t, t_i)$. Due to the Lévy Characterization Theorem the integral above can be extended to all processes C having the properties of being progressively measurable and $\int_0^t C_s^2 ds < \infty$ for all t (Aït-Sahalia & Jacod 2014). We call $\int_0^t C_s dW_s$ a stochastic integral and more precisely a Brownian stochastic integral.

Definition 20. *If, for any $t \geq 0$, the function $(\omega, s) \mapsto X_s(\omega)$ is $\mathcal{F}_t \times \mathcal{B}([0, t])$-measurable on $\Omega \times [0, t]$, then the process $(X_t)_{t\geq 0}$ is said to be progressively measurable.*

Definition 21. *Let $\mathcal{B}([0,t])$ denote the Borel σ-field of the interval $[0,t]$.*

A simple example

$$\int_0^t \mathrm{d}X_s = \int_0^t \mu \mathrm{d}s + \int_0^t \sigma \mathrm{d}W_s$$

gives

$$X_t = \mu t + \sigma W_t$$
$$X_0 = 0.$$

The stochastic integral can be encounted as an Itô integral as well. The stochastic integral is a stochastic variable or, more precisely, a stochastic process. The integrand process, C can be integrated against all types of semimartingales (see the next section). Semimartingales are the largest class of stochastic processes in which the Itô integral can be defined.

An important subject in this thesis is simulation and numerical integration of stochastic differential equations. The focus will be on the Euler-Maruyama approximation method. It is a Euler scheme method and requires the SDE to be on an Itô form; see the next subsection.

Let us assume that $(X_t)_{t\in[0,T]}$ is a solution to the diffusion process defined by the SDE in definition 16. Integrating on $[0,T]$ we get

$$X_t - x = \int_0^T \mu_s(X_s)ds + \int_0^T \sigma_s(X_s)dW_s, \quad X_0 = x.$$

We discretize the interval according to $\Delta_n = T/n$ for some integer n, and with $t_i = i\Delta_n$. Denoting $X_{t_i} = X_i$, the Euler-Maruyama approximation of X is given by the discrete stochastic process satisfying the iterative scheme (Morlanes 2017):

$$X_i = X_{i-1} + \mu_{i-1}\left(X_{i-1}\right)\Delta_n + \sigma_{i-1}\left(X_{i-1}\right)\left(W_i - W_{i-1}\right).$$

Between any two points t_i and t_{i-1} in $[0,T]$, linear interpolation is applied to get:

$$X_t = X_{i-1} + \left(X_i - X_{i-1}\right)\frac{t - t_{i-1}}{t_i - t_{i-1}}, \quad t \in [t_{i-1}, t_i).$$

In order to simulate X, it is only necessary to generate the increments $(W_i - W_{i-1})$ of the standard Brownian motion W. These non-overlapping increments are independent and Gaussian with mean zero and variance one.

2.4 Semimartingales

The fundamental asset pricing theorem states that in the absence of arbitrage opportunities, the price process must adhere to the properties of a semimartingale (Aït-Sahalia & Jacod 2014). A semimartingale process is defined as:

Definition 22. *On the filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$, a real-valued process X is called a semimartingale if it can be written as sum of a local martingale M and an adapted càdlàg process A "with finite variation"*

$$X_t = M_t + A_t. \tag{4}$$

In this definition, the local martingale M is the randomness component and A is the drift component.

A local martingale is defined by:

Definition 23. *(Tankov 2003) A process $(X_t)_{t\in[0,T]}$ is a local martingale if there exist a sequence of stopping times $(T_n)_{n\geq 1}$ with $T_n \to \infty$ (with probability 1), such that $(X_{t\wedge T_n})_{t\in[0,T]}$ is a martingale.*

In other words, a local martingale confirms to the properties of a martingale up to time T_n. A stopping time T_n (Tankov 2003) is a positive random time with

$$\forall t \geq 0 \quad \{T_n \leq t\} \in \mathcal{F}_t.$$

So given the information $\mathcal{F}_t$ we know that the stopping time T_n has occurred. The process $(X_{t\wedge T_n})_{t\in[0,T]}$ is defined as

$$X_{t\wedge T_n} = X_t \text{ if } t < T_n, \quad X_{t\wedge T_n} = X_{T_n} \text{ if } t \geq T_n.$$

Every martingale is a local martingale by definition. We now define the concept of finite variation.

Definition 24. *(Mykland & Zhang 2012) A process A with finite variation fulfills*

$$\sup_{n\to\infty} \sum_{i=1}^{n-1} |A_{t_{i+1}} - A_{t_i}| < \infty.$$

The standard Brownian motion is the most simple form of a semimartingale and

$$X_t = X_0 + \int_0^t \mu ds + \int_0^t \sigma dW_s,$$

then X is semimartingale with:

$$M_t = X_0 + \int_0^t \sigma dW_s \text{ and } A_t = \int_0^t \mu ds.$$

An important class of semimartingales are Itô processes or Itô semimartingales. Above in definition 16 we defined a difussion process.

Definition 25. *Let $\sigma = (\sigma_t)_{t\geq 0}$ be an square integrable process, $\mu = (\mu_t)_{t\geq 0}$ be another progressively measurable process such that $\int_0^t |\mu_s|\, ds < \infty$ for all $t \geq 0$ and W is a Brownian motion. An $(\mathcal{F}_t)$-adapted process X written as*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \tag{5}$$

is a continuous Itô semimartingale (general diffusion).

The process $(W_t)_{t\geq 0}$ is the driving process of $(X_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$ is the rate of return while $(\sigma_t)_{t\geq 0}$ is the spot volatility. In mathematical finance, Itô processes contribute to most processes used for modeling empirical financial data. Let us consider the drift process $\mu = (\mu_t)_{t\geq 0}$ in an Itô process. The process is said to be predictable if:

Definition 26. *(Jacod 2012) $\mu = (\mu_t)_{t\geq 0}$ is predictable if it can be written as a limit of simple functions $\mu_t^{(n)}$*

$$\mu_t^{(n)}(\omega) \to \mu_t(\omega), \quad n \to \infty$$
$$\forall (t,\omega) \in [0,T] \times \Omega.$$

All adapted and left continuous processes are predictable. Another important class of semimartingales are Lévy processes. A Lévy process is defined by:

Definition 27. *(Aït-Sahalia & Jacod 2014) A filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ being given, a Lévy process relative to the filtration $(\mathcal{F}_t)_{t\geq 0}$, or an $(\mathcal{F}_t)$-Lévy process, is an $\mathbb{R}$-valued process X satisfying the following three conditions:*
1. Its paths are right-continuous with left limit everywhere (càdlàg), with $X_0 = 0$
2. It is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$
3. For all $s, t \geq 0$ the variable $X_{t+s} - X_t$ is independent of the σ-field $\mathcal{F}_t$, and its law only depends on s.

The càdlàg property is important, as it supports the concept of discontinuities, "jumps". The càdlàg property limits the total number of jumps to be countable, and it also limits the number of jumps in a fixed interval to be finite (Tankov 2003). Discontinuities are allowed at random times, but the probability of a jump at a given time t is zero.

For any $t > 0$, the distribution of X_t, where X is a Lévy process, is infinitely divisible.

Definition 28. *(Tankov (2003)) A probability distribution D on $\mathbb{R}$ is said to be infinitely divisible if, for any integer $n \geq 2$, there exists n i.i.d. random variables $X_1, \ldots X_n$ such that $X_1 + \ldots + X_n$ has distribution D.*

As an example of a simple Lévy process, let us consider a compound Poisson process. A compound Poisson process is a stochastic process Z defined as

$$Z_t = \sum_{n=1}^{N_t} J_n, \quad Z_0 = 0, \quad t \geq 0,$$

where jumps sizes J_t are i.i.d. with a common distribution, e.g Gaussian and $(N_t)_{t\geq 0}$ is a Poisson counting process with parameter λ. N_t and J_t are independent. A compound Poisson distribution can be written as

$$dZ_t = J_t dN_t, \quad t \geq 0$$

and the Poisson process is written as $N_t = \sum_{n\geq 1} 1_{\{T_n \leq t\}}$, where T_n is a sequence of positive stopping times that strictly increases and approaches $+\infty$ as its limit.

The properties of a compound Poisson process are:

- $N_t = 0 \Rightarrow Z_t = 0$
- Càdlàg
- Adapted Given information up to time t
- $Z_{t+s} - Z_t$: independent of $Z_u, 0 \leq u \leq t$
- $Z_{t+s} - Z_t$: same distribution as Z_s.

Lévy processes are a class of purely continuous processes, purely discontinuous or continuous with discontinuities processes. Other simple examples of Lévy processes are the standard Brownian motion and the Poisson process. Lévy processes are divided into 3 independent parts: Linear drift, a Brownian motion, and the Lévy jump process. This implies that the only continuous non-deterministic Lévy process is a Brownian motion with drift (Tankov 2003). The sample paths have continuous paths. All other Lévy processes contain discontinuities with non-zero probability (jumps).

So, with Lévy processes introduced, we can write semimartingales as

$$X = X_0 + X^c + X^d, \quad X_0^c = X_0^d = 0,$$

where X^c is the continuous component and X^d is the discrete component. A simple example of this is

$$dX_t = \mu dt + \sigma dW_t + J_t dN_t, \quad t \geq 0.$$

Here

$$X_t^c = \int_0^t \mu ds + \int_0^t \sigma dW_s$$

$$X_t^d = \int_0^t J_s dN_s.$$

2.5 Fractional Brownian motion

In stochastic volatility modeling, the main focus is on the diffusion term in definition 16. Standard SV models use the standard Brownian motion to drive the volatility over time. We will now present a generalization of the Brownian motion, which will be the basis for the SV processes in this thesis.

Introduced by Mandelbrot & Van Ness (1968) the fractional Brownian motion (fBm) is defined as a centered Gaussian process:

$$B_t^H = B_0^H + \frac{1}{\Gamma\left(H + \frac{1}{2}\right)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-1/2} \ -(-s)^{H-1/2} \right] dB_s + \int_0^t (t-s)^{H-1/2} dB_s \right\} \tag{6}$$

where B is a standard Brownian motion and Γ is the gamma function. The Hurst parameter, denoted as $H \in (0,1)$, both defines the sign of the correlation of the increments and accounts for the regularity present in the sample paths. With $H > 0.5$ the increments are positively correlated and with $H < 0.5$ the increments are negatively correlated. At $H = 0.5$ the fBm process reduces to the standard Brownian motion.

The fBm is not a semimartingale when $H \neq 0.5$. This implies that conventional stochastic calculus is not applicable (Decreusefond & Üstünel 1999).

The fBm process has the covariance function

$$\operatorname{Cov}\left(B_t^H, B_s^H\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \quad t, s \in \mathbb{R}. \tag{7}$$

In particular $\operatorname{Var}\left(B_t^H\right) = |t|^{2H}$.

The process $\left(B_t^H\right)_{t\geq 0}$ exhibits self-similarity. So for any constant $c > 0$, the rescaled processes $\left(c^{-H} B_{ct}^H\right)_{t\geq 0}$ share the same probability distribution as $\left(B_t^H\right)_{t\geq 0}$. This property directly follows from the homogeneity of the covariance function with respect to $2H$, expressed as (Morlanes 2017):

$$\operatorname{Cov}\left(B_{ct}^H, B_{ct}^H\right) = c^{2H} \operatorname{Cov}\left(B_t^H, B_s^H\right), \quad c > 0.$$

The increments $\left(B_t^H - B_s^H\right)$ are stationary with a zero mean Gaussian distribution, and their variance is given by $\operatorname{Var}\left(B_t^H - B_s^H\right) = |t-s|^{2H}$ (Morlanes 2017). In fact, the fBm is the only self-similar Gaussian process with stationary increments (Marquardt 2006).

3 Jumps in financial econometrics

Up until now, we have mainly focused on stochastic processes where the models have been defined as continuous processes over time. In the Introduction, we mentioned that empirical evidence suggests the presence of discontinuities in both the log-price and volatility process.

Large random returns in the price process can be used to model the effects of events in financial markets. The large returns have no impact on the future distribution of returns (Eraker et al. 2003) and can be modeled by including a jump process in prices. A volatility model driven by a Brownian motion alone cannot model a rapid increase that is persistent for a period of time. A jump process in the volatility model can provide this proven empirical feature Eraker et al. (2003). Jumps in volatility can be more frequent, and the effect of a positive-valued jump will have a longer effect, in which volatility will revert back to the long run mean-level over time (Eraker et al. 2003). In an Ornstein-Uhlenbeck model, the volatility process will, between two jumps, decay exponentially as an effect of the linear damping term, then the jump activity imply an asymmetric behavior of positive jumps.

Jumps in prices are likely to be negative and have a very short term effect on the price process.

Furthermore, jumps in both price and volatility at the exact same time have empirical evidence. Often these jumps have opposite jump sizes with a negative correlation (Bandi & Renò 2012).

Jumps can be an important factor in a complete description of market risk, but they involve complications and less tractable analysis (Tankov 2003).

3.1 Lévy processes - building blocks for jumps

The Poisson and Wiener processes are Lévy processes. In fact, every Lévy process is a superposition of a Wiener process and a (possibly infinite) number of independent Poisson processes (Tankov 2003). For this discussion, we need to introduce two random measures; Jump measure and Lévy measure.

Definition 29. *(Tankov 2003) Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}$. The measure v on $\mathbb{R}$*

$$v(A) = E\left[\#\left\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\right\}\right]$$

$A \in \mathcal{B}(\mathbb{R})$, is the Lévy measure of X.

In other words it is the expected number of jumps with sizes in A, expectation is per unit time. $\mathcal{B}$ is the Borel σ-field on $\mathbb{R}$, and $\Delta X_t = X_t - X_{t-}$ is the jump size at time t. X_{t-} is the left limit of X. For every compound Poisson process $(X_t)_{t\geq 0}$ (see section 2.4) on $\mathbb{R}$ we can associate a random measure to describe the jumps

$$J_X(B) = \#\{(t, \Delta X_t \in B\}$$

for any measuable set $B \in [0,\infty) \times \mathbb{R}$. Then for every measureable set $A \in \mathbb{R}$

$$J_X\left([0,t] \times A\right)$$

is the number of jumps in $[0,t]$ with sizes in A (Tankov 2003).

A Lévy process has the following Lévy-Itô decomposition:

Proposition 1. *(Tankov 2003) Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}$ and v the belonging Lévy measure defined in Definition 29.*

- *Then v satisfy*
$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty, \quad \int_{|x|\geq 1} \nu(dx) < \infty$$
for $A\times \in \mathbb{R}\backslash\{0\}$
- *The J_X measure is a Poisson random measure on $[0,\infty) \times \mathbb{R}$ with intensity $v(dx)dt$*
- *There exist a scalar μ and a Brownian motion $(B_t)_{t\geq 0}$ such that X can be written like*
$$X_t = \mu t + B_t + X_t^l + \lim_{\epsilon \to 0} \tilde{X}_t^\epsilon$$
with
$$X_t^l = \int_{|x|\geq 1} \int_0^t x J_X(ds \times dx)$$
$$\tilde{X}_t^\epsilon = \int_{\epsilon \leq |x| < 1} \int_0^t x \left[J_X(ds \times dx) - \nu(dx)ds\right]$$

The terms in the above formula are independent and the convergence is with probability 1 in $t \in [0,T]$. The cutoff value of 1, is arbitrary.

The term $\mu t + B_t$ is a continuous Gaussian Lévy process. The two integrals describe the jumps of X. The first integral represents the large jump process, the second is a small jump process. Because X has a finite number of jumps, given by $\int_{|x|\geq 1} \nu(dx) < \infty$, then

$$X_t^l = \sum_{0\leq s\leq t}^{|\Delta X_s|\geq 1} \Delta X_s$$

has a finite number of terms, $(X_t^l)_{t\geq 0}$ is a compound Poisson process. Since v can have a singularity at 0, infinitely many small jumps

$$\lim_{\epsilon\to 0} X_t^\epsilon = \lim_{\epsilon \to 0} \sum_{0\leq s\leq t}^{1>|\Delta X_s|\geq \epsilon} \Delta X_s = \lim_{\epsilon \to 0} \int_{\epsilon \leq |x| < 1} \int_0^t x J_X(ds \times dx)$$

does not necessarily converge. However, if we write $\tilde{X}_t^\epsilon$ as the infinite sum of centered (compensated) Poisson processes, convergence can be obtained (Tankov 2003).

In this thesis, we will consider the addition of compound Poisson processes to describe jumps in both prices and volatility. As the proposition says, addition of a independent compound Poisson process to a Lévy process is again a Lévy process. So we focus on the case where we have an independent compound Poisson process with finite activity, i.e. finite number of jumps on $[0,T]$ (that is if $v(\mathbb{R}\backslash\{0\}) < \infty$).

A Poisson process is a compound Poisson process on $\mathbb{R}$ with all $(J_i)_{i\geq 1}$ equal to 1. More precisely, in the following we denote a compound Poisson process by $(Z_t)_{t\geq 1}$ and it can be written as the sum

$$Z_t = \sum_{k=1}^{N_t} J_k, \quad N_0 = 0$$

where N_t is the random number of jumps up to time t, J_n is the random size of the n'th jump. It is assumed that

$$N_t \sim \text{Pois}(\lambda), \quad \forall t$$

and that the jump sizes are independent and follow the same distribution, e.g. a Gaussian distribution. The process is piecewise constant. The jump process can be written in differential form

$$dZ_t = J_t dN_t, \quad Z_0 = 0$$

implying that

$$Z_t = \int_0^t J_s dN_s = \sum_{k=1}^{N_t} J_k.$$

A pure jump process is just a compound Poisson process. A driftless Lévy process with jumps is

$$X_t = \sigma W_t + \sum_{k=1}^{N_t} J_k.$$

In practice, a very useful model is an Itô semimartingale added a compound Poisson process

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \sum_{k=1}^{N_t} J_k$$

can be written as

$$X_t^c = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

$$X_t^d = \int_0^t J_s dN_s = \sum_{n=1}^{N_t} J_n.$$

The continuous part X_t^c is decomposed into $A_t^c + M_t^c$, with M_t^c being the continuous local martingale and A_t^c being the continuous process of finite total variation (see section 2.4).

Finally, we will add a jump process component to a process that is not a semimartingale. The model setting studied in this thesis includes a fractional OU process. Later in this thesis we study the process

$$Y_t = v \int_0^t e^{-\lambda(t-s)} \mathrm{d}B_s^H + \sum_{n=1}^{N_t} J_n.$$

Since $Y_t = log(\sigma_t^2)$, we get

$$\sigma_t^2 = \xi \exp\left(Y_t - \frac{1}{2}\kappa(0)\right) \prod_{n=1}^{N_t} exp(J_n).$$

3.2 Effects of adding jumps

A jump in the log-price process at time t will result in a large change in return around that time. After the jump, the process continues along a "standard" path governed by its continuous component but at a different level than before the jump. Returns are back at their sizes according to the continuous process. A jump process in volatility has a persistent effect. The jump at time t will change the level of volatility, but the volatility process will revert back to the average volatility over time, indicating a changed dynamic. Co-jumps are from a jump process in both the price and volatility process. Co-jumps share the same Poisson process, i.e. jumps at exactly the same time in both processes. The jump size distribution are not the same, but normally it is assumed that they are correlated. Most empirical studies indicate a negative correlation.

4 Introduction to volatility estimation

A main topic in the analysis of high-frequency financial data is the estimation of volatility. Important concepts in volatility estimation are integrated variation, quadratic variation, and realized volatility. We will first introduce these concepts in a simple setting, assuming that X is a continuous time stochastic process observed on a bounded time interval $[0, T]$ without microstructure noise and equidistant time points. In later sections, we will cover the effect of addition of microstructure noise, as well as the effects of jumps and roughness on volatility estimation.

This subsection is inspired by Chapters 1, 3, 4, and 6 of Aït-Sahalia & Jacod (2014). In analysing the volatility, we make use of four properties regarding the setup with high frequency data:

- Data are sampled at discrete time points, with small to very small distance
- Only a single path is observed
- Sampling is performed on a finite time horizon
- High frequency asymptotics with the time interval Δ between sampling tending to zero

Let us assume that we have an Itô process of the form

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \quad t \geq 0, \tag{8}$$

and that the process have been observed at discrete time points in the interval $[0, T]$. We will in this section apply the notation from Aït-Sahalia & Jacod (2014) for the observation times

$$i\Delta_n, \quad i = 0, 1, ..., \lfloor T/\Delta_n \rfloor .$$

Here Δ_n is the length of the time interval between $i\Delta_n - (i-1)\Delta_n$ and $\lfloor T/\Delta_n \rfloor$ is the largest integer less than or equal to T/Δ_n. In later sections, we will use specific examples of this notation, e.g., with $T = 1$ and $\Delta_n = \frac{1}{n}$, then $i\Delta_n = \frac{i}{n}$ for $i = 0, 1, ..., n$. Or n observations in $[t-1, t]$, then again $\Delta_n = \frac{1}{n}$ and observation times $t-1, t-1+\frac{1}{n}, ..., t-1+\frac{n-1}{n}, t$. Then $i\Delta_n = \frac{n(t-1)+i}{n}$.

In estimating the volatility, the main interest is the integrated variance (IV) which is the integral

$$IV_T = \int_0^T \sigma_s^2 ds.$$

In general IV_T is a random variable and simple Itô calculus gives

$$\left| \int_0^T \sigma_s dW_s \right|^2 = \int_0^T \sigma_s^2 \left(dW_t\right)^2 = \int_0^T \sigma_s^2 ds.$$

If we assume the following mild conditions (Barndorff-Nielsen & Shephard 2002):

1. Processes μ and σ are pathwise of local bounded variation on $[0, \infty)$:

$$\int_0^t \mu^2 ds < \infty, \quad \int_0^t \sigma^2 ds < \infty.$$

2. μ has the Lipschitz continuity property:

$$\forall t_1, t_2, \exists K$$

$$|\mu_{t_1} - \mu_{t_2}| \leq K|t_1 - t_2|, \quad K > 0.$$

Then

$$[X,X]_T = \int_0^T \sigma_s^2 ds = IV_T,$$

where $[,]$ stands for quadratic variation. Quadratic variation is of special interest and an important tool in the analysis of stochastic processes, especially with respect to volatility.

For a pure continuous process you will also encounter the notation $\langle X,X\rangle_T$. For a continuous martingale or semimartingale

$$\langle X,X\rangle_T = [X,X]_T.$$

The quantity $\langle,\rangle_T$ is called the predictable quadratic variation or locally square integrable martingales.

The quadratic variation is a measure of how the randomness of a process adds up over time. For a standard Brownian motion W

$$[W,W]_T = \int_0^T dW_s = T,$$

the quadratic variation (QV) is just the normal time. That is why QV can be thought of as the internal clock of the process. QV is an increasing process for $t \geq 0$.

Theorem 4. *(Aït-Sahalia & Jacod (2014)) Let X and Y be two semimartingales. For each n, let $T(n,0) = 0 < T(n,1) < T(n,2) < \cdots$ be a strictly increasing sequence of stopping times with infinite limit, and suppose that the mesh $\pi_n(t) = \sup_{i\geq 1}(T(n,i)\wedge t - T(n,i-1)\wedge t)$ goes to 0 for all t, as $n \to \infty$. Then we have the following convergence in probability:*

$$\sum_{i\geq 1}\left(X_{T(n,i)\wedge t} - X_{T(n,i-1)\wedge t}\right)\left(Y_{T(n,i)\wedge t} - Y_{T(n,i-1)\wedge t}\right) \overset{u.c.p.}{\Longrightarrow} [X,Y]_t.$$

The term "ucp" is local uniform convergence in probability.
Examples of quadratic variations:

- An Itõ process with constant drift and constant volatility

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s.$$

We get

$$[X,X]_T = \sigma^2 T.$$

If

$$X_t = X_0 + \int_0^t \mu_s ds$$

then $[X,X]_T = 0$.

In general, if X has finite variation, then

$$[X,X]_T = 0.$$

- For a compound Poisson process

$$Z_t = \sum_{k=1}^{N_t} J_k,$$

the quadratic variation is

$$[Z,Z]_T = \sum_{k=1}^{N_T} J_k^2 = \sum_{0\leq s\leq T} (\Delta Z_s)^2,$$

where the last expression holds for every càdlàg process with finite variation.

- Then for a Lévy process we have

$$[X,X]_T = \sigma^2 T + \sum_{0\le s\le T} (\Delta X_s)^2 .$$

Theorem 4 implies that a sum of squared returns is a good candidate for a consistent estimator of QV. For the discrete and equidistant observations of X, the return over the i'th interval is denoted by

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}.$$

In Barndorff-Nielsen & Shephard (2002) it is proved that realized volatility (RV) as defined by

$$RV\left(\Delta_n\right)_T = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\Delta_i^n X\right)^2$$

is a consistent estimator for $QV_T = [X,X]_T$,

$$RV\left(\Delta_n\right)_T \xrightarrow{P} QV_T \quad as \quad \Delta_n \to 0.$$

This convergence result is valid for the class of semimartingales. For continuous Itô semimartingales we proved that $QV_T = IV_T$ so $RV(\Delta_n)_T$ is a consistent estimator for IV_T, implying that realized volatility can be used to estimate the spot variance of Itô processes. For Itô processes, $C_t = [X,X]_t$, $t \ge 0$ is a continuous adapted process with $C_0 = 0$. The integrated variation for $t \ge 0$, IV_t can also be viewed as a continuous process. We should note that a single path $t \mapsto X_t(\omega)$ is partly observed up to time T and IV_T is estimated for this particular path ω. Different paths will result in different estimates for IV. The second remark is that of RV_T is fill-in as $\Delta_n \to \infty$ and not as a convergence in time with $T \to \infty$.

4.1 IV estimation with no noise

In this subsection, we will briefly touch on the estimation of IV when there is no microstructure noise. All consistency theorems and the central limit theorems are proved in Aït-Sahalia & Jacod (2014), Barndorff-Nielsen & Shephard (2002), Barndorff-Nielsen & Shephard (2006) and Mancini (2009).

We want to estimate IV_T at a given time T upon observing X at discrete time $i\Delta_n$ for $i = 0, 1, \ldots, \lfloor T/\Delta_n \rfloor$ and when the partition Δ_n goes to 0, which is the same as the frequency of sampling tends to infinity.

Let us assume that the Itô process is continuous, i.e.

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^s \sigma_s dW_s, \quad t \ge 0. \tag{9}$$

If we define integrated quarticity as

$$IQ_T = \int_0^T \sigma_s^4 ds,$$

then the following central limit theorem (CLT) holds

$$\sqrt{n}\left(RV(\Delta_n)_T - IV_T\right) \xrightarrow{D} N(0, 2IQ_T), \quad \text{as } \Delta_n \to 0.$$

Next, let us assume that X is a discontinuous Itô process, and for simplicity we assume a driftless process

$$X_t = X_0 + \int_0^t \sigma_s dW_s + Z_t, \quad t \ge 0 \tag{10}$$

where $(Z_t)_{t\ge 0}$ is a compound Poisson process. In this model we have that the quadratic variation is

$$[X,X]_T = IV_T + \sum_{k=1}^{N_T} J_k^2$$

when $Z_T = \sum_{k=1}^{N_T} J_k$. We now have the following convergence in probability

$$RV(\Delta_n)_T \xrightarrow{P} IV_T + JV_T,$$

with

$$JV_T = \sum_{k=1}^{N_T} J_k^2.$$

We note that RV no longer is a consistent estimator for IV. In order to estimate IV consistenly we introduce the bipower variation defined as

$$B\left(1,1,\Delta_n\right)_T = \frac{\pi}{2} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor - 1} \left|\Delta_i^n X\right| \left|\Delta_{i+1}^n X\right|.$$

It can be proved that

$$B\left(1,1,\Delta_n\right)_T \xrightarrow{P} IV_T,$$

The last convergence in probability can loosely be justified, since the two returns in the bipower variation in the limit cannot both have jumps since they are rare events.

Now, let us consider another way to deal with jumps in order to estimate IV consistenly. For this, we introduce the truncated realized volatility defined as

$$RV\left(\Delta_n, u_n\right)_T = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i^n X)^2 1_{\left\{|\Delta_i^n X| \leq u_n\right\}}. \tag{11}$$

In the definition, the sequence (u_n) converges to 0, as $n \to \infty$, and $1_{\{\}}$ is the indicator function. The truncation removes the increments including jumps, and if (u_n) is not chosen too small it will still preserve the continuous part of the increments. Given that σ is càdlàg then

$$RV\left(\Delta_n, u_n\right) \xrightarrow{P} IV$$

for

$$u_n / \sqrt{\Delta_n \log\left(1/\Delta_n\right)} \to \infty$$

and

$$u_n \to 0.$$

In order to define u_n the following property can be used

$$u_n \asymp \Delta_n^{\varpi} \text{ for a } \varpi \in \left(0, \frac{1}{2}\right)$$

where

$$u_n \asymp v_n$$

implies that both u_n/v_n and v_n/u_n are bounded in size.

Choosing the right truncation level is crucial in using truncated realized variation. We note that, if

$$u_n = \alpha \Delta_n^{\varpi}$$

for an $\alpha > 0$, the consistency result holds regardless of the choice of α. If α is increased, the empirical value of $RV(\Delta_n, u_n)_T$ will be closer to the non-truncated $RV(\Delta_n)_T$, so the value of u_n in relation to Δ_n and the total number of observations matters in the practical use of truncated realized variation.

We will later in the thesis apply the truncated realized variation and define the truncation level we use.

For estimating the jump variation, we note that

$$RV(\Delta_n)_T - B(1,1,\Delta_n)_T \xrightarrow{P} JV_T$$

and the proportion of jumps can be estimated according to

$$1 - \frac{B(1,1,\Delta_n)_T}{RV(\Delta_n)_T} \xrightarrow{P} \frac{JV_t}{[X,X]_T}.$$

In order to estimate IQ_T we introduce power variations defined as

$$RV(p,\Delta_n)_T = \frac{1}{n}\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} |\sqrt{n}\Delta_i^n X|^p$$

with $p > 0$. Note that

$$RV(2,\Delta_n)_T = RV(\Delta_n)_T.$$

The general consistency theorem holds:

$$RV(p,\Delta_n)_T \xrightarrow{P} m_p \int_0^1 |\sigma|^p ds.$$

Here

$$m_p = \mathbb{E}[|Y|^p] = 2^{\frac{p}{2}} \frac{\Gamma(\frac{1+p}{2})}{\sqrt{\pi}}$$

where $Y \sim N(0,1)$. Then

$$\frac{2}{3} RV(4,\Delta_n)_T \xrightarrow{p} 2IQ_T.$$

This holds for X being a continuous process. If X is discontinuous, the following hold,

$$\sum_{i=1}^{\lfloor T/\Delta_n \rfloor} |\Delta_i^n X|^p \to JV(p)_T \quad p > 2,$$

where

$$JV(p)_T = \sum_{k=1}^{N_T} |J_k|^p.$$

5 Microstructure noise

In the previous sections, we assumed that we observed the prices of a financial asset without error. However, this is not the case when we look at the data. With increasing sampling frequency, the so-called microstructure noise will have a larger effect on realized volatility (Hansen & Lunde (2006), Aït-Sahalia & Jacod (2014) and Jacod et al. (2009)). The presence of microstructure noise in the observed log-prices poses problems in using RV as an estimator for QV.

The main reference for this section is Chapter 7 of Aït-Sahalia & Jacod (2014).

We will discuss three types of errors, i.e. noise when observing the log-price X.

Firstly, factual errors might arise in the recording of prices. These are pretty easy to eliminate by cleaning the data, which is detailed in the section of empirical work in this thesis.

Secondly, rounding errors are present. This is because prices are recorded at the nearest cent. A small number of consecutive observations is therefore often unchanged, which is not agreeable with a semimartingale model. We therefore have to account for this when estimating volatility.

Thirdly, a no arbitrage constraint is often used in finance, which limits the price process to that of a semimartingale (also assumed is that the process is continuous). Prices are, however, determined by transactions and nothing exists between transactions; therefore, having a continuous-time model only makes sense as a scaling limit. Microstructure, in this case, therefore refers to this microeconomic mechanism of transactions (or quotes for that matter). The combination of factual

errors, rounding errors, and the difference between the microstructure and its scaling limit is what we call microstructure noise, in which the efficient price is contaminated.

In this section, we will consider the underlying process of the form:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \quad t \geq 0 \tag{12}$$

and we will throughout assume that μ is bounded and σ is càdlàg.

5.1 Microstructure noise models

Let us start by making the following assumptions. X is at least a Itô semimartingale. We observe equidistant observations $i\Delta_n$ for $i = 0, 1, \ldots$ over a finite time interval $[0, T]$. We now do not observe the efficient price, but instead

$$Y_i^n = X_{i\Delta_n} + \varepsilon_i^n \tag{13}$$

where ε_i^n is the noise. The upper index n defines the frequency of the observations, while the lower index defines the rank of the observation.

The first model we will discuss is that of pure rounding noise. This is the first thing that becomes apparent after the data is cleaned. We will here take advantage of how prices are observed. Returns and volatility might become zero in short intervals, which is of zero probability for a model that has a Brownian semimartingale component. Consider the case where we observe the log-price of an asset. Fixing a level of rounding $\beta > 0$, taking the closest from below rounded-off value the price and we observe

$$Y_i^n = \log\left(\beta \left\lfloor e^{X_{i\Delta_n}}/\beta \right\rfloor\right)$$

where β is the real rounding level and $\lfloor x \rfloor$ stands for the integer part of $x \in \mathbb{R}$. We could have a version of shrinking noise where $\beta = \beta_n$ goes to 0 as $\Delta_n \to 0$. While rounding is a feature of financial data, the pure rounding gives rise to statistical problems which have no solution.

The second model we consider is the additive noise case where, for every step n, the variables $(\varepsilon_i^n)_{i\geq 0}$ remain completely independent of the process X and take the form

$$\varepsilon_i^n = \alpha_n \chi_{i\Delta_n}$$

where $(\chi_t)_{t\geq 0}$ is a white noise. This gives rise to two different forms: One where $\alpha_n = 1$ for all n, that is we are in the case of fixed noise, and another where we have shrinking noise, that is $\alpha_n \to 0$ as $\Delta_n \to 0$. In both cases, we assume that $\mathbb{E}[\chi_t] = 0$ (centered noise).

The third model we will consider combines rounding and the fact that the efficient price is only a limiting case for when the transaction frequency goes to infinity. Note here that the difference between the effcient price and the real price is one type of noise, while rounding is another. With that in mind, we can consider the case where the observations has the following form

$$Y_i^n = \log\left(\beta \left\lfloor e^{X_{i\Delta_n} + \alpha_n \chi_{i\Delta_n}}/\beta \right\rfloor\right).$$

Here the white noise process is, as before, independent of X.

We might ask ourselves what happens to realized volatility when noise is present. In this case the realized volatility in the non-corrected case takes the form

$$RV^{\text{noisy}}(\Delta_n)_t = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i^n Y)^2, \quad \text{where } \Delta_i^n Y = Y_i^n - Y_{i-1}^n.$$

For the case of additive white noise, for both α_n being 1 or going to zero, it is possible to express RV as

$$\begin{array}{l} RV^{\text{noisy}}(\Delta_n)_t = \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i^n X)^2 + 2\alpha_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \Delta_i^n X \left(\chi_{i\Delta_n} - \chi_{(i-1)\Delta_n}\right) \\ -2\alpha_n^2 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \chi_{(i-1)\Delta_n}\chi_{i\Delta_n} + \alpha_n^2 \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \left(\left(\chi_{i\Delta_n}\right)^2 + \left(\chi_{(i-1)\Delta_n}\right)^2\right). \end{array}$$

While the first term goes towards QV the last three term diverge such that, when the noise size is constant, RV tends to $+\infty$, at rate $1/\Delta_n$. One way to deal with the divergence is to do sparse sampling, which consists of forming returns at lower frequencies. One might sample at a 5-minute interval, thus having 78 observations in a 6.5-hour trading day. One has to be worried about sampling to sparsely as this has the effect of increasing the variance of the estimator via the discretization effect. One can formalize an optimal frequency by, e.g. minimizing the mean square error of the sparsely sampled RV (although not feasible in practice as the expected value of the error and integrated quaticity are not known).

In the case of pure rounding noise, the general behavior is not known.

5.2 Noise assumption

We will bring the main assumption for the mixed noise case. In this case, additional randomness is added to the randomness that describes the evolution of X (which is contained in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$). We allow the noise to depend on X in the mixed case, so a model will have the noise defined on an auxiliary space $(\Omega', \mathcal{F}')$, which has a transition probability $\mathbb{Q}(\omega, d\omega')$, making the "process + noise" pair defined on the product space $\widetilde{\Omega} = \Omega \times \Omega'$, evolving according to $\mathbb{P}(d\omega)\mathbb{Q}(\omega, d\omega')$. We also assume that, at each t, the noise are defined, all while well knowing the noise only has psysical existence at observation times.

At stage n we have

$$Y_i^n = \log\left(\beta \left\lfloor e^{X_{i\Delta_n} + \alpha_n \chi_{i\Delta_n}} / \beta \right\rfloor\right), \quad \varepsilon_i^n = \alpha_n \chi_{i\Delta_n}, \text{ either } \alpha_n \equiv 1 \text{ or } \alpha_n \to 0.$$

Our main assumption on the noise is:

Assumption 1. *Conditionally on $\mathcal{F}$, all variables $(\chi_t : t \geq 0)$ are independent, and we have*
- $\mathbb{E}(\chi_t \mid \mathcal{F}) = 0$
- for all $p > 0$ the process $\mathbb{E}(|\chi_t|^p \mid \mathcal{F})$ is $(\mathcal{F}_t)$-adapted and locally bounded
- the (conditional) variance process $\gamma_t = \mathbb{E}\left(|\chi_t|^2 \mid \mathcal{F}\right)$ is càdlàg.

5.3 Pre-averaging

Here we assume Assumption 1 for the noise. The noise may possibly shrink as the noise takes the form $\alpha_n = \Delta_n^\eta$ for some known $\eta \in \left[0, \frac{1}{2}\right)$. Observations are now

$$Y_i^n = \log\left(\beta \left\lfloor e^{X_{i\Delta_n} + \alpha_n \chi_{i\Delta_n}} / \beta \right\rfloor\right), \quad \varepsilon_i^n = \Delta_n^\eta \chi_{i\Delta_n}.$$

The most important case is $\eta = 0$.

Heuristically, the idea is to consider the averages

$$Y_i^{\mathrm{av},n} = \frac{1}{k_n'} \sum_{j=0}^{k_n'-1} Y_{i+j}^n$$

$$X_i^{\mathrm{av},n} = \frac{1}{k_n'} \sum_{j=0}^{k_n'-1} X_{(i+j)\Delta_n}$$

$$\varepsilon_i^{\mathrm{av},n} = \frac{1}{k_n'} \sum_{j=0}^{k_n'-1} \varepsilon_{(i+j)\Delta_n}$$

where k_n' is a sequence of integers such that $k_n' \to \infty$ and $k_n' \Delta_n \to 0$. As $k_n' \to \infty$, then $\varepsilon_i^{\mathrm{av},n}$ goes to 0, so we expect $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor - 2k_n'} \left(Y_{i+k_n'}^{\mathrm{av},n} - Y_i^{\mathrm{av},n}\right)^2$ to be a good estimator for IV (after normalization).

In reality, we need a bit more for it to work. Let us provide the two ingredients:

1. We need a sequence k_n of integers which satisfy

$$k_n = \frac{1}{\theta \Delta_n^{\eta'}} + o\left(\frac{1}{\Delta_n^{(3\eta'-1)/2}}\right) \text{ as } n \to \infty$$
$$\text{where } \theta > 0, \eta' \geq \tfrac{1}{2} - \eta.$$

2. We need a weight function g, which is real valued on $\mathbb{R}$ that satisfy: g is continuous, null outside $(0,1)$ and piecewise C^1 with a piecewise Lipschitz derivative g'.

Setting

$$\begin{aligned}\phi(f,h \mid t) &= \textstyle\int_t^1 f(s-t)h(s)ds \\ \phi(f) &= \phi(f,f \mid 0) = \textstyle\int_0^1 f(t)^2 dt \\ \phi(f,h) &= \textstyle\int_0^1 \phi(f,f \mid t)\phi(h,h \mid t)dt\end{aligned}$$

and

$$\begin{aligned}\phi_{k_n}(h) &= \sum_{i=1}^{k_n} h\left(\frac{i}{k_n}\right)^2 \\ \phi'_{k_n}(h) &= \sum_{i=1}^{k_n}\left(h\left(\frac{i}{k_n}\right) - h\left(\frac{i-1}{k_n}\right)\right)^2\end{aligned}$$

where f, h is any bounded functions on $\mathbb{R}$ with support in $[0,1]$ and where all integers $n, i \geq 1$ and all $t \in [0,1]$. Given the real-valued weight function satisfy the condition above, for any g, where $n \to \infty$, we get

$$\begin{aligned}\phi_{k_n}(g) &= k_n\phi(g) + \mathrm{O}(1) \\ \phi'_{k_n}(g) &= \frac{1}{k_n}\phi(g') + \mathrm{O}\left(\frac{1}{k_n^2}\right).\end{aligned}$$

With all this in mind, our observations becomes

$$\begin{aligned}\bar{Y}_i^n &= \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)\Delta_{i+j-1}^n Y = -\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right)Y_{i+j-2}^n \\ \widehat{Y}_i^n &= \sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right)^2\left(\Delta_{i+j-1}^n Y\right)^2.\end{aligned}$$

Suitable estimators for IV_T is then (as soon as $t > k_n\Delta_n$)

$$RV^{\text{Preav}}\ (\Delta_n, k_n, g)_t = \frac{1}{\phi_{k_n}(g)}\frac{t}{t-k_n\Delta_n} \quad \times \sum_{i=1}^{\lfloor T/\Delta_n\rfloor - k_n + 1}\left(\left(\bar{Y}_i^n\right)^2 - \frac{1}{2}\widehat{Y}_i^n\right).$$

This leads to the formulation of the following theorem:

Theorem 5. *Let Assumption 1 be satisfied and choose k_n and g as given above. Then for each $T > 0$ we have the convergence in probability*

$$RV^{Preav}\ (\Delta_n, k_n, g)_T \xrightarrow{\mathbb{P}} IV_T.$$

6 Testing for jumps

We have mentioned earlier that empirical studies have observed jumps in financial data and that jumps are due to certain events in the market. The ability to separate and distinguish jumps and continuous paths have significance in e.g. allocation of assets and management of risk (Barndorff-Nielsen & Shephard 2006). High-frequency data can be used to estimate how large a contribution jumps and the continuous part have to the quadratic variation process. Since data are observed at discrete times of what is essential a continuous-time model, the existence of jumps can be hard to prove, but high-frequency data can be useful in this aspect.

First of all, it is important to find out if jumps are present in the stochastic process we are studying. A number of different tests for jumps in the price process have been published during the 2000's; e.g. Barndorff-Nielsen & Shephard (2006), Lee & Mykland (2008) and Aït-Sahalia & Jacod (2009). All these jump test methods have the assumption of a semimartingale volatility process.

The jump tests can also be used for testing for discontinuities in the volatility process in a SV-model, again in a semimartingale setup. Furthermore, co-jumps, that is, jumps occurring at

exactly the same time in both the price and volatility processes, can be found using a method described in Aït-Sahalia & Jacod (2012).

Jump test are constructed under the general hypothesis setup

$$\begin{aligned} H_0 &= \{\omega \in \Omega : t \mapsto X_t(\omega) \text{ is continuous on } [0,T]\} \\ H_1 &= \{\omega \in \Omega : t \mapsto X_t(\omega) \text{ is discontinuous on } [0,T]\}. \end{aligned} \tag{14}$$

The test should answer if the observed outcome of X during $[0,T]$ is in H_0 or H_1. The tests are non-parametric tests of the hypothesis that X have continuous sample paths (H_0). If we denote the jump process by $(Z_t)_{t\geq 0}$, we can reformulate the two hypothesis to

$$\begin{aligned} H_0 &= \{\omega \in \Omega : Z_t(\omega) = 0, \forall t \in [0,T]\} \\ H_1 &= \{\omega \in \Omega : \exists t \in [0,T] : \Delta Z_t(\omega) \neq 0\} \end{aligned} \tag{15}$$

and as in standard statistical hypothesis testing we can formulate one-sided tests, e.g. testing only for positive jumps. The H_1 is then modified according to this.

Applying jump tests can help decide which components to include in the models for data, i.e. only continuous parts, jump processes, and finite/infinite activity. Besides the test for presence of jump, there are methods for finding the location of jumps, see, e.g., Lee & Mykland (2008) and Aït-Sahalia & Jacod (2009).

In our fSV model setup where the log spot variance is driven by a fractional Brownian motion with Hurst parameter in $(0,1)$ the resulting process is not a semimartingale, unless $H = 0.5$, so we need a jump test for when the jumps are added to a fractional process with unknown H. This jump test is presented in a recent paper by Bibinger & Sonntag (2023).

6.1 Review of the method in Bibinger & Sonntag (2023)

We give a short review of the test for jumps in the fractional setup. Not only is a jump test presented, but also an estimator for H. We will not go into details regarding assumptions and proofs of theorems, corollaries, etc., but simply list the models, test statistics, the estimator for H and the asymptotic results. The paper should be consulted for detailed proofs and simulation/empirical results. In Sections 9 and 10 we will use the test both on simulated data and on empirical data.

Bibinger & Sonntag (2023) assume that the processes are observed on the interval $[0,1]$; while we assume observations in $[0,T]$. This is of course no problem, so in this subsection we shall go with $[0,1]$. We consider the following model for X

$$X_t = Y_t + Z_t = \int_0^t \sigma_s dB_s^H + Z_t, \quad t \geq 0,$$

with $(B_t^H)_{t\geq 0}$ being a fractional Brownian motion and $(Z_t^H)_{t\geq 0}$ being a càdlàg jump process, e.g. a compound Poisson process. In applications, X could be both a model for log-price or for log-spot variance.

X is observed at equidistant points in $[0,1]$, $n+1$ times in total, i.e.

$$X_{i/n} = Y_{i/n} + Z_{i/n}, \quad i = 0, \ldots, n.$$

It is noted that the paper assumes a driftless model. The reason for this is that a drift component will have no effect on the asymptotic fill-in results. Testing for jumps are based on second-order increments defined as

$$\Delta_{n,i}^{(2)} X = \left(X_{i/n} - 2X_{(i-1)/n} + X_{(i-2)/n}\right) = \Delta_{n,i}^{(2)} Y + \Delta_{n,i}^{(2)} Z, \quad i = 2, \ldots, n.$$

Since we assume $Z_0 = 0$, testing for jumps is of course not relevant at $t = 0$. The test statistic for the double-sided test is

$$R_n = \max_{k,j} \left(\frac{\left|\Delta_{n,j}^{(2)} X\right|}{\left(h_n^{-1} m_p^{-1} \sum_{i=(k-1)h_n+2}^{kh_n} \left|\Delta_{n,i}^{(2)} X\right|^p\right)^{\frac{1}{p}}} \right).$$

Here $m_p = \mathbb{E}(|X|^p)$, with $X \sim N(0,1)$ and $(h_n)_{n\in\mathbb{N}}$ is a sequence of natural numbers with

$$\lim_{n\to\infty} h_n = \infty$$

$$\lim_{n\to\infty} \frac{h_n}{n} = 0.$$

The number h_n/n is the bandwidth for the test. It is suggested that

$$p \leq 1$$

$$h_n \propto n^{\beta}$$

$$pH < \beta < 1.$$

Finally, the $\max_{k,j}$ symbol is defined as a calculation of a maximum (extreme) value. It is defined as

$$\max_{k,j} z_{i,j} := \max_{2\leq j<h_n} z_{1,j} \vee \max_{2\leq k\leq m} \left(\max_{(i-1)h_n \leq j < ih_n} z_{i,j} \right) \vee \max_{mh_n \leq j \leq n} z_{m+1,j}$$

where

$$a \vee b = \max\{a, b\}$$

$$\max_k z_k := \max_{1\leq k\leq m} z_k$$

$$m = \lfloor n/h_n \rfloor .$$

The test for jumps are evaluated with a Gumbel distribution. If we let

$$q_\alpha = -\log(-\log(1-\alpha))$$

be the $(1-\alpha)$-quantile of the Gumbel distribution the H_0 is rejected if

$$c_n (R_n - d_n) \geq q_\alpha$$

where

$$c_n = \sqrt{2\log(2n)}$$

$$d_n = \sqrt{2\log(2n)} - \frac{\log(\log(2n)) + \log(4\pi)}{2\sqrt{2\log(2n)}}.$$

The R_n-statistic is all about finding the maximum of rescaled absolute-value increments. The rescaling is with respect to the spot volatility estimate, calculated using standardized second-order increments (Bibinger & Sonntag 2023). The paper mentions that, if the true H is small, then only large jumps can be detected, whereas if H is large, the test can find smaller jumps.

The paper also introduces an estimator for H which is

$$\hat{H}_n = \frac{1}{2\log(2)} \log\left(\frac{\sum_{j=0}^{n-2} \left(X_{(j+2)/n} - X_{j/n}\right)^2}{\sum_{j=0}^{n-1} \left(X_{(j+1)/n} - X_{j/n}\right)^2} \right).$$

It is assumed that the jump process is an Itô semimartingale with finite quadratic variation and bounded jump sizes. Furthermore, Y and Z should be independent. Then it is proved that $\hat{H}_n$ is a consistent estimator of H, when $H \in (0, 1/2)$. If $H \in [1/2, 1)$ then $\hat{H}_n$ is not consistent, since, for Z being a Itô semimartingale with bounded jump sizes,

$$\hat{H}_n \to \frac{1}{2}, \quad n \to \infty.$$

Regarding the estimation of H, they find that standard estimates are asymptotically robust, but for observations on a finite time interval, the bias is positive and seems to be larger with greater H.

Finally, the paper also introduces an estimator for the location of jumps. Lee & Mykland (2008) have a similar estimator in the non-fractional case. The estimator in the paper uses their jump test statistic. The jump time estimator is given in the following proposition.

Proposition 2. *Under H_1, if there is one jump at time $\tau \in (0,1)$, and $\Delta Z_t = 0$, for all $t \neq \tau$, the estimator of the jump time is*

$$\hat{\tau}_n = \frac{1}{n} \operatorname{argmax}_{2 \leq j \leq n} \frac{\left|\Delta_{n,j}^{(2)} X\right|}{\left(h_n^{-1} m_p^{-1} \sum_{i=\lfloor j h_n^{-1} \rfloor}^{\lfloor j h_n^{-1} \rfloor + h_n} \left|\Delta_{n,j}^{(2)} X\right|^p\right)^{1/p}}.$$

Since we only observe X at discrete points in time, we can only say that jump time is in $[\frac{1}{n} - \tau, \tau + \frac{1}{n}]$. The jump time estimator can be used to filter out jumps, and the paper shows that there is a slight increase in $\hat{H}_n$ when done so. The paper also concludes that jumps do not change the belief of rough volatility, and it is possible to have robust inference on H and σ^2, for small H.

Note, if the main focus is to estimate the Hurst exponent H, then one has to see if the estimation routine is robust to the addition of a jump process or to filter out identified jumps (Bibinger & Sonntag 2023). If jumps (intensity and sizes) are the focus as well, we need elaborate estimation procedures to estimate the jump process parameters.

7 The fSV model setting

In this section, we will present and define in more detail the fSV model, which we will use as the basic model for stochastic volatility. The model setting was already presented at the beginning of Section 2. So, here the focus will be on the fractional Ornstein-Uhlenbeck (fOU) process.

The classical Ornstein-Uhlenbeck (OU) process has the Brownian motion as the driving process. As the OU process, the fOU process can be obtained either as the solution to the Langevin equation or as the Lamperti transformation (Cheridito et al. 2003). In addition to being stationary, it is ergodic and mean reverting, that is, as $T \to \infty$ the process will approach the long term mean.

The fOU has 3 parameters v, λ and H which are in $\phi \in \Phi \subset \mathbb{R}^3$. It is assumed that $v, \lambda > 0$ and $H \in (0,1)$. According to Garnier & Sølna (2018) the auto-covariance function (acf) can be expressed as

$$\kappa(\ell) = \frac{v^2}{2\lambda^{2H}} \left(\frac{1}{2} \int_{-\infty}^{\infty} e^{-|y|} |\lambda\ell + y|^{2H} \, \mathrm{d}y - |\lambda\ell|^{2H} \right), \quad \ell \geq 0. \tag{16}$$

When $H = 0.5$ we have

$$\kappa(\ell) = \frac{v^2}{2\lambda} e^{-\lambda\ell},$$

indicating an exponential decay for the acf. For other H-values the decay is hyperbolic (see Appendix A.2). From equation (16) we have that the variance of Y_t, $t \geq 0$ is

$$\kappa(0) = \frac{v^2}{2\lambda^{2H}} \Gamma(1 + 2H). \tag{17}$$

In Section 4 we have discussed quadratic variance and how to estimate the integrated variance (IV) defined as

$$IV = \int_0^t \sigma_s^2 ds, \quad t \geq 0.$$

According to this definition, the IV process will have information on all parameters of the fSV model (Bolko et al. 2023). We cannot measure IV directly, so it has to be replaced by the realized measure of volatility, see Section 4.

In this thesis, as in Bolko et al. (2023), we will define an estimation procedure where we use IV on a daily basis

$$IV_t = \int_{t-1}^{t} \sigma_s^2 \, \mathrm{d}s, \quad t \in \mathbb{Z}. \tag{18}$$

This defines a discrete process $(IV_t)_{t \in \mathbb{Z}}$.

7.1 General jump setup

In this thesis we will mainly add a compound Poisson process to either the price process, volatility process, or both. Jump sizes will follow a Gaussian law. According to the model setting described in Section 2, we will define the new processes as

$$X_t = \int_0^t \sigma_s dW_s + \sum_{n=1}^{N_t^{(X)}} J_n^{(X)} + \sum_{n=1}^{N_t^{(XY)}} J_n^{(XY)}$$

$$Y_t = v \int_0^t e^{-\lambda(t-s)} \mathrm{d}B_s^H + \sum_{n=1}^{N_t^{(Y)}} J_n^{(Y)} + \sum_{n=1}^{N_t^{(XY)}} J_n^{(YX)},$$

where we assume the the processes $W, B^H, N^{(X)}, N^{(Y)}, N^{(XY)}, J^{(X)}, J^{(Y)}$ are independent of each other.

$$N_t^{(X)} \sim Pois(\lambda^{(X)}), \quad N_0^{(X)} = 0$$
$$N_t^{(Y)} \sim Pois(\lambda^{(Y)}), \quad N_0^{(Y)} = 0$$
$$N_t^{(XY)} \sim Pois(\lambda^{(XY)}), \quad N_0^{(XY)} = 0$$
$$J_t^{(X} \overset{\text{iid}}{\sim} N\left(\mu^{(X)}, \sigma^{(X)^2}\right)$$
$$J_t^{(Y)} \overset{\text{iid}}{\sim} N\left(\mu^{(Y)}, \sigma^{(Y)^2}\right)$$
$$J_t^{(XY)} \overset{\text{iid}}{\sim} N\left(\mu^{(XY)}, \sigma^{(XY)^2}\right)$$
$$J_t^{(YX)} \overset{\text{iid}}{\sim} N\left(\mu^{(YX)}, \sigma^{(YX)^2}\right).$$

$J_t^{(XY)}$ and $J_t^{(YX)}$ are possibly correlated with correlation coefficient $\rho \in [0, 1]$. All parameters are homogeneous in time.

7.2 Moments of the $(IV_t)_{t\in\mathbb{N}}$ process

Within the model setting defined in (**??**) and (**??**) it is possible to derive the structure of the theoretical moments for the IV process. This is covered in great detail in Bolko et al. (2023), so here we will just give the general formula for calculating $\mathbb{E}[IV_t^r]$, the r'th moment of IV_t, with $r \geq 2$. Although IV_t is a sum of correlated log-normal distributed random variables implying that the determination of its probability distribution would be challenging, it is possible to find that

$$\mathbb{E}[IV_t^r] = r!\xi^r \int_0^1 \int_0^{x_1} \cdots \int_0^{x_{r-2}} (1 - x_1)\, g(x_1, \ldots, x_{r-1})\, \mathrm{d}x_{r-1} \ldots \mathrm{d}x_1, \tag{19}$$

see Theorem 2.1 in Bolko et al. (2023). The function g is

$$g(x_1, \ldots, x_{r-1}) = \exp\left(\sum_{1 \leq i < j \leq r-1} \kappa(|x_i - x_j|) + \sum_{i=1}^{r-1} \kappa(|x_i|) \right). \tag{20}$$

Furthermore

$$\mathbb{E}[IV_t] = \xi, \tag{21}$$

since $(IV_t)_{t\in\mathbb{N}}$ is stationary.

For the model setting considered in this thesis, the integrals in (19) need to be approximated or solved numerically. The acf of the spot variance has a steep incline near the origin. An incline that gets steeper for small H. This has the effect that computation of the third and higher moments is difficult to work with in practice. Furthermore, because of the complicated distribution of IV, it is difficult to estimate the moments, this becomes even more true when we are in a noisy setting where we experience error. The GMM estimation procedure, which we will present in the following sections, therefore rely on lower-order moments.

7.3 Jumps in log-price

We assume a model with a jump process in the log-price process X, i.e.

$$X_t = \int_0^t \sigma_s dW_s + Z_t, \quad t \geq 0,$$

and the Y process remains unchanged, so we assume $\lambda_J^{(Y)} = \lambda_J^{(XY)} = 0$. The fSV model has 7 parameters in total:

$$\xi, \lambda, v, H, \lambda_J^{(X)}, \mu_J^{(X)}, \sigma_J^{(X)}.$$

Let the jump parameters be given by $(\lambda_J^X, \mu_J^X, \sigma_J^X) \in \varsigma \subset \mathbb{R}^3$. We define the compact parameter space as $\vartheta = \Xi \times \Phi \times \varsigma \subset \mathbb{R}^{3+1+3}$. For no risk of confusion, we write $\theta = (\xi, \phi, \varsigma) \in \vartheta$.

According to Section 4, we have the following convergence in probability

$$RV_T \xrightarrow{P} [X,X]_T = QV_T = IV_T + JV_T = \int_0^T \sigma_s^2 ds + \sum_{k=1}^{N_T^{(X)}} {J_k^{(X)}}^2. \tag{22}$$

In this model setup RV can be used as an estimate for $[X,X]$, but not for IV as in the case where X has no jump process added. Define now as before

$$IV_t = \int_{t-1}^t \sigma_s^2 \, \mathrm{d}s, \quad t \in \mathbb{Z}$$

and additionally

$$JV_t = \sum_{k=N_{t-1}^{(X)}+1}^{N_t^{(X)}} {J_k^{(X)}}^2 = \sum_{k=1}^{N_t^{(X)} - N_{t-1}^{(X)}} {J^{(X)}}^2_{k+N_{t-1}^{(X)}} = \sum_{k=1}^{N_t^{(X)}} {J_k^{(X)}}^2 - \sum_{k=1}^{N_{t-1}^{(X)}} {J_k^{(X)}}^2.$$

Taking the expectation on the right side of (22) we get the first order moment

$$\mathbb{E}[QV_t] = \xi + \lambda_J^{(X)} \left({\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 \right).$$

Proof. First, since W and Z are independent we can write

$$\mathbb{E}\left[QV_t\right] = \mathbb{E}\left[IV_t\right] + \mathbb{E}\left[JV_t\right].$$

Note that as in Bolko et al. (2023), due to the stationarity of the variance process, the use of Fubini's theorem leads to

$$\mathbb{E}[IV_t] = \xi.$$

Next, we have that

$$\begin{aligned}
\mathbb{E}[JV_t] &= \mathbb{E}\left[\sum_{k=N_{t-1}^{(X)}+1}^{N_t^{(X)}} {J_k^{(X)}}^2 \right] = \mathbb{E}\left[\sum_{k=1}^{N_t^{(X)} - N_{t-1}^{(X)}} {J^{(X)}}^2_{k+N_{t-1}^{(X)}} \right] \\
&= \mathbb{E}\left[N_t^{(X)} - N_{t-1}^{(X)} \right] \mathbb{E}\left[{J_k^{(X)}}^2 \right] = \lambda_J^{(X)} \left({\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 \right).
\end{aligned}$$

Have we have used change of summation limits for the random sum of jump sizes, series summations, and the properties for the Poisson and compound Poisson process (see Appendix A.1). □

Regarding the second-order moments we have; for $\ell = 0$

$$\begin{aligned}
\mathbb{E}\left[QV_t^2\right] = 2\xi^2 \int_0^1 (1-y)\exp(\kappa(y))\mathrm{d}y + \lambda_J^{(X)} \left({\mu_J^{(X)}}^4 + 6{\mu_J^{(X)}}^2 {\sigma_J^{(X)}}^2 + 3{\sigma_J^{(X)}}^4 \right) \\
+ {\lambda_J^{(X)}}^2 \left({\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 \right)^2 + 2\xi\lambda_J^{(X)} \left({\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 \right)
\end{aligned} \tag{23}$$

and for $\ell > 0$

$$\mathbb{E}[QV_t QV_{t+\ell}] = \xi^2 \int_0^1 (1-y)[\exp(\kappa(\ell+y)) + \exp(\kappa(|\ell-y|))]\mathrm{d}y + {\lambda_J^{(X)}}^2 \left({\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 \right)^2 + 2\xi\lambda_J^{(X)} \left({\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 \right).$$

Proof. Independence give

$$\mathbb{E}\left[QV_t^2\right] = \mathbb{E}\left[IV_t^2\right] + 2\mathbb{E}[IV_t]\mathbb{E}[JV_t] + \mathbb{E}\left[JV_t^2\right].$$

By standard probability theory we have

$$\mathbb{E}\left[JV_tJV_t\right] = Var(JV_t) + \left(\mathbb{E}[JV_t]\right)^2$$

and

$$Var(JV_t) = \mathbb{E}\left[\sum_{k=1}^{N_t^{(X)}-N_{t-1}^{(X)}} J_{k+N_{t-1}^{(X)}}^{(X)^4}\right] = \lambda_J^{(X)}\mathbb{E}\left[J_t^{(X)^4}\right] = \lambda_J^{(X)}\left(\mu_J^{(X)^4} + 6\mu_J^{(X)^2}\sigma_J^{(X)^2} + 3\sigma_J^{(X)^4}\right).$$

Next for $\ell > 0$

$$\mathbb{E}\left[QV_tQV_{t+\ell}\right] = \mathbb{E}\left[IV_tIV_{t+\ell}\right] + \mathbb{E}\left[IV_tJV_{t+\ell}\right] + \mathbb{E}\left[IV_{t+\ell}JV_t\right] + \mathbb{E}\left[JV_tJV_{t+\ell}\right].$$

Since jumps are independent we have

$$\mathbb{E}\left[JV_tJV_{t+\ell}\right] = \lambda_J^{(X)^2}\left(\mu_J^{(X)^2} + \sigma_J^{(X)^2}\right)^2,$$

which complete the proof. □

7.4 Jumps in volatility

Now we turn to the case where the fSV model has a jump process in the volatility process. It is assumed that

$$\lambda_J^{(X)} = \lambda_J^{(XY)} = 0.$$

The fSV model has 7 parameters:

$$\xi, \lambda, v, H, \lambda_J^{(Y)}, \mu_J^{(Y)}, \sigma_J^{(Y)}.$$

Les us denote

$$\sigma_t^{c^2} = \xi \exp\left(Y_t^c - \frac{1}{2}\kappa(0)\right), \quad t \geq 0$$

and

$$\sigma_t^{d^2} = \prod_{k=1}^{N_t^{(Y)}} \exp\left(J_k^{(y)}\right) = exp\left(\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right), \quad t \geq 0$$

with

$$Y_t^c = v\int_0^t e^{-\lambda(t-s)}\mathrm{d}B_s^H, \quad t \geq 0$$
$$\kappa(t) = \mathrm{cov}\left(Y_0^c, Y_t^c\right), t \geq 0.$$

The $(\sigma_t^{c^2})_{t\geq 0}$ process is exactly the same as the volatility process in Bolko et al. (2023), Section 2. Regarding the moments for $\sigma_t^{d^2}$ we have

$$\mathbb{E}\left[\sigma_t^{d^2}\right] = \exp\left(\lambda_J^{(Y)} \cdot t\left(\exp\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right) - 1\right)\right) \tag{24}$$

$$\mathbb{E}\left[\sigma_t^{d^2} \cdot \sigma_t^{d^2}\right] = \exp\left(\lambda_J^{(Y)} \cdot t\left(\exp\left(2\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)}\right)\right) - 1\right)\right) \tag{25}$$

$$\begin{aligned}\mathbb{E}\left[\sigma_t^{d^2} \cdot \sigma_{t+\ell}^{d^2}\right] =& \exp\left(\lambda_J^{(Y)} \cdot t\left(\exp\left(2\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right)\right) - 1\right)\right) \\ &\cdot \exp\left(\lambda_J^{(Y)} \cdot \ell\left(\exp\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right)^2 - 1\right)\right).\end{aligned} \tag{26}$$

Proof. Equation (24) is proved by

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right)\right] &= \sum_{n\geq 0}\mathbb{E}\left[\exp\left(\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right) \mid N_t^{(n)} = n\right]\cdot P\left(N_t^{(Y)=n}\right) \\
&= e^{-\lambda_J^{(Y)}\cdot t}\sum_{n\geq 0}\frac{\lambda_j^{(y)^n}}{n!}t^n \cdot E\left[\exp\left(\sum_{k=1}^{n} J_k^{(Y)}\right)\right] \\
&= e^{-\lambda_J^{(Y)}\cdot t}\sum_{n\geq 0}\frac{\lambda_J^{(Y)^n}}{n!}t^n \cdot \prod_{k=1}^{n} E\left[e^{J_k}\right] \\
&= e^{-\lambda_J^{(Y)}\cdot t}\sum_{n\geq 0}\frac{\lambda_J^{(Y)^n}}{n!}\cdot\left(E\left[e^{J_k}\right]\right)^n \\
&= \exp\left(\lambda_J^{(Y)} t\left(\exp\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right) - 1\right)\right).
\end{aligned}$$

Equation (25) is proved is similar fashion by noting that

$$\mathbb{E}\left[\exp\left(\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right)\right]\mathbb{E}\left[\exp\left(\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right)\right] = \mathbb{E}\left[\exp\left(2\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right)\right]$$

Finally, Equation (26) is obtained by

$$\mathbb{E}\left[\exp\left(\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right)\right]\mathbb{E}\left[\exp\left(\sum_{k=1}^{N_{t+\ell}^{(Y)}} J_k^{(Y)}\right)\right] = \mathbb{E}\left[\exp\left(2\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}\right)\cdot\exp\left(\sum_{k=N_t^{(Y)}+1}^{N_{t+\ell}^{(Y)}} J_k^{(Y)}\right)\right]$$

using the independence of jump sizes. □

A few calculations regarding the $(Y_t)_{t\geq 0}$ process

$$\mathbb{E}[Y_t] = t\lambda_J^{(Y)}\mu_J^{(Y)}$$

$$\mathbb{E}[Y_t + Y_s] = (t+s)\lambda_J^{(Y)}\mu_J^{(Y)} = \mathbb{E}[Y_t + s].$$

The covariance function of $(Y_t)_{t\geq 0}$ is, with $\sum_{k=1}^{N_t^{(Y)}} J_k^{(Y)}$

$$Cov(Y_t, Y_{t+\ell}) = \kappa(\ell) + Cov(Z_t, Z_{t+\ell}) = \kappa(\ell) + t\lambda_J^{(Y)}\left(\mu_J^{(Y)^2} + \sigma_J^{(Y)^2}\right).$$

The challenge is to calculate the moments for IV. The moments in which we are interested are $\mathbb{E}[IV_t]$ and $\mathbb{E}[IV_t IV_{t+\ell}]$, $\ell \geq 0$. Being that they now depend on t, each moment has to be calculated at every point in time. We have tried a modified GMM procedure with the following theoretical moments

$$\mathbb{E}[IV_t] = \xi\exp\left(\lambda_J^{(Y)}\left(\exp\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right) - 1\right)\right)$$

$$\mathbb{E}[IV_t^2] = \mathbb{E}[IV_t^{c^2}]\cdot\exp\left(\lambda_J^{(Y)}\left(\exp\left(2\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)}\right)\right) - 1\right)\right)$$

$$\begin{aligned}
\mathbb{E}[IV_t IV_{t+\ell}] =& \mathbb{E}[IV_t^c IV_{t+\ell}^c]\cdot\exp\left(\lambda_J^{(Y)}\left(\exp\left(2\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right)\right) - 1\right)\right) \\
&\cdot\exp\left(\lambda_J^{(Y)}\cdot\ell\left(\exp\left(\mu_J^{(Y)} + \frac{1}{2}\sigma_J^{(Y)^2}\right)^2 - 1\right)\right).
\end{aligned}$$

These moments are probably not the correct values of the complicated integrals; but we are interested in running GMM to see if the estimates were close to the correct values. Unfortunately, the results on the simulated data were not satisfactory.

7.5 Note regarding a fSV model with co-jumps

The model with co-jumps. i.e., adding a jump process in both the X and Y processes with jumps occurring at exactly the same time for both processes, will not be treated here in great detail, but will be left for future work. Here, we will just define the model with a few assumptions.

The jumps sizes are assumed to be dependent, i.e. we assume

$$X_t = \int_0^t \sigma_s dW_s + \sum_{k=1}^{N_T^{(XY)}} J_k^{(XY)} \quad t \geq 0$$

and

$$Y_t = v \int_0^t e^{-\lambda(t-s)} \mathrm{d}B_s^H + \sum_{k=1}^{N_i^{XY}} J_k^{(YX)} \quad t \geq 0.$$

When the jump sizes are dependent, we have

$$Cov\left(J_t^{(XY)}, J_t^{(YX)}\right) = \rho_j \sigma^{(XY)} \sigma^{(YX)}.$$

The jump sizes can have different distributions but occur at the same time according to $(N_t^{(XY)})_{t\geq 0}$. This fSV model has 10 parameters:

$$\left(\xi, \lambda, v, H, \lambda_J^{(XY)}, \rho_j, \mu_J^{(XY)}, \mu_J^{(YX)}, \sigma_J^{(XY)}, \sigma_J^{(YX)}\right).$$

The model can be even more general with additional jump processes in both X and Y, which are independent of the co-jump processes.

8 GMM estimation procedure

In this section, we will now define the GMM estimation setup that we will be utilized for Monte Carlo simulations and empirical analysis. The first subsection follows from Bolko et al. (2023), while the method is modified to include the presence of price jumps in the second subsection. A general treatment of GMM for time series is in Appendix B.

8.1 GMM estimation presented in Bolko et al. (2023)

In Section 7 we defined a discrete process $(IV_t)_{t\in\mathbb{Z}}$ that is based on integrated variance for one day t. The integrated variance is the integral of the spot variance process $(\sigma_t^2)_{t\geq 0}$ and the spot variance is the exponential of a fOU process Y. First consider the Gaussian process Y, which has the following main assumptions (Assumption 1 in Bolko et al. (2023)).

Assumption 2. *The Gaussian process Y and its autocovariance function κ satisfy the following conditions:*

(i) Y has continuous sample paths for any $\phi \in \Phi$

(ii) $(u, \phi) \mapsto \kappa_\phi(u)$ is a continuous function.

Condition (i) is to make sure that Y is bounded. Condition (ii) guarantees continuous moments of the Y process with respect to θ. When Y is an fOU process, Proposition 3.4 of Kaarakka & Salminen (2011) shows condition (i) and applying the dominated convergence theorem to Equation (16) verifies condition (ii).

Stationarity and ergodicity of the fOU process is proven in Cheridito et al. (2003). Given that Y is a stationary Gaussian process, σ and IV are stationary and ergodic as well.

The spot variance $\sigma^2 = \left(\sigma_t^2\right)_{t\geq 0}$ depends on the parameter vector $\xi, \phi) \in \Theta$ and $\mathcal{F}^\sigma$ denotes the σ-algebra generated by σ^2. $\mathbb{E}_\theta$ is the expectation operator induced by θ and $\mathbb{P}_\theta$ is the corresponding

probability measure. $(IV_t)_{t\in\mathbb{Z}}$ is a latent variable and following Bolko et al. (2023) we postulate that we observe a noisy proxy of IV_t, which we will denote $\widehat{IV}_t$ and assume can be written as

$$\widehat{IV}_t = IV_t + \varepsilon_t.$$

The random variable ε_t will represent the random size of measurement error. The error process $(\varepsilon_t)_{t\in\mathbb{Z}}$ must satisfy the conditions in the following assumption (Assumption 2 in Bolko et al. (2023)):

Assumption 3. *The processes $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ satisfy the following conditions:*

(i) $(IV_t, \varepsilon_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic process under $\mathbb{P}_\theta$ for any $\theta \in \Theta$

(ii) $\theta \mapsto c(\theta) \equiv \mathbb{E}_\theta\left[\varepsilon_1^2\right]$ is a finite-valued, continuous function on Θ

(iii) $\mathbb{E}_\theta\left[\varepsilon_t \mid \mathcal{F}_{t-1}^{\sigma,\varepsilon}\right] = 0$ for any $t \in \mathbb{Z}$ and any $\theta \in \Theta$.

If $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are weakly mixing (Lindgren 2006), then condition (i) is fulfilled. The stationarity and ergodicity of $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ as separate processes follow directly from the same properties of Y. In Assumption 3 the filtration $(\mathcal{F}_t^{\sigma,\varepsilon})_{t\in\mathbb{Z}}$ is defined as

$$\mathcal{F}_t^{\sigma,\varepsilon} = \mathcal{F}_t^{\sigma} \vee \mathcal{F}_t^{\varepsilon}$$

with

$$\mathcal{F}_t^{\varepsilon} = \sigma\left(\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}\right), t \in \mathbb{Z},$$

being the σ-algebra generated by the errors up to time t, see Section 2. Assumption 3 gives the long-term properties of the joint process $(IV_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$.

Condition (iii) ensures that $\widehat{IV}_t$ is unbiased. Condition (ii) can be checked on a case-by-case basis and for realized variance and bipower variation it is already done in Bolko et al. (2023). We provide some evidence for condition (ii) for the case of truncated realized variance in the following example.

Example 1. *Truncated Realized Variance*

Suppose that we want to estimate IV using truncated realized variance:

$$RV_t^{u_n,n} = \sum_{i=1}^{n}\left(X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i}{n}-1}\right)^2 1_{\left\{|\Delta_i^n X| \le u_n\right\}}. \tag{27}$$

We need to establish the asymptotic distribution of

$$RV_t^{u_n,n} - IV_t.$$

By using theorems and propositions in Section 3.2 of Mancini (2009) we get

$$\sqrt{n}\left(RV_t^{u_n,n} - IV_t\right) \xrightarrow{D} N(0, 2IQ_t).$$

This entails that we can apply the same correction for the measurement error as for realized variance in Example 3.1 in Bolko et al. (2023), namely:

$$c(\theta) = \frac{2\xi^2}{n}\exp\left(\kappa_\phi(0)\right). \tag{28}$$

8.1.1 Setting up the GMM estimation procedure for the fSV model

In this segment, we establish the GMM framework and deal with consistency of the fSV model defined earlier in this thesis. Definitions, assumptions, and theorems follow closely Bolko et al. (2023).

We start by assuming $\theta \in \Theta, t \in \mathbb{Z}$ and $\ell \in \mathbb{Z}$. We introduce the moment structure needed for the GMM procedure. We define a number of functions as

$$g_0^{(1)}(\theta) = \mathbb{E}_\theta\left[IV_t\right], \quad g_0^{(2)}(\theta) = \mathbb{E}_\theta\left[IV_t^2\right], \quad g_\ell(\theta) = \mathbb{E}_\theta\left[IV_t IV_{t-\ell}\right]$$

with index $\ell = 1, \ldots, k$ indicating days of lag.

We arrange these functions in a column vector denoted

$$G(\theta) = \left(g_0^{(1)}(\theta), g_0^{(2)}(\theta), g_1(\theta), \ldots, g_k(\theta)\right)^\top$$

where $\top$ is the transpose operator. In the same way, we define two column vectors regarding IV as

$$\begin{aligned}\mathbb{IV}_t &= \left(IV_t, IV_t^2, IV_t IV_{t-1}, \ldots, IV_t IV_{t-k}\right)^\top \\ \widehat{\mathbb{IV}}_t &= \left(\widehat{IV}_t, \widehat{IV}_t^2, \widehat{IV}_t \widehat{IV}_{t-1}, \ldots, \widehat{IV}_t \widehat{IV}_{t-k}\right)^\top .\end{aligned}$$

These two IV-processes can be seen to be stationary and ergodic according to condition (i) in Assumption 3. Applying condition (ii)-(iii) in the same assumption we get that

$$\begin{aligned}\mathbb{E}_\theta\left[\widehat{IV}_t\right] &= g_0^{(1)}(\theta) \\ \mathbb{E}_\theta\left[\widehat{IV}_t \widehat{IV}_{t-\ell}\right] &= \begin{cases} g_0^{(2)}(\theta) + c(\theta), & \ell = 0 \\ g_\ell(\theta), & \ell \neq 0. \end{cases}\end{aligned}$$

Here the function c is from the proxy with measurement error and this will only impact the second moment of $\widehat{IV}_t$ according to the assumption regarding the noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$. This fact was noted in Bolko et al. (2023), which includes the error directly in the model compared to Bollerslev & Zhou (2002), which adds a nuisance parameter.

Because it is assumed that errors are serially uncorrelated and mean zero together with the linearity property of the expectation operator ensuring unbiasedness of first and second order moments (other than $g_0^{(2)}(\theta)$), we could avoid the negative impact of measurement errors. To avoid systematic bias in the estimated parameter values, it is necessary to correct the moments as defined above. Moments with low orders, like variance and the absolute value, will contain a great deal of information on the parameters, which is why we include the corrected second-order moment which is easier to work with than the absolute value. So we collect the function G as follows

$$G_c(\theta) = G(\theta) + (0, c(\theta), 0, \ldots, 0)^\top$$

to be used with $(\widehat{IV}_t)_{t\in\mathbb{Z}}$. Define the random function $\widehat{m}_T(\theta)$ as

$$\widehat{m}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \widehat{\mathbb{IV}}_t - G_c(\theta).$$

Taking the expectation with respect to θ_0 which is the true value of θ then

$$\mathbb{E}_{\theta_0}\left[\widehat{m}_T(\theta)\right] = G_c\left(\theta_0\right) - G_c(\theta) \equiv m(\theta).$$

We have that $m\left(\theta_0\right) = 0$, and the GMM estimator can now be defined as

$$\widehat{\theta}_T = \arg\min_\theta \ \widehat{m}_T(\theta)^\top \mathbb{W}_T \widehat{m}_T(\theta)$$

where $(\widehat{IV}_t)_{t\in\mathbb{Z}}$ is "observed" for T days. W_t is a random $(k+2)\times(k+2)$ weight matrix. The consistency of $\hat{\theta}_t$ is ensured by some additional conditions. The first is the same as Assumption 3 in Bolko et al. (2023)

Assumption 4. *$\mathbb{W}_T = A_T^\top A_T$ for a random $(k+2)\times(k+2)$ matrix A_T, which under $\mathbb{P}_{\theta_0}$ converges almost surely to a non-random matrix A as $T \to \infty$.*

This condition define the limiting behavior of $\mathbb{W}$. Recognizing that a matrix $\mathbb{W}_T$ is positive-definite if there exists an invertible matrix A such that $\mathbb{W}_T = A_T^\top A_T$.

Assumption 5. *$Am(\theta) = 0$ if and only if $\theta = \theta_0$.*

This assumption ensure a unique solution, but since we do not have the moments on a algebraic form the uniqueness is hard to check. Appendix B in Bolko et al. (2023) provides a partial identification result.

Consistency can now be formulated by the following theorem:

Theorem 6. *Suppose Assumptions 2-5 hold. As* $T \to \infty$

$$\widehat{\theta}_T \xrightarrow{P} \theta_0. \tag{29}$$

The number of observations per day, n, is fixed in the analysis above. This also relies on the noisy proxy idea (Bolko et al. 2023).

A short proof of Theorem 6 is provided in the Appendix of Bolko et al. (2023).

We will now cover the theory for the case where, in a double-asymptotic setting $T \to \infty$ and $n \to \infty$. Define

$$\mathbb{V}_t^n = \left(V_t^n, (V_t^n)^2, V_t^n V_{t-1}^n, \ldots, V_t^n V_{t-k}^n\right)^\top$$

as a consistent realised measure of integrated variance for a fixed $k \in \mathbb{Z}$. This has the associated sample moments

$$\tilde{m}_{n,T}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{V}_t^n - G(\theta).$$

Note that we use the function $G(\theta)$ and not $G_c(\theta)$, because as $n \to \infty$ the error vanishes and the result is therefore unaffected by this choice. The GMM estimator is now

$$\tilde{\theta}_{n,T} = \arg\min_{\theta \in \Theta} \tilde{m}_{n,T}(\theta)^\top \mathbb{W}_{n,T} \tilde{m}_{n,T}(\theta).$$

We can now replace Assumption 3 with the following (Bolko et al. 2023):

Assumption 6. *The processes* $(IV_t)_{t\in\mathbb{Z}}$ *and* $(V_t^n)_{t\in\mathbb{Z},n\in\mathbb{N}}$ *admit the following:*
(i) $(IV_t)_{t\in\mathbb{Z}}$ *is a stationary and ergodic process under* $\mathbb{P}_\theta$ *for any* $\theta \in \Theta$,
(ii) $\sup_{t\in\mathbb{Z}} \mathbb{E}_{\theta_0}\left[(V_t^n - IV_t)^2\right] \to 0$ *as* $n \to \infty$.

And the modified consistency theorem.

Theorem 7. *Suppose Assumptions 2, 4-6 hold. As* $T \to \infty$ *and* $n \to \infty$

$$\tilde{\theta}_{n,T} \xrightarrow{\mathbb{P}} \theta_0. \tag{30}$$

In Bolko et al. (2023) it is demonstrated that

$$\sup_{t\in\mathbb{Z}} \mathbb{E}\left[(RV_t^n - IV_t)^2\right] \leq Cn^{-1}$$

for some $C > 0$ and subject to the boundedness condition imposed on both the drift and volatility. If this is obtained, then Assumption 6 is satisfied for realized variance and truncated realized variance.

8.1.2 Asymptotic normality

As Bolko et al. (2023), we assume that under $\mathbb{P}_{\theta_0}$ the Gaussian process Y admits a causal moving average representation

$$Y_t = \int_{-\infty}^{t} K(t-u)\mathrm{d}B_u, \quad t \in \mathbb{R} \tag{31}$$

for a two-sided standard Brownian motion $B = (B_t)_{t\in\mathbb{R}}$, that defined as

$$B_t = \begin{cases} B_t^1, & t > 0 \\ 0, & t = 0 \\ B_{-t}^2, & t < 0 \end{cases}$$

where B^1 and B^2 represent two independent Brownian motions, and a measurable kernel $K{:}(0,\infty) \to \mathbb{R}$ such that $\int_0^\infty K(u)^2 \, \mathrm{d}u < \infty$. We do this for technical reasons to establish the asymptotic normality of our GMM estimator (Bolko et al. 2023). The fOU process can be represented in this manner; consequently, the fSV model presented in this thesis adheres to this structure. The long-term memory of Y is governed by the asymptotic behavior of $K(u)$ as $u \to \infty$ (Bolko et al. 2023). Therefore, we restrict this memory to derive the asymptotic normality of the GMM estimator.

Assumption 7. *$K(u) = O\left(u^{-\gamma}\right)$ as $u \to \infty$ for some $\gamma > 1$.*

Consequently, the fSV model requires $H \leq 1/2$ to comply with Assumption 5, thereby only allowing for the case of rough volatility (Garnier & Sølna 2018). For a discussion of why this constraint is nearly optimal, see Bolko et al. (2023).

We now introduce a stronger assumption about the error process $(\varepsilon_t)_{t\in\mathbb{Z}}$. Following the notation of Bolko et al. (2023), we write $\|X\|_{L^2(\mathbb{P}_\theta)} = \mathbb{E}_\theta\left[X^2\right]^{1/2}$ for any square integrable random variable X and work with the filtrations $\mathcal{F}_t^{\widehat{\mathbb{IV}}} = \sigma\left\{\widehat{\mathbb{IV}}_t, \widehat{\mathbb{IV}}_{t-1}, \ldots\right\}$ and $\mathcal{F}_t^{B,\varepsilon} = \sigma\left\{\varepsilon_t, \varepsilon_{t-1}, \ldots\right\} \vee \sigma\left\{B_u : u \leq t\right\}, t \in \mathbb{Z}$.

Assumption 8. *The processes B and $(\varepsilon_t)_{t\in\mathbb{Z}}$ satisfy the following conditions:*
(i) $\mathbb{E}\left[\varepsilon_1^4\right] < \infty$
(ii) $\left\|\mathbb{E}_{\theta_0}\left[\varepsilon_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{IV}}}\right] - \mathbb{E}_{\theta_0}\left[\varepsilon_1^2\right]\right\|_{L^2\left(\mathbb{P}_{\theta_0}\right)} = O\left(r^{-\gamma+1/2}\right)$ *as $r \to \infty$*
(iii) B has independent increments with respect to $\left(\mathcal{F}_t^{B,\varepsilon}\right)_{t\in\mathbb{Z}}$ (i.e., for any $t \in \mathbb{Z}$ the process $\left(B_u - B_t\right)_{u\geq t}$ is independent of $\mathcal{F}_t^{B,\varepsilon}$).

Condition (ii) imposes constraints on the memory embedded in the squared measurement error (Bolko et al. 2023).

Proposition 3. *Condition (ii) in Assumption 8 is valid for Example 1, if Assumptions 2 and 7 hold.*

For proof, see Proposition A.6 in Bolko et al. (2023). As noted in Example 1, the error term is the same as for Example 3.1 in Bolko et al. (2023).

Let us now, for the sample mean of our statistic, present the CLT:

Proposition 4. *Suppose that Assumptions 2, 3, 7 and 8 hold. Then, as $T \to \infty$, under $\mathbb{P}_{\theta_0}$*

$$\sqrt{T}\widehat{m}_T\left(\theta_0\right) \xrightarrow{D} N\left(0, \Sigma_{\widehat{\mathbb{IV}}}\right) \tag{32}$$

where

$$\Sigma_{\widehat{\mathbb{IV}}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\widehat{\mathbb{IV}}}(\ell) \tag{33}$$

and

$$\Gamma_{\widehat{\mathbb{IV}}}(\ell) = \mathbb{E}_{\theta_0}\left[\left(\widehat{\mathbb{IV}}_1 - G_c\left(\theta_0\right)\right)\left(\widehat{\mathbb{IV}}_{1+\ell} - G_c\left(\theta_0\right)\right)^\top\right]. \tag{34}$$

The final assumption governing the CLT for our GMM estimator is introduced. First, we introduce the function $\mathbf{g} : \mathbb{R}^{k+2} \times \Theta \to \mathbb{R}$ via $\mathbf{g}(x, \theta) = x - G_c(\theta)$.

Assumption 9. *It holds that*
(i) θ_0 is an interior point of Θ
(ii) $J^\top \mathbb{W} J$ is non-singular, where $J = \mathbb{E}_{\theta_0}\left[\nabla_\theta \mathbf{g}\left(\widehat{\mathbb{IV}}_1, \theta_0\right)\right]$ and $\mathbb{W} = A^\top A$
(iii) The function $\theta \mapsto \mathbf{g}(x, \theta)$ is continuously differentiable. In addition, $\mathbb{E}_{\theta_0}\left[\left\|\mathbf{g}\left(\widehat{\mathbb{IV}}_1, \theta_0\right)\right\|^2\right] < \infty$ and $\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta}\left\|\nabla_\theta \mathbf{g}\left(\widehat{\mathbb{IV}}_1, \theta\right)\right\|\right] < \infty$.

Condition (i) ensures that there exists a centered open ball around θ_0 completely contained in Θ. This will help us to determine that θ_0 is indeed a minimum. The same condition is assumed in Barboza & Viens (2017) and Newey & McFadden (1994). For a proof of condition (ii), see Barboza & Viens (2017). Note that J does not depend on t as IV is stationary. Condition (iii) is presented as a condition in Theorem 3.4 in Newey & McFadden (1994).

We can now present the asymptotic distribution of $\widehat{\theta}_T$.

Theorem 8. *Suppose Assumptions 2-5 and 7-9 hold. As $T \to \infty$,*

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{D} N\left(0, \left(J^\top \mathbb{W} J\right)^{-1} J^\top \mathbb{W} \Sigma_{\widehat{\mathbb{V}}} \mathbb{W} J \left(J^\top \mathbb{W} J\right)^{-1}\right). \tag{35}$$

In the customary fashion, to minimize the asymptotic variance in (35) and obtain the efficient GMM estimator, given the moment conditions, we opt for an optimal weight matrix represented by the inverse of a consistent estimator for $\Sigma_{\mathbb{V}}$. Following Bolko et al. (2023) we propose the following HAC-type estimator:

$$\widehat{\Sigma}_T = \widehat{\Gamma}(0) + \sum_{\ell=1}^{T-1} w(\ell/L)\left[\widehat{\Gamma}(\ell) + \widehat{\Gamma}(\ell)^\top\right] \tag{36}$$

where w is a weight function, $L = o\left(T^{1/2}\right)$ is the lag length, and

$$\widehat{\Gamma}(\ell) = \frac{1}{T}\sum_{t=1}^{T-l}\left(\widehat{\mathbb{V}}_t - G_c\left(\widehat{\theta}_T\right)\right)^\top \left(\widehat{\mathbb{V}}_{t+\ell} - G_c\left(\widehat{\theta}_T\right)\right). \tag{37}$$

We impose the same weak regularity conditions on w as Bolko et al. (2023).

Assumption 10. *It holds that*
(i) $w(0) = 1$ *and* $\sup_{x\geq 0}|w(x)| < \infty$
(ii) w *is continuous at 0*
(iii) $\int_0^\infty \bar{w}(x)\mathrm{d}x < \infty$, *where* $\bar{w}(x) = \sup_{y\geq x}|w(y)|$.

These follow almost directly from Davidson (2020), which also assume $w(-x) = w(x)$, for all $x \geq 0$. The assumption 10 is required for the consistency of the HAC estimator. The assumption will hold for both the Bartlett and Parzen kernels, which are the two kernels that we consider and use in this thesis. The kernels have the following form (Andrews 1991):

$$\begin{aligned}
&\text{Bartlett:} \quad w_{BT}(x) = \begin{cases} 1-|x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \\
&\text{Parzen:} \quad w_{PR}(x) = \begin{cases} 1-6x^2+6|x|^3 & \text{for } 0 \leq |x| \leq 1/2 \\ 2(1-|x|)^3 & \text{for } 1/2 < |x| \leq 1 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{38}$$

Theorem 9. *Suppose Assumptions 2-5 and 7-10 hold. As $T \to \infty$*

$$\widehat{\Sigma}_T \xrightarrow{\mathbb{P}} \Sigma_{\widehat{\mathbb{V}}}. \tag{39}$$

Following with setting $\mathbb{W}_T = \widehat{\Sigma}_T^{-1}$ we have

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{D} N\left(0, \left(J^\top \Sigma_{\mathbb{V}}^{-1} J\right)^{-1}\right). \tag{40}$$

Plugging in $\widehat{\Sigma}_T^{-1}$ into (40), we can make inference on the parameters (Bolko et al. 2023).

Given we have quadratic forms of multivariate normal variables, the asymptotic distribution of the product of the minimized objective function value and the sample size is chi-square (Bolko et al. 2023)

$$\mathcal{J}_{\text{HS}} = T\widehat{m}_T\left(\widehat{\theta}_T\right)^\top \widehat{\Sigma}_T^{-1}\widehat{m}_T\left(\widehat{\theta}_T\right) \xrightarrow{D} \chi^2(k-p+1) \tag{41}$$

where $k-p+1$ is the number of overidentifying restrictions. This enables a Sargan–Hansen omnibus specification test for the model (Bolko et al. 2023).

Closing this section, in the double-asymptotic setting, where $T \to \infty$ and $n \to \infty$ (such that the discretization error is negligible), we study the CLT of the GMM estimator (Bolko et al. 2023).

Assumption 11. *The processes* $(IV_t)_{t\in\mathbb{Z}}$ *and* $(V_t^n)_{t\in\mathbb{Z},n\in\mathbb{N}}$ *satisfy the following conditions:*
(i) $(IV_t)_{t\in\mathbb{Z}}$ *is a stationary and ergodic process under* $\mathbb{P}_\theta$ *for any* $\theta\in\Theta$.
(ii) $\sup_{t\in\mathbb{Z}}\mathbb{E}_{\theta_0}\left[\left(\sqrt{T}\left(V_t^n-IV_t\right)\right)^2\right]\to 0$ *as* $T\to\infty$ *and* $n\to\infty$.

Introducing a HAC-type estimator $\widehat{\Sigma}_{n,T}$, which merely substitutes $\widehat{\mathbb{IV}}_t - G_c$ with $\mathbb{V}_t^n - G$ and then we take $\mathbb{W}_{n,T}=\widehat{\Sigma}_{n,T}^{-1}$ we arrive the following theorem (Bolko et al. 2023):

Theorem 10. *Suppose Assumptions 2,4-5,7,9-11 hold. As* $T\to\infty$ *and* $n\to\infty$

$$\sqrt{T}\left(\tilde{\theta}_{n,T}-\theta_0\right)\xrightarrow{d}N\left(0,\left(\widetilde{J}^\top\Sigma_{\mathbb{IV}}^{-1}\widetilde{J}\right)^{-1}\right), \tag{42}$$

where

$$\Sigma_{\mathbb{IV}}=\sum_{\ell=-\infty}^{\infty}\Gamma_{\mathbb{IV}}(\ell) \tag{43}$$

and

$$\Gamma_{\mathbb{IV}}(\ell)=\mathbb{E}_{\theta_0}\left[\left(\mathbb{IV}_1-G\left(\theta_0\right)\right)\left(\mathbb{IV}_{1+\ell}-G\left(\theta_0\right)\right)^\top\right] \tag{44}$$

with $\widetilde{J}=\mathbb{E}_{\theta_0}\left[\nabla_\theta \mathbf{g}\left(\mathbb{IV}_1,\theta_0\right)\right]$.

Note that Theorems 7 and 10 also hold for the bias-corrected estimator $\hat{\theta}_T$, although it has to fulfill the additional conditions imposed by Assumption 6(ii) for consistency and 11(ii) for asymptotic normality, as $T\to\infty$ and $n\to\infty$ (Bolko et al. 2023). Here, the bias correction disappears under the double-asymptotic framework. As in Bolko et al. (2023), we advocate for the inclusion of the correction as a precautionary measure, as its addition does not entail any loss.

8.2 GMM estimation procedure with jumps in X

The empirical evidence of jumps in the log-price process led us to investigate whether the GMM approach of Bolko et al. (2023) can be extended to provide estimates for the jump process parameters and also to improve the estimation of H. The jump parameters could be important in some circumstances, even though volatility is often the main focus. Let us start by defining the jump-diffusion model for the log-price X:

$$X_t=\int_0^t\sigma_s dW_s+Z_t,\quad t\geq 0 \tag{45}$$

where $(Z_t)_{t\geq 0}$ is a compound Poisson process. Looking at Equation (45), it would be natural to continue along the path of using quadratic variation in the GMM estimation. For jump-diffusion we have (see Section 4)

$$RV_t^n\xrightarrow{P}QV_T=IV_T+JV_T.$$

We then define

$$QV_t=IV_t+JV_t=\int_0^t\sigma_s^2 ds+\sum_{k=1}^{N_T^{(X)}}J_k^{(X)^2},$$

for $t\in\mathbb{Z}$.

The moment structure of $(QV_t)_{t\in\mathbb{Z}}$ is already calculated in Section 7. Our main focus now is on QV, but we assume that it cannot be observed directly, but in the form of a noisy proxy for QV, namely

$$\widehat{QV}_t=QV_t+\varepsilon_t,\quad t\in\mathbb{Z}$$

where ε_t captures the measurement noise. Regarding the stationarity and ergodicity requirements for QV_t, we know that IV_t was proven to fulfill these properties in Section 8.1. For $(JV_t)_{t\in\mathbb{Z}}$ we note that the jump variation process is constructed by differencing

$$JV_t = \sum_{k=1}^{N_t^{(X)}} J_k^{(X)^2} - \sum_{k=1}^{N_{t-1}^{(X)}} J_k^{(X)^2}, \quad t \in \mathbb{Z}.$$

We have already noted earlier that

$$\mathbb{E}\left[JV_t\right] = \lambda_J^{(X)^2}\left(\mu_J^{(X)^2} + \sigma_J^{(X)^2}\right)^2 = \mathbb{E}\left[JV_{t+\ell}\right], \quad \forall t \in \mathbb{Z},\ \forall \ell > 0.$$

Furthermore

$$Cov\left(JV_t, JV_{t+\ell}\right) = 0, \quad \forall t \in \mathbb{Z},\ \forall \ell > 0,$$

and

$$\begin{aligned} Var(JV_t) &= \lambda_J^{(X)}\left(\mu_J^{(X)^4} + 6\mu_J^{(X)^2}\sigma_J^{(X)^2} + 3\sigma_J^{(X)^4}\right) - \lambda_J^{(X)^2}\left(\mu_J^{(X)^2} + \sigma_J^{(X)^2}\right)^2 \\ &= Var(JV_{t+\ell}), \quad \forall t \in \mathbb{Z},\ \forall \ell > 0. \end{aligned}$$

Stationarity is fulfilled since mean, variance, and covariance are independent of t and are all the same expression for t and for $t+\ell$.

With respect to ergodicity of $(JV_t)_{t\in\mathbb{Z}}$, we note that both $\mathbb{E}\left[JV_t\right]$ Cov $\left(JV_t, JV_{t+\ell}\right)$ and Var $\left(JV_t\right)$ do not depend on t, or ℓ. This indicates that ergodicity in both mean and autocovariance can be proved. Ergodicity in mean is defined as $\frac{1}{T}\int_0^T JV_s ds$ converges, as $T \to \infty$ to $\mathbb{E}\left[JV_t\right]$ in squared mean, i.e.

$$\lim_{T\to\infty} E\left[\left(\frac{1}{2}\int_0^T JV_s ds - \mathbb{E}\left[JV_t\right]\right)^2\right] = 0.$$

The proofs are tedious and involves use of Fubini's Theorem.

Next, we present the correction factor when dealing with quadratic variance when jumps in X are present.

Example 2. *Realized Quadratic Variance*

Suppose that we want to estimate the quadratic variance QV using Realized variance

$$RV_t^n = \sum_{i=1}^{n}\left(X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}\right)^2.$$

We have

$$QV_t = IV_t + JV_t + \varepsilon_t, \quad t \in \mathbb{N}.$$

We have that

$$\sqrt{n}\left(RV_t^n - QV_t\right)$$

converges in law as $n \to \infty$, to a stochastic variable distributed as mixed normal with mean zero and a conditional asymptotic variance given by (Aït-Sahalia & Jacod 2014) (Veraart 2011)

$$V_t = 2IQ + 4\sum_{k=N_{t-1}^{(X)}+1}^{N_t^{(X)}} \sigma_k^2 J_k^{(X)}.$$

Then the proxy for QV_t is

$$\widehat{QV}_t = QV_t + \varepsilon_t$$

with

$$\varepsilon_t = \left(\frac{2}{n}IQ + \frac{4}{n}\sum_{k=N_{t-1}^{(X)}+1}^{N_t^{(X)}} \sigma_k^2 J_k^{(X)^2}\right)^{\frac{1}{2}}$$

and a correction function defined as

$$c(\theta) = \frac{2}{n}exp\left(\kappa_\phi(0)\right) + \frac{4}{n}\xi\lambda_J^{(X)}\left(\mu_J^{(X)^2} + \sigma_J^{(X)^2}\right). \tag{46}$$

We setup the GMM with $(QV_t)_{t\in\mathbb{Z}}$ exactly in the same way as for $(IV_t)_{t\in\mathbb{Z}}$. Regarding the consistency of the GMM estimator we will not go into details for now, but rather point at a potential problem with the proposed GMM procedure.

In the population moment conditions, the mean in the jump size distribution $\mu_J^{(X)}$ is only present in power 2 and 4. Since $\mu_J^{(X)}$ can take both positive and negative values in theory, it cannot be identified by only using even powers. We use QV, potential jumps emerge as large positive deviations. Most likely it will not have an effect on the jump intensity parameter $\lambda_J^{(X)}$ since all jumps still will be recorded. The long-run mean of σ^2 and ξ will probably not be influenced as well. Since all returns are squared, the sign of potential jumps vanishes and we suspect that the GMM approach, in fact, estimates the mean and variance in a folded normal distribution $|J_t|$. In general, if $J_t \sim N(\mu_J^{(X)}, {\sigma_J^{(X)}}^2)$ then

$$\mu_{|J_t|} = \sqrt{\frac{2}{\pi}}\sigma e^{\left(-\frac{{\mu_J^{(X)}}^2}{2{\sigma_J^{(X)}}^2})\right)} + \mu_J^{(X)}\left[1 - 2\Phi\left(-\frac{\mu_J^{(X)}}{{\sigma_J^{(X)}}^2}\right)\right]$$

and

$$\sigma^2_{|J_t|} = {\mu_J^{(X)}}^2 + {\sigma_J^{(X)}}^2 - \mu^2_{|J_t|}$$

where Φ is the cdf of a standard normal distribution.

For two different jump size distributions

1. $J_t^{(X)} \sim N(0, 0.01)$
2. $J_t^{(X)} \sim N(0, 1)$

We find that for case 1

$$\mu_{|J_t|} = 0.008$$
$$\sigma^2_{|J_t|} = 0.01$$

and for case 2

$$\mu_{|J_t|} = 0.8$$
$$\sigma^2_{|J_t|} = 0.36.$$

The change in mean and variance is almost none in case 1 (jumps are small), while in case 2 the bias is substantial.

9 Simulation

This section deals with the simulation and estimation of the fSV model defined in sections 2 and 7. We simulate different versions of the fSV model in order to see the evolving pattern of the fSV model for different values of the Hurst parameter and to check the robustness of the GMM estimation method for processes with known values of the parameters. The section is organized as follows. First, we will introduce the simulation approach both for the fBM, the fSV model specified in Bolko et al. (2023) and in a general jump setup. Second, we will replicate the GMM estimation of the parameters in the fSV model, using the method described in Bolko et al. (2023). Third, we test the GMM approach under different setups, which include sampling size and frequency, kernel specification, and moment structure. Lastly, section also includes subsections dedicated to the estimation of the fSV model when affected by microstructure noise, as well as the estimation of the fSV model when incorporating jumps. Note that Tables 7-12 are placed in the Appendix together with Figures 9-14.

9.1 Fractional Brownian Motion

In the first part of this section, we will define how a fBM can be simulated. As mentioned in Section 2.5, we note that the fBm satisfies a rescalling property so that the process $\left\{c^{-H}B_{ct}^H, t \geq 0\right\} \stackrel{D}{=} \left\{B_t^H, t \geq 0\right\}$ for all $c > 0$ (Kroese & Botev 2014). The fBm process is the only self-similar Gaussian process with stationary increments, which entails from the rescalling property.

The fractional Brownian motion (fBm) is simulated on the interval $[0, 1]$ using a uniformly spaced grid, defined as $0 = t_0 < t_1 < t_2 < \cdots < t_n = 1$. This grid is used to generate a discrete fractional Gaussian noise process $\{X_1, X_2, \ldots, X_n\}$ with $X_i = B_i^H - B_{i-1}^H$ and i is understood as t_i, see Subsection 2.3. According to Section 2.5, the fractional Gaussian noise has

$$Var[X_i] = 1, \quad i = 0, 1, \ldots, n$$

and the covariance function

$$\text{Cov}\left(X_i, X_{i+k}\right) = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right), \quad k = 0, 1, 2, \ldots \tag{47}$$

Then we get that the discrete fractional Brownian motion on $i = 0, 1, \ldots, n$ can be calculated as the cumulative sum

$$B_i^H = C^H \sum_{k=1}^{i} X_k, \quad \text{with } C = \frac{1}{n}.$$

Since, as mentioned in Subsection 2.5 and pointed out by Kroese & Botev (2014), the fractional Gaussian noise process has a constant expectation over time and the covariance function is invariant under translations (i.e. stationary in nature), a circulant embedding approach can be used.

The circulant embedding approach can be explained as follows. Consider a symmetric $(n+1) \times (n+1)$ covariance matrix Ω of Toeplitz form with elements

$$\Omega_{i+1,j+1} = \text{cov}\left(X_i, X_j\right), i, j = 0, \ldots, n.$$

Here Ω is not to be confused with the concept of sample space. We note that Ω is completely defined by the first row which we will denote $(r_1, \ldots, r_{n+1})$. Let Σ denote a non-negative definite and symmetric $2n \times 2n$ circulant matrix that embeds Ω in the upper left $(n+1) \times (n+1)$ corner. Then the first row of Σ is

$$r = \left(r_1, \ldots, r_{n+1}, r_n, r_{n-1}, \ldots, r_2\right).$$

Next, let λ denote the one-dimensional fast Fourier transform of r, and it is defined as the linear transformation $\lambda = Fr$, where F is a $2n \times 2n$ matrix with

$$F_{j,k} = \exp\left(\frac{-i \cdot 2\pi \cdot j \cdot k}{2n}\right)/\sqrt{2n}, \quad j, k = 0, 1, \ldots, 2n-1.$$

Dividing by $2n$ creates the eigenvalues. Finally, calculate

$$X = F^* \cdot diag(\sqrt{\lambda}) \cdot Z$$

with Z being a $2n$ complex-valued Gaussian vector, and F^* being the complex conjugate transpose of F. The $2n \times 2n$ matrix $diag(\sqrt{\lambda})$ is a diagonal matrix with $\sqrt{\lambda}$ in the diagonals. The vector Z is constructed by $2n$ random draws of a standard normal distribution denoted N and we define Z as

$$Z = N + i \cdot N.$$

Then the real and imaginary parts of the first $n+1$ components of X yield two independent realizations of the fractional Brownian noise. Only the real valued part will be used. Then

$$B_i^H = n^{-H} \sum_{k=1}^{i} real(X_k).$$

This simulation method and a simple implementation using `MATLAB` can be found in Kroese & Botev (2014), where it is described in more detail. In this thesis, we will give a simple algorithm (Algorithm 1 in the Appendix) that can be implemented in whatever software the reader sees fit. Three realizations of the fBm with different Hurst values is shown in Figure 1.

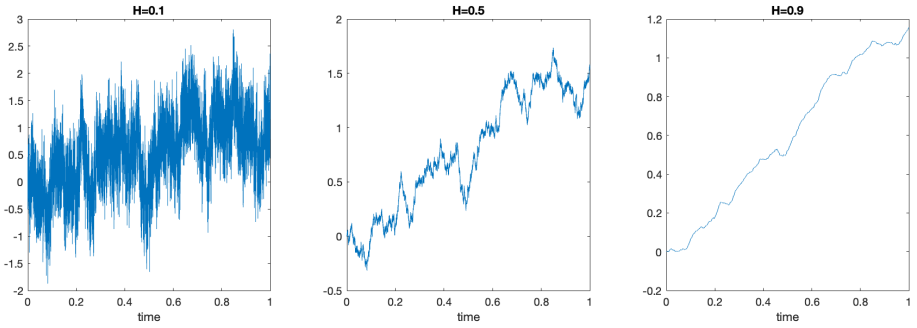


Figure 1: Three examples of paths for a Fractional Brownian motion with different Hurst parameters, showing the different correlations of the increments. From left to right $H = 0.1, 0.5, 0.9$.

9.2 Simulation of the fSV model as defined in Section 2 and in Bolko et al. (2023)

In this subsection assumed that the log-price, $(X_t)_{t\geq 0}$, evolves as a driftless Itô process (see Section 2.4) defined as

$$\mathrm{d}X_t = \sigma_t \, \mathrm{d}W_t, \quad t \geq 0. \tag{48}$$

The initial condition $X_0 \equiv 0$, W_t is a standard Brownian motion, and σ_t is the spot volatility. For the simulation, we will discretize X according to a Euler scheme (see Subsection 2.3). Since we will mainly be interested in volatility, it is not restrictive to assume a driftless process. The log-variance process defined as

$$Y_t = log\left(\sigma_t^2\right), \quad t \geq 0$$

is assumed to evolve as a fOU process

$$\mathrm{d}Y_t = -\lambda\left(Y_t - \eta\right)\mathrm{d}t + v \, \mathrm{d}B_t^H, \quad t \geq 0 \tag{49}$$

where B_t^H is a fBm and $\eta = \log(\xi) - 0.5 \cdot \mathrm{Var}(Y_t)$. It is assumed that we have no leverage effect, which implies $W \perp B^H$, i.e. the two processes are independent.

Solving the SDE in (49) we get the following expression for Y.

$$Y_t = \eta + \left(Y_{t-\Delta} - \eta\right) e^{-\lambda\Delta} + v \int_{t-\Delta}^{t} e^{-\lambda(t-s)} \mathrm{d}B_s^H. \tag{50}$$

This implies that we need to be able to approximate the stochastic integral

$$\int_{t-\Delta}^{t} e^{-\lambda(t-s)} \mathrm{d}B_s^H.$$

An approximation is given by

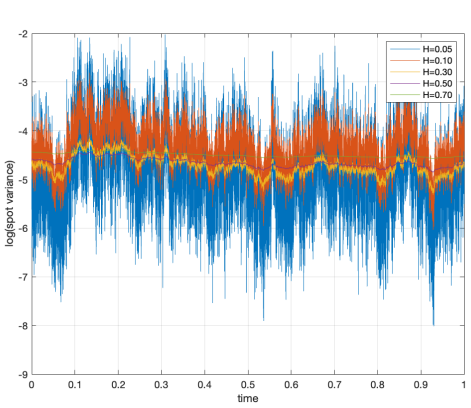
$$e^{-\lambda\Delta/2}\left(B_t^H - B_{t-\Delta}^H\right)$$

requiring discretely sampled fBm increments (Subsection 9.1, Algorithm 1).

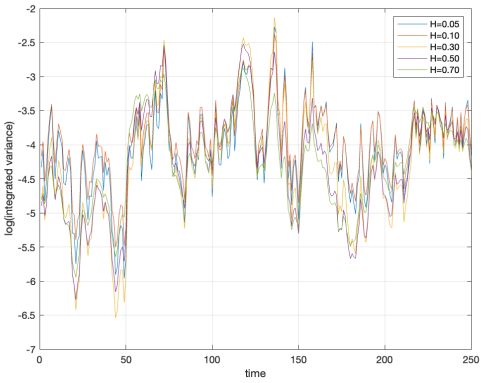
For the simulation of the fSV model, we draw 250 independent replications of the model with path length $T = 1000$. The length of the path and the number of independent replications are lower than those of Bolko et al. (2023). The reason for this is time constraints, as each full replication takes around 5 minutes on the machine used.

In each independent simulation, the log-variance process Y is randomly initialized from the stationary distribution $Y_0 \sim N\left(\eta, \mathrm{var}\left(Y_t\right)\right)$. The variance of Y_t is κ_0 which is given in (17). In order to achieve an almost continuous time simulation of the processes Y, σ^2 and X, as well as minimizing the dicretization bias, we partition $[t-1, t]$, for $t = 1, \ldots, T$, into $N = 23400$ discrete subintervals of length $\Delta = 1/N$ as in Bolko et al. (2023). In total, a simulated path is equal to a 4-year sample of a stock price recorded every second on the US equity market.

We set the model parameters in coherence with Bolko et al. (2023), which is



(a) Simulated path of log(spot variance) for one day.



(b) Simulated path of log(Integrated variance) over 250 days.

Figure 2: Example trajectory of spot variance and integrated variance.

$$\xi = \mathbb{E}[\sigma_t^2] = 0.0225$$
$$\lambda = [0.005, 0.01, 0.015, 0.035, 0.07]$$
$$v = [1.25, 0.75, 0.5, 0.3, 0.2]$$
$$H = [0.05, 0.1, 0.3, 0.5, 0.7].$$

These values are combined in 5 different parameter sets (θ). The constant value of ξ ensures that the unconditional mean of the variance process is identical for all 5 model settings. The 5 different values of the Hurst exponent H cover the rough, "standard" and long-memory fBm processes.

In Figure 2 we present simulated sample paths of the spot and integrated variance for the 5 parameter sets. The paths are presented for one day for spot variance and for 250 days for integrated variance. Figure 2 shows that our simulations exhibits the same pathwise properties of volatility as in Bolko et al. (2023). On the microscopic scale (spot variance during one day) the paths are very different with respect to the overall variation, but when integrated up to the daily horizon the paths are much harder to differentiate.

9.3 Simulation of the fSV model in the general setup

In this subsection we explain how the fractional OU model with jumps (fOUJ) is simulated. We consider five ways to incorporate jumps:

1. In the log-price process X.
2. In the log-variance process Y.
3. In both X and Y defined as co-jumps.
4. In both X and Y as independent jumps.
5. In both independent jumps and co-jumps.

Note that the parameter vector is now

$$\theta = \left(\xi, \lambda, v, H, \sigma_J^{(X)}, \lambda_J^{(X)}, \mu_J^{(X)}, \sigma_J^{(Y)}, \lambda_J^{(Y)}, \mu_J^{(Y)}, \lambda_J^{(XY)}, \rho_J, \mu_J^{(XY)}, \mu_J^{(YX)}, \sigma_J^{(XY)}, \sigma_J^{(YX)}\right).$$

Each different jump specification can be set by either setting some jump parameters to zero or giving them a value not equal to zero. The simulation of the fSV model in Bolko et al. (2023) is then simulated by setting $\lambda_J^{(X)}$, $\lambda_J^{(Y)}$ and $\lambda_J^{(XY)}$ to zero. In this case, it is easy to specify different jump models when the jump process is a compound Poisson process. The method is outlined in Algorithm 2 in the Appendix.

9.4 GMM estimation of replication study

For the GMM estimation procedure, we gather realized variance with 5-minute data, that is, $n = 78$. The input data to the optimizer can be either the values of $(IV_t)_{t=1}^T$ or $(RV_t^n)_{t=1}^T$. In this subsection, we will focus on the implementation of GMM for the integrated variance series. The implementation with realized variance follows exactly the same steps. For a description with $(RV_t^n)_{t=1}^T$ we refer the reader to Bolko et al. (2023). In Algorithm 3 in the Appendix we present the computational algorithm used to estimate the initial values of the 4 parameters in the fSV model of Bolko et al. (2023). The initial values are used as starting values in the GMM estimation procedure, which is described in Algorithm 4 in the Appendix.

Theoretical moments are calculated from equations (21) and (19) for $r = 2$ using numerical integration and lag values set to $\ell = [0, 1, 2, 3, 5, 20, 50]$. In order to do this, the `integrate` function in MATLAB is used twice (one inside the other). The theoretical moments are then subtracted from the overall mean of the sample moments. This difference is multiplied by a weighting matrix W_T, and the result is a vector of length $\ell + 1$. This vector is minimized with respect to the parameters using the `lsqnonlin` function, which is a nonlinear least-squares solver that we set to use the 'trust-region-reflective' algorithm. The algorithm requires a vector valued function with at least as many elements as the parameters to optimize. Importantly, we do not specify the objective function as a scalar but as a vector as this is required by the optimizer. We therefore use the form

$$\widehat{\theta}_T \arg\min_{\theta} \mathbb{W}_T \widehat{m}_T(\theta).$$

We set lower and upper bounds on the parameters, in particular, the bounds for $H = [0.001, 0.999]$.

We iterate the GMM procedure three times. The first time the weighting matrix $\mathbb{W}_T$ is set as the identity matrix. In the second and third iterations, the weighting matrix is calculated using equation (36) with a Parzen kernel and automatic lag selection based on the method presented in Andrews (1991), in which we use an approximate ARMA(1,1) structure for the volatility measure. For speed, we do a Cholesky decomposition on $\mathbb{W}_T$ in the second and third iteration. This is possible according to Assumption 5.

9.4.1 Results of replication study

The results of the replication of Bolko et al. (2023) are shown in Table 8. We start by commenting on the initial values and the estimation results for the integrated variance. For the initial values of ξ, we see that the initial values come close to the true values, whereas for λ and v, they largely deviates from the true value. For H, initial values have an upward bias when $H < 0.5$. In the estimation results, we see a clear underestimation of ξ, an overestimation of λ, while both v and H are very close to the true values. ξ and λ are both drift parameters (being the average level of the spot volatility process and the speed of mean reversion, respectively) and affect the acf of the integrated variance in opposite directions. We note that 250 replications are a relatively small replication size, and we will expect that result may vary from what would be obtained with a higher number of replications. We note that we see the same patterns in our result as those of Bolko et al. (2023). All estimates remain within one Monte Carlo standard deviation from the true value.

Looking at the initial values and the estimation results of the realized variance, we observe that the initials have patterns similar to the initials of integrated variance but notably a much worse value for the case of a large H. The uncorrected estimates perform poorly for v and H and some estimates are more than one standard error away from the true values. H is underestimated quite heavily and is closer to half the true value. Being a noisy proxy for integrated variance, realized variance induces illusive roughness. Correcting for the measurement error, we get much better estimates of H and only for $H = 0.7$ the estimates are off by over one standard error.

One thing to note is that Bolko et al. (2023) has a downward bias in their H estimates. This is not the case for our replication study.

9.5 GMM estimation with fewer days observed

As the first robustness check, we run the estimation procedure, but only with $T = 500$. This provides an intuition for whether the model works for assets where a lower number of days is

available. The results are presented in Table 9.

Comparing the results of Table 9 with Table 8, we first notice that the initial values for both the integrated variance and the realized variance have larger standard deviations.

Next, we observe noticeably worse estimates and higher standard deviations of v and H in rough ($H < 0.5$) cases for the integrated variance process.

For the corrected estimates, H is farther away from the true parameter values, except when $H = 0.3$. This leads us to the conclusion that the number of days observed has an impact on the reliability of the estimate of H. We see that the v estimates are further away from the true values except for $v = 0.5$, while all estimates of λ are further away from the true value compared to Table 8. Therefore, we also believe that the estimates for the replication study in Table 8 would improve if $T > 1000$. This is in line with Bolko et al. (2023) where $T = 4000$ and their estimates are closer to the true parameter values.

9.6 Kernel specification in the weight function

We now turn our attention to the kernel specification used by the weight function in the HAC estimator. The results are presented in Table 10.

The same sample for IV and RV as in the replication study is used for this robustness check, resulting in the initial values being equal to each other across the two tables.

Compared to Table 8, we do not see a significant variation in the estimates. This is in line with what we would expect. Both kernels satisfies our assumption for the weight function, so asymptotically it should not matter which kernel we choose. However, we do see a slight improvement in the estimates of H for IV but for RV it is inconclusive. Comparing IV estimates of v we see that the Barlett kernel has a downward bias, whereas the Parzen kernel generally has an upward bias. Regarding IV estimates of ξ the Parzen kernel clearly underestimates more than the Barlett kernel. We also see this to be the case for the RV estimates of ξ. We see a slight improvement in the estimates of λ for RV using the Barlett kernel.

We conclude that in first glance the slight variation between the two kernels is most likely due to the amount of replications being small. We will provide evidence that the Barlett kernel has better distributional attributes later in the section.

9.7 Lower number of moments

We now turn our attention to the number of moments used in the estimation procedure. A possible way to improve the computation time of the estimation (at least for calculating the moments) is to reduce the number of moments. The selected lags are now $\ell = (1, 2, 5, 10)$. The results are presented in Table 11.

For integrated variance, the estimates for H are only slightly worse. The standard deviations are larger, which is expected since we now include less information in the estimation procedure. This in turn makes the estimates less precise. We note that estimates of ξ are closer to the real value than for the replication of Bolko et al. (2023). Estimates of v and λ are determined by their ratio, so although estimates for v are closer to the true value, estimates for λ are further away, making the result inconclusive.

Overall, the estimates are worse with a lower number of moments, in line with our expectation.

9.8 The sampling frequency of RV

We now focus on the sampling frequency of our log price differences in the computation of our estimator. Previously, we computed the realized variance using data sampled at 5-minute intervals. In this study, we increase the sampling frequency to 1-minute intervals. The findings of this robustness check are detailed in Table 12. We expect the results to be more accurate because of lower discretization bias.

First, it is important to note that variations in the IV parameter estimates relative to Table 8 arise solely from the use of another set of replications.

We observe that we are closer to the true value with the RV estimates of H when $H \geq 0.5$ compared to the replication study. For the very rough cases, the estimates are slightly farther away

from the true value. We also observe smaller standard deviations for the H estimates, especially for higher H. We observe better estimates for ξ in both the uncorrected case and the corrected cases. We believe that this is due to higher means of RV pushing ξ upwards, and not due to the higher sampling frequency. We notice that the uncorrected estimates for v and H are now much closer to the true value, leading us to conclude that some of the measurement noise comes from the discretization bias inherent in the sampling of RV. We also conclude that increasing the sampling rate is beneficial when dealing with observations that do not exhibit microstructure noise. Next, we will consider the case where microstructure noise is incorporated.

9.9 Fractional stochastic volatility model with microstructure noise

In this subsection we specify how noise enters the price process and how we go about eliminating this noise for our estimators.

The simulation steps for Y are the same as in the previous subsection. Now, we do not observe the efficient price process, but a noisy observation that has the following form

$$Z_t = X_t + \varepsilon_t, \quad \varepsilon_t = \alpha_t \chi_t, \quad t \in \mathbb{N}. \tag{51}$$

Where the noise has the following form

$$\chi_t = \log\left(\beta\left[e^{X_t + \chi_t'}/\beta\right]\right) - X_t.$$

We set $\alpha_t = 1,\ \forall t \in \mathbb{Z}$, so we are in a non-shrinking noise case. The additive noise $\chi_t' \sim i.i.d.\ N(0, 0.0005^2)$. This is on par with the additive noise component in the simulations of Jacod et al. (2009). The rounding parameter β is set to be a rounding to the nearest cent. We choose $X_0 = 5.225$, which is roughly equal to the (log) price of the most traded asset in the DJIA index (namely AAPL).

We again draw 250 simulations each with 1000 days and 23400 observations per day. Each parameter used by Y is set to the same values as in the previous subsection, since the noise does not have any influence on the log-variance process. In this case plots of the (log) spot variance and (log) integrated variance have the same characteristics (and the exact same path in the case where seeds are set in the right order). The observed price process now looks much different and resembles that of empirical data much more closely. In Figure 9, we see that the observed price process exhibits rounding errors. The additive noise is not as visible to the naked eye as the rounding noise. Figure 10 shows the time series of the price of INTC during trading hours on our last observed day in our dataset.

In addition to IV, we also collect pre-averaged RV which is specified in the following way and in a fashion similar to Christensen et al. (2014).

Let the returns on the observed price series pre-averaged in a local neighborhood of K observations be of the form

$$r_{i,K}^{preav} = \sum_{j=1}^{K-1} g\left(\frac{j}{K}\right) r_{i+j}^*$$

where $r_i^* = Z_{i/N} - Z_{(i-1)/N}$, and $g : [0,1] \to \mathbb{R}$ represent a weighting function. This function is continuous and has sections that are continuously differentiable. Its derivative, g', is piecewise Lipschitz. Furthermore, g satisfies $g(0) = g(1) = 0$ and the integral $\int_0^1 g^2(u)du$ is greater than zero. We define the following

$$\psi_1 = \int_0^1 \left(g'(s)\right)^2 \, \mathrm{d}s, \quad \psi_2 = \int_0^1 g^2(s)\mathrm{d}s.$$

Noise-robust pre-averaged RV is then defined as

$$RV^{preav} = \frac{N}{N-K+2}\frac{1}{K\psi_2^K}\sum_{i=0}^{N-K+1}\left|r_{i,K}^{preav}\right|^2 - \frac{\widehat{\omega}^2}{\theta^2}\frac{\psi_1^K}{\psi_2^K}.$$

Here we specify Riemann approximations to ψ_1 and ψ_2 as ψ_1^K and ψ_2^K

$$\psi_1^K = K\sum_{j=1}^{K}\left[g\left(\frac{j}{K}\right) - g\left(\frac{j-1}{K}\right)\right]^2, \quad \psi_2^K = \frac{1}{K}\sum_{j=1}^{K-1} g^2\left(\frac{j}{K}\right).$$

We simplify the setup by assuming that K is even and we use the weight function

$$g(x) = \min(x, 1-x).$$

The pre-averaged returns is

$$r_{i,K}^{preav} = \frac{1}{K}\left(\sum_{j=K/2}^{K-1} Z_{(i+j)/N} - \sum_{j=0}^{K/2-1} Z_{(i+j)/N}\right).$$

Noting that $\psi_1^K = 1$ and $\psi_2^K = \psi_K = \left(1 + 2K^{-2}\right)/12$, we have the noise robust version of RV

$$RV^{preav} = \frac{N}{N-K+2}\frac{1}{K\psi_K}\sum_{i=0}^{N-K+1}\left|r_{i,K}^{preav}\right|^2 - \frac{\widehat{\omega}^2}{\theta^2\psi_K}.$$

Here, the term $\widehat{\omega}^2/\theta^2\psi_K$ is a bias-correction term added to account for residual microstructure noise still present after pre-averaging. In our case, we use an estimator proposed by Oomen (2006), which has the form

$$\hat{\omega}^2 = -\frac{1}{N-1}\sum_{i=2}^{N} r_i^* r_{i-1}^*.$$

We choose $\theta = 1$ for simplicity. K is assumed to be $K = \theta\sqrt{N} + \mathrm{o}\left(N^{-1/4}\right)$ meaning that the pre-averaging period expands with an increase in sampling frequency, albeit at a diminished pace.

Lastly, the initial values of the parameter set are determined by the same procedure as in the subsection above.

9.9.1 Practical concerns in calculating initial values

A drawback of using the pre-averaging approach is that RV^{preav} may take negative values. This is also the case for other estimators dealing with microstructure noise (Two-Scales Realized Volatility, Flat Top Realized Kernels and Multi-scales Estimators) (Aït-Sahalia & Jacod 2014). In order to calculate initial values for λ, v and H, we need $RV_t^{preav} > 0$, $\forall t$. This is done by removing the estimates that are ≤ 0 when calculating the initial values.

9.9.2 Comment on the results

We begin by examining the estimates for the integrated variance. We expect the results to look almost the same as the previous IV estimates of the replication study, as nothing has changed in the simulation and estimation (although we now have 250 new replications). Indeed, we see similar results to the model before with one exception, which is that the standard deviations on v is smaller by almost one half in Panels A and B. This is due to the replication amount being small.

For the parameter values for the realized measures, we see that pre-averaged and 5-minute realized variance have similar performance. All fail to estimate H correctly, although the estimates are within one standard deviation except for $H = 0.7$. We generally see lower standard deviations for H when using pre-averaged RV (except for $H = 0.05$). We also see that for the corrected estimates of H for 5-minute RV is closer to those of IV. For lower H, pre-averaged RV is closer to the true values. However, this might be a coincidence because of the low amount of replications.

We conclude that both pre-averaged RV and 5-minute RV are equally good at estimating the fSV model in the presence of microstructure noise. One has to be careful interpreting the estimates as they are now further away from the true parameter values. This is especially true when data are very rough or exhibit long memory in the volatility.

Parameter	Value	Integrated variance		Pre-averaged Realized variance		5-minute Realized variance			
							Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:									
ξ	0.0225	0.0258 (0.0147)	0.0208 (0.0074)	0.0341 (0.0409)	0.0286 (0.0363)	0.0364 (0.0591)	0.0266 (0.0354)	0.0278 (0.0372)	0.0286 (0.0412)
λ	0.0050	0.0286 (0.0144)	0.0167 (0.0161)	0.0248 (0.0152)	0.0170 (0.0182)	0.0212 (0.0137)	0.0157 (0.0179)	0.0174 (0.0201)	0.0175 (0.0194)
v	1.2500	0.4582 (0.0118)	1.1112 (0.3548)	0.5040 (0.1124)	1.3334 (0.4828)	0.5323 (0.0426)	1.7293 (0.4882)	1.2151 (0.4698)	1.0546 (0.3665)
H	0.0500	0.2675 (0.0198)	0.0816 (0.0462)	0.2409 (0.0436)	0.0716 (0.0667)	0.2147 (0.0273)	0.0430 (0.0366)	0.0787 (0.0632)	0.0870 (0.0440)
Panel B:									
ξ	0.0225	0.0225 (0.0041)	0.0198 (0.0031)	0.0296 (0.0316)	0.0253 (0.0274)	0.0292 (0.0327)	0.0232 (0.0197)	0.0240 (0.0203)	0.0242 (0.0209)
λ	0.0100	0.0366 (0.0126)	0.0239 (0.0178)	0.0306 (0.0147)	0.0209 (0.0212)	0.0250 (0.0129)	0.0169 (0.0183)	0.0242 (0.0214)	0.0242 (0.0206)
v	0.7500	0.0366 (0.0094)	0.7011 (0.1472)	0.4159 (0.1155)	0.9492 (0.3967)	0.4510 (0.0502)	1.3654 (0.4916)	0.7942 (0.3729)	0.7764 (0.4025)
H	0.1000	0.3008 (0.0204)	0.1299 (0.0535)	0.2664 (0.0502)	0.0929 (0.0500)	0.2284 (0.0312)	0.0551 (0.0526)	0.1263 (0.0780)	0.1276 (0.0591)
Panel C:									
ξ	0.0225	0.0223 (0.0069)	0.0160 (0.0051)	0.0283 (0.0249)	0.0209 (0.0215)	0.0297 (0.0374)	0.0195 (0.0180)	0.0200 (0.0181)	0.1276 (0.0188)
λ	0.0150	0.0335 (0.0094)	0.0314 (0.0199)	0.0276 (0.0117)	0.0251 (0.0212)	0.0218 (0.0095)	0.0164 (0.0197)	0.0331 (0.0274)	0.0328 (0.0272)
v	0.5000	0.3631 (0.0091)	0.4968 (0.1000)	0.4139 (0.1313)	0.6429 (0.3924)	0.4393 (0.0488)	0.7962 (0.3965)	0.6003 (0.3869)	0.6060 (0.4024)
H	0.3000	0.4386 (0.0215)	0.3073 (0.0700)	0.3909 (0.0761)	0.2534 (0.0960)	0.3475 (0.0439)	0.1799 (0.0950)	0.2877 (0.1128)	0.2860 (0.1106)
Panel D:									
ξ	0.0225	0.0217 (0.0066)	0.0163 (0.0051)	0.0269 (0.0226)	0.0198 (0.0193)	0.0276 (0.0302)	0.0178 (0.0170)	0.0189 (0.0176)	0.0192 (0.0190)
λ	0.0350	0.0447 (0.0088)	0.0474 (0.0183)	0.0343 (0.0123)	0.0349 (0.0190)	0.0235 (0.0088)	0.0222 (0.0202)	0.0483 (0.0254)	0.0486 (0.0255)
v	0.3000	0.2489 (0.0063)	0.2920 (0.0378)	0.3047 (0.1226)	0.4291 (0.4148)	0.3456 (0.0532)	0.5180 (0.4002)	0.4303 (0.4253)	0.4281 (0.4182)
H	0.5000	0.5717 (0.0213)	0.4978 (0.0663)	0.4905 (0.0961)	0.4079 (0.1172)	0.4065 (0.0560)	0.2937 (0.1211)	0.4373 (0.1282)	0.4392 (0.1293)
Panel E:									
ξ	0.0225	0.0214 (0.0080)	0.0180 (0.0070)	0.0260 (0.0214)	0.0203 (0.0165)	0.0263 (0.0272)	0.0181 (0.0174)	0.0190 (0.0175)	0.0195 (0.0193)
λ	0.0700	0.0562 (0.0092)	0.0616 (0.0177)	0.0378 (0.0133)	0.0410 (0.0186)	0.0212 (0.0081)	0.0236 (0.0183)	0.0574 (0.0293)	0.0573 (0.0293)
v	0.2000	0.1696 (0.0047)	0.1877 (0.0193)	0.2339 (0.1187)	0.2989 (0.3806)	0.2899 (0.0548)	0.3626 (0.2881)	0.2945 (0.3295)	0.3186 (0.3774)
H	0.7000	0.6776 (0.0196)	0.6442 (0.0561)	0.5407 (0.1103)	0.5150 (0.1206)	0.4055 (0.0592)	0.3558 (0.1155)	0.5306 (0.1354)	0.5233 (0.1455)

Table 1: Initial values and estimation results of the fSV model with microstructure noise.

9.10 Jump model

We now perform a robustness check on the GMM approach in the presence of jumps.

9.10.1 Truncation approach

If the jump parameters are not of interest, the fSV model can be estimated using a robust jump estimator. In our case, we have chosen the truncated variance as it does not depend on the volatility having a semimartingale structure, which the fSV model does not have for $H \neq 0.5$. We have already shown in Example 1 that the correction factor is the same as in Example 3.1 of Bolko et al. (2023). The results are presented in Table 2. At first glance, we notice an illusive roughness in the long memory case ($H = 0.7$). For other values of H, truncated variance performs quite well. We notice a slight underestimation of v in the very rough case ($H = 0.05$). We conclude that truncated variance can be used to estimate H except when dealing with a long-memory process.

As seen, truncated variance does a good job of estimating the 4 parameters in the presence of jumps in X.

9.10.2 When using Realized variance on jump model

In Table 3 we estimate the fSV model using the GMM approach of Bolko et al. (2023) on both a jump in X model and a jump in Y model using realized variance.

Starting with the jump in X model, we see that estimates of H are now largely underestimated, especially for larger H. This in turn causes the model to seem rougher than it actually is. We see that estimates of v are now largely overestimated. ξ is seemingly unaffected and still underestimates compared to the mean volatility, which is now a bit higher due to the extra variation price jumps cause. λ is also far away from its true value.

For the jumps in Y we see another picture. For $H \leq 0.5$, the estimates are largely overestimated and do not come close in the very rough cases. We now see ξ being significantly affected by volatility jumps. This makes intuitive sense, as the jumps now directly effect volatility. We also see that the estimates of v and λ are inaccurate.

9.10.3 Initial values

Initial estimates for $\lambda_J^{(X)}$ and $\sigma_J^{2(X)}$ are calculated as follows:

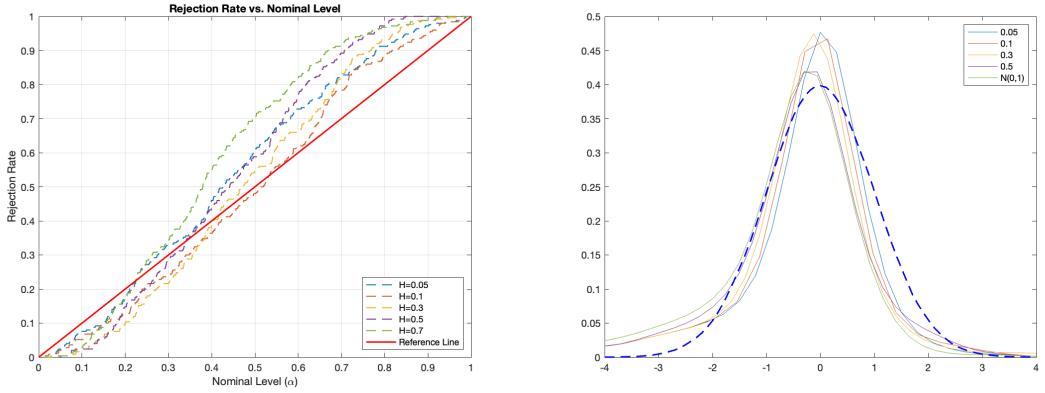
$$\lambda_J^{(X)}(initial) = \frac{\sum_{t=1}^{T}(RV_t - RV_t^T)}{T}$$

Parameter	Value	Integrated variance		Truncated variance		
		Initial value	Estimate	Initial value	Uncorrected estimate	Corrected estimate Approximate
Panel A:						
ξ	0.0225	0.0258 (0.0148)	0.0209 (0.0077)	0.0253 (0.0145)	0.0203 (0.0073)	0.0210 (0.0077)
λ	0.0050	0.0306 (0.0155)	0.0196 (0.0190)	0.0242 (0.0141)	0.0166 (0.0163)	0.0177 (0.0181)
v	1.2500	0.4560 (0.0120)	1.0828 (0.3405)	0.5258 (0.0151)	1.6962 (0.4616)	1.0140 (0.1982)
H	0.0500	0.2676 (0.0192)	0.0857 (0.0513)	0.2193 (0.0190)	0.0418 (0.0333)	0.0848 (0.0429)
Panel B:						
ξ	0.0225	0.0231 (0.0045)	0.0202 (0.0033)	0.0228 (0.0044)	0.0196 (0.0031)	0.0202 (0.0033)
λ	0.0100	0.0363 (0.0132)	0.0224 (0.0170)	0.0267 (0.0114)	0.0146 (0.0135)	0.0220 (0.0182)
v	0.7500	0.3673 (0.0099)	0.7047 (0.1516)	0.4421 (0.0125)	1.3335 (0.4289)	0.7200 (0.1647)
H	0.1000	0.3008 (0.0186)	0.1257 (0.0484)	0.2339 (0.0183)	0.0499 (0.0359)	0.1227 (0.0512)
Panel C:						
ξ	0.0225	0.0224 (0.0077)	0.0160 (0.0049)	0.0221 (0.0076)	0.0156 (0.0047)	0.0162 (0.0050)
λ	0.0150	0.0332 (0.0102)	0.0298 (0.0209)	0.0233 (0.0088)	0.0143 (0.0158)	0.0278 (0.0228)
v	0.5000	0.3617 (0.0088)	0.4813 (0.0713)	0.4313 (0.0115)	0.7160 (0.1638)	0.5112 (0.0853)
H	0.3000	0.4360 (0.0209)	0.3057 (0.0731)	0.3564 (0.0216)	0.1764 (0.0739)	0.2823 (0.0856)
Panel D:						
ξ	0.0225	0.0227 (0.0072)	0.0168 (0.0048)	0.0224 (0.0071)	0.0151 (0.0043)	0.0163 (0.0048)
λ	0.0350	0.0441 (0.0090)	0.0483 (0.0190)	0.0250 (0.0070)	0.0223 (0.0212)	0.0460 (0.0262)
v	0.3000	0.2485 (0.0061)	0.2869 (0.0312)	0.3347 (0.0085)	0.4385 (0.1145)	0.3303 (0.0622)
H	0.5000	0.5709 (0.0210)	0.5029 (0.0624)	0.4214 (0.0221)	0.2977 (0.0993)	0.4455 (0.0974)
Panel E:						
ξ	0.0225	0.0223 (0.0084)	0.0187 (0.0070)	0.0220 (0.0083)	0.0158 (0.0059)	0.0170 (0.0063)
λ	0.0700	0.0566 (0.0101)	0.0620 (0.0187)	0.0231 (0.0069)	0.0278 (0.0231)	0.0618 (0.0279)
v	0.2000	0.1691 (0.0047)	0.1867 (0.0199)	0.2788 (0.0140)	0.3124 (0.0594)	0.2360 (0.0353)
H	0.7000	0.6761 (0.0187)	0.6415 (0.0514)	0.4184 (0.0275)	0.3699 (0.1087)	0.5523 (0.0991)

Table 2: Initial values and estimation results of the jump model using truncated variance.

Parameter	Value	Realized variance: Jumps in X		Realized variance: Jumps in Y	
		Initial value	Estimate	Initial value	Estimate
Panel A:					
ξ	0.0225	0.0243 (0.0042)	0.0226 (0.0031)	3.5196 (13.4997)	0.5122 (0.3293)
λ	0.0050	0.0176 (0.0073)	0.0075 (0.0085)	0.0044 (0.0022)	0.0092 (0.0206)
v	1.2500	0.5785 (0.0181)	1.3550 (0.1714)	0.4938 (0.0100)	0.6921 (0.2473)
H	0.0500	0.1846 (0.0144)	0.0433 (0.0176)	0.3343 (0.0147)	0.2095 (0.1152)
Panel B:					
ξ	0.0225	0.0233 (0.0018)	0.0221 (0.0015)	0.2976 (0.4645)	0.1229 (0.0779)
λ	0.0100	0.0207 (0.0062)	0.0078 (0.0055)	0.0109 (0.0036)	0.0190 (0.0269)
v	0.7500	0.4928 (0.0123)	1.1662 (0.1902)	0.4117 (0.0110)	0.4784 (0.1904)
H	0.1000	0.1934 (0.0128)	0.0463 (0.0237)	0.3786 (0.0185)	0.3190 (0.1558)
Panel C:					
ξ	0.0225	0.0232 (0.0039)	0.0191 (0.0030)	0.1413 (0.1788)	0.0593 (0.0330)
λ	0.0150	0.0122 (0.0043)	0.0096 (0.0082)	0.0239 (0.0051)	0.0489 (0.0377)
v	0.5000	0.5466 (0.0264)	0.7361 (0.2158)	0.4073 (0.0101)	0.4615 (0.2109)
H	0.3000	0.2589 (0.0222)	0.1688 (0.0739)	0.4697 (0.0144)	0.4483 (0.1762)
Panel D:					
ξ	0.0225	0.0228 (0.0045)	0.0186 (0.0039)	0.0481 (0.0185)	0.0282 (0.0089)
λ	0.0350	0.0100 (0.0036)	0.0136 (0.0126)	0.0421 (0.0054)	0.0824 (0.0295)
v	0.3000	0.4717 (0.0295)	0.6228 (0.3579)	0.3109 (0.0124)	0.3700 (0.1378)
H	0.5000	0.2730 (0.0246)	0.2275 (0.1322)	0.5647 (0.0133)	0.5707 (0.1291)
Panel E:					
ξ	0.0225	0.0230 (0.0063)	0.0192 (0.0057)	0.0328 (0.0118)	0.0219 (0.0069)
λ	0.0700	0.0078 (0.0032)	0.0151 (0.0156)	0.0526 (0.0062)	0.0826 (0.0295)
v	0.2000	0.4050 (0.0360)	0.5846 (0.4629)	0.2513 (0.0148)	0.3171 (0.1635)
H	0.7000	0.2641 (0.0327)	0.2561 (0.1710)	0.6081 (0.0155)	0.5839 (0.1393)

Table 3: Initial values and estimation applying GMM approach of Bolko et al. (2023). For jumps in X the parameters set as $\lambda_J^{(X)} = 0.08$, $\mu_J^{(X)} = 0$ and $\sigma_J^{(X)} = 0.10$. For the jumps in Y the parameters are set as: $\lambda_J^{(Y)} = 0.025$, $\mu_J^{(Y)} = 0$ and $\sigma_J^{(Y)} = 1.25$.



(a) **(b)**

Figure 3: (a) Distribution of the J-test of overidentifying restrictions. P-P plots are created to highlight the J-test variation for the GMM jump in X model corrected. (b) Kernel density plot of standardized H for jump in X model.

$$\sigma_J^{(X)^2}(initial) = \frac{\bar{RV} - \xi(initial)}{\lambda_J^{(X)}(initial)}.$$

In the simulations the initial value of $\mu_J^{(X)}$ was always set to zero, but an alternative value could be

$$\mu_J^{(X)}(initial) = \frac{\sum_{t=1}^{T}(X_t - X_{t-1})}{T} - \frac{\xi(initial)}{\lambda_J^{(X)}(initial)}$$

9.10.4 Results

We now simulate 100 replications of the time length $T = 2500$ of the fSV plus jump in X. Table 4 reports the results.

Looking at 7 we see that we can effectively estimate using quadratic variation as an estimator by applying the modified GMM approach. Looking at H we see just as good results as in the replication study and when using truncated variance. Now, the benefit is that we have parameter estimates for the jump process. Looking at σ_J we observe a slight downward bias, which is unexpected since the bias should be really small in this case. Estimates of μ_j are positive, indicating that we are in fact in a folded normal distribution case. λ_J are in all but one case closely estimated to the true value.

In Figure 3a we show a P-P plot of the distribution of the J-test of overidentifying restrictions for each set of parameters. We see a similar pattern for the distribution as in Bolko et al. (2023), that is, in the lower portion of the diagram, the curves appear to the right of the reference line, while in the upper portion, they appear to the left. This indicates that there is a higher concentration of mass in the middle of the empirical distribution, implying that the test statistic exhibits a minor underdispersion (Bolko et al. 2023). In the appendix, we show the values the GMM approach estimates jump parameters using the replication sample without jumps.

In Figure 3b we show the kernel density plot of the standardized H. Overall, we see a good fit over the standard normal distribution, though with a heavy left tail.

9.11 A note on convergence

In testing and simulation of the models, we encountered multiple convergence problems. This section attempts to justify choices such as the path length of the simulations.

Andersen & Sørensen (1996) investigates the small-sample properties of the log-normal stochastic model when performing GMM estimation. They address the choice of number of moments to include in the estimation procedure and find that it depends heavily on the sample size. They find that as sample size increase one should exploit additional moment restrictions. Inclusion of

Parameter	Value	Quadratic variance	
		Initial value	Corrected estimate
Panel A:			
ξ	0.0225	0.0220 (0.0038)	0.0205 (0.0028)
λ	0.0050	0.0204 (0.0081)	0.0147 (0.0121)
v	1.2500	0.5480 (0.0090)	0.9552 (0.2493)
H	0.0500	0.2038 (0.0124)	0.0987 (0.0609)
σ_J	0.1000	0.0234 (0.0039)	0.0866 (0.0284)
λ_J	0.1000	0.0629 (0.0065)	0.1021 (0.0625)
μ_J	0.0000	0.0000 (0.0000)	0.0597 (0.0274)
Panel B:			
ξ	0.0225	0.0222 (0.0016)	0.0213 (0.0014)
λ	0.0100	0.0241 (0.0072)	0.0189 (0.0258)
v	0.7500	0.4658 (0.0070)	0.7304 (0.0258)
H	0.1000	0.2151 (0.0119)	0.1243 (0.0799)
σ_J	0.1000	0.0233 (0.0016)	0.0885 (0.0190)
λ_J	0.1000	0.0500 (0.0044)	0.0545 (0.0162)
μ_J	0.0000	0.0000 (0.0000)	0.0660 (0.0235)
Panel C:			
ξ	0.0225	0.0223 (0.0037)	0.0181 (0.0029)
λ	0.0150	0.0183 (0.0047)	0.0315 (0.0456)
v	0.5000	0.4619 (0.0079)	0.5364 (0.1200)
H	0.3000	0.3265 (0.0139)	0.2960 (0.1347)
σ_J	0.1000	0.0234 (0.0037)	0.0683 (0.0452)
λ_J	0.1000	0.0475 (0.0078)	0.1182 (0.1194)
μ_J	0.0000	0.0000 (0.0000)	0.0720 (0.0380)
Panel D:			
ξ	0.0225	0.0228 (0.0045)	0.0181 (0.0034)
λ	0.0350	0.0186 (0.0042)	0.0557 (0.0539)
v	0.3000	0.3714 (0.0059)	0.3417 (0.0931)
H	0.5000	0.3726 (0.0141)	0.4695 (0.1733)
σ_J	0.1000	0.0239 (0.0045)	0.0871 (0.0457)
λ_J	0.1000	0.0445 (0.0074)	0.0906 (0.1006)
μ_J	0.0000	0.0000 (0.0000)	0.0731 (0.0296)
Panel E:			
ξ	0.0225	0.0204 (0.0058)	0.0165 (0.0049)
λ	0.0700	0.0152 (0.0038)	0.0725 (0.0358)
v	0.2000	0.3197 (0.0059)	0.2225 (0.0320)
H	0.7000	0.3589 (0.0166)	0.6176 (0.1243)
σ_J	0.1000	0.0215 (0.0058)	0.0801 (0.0470)
λ_J	0.1000	0.0506 (0.0113)	0.1163 (0.1321)
μ_J	0.0000	0.0000 (0.0000)	0.0669 (0.0322)

Table 4: Initial values and estimation results of the fOU with jumps in X model.

excessive number of moments in small samples results in pronounced bias, so we want to choose a sample size which is appropriate.

Considering a small sample size of $T = 400$ we experienced a lot of convergence problems. This is in line with Andersen & Sørensen (1996) which had a 'fail to converse' percentage of approximately one third when considered a path length of $T = 500$ with their GMM procedure on a standard SV model. They also noticed a substantial bias in the estimates.

Andersen & Sørensen (1996) noticed an improvement of quality in interference when the sample size increased. The nonconvergence cases were almost nonexistent when setting $T = 1000$ or above (at least when having 9 or more moments). Therefore, we set the path length $T \geq 1000$ to avoid cases where the estimation procedure does not converge.

We note that the quality of inference is very sensitive to the number of moments relative to the sample size.

9.12 Kernel density estimate of standardized H under the different model settings

In this subsection, we present the kernel density estimates of standardized H in the different model settings described above.

We note that the statistic

$$\frac{\widehat{H} - H}{\widehat{\text{se}}(\widehat{H})} \xrightarrow{d} N(0,1). \tag{52}$$

We let J denote the Jacobi matrix from the optimizer. It can be shown that

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(0, \left(J^\top \Sigma_{\mathbb{IV}}^{-1} J\right)^{-1}\right) \tag{53}$$

when $T \to \infty$. Here

$$\Sigma_{\mathbb{IV}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\mathbb{IV}}(\ell) \quad \text{(Variance matrix)}$$

$$\Gamma_{\mathbb{IV}}(\ell) = \mathbb{E}_{\theta_0}\left[\left(\mathbb{IV}_1 - G\left(\theta_0\right)\right)\left(\mathbb{IV}_{1+\ell} - G\left(\theta_0\right)\right)^\top\right].$$

Since

$$\widehat{\Sigma}_T \xrightarrow{\mathbb{P}} \Sigma_{\mathbb{IV}}, \quad T \to \infty.$$

We can plug in $\widehat{\Sigma}_T^{-1}$ in (53) when we do inference on the parameters.

Figures 11-14 present kernel smoothed densities of the statistic presented in Equation (52). A standard normal distribution is also shown in blue as a reference distribution.

From the replication study of Bolko et al. (2023), the standard normal distribution appears to be a poor fit for the standardized H. The tails are considerably larger. For lower number of observations, we see an even worse fit. This shows the effect that a smaller T have on the distribution. Quite surprisingly, for the robustness check with $\ell = [0, 1, 2, 5, 10]$ (Figure 13, the standard normal distribution is a better fit, at least for all H except for $H = 0.1$. Lastly, for the robustness check with the Barlett kernel, the distributions show a good fit at least for all H expect for $H = 0.1$. This suggests that the Barlett kernel performs better than the Parzen kernel.

9.13 Jump testing

In this subsection we apply the jump test to 250 replications of a fSV model with jumps in Y An `R` script is available from Bibinger & Sonntag (2023) (linked in their paper). Specifically, they provide 3 simple functions which can run the tests, outputting the test-statistic. We have tried it by simply translating the code into `MATLAB` but it needed some modifications to make it work properly.

We apply the jump test where each parameter set follows from Bolko et al. (2023), with parameter set 1 being panel A and so on and with jump parameters $\lambda_J^{(Y)} = 0.0279$, $\sigma_J^{(Y)} = 1.25$ and $\mu_J^{(Y)} = 0$. The jump proportion is shown in Figure 4 for each replication. It clearly shows that for lower values of H the jump proportion increases.

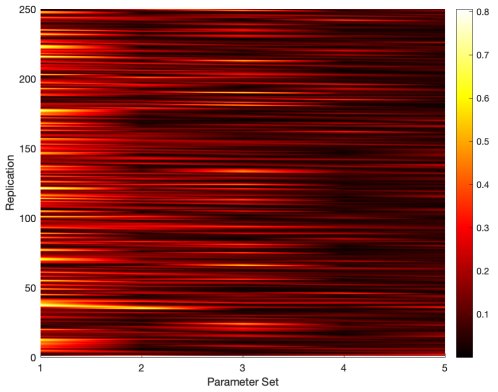


Figure 4: Heat map of jump proportion for each simulation. On the x-axis each parameter set of Bolko et al. (2023) is shown with 1 corresponding to Panel A and so on. For all parameter sets $\lambda_J^{(Y)} = 0.0279$.

10 Empirical

In this section, we will do extensive data analysis on all 30 stocks in the Dow Jones Industrial Average (DJIA). We will perform GMM estimation for each model considered (both with or without jumps). Note that Tables 5-6 and Figures 15-19 are placed in the appendix.

10.1 Data

Data are obtained by accessing the NYSE Trade and Quote (TAQ) database. From 2014, the transactions are recorded at the nanosecond level, so we exclude earlier observations (before 2014), as they are recorded at a much lower sampling rate. This way we can assume a much closer approximation to the underlying continuous time model.

We rely on data sampled at times $i\Delta_n$ for $i = 0, ..., \lfloor T/\Delta_n \rfloor$ at sometimes unevenly spaced times. We do not assume observance of the full continuous-time sample path.

The total sample size for each stock is 2535 days except for CVX which is missing one observation in 2018 due to a data error in the corresponding TAQ file. It is hard to say the magnitude of impact the missing observation will have on the estimates, but given that we have $T > 2500$, we conclude that one missing observation generally will have a negligible effect on the overall analysis.

10.1.1 Data cleaning

A crucial element in estimating the volatility of high-frequency data is effective data cleaning. The initial step of our cleaning procedure is detailed in Barndorff-Nielsen et al. (2009). The steps are as follows. We delete transactions with the timestamp outside the 9:30 am–4 pm window where the exchange is closed. Next, we delete entries where the transaction price, ask or bid is equal to zero, correlated trades which constitute serious errors (trades with a correction indicator, CORR not equal to 0) or abnormal sale conditions which the TAQ database has flagged as a problem (this is specified where trades have a letter code in the COND variable except for 'E' or 'F'). Deletion of entries with transaction price equal zero and correlated trades constitutes serious database errors such as misrecording of prices. This is followed by the use of the median price when multiple transactions have the same time stamp.

Barndorff-Nielsen et al. (2009) has a deletion step named T4, where entries such as excessive bid-ask/ask-bid spread are deleted. This has not been applied here, as the quote data have not been obtained by us. We note that this is rarely a problem in practice, as this is rarely activated and the RV estimator is insensitive to the use of T4 (shown in Table 1 and Table 2 in Barndorff-Nielsen et al. (2009)). We implement this procedure to exclude obvious noisy observations, which would degrade the realized volatility estimator, as it "treats" all observations equally.

To further clean for outliers we do as in Bolko et al. (2023) and remove RV estimates that deviate by more than 30 mean absolute deviations from the average RV computed from a rolling

window of 50 observations centered on, but not including, the data point being analyzed. This is a filter which does not remove many observations, but has a big effect on the autocorrelation of the RV estimates and also a large effect on our parameter estimates. We believe that this procedure will not remove jumps but only remove the most extreme data. On average this filter removes 1.8 observation per stock.

10.2 Empirical evidence for roughness

We first start by looking at the empirical estimates of the model by Bolko et al. (2023). Given that we are working with equity data, the presence of both drift and leverage is anticipated, therefore the approximate correction for RV is utilized. We use the same specification for the GMM estimation as in Section 9 only with the Barlett kernel, which we showed had better distributional results in finite samples than the Parzen kernel. The results of the GMM estimation are reported in Table 5. In the fSV model ξ corresponds to the average level of the spot variance process, λ indicates the speed of mean reversion, v represents the volatility of volatility and H is the Hurst exponent. The estimate of $\bar{\xi} = 0.0115$ translates to approximately 19.5% annualized volatility for the overall DJIA index. We find that the model under estimates ξ which is in line with our finding from the simulation analysis and is the same conclusion as Bolko et al. (2023). For λ, we estimate a much larger average of 0.9423 compared to 0.214. This shows that when looking at the DJIA constituencies, there is a much higher mean reversion speed than for the stock market indices estimated by Bolko et al. (2023). Comparing $\bar{v} = 1.8452$ with $\bar{v}_{Bolko} = 1.980$, we estimate that v is slightly higher but insignificant. We similarly observe very rough volatility with an average $\bar{H} = 0.0237$ that is approximately the same as Bolko et al. (2023) which found an average estimate of 0.024. We observe very rough volatility on every stock, with no stock having $H > 0.07$. The scatter plot of the estimated H versus the first-order autocorrelation coefficient ρ_1 with a fitted OLS regression line is shown in Figure 15, indicating a positive relationship between H and ρ_1 with a $\widehat{\beta} = 0.1005$ ($\widehat{\alpha} = -0.0514$).

Figure 16 shows the autocorrelation function of AAPL and DJIA. For AAPL, we observe significant memory, but with a much quicker decay for the first 50 lags than for S&P 500 index (Figure 5 in Bolko et al. (2023)). With could indicate even more roughness or the presence of jumps. Noticeably, we observe an increase in sample autocorrelation from about lag 75 to lag 125 that quickly decays from lag 150. This is quite different than for the S&P 500 index. This leads us to conclude that there may be some memory of the process in that lag area. This will be important for us when estimating the roughness of volatility in the presence of jumps, where we have an increase in moments that has to be estimated (due to more parameters).

To assess the statistical adequacy of the model, we compute the J-test for overidentifying restrictions. The results are displayed in the final column of Table 5. In general, the high P-values indicate that the fSV process effectively represents the data.

We might suspect that there is microstructure noise in the empirical data, so we perform the same GMM estimation on pre-averaged variance estimates of the DJIA constituents. The results are reported in Table 14. The first thing we notice is that λ estimates is much lower indicating a lower mean reversion speed. The second thing we notice is that $\widehat{H}$ is on average the same. Third, we notice a clear outlier in ξ, namely AAPL with an $\xi = 0.0620$.

As a robustness check, we also find evidence for roughness in SPX data using the method described in Gatheral et al. (2022) with an H in line with the paper (shown in Appendix A.4).

10.3 Empirical evidence for jumps

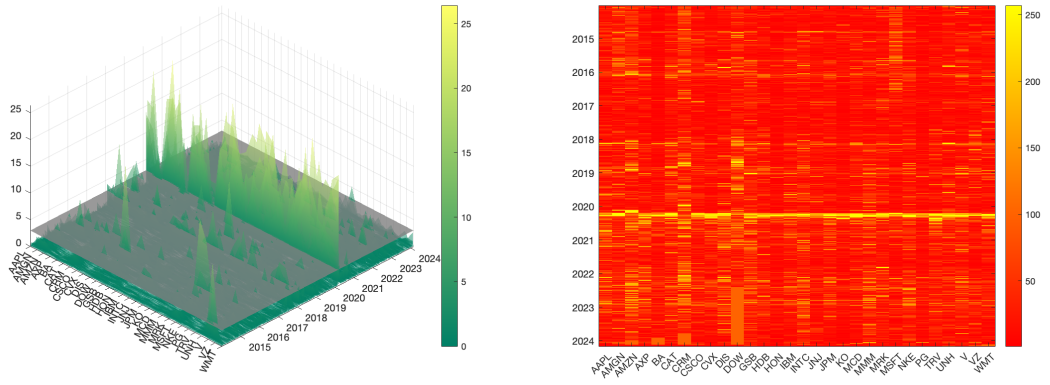
In this subsection, we will provide evidence for jumps in both prices and volatility.

10.3.1 Jumps in prices

In Figure 17 we provide a scatter plot of RV versus BV for AAPL and DJIA. The regressions show a $\beta < 1$ which indicates a significant presence of jumps. Without the presence of jumps, IV should be QV (and therefore $BV = RV$, at least under a semimartingale volatility assumption), and estimates are expected to align along a 45-degree line.

Symbol	$\overline{RV}$	ρ_1	GMM Estimate				$\mathcal{J}_{\text{HS}}$
			ξ	$\lambda \times 100$	v	H	
AAPL	0.0145	0.6346	0.0123	0.1268	1.8364	0.0218	0.3640
AMGN	0.0160	0.6139	0.0140	0.2006	1.8992	0.0193	0.9127
AMZN	0.0198	0.6478	0.0171	0.0878	1.4903	0.0289	0.5223
AXP	0.0146	0.8046	0.0107	3.3269	1.8522	0.0306	0.4657
BA	0.0277	0.7047	0.0176	0.0010	2.0029	0.0132	0.6522
CAT	0.0173	0.6716	0.0155	0.0192	1.9671	0.0126	0.8359
CRM	0.0233	0.7172	0.0189	0.0649	1.5379	0.0207	0.8210
CSCO	0.0122	0.7591	0.0103	0.0010	2.0244	0.0096	0.9094
CVX	0.0165	0.8307	0.0128	0.2282	1.3186	0.0545	0.7580
DIS	0.0143	0.8000	0.0120	0.0010	1.7831	0.0166	0.8480
DOW	0.0226	0.7236	0.0161	0.0327	1.6789	0.0160	0.4555
GS	0.0156	0.8455	0.0127	3.4937	1.4362	0.0433	0.4423
HD	0.0126	0.8413	0.0103	0.0027	1.7417	0.0160	0.7426
HON	0.0110	0.6688	0.0087	0.0010	2.1647	0.0112	0.7165
IBM	0.0098	0.8003	0.0082	0.0822	1.9553	0.0187	0.7659
INTC	0.0188	0.7602	0.0160	0.1837	1.7810	0.0254	0.5970
JNJ	0.0079	0.7233	0.0068	0.0540	2.1764	0.0183	0.9107
JPM	0.0132	0.8002	0.0111	0.0011	1.7232	0.0167	0.7928
KO	0.0075	0.7087	0.0058	0.0011	1.9809	0.0095	0.7928
MCD	0.0088	0.8922	0.0062	5.5421	1.5728	0.0696	0.5394
MMM	0.0105	0.7996	0.0089	0.1247	1.9266	0.0213	0.6926
MRK	0.0107	0.6651	0.0096	0.1029	2.2242	0.0138	0.8732
MSFT	0.0224	0.7493	0.0196	0.2448	1.5811	0.0352	0.8344
NKE	0.0142	0.7834	0.0124	0.0098	1.8951	0.0140	0.9133
PG	0.0082	0.7594	0.0063	0.0010	2.1449	0.0096	0.8520
TRV	0.0108	0.8114	0.0081	9.6701	1.9717	0.0328	0.4626
UNH	0.0144	0.7324	0.0127	0.1113	1.8752	0.0296	0.8503
V	0.0121	0.7573	0.0103	0.0049	2.1346	0.0115	0.8735
VZ	0.0088	0.6903	0.0076	0.6149	1.9582	0.0191	0.6577
WMT	0.0088	0.7323	0.0072	3.9320	1.7207	0.0525	0.3046
Average	0.0143	0.7476	0.0115	0.9423	1.8452	0.0237	0.7017
Std	(0.0119)	(0.0682)	(0.0039)	(2.1732)	(0.2318)	(0.0147)	

Table 5: Empirical estimates of model by Bolko et al. (2023). $\overline{RV}$ represents the sample mean of realized variance. ρ_1 denotes the first-order autocorrelation of realized variance. $\mathcal{J}_{\text{HS}}$ presents the P-value of the Sargan–Hansen test for overidentifying restrictions, which is asymptotically $\chi^2(4)$-distributed. λ is scaled by 100 for better readability. The bottom row shows the standard deviation while the row above shows the cross-sectional average.



(a) **(b)**

Figure 5: (a): Sequential jump test with an imposed 5% critical value plane in grey. (b): Heat map of the sequential jump test with test statistic scaled by 10 for color correction.

10.3.2 Jumps in volatility

In Figure 18 standardized second-order increments of RV for AAPL are plotted with blue triangles that indicate specific jumps pinpointed by the sequential jump test of Bibinger & Sonntag (2023). As seen, we do not observe many jumps on a 5% significance level. The test show a cluster of jump around the corona pandemic in 2020.

In Table 13 we provide estimates of the jump proportion of each stock in the DJIA, the individual jump count, and the market beta value of the specific stock. As seen, there is a great deal of difference in the jump count, with the highest (BA) being 16 times as frequent as the smallest (VZ). This shows a huge spread in the intensity of jumps across the DJIA constituents. We might expect stocks with high jump intensity to be more sensitive to market changes. Comparing the beta value to the jump count we see a positive correlation with a significant regression coefficient at 6% increase in beta per jump.

In Figures 5a and 5b the sequential jump test is performed on all 30 DJIA constituencies using RV estimates. Figure 5 shows that the jumps in volatility occur both independently and in clusters. The heat map in Figure 5b shows how the DJIA constituencies react to global events which we expect will cause a jump in volatility. We see the first (small) cluster around the end of 2015 which coincides with the presidential election in the USA. The second cluster occurs in the beginning of 2018, which could be due to the US government entering a federal government shutdown (January) and the beginning of a trade war between the USA and China (March). The biggest cluster occurs around the corona pandemic outbreak in 2020 with a huge spark in the test statistic over multiple days. We also see a general higher jump statistic from around 2022, which indicates a more unstable volatility process with more jumps. This coincides with the invasion of Ukraine.

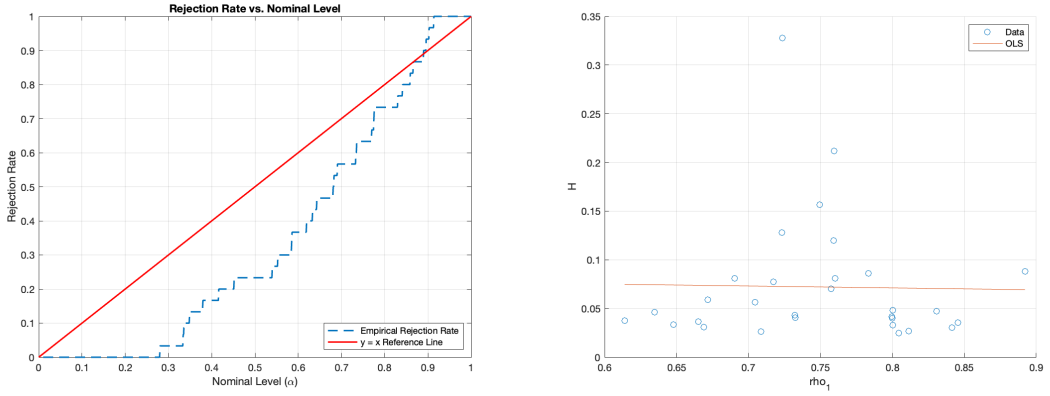
10.4 Empirical analysis of the fSV model in the presence of price jumps

The results are shown in Table 6. Compared to Table 5, we now observe a much higher average H estimate indicating that volatility is less rough. We argue that some of what is perceived as roughness is actually price jumps. This is shown by the estimates of $\lambda_J^{(X)}$ which are all different from zero. The average H estimates are more than three times as large as we previously observed. All $\mu_J^{(X)}$ estimates are positive, which, as discussed in Section 8.2, indicates a possibly positive bias in the estimates. We observe $\lambda_J^{(X)}$, which is on average 0.0832 indicating that we do experience price jumps (about 20 per year).

In Figure 6a a P-P plot is shown and on Figure 6b, we now see no correlation between ρ_1 and H. We also observe a larger spread in H and a clear outlier at $H > 0.3$. This is interesting and we would suspect a positive correlation, since ρ_1 explains the correlation between increments just

Symbol	GMM Estimate							$\mathcal{J}_{\text{HS}}$
	ξ	$\lambda \times 100$	v	H	$\sigma_J^{(X)}$	$\lambda_J^{(X)}$	$\mu_J^{(X)}$	
AAPL	0.0108	0.0835	1.4731	0.0388	0.0242	0.1211	0.1197	0.4158
AMGN	0.0124	0.2147	1.3092	0.0501	0.1167	0.1232	0.0655	0.7326
AMZN	0.0160	0.4687	1.2266	0.0763	0.0434	0.2578	0.1098	0.7354
AXP	0.0108	0.0899	0.9206	0.0879	0.1330	0.0495	0.0187	0.5392
BA	0.0176	0.0007	1.4952	0.0324	0.1035	0.0872	0.1411	0.3787
CAT	0.0144	0.0822	1.3236	0.0345	0.0962	0.0747	0.0926	0.6807
CRM	0.0195	0.4743	1.1762	0.0721	0.1052	0.1488	0.0715	0.6829
CSCO	0.0092	0.0993	0.8448	0.0835	0.0560	0.0704	0.1184	0.8959
CVX	0.0117	0.0429	1.3506	0.0430	0.0927	0.0357	0.0500	0.3334
DIS	0.0108	0.3709	0.4822	0.2227	0.0000	0.0678	0.1718	0.9027
DOW	0.0118	0.0867	0.1252	0.3523	0.0103	0.1743	0.1804	0.7741
GS	0.0129	0.0457	1.3973	0.0302	0.0303	0.0405	0.0703	0.3489
HD	0.0094	0.0041	1.4645	0.0238	0.0201	0.0482	0.1236	0.5856
HON	0.0078	0.0011	1.4997	0.0226	0.1060	0.0713	0.0289	0.6193
IBM	0.0072	0.3802	0.3844	0.2382	0.0405	0.0784	0.1135	0.8299
INTC	0.0163	0.2981	0.7200	0.1180	0.0268	0.0042	0.4149	0.8408
JNJ	0.0060	0.3055	0.8337	0.1170	0.0642	0.0473	0.1146	0.8895
JPM	0.0099	0.0633	1.0209	0.0635	0.0234	0.0488	0.1312	0.4510
KO	0.0056	0.0005	0.7263	0.0462	0.0852	0.0195	0.0740	0.6431
MCD	0.0061	2.3623	1.3881	0.0880	0.0023	0.0166	0.0367	0.2804
MMM	0.0086	0.0747	1.3277	0.0416	0.0145	0.0096	0.2000	0.5528
MRK	0.0082	0.1587	1.4995	0.0365	0.0669	0.1254	0.0802	0.7757
MSFT	0.0160	0.6564	0.7498	0.1567	0.0548	0.4674	0.0750	0.6901
NKE	0.0117	0.1058	0.7556	0.0862	0.0819	0.0338	0.1265	0.9130
PG	0.0056	0.1542	0.3364	0.2116	0.0460	0.0437	0.1207	0.8585
TRV	0.0075	0.0005	1.4050	0.0269	0.0349	0.1314	0.0962	0.6331
UNH	0.0117	0.0841	1.4997	0.0406	0.1137	0.0440	0.0314	0.5839
V	0.0096	0.0802	0.9021	0.0704	0.0820	0.0254	0.1526	0.8653
VZ	0.0073	0.2591	0.7979	0.0808	0.1109	0.0262	0.0385	0.7699
WMT	0.0070	0.0623	1.1679	0.0434	0.0000	0.0037	0.2326	0.3357
Average	0.0106	0.2370	0.2370	0.0879	0.0595	0.0832	0.1134	0.6513
Std	(0.0037)	(0.4346)	(0.3982)	(0.0770)	(0.0770)	(0.0921)	(0.0771)	

Table 6: Empirical results of our GMM estimation with the assumption of jumps in log prices. $\mathcal{J}_{\text{HS}}$ indicates the P-value obtained from the Sargan–Hansen test for overidentifying restrictions, which follows an asymptotic $\chi^2(4)$ distribution. The variable λ has been scaled by a factor of 100 for enhanced readability. The bottom row represents the standard deviation, whereas the row immediately above it displays the cross-sectional mean.



(a) **(b)**

Figure 6: (a) Distribution of the J-test of overidentifying restrictions on DJIA stocks. (b) OLS regression of ρ_1 on Estimated H.

as H does. We however note that we only have 30 estimates and therefore one for investigate this on a larger scale before coming to conclusions.

11 Conclusion

We study a stochastic volatility model with a fractional OU process driving the log-spot variance. The main purpose is to analyse roughness in volatility. An analysis of the 30 constituents of the Dow Jones Industrial Average confirms the notion of rough volatility in stocks.

Another subject of the thesis is analysing the robustness of the GMM approach for estimating the volatility parameters we are applying.

The performance of the GMM estimation under different data and model specifications was as expected. The effect of a smaller sample size was increased uncertainty and poorer estimates. Increasing the sampling frequency gave more precise estimates. Fewer moments resulted in a deterioration of estimates and precision. Pre-averaging in order to minimize microstructure noise was not superior to using 5-minute realized variance.

The GMM method is not robust to the presence of jumps in prices or in log-variance using realized variance. Applying truncated realized variance is robust toward jumps in prices and gave precise and consistent estimates. We conclude that one has to be careful in implementing the estimation procedure without data considerations in mind. In order to do the robustness checks for jumps we develop a general fractional stochastic volatility plus jumps simulation setup. The setup includes five different ways of including jumps. We give algorithms for implementing the simulation as well as the GMM estimation procedure.

Another study subject considered is testing for presence of jumps in a model setting with a fractional volatility process. We use a test developed for a non-semimartingale environment and show a significant presence of jumps in the empirical data. The results are illustrated in two- and three-dimensional figures and tables. In the empirical data, jump intensity of volatility jumps has an average value of 1.3 per year, much smaller than the amount when compared to a non-fractional setup. We also show that, under the jump-robust setup, estimates or the Hurst parameter increase almost fourfold to a value of 0.09. This suggests that jumps do not cancel out the presence of rough volatility. Not only do we study estimation of the Hurst parameter under the presence of jumps, but we are also interested in estimating the parameters of the jump process along with volatility parameters.

Another important contribution is that we have taken the first steps to introduce modifications of the GMM approach regarding jumps in prices and log-spot variance in a fractional setup. The modification demonstrates potential with estimates of H that are not far from true values. Regarding jumps in prices, we assume that estimates of jump size mean and variance are slightly

biased. The bias is close to the expected theoretical bias, and we expect that it will be possible to bias-correct the estimates.

We suggest a number of future study subjects, which have emerged from the work with this thesis. The list includes comparison of different jump test methods, the effect of filtering out jumps, study co-jumps models with respect to estimation, and the connection between jumps in prices and volatility. Finally, more work is needed regarding modification of GMM to include estimation of jump processes, both theoretically and analytically. In addition, introducing more complicated fractional Lévy processes to simulate jump behavior could be of interest when using GMM.

References

Aït-Sahalia, Y. & Jacod, J. (2009), 'Testing for jumps in a discretely observed process'.

Aït-Sahalia, Y. & Jacod, J. (2012), 'Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data', *Journal of Economic Literature* **50**(4), 1007–1050.

Alos, E., León, J. A. & Vives, J. (2007), 'On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility', *Finance and stochastics* **11**(4), 571–589.

Andersen, T. G. & Sørensen, B. E. (1996), 'Gmm estimation of a stochastic volatility model: A monte carlo study', *Journal of Business & Economic Statistics* **14**(3), 328–352.

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica: Journal of the Econometric Society* pp. 817–858.

Aït-Sahalia, Y. & Jacod, J. (2014), *High-Frequency Financial Econometrics*, Princeton University Press, Princeton.
URL: *https://doi.org/10.1515/9781400850327*

Back, K. (1991), 'Asset pricing for general processes', *Journal of Mathematical Economics* **20**(4), 371–395.
URL: *https://www.sciencedirect.com/science/article/pii/030440689190037T*

Bakshi, G., Cao, C. & Chen, Z. (1997), 'Empirical performance of alternative option pricing models', *The Journal of finance* **52**(5), 2003–2049.

Bandi, F. M. & Renò, R. (2012), 'Time-varying leverage effects', *Journal of Econometrics* **169**(1), 94–113.

Bandi, F. M. & Reno, R. (2016), 'Price and volatility co-jumps', *Journal of Financial Economics* **119**(1), 107–146.

Barboza, L. A. & Viens, F. G. (2017), 'Parameter estimation of gaussian stationary processes using the generalized method of moments'.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Estimating quadratic variation using realized variance', *Journal of Applied econometrics* **17**(5), 457–477.

Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Power and bipower variation with stochastic volatility and jumps', *Journal of financial econometrics* **2**(1), 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), 1–30.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realized kernels in practice: trades and quotes', *The Econometrics Journal* **12**(3), C1–C32.
URL: *https://doi.org/10.1111/j.1368-423X.2008.00275.x*

Bates, D. S. (2000), 'Post-'87 crash fears in the s&p 500 futures option market', *Journal of econometrics* **94**(1-2), 181–238.

Bayer, C., Friz, P. & Gatheral, J. (2016), 'Pricing under rough volatility', *Quantitative Finance* **16**(6), 887–904.

Bennedsen, M., Lunde, A. & S, P. M. (2022), 'Decoupling the short-and long-term behavior of stochastic volatility', *Journal of Financial Econometrics* **20**(5), 961–1006.

Bibinger, M. & Sonntag, M. (2023), 'Testing for jumps in processes with integral fractional part and jump-robust inference on the hurst exponent', *arXiv preprint arXiv:2305.01751* .

Billingsley, P. (2017), *Probability and measure*, John Wiley & Sons.

Bolko, A. E., Christensen, K., Pakkanen, M. S. & Veliyev, B. (2023), 'A gmm approach to estimate the roughness of stochastic volatility', *Journal of Econometrics* **235**(2), 745–778.

Bollerslev, T. & Zhou, H. (2002), 'Estimating stochastic volatility diffusion using conditional moments of integrated volatility', *Journal of Econometrics* **109**(1), 33–65.

Cheridito, P., Kawaguchi, H. & Maejima, M. (2003), 'Fractional ornstein-uhlenbeck processes'.

Christensen, K., Oomen, R. C. & Podolskij, M. (2014), 'Fact or friction: Jumps at ultra high frequency', *Journal of Financial Economics* **114**(3), 576–599.

Clark, P. K. (1973), 'A subordinated stochastic process model with finite variance for speculative prices', *Econometrica: journal of the Econometric Society* pp. 135–155.

Comte, F. & Renault, E. (1996), 'Long memory continuous time models', *Journal of Econometrics* **73**(1), 101–149.
URL: *https://www.sciencedirect.com/science/article/pii/0304407695017356*

Comte, F. & Renault, E. (1998), 'Long memory in continuous-time stochastic volatility models', *Mathematical Finance* **8**(4), 291–323.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9965.00057*

Cont, R. & Das, P. (2022), 'Rough volatility: fact or artefact?', *arXiv preprint arXiv:2203.13820* .

Davidson, J. (2020), 'A new consistency proof for hac variance estimators', *Economics Letters* **186**, 108811.

Decreusefond, L. & Üstünel, A. S. (1999), 'Stochastic analysis of the fractional brownian motion', *Potential analysis* **10**, 177–214.

Duffie, D., Pan, J. & Singleton, K. (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica* **68**(6), 1343–1376.

Eraker, B., Johannes, M. & Polson, N. (2003), 'The impact of jumps in volatility and returns', *The Journal of Finance* **58**(3), 1269–1300.

Fama, E. F. (1963), 'Mandelbrot and the stable paretian hypothesis', *The Journal of Business* **36**(4), 420–429.
URL: *http://www.jstor.org/stable/2350971*

Fukasawa, M. (2017), 'Short-time at-the-money skew and rough fractional volatility', *Quantitative Finance* **17**(2), 189–198.

Fukasawa, M., Takabatake, T. & Westphal, R. (2022), 'Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics', *Mathematical Finance* **32**(4), 1086–1132.

Garnier, J. & Sølna, K. (2018), 'Option pricing under fast-varying and rough stochastic volatility', *Annals of Finance* **14**(4), 489–516.

Gatheral, J., Jaisson, T. & Rosenbaum, M. (2022), Volatility is rough, *in* 'Commodities', Chapman and Hall/CRC, pp. 659–690.

Hall, A. (2005), *Generalized Method of Moments*, Advanced texts in econometrics, Oxford University Press.
URL: *https://books.google.dk/books?id=8YkSDAAAQBAJ*

Hall, A. R. (2013), 'Generalized method of moments', *Handbook of research methods and applications in empirical macroeconomics* pp. 313–333.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica: Journal of the econometric society* pp. 1029–1054.

Hansen, L. P. & Hodrick, R. J. (1980), 'Forward exchange rates as optimal predictors of future spot rates: An econometric analysis', *Journal of political economy* **88**(5), 829–853.

Hansen, L. P. & Singleton, K. J. (1982), 'Generalized instrumental variables estimation of nonlinear rational expectations models', *Econometrica: Journal of the Econometric Society* pp. 1269–1286.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business & Economic Statistics* **24**(2), 127–161.

Jaber, E. A. & Carvalho, N. D. (2023), 'Reconciling rough volatility with jumps', *Available at SSRN 4387574* .

Jacod, J. (2012), 'Statistics and high frequency data', *Statistical methods for stochastic differential equations* **124**, 191–310.

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**(7), 2249–2276.

Jacod, J. & Todorov, V. (2010), 'Do price and volatility jump together?'.

Kaarakka, T. & Salminen, P. (2011), 'On fractional ornstein-uhlenbeck processes', *Communications on Stochastic Analysis* **5**(1), 8.

Kroese, D. P. & Botev, Z. I. (2014), Spatial process simulation, *in* 'Stochastic geometry, spatial statistics and random fields: Models and algorithms', Springer, pp. 369–404.

Lee, S. S. & Mykland, P. A. (2008), 'Jumps in financial markets: A new nonparametric test and jump dynamics', *The Review of Financial Studies* **21**(6), 2535–2563.

Lindgren, G. (2006), 'Lectures on stationary stochastic processes', *PhD course Lund University* .

Mancini, C. (2009), 'Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps', *Scandinavian Journal of Statistics* **36**(2), 270–296.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *The Journal of Business* **36**(4), 394–419.
URL: *http://www.jstor.org/stable/2350970*

Mandelbrot, B. B. & Van Ness, J. W. (1968), 'Fractional brownian motions, fractional noises and applications', *SIAM review* **10**(4), 422–437.

Marquardt, T. (2006), 'Fractional lévy processes with an application to long memory moving average processes', *Bernoulli* **12**(6), 1099–1126.

Morlanes, J. I. (2017), Some Extensions of Fractional Ornstein-Uhlenbeck Model: Arbitrage and Other Applications, PhD thesis, Department of Statistics, Stockholm University.

Mykland, P. A. & Zhang, L. (2012), 'The econometrics of high frequency data', *Statistical methods for stochastic differential equations* **124**, 109.

Newey, W. K. & McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Omar, E. E., Masaaki, F. & Mathieu, R. (2016), 'The microstructural foundations of leverage effect and rough volatility', *arXiv preprint arXiv:1609.05177* .

Oomen, R. C. A. (2006), 'Comment', *Journal of Business & Economic Statistics* **24**(2), 195–202. **URL:** *https://doi.org/10.1198/073500106000000125*

Pan, J. (2002), 'The jump-risk premia implicit in options: Evidence from an integrated time-series study', *Journal of financial economics* **63**(1), 3–50.

Podolskij, M. & Vetter, M. (2009*a*), 'Bipower-type estimation in a noisy diffusion setting', *Stochastic processes and their applications* **119**(9), 2803–2831.

Podolskij, M. & Vetter, M. (2009*b*), 'Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps', *Bernoulli* **15**(3), 634 – 658. **URL:** *https://doi.org/10.3150/08-BEJ167*

Press, S. J. (1967), 'A compound events model for security prices', *Journal of business* pp. 317–335.

Renault, E. (2009), 'Moment–based estimation of stochastic volatility models', *Handbook of Financial Time Series* pp. 269–311.

Resnick, S. (2019), *A probability path*, Springer.

Rogers, L. (2023), Things we think we know, *in* 'Options—45 years since the Publication of the Black–Scholes–Merton Model: The Gershon Fintech Center Conference', World Scientific, pp. 173–184.

Rogers, L. C. G. (1997), 'Arbitrage with fractional brownian motion', *Mathematical finance* **7**(1), 95–105.

Tankov, P. (2003), *Financial modelling with jump processes*, Chapman and Hall/CRC.

Veraart, A. E. (2011), 'How precise is the finite sample approximation of the asymptotic distribution of realised variation measures in the presence of jumps?', *AStA Advances in Statistical Analysis* **95**, 253–291.

Wang, X., Xiao, W. & Yu, J. (2021), 'Modeling and forecasting realized volatility with the fractional ornstein–uhlenbeck process', *Journal of Econometrics* .

A Appendix

A.1 Properties of a Poisson and compound Poisson process

The mean and variance for Z_t is:

$$\mathbb{E}[Z_t] = \mathbb{E}[N_t] \cdot \mathbb{E}[J] = \lambda \cdot t \cdot \mathbb{E}[J]$$

$$\text{Var}[Z_t] = \mathbb{E}[N_t] \cdot \mathbb{E}[J^2] = \lambda \cdot t \cdot \mathbb{E}[J^2].$$

These can be obtained by applying the laws of total mean and total variance.

The moments for the Poisson process $(N_t)_{t>0}$ is:

$$\mathbb{E}[N_t] = \lambda \cdot t$$

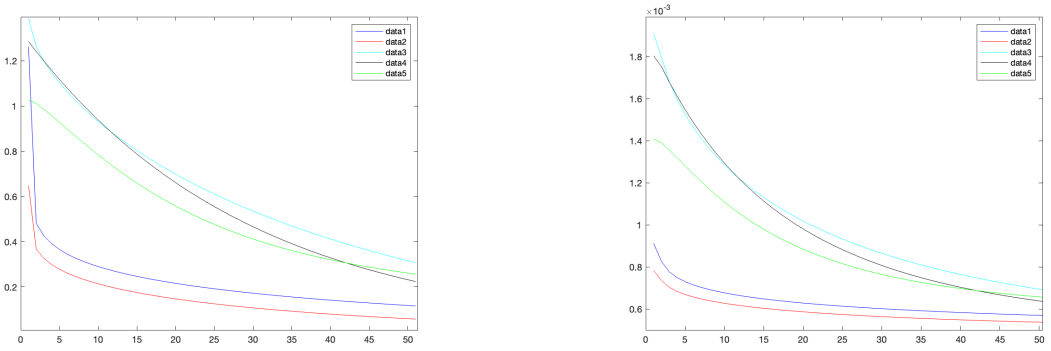
$$\mathbb{E}[N_t - N_s] = \mathbb{E}[N_{t-s}] = \lambda \cdot (t - s).$$

Thus

$$\mathbb{E}[N_t - N_{t-1}] = \lambda.$$

Furthermore $(N_t - N_{t-1})$ and $N_{t+l} - N_{t+l-1}$ are independent for $l > 0$ and

$$\mathbb{E}[N_t^2] = \lambda \cdot t + \lambda^2 \cdot t^2$$



(a) **(b)**

Figure 7: Each parameter set of Bolko et al. (2023) with data1=Panel A and so forth with (a) $\kappa(\ell)$ (b) $\mathbb{E}[IV_t IV_{t+\ell}]$ and ℓ on the x axis.

$$\mathbb{E}[N_t \cdot N_{t+l}] = \lambda t + \lambda^2 t \cdot (t + l).$$

In general $\mathbb{E}[N_t \cdot N_s] = \lambda min(t, s) + \lambda^2 ts$.

The moment generating function for Z is

$$\mathbb{E}[e^{(Z_T - Z_t)\alpha}] = \exp((T - t)\lambda(\mathbb{E}[e^{\alpha J}] - 1)), \quad \alpha \in \mathbb{R}$$

with

$$Z_T - Z_t = \sum_{k=N_t+1}^{N_T} J_k.$$

The autocovariance function for $(Z_t)_{t>0}$ is

$$\mathrm{Cov}(Z_t, Z_s) = \lambda \mathbb{E}[J_t^2] min(t, s)$$

with

$$\mathrm{Cov}(Z_t, Z_{t+l}) = \lambda t \mathbb{E}[J_t^2].$$

A.2 Plots of $\kappa(\ell)$ and $\mathbb{E}[IV_t IV_{t+\ell}]$

In Figure 7, we show the rate of decay of $\kappa(\ell)$ and $\mathbb{E}[IV_t IV_\ell]$.

A.3 Algorithms

Algorithm 1 An simple algorithm for simulating a discrete-time fractional Brownian motion

Require: A Hurst parameter $H \in (0,1)$, an integer indicating the number of grid/simulation points within one day $N \geq 1$, and an integer giving the number of simulated days $T \geq 1$.

$r \leftarrow NaN_{1:n}$ ▷ initialize the vector $r_{1:n}$ with NaN's.

$r_1 = 1$

for k in 1:n **do**

$r_{k+1} = 0.5 \cdot \left((k+1)^{2\cdot H} - 2 \cdot k^{2\cdot H} + (k-1)^{2\cdot H}\right)$

end for

$x = r_{n:2}$ ▷ Flip the vector r so it runs from $n \to 2$.

$r = [r, x]$ ▷ Elongate the vector r with x.

$\lambda \leftarrow \text{FFT}(r)/2n$ ▷ Create eigenvalues by taking the fast Fourier transform (FFT) of r divided by $2n$.

$W \leftarrow \text{FFT}\left(\sqrt{\lambda} \odot (N_{1\times 2n} + 1i \cdot N_{1\times 2n})\right)$ ▷ $N_{1\times 2n}$: $2n$ random draws of N(0,1)

▷ $\odot$: Hadamard product

$W \leftarrow n^{-H} \cdot \text{cumsum}(\Re(W)$ ▷ cumsum = cumulative summation function,

▷ $\Re(.)$) = real part function.

$W \leftarrow t^H W$ ▷ For rescalling.

return W

Algorithm 2 An algorithm for simulation of the fSV model with co-jumps

Require: A parameter vector θ, X_0, a number of observations per day N and a number of days T.

Set κ_0 according to equations (17).

$\Delta \leftarrow 1/N$

$\eta \leftarrow \log(\xi) - 0.5 \cdot \kappa_0$

$Y_0 \sim N(\eta, \kappa_0)$ ▷ Draw Y_0 from $N(\eta, \kappa_0)$.

$\sigma_0 \leftarrow \exp(Y_0/2)$

$B^H_{1:N\times T} \sim \text{fBm}(H, N, T)$ ▷ Draw $N \times T$ increments from the fBm algorithm.

$W_{1:N\times T} \sim N(0, \Delta)$ ▷ Draw $N \times T$ random samples from a $N(0, \Delta)$.

for $t = 1 : T \times N$ **do**

$P_1 \sim pois(\lambda_J^{(X)} \cdot \Delta)$ ▷ Draw one random sample from Poisson distribution

$P_2 \sim pois(\lambda_J^{(Y)} \cdot \Delta)$

$P_{12} \sim pois(\lambda_J^{XY} \cdot \Delta)$

$N_1 \sim N(\mu_J^{(X)}, \sigma_j^{(X)^2})$ ▷ Draw one random sample from normal distribution

$N_2 \sim N(\mu_J^{(Y)}, \sigma_j^{(Y)^2})$

$x = \rho_J \cdot \sigma_J^{(XY)} \cdot \sigma_J^{(YX)}$

$N_{12} \sim N\left(\begin{bmatrix} \mu_J^{(XY)} \\ \mu_J^{(YX)} \end{bmatrix}, \begin{bmatrix} \sigma_J^{(XY)^2} & x \\ x & \sigma_J^{(YX)^2} \end{bmatrix}\right)$ ▷ Multivariate normal distribution.

$Y_t \leftarrow \eta + (Y_{t-1} - \eta)\, e^{-\lambda\Delta} + ve^{-\lambda\Delta/2}\left(B_t^H - B_{t-1}^H\right) + P_2 N_2 + P_{12} N_{12}(2)$

▷ $N_{12}(2) takes second output of N_{12}$.

$\sigma_t \leftarrow exp(Y_t/2)$

$X_t \leftarrow X_{t-1} + \sigma_t\,(W_t - W_{t-1}) + P_1 N_1 + P_{12} N_{12}(1)$ ▷ Euler scheme

end for

return Y, σ, X ▷ Return time series of Y, σ and X.

Algorithm 3 An algorithm for calculating initial values of the parameter vector θ

Require: A time series of volatility measures $\mathbf{V}_T$, a integer q,a integer series of lags $h = 1, ..., m$ and an integer indicating the number of days T.

$\xi \leftarrow T^{-1} \sum_{t=1}^{T} \mathbf{V}_t$

$K_q \leftarrow 2^{q/2} \cdot \frac{\Gamma((q+1)/2)}{\sqrt{\pi}}$

$m \leftarrow h_m$ ▷ last element in the vector h.

for i in h **do**

$\hat{\gamma_h} = \frac{1}{T-m} \sum_{t=1}^{T-m} \left| \ln\left(\mathbf{V}_{t+i}^n\right) - \ln\left(\mathbf{V}_t^n\right) \right|^q$

end for

Model $\leftarrow$ OLS($ln(h), ln(\hat{\gamma})$) ▷ Use a OLS regression algorithm for $ln(h)$ on $ln(\hat{\gamma})$.

$v \leftarrow exp\left(\frac{\text{Model}_\alpha - ln(K_q)}{q}\right)$ ▷ Use α coefficient from OLS regression.

$\zeta_q \leftarrow \text{Model}_\beta$ ▷ Use β coefficient from OLS regression.

Model $\leftarrow$ OLS(q, ζ_q)

$H \leftarrow \text{Model}_\beta$

$\lambda \leftarrow \left(\frac{v^2}{2 \cdot var(ln(\mathbf{V}_T))} \Gamma\left(1 + 2H\right)\right)^{\frac{1}{2H}}$

return ξ, λ, v, and H.

Algorithm 4 An algorithm for GMM estimation

Require: A parameter vector θ, a empirical sample matrix $\mathbb{V}$ and a vector of lag values ℓ

$W \leftarrow \mathbf{I}(\ell + 5)$ ▷ Create weighting matrix using identity matrix

for i = 1 **do** ▷ Optimize θ using the lsqnonlin function using the following object function

$g \leftarrow G(\theta, \ell)$ ▷ Calculate the theoretical moments using θ and ℓ

$\bar{m} \leftarrow mean(\mathbb{V}_{1:T})$ ▷ Take mean of sample matrix along time dimension

$\hat{m} \leftarrow \bar{m} - g$

$f \leftarrow W \cdot \hat{m}$

return $\theta \leftarrow \min_f$

end for

for i= 2:3 **do**

$g \leftarrow G(\theta, \ell)$

$\widehat{\Gamma}(0) \leftarrow \frac{1}{T} \sum_{t=1}^{T-l} \left(\widehat{\mathbb{V}}_t - g\right)^\top \left(\widehat{\mathbb{V}}_{t+\ell} - g\right)$

$\widehat{\Sigma}_T \leftarrow \widehat{\Gamma}(0) + \sum_{\ell=1}^{T-1} w(\ell/L) \left[\widehat{\Gamma}(\ell) + \widehat{\Gamma}(\ell)^\top\right]$

Optimize θ as done when $i = 1$

end for

Algorithm 5 An algorithm for calculating the HAC estimator when there is jumps in X

Require: A sample matrix $\mathbb{V}$, a time series of jump-robust volatility measures $\bar{V}$, a vector of parameters $\theta = (\xi, \lambda, v, H, \sigma_J, \lambda_J, \mu_J)$ and a vector of lags ℓ

$g \leftarrow G(\theta, \ell)$ ▷ Creates a 12×1 vector of theoretical moments

$L \leftarrow \text{Lag}(\bar{V})$ ▷ Use lag function to create lag value

$T \leftarrow \text{length}(\bar{V})$ ▷ Set T to number of elements in $\bar{V}$

$w_{1:T-1} \leftarrow \text{Weight}((1 : T-1)/L)$ ▷ Use weight function to create vector of weights

$\widehat{\Sigma} \leftarrow \frac{1}{T} \sum_{t=1}^{T} \left(\widehat{\mathbb{V}}_t - g\right)^\top \left(\widehat{\mathbb{V}}_t - g\right)$

for i in 1:T-1 **do**

$\widehat{\Gamma}_i \leftarrow \frac{1}{T} \sum_{t=1}^{T-i} \left(\widehat{\mathbb{V}}_t - g\right)^\top \left(\widehat{\mathbb{V}}_{t+i} - g\right)$

$\widehat{\Sigma} \leftarrow \widehat{\Sigma} + w(i) \left[\widehat{\Gamma}_i + \widehat{\Gamma}_i^\top\right]$

end for

return $\widehat{\Sigma}$

A.4 Replication of the two-stage procedure as presented in Gatheral et al. (2022)

The two-stage procedure can be used to estimate the parameters H and v. In this thesis we apply the procedure to determine and set the initial values of H and v (Section 9). The procedure starts out with applying the scaling law (applied on log-volatility increments) formulated in our model setting as

$$\gamma_h = \mathbb{E}[|\log(\sigma^2_{t+h}) - \log(\sigma^2_t)|] = \mathbb{E}[|Y_{t+h} - Y_t|] \to k_q v^q |h|^{qH}, \quad \text{when } h \to 0. \tag{54}$$

Here

$$k_q = 2^{q/2} \frac{\Gamma\left(\frac{q+1}{2}\right)}{\sqrt{\pi}},$$

which is exactly

$$\mathbb{E}[|X|^q] \quad \text{with } X \sim N(0,1).$$

Taking the logarithm on both sides of (54) results in the linear log-log equation

$$\log(\gamma_h) = \log(k_q v^q) + qH \cdot \log(|h|).$$

This can be used to estimate v and H in a linear regression defined as

$$Y = \alpha + \beta X,$$

where we have that

$$\alpha = \log(k_q v^q)$$
$$\beta = qH.$$

Using ordinary least squares (OLS) with $h = 1, 2, 3, 4, 5, 6$ and $q = 2$ we get

$$\widehat{H} = \frac{\widehat{\beta}}{2}$$

and

$$\widehat{v} = \left[\frac{e^{\widehat{\alpha}}}{k_2}\right]^{\frac{1}{2}}$$

with

$$k_2 = \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right).$$

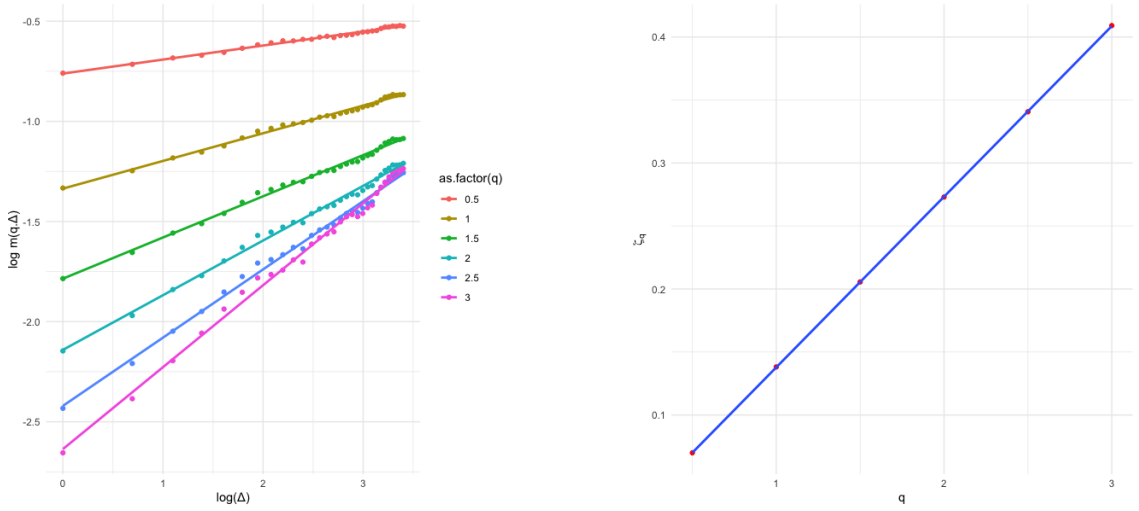
For the values of γ_h we apply

$$\widehat{\gamma}_h = \frac{1}{T-6} \sum_{t=1}^{T-6} |\log(RV^n_{t+h}) - \log(RV^n_t)|^2$$

with RV^n_t as an estimate of the variance on day t. An example of this is provided in figure 8.

A.5 GMM with no jumps in X

Presented without comment.



(a) **(b)**

Figure 8: Linear regressions as in Gatheral et al. (2022) with SPX data. $H = 0.1354$.

Parameter	Value	Realized variance	
		Initial value	Corrected estimate
Panel A:			
ξ	0.0225	0.0240 (0.0104)	0.0199 (0.0072)
λ	0.0050	0.0209 (0.0132)	0.0262 (0.0349)
v	1.2500	0.5411 (0.0150)	0.7904 (0.2073)
H	0.0500	0.2075 (0.0200)	0.1376 (0.0851)
σ_J	0.0000	0.0248 (0.0107)	0.0662 (0.0462)
λ_J	0.0000	0.0291 (0.0042)	0.0379 (0.0324)
μ_J	0.0000	0 (0)	0.0160 (0.0719)
Panel B:			
ξ	0.0225	0.0222 (0.0037)	0.0191 (0.0029)
λ	0.0100	0.0250 (0.0126)	0.0391 (0.0464)
v	0.7500	0.4594 (0.0130)	0.5750 (0.1307)
H	0.1000	0.2187 (0.0203)	0.1887 (0.1050)
σ_J	0.0000	0.0226 (0.0037)	0.0480 (0.0328)
λ_J	0.0000	0.0197 (0.0035)	0.0337 (0.0656)
μ_J	0.0000	0 (0)	0.0010 (0.0708)
Panel C:			
ξ	0.0225	0.0219 (0.0071)	0.0153 (0.0045)
λ	0.0150	0.0221 (0.0092)	0.0448 (0.0501)
v	0.5000	0.4477 (0.0119)	0.4727 (0.0807)
H	0.3000	0.3401 (0.0219)	0.3472 (0.1264)
σ_J	0.0000	0.0222 (0.0072)	0.0407 (0.0370)
λ_J	0.0000	0.0143 (0.0049)	0.1063 (0.1811)
μ_J	0.0000	0 (0)	0.0013 (0.0621)
Panel D:			
ξ	0.0225	0.0215 (0.0068)	0.0154 (0.0048)
λ	0.0350	0.0227 (0.0076)	0.0583 (0.0386)
v	0.3000	0.3547 (0.0089)	0.3254 (0.0475)
H	0.5000	0.3925 (0.0224)	0.4884 (0.1127)
σ_J	0.0000	0.0220 (0.0075)	0.0327 (0.0323)
λ_J	0.0000	0.0133 (0.0047)	0.1077 (0.1798)
μ_J	0.0000	0 (0)	0.0028 (0.0574)
Panel E:			
ξ	0.0225	0.0215 (0.0081)	0.0162 (0.0063)
λ	0.0700	0.0193 (0.0066)	0.0597 (0.0334)
v	0.2000	0.3013 (0.0073)	0.2613 (0.0398)
H	0.7000	0.3820 (0.0230)	0.5273 (0.1024)
σ_J	0.0000	0.0218 (0.0082)	0.0253 (0.0232)
λ_J	0.0000	0.0132 (0.0041)	0.0948 (0.1764)
μ_J	0.0000	0 (0)	-0.0021 (0.0549)

Table 7: Initial values and estimation results of the jump model.

A.6 Figures

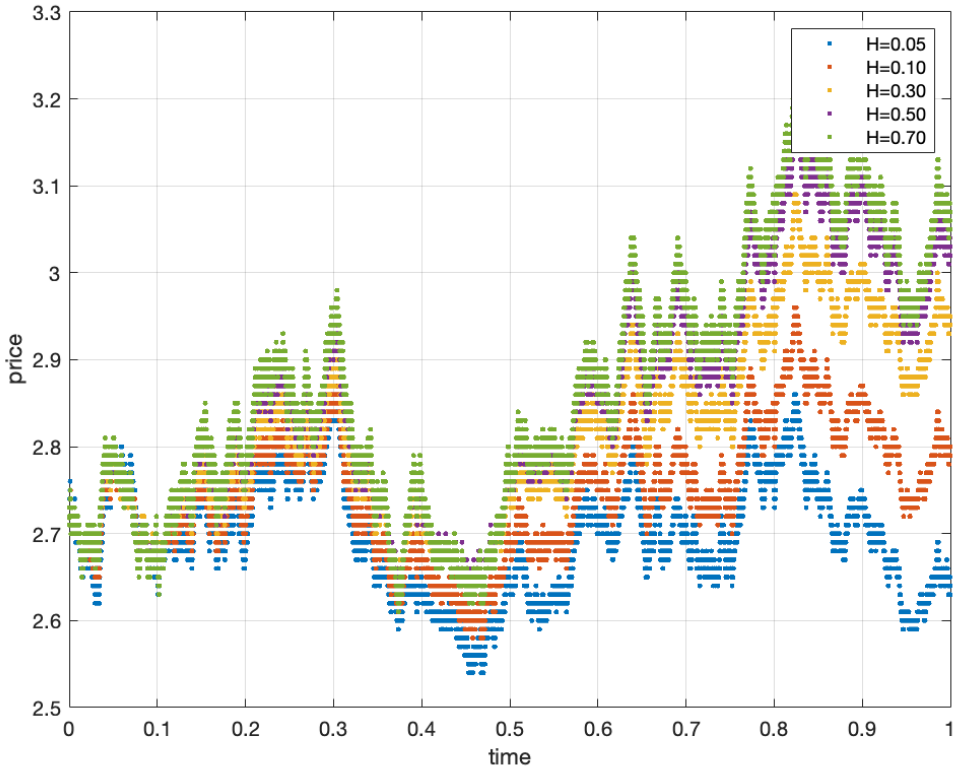


Figure 9: Simulated price for one day with microstructure noise.

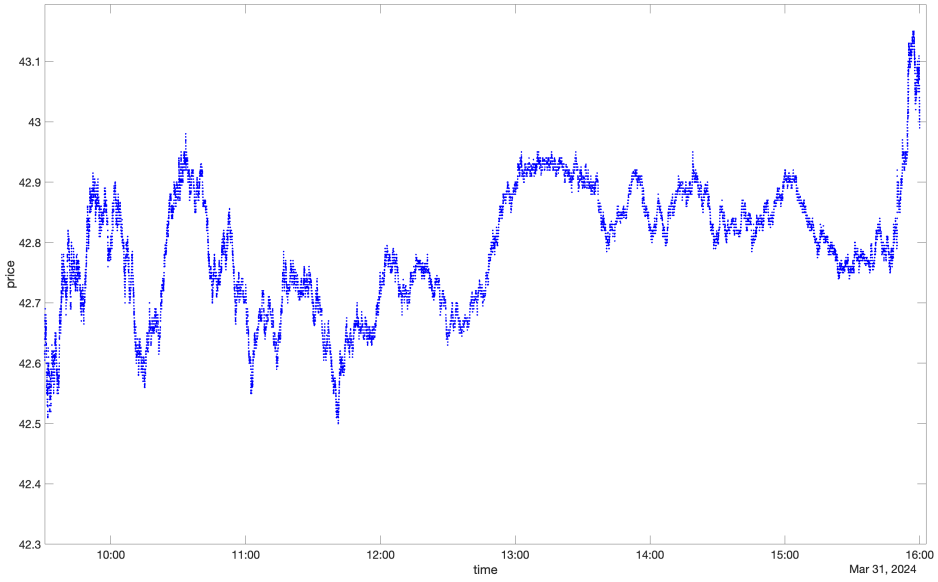
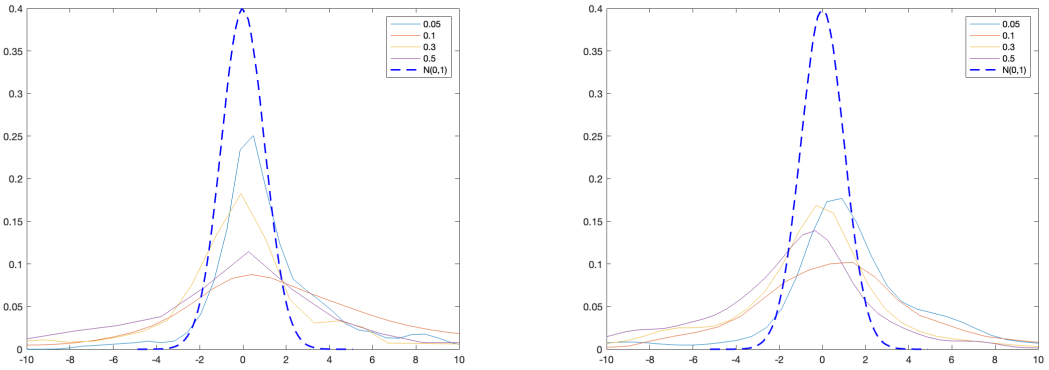
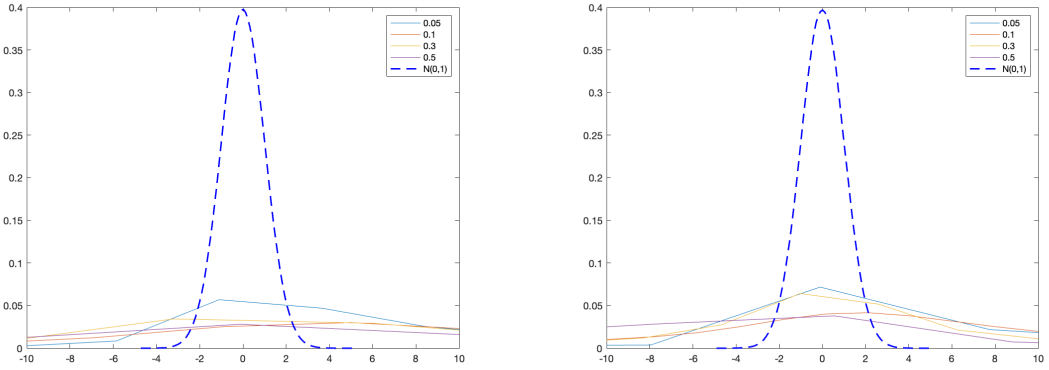


Figure 10: Price of INTC 29/02-2024.



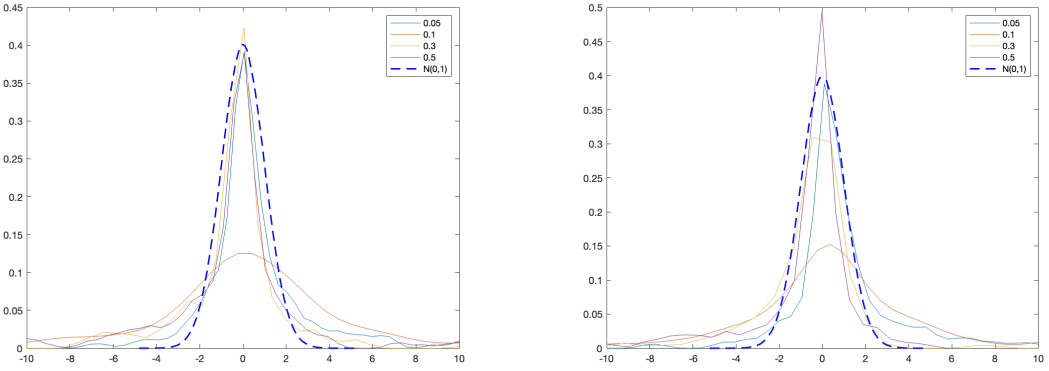
(a) Integrated variance **(b)** Realized variance

Figure 11: Kernel density estimate of standardized H for replication study of Bolko et al. (2023).



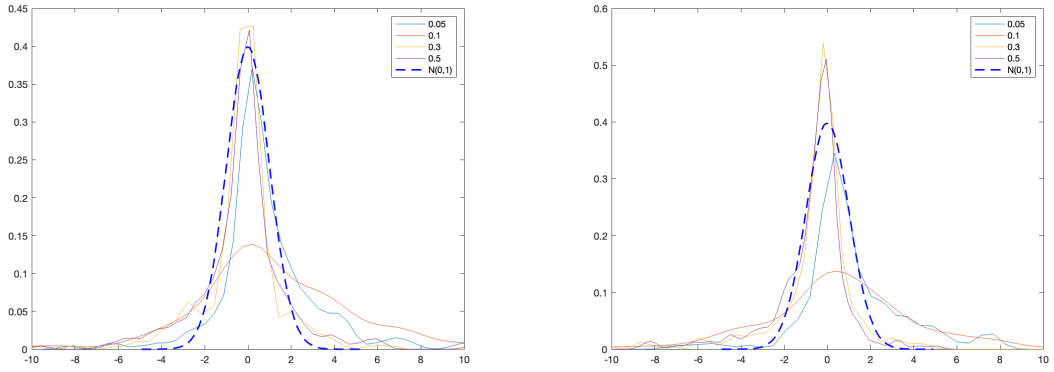
(a) Integrated variance **(b)** Realized variance

Figure 12: Kernel density estimate of standardized H with $T = 500$.



(a) Integrated variance **(b)** Realized variance

Figure 13: Kernel density estimate of standardized H with $\ell = [0, 1, 2, 5, 10]$.



(a) Integrated variance **(b)** Realized variance

Figure 14: Kernel density estimate of standardized H with Barlett kernel.

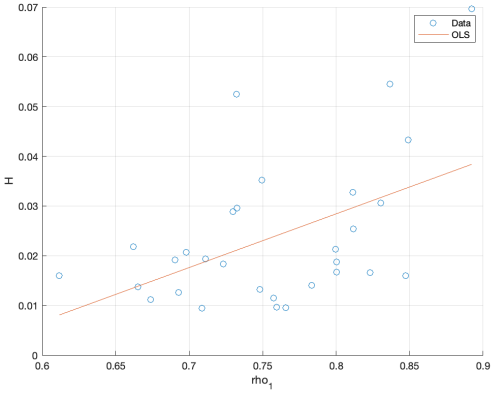
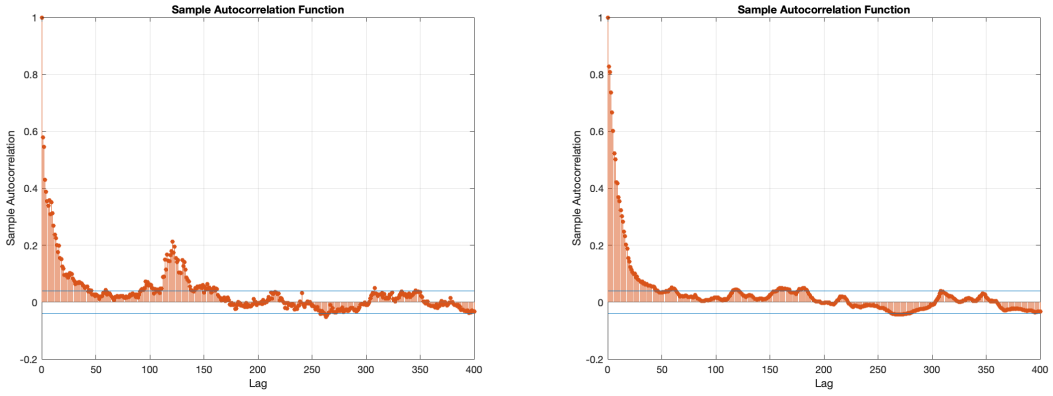
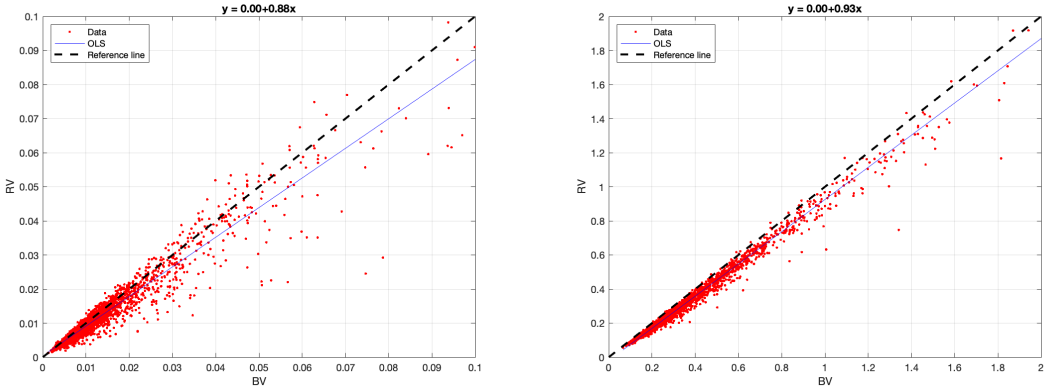


Figure 15: Scatter plot of estimated H versus first-order autocorrelation coefficient ρ_1 fitted with a regression line: $\widehat{H} = \hat{\alpha} + \hat{\beta}\rho_1$.



(a) The sample auto-correlation function of AAPL using realized variance. **(b)** The sample auto-correlation function of AAPL using realized variance.

Figure 16: ACF plot with 95%-confidence band for (a): AAPL (b): DJIA.



(a) **(b)**

Figure 17: BV/RV plot of (a) AAPL (b) DJIA. Plotted with regression line in blue and a 45 degree reference line in black which is a reference point to the null hypothesis of no jumps in X.

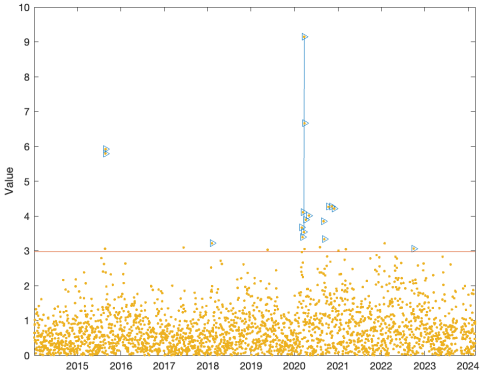


Figure 18: Daily standardized second-order increments of RV with sequential jump location test value on AAPL. Blue markers signal a jump for the given dag and blue lines signals jumps consecutive days. The 5% critical level is marked with a orange line.

A.7 Tables

Parameter	Value	Integrated variance		Realized variance			
					Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:							
ξ	0.0225	0.0251 (0.0109)	0.0207 (0.0071)	0.0248 (0.0107)	0.0203 (0.0074)	0.0210 (0.0075)	0.0212 (0.0077)
λ	0.0050	0.0285 (0.0151)	0.0172 (0.0171)	0.0222 (0.0135)	0.0162 (0.0196)	0.0158 (0.0174)	0.0173 (0.0220)
v	1.2500	0.4562 (0.0124)	1.3261 (0.6884)	0.5238 (0.0148)	2.3533 (1.0458)	1.2186 (0.4439)	1.0256 (0.2305)
H	0.0500	0.2659 (0.0216)	0.0686 (0.0450)	0.2180 (0.0200)	0.0284 (0.0260)	0.0734 (0.0490)	0.0833 (0.0466)
Panel B:							
ξ	0.0225	0.0229 (0.0038)	0.0201 (0.0032)	0.0226 (0.0037)	0.0194 (0.0032)	0.0202 (0.0031)	0.0202 (0.0031)
λ	0.0100	0.0367 (0.0142)	0.0231 (0.0192)	0.0270 (0.0128)	0.0180 (0.0223)	0.0222 (0.0192)	0.0241 (0.0222)
v	0.7500	0.3670 (0.0099)	0.7740 (0.3032)	0.4396 (0.0124)	1.7517 (0.8930)	0.7757 (0.2400)	0.7157 (0.1557)
H	0.1000	0.2993 (0.0220)	0.1128 (0.0482)	0.2332 (0.0203)	0.0400 (0.0537)	0.1149 (0.0599)	0.1232 (0.0587)
Panel C:							
ξ	0.0225	0.0225 (0.0074)	0.0159 (0.0050)	0.0222 (0.0072)	0.0152 (0.0046)	0.0161 (0.0046)	0.0161 (0.0046)
λ	0.0150	0.0342 (0.0110)	0.0361 (0.0437)	0.0243 (0.0095)	0.0185 (0.0263)	0.0303 (0.0241)	0.0377 (0.0598)
v	0.5000	0.3612 (0.0092)	0.5096 (0.0720)	0.4283 (0.0114)	0.8531 (0.4659)	0.5174 (0.0912)	0.5183 (0.0896)
H	0.3000	0.4373 (0.0224)	0.2902 (0.0824)	0.3590 (0.0219)	0.1551 (0.0916)	0.2840 (0.0888)	0.2931 (0.1098)
Panel D:							
ξ	0.0225	0.0221 (0.0070)	0.0156 (0.0049)	0.0218 (0.0069)	0.0140 (0.0044)	0.0158 (0.0047)	0.0158 (0.0047)
λ	0.0350	0.0454 (0.0101)	0.0578 (0.0336)	0.0261 (0.0080)	0.0347 (0.0539)	0.0506 (0.0269)	0.0593 (0.0468)
v	0.3000	0.2475 (0.0061)	0.2995 (0.0427)	0.3321 (0.0084)	0.4687 (0.1662)	0.3236 (0.0514)	0.3252 (0.0498)
H	0.5000	0.5709 (0.0214)	0.4987 (0.0845)	0.4225 (0.0223)	0.2946 (0.1506)	0.4606 (0.0969)	0.4742 (0.1129)
Panel E:							
ξ	0.0225	0.0221 (0.0083)	0.0174 (0.0067)	0.0218 (0.0082)	0.0144 (0.0054)	0.0166 (0.0063)	0.0166 (0.0063)
λ	0.0700	0.0568 (0.0104)	0.0729 (0.0282)	0.0235 (0.0072)	0.0407 (0.0494)	0.0607 (0.0273)	0.0699 (0.0381)
v	0.2000	0.1686 (0.0044)	0.1983 (0.0246)	0.2753 (0.0068)	0.3418 (0.1193)	0.2372 (0.0365)	0.2412 (0.0392)
H	0.7000	0.6772 (0.0194)	0.6416 (0.0696)	0.4222 (0.0232)	0.3665 (0.1546)	0.5513 (0.0998)	0.5651 (0.1078)

Table 8: Initial values and estimation results of replication of simulation study by Bolko et al. (2023).

Parameter	Value	Integrated variance		Realized variance			
					Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:							
ξ	0.0225	0.0275 (0.0200)	0.0213 (0.0140)	0.0272 (0.0198)	0.0212 (0.0140)	0.0218 (0.0145)	0.0219 (0.0146)
λ	0.0050	0.0327 (0.0200)	0.0275 (0.0262)	0.0262 (0.0184)	0.0231 (0.0239)	0.0271 (0.0268)	0.0275 (0.0275)
v	1.2500	0.4559 (0.0168)	0.9830 (0.3394)	0.5240 (0.0186)	1.5060 (0.5012)	1.0799 (0.4012)	0.9547 (0.2463)
H	0.0500	0.2633 (0.0300)	0.1045 (0.0609)	0.2153 (0.0279)	0.0566 (0.0523)	0.0944 (0.0608)	0.1021 (0.0573)
Panel B:							
ξ	0.0225	0.0232 (0.0070)	0.0191 (0.0054)	0.0229 (0.0069)	0.0188 (0.0054)	0.0192 (0.0055)	0.0193 (0.0573)
λ	0.0100	0.0386 (0.0191)	0.0334 (0.0256)	0.0290 (0.0175)	0.0231 (0.0222)	0.0328 (0.0280)	0.0337 (0.0287)
v	0.7500	0.3667 (0.0133)	0.6347 (0.1920)	0.4396 (0.0154)	1.1357 (0.4375)	0.7180 (0.2242)	0.6776 (0.1601)
H	0.1000	0.2965 (0.0305)	0.1600 (0.0683)	0.2306 (0.0285)	0.0739 (0.0563)	0.1381 (0.0719)	0.1435 (0.0688)
Panel C:							
ξ	0.0225	0.0222 (0.0110)	0.0152 (0.0070)	0.0220 (0.0109)	0.0145 (0.0065)	0.0149 (0.0066)	0.0149 (0.0066)
λ	0.0150	0.0375 (0.0147)	0.0399 (0.0255)	0.0272 (0.0131)	0.0245 (0.0252)	0.0391 (0.0297)	0.0393 (0.0297)
v	0.5000	0.3607 (0.0127)	0.4647 (0.0761)	0.4280 (0.0147)	0.6561 (0.1705)	0.5124 (0.0938)	0.5104 (0.0909)
H	0.3000	0.4342 (0.0314)	0.3364 (0.0805)	0.3559 (0.0312)	0.2162 (0.0983)	0.3022 (0.0986)	0.3033 (0.0974)
Panel D:							
ξ	0.0225	0.0215 (0.0093)	0.0158 (0.0063)	0.0212 (0.0092)	0.0144 (0.0059)	0.0148 (0.0060)	0.0148 (0.0060)
λ	0.0350	0.0485 (0.0135)	0.0508 (0.0208)	0.0286 (0.0111)	0.0292 (0.0230)	0.0486 (0.0268)	0.0487 (0.0271)
v	0.3000	0.2470 (0.0087)	0.2815 (0.0418)	0.3318 (0.0111)	0.4006 (0.0902)	0.3372 (0.0526)	0.3371 (0.0520)
H	0.5000	0.5679 (0.0305)	0.5098 (0.0615)	0.4193 (0.0327)	0.3438 (0.0999)	0.4388 (0.0846)	0.4387 (0.0858)
Panel E:							
ξ	0.0225	0.0214 (0.0100)	0.0178 (0.0083)	0.0211 (0.0098)	0.0159 (0.0076)	0.0160 (0.0074)	0.0160 (0.0074)
λ	0.0700	0.0610 (0.0136)	0.0585 (0.0206)	0.0264 (0.0100)	0.0283 (0.0172)	0.0525 (0.0275)	0.0524 (0.0275)
v	0.2000	0.1682 (0.0064)	0.1866 (0.0265)	0.2752 (0.0093)	0.3137 (0.0599)	0.2608 (0.0462)	0.2611 (0.0464)
H	0.7000	0.6747 (0.0281)	0.6181 (0.0641)	0.4192 (0.0343)	0.3789 (0.0801)	0.4926 (0.0918)	0.4920 (0.0914)

Table 9: Initial values and estimation results of robustness check using 500 days per repetition.

Parameter	Value	Integrated variance		Realized variance			
					Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:							
ξ	0.0225	0.0251 (0.0109)	0.0213 (0.0077)	0.0248 (0.0107)	0.0207 (0.0078)	0.0213 (0.0079)	0.0214 (0.0079)
λ	0.0050	0.0285 (0.0151)	0.0134 (0.0121)	0.0222 (0.0135)	0.0120 (0.0123)	0.0131 (0.0123)	0.0133 (0.0123)
v	1.2500	0.4562 (0.0124)	1.1296 (0.3032)	0.5238 (0.0148)	1.7627 (0.4786)	1.2159 (0.4188)	1.0202 (0.2064)
H	0.0500	0.2659 (0.0216)	0.0739 (0.0373)	0.2180 (0.0200)	0.0369 (0.0239)	0.0706 (0.0427)	0.0813 (0.0378)
Panel B:							
ξ	0.0225	0.0229 (0.0038)	0.0205 (0.0032)	0.0226 (0.0037)	0.0197 (0.0031)	0.0203 (0.0031)	0.0203 (0.0031)
λ	0.0100	0.0367 (0.0142)	0.0195 (0.0131)	0.0270 (0.0128)	0.0131 (0.0107)	0.0190 (0.0143)	0.0197 (0.0147)
v	0.7500	0.3670 (0.0099)	0.7137 (0.0131)	0.4396 (0.0124)	1.3887 (0.4352)	0.7714 (0.2103)	0.7157 (0.1429)
H	0.1000	0.2993 (0.0220)	0.1198 (0.0416)	0.2332 (0.0203)	0.0448 (0.0281)	0.1116 (0.0522)	0.1193 (0.0487)
Panel C:							
ξ	0.0225	0.0225 (0.0220)	0.0172 (0.0050)	0.0222 (0.0072)	0.0163 (0.0047)	0.0169 (0.0048)	0.0169 (0.0048)
λ	0.0150	0.0342 (0.0110)	0.0266 (0.0175)	0.0243 (0.0095)	0.0144 (0.0137)	0.0256 (0.0196)	0.0255 (0.0193)
v	0.5000	0.3612 (0.0092)	0.4970 (0.0628)	0.4283 (0.0114)	0.7097 (0.1354)	0.5182 (0.0954)	0.5165 (0.0926)
H	0.3000	0.4373 (0.0224)	0.2926 (0.0632)	0.3590 (0.0219)	0.1809 (0.0656)	0.2801 (0.0812)	0.2804 (0.0795)
Panel D:							
ξ	0.0225	0.0221 (0.0070)	0.0174 (0.0051)	0.0218 (0.0069)	0.0158 (0.0046)	0.0167 (0.0048)	0.0167 (0.0048)
λ	0.0350	0.0454 (0.0101)	0.0483 (0.0192)	0.0261 (0.0080)	0.0226 (0.0156)	0.0465 (0.0235)	0.0464 (0.0234)
v	0.3000	0.2475 (0.0061)	0.2963 (0.0310)	0.3321 (0.0084)	0.4429 (0.0800)	0.3201 (0.0437)	0.3200 (0.0432)
H	0.5000	0.5709 (0.0214)	0.4892 (0.0611)	0.4225 (0.0223)	0.3005 (0.0821)	0.4562 (0.0862)	0.4563 (0.0855)
Panel E:							
ξ	0.0225	0.0221 (0.0083)	0.0186 (0.0069)	0.0218 (0.0082)	0.0162 (0.0060)	0.0175 (0.0065)	0.0175 (0.0065)
λ	0.0700	0.0568 (0.0104)	0.0641 (0.0175)	0.0235 (0.0072)	0.0238 (0.0144)	0.0619 (0.0252)	0.0620 (0.0253)
v	0.2000	0.1686 (0.0044)	0.1956 (0.0173)	0.2753 (0.0068)	0.3295 (0.0603)	0.2309 (0.0281)	0.2309 (0.0280)
H	0.7000	0.6772 (0.0194)	0.6300 (0.0514)	0.4222 (0.0232)	0.3502 (0.0800)	0.5626 (0.0889)	0.5626 (0.0885)

Table 10: Initial values and estimation results of robustness check with Barlett kernel.

Parameter	Value	Integrated variance		Realized variance			
					Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:							
ξ	0.0225	0.0251 (0.0109)	0.0218 (0.0079)	0.0248 (0.0107)	0.0209 (0.0080)	0.0219 (0.0083)	0.0220 (0.0083)
λ	0.0050	0.0285 (0.0151)	0.0140 (0.0128)	0.0222 (0.0135)	0.0150 (0.0143)	0.0142 (0.0140)	0.0145 (0.0144)
v	1.2500	0.4562 (0.0124)	1.2239 (0.4050)	0.5238 (0.0148)	1.8082 (0.4710)	1.2621 (0.4410)	1.0326 (0.2154)
H	0.0500	0.2659 (0.0216)	0.0696 (0.0411)	0.2180 (0.0200)	0.0363 (0.0277)	0.0691 (0.0467)	0.0823 (0.0427)
Panel B:							
ξ	0.0225	0.0229 (0.0038)	0.0211 (0.0033)	0.0226 (0.0037)	0.0201 (0.0035)	0.0209 (0.0467)	0.0210 (0.0033)
λ	0.0100	0.0367 (0.0142)	0.1135 (0.0148)	0.0270 (0.0128)	0.0154 (0.0154)	0.0194 (0.0171)	0.0201 (0.0168)
v	0.7500	0.3670 (0.0099)	0.7607 (0.1775)	0.4396 (0.0124)	1.4811 (0.4728)	0.8060 (0.2612)	0.7353 (0.1488)
H	0.1000	0.2993 (0.0220)	0.1098 (0.0444)	0.2332 (0.0203)	0.0415 (0.0356)	0.1060 (0.0510)	0.1144 (0.0478)
Panel C:							
ξ	0.0225	0.0225 (0.0074)	0.0175 (0.0053)	0.0222 (0.0072)	0.0166 (0.0051)	0.0177 (0.0055)	0.0177 (0.0055)
λ	0.0150	0.0342 (0.0110)	0.0288 (0.0201)	0.0243 (0.0095)	0.0161 (0.0211)	0.0286 (0.0240)	0.0289 (0.0241)
v	0.5000	0.3612 (0.0092)	0.5017 (0.0718)	0.4283 (0.0114)	0.7490 (0.2000)	0.5190 (0.1047)	0.5162 (0.1020)
H	0.3000	0.4373 (0.0224)	0.2892 (0.0718)	0.3590 (0.0219)	0.1722 (0.0994)	0.2820 (0.0984)	0.2839 (0.0971)
Panel D:							
ξ	0.0225	0.0221 (0.0070)	0.0177 (0.0056)	0.0218 (0.0069)	0.0153 (0.0048)	0.0174 (0.0056)	0.0174 (0.0056)
λ	0.0350	0.0454 (0.0101)	0.0497 (0.0056)	0.0261 (0.0080)	0.0254 (0.0270)	0.0509 (0.0290)	0.0507 (0.0056)
v	0.3000	0.2475 (0.0061)	0.2949 (0.0431)	0.3321 (0.0084)	0.4642 (0.1664)	0.3186 (0.0547)	0.3181 (0.0533)
H	0.5000	0.5709 (0.0214)	0.4852 (0.0788)	0.4225 (0.0223)	0.2836 (0.1341)	0.4609 (0.1086)	0.4608 (0.1088)
Panel E:							
ξ	0.0225	0.0221 (0.0083)	0.0186 (0.0071)	0.0218 (0.0082)	0.0154 (0.0060)	0.0177 (0.0065)	0.0177 (0.0065)
λ	0.0700	0.0568 (0.0104)	0.0712 (0.0221)	0.0235 (0.0072)	0.0316 (0.0285)	0.0711 (0.0292)	0.0711 (0.0292)
v	0.2000	0.1686 (0.0044)	0.1943 (0.0206)	0.2753 (0.0068)	0.3348 (0.1158)	0.2241 (0.0401)	0.2241 (0.0400)
H	0.7000	0.6772 (0.0194)	0.6462 (0.0671)	0.4222 (0.0232)	0.3527 (0.1417)	0.5884 (0.1136)	0.5884 (0.1135)

Table 11: Initial values and estimation results of robustness check with lower number of moments.

Parameter	Value	Integrated variance		Realized variance			
					Uncorrected	Corrected estimate	
		Initial value	Estimate	Initial value	estimate	Exact	Approximate
Panel A:							
ξ	0.0225	0.0249 (0.0103)	0.0210 (0.0076)	0.0249 (0.0102)	0.0207 (0.0073)	0.0208 (0.0073)	0.0209 (0.0073)
λ	0.0050	0.0275 (0.0145)	0.0181 (0.0198)	0.0260 (0.0142)	0.0176 (0.0177)	0.0179 (0.0189)	0.0184 (0.0195)
v	1.2500	0.4580 (0.0112)	1.1379 (0.3402)	0.4730 (0.0115)	1.1406 (0.2378)	1.0678 (0.2351)	1.0447 (0.2334)
H	0.0500	0.2650 (0.0199)	0.0784 (0.0472)	0.2536 (0.0201)	0.0733 (0.0403)	0.0822 (0.0455)	0.0859 (0.0541)
Panel B:							
ξ	0.0225	0.0230 (0.0035)	0.0203 (0.0031)	0.0229 (0.0035)	0.0202 (0.0031)	0.0203 (0.0031)	0.0203 (0.0031)
λ	0.0100	0.0355 (0.0135)	0.0229 (0.0172)	0.0332 (0.0133)	0.0203 (0.0183)	0.0229 (0.0185)	0.0240 (0.0200)
v	0.7500	0.3685 (0.0090)	0.7056 (0.1315)	0.3841 (0.0093)	0.8184 (0.1962)	0.7119 (0.1415)	0.7024 (0.1296)
H	0.1000	0.2982 (0.0203)	0.1257 (0.0495)	0.2824 (0.0206)	0.1024 (0.0522)	0.1245 (0.0514)	0.1268 (0.0509)
Panel C:							
ξ	0.0225	0.0229 (0.0069)	0.0163 (0.0046)	0.0229 (0.0068)	0.0165 (0.0047)	0.0165 (0.0047)	0.0165 (0.0047)
λ	0.0150	0.0330 (0.0104)	0.0229 (0.0234)	0.0307 (0.0102)	0.0266 (0.0216)	0.0326 (0.0244)	0.0323 (0.0241)
v	0.5000	0.3626 (0.0090)	0.4883 (0.0665)	0.3767 (0.0092)	0.5276 (0.0744)	0.4953 (0.0686)	0.4953 (0.0686)
H	0.3000	0.4362 (0.0210)	0.3101 (0.0790)	0.4183 (0.0215)	0.2725 (0.0700)	0.3031 (0.0725)	0.3027 (0.0727)
Panel D:							
ξ	0.0225	0.0227 (0.0065)	0.0168 (0.0046)	0.0226 (0.0065)	0.0164 (0.0046)	0.0165 (0.0047)	0.0165 (0.0047)
λ	0.0350	0.0441 (0.0097)	0.0486 (0.0204)	0.0389 (0.0093)	0.0381 (0.0201)	0.0492 (0.0227)	0.0492 (0.0227)
v	0.3000	0.2484 (0.0065)	0.2927 (0.0358)	0.2671 (0.0065)	0.3139 (0.0417)	0.3000 (0.0329)	0.3000 (0.0328)
H	0.5000	0.5698 (0.0203)	0.4990 (0.0652)	0.5315 (0.0213)	0.4464 (0.0815)	0.4904 (0.0723)	0.4905 (0.0723)
Panel E:							
ξ	0.0225	0.0226 (0.0077)	0.0189 (0.0065)	0.0226 (0.0077)	0.0181 (0.0061)	0.0178 (0.0060)	0.0178 (0.0060)
λ	0.0700	0.0557 (0.0100)	0.0608 (0.0194)	0.0389 (0.0094)	0.0444 (0.0157)	0.0582 (0.0202)	0.0582 (0.0202)
v	0.2000	0.1692 (0.0048)	0.1884 (0.0212)	0.1946 (0.0048)	0.2079 (0.0247)	0.2060 (0.0206)	0.2059 (0.0206)
H	0.7000	0.6763 (0.0185)	0.6370 (0.0610)	0.5979 (0.0210)	0.5610 (0.0649)	0.6015 (0.0596)	0.6015 (0.0597)

Table 12: Initial values and estimation results of robustness check with 1 minute sampling.

Symbol	PJ	Jump count	Beta (5y)
AAPL	0.0067	17	1.24
AMGN	0.0036	9	0.61
AMZN	0.0012	3	1.15
AXP	0.0114	29	1.21
BA	0.0126	32	1.58
CAT	0.0047	12	1.09
CRM	0.0024	6	1.30
CSCO	0.0032	8	0.82
CVX	0.0075	19	1.08
DIS	0.0051	13	1.40
DOW	0.0055	14	1.21
GS	0.0020	5	1.35
HD	0.0059	15	1.00
HON	0.0063	16	1.03
IBM	0.0016	4	0.69
INTC	0.0055	14	1.02
JNJ	0.0047	12	0.52
JPM	0.0047	12	1.11
KO	0.0039	10	0.61
MCD	0.0110	28	0.72
MMM	0.0103	26	1.00
MRK	0.0016	4	0.40
MSFT	0.0016	4	0.90
NKE	0.0051	13	1.01
PG	0.0067	17	0.41
TRV	0.0071	18	0.63
UNH	0.0043	11	0.60
V	0.0051	13	0.96
VZ	0.0016	2	0.41
WMT	0.0047	12	0.51
Average	0.0053	13.3333	

Table 13: Proportion of jumps (PJ), jump count for DJIA constituents and 5-year Beta value (source: Yahoo Finance).

Symbol	$\overline{RV}$	ρ_1	GMM Estimate				$\mathcal{J}_{\text{HS}}$
			ξ	$\lambda \times 100$	v	H	
AAPL	0.0796	0.7053	0.0620	1.1669	2.6392	0.0315	0.5895
AMGN	0.0157	0.7128	0.0140	0.6200	1.7413	0.0273	0.7372
AMZN	0.0198	0.7181	0.0194	0.3714	1.3109	0.0524	0.6273
AXP	0.0144	0.8533	0.0121	0.0027	1.6905	0.0237	0.5948
BA	0.0275	0.7601	0.0202	0.0010	1.4021	0.0311	0.6510
CAT	0.0173	0.7805	0.0152	0.2546	1.7135	0.0260	0.5723
CRM	0.0230	0.7467	0.0213	0.1724	1.4421	0.0341	0.8346
CSCO	0.0123	0.7923	0.0103	0.0011	1.7594	0.0124	0.8414
CVX	0.0162	0.8556	0.0130	0.0526	1.3702	0.0377	0.8479
DIS	0.0142	0.8510	0.0113	0.0432	1.6992	0.0234	0.6439
DOW	0.0212	0.7325	0.0167	0.0011	1.3522	0.0215	0.7520
GS	0.0155	0.8714	0.0128	0.8285	1.3192	0.0436	0.3790
HD	0.0134	0.8133	0.0104	0.0027	1.5788	0.0208	0.8471
HON	0.0107	0.7501	0.0080	2.7088	1.8061	0.0354	0.5623
IBM	0.0098	0.8078	0.0084	0.0151	1.7513	0.0221	0.8029
INTC	0.0187	0.8162	0.0164	0.0850	1.6185	0.0251	0.5431
JNJ	0.0084	0.6667	0.0065	0.0018	1.9183	0.0137	0.7146
JPM	0.0144	0.3782	0.0111	0.0010	0.8868	0.0438	0.7634
KO	0.0076	0.7673	0.0059	0.0010	1.6348	0.0163	0.8767
MCD	0.0090	0.8617	0.0064	0.0010	1.7813	0.0166	0.7422
MMM	0.0104	0.8259	0.0086	0.1340	1.7190	0.0252	0.5160
MRK	0.0104	0.8259	0.0086	0.1340	1.7190	0.0252	0.5160
MSFT	0.0140	0.7891	0.0121	0.0250	1.4028	0.0280	0.6985
NKE	0.0147	0.8391	0.0122	0.0382	1.6038	0.0228	0.7889
PG	0.0087	0.8426	0.0063	0.0011	1.7016	0.0147	0.8769
TRV	0.0102	0.8201	0.0077	0.3680	1.8248	0.0214	0.6637
UNH	0.0146	0.8536	0.0115	0.2870	1.2231	0.0558	0.5095
V	0.0123	0.8400	0.0103	0.0010	1.7454	0.0135	0.7847
VZ	0.0095	0.7755	0.0075	0.1523	1.6987	0.0229	0.7267
WMT	0.0091	0.7417	0.0073	0.0281	1.8570	0.0173	0.4215
Average	0.0161	0.6809	0.0131	0.2500	1.6304	0.0268	0.6809
Std	(0.0005)	(0.1259)	(0.0101)	(0.5388)	(0.2963)	(0.0110)	

Table 14: Empirical estimates of model by Bolko et al. (2023) with microstructure robust estimator.

B Generalized method of moments

This section includes an introductory description in a method for estimating the parameters of a time series model. In the continuous SV model setup studied in this thesis, we have 4 parameters to estimate $\theta = (\xi, \lambda, v, H)$, see Sections 2 and 7. In order to estimate these parameters, we will apply the generalized method of moments (GMM). As the name indicates, it is a generalization of the well-known method of moments and can be characterized as an estimating function method. This appendix is general in the sense that many of the assumptions and theorems hold for models which exist outside the fSV domain. As such, when we dive deeper into the fSV model, many assumptions will look slightly different. We will refer back to this appendix to compare the assumptions.

The GMM provides a set of estimates that apply both sample moment conditions and orthogonality conditions (population moment conditions) and match these two types of conditions for the model studied. A GMM estimator is the one that suggests, as an estimate, the element in the parameter space where the sample and population moments are as close to each other according to a defined measure of distance. When there are more population conditions than parameters to estimate, a weighting matrix becomes important and needs to be optimal in an asymptotic sense.

GMM was first described by Hansen (1982), which also gave the statistical theory of the method. In two contemporaneous papers, Hansen & Singleton (1982) and Hansen & Hodrick (1980), the method was proved to be very useful for the empirical analysis of asset pricing and foreign exchange markets. In general, the method has been successfully applied in many different areas of econometrics but mostly in empirical finance (Hall 2013).

This appendix will give a review of GMM, describe the method mathematically, and describe the framework for how it can be used for estimation and inference. It will also include how it can be implemented computationally, as the method is effective (computational-wise) for calculating parameter estimates when implemented with an optimization routine. We apply the GMM framework to the fSV model briefly described in Sections 2 and 7.

It should be noted that this whole appendix assumes that we have a correctly specified model. For GMM estimation of misspecified models, we refer to Hall (2013) chapter 4.

GMM estimates are consistent, and their estimators are asymptotic Gaussian distributed. Consistency and asymptotic normality are the two large sample properties that we will consider in this appendix. GMM allows for estimation of parameters from a partial specified model. We do not need to assume a joint distribution of variables a priori. GMM also has asymptotics that are easy to identify.

This appendix is mainly based on Hall (2013) and the more detailed description in Hall (2005), which is comprehensive with respect to both theory and empirical examples from finance.

Furthermore, the treatment here will mainly focus on GMM applied to time series data, although the method is more general.

The GMM is attractive, compared to the Maximum Likelihood (ML), since it is not necessary to assume any joint probability distribution of the data. ML is the most optimal method when a probability distribution can be assumed to be of the correct form. This is often not the case in SV models, at least not in closed form, which makes ML not so useful. In general, for SV models the distributional features can be quite infeasible for application of ML.

This is the reason why utilizing moment conditions is considerably simpler; they are frequently provided for theoretical finance models.

B.1 Identification

First of all, let us start by defining the notation that we will use in this appendix. Let $\theta \in \Theta$ denote a vector of parameters for a model, e.g. $\theta = (\xi, \lambda, v, H)$ for the continuous SV model in Sections 2 and 7. Note that θ is assigned to be an arbitrary element of Θ. Let θ_0 denote the true values of the parameters, and the objective is to estimate θ_0 from a sample. Let $v_t, t \in \mathbb{Z}$ be a vector of random variables, which in our case will be a time series of volatility measures, and let f be a vector of functions. Assume that θ is k-dimensional and f is q-dimensional with $q \geq k$.

We employ certain assumptions before introducing the population moment conditions:

Assumption 12. *Strict stationarity:* $(v_t)_{t\in\mathbb{Z}}$ *is a strictly stationary process.*

Assumption 13. $f(.)$ *is continuous on* Θ, $\mathbb{E}[f(v_t;\theta)]$ *exists and is finite for every* $\theta \in \Theta$, *and* $\mathbb{E}[f(v_t;\theta)]$ *is continuous on* Θ.

Formally, we need $f(.)$ to be a measurable function.

Assumption 14. *(Hall 2013) The population moment condition is*

$$\mathbb{E}[f(v_t, \theta_0)] = 0, \forall t.$$

If this hold, then we say that θ_0 is identified by the condition

$$\mathbb{E}[f(v_t, \theta_0)] = 0.$$

Given that GMM estimators rely on moment conditions, our emphasis here is on identification based on these moment conditions. In nonlinear models identification may fail because of the properties of: The function $f(.)$ or the data (or due to an interaction between $f(.)$ and the data) (Hall 2005). Because of this, we introduce two concepts, namely global and local identification.

Assumption 15. *Global identification:* $\mathbb{E}[f(v_t, \theta)] \neq 0, \forall \theta \in \Theta \backslash \{\theta_0\}$.

In nonlinear models, it is rarely feasible to derive a convenient condition for global identification, but in the case where failure is caused by the nature of $f(.)$, one could pinpoint the failure by using the condition in Assumption 15. Note that, when the moments are linear in the parameters, then the necessary and sufficient condition for identification is a simple rank condition.

This leads us to formulate a first-order condition for the local identification, which is

$$\text{rank}(G(\theta_0)) = k,$$

where $G(\theta)$ is a $q \times k$ matrix defined by

$$G(\theta) = \mathbb{E}\left[\frac{\partial f(v_t, \theta)}{\partial \theta'}\right],$$

which is the mean of the partial derivatives, with respect to the elements of θ. The partial derivatives should be continuous and exist. It is immediately observed that $q \geq k$ should hold in order to obtain identification.

Let us now define

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(v_t, \theta),$$

which is a $q \times 1$ vector, i.e. the sample moment analogous to the population moment $\mathbb{E}\left[f\left(v_t, \theta\right)\right]$. Here, T is the sample size, e.g. the number of days. Now, the GMM is defined as

Definition 30. *(Hall 2013) The GMM estimate is the value of θ minimizing*

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta).$$

The $q \times q$ matrix W_T is called the weighting matrix. It must be positive semidefinite and fulfill that $W_T \xrightarrow{P} W$. Here, W is a positive definite matrix of constants. Later in this section, we will discuss what the best choice of W_T is. $Q_T(\theta)$ measure the distance from $g_T(\theta)$ to zero and that is the reason why W_T has to be positive semidefinite.

The GMM minimization problem results in the following first-order conditions:

$$G_T(\hat{\theta_T})' W_T g_T(\hat{\theta_T}) = 0, \tag{55}$$

with

$$G_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial f\left(v_t, \theta\right)}{\partial \theta'}.$$

$G_T(\theta)$ is a $q \times k$ matrix, and (55) indicate that $\hat{\theta_T}$ should be calculated as the value of θ that results in the k linear combinations of g_T are zero.

If the number of moment conditions is equal to the length of the parameter vector θ, then it would be possible to set $g_T = 0$ and solve for θ and thus obtain a consistent estimator. However, this may not be feasible if the function f is non-linear in θ, making the analytical solution unavailable. If the moment conditions are less than the length of θ, then θ is non-identifiable. If the moment conditions are larger, then no unique solution exist, but instead we have to choose our estimator for θ so that $Q_T(\theta)$ is as close to zero as possible.

If we have more moment conditions than parameters, $q > k$, then we can conclude that the first order conditions are not the same as solving the sample moment condition. When $q > k$, the estimate $\hat{\theta_T}$ is based on

$$G(\theta_0)' W \mathbb{E}\left[f\left(v_t, \theta_0\right)\right] = 0.$$

For $q > k$ the original moment condition is divided into two parts:

1. Identifying restrictions
2. Overidentifying restrictions

The identifying restrictions are used for estimating θ_0, i.e. associated with $\hat{\theta_T}$, and $g_T(\hat{\theta_T})$, the estimated sample moments, are a function of the overidentifying restrictions. The overidentifying restrictions are used for inference regarding the validity of the model.

B.2 Weighting matrix

An important feature of the GMM estimator is the weighting matrix. In general, the GMM estimator will depend on W_T, so how W_T is chosen is an important issue.

For consistency of our estimator, we need an optimal choice of our weighting matrix, which we allow to have a possible dependence on the sample data. We can think of the optimal choice of the weighting matrix as a sequence of weighting matrices that will converge to a matrix of constants as $T \to \infty$. We use our weighting matrices together with our population moment conditions to get a sample objective function that is minimized. We consider the weighting matrix as the key metric that ensures that the sample moment conditions are minimized to be as close to zero as possible.

Standard properties of the weighting matrix, see Hall (2005):

Assumption 16. *W_T is a positive semidefinite matrix which converges in probability to the positive definite matrix of constants W.*

In the case where $q = k$, the estimator is determined by solving

$$g_T(\hat{\theta_T}) = 0,$$

which is standard method of moments, and imply that the estimator is independent of the weighting matrix and therefore (Hall 2013)

$$V = [G(\theta_0)' S(\theta_0)^{-1} G(\theta_0)]^{-1} = \lim_{T\to\infty} Var(\sqrt{T}(\widehat{\theta}_T - \theta_0)),$$

where $S(\theta) = \lim_{T\to\infty} Var(\sqrt{T} g_T(\theta))$.

In the case of $q > k$, the estimate $\hat{\theta_T}$ depends on W_T through the order conditions in equation (55). This can be seen as unattractive, but the asymptotics show that only the covariance matrix V is dependent on W_T. In Hansen (1982) it is proved that by setting

$$W_T = \widehat{S}_T^{-1}$$

with $\widehat{S}_T$ being a consistent estimator of $S(\theta_0)$, then V is minimized.

Now, the main question is how to choose $\hat{S_T}$ so that it is a consistent estimator of $S(\theta_0)$. If we define

$$\Gamma_i(\theta_0) = Cov\left(f(v_t, \theta_0), f(v_{t-i}, \theta_0)\right)$$

then it can be proved that (Andrews 1991):

$$S(\theta_0) = \Gamma_0(\theta_0) + \sum_{i=1}^{\infty} (\Gamma_i(\theta_0) + \Gamma_i(\theta_0)').$$

It can be shown that the class of heteroscedasticity autocorrelation covariance (HAC) estimators

$$\hat{S}_{HAC,T} = \hat{\Gamma_0} + \sum_{i=1}^{T-1} w(i; b_T)(\hat{\Gamma_i} + \hat{\Gamma_i}') \tag{56}$$

with

$$\hat{\Gamma_j} = \frac{1}{T} \sum_{t=j+1}^{T} \hat{f}_t \cdot \hat{f}_{t-j}, \quad \hat{f}_t = f(v_t, \hat{\theta_T})$$

can estimate the variance V in the long run. In equation (56), w is a kernel and b_t is the bandwidth of the kernel. The kernel and bandwidth can ensure consistency and the positive semidefinite condition.

B.3 Consistency and asymptotic normality

The treatment here follow Hall (2013). In order to allow the use of Laws of Large Numbers as well as the Central Limit Theorem, a number of aassumption/restrictions on (v_t) and f has to be listed.

The first asymptotic result is the consistency of the GMM estimator. If the GMM estimator is not consistent it would not be adequate to use GMM for estimation since the estimate is not guaranteed to be close to the true parameter value as $T \to \infty$. Additional to the assymptions already mentioned earlier in this section (Assumtions 12-16) we need a few more.

Assumption 17. *The process v_t is ergodic.*

Assumption 18. *The parameter space Θ is compact.*

It is possible to relax the assumption of compact parameter space, which might be justified in practice, and for a discussion of this, see Newey & McFadden (1994). We need to assume that $f(v_t, \theta)$ is bounded by a function with finite expectation i.e.

Assumption 19. $\mathbb{E}\left[sup_{\theta\in\Theta} ||f(v_t, \theta)||\right] < \infty$.

The consistency of $\widehat{\theta}_T$ can be formulated as

Theorem 11. *Suppose Assumption 12-19 hold, then*

$$\widehat{\theta}_T \xrightarrow{P} \theta_0, \quad as\ T \to \infty. \tag{57}$$

Next, we will look at the asymptotic distribution of g_T. For this, we need one additional assumption. We need properties for the sample moment variance.

Assumption 20. *(i) $\mathbb{E}[f(v_t, \theta_0) \cdot f(v_t, \theta_0)']$ exist and is finite; (ii) $S = \lim_{T\to\infty} \mathrm{Var}\left(\sqrt{T} g_T(\theta_0)\right)$ exist and is a finite valued positive definite matrix.*

Then we have the theorem

Theorem 12. *If Assumptions 12, 14, 17,and 20 hold, then $\sqrt{T} g_T(\theta_0) \xrightarrow{D} N(0, S), \quad T \to \infty$.*

Finally, we will discuss the asymptotic distribution of the parameter estimator $\widehat{\theta}_T$. Again, additional assumptions must be made.

Assumption 21. *(i) The derivative matrix $\partial f(v_t, \theta)/\partial\theta'$ exists and is continuous on Θ for v_t; (ii) θ_0 is an interior point of Θ; (iii) $E\left[\partial f(v_t, \theta_0)/\partial\theta'\right]$ exists and is finite.*

Assumption 22. *(Hall 2013) $\mathbb{E}[\partial f(v_t, \theta)/\partial\theta']$ is continuous in some neighborhood N_ϵ of θ_0.*

Assumption 23. *(Hall 2013) $\sup_{\theta\in N_\epsilon} \|G_T(\theta) - E\left[\partial f(v_t, \theta)/\partial\theta'\right]\| \xrightarrow{P} 0$*

Then first of all we have:

Lemma 1. *If Assumptions 12-19 and 22-23 hold then $G_T\left(\hat{\theta}_T\right) \xrightarrow{P} G_0$*

with

$$G_0 = \mathbb{E}[\partial f(v_t, \theta_0)/\partial\theta']$$

and secondly

Theorem 13. *If Assumptions 12-23 hold, then $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{D} N(0, MSM')$ where $M = \left(G_0' W G_0\right)^{-1} G_0' W$.*

B.4 Testing over-identifying restrictions

Next, we will introduce a test for a correctly specified model. For this to be the case the equation

$$\mathbb{E}[f(v_t, \theta_0)] = 0 \quad \forall t,$$

contains useful information regarding θ_0. For practical use, we can translate this to the question whether the observed data are consistent with the population moment condition. As noted above, if $q = k$, then $g_T(\widehat{\theta}_T) = 0$, so we cannot test for a correctly specified model in this case using $g_T(\widehat{\theta}_T)$. When $q > k$ we have that

$$g_T(\widehat{\theta}_T) \neq 0.$$

This is because only the identifying restrictions are used to estimate θ_0 (the number of identifying restrictions are k). The rest of $(q - k)$ overidentifying restrictions are true, if

$$\mathbb{E}[f(v_t, \theta_0)] = 0.$$

So applying GMM when $q > k$, we note that

1. The distance from $g_T(\widehat{\theta}_T)$ to zero is measured by $Q_T(\widehat{\theta}_T)$.
2. The estimate $g_T(\widehat{\theta}_T)$ of the sample moments will have information on the $q-k$ overidentifying restrictions.

A model diagnostic regarding a correctly specified model can be formed as a χ-test in the following way with the statistic

$$J_T = T g_T(\widehat{\theta}_T)' \widehat{S}_T^{-1} g_T(\widehat{\theta}_T).$$

The statistic is asymptotically χ^2_{q-k} distributed, and is the sample size times an expression which is the expression GMM needs to be minimized and evaluated in $\widehat{\theta}_T$. The expression is used in the numerical optimization method described next.

B.5 Numerical optimization

Analytically $\hat{\theta}$ is not often found, and instead we have to use numerical methods.

By Definition 30, the GMM estimator of θ_0 is

$$\widehat{\theta}_T = \underset{\theta \in \Theta}{\arg\min} Q_T(\theta). \tag{58}$$

First order conditions for minimization require $\partial Q_T\left(\widehat{\theta}_T\right)/\partial\theta' = 0$. In nonlinear models, it is mostly impossible to obtain a closed-form solution. Instead, we can use numerical methods to calculate $\hat{\theta}_T$. This is done by trial and error, where a computer-based routine tries to find the value of θ that minimizes $Q_T(\theta)$. The procedure starts by taking an initial value of θ, say $\bar{\theta}_0$, which is then used as a benchmark to find the next value of θ that satisfies $Q_T(\bar{\theta}_1) < Q_T(\bar{\theta}_0)$ (which is only possible if $\bar{\theta}_0$ is not the value that minimizes Q_T). After finding $\bar{\theta}_1$, the routine is then to find the value that further lowers the value of $Q_T(\theta)$ and this updating process continues until the value of θ converges to a minimum which satisfies a convergence criterion. Different routines do different iterative searches. In this thesis we use gradient based routines. Let the value of θ at the i'th step be:

$$\bar{\theta}_i = \bar{\theta}_{i-1} + \lambda_i D(\bar{\theta}_{i-1}),$$

where λ_i is the step size and $D(.)$ is the step direction which is a function of the gradient, $\partial Q_T(\bar{\theta}_{i-1})/\partial\theta'$. λ_i determines how far to go in the direction.

Many iterative minimizing/maximizing routines may only converge to a local minimum, whereas consistency results require a global minimum to be found. It is therefore important to find good starting values. One might also start the optimizer from multiple starting values and pick the best $\hat{\theta}$.