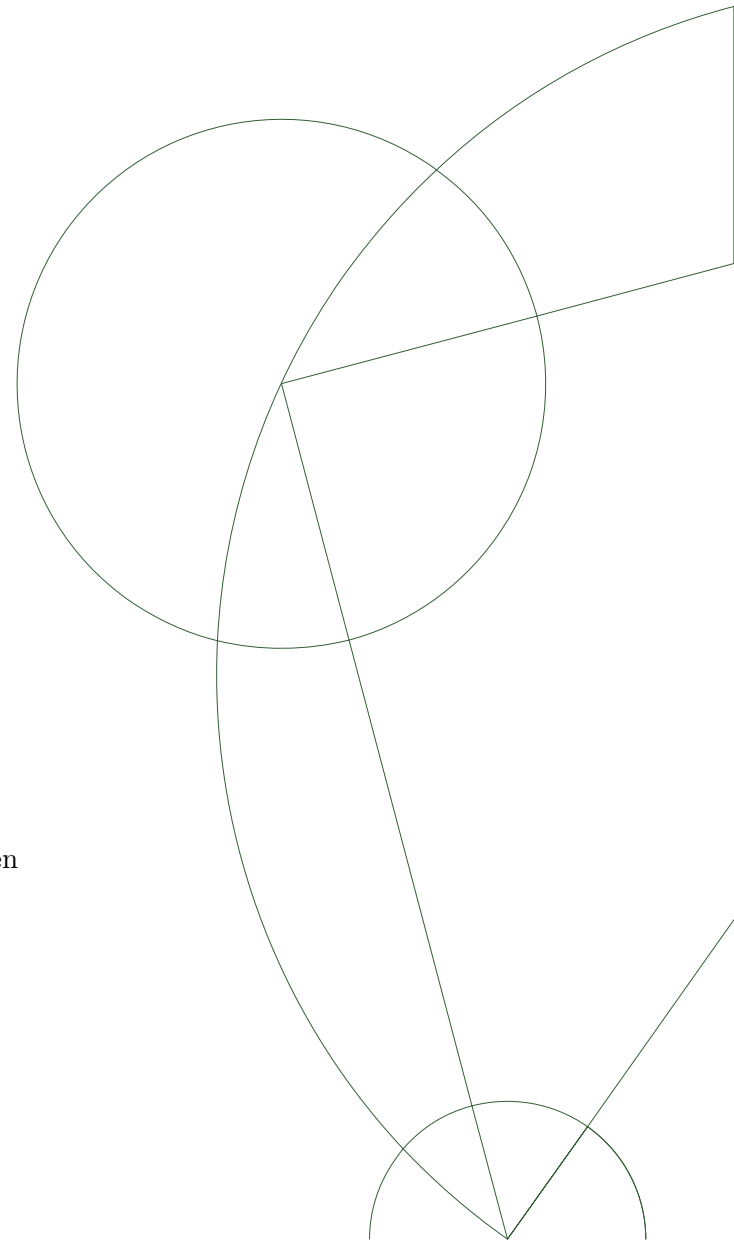
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Master's Thesis in Mathematics-Economics

Sebastian G. Pedersen
Mathias Wilken
Frederik Skjøtt

Investigation of the Burst Hypothesis and Intraday Volatility Behaviour



Supervisor: Rolf Poulsen, University of Copenhagen

Submitted on: 6 August 2018

Abstract

This thesis expands the theory and mathematical foundation behind the existence of short-lived explosive trends in the financial markets, split into the explosion of the volatility and explosion of drift as first described in Christensen et al. (2016). We add finite sample corrections and extra parameter tuning to the estimators, allowing us to improve the detection of periods with extremely high volatility and possibly drift. With fast computational algorithms, we create thorough simulation studies showing that these periods cannot be described by the standard Heston model with a jump component. However, we also provide evidence that conclusions drawn from the Heston model does not necessarily extend to financial markets, particularly S&P 500. To remedy this, a thorough study of volatility is conducted. The empirical behaviour of volatility is investigated and volatility is modelled utilizing Brownian semi-stationary processes as introduced in Barndorff-Nielsen & Schmiegel (2009) in an attempt to capture both roughness and persistence along with intraday seasonality.

Ultimately, the simulation study is extended to a model featuring a Brownian semi-stationary process, which strengthens the hypothesis of the existence of the short-lived explosive periods as even such models fall short of explaining these extreme events, denoted as *bursts*. This allows us to show the presence of bursts on S&P 500 but question whether drift bursts are present or the bursts are in fact all explained by an explosion in the volatility. We also apply the theory on Bitcoin and justify the existence of both volatility bursts and drift bursts in this market, though high-frequency frictions on the exchanges make these conclusions uncertain.

Contents

1 Introduction[1]

In recent times, the availability of high frequency data in financial markets allows for detailed examination and modelling of intraday behaviour. This has resulted in the literature expanding on the shortcomings of otherwise well established models such as the Heston model. The focus in this thesis will be to examine these shortcomings with outset in the common setting of an Itô drift-diffusion process with an added jump component.

Adding a jump and a discontinuity to the Ito-process has been the predominant way of creating a process that would be able to describe large sudden changes in the price process. These are often associated with events that changes the demand for the underlying asset. Such events could be the news of a successful or failed drug trial, annual reports etc.. Another example of such an event is the announcements from the United States Federal Reserve, which we will examine in greater depths later in this thesis. In Figure 1 we see the price process of the SPY, which tracks S&P 500, on the 19th of March 2014. On an daily view the process appears to jump around 15:00.

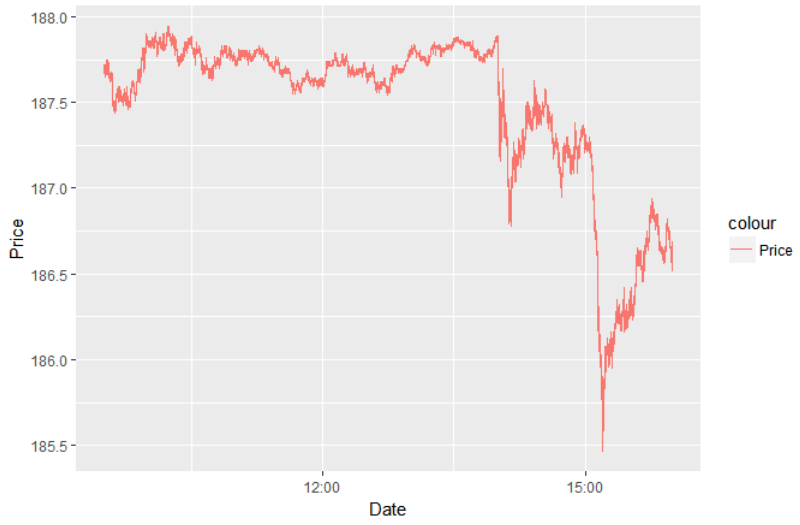


Figure 1: Price process of SPY on 19th of March 2014 following the Federal Reserve's announcement of increased interest rates.

When a jump component is added to the stochastic process, it appears to mimic the jump-like movement seen in Figure 1, and the jump models may seem very appropriate at describing the processes.
However, up-close the price process looks nothing like a jump. Figure 2 portrays an intraday close-up of the price process at the point of the supposed *jump*. Instead, it looks more like an Ito-process with a strong downwards trend.

Although it may be tolerable to assume that the price around 15:00 follows an Ito-process with a jump component when studying the process on larger time scales, it is not when studying on an intraday level. In this thesis, we will show that the price process of S&P 500 contains very rapid movements like these, which do not seem to be described by neither Heston dynamics nor jumps.

Christensen et al. (2016) formulate an alternative to jumps as a possible explanation to the sudden intraday price movement seen in Figure 2. Christensen et al. (2016) presents the *Drift Burst Hypothesis*, which postulates that the price processes instead of jumps contains *short-lived locally explosive trends*. These trends can be described as changes in the dirft term and/or the diffusion term in an Itô process - respectively termed

[1]Everyone is responsible for this section and every section where nothing else is stated.

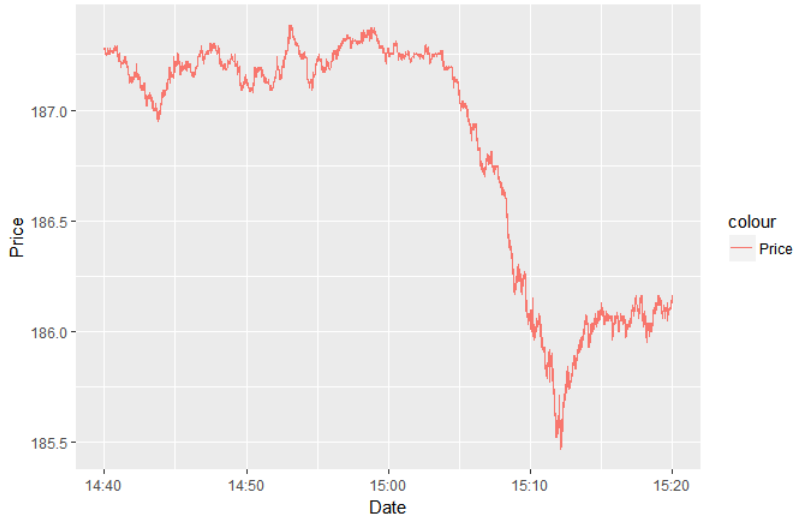


Figure 2: Intraday view of the price process at 15:00 on 19th of March 2014.

drift bursts and volatility bursts. In order to support this hypothesis, Christensen et al. (2016) create an estimator to identify explosive trends that cannot be explained by an Itô-process with jumps.

In this thesis we show new convergence result of the estimator formulated in Christensen et al. (2016), both in in presence of bursts and in the presence of microstructure noise. These are both shown theoretically and validated in simulation studies. We also create finite sample corrections and construct algorithms for fast calculations of the estimators, without which the simulations could not have been carried out in manageable time. Also, importantly, the issues related to distinguishing between drift burst and/or volatility bursts is heavily discussed.

As will be discovered later, accurate modelling of intraday volatility is detrimental when attempting to evaluate the results of the developed estimators. Recently, there has been development in the literature regarding intraday volatility estimation. In particular, the focus on rough volatility models has increased and several models have been proposed.

In Bennedsen et al. (2016) it is suggested to utilize a so-called Brownian semistationary process to describe the volatility process. In accordance with the methodology presented therein, we examine the empirical behavior of volatility in an attempt to quantify potentially important features of volatility - roughness and persistence of the process of interest. Examples of Brownian semistationary processes are fitted to data, and their predictive power is evaluated, which further allows for the importance of roughness and persistence.

Ultimately, we are able to interpret the results of applying the burst estimator on real data. First on the S&P-500 index, we find a burst every third day, but also that the existence of drift bursts is questionable, as they seem to be overshadowed by the volatility bursts. We find bursts more frequently than previously discovered by Christensen et al. (2016). This is most likely due to tuning and finite sample corrections of estimators, which to our knowledge has not been done previously.

We also apply the estimator on data on three major Bitcoin exchanges. Here we find significantly more bursts than in the S&P-500 market and where the existence of drift bursts in S&P-500 was questionably, drift bursts clearly seems to be present in the Bitcoin market.

1.1 Structure

The composition of the thesis is as follows.

Section 2: A description of the data used in the thesis. We use SPY trades data for 2014 and Bitcoin quotes from three different Bitcoin Exchanges from 26th of March 2018 and ends at 13th of June 2018.

Section 3: The introduction of the common notation and assumptions that is used throughout the thesis.

Section 4: We present the *Drift Burst hypothesis* that is described in Christensen et al. (2016), and introduce burst and jump components to the Heston price dynamics.

Section 5: Creation of the T-estimator used for detecting bursts in the absence of microstructure noise. This section introduces our choice of kernel, and forms the basis of the burst detection. We examine the convergence results of the estimator as the number of observations increases.

Section 6: The microstructure noise in data disrupts the quality of the estimators that were used in the section without noise. In this section, we develop new estimators that are much more robust to the microstructure noise. This includes a walk-through of pre-average methods and the kernel-based approach. We study the convergence of the T-estimator with the presence of microstructure noise.

Section 7: We examine the properties and behaviour of the T-estimator when a certain degree of bias from the microstructure noise is preserved in the estimation. We show that allowing some bias in exchange for better convergence might be the best approach in regards to detection of bursts.

Section 8: We create a quantile to evaluate the T-estimator against, such that we overcome the multiple testing problem and can control the amount of false positive burst detection.

Section 9: Through simulation studies, we investigate the optimal choice of bandwidth of the T-estimator such that the detection is as good as possible. Furthermore we conduct simulation studies that allow us to draw conclusions on the fraction of false-detections that is expected to occur in data.

Section 10: The theory and the simulation studies have all been based on equidistant observations. Here we prepare ourselves for the estimation on data by examining the performance of the T-estimator in a setting without equidistant observations. Additionally, we conclude that the intraday seasonality in the number of trades has an insignificant impact on the performance of the T-estimator.

Section 11: We optimize the computational calculations of the estimators such that simulation and data studies can be carried out in manageable time. We exploit our choice of kernel to rewrite the estimators on a recursive form that speeds up the estimation remarkably.

Section 12: The burst detection is carried out in practice on the SPY and Bitcoin data. We identify bursts and interpret the results. We discover that the intraday behavior of SPY is significantly different from the simulation study, and that this effects the T-estimator in a way, that might invalidate the conclusions. This leads to a necessity for a better intraday volatility model.

Section 13: We introduce the necessary framework for investigating the volatility.

Section 14: The volatility of data is examined. We quantify the intraday seasonality of volatility and estimate persistence and roughness of the deseasonalized log volatility. The findings leads us to the introduction of a model that encompasses both roughness and the potential for long term memory.

Section 15: The introduction of the $\mathcal{BSS}$-framework, specifically the two the Gamma-$\mathcal{BSS}$ and the Power-$\mathcal{BSS}$ model. The properties of these models are stated, and methods for fitting the parameters of the models are stated.

Section 16: The $\mathcal{BSS}$-models are fitted to data and their predictive power is examined in a volatility prediction scheme. The results are analyzed in an attempt to further enlighten the extend of the models ability to describe the observed volatility.

Section 17: We return to the burst analysis an behavior of T-estimator in the $\mathcal{BSS}$-framework. We find that the distribution of T-estimator under the $\mathcal{BSS}$-model imitates the behavior on SPY extremely well. The findings in these renewed studies show that the results in Section 12 are reliable and conclusions in fact can be drawn from these. Additionally, we achieve a better estimate of the fraction of false-positives that we should see in data.

Section 18: In this section we follow along the lines of Christensen et al. (2012) to evaluate the performance of the BV^*-estimator during various bursts to determine, whether the estimation of the integrated variance is reliable in presence of bursts.

Section 19: We discuss various interesting topics that has been left out. These include a discussion of the reversion study carried out in Christensen et al. (2016), an analysis of Bitcoin specific effects that affect our burst detection. Furthermore, we briefly discuss a possible altercation to the burst estimation that could compensate for seasonality.

Section 20: We conclude on our findings throughout the thesis.

2 Data description

The data we have available is tick-by-tick trades of SPDR S&P 500 ETF TRUST (SPY) for 2014 and Bitcoin from March 2018 to June 2018 on three major European bitcoin exchanges.

2.1 SPDR S&P 500 ETF

The SPDR S&P 500 ETF (SPY) is an Exchange-traded-fund that tracks the Standard and Poor's 500 (S&P 500) and is one of the largest ETFs in the world.
The SPY data consists of 33,117,626 trades across 249 trading days from 2nd of January 2014 to 31st of December 2014. Each trading day starts at 09:30:00 and ends at 16:00:00.

Table 1: SPDR S&P 500

Price primo 2014	$ 183.98
Price ultimo 2014	$ 205.52
Yearly return (2014)	11.07 %
Highest (2014)	$ 208.97
Lowest (2014)	$ 173.71
Trades (2014)	33,117,626

A very evident feature in the data is that the intraday trades are not evenly distributed throughout the trading day. To plot the distribution throughout the day, we create 5-minute *buckets* and average the number of trades within each bucket across all the trading days. This means that the bucket at 09:35 includes all the trades within the first five minutes of each day - 09:00 to 09:35.
The amount of trades in the beginning and end of the day is remarkably higher than that of the mid-day. This can be the result of several factors. One of the reasons behind this could be due to overnight news. We can only trade the asset 6.5 hours per trading day, but the rest of the world does not stand still outside of this window. Events that occur would accumulate up until the markets open. This phenomenon would also explain the large amount of trades during the final minutes of the day. Many agents on the market may wish to hedge or close their positions at the end of the day to avoid the risk that arises from above scenario, resulting in a large amount of trades at the end of the day as well. This feature of the data has to be taken

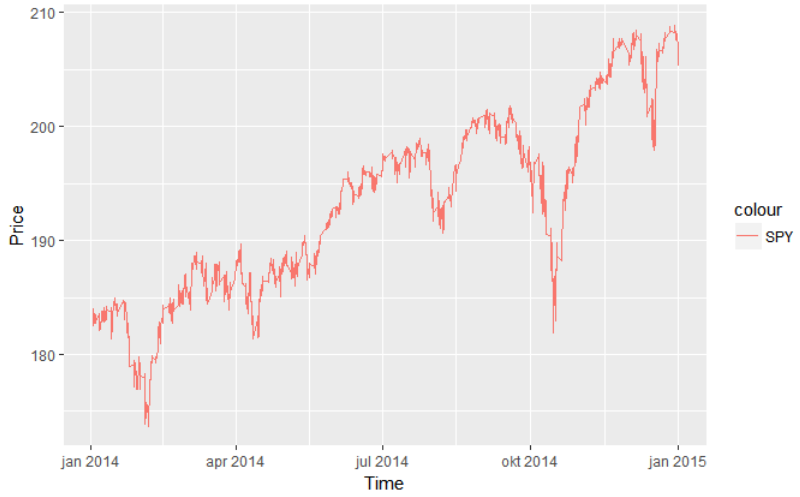


Figure 3: The price of the SDPR S&P 500 ETF throughout 2014.

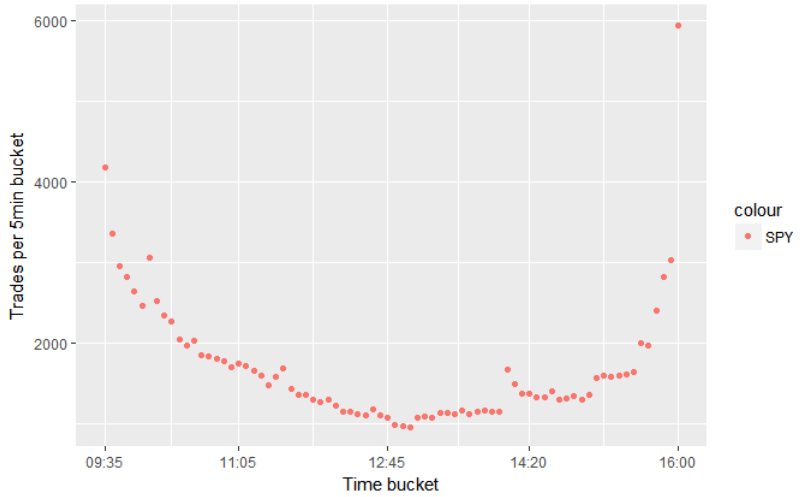


Figure 4: Distribution of trades throughout the day

into account in the subsequent modelling.

2.2 Bitcoin

Bitcoin is the largest cryptocurrency with a total market cap of $ 146.5 billion at the beginning of the 26th of Marhc 2018[2]. At the beginning of the thesis we began extracting high frequency data on quotes of three major Bitcoin exchanges - Bitfinex, BitMEX and Kraken.
Our tick-by-tick data starts at the 26th of March 2018 and ends at 13th of June 2018 and a summary of the data can be seen in Table 2.

Interestingly, the prices are not always equivalent across the exchanges. This may seem like an obvious arbitrage, but transferring Bitcoin from one exchange to another is costly and not instantaneous. Besides,

[2]https://coinmarketcap.com/currencies/bitcoin/

Table 2: Bitcoin 2018

	Bitfinex	BitMEX	Kraken
Price at beginning	$ 8445.15	$ 8444.25	$ 8449.95
Price at end	$ 6559.25	$ 6265.1	$ 6562.4
Return (Interval)	-25.3 %	-25.1 %	-25.3%
Highest (Interval)	$ 9989.95	$ 9969.25	$ 9947.7
Lowest (Interval)	$ 6430.05	$ 6405.5	$ 6265.1
Trades (Interval)	6,186,183	8,094,341	3,345,264

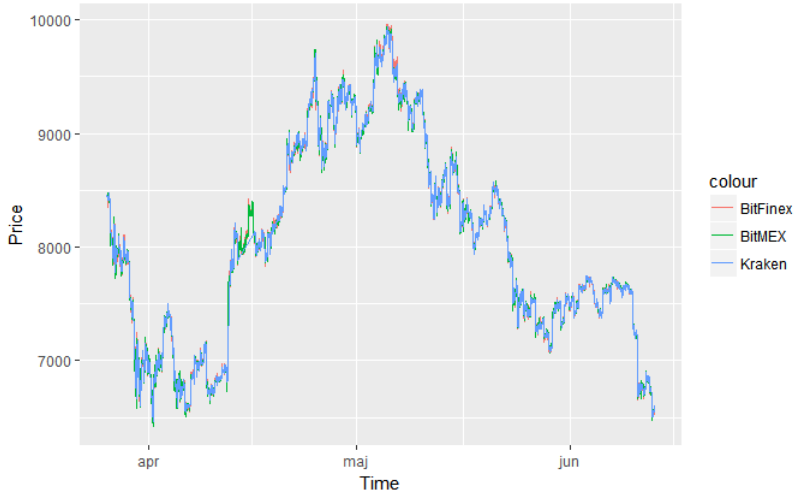


Figure 5: Bitcoin from 26th of March 2018 to 13th of June 2018

different exchanges have different spreads and broker fees such that a smaller difference between the platforms is possible without introducing arbitrages. Lastly, the credit risk, security levels and country regulations may vary greatly from exchange to exchange and in this market these are very important risk factors when trading on crypto exchanges.

From Figure 5 displaying the Bitcoin price development throughout the 80 days of data, we can see that the process is very volatile.

The quotes on Bitfinex and BitMEX are collected with a subscription to the WebSocket API on the exchange. This means that they push every quote update to our server, and we save the mid-quote at every quote update. The mean number of quote-updates during the day can be seen in Figure 6. We see that these are quite evenly distributed throughout the day, and we do not find any intraday seasonality in these. Note however that this does not necessarily mean that there is no seasonality in the number of trades.
The cryptoexchange Kraken does not have a WebSocket API. As a result of this, the data is collected through a REST API with an HTTP request. This means that while we received every quote update from the two other exchanges, we receive the quotes on Kraken when we requested for it. Because of a rate limit on the exchange, we requested a new quote every 2nd second and saved the mid-quote. The quote-updates on this exchange are therefore very evenly distributed throughout the day with approximately 150 updates per 5min interval as seen in Figure 6.

The different methods used on Kraken and the two others has the important consequence, that the observations are not equidistant on Bitfinex and BitMEX, whereas there are close to 2 seconds between every observation on Kraken. How this affects the study is investigated later.

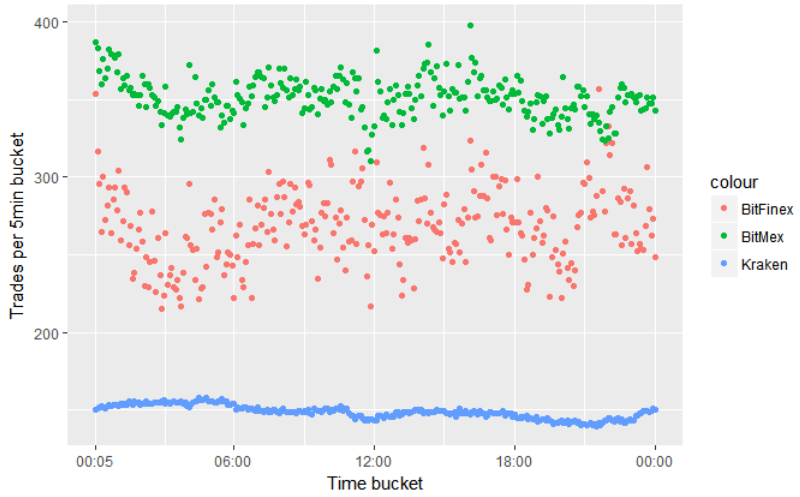


Figure 6: Distribution of quote updates across the day for the three exchanges

3 Notation and assumptions

In the following, we will be working with one or more stochastic processes. These will all be defined on the filtered probability space $\left(\Omega, \mathbb{F}, (\mathbb{F})_{t\geq 0}, P\right)$. Time is denoted t for which it is assumed that $0 \leq t \leq T$ for some $T > 0$. Furthermore, the filtered probability space is $P-$complete and the filtration is right-continuous to ensure conditions in referenced papers are satisfied.

We denote the time points where the log-price is observed as $0 = t_0 < t_1 < ... < t_n = T$.

For a process X sampled at the time points above, we will often work with the *delta* sequence defined as

$$\Delta X_{i,n} = X_{t_i} - X_{t_{i-1}}.$$

If X_t is an adapted process then this is naturally also an adapted process.

4 The bursts hypothesis

We will work with an Itô process X, which we assume has a stochastic differential given by

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t \tag{1}$$

where $(\mu_t)_{t\geq 0}$ is a predictable process, $(\sigma_t)_{t\geq 0}$ is an adapted strictly positive càdlàg process, and $(W_t)_{t\geq 0}$ is a standard Brownian motion.
dJ_t is a jump component where $dJ_{\tau_J} = J$ is a random variable representing a jump with size J at time $t = \tau_J$. We assume that there is a finite number of jumps.

For the Itô processes, we usually have

$$\int_{t-\bar{\Delta}}^{t} \mu_s ds = O_p(\bar{\Delta}) \quad \text{and} \quad \int_{t-\bar{\Delta}}^{t} \sigma_s dW_s = O_p(\sqrt{\bar{\Delta}})$$

with $0 < \bar{\Delta} < t$. This is the reason why the drift is often neglected in the literature, as the volatility prevails when looking at short intervals.
The hypothesis first formulated in Christensen et al. (2016) is that the drift and volatility can explode over short time periods such that

$$\int_{t-\bar{\Delta}}^{t} \mu_s ds = O_p(\bar{\Delta}^{1-\alpha}) \quad \text{and/or} \quad \int_{t-\bar{\Delta}}^{t} \sigma_s dW_s = O_p(\bar{\Delta}^{1/2-\beta}) \tag{2}$$

with $0 < \alpha < 1$ and $0 < \beta < 1/2$ for some limited period of time.

This can be achieved by adding a drift and volatility as an extra component to the Itô process such that the stochastic differential equations instead are given by

$$dX_t = (\mu_t + \tilde{\mu}_t)dt + (\sigma_t dW_{1t} + \tilde{\sigma}_t dW_{2t}) + dJ_t \tag{3}$$

where $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ at one or several time points follow the convergence rates from (2).

In this thesis, we adopt the canonical example from Christensen et al. (2016) and assume that these components are parameterized by

$$\tilde{\mu}_t = \frac{c_1}{(\tau - t)^\alpha} \quad \text{and} \quad \tilde{\sigma}_t = \frac{c_2}{(\tau - t)^\beta} \quad \text{for} \quad t \in]0, \tau[.$$

With this parameterization the equations in (2) are satisfied at time point τ, while the total drift and variation remains finite.

An example of a plausible situation for a volatility burst is right up to an event such as an annual report of a company or on larger scale a statement from FED regarding new interest rates. The volatility increases gradually up to these known events, and this could be incorporated into the Itô process for example as $\tilde{\sigma}_t$ with τ as event time.
A reason for a drift burst could be liquidity crisis with forced liquidation or very short periods with arbitrage. When the market agents cannot react instantaneously to the news - either due to limitations or skill - the price will move quickly but not instantaneously.

The purpose of the following chapters is to create an estimator that can distinguish periods of bursts from periods where X is solely an Itô process with jumps as given by the dynamics in (1). It will then be possible to determine whether the existence of bursts in the financial markets are plausible or not. First the estimator is created in absence of noise, thereafter it will be theoretically expanded to handle microstructure noise. After this it is optimized for best practical use.

5 The estimator in absence of noise[3]

In this section we create the estimator to distinguish between periods with bursts and periods where X follows an Itô-process with jumps in the framework of absence of noise. Here we follow the same procedure as in Christensen et al. (2016), though with a different convergence result in presence of a burst along with important additional considerations.

[3]Responsible: Sebastian, Frederik

We wish to create an estimator to detect drift bursts, and thus we need to be able to estimate the drift of the process X. An obvious option is to estimate it as the mean drift of the process, i.e.

$$\hat{\mu}_{\text{naive}} = \frac{X_n - X_0}{T} = \frac{1}{T}\sum_{i=1}^{n} \Delta X_{i,n}.$$

which estimates the integrated drift $\int_0^T \mu_t dt$. This is however not desirable as we wish to detect drift burst on a much more granular level. For that we want the estimator to estimate the instantaneous drift. This can be achieved with a kernel that weights the latest returns higher than older. Hence we let the drift estimator be given as

$$\hat{\mu}_t^n = \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right)\Delta X_{i,n} \tag{4}$$

where K is a left sided kernel satisfying suitable regularity conditions (Assumption 3 in Christensen et al. (2016)), and h_n is the bandwidth in the kernel.

This estimator is simply an average of the process ΔX weighted with the kernel K. If we let the kernel be increasing for negative values of x and zero for positive, i.e.

$$\begin{aligned} &K(x) > 0 \text{ for } x \leq 0 \\ &K(x) = 0 \text{ for } x > 0 \\ &K'(x) > 0 \text{ for } x < 0 \end{aligned}$$

the newest observations will have the most weight as desired, and because the estimator only relies on past and present observations, we are able to use the estimator in *real time*. Throughout this thesis, we will use the following kernel:

$$K(x) = \begin{cases} \exp(x) & \text{for } x \leq 0 \\ 0 & \text{for } x > 0. \end{cases}$$

This choice of kernel is also the one used in Christensen et al. (2016) and satisfies both the necessary assumptions and desirable conditions above. In addition to this, it will later be shown that the specific properties of the exponential kernel allow us to speed up the calculations in practice significantly.

Figure 7 shows how the weights of the $\Delta X_{i,n}$ changes as the bandwidth decreases. Note that these are indeed weights, as the total weight is approximately the same as in the naive estimator, i.e.

$$\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right) \approx \frac{1}{T}\sum_{i=1}^{n} = \frac{n}{T}$$

which comes from the fact, that

$$\sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right)\frac{T}{n \cdot h_n} \to \int_{-\infty}^{0} K(x)dx = 1$$

by Riemann approximation of the integral (if we have $h_n n \to \infty$). We see that as the bandwidth h_n decreases, the newest observations is assigned more and more weight.

We make the conventions

$$K_2 = \int_{-\infty}^{0} K(x)^2 dx$$

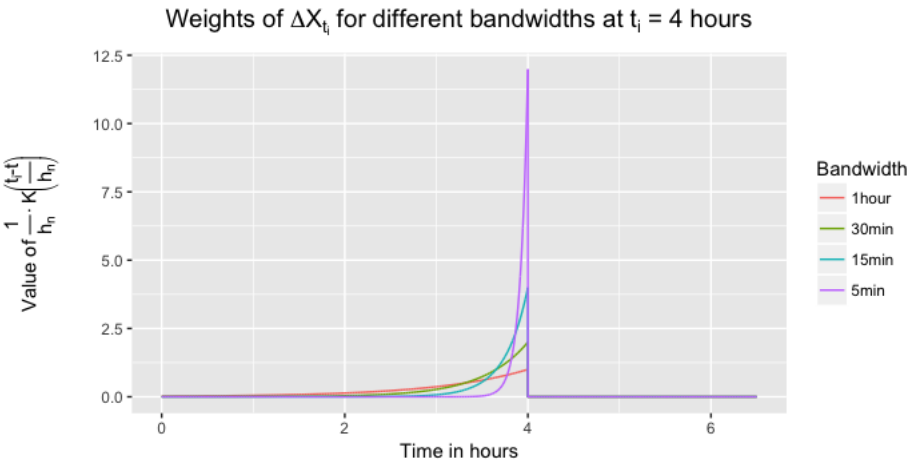


Figure 7: An illustration of the weight $\frac{1}{h_n}K\left(\frac{t_i-t}{h_n}\right)$ of every ΔX_{t_i} at for the drift estimator at time point $t = 4$hours. The weight is showed for four different bandwidth lengths.

$$\mu_{t-} = \lim_{h\to 0^+} \mu_{t-h} \quad \text{and} \quad \sigma_{t-} = \lim_{h\to 0^+} \sigma_{t-h}$$

and have the following distribution result for the drift-estimator:

Theorem 1. *If X follows the dynamics in (1) with no jumps, it holds that*

$$\sqrt{h_n}(\hat{\mu}_t^n - \mu_{t-}) \xrightarrow{d} N(0, K_2\sigma_{t-}^2) \tag{5}$$

as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$. The convergence is stable in law.

Proof. See Appendix A in Christensen et al. (2016). For an elaboration of stably convergence in law and the importance of this, see e.g. Section 3.2.1 in Barndorff-Nielsen et al. (2008). □

The theorem above shows that $\hat{\mu}_t^n$ is asymptotically an estimator for μ_{t-}. The idea is that we can use this result to make the following approximation:

$$\sqrt{h_n}\hat{\mu}_{t-} \approx N\left(\sqrt{h_n}\mu_{t-}, K_2\sigma_{t-}^2\right).$$

Under normal circumstances it would hold that $\sqrt{h_n}\mu_{t-} \to 0$ as $h_n \to 0$, however at a drift burst one could hope that this term might instead explode if μ_{t-} is *sufficiently large*. Unfortunately the estimator includes the unknown parameter σ_{t-} in the variance. In order to introduce useful concepts such as quantiles and confidence intervals, we would need an estimator to quantify σ_{t-}. As we need to estimate the instantaneous volatility, we include the left-sided kernel just as in the drift estimator. We construct the following estimator

$$(\hat{\sigma}_t^n)^2 = \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right)(\Delta X_{i,n})^2. \tag{6}$$

Theorem 2. *With $(\hat{\sigma}_t^n)^2$ defined as in (6), we have in absence of jumps*

$$(\hat{\sigma}_t^n)^2 \xrightarrow{p} \sigma_{t-}^2 \tag{7}$$

as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$.

Proof. See Theorem 2.2 in Mattiussi & Renò (2009). □

It is now possible to define an estimator with a known mean and variance, and we define

$$T_t^n = \sqrt{\frac{h_n}{K_2}} \frac{\hat{\mu}_t^n}{\hat{\sigma}_t^n}. \tag{8}$$

Theorem 3. *If X follows the dynamics in (1), then it at a time without jumps it holds that*

$$T_t^n \xrightarrow{d} N(0,1) \tag{9}$$

as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$.

Proof. Since the convergence of the drift-estimator in Theorem 1 is stable in law, and the volatility-estimator in Theorem 2 converges in probability, the result follows from the equivalence of Slutsky's theorem for stable convergence. Note that the convergence in law of the drift estimator is necessary as σ_{t-} is a stochastic variable. □

The theorem above shows that we have created an estimator which is normally distributed if the process X follows an Itô-process. In the following sections we will investigate the distribution and behavior of this estimator when the underlying process differs from an Itô-process.

5.1 In presence of jumps

In presence of jumps, the drift estimator diverges at jump time. This is quite easy to see as we can write the estimator at jump time as

$$\hat{\mu}_\tau^n = \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_i - \tau}{h_n}\right) (\Delta X_{i,n} + 1_{\{t_i = \tau\}} J_\tau) = \frac{1}{h_n} J_\tau + \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_i - \tau}{h_n}\right) \Delta X_{i,n}.$$

and as $h_n \to 0$ we assign more and more weight to the newest change in the process as seen in Figure 7.

The volatility estimator also diverges in presence of jumps as

$$(\hat{\sigma}_\tau^n)^2 = \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_i - \tau}{h_n}\right) (\Delta X_{i,n} + 1_{\{t_i = \tau\}} J_\tau)^2 = \frac{1}{h_n} J_\tau^2 + \frac{2}{h_n} J_\tau \Delta X_\tau + \frac{1}{h_n} \sum_{i=1}^n K\left(\frac{t_i - \tau}{h_n}\right) (\Delta X_{i,n})^2$$

and as $h_n \to 0$ the term $\frac{1}{h_n} J_\tau^2$ will dominate.

These will however cancel out in the T-estimator from (8) such that the following theorem holds at jump time:

Theorem 4. *Let X follow the dynamics in (1) and let τ define a time point where $dJ_\tau = J$ with $J \neq 0$. We then have*

$$T_\tau^n \xrightarrow{p} \sqrt{\frac{K(0)}{K_2}} \cdot sign(J) \tag{10}$$

as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$.

Proof. The result follows from Appendix A in Christensen et al. (2016). □

For the exponential kernel $K_2 = \frac{1}{2}$ and $K(0) = 1$ which means that $T_\tau^n \xrightarrow{p} \pm\sqrt{2}$.

5.1.1 Note on jumps

Though it has not been done in previous work, it is important to investigate both how a jump's impact on the T-estimator depends on the jump size and how it affects the T-estimator post-jump. From the intuition in last section and formal proof of Theorem 4, it is apparent that the distribution of the estimator at jump time $t = \tau$ can be written as

$$T_\tau = \frac{\sqrt{h_n}\hat{\mu}_\tau^n}{\sqrt{K_2}\hat{\sigma}_\tau^n} = \frac{1}{\sqrt{K_2}} \frac{\frac{1}{\sqrt{h_n}}J + N(\sqrt{h_n}\hat{\mu}_t, K_2\sigma_{t-}^2) + o_p(1)}{\sqrt{\frac{1}{h_n}J(J + 2\Delta X_{\tau,n}) + \sigma_{t-}^2 + o_p(1)}}.$$

We see that for $h_n \to \infty$, the term $\frac{1}{\sqrt{h_n}}J$ dominates in the numerator, and the term $\frac{1}{h_n}J^2$ in the denominator. From this follows, that we for every jump-size J have in probability convergence towards $\pm\sqrt{2}$ just as stated in Theorem 4. However it also holds that for every h_n and constant k_1, there exists a jump J such that $\frac{1}{\sqrt{h_n}}J = k_1$. We can then approximate the equation above as

$$T_\tau \approx \frac{\frac{1}{\sqrt{K_2}}k_1 + N(0, \sigma_{t-}^2)}{\sqrt{k_1^2 + \sigma_{t-}^2}} = \frac{1}{\sqrt{k_1^2 + \sigma_{t-}^2}} N\left(\frac{1}{\sqrt{K_2}}k_1, \sigma_{t-}^2\right).$$

For a given jump-size J, $k_1 \to \infty$ as $h_n \to 0$, and the approximation above converges in probability to the expression from Theorem 4. At the same time for a very small jump, k_1 is close to zero and the approximation is a standard normal distribution which is the distribution in absence of a jump.

So for very large jumps, the term is close to $\pm\sqrt{2}$ and for very small jumps, the estimator is closer to $N(0, 1)$ distribution. However there exists jump-sizes, where the term is a mixture. Since we do not want our estimator to diverge in the presence of jumps, it is vital that we ensure that the estimator is bounded - no matter the size of the jump.

Proposition 1. *Let T follow the distribution*

$$T \sim \frac{1}{\sqrt{k^2 + \sigma^2}} N(c \cdot k, \sigma^2)$$

then for all $k > 0$, we have

$$P(T > q) \leq P(X > \sqrt{q^2 - c^2}) \tag{11}$$

where $X \sim N(0, 1)$. Furthermore there exists a k where it holds with equality.

Proof. See appendix. □

We now have a bound on the percentile of the T-estimator for every jump size. The theorem also states that this is the best bound achievable.

This bound can also be used to bound the estimator post-jump. The reason is that after a jump the estimator can be written (omitting the error terms) as

$$T_{\tau+t} \approx \frac{K\left(\frac{-t}{h_n}\right)\frac{1}{\sqrt{K_2}}k_1 + N(0, \sigma_{t-}^2)}{\sqrt{K\left(\frac{-t}{h_n}\right)k_1^2 + \sigma_{t-}^2}} = \frac{1}{\sqrt{\left(\sqrt{K\left(\frac{-t}{h_n}\right)}k_1\right)^2 + \sigma_{t-}^2}} N\left(\frac{\sqrt{K\left(\frac{-t}{h_n}\right)}}{\sqrt{K_2}} \cdot \left(\sqrt{K\left(\frac{-t}{h_n}\right)}\ k_1\right), \sigma_{t-}^2\right).$$

Rewriting $k = \sqrt{K(\frac{-t}{h_n})}k_1$, we see that Proposition 1 can be used to bound the percentile post-jump as well. Since c is smaller post-jump due to the fact that $\sqrt{K(\frac{-t}{h_n})} < 1$, the bound is even tighter. Hence the bound still holds.The influence of this result on the T-estimator will be investigated in simulation studies in later sections.

5.2 In presence of drift- and volatility bursts[4]

We now investigate the behaviour of the T-estimator at the peak of a drift and/or volatility burst for the process X when it follows the dynamics in (3).

Theorem 5. *Assume that X follows the dynamics in eq. (3) with $0 < \beta < 1/2 < \alpha < 1$. Assume that there is a drift- and volatility burst with a peak at time τ_b. It then holds that*

$$|T_{\tau_b}| \xrightarrow{p} \infty \qquad \text{for } \beta + 1/2 < \alpha \tag{12}$$

$$T_{\tau_b} \xrightarrow{d} N(0, 2^{2\beta}) \qquad \text{for } \beta + 1/2 > \alpha \tag{13}$$

as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$.

Proof. See Appendix 1. □

This result is to our knowledge a new result. It shows that if the drift burst is significantly stronger than the volatility burst then estimator will diverge. However if the volatility is too strong, it will dominate the drift, and the convergence will be quite similar to the normal case where the volatility prevails.

The intuition behind the necessity of $\alpha > 1/2$ comes from the result in Theorem 1 that

$$\sqrt{h_n}\hat{\mu}_t^n \xrightarrow{d} N(\sqrt{h_n}\mu_{t-}, K_2\sigma_{t-}^2).$$

For this to diverge it is necessary that $\sqrt{h_n}\mu_{t-} \to \infty$, which is only the case if the drift diverges faster than $1/\sqrt{h_n}$. This happens if exactly $\alpha > 1/2$. The further condition that it is necessary that $\alpha > 1/2 + \beta$ to ensure the divergence of the entire T-estimator comes from the fact, that the volatility estimator in the denominator dominates over the drift otherwise, and the mean would then converge to zero.

We have now constructed an estimator, which only diverges in the presence of a strong drift burst. In the case of the presence of only one type of burst, we have the following corollary:

Corollary 1. *Assume that X follows the dynamics in eq. (3), and we have either a volatility burst or a drift burst with peak at time τ_b but not both. Then the convergence is as in Theorem 5 where we let $\beta = 0$ in case of absence of volatility burst and $\alpha = 0$ in the case of absence of drift burst.*

Proof. The proof follows exactly as the proof in Theorem 5 with the absence of one of the bursts. □

It is now shown that the estimator can only diverge in the case of drift bursts, and we seem to have a valid estimator for testing the existence of drift bursts in absence of microstructure noise.

[4]Responsible: Sebastian

5.3 Notes on arbitrage

We have shown that in presence of drift- and volatility bursts, the drift burst would prevail over the variance and the T-estimator would therefore diverge if $\alpha > \beta + 1/2$.
As showed in Christensen et al. (2016), this coincides with the presence of a free lunch with vanishing risk as defined in Definition 10.6 in Björk (2009). This estimator is therefore asymptotically only capable of detecting drift bursts in case of arbitrage. We will return to this later.

5.4 Simulation study

To test the convergence/divergence of the estimator, we let the price process X follow the dynamics of a Heston-model without rate or dividend - i.e. the dynamics of $\log(X)$ is given by

$$dX_t = c\sigma_t^2 dt + \sigma_t dW_t$$
$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \xi\sigma_t dB_t$$

where W and B are Brownian motions with correlation ρ. We choose the annualized parameters to be given by $(\theta, \kappa, \xi, c, \rho) = (0.0457, 5.07, 0.48, 1.11, -0.767)$. These parameters are the values estimated by Aït-Sahalia & Kimmel (2007) on data from S&P-500 index from 1990-2003.

We simulate one trading day corresponding to 6.5 hours and let $t_0 = 0$ and $T = 6.5$ hours.

5.4.1 Simulation of Heston dynamics[5]

The Heston stock dynamics under the risk-neutral Q-measure are given by:

$$dS = \sigma_t S_t dB_t^s$$
$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \xi\sigma_t dB_t^\sigma$$
$$E(dB_t^s \; dB_t^\sigma) = \rho dt$$

The log dynamics are found using Ito's lemma:

$$ds_t = d\log(S_t) = \frac{1}{S_t}\left(\sigma_t S_t dB_t^s\right) + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)\left(\sigma_t S_t dB_t^s\right)$$
$$= -\frac{1}{2}\sigma_t^2 dt + \sigma_t dB_t^s$$

We can rewrite the two correlated brownian motions using two independent brownian motions:

$$\begin{bmatrix} dB_t^s \\ dB_t^\sigma \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{bmatrix} \begin{bmatrix} dW_t^s \\ dW_t^\sigma \end{bmatrix}, \qquad E(dW_t^s \; dW_t^\sigma) = 0$$

Thus we can rewrite the dynamics using the above:

$$ds = -\frac{1}{2}\sigma_t^2 dt + \sigma_t\left[\sqrt{1-\rho^2}\rho dW_t^s + dW_t^\sigma\right]$$
$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \xi\sigma_t dW_t^\sigma$$

The purpose of our simulations is to simulate realistic stock dynamics and not to price derivatives. This means that we are interested in the P-dynamics. For this we use the Girsanov theorem which yields the following relation between the Q and P measure:

$$d\bar{W}_t^s = \phi_t^s dt + dW_t^s$$
$$d\bar{W}_t^\sigma = \phi_t^\sigma dt + dW_t^\sigma$$

[5]Responsible: Frederik

Where $d\bar{W}_t^s$ and $d\bar{W}_t^s$ denote the brownian motion under the P-dynamics. We define the Girsanov kernels, ϕ_t^s and ϕ_t^σ, as:

$$\phi_t^s = \lambda^s\sqrt{1-\rho^2}\sigma_t, \quad \phi_t^\sigma = \lambda^\sigma\sigma_t$$

The final Heston log-dynamic under the P-measure is given by:

$$\begin{aligned} ds &= -\frac{1}{2}\sigma_t^2 dt + \sigma_t\left[\sqrt{1-\rho^2}\left(d\bar{W}_t^s - \phi_t^s dt\right) + \rho\left(d\bar{W}_t^\sigma - \phi_t^\sigma dt\right)\right] \\ &= -\frac{1}{2}\sigma_t^2 dt + \sigma_t\left[\sqrt{1-\rho^2}\left(d\bar{W}_t^s - \lambda^s\sqrt{1-\rho^2}\sigma_t dt\right) + \rho\left(d\bar{W}_t^\sigma - \lambda^\sigma\sigma_t dt\right)\right] \\ &= C\sigma_t^2 dt + \sigma_t\left[\rho d\bar{W}_t^s + \sqrt{1-\rho^2}d\bar{W}_t^\sigma\right], \quad C = -\frac{1}{2} - (1-\rho^2)\lambda^s - \rho\lambda^\sigma \end{aligned}$$

$$\begin{aligned} d\sigma_t^2 &= \kappa(\theta - \sigma_t^2)dt + \xi\sigma_t\left(d\bar{W}_t^\sigma - \phi_t^\sigma dt\right) = \kappa(\theta-\sigma_t^2)dt + \xi\sigma_t\left(d\bar{W}_t^\sigma - \lambda^\sigma\sigma_t dt\right) \\ &= \kappa^*(\theta^* - \sigma_t^2)dt + \xi\sigma_t d\bar{W}_t^\sigma \end{aligned}$$

Where $\kappa^* = \kappa - \xi\lambda^\sigma,\ \theta^* = \left(\frac{\kappa^* + \xi\lambda^\sigma}{\kappa^*}\right)\theta$.
The log-price process can then be simulated using the following scheme:

$$\begin{aligned} s_{t+1} &= s_t + C\sigma_t^2 dt + \sigma_t\sqrt{dt}\left(\rho Z_t^\sigma + \sqrt{1-\rho^2}Z_t^s\right), && Z_t^s \sim N(0,1) \\ \sigma_{t+1}^2 &= \sigma_t^2 + \kappa^*\left(\theta^* - \sigma_t^2\right)dt + \sigma_t\sqrt{dt}Z_t^\sigma, && Z_t^\sigma \sim N(0,1) \end{aligned}$$

In practice, we may risk that the squared volatility, σ^2, becomes negative, and our simulation breaks down as we square root the squared volatility to obtain the volatility, σ. In order to avoid this, we adjust the volatility process by taking the absolute value - this is also known as reflection:

$$\hat{\sigma}_{t+1}^2 = |\hat{\sigma}_t^2 + \kappa^*\left(\theta^* - \hat{\sigma}_t^2\right)dt + \hat{\sigma}_t\sqrt{dt}Z_t^\sigma)|, \qquad Z_t^\sigma \sim N(0,1)$$

5.4.2 Burst parameters

To illustrate and clarify both of the results that follow from Theorem 5, we simulate two different bursts. The first burst is given by $(\alpha, \beta) = (0.55, 0.45)$ and the second by $(\alpha, \beta) = (0.8, 0.1)$ - such that one satisfies $a > \beta + 1/2$ and one does not.

We let the bursts last 10 minutes in the middle of the day, i.e. $\tau_b = 1/2 \cdot 6.5$ hours and set the burst estimators to

$$\mu_t^b = \frac{c_1}{(\tau_b - t)^\alpha}, \quad \sigma_t^b = \frac{c_2}{(\tau_b - t)^\beta} \ \text{ for } t \in [\tau_b - 10\text{min}; \tau_b[. \tag{14}$$

From the equations above, it is clear that with a timescale larger than 10min, a higher α yields a higher drift and thus a larger total drift burst. The same thing is true for the volatility and β parameter. However we want to show that the results of Theorem 5 follows from the *steepness* of the burst and not as a result of a larger burst size for larger values of α and β. We therefore choose a set of c_1 and c_2 such that the total bursts are of the same size in the two cases.
The constant c_1 is chosen such that $\int_{\tau_{db}-10\text{min}}^{\tau_{db}} \mu_t^{db} dt = -0.5\%$ which approximately corresponds to a 0.5% drop in the stock-price in 10 minutes. We choose the constant c_2 such that the volatility bursts contributes to a standard deviation of the X process of approximately 0.1% (in addition to the variance from the Heston process). Such that,

$$E\left[\left(\int_{\tau_{db}-10\text{min}}^{\tau_{db}} \sigma_t^{vb} dW_t\right)^2\right]^{1/2} = \left(\int_{\tau_{db}-10\text{min}}^{\tau_{db}} (\sigma_t^{vb})^2 dt\right)^{1/2} = 0.1\%.$$

Both the drift and volatility integrals are finite as $\alpha < 1$ and $\beta < 1/2$. Solving these integrals yields

$$c_1 = \frac{(1-\alpha)\cdot(-0.5\%)}{(10\text{min})^{1-\alpha}}, \quad \text{and} \quad c_2 = \left(\frac{(1-2\beta)\cdot(-0.1\%)^2}{(10\text{min})^{1-2\beta}}\right)^{1/2}. \tag{15}$$

The constants depend on the time scale. Since the Heston parameters are given in years we keep the simulation in years. We assume 252 trading days per as Aït-Sahalia & Kimmel (2007) uses 252 business days in the year. We also let the day be of length 6.5hours as this is the NYSE opening hours. We simulate a single day and thus set $t_0 = 0$ and $T = \frac{1}{252}$ and the burst duration of 10 minutes as $\frac{10}{60\cdot 6.5\cdot 252}$.

5.4.3 Jump parameter

In addition to simulating bursts, we simulate a process with a jump at the same time as the peak of the drift bursts. To make the two processes comparable, we let the jump be of the same size as the integrated drift bursts, i.e. $J = 0.5\%$.

5.4.4 Simulated paths

Figure 8 and 9 show a pure Heston-simulated path, along with three alternative paths that contain either a volatility burst, a jump or both a volatility and drift burst. In the first figure, $(\alpha, \beta) = (0.55, 0.45)$ and in the second $(\alpha, \beta) = (0.8, 0.1)$. To the naked eye the two bursts look very similar (because we adjusted the constants so the total size were the same), however by Theorem 5 we should expect that the T-estimator converges in distribution in the first case and diverges in the second.

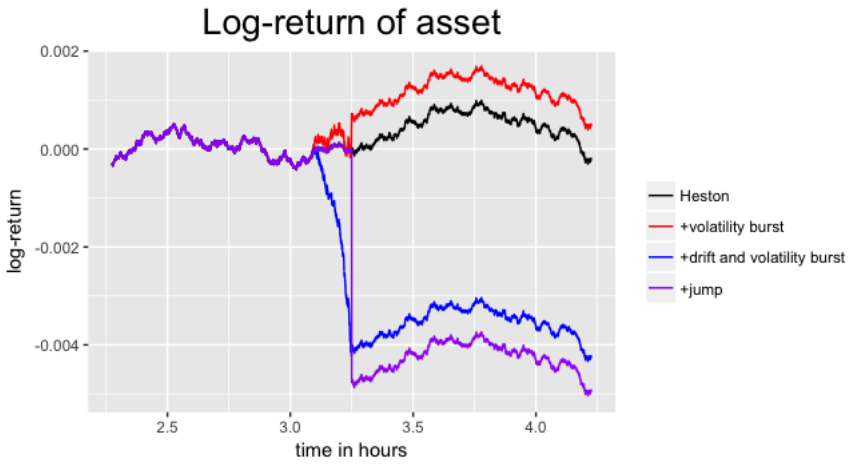


Figure 8: 4 simulated paths. The first is log-price in a standard Heston with annualized parameters $(\theta, \kappa, \xi, c, \rho) = (0.0457, 5.07, 0.48, 1.11, -0.767)$. In the second a volatility burst from eq. (14) is added with $c_2 = 5.01 \cdot 10^{-4}$ and $\beta = 0.45$. In the third both a volatility and drift burst with from eq. (14) with parameters $(c_1, c_2, \alpha, \beta) = (-2.99, 5.44 \cdot 10^{-4}, 0.55, 0.45)$ is added to the Heston model. In the fourth path a jump with the same size as the drift burst is added to the Heston simulation.

5.4.5 Convergence of T-estimator

We test the convergence of the T-estimator (at event time $t = \tau$) for 4 different processes:

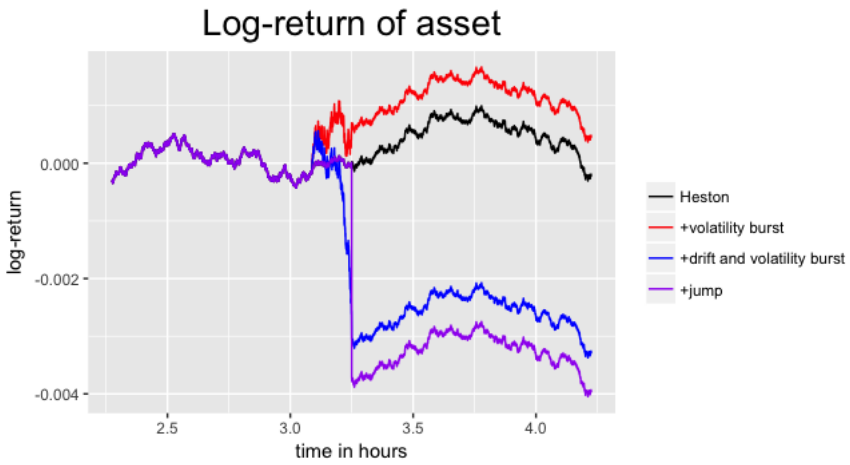


Figure 9: The same simulated paths as in Figure 8 but with burst parameters $(c_1, c_2, \alpha, \beta) = (-0.009, 0.069, 0.8, 0.1)$.

1) Heston
2) Heston + volatility burst
3) Heston + jump
4) Heston + volatility burst + drift burst

We use the same burst settings as in Figure 8 and 9.

To visualize the convergence, we increase the number of observations on the day gradually from $n = 50$ to $n = 60,000$. Since the convergence requires $h_n \to 0$ and $nh_n \to \infty$, we set $h_n = c\sqrt{dt}$. When choosing the constant c, we need to let it be large enough to have enough observations with significant weight but simultaneously small enough to detect changes in drift and/or volatility and thus detect the bursts. Heuristically, we let $c = 0.01$ as it then holds that $h_n/dt = 2.6$ for $n = 50$ and $h_n/dt = 89.8$ for $n = 60,000$. This corresponds to approximately 3 observations within a bandwidth for the lowest number of observations and 90 observations within a bandwidth for the largest number of observations.

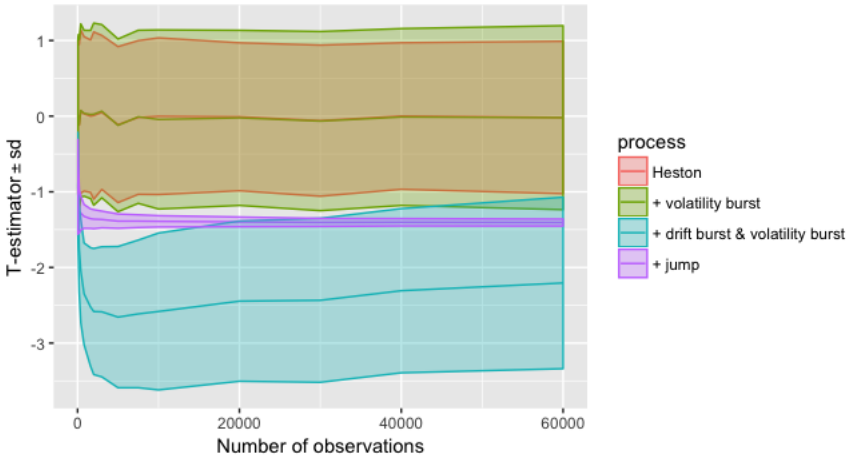


Figure 10: Convergence of the $T-$estimator at peak of event time ($\tau = 3.25$hours) for different processes (with the parameters from Figure 8). The ribbon is the standard deviation of the estimator. Note that the drift burst converges because $\beta + 1/2 > \alpha$.

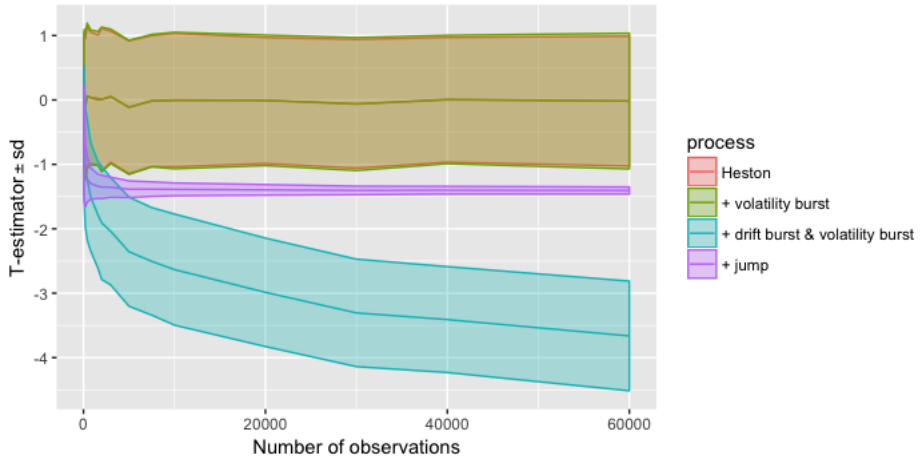


Figure 11: Convergence of the $T-$estimator at peak of event time ($\tau = 3.25$hours) for different processes (with the parameters from Figure 9). The ribbon is the standard deviation of the estimator. Note that the drift burst diverges as $\beta + 1/2 < \alpha$.

The value of T_τ^n is calculated 500 times for every n and the mean and standard deviation is illustrated as a ribbon around the mean. The results can be seen in Figure 10 and 11. We see that the pure Heston (which is the brown ribbon contained in the green) has a mean of zero and standard deviation of one in both cases. This is consistent with the result that it should be asymptotically $N(0,1)$ distributed. The Heston with a volatility burst has a slightly increased volatility in Figure 10. This fits very well as it should be asymptotic normal with standard devation $\sqrt{2^{2\cdot 0.45}} \approx 1.37$ according to the corollary. With $\beta = 0.1$, the variance is so close to 1 that it is almost indistinguishable from the standard deviation of the Heston as Figure 11 shows. This also fits well with the result that the standard deviation should converge to $\sqrt{2^{2\cdot 0.1}} \approx 1.07$. The jump converges fast in probability exactly as desired.

Though the two drift bursts seem very similar in the simulations (Figure 8 and 9), there is a huge difference in the T-statistics. The drift burst with $\alpha > \beta + 1/2$ clearly diverges, while the drift burst with $\alpha < \beta + 1/2$ peaks at $n \approx 5,000$ and decreases hereafter.
We notice that the drift burst with $\alpha < \beta + 1/2$ converges extremely slowly even despite $\alpha = 0.45$ and $\beta = 0.55$. This is because the convergence of the mean is very slow when a drift burst is present, and on small samples the drift burst contributes to a very large drift estimate. Because of this, depending on the threshold, there is a decent probability of detecting the drift burst despite $\alpha < \beta + 1/2$ and absence of arbitrage in the finite case.

5.4.6 The maximal T-estimator with jumps

Though it is true that the estimator at jump-time converges in probability to $\sqrt{\frac{K(0)}{K_2}}$, we have shown earlier that for every finite choice of h_n and n, there exists a jump-size where the T-estimator has higher probability of being above a certain threshold.

To test the bound in Theorem 1 in practice we use the simulated Heston-path from Figure 8 with $n = 23,400$ and bandwidth $h_n = 5$min. These are chosen based on the choices in Christensen et al. (2016) (the choice of bandwidth will be investigated in depth later).

We let the threshold be given by $q_{0.975}$, i.e. the 97.5% quantile of a normal distribution. Since the T-

estimator with our exponential kernel has $c = \sqrt{2}$, Theorem 1 yields a theoretical maximum percentile of $P(T_\tau > q_{97.5\%}) \leq P\left(T > \sqrt{(q_{0.975})^2 - 2}\right) \approx 8.7\%$.

Figure 12 below shows the empirical result of $P(T > q_{0.975})$ based on 500 simulations. The figure is as expected. Without any jumps the distribution is very close to a normal distribution and the rejection rate is very close to 2.5%. For large jump sizes, we have convergence in probability and thus do not have any values of the T-statistic above $q_{0.975}$. However there exists jump sizes with increased probability of $T > q_{0.975}$. It does not reach the theoretical maximum, which can be explained by the T-statistic not being fully normally distributed but on the conservative side as will be discovered later. It shows, however, that there exists a jump-size where $P(T > q_{0.975}) > 2.5\%$. This means that we cannot solely focus on the convergence theorem of the estimator for jumps but have to consider all different jump sizes. We will return to this when the final estimator is constructed.

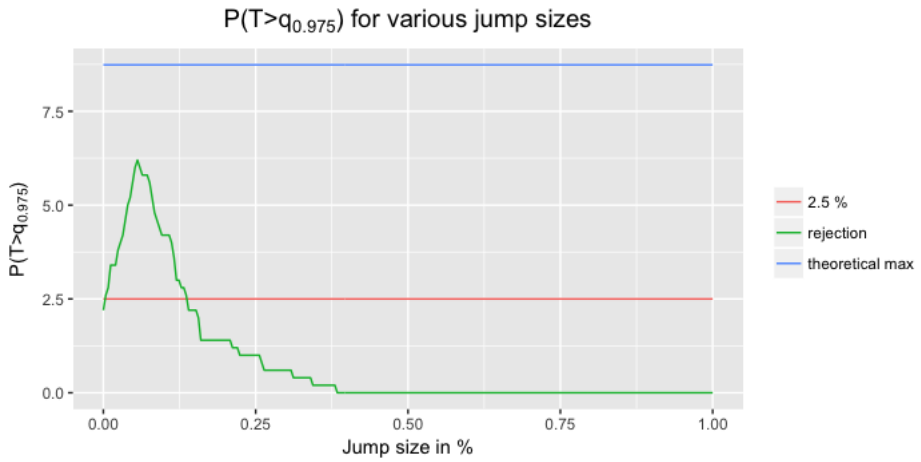


Figure 12: $P(T_\tau > q_{0.975})$ for the Heston simulation from Figure 8 with various jump sizes added. The parameter choices are $(n, h_n) = (23400, 5\text{min})$. We see, that it is above the detection rate for the normal distribution for some jump sizes, but below the theoretical maximum from Proposition 1.

6 Microstructure noise[6]

On real data the underlying price process X is hidden and what we observe is a noisy, polluted, version of this process. In practice the observed process contains microstructure effects for several reasons. Some of the elements contributing to the microstructure noise are price discreteness, rounding, bid-ask bounces and gradual response of prices to a block trade.

We assume the observed process Y is given by

$$Y_t = X_t + \epsilon_t \tag{16}$$

where ϵ_t is a noise process with mean zero and finite moments. We also assume that the noise process is independent of the X process.

[6]Responsible: Sebastian

Regarding the dependence of ϵ_t we assume that the sum of absolute covariances for all t are bounded, i.e.

$$\sum_{i=0}^{\infty} |\text{cov}(\epsilon_t, \epsilon_{t+i})| < C$$

for some constant C. When more strict assumptions are required, these will be stated explicitly.

6.1 Complications of noise

The estimators created in Christensen et al. (2016) are not robust to microstructure noise. Robustness to microstructure noise is important when modelling high-frequency data, and therefore we build the theoretical foundation for handling microstructure noise in this chapter.

Microstructure noise complicates the estimation significantly. The reason is that while the increments of the X-process converge in probability to zero as the distance between the observations decreases, the variance of the microstructure noise remains unchanged. Thus the quadratic variance of the observed price process is unbounded.

If we were to use the same estimator for the drift, as without noise, we would have the estimator

$$\hat{\mu}_t^n = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta Y_{i,n}. \tag{17}$$

The theorem below states the convergence of the estimator in presence of microstructure noise. The proof of the theorem highlights the general problem of the microstructure noise very well.

Theorem 6. *With the $\hat{\mu}$ estimator from (17) we have*

$$h_n \hat{\mu}_t^n \xrightarrow{p} \epsilon_n \tag{18}$$

Proof. We rewrite

$$\begin{aligned} h_n \hat{\mu}_t^n &= \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta Y_{i,n} \\ &= \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) (\Delta X_{i,n} + \Delta \epsilon_{i,n}) \\ &= \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta X_{i,n} + \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta \epsilon_{i,n} \end{aligned}$$

From the convergence of the drift estimator in absence of noise, we know that

$$\frac{1}{\sqrt{h_n}} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta X_{i,n} \xrightarrow{d} N(0, K_2 \sigma_{t-}^2)$$

and from this follows that

$$\sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta X_{i,n} \xrightarrow{p} 0.$$

For the noise term we have

$$\sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \Delta \epsilon_{i,n} =$$

$$K\left(\frac{t_{n-1}-t}{h_n}\right)\epsilon_n - K\left(\frac{t_0-t}{h_n}\right)\epsilon_0 + \sum_{i=1}^{n-1}\left(K\left(\frac{t_{i-1}-t}{h_n}\right) - K\left(\frac{t_i-t}{h_n}\right)\right)\epsilon_i. \tag{19}$$

We investigate the convergence of the expression above - one term at a time.

For the first term we have $K(\frac{t_{n-1}-t}{h_n}) = e^{-\frac{t}{nh_n}} \to K(0) = 1$. Using Slutsky, this means that $K(\frac{t_{n-1}-t}{h_n})\epsilon_n \xrightarrow{p} \epsilon_n$.

For the second term, we have $K(\frac{t_0-t}{h_n}) = e^{-\frac{t}{h_n}} \to 0$, and from this it follows that $K(\frac{t_0-t}{h_n})\epsilon_0 \xrightarrow{p} 0$.

Before investigating the convergence of the last term in (19), we make the convention $\Delta K_i = K(\frac{t_{i-1}-t}{h_n}) - K(\frac{t_i-t}{h_n})$.

We naturally have an expectation of zero in the last term due to linearity of expectations. Regarding the variance of the last expression we have

$$\text{Var}\left(\sum_{i=1}^{n-1}\Delta K_i\epsilon_i\right) = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\text{Cov}(\Delta K_i\epsilon_i, \Delta K_j\epsilon_j) = \sum_{i=1}^{n-1}\Delta K_i\sum_{j=1}^{n-1}\Delta K_j\text{Cov}(\epsilon_i,\epsilon_j).$$

Exploiting that $\sum_{i=1}^{\infty}|\text{Cov}(\epsilon_i,\epsilon_j)| < c$ for some constant c, we can create an upper limit and then as $1 > \Delta K_i > 0 \ \forall \ i$ we have

$$\sum_{i=1}^{n-1}\Delta K_i\sum_{j=1}^{n-1}\Delta K_j\text{Cov}(\epsilon_i,\epsilon_j) \leq \sum_{i=1}^{n-1}\Delta K_i\sum_{j=1}^{n-1}\Delta K_j|\text{Cov}(\epsilon_i,\epsilon_j)| \leq$$

$$(\max_j \Delta K_j)\sum_{i=1}^{n-1}\Delta K_i\sum_{j=1}^{n-1}|\text{Cov}(\epsilon_i,\epsilon_j)| \leq (\max_j \Delta K_j)c\sum_{i=1}^{n-1}\Delta K_i.$$

Having that that $\sum_{i=1}^{n-1}\Delta K_i = K(0) - K(\frac{t}{h_n}) < K(0) = 1$, we get

$$\text{Var}\left(\sum_{i=1}^{n-1}\Delta K_i\epsilon_i\right) \leq c\cdot(\max_j \Delta K_j) \to 0$$

since $\max_j \Delta K_j \to 0$.

Hence the variance of the third part in (19) converges to zero, and we can conclude that it converges in probability to zero. Then, by using Slutsky's equation we can collect the terms to achieve

$$h_n\hat{\mu}_t^n \xrightarrow{p} \epsilon_n$$

as we desired. □

It is clear from the proof that the problem with the microstructure noise is at the end points. As long as ΔK_i converges "fast enough" to zero, the weights of the noises in the middle are so small that the influence is negligible. The only way we are able to make the estimator converge is by somehow removing the bias that comes from the last microstructure noise term.

The same problem exists for the volatility estimator. Consistent volatility estimation under microstructure noise is studied much more in the literature than drift estimation. We will therefore borrow ideas from volatility estimation in literature to construct our estimator. We will in this thesis examine two of these methods. The objective of both methods is to *fade out* the noise in the edges - the end points - such that the first and last noise term is allocated a weight of zero.

6.2 Pre-averaging

The first method to make the estimators robust to microstructure noise that we will examine is pre-averaging. One of the pioneers of this is Jacod et al. (2009) who construct a consistent estimator of the integrated variance using this approach.

The idea behind pre-averaging is to sum over several observations and treat this as a single observation. This allows for the microstructure noise to cancel out or "fade away".

The intuition can be viewed in several ways. In this thesis, we adopt the general notation from Jacod et al. (2009).
We define the pre-averaged process $\Delta\bar{Y}_i$ as

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

where g is a kernel, which is positive on the interval $(0,1)$ with $g(x) \to 0$ for $x \to 0^+$ and $g(x) \to 0$ for $x \to 1^-$. Throughout the thesis we will use the following kernel for pre-averaging

$$g(x) = \begin{cases} \min(x, 1-x) & \text{for } x \in [0,1] \\ 0 & \text{for } x \notin [0,1] \end{cases} \tag{20}$$

which satisfies the conditions.

Figure 13 shows the weight of the individual ΔY_{i-j} in every pre-averaged observation $\Delta\bar{Y}_i$ along with the total weight of the delta process in the sum $\sum_{i=1}^{n} \Delta\bar{Y}_i$. It is non-trivial that the total weight is of any significance, however it will be shown later, that the total sum of the weights (the blue line in the figure) is of importance.

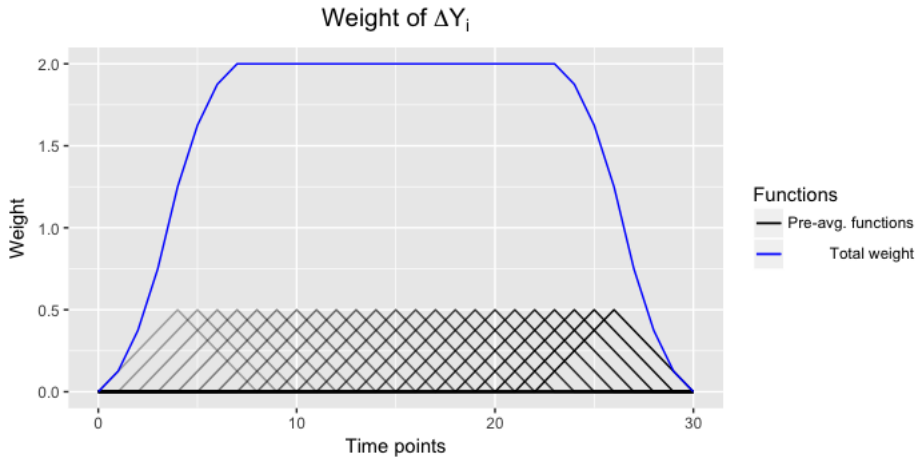


Figure 13: An illustration of the weights from the pre-averaging kernel $g(x) = \min(1-x, x)$ on the ΔY_i process plotted along with the sum of the weights. The parameters are $k_n = 8$ and $n = 30$.

To get an intuitive understanding of the behavior of the estimator, we rewrite it as

$$\begin{aligned}\Delta\bar{Y}_i &= \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)(\Delta X_{i+j+1} + \Delta\epsilon_{i+j+1}) \\ &= \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)(\Delta X_{i+j+1}) + \sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j+1}{k_n}\right)\right)\epsilon_{i+j} \\ &\approx \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)(\Delta X_{i+j+1}) + \frac{1}{k_n}\sum_{j=1}^{k_n} -g'\left(\frac{j}{k_n}\right)\epsilon_{i+j}.\end{aligned}$$

The point is that $\frac{1}{k_n}\sum_{j=1}^{k_n} -g'(\frac{j}{k_n})\epsilon_{i+j} \xrightarrow{p} 0$ as $k_n \to \infty$ from the law of large numbers (as the noise process is mixing). Hence the noise part becomes insignificant.

In other articles such as Christensen et al. (2012), the pre-averaging is written as

$$r_i^* = \frac{1}{k_n}\left(\sum_{j=k_n/2}^{k_n} Y_{i+j} - \sum_{j=1}^{k_n/2-1} Y_{i+j}\right)$$

where $k_n \geq 2$ and even.
The two ways of pre-averaging are equivalent with the choice of pre-average kernel from (20). This can be seen from the following:

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

If k_n is even, then $g(\frac{j}{k_n}) - g(\frac{j+1}{k_n}) = -\frac{1}{k_n}$ for $j \leq k_n/2-1$ and $g(\frac{j}{k_n}) - g(\frac{j+1}{k_n}) = \frac{1}{k_n}$ for $j \geq k_n/2$. Hence

$$\Delta\bar{Y}_i = \frac{1}{k_n}\left(\sum_{j=k_n/2}^{k_n} Y_{i+j} - \sum_{j=1}^{k_n/2-1} Y_{i+j}\right) = r_i^*.$$

We will not go into specific examples of estimators using the pre-average approach in this section, however we will mention that with the pre-averaging approach, it is possible to make different estimators of the integrated volatility. We will divide these into two groups.

The first group of estimators are estimators with $k_n = \theta\sqrt{n}$ for some konstant θ. Estimators of this group can be found in articles such as Jacod et al. (2009) or Christensen et al. (2012) and will be further elaborated later. They have a convergence rate of $O(n^{-1/4})$ but the estimators do not necessarily yield positive values.

The second group of estimators are estimators with $k_n/\sqrt{n} \to \infty$. These can achieve a convergence rate up to $O(n^{-1/5})$ as seen in e.g. Christensen et al. (2010) and be almost surely positive.

6.3 Kernel-based approach

A second approach to volatility estimation is the realized kernel based approach in Barndorff-Nielsen et al. (2008). The idea behind this is that we can estimate the integrated variance as

$$\int_0^T \sigma(s)^2 ds = \text{Var}(X_T - X_0) = \text{Var}\left(\sum_{i=1}^{T}(X_i - X_{i-1})\right) = \text{Var}\left(\sum_{i=1}^{T}\Delta X_i\right)$$

If we do not assume independence but assume mean zero, we get

$$\text{Var}\left(\sum_{i=1}^{T}\Delta X_i\right) = \sum_{i=1}^{T}\sum_{j=1}^{T}\text{cov}(\Delta X_i, \Delta X_j) = \sum_{i=1}^{T}\sum_{j=1}^{T}E(\Delta X_i \cdot \Delta X_j).$$

if we know the process only has a finite number L of non-zero autocorrelations (i.e. $E(X_i X_{i+n}) = 0$ for $n > L$), we have

$$\sum_{i=1}^{T}\sum_{j=1}^{T}E(\Delta X_i \cdot \Delta X_j) = \gamma_0 + 2\sum_{l=1}^{L}\gamma_l \qquad \text{where} \qquad \gamma_l = \sum_{i=1+l}^{T}E(\Delta X_i \cdot \Delta X_{i-l}).$$

In the literature, consistent estimation of heteroskedasticity and autocorrelated (HAC) matrices has already been studied a great deal, and this is very similar to the expression above.
Newey & West (1987) shows that if X_i is stationary and mixing, and we know L (the number of non-zero autocorrelations), we can estimate $\frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}E(\Delta X_i \cdot \Delta X_j)$ with the sample autocovariances

$$\hat{\sigma} = \gamma_0 + 2\sum_{l=1}^{L}\gamma_l$$

$$\gamma_l = \frac{1}{T}\sum_{i=1+l}^{T}\Delta X_i \cdot \Delta X_{i-l}$$

and this estimate is consistent. The estimator has two disadvantages. The first is that L has to be finite and known a priori. The second is that the estimator might not necessarily yield positive estimates.

Newey & West (1987) show that it is possible to create an estimator which handles both of the disadvantages. By letting L be a function of T, which increases slowly enough ($L \to \infty$ but $LT^{-1/4} \to 0$), it is possible to solve the first issue and still estimate the variance consistently.
By introducing a proper kernel-function in the estimation, it is possible to solve the second. They suggest the following estimator

$$\hat{\Sigma} = \gamma_0 + 2\sum_{l=1}^{L(T)}w(l, L(T))\gamma_l$$

$$\gamma_l = \frac{1}{T}\sum_{i=1+l}^{T}\Delta X_i \cdot \Delta X_{i-l}$$

which under suitable conditions (Theorem 2 in Newey & West (1987)) is consistent and almost surely positive. They show that if $w(l, L(T)) = 1 - \frac{l}{L+1}$, both non-negativity and consistency is achieved.

Barndorff-Nielsen et al. (2008) has worked on an estimator very similar to this for estimating integrated volatility in presence of microstructure noise. This is also found in Barndorff-Nielsen et al. (2009) and

Barndorff-Nielsen et al. (2011). The estimator in the first article is different from the estimator in the subsequent articles. Both estimators use a kind of 'jittering' in the edges to fade out noise effects. We will shortly review both estimators including the intuition of how noise bias at the edges is avoided.

The problem with the estimator from Newey & West (1987) is that it cannot be directly transferred to the integrated volatility. This is due to the fact that if we present the estimator as

$$\hat{\Sigma}_0 = \gamma_0 + 2\sum_{l=1}^{L(T)} w(l, L(T))\gamma_l$$

$$\gamma_l = \sum_{i=1+l}^{T} \Delta X_i \cdot \Delta X_{i-l}$$

then with correct convergence rates and a suitable kernel function (which will be elaborated on later), it holds that

$$\hat{\Sigma} \xrightarrow{p} \epsilon_0^2 + \epsilon_n^2 + \int_0^T \sigma_s^2 ds.$$

This follows e.g. from Section 2.2 in Barndorff-Nielsen et al. (2009). This shows, that we have problems with end effects of the noise just as without pre-averaging, if the estimator is not modified further.

In Barndorff-Nielsen et al. (2008) this is solved by changing the indices in the ends and instead use the estimator

$$\hat{\Sigma}_1 = \sum_{l=-L(T)}^{L(T)} w(l, L(T))\gamma_l$$

$$\gamma_l = \sum_{i=1}^{T} \Delta X_i \cdot \Delta X_{i+l}.$$

The difference is that with the indices now ranging from 1 instead of from $1+l$, the observations $X_0, X_{-1}, ..., X_{-L+1}$ are used to eliminate the noise effects in the beginning and similar $X_{T+1}, X_{T+2}, ..., X_{T+L}$ are used to eliminate the noise effects in the end.

This estimator can be consistent if the appropriate weight function is used, and it thus solves the problem regarding the microstructure noise. They show that this estimator can have a rate of convergence up to $O(n^{-1/4})$. Unfortunately the estimator cannot be guaranteed to be positive when optimal kernels are chosen.
They therefore create a new estimator in Barndorff-Nielsen et al. (2011) and Barndorff-Nielsen et al. (2009). Here, the end points are first redefined as

$$X_0 = \frac{1}{m}\sum_{i=0}^{m-1} X_{-i} \quad \text{and} \quad X_n = \frac{1}{m}\sum_{i=0}^{m-1} X_{n+i} \tag{21}$$

That way the bias from the ends are eliminated (as $m \to \infty$), and the estimator $\hat{\Sigma}_0$ can be used afterwards, which gives the estimator

$$\hat{\Sigma}_2 = \gamma_0 + 2\sum_{l=1}^{L(T)} w(l, L(T))\gamma_l \qquad \text{with} \qquad \gamma_l = \sum_{i=1+l}^{T} \Delta X_i \cdot \Delta X_{i-l} \tag{22}$$

where X_0 and X_n are given by (21). This estimator can, with the right kernel and rates of convergences of the parameters, be ensured to be non-negative and consistent at the same time. It is also more robust to dependent and endogenous noise than the estimator previous estimator (see Section 6.1 in Barndorff-Nielsen et al. (2011)), and this is why they advocate using this estimator instead of the old. The disadvantage is that the best rate of convergence is only $O(n^{-1/5})$.

6.4 Comparison of methods

The estimators based on pre-averaging and the estimators from the kernel-based approach seem at first glance quite different. However once understood, they are quite similar. We will not show it here, but the pre-averaged volatility estimator can be rewritten on a form that is very similar to the kernel-based estimator. The k_n plays the role of L and a transformation of the pre-averaging kernel $g(x)$ replaces the weight-function $w(l, L(T)))$. It is shown e.g. in Jacod et al. (2009) that the only significant difference is how the edge effects are handled.

With the pre-averaging approach, it was possible to either create estimators with a convergence rate of $O_p(n^{-1/4})$, which though was not surely positive or estimators with a slower rate of convergence of $O_p(n^{-1/5})$, which on the other hand was guaranteed to be non-negative.
Similarly, the estimator from Barndorff-Nielsen et al. (2008) had a rate of convergence of $O_p(n^{-1/4})$ but was not guaranteed to be positive, while the estimator from Barndorff-Nielsen et al. (2009) and Barndorff-Nielsen et al. (2011) had a convergence rate of $O_p(n^{-1/5})$ but was almost surely positive.

One disadvantage of using many of the fast convergence estimators based on the pre-averaging approach (e.g. as in Jacod et al. (2009) or Christensen et al. (2012)) is that the variance of the microstructure noise has to be estimated. This is not the case for the fast converging estimator in Barndorff-Nielsen et al. (2008), which suggests these are easier to implement in practice.
However with the pre-average approach, it is also possible to create estimators that are robust to jumps. This has not been done for the kernel-based approach, although it is likely possible to do so using the inspiration from the pre-averaged estimator.
The estimators with slower convergence rates (both with pre-average and kernel-based approach) are more robust to noise that are serially independent and dependent of X. For the kernel-based estimators, we have from Barndorff-Nielsen et al. (2011), that consistency holds with the weak assumption that $\sum_{i=0}^{\infty} |\text{cov}(\epsilon_t, \epsilon_{t+i})| < \infty$, and the noise is in fact allowed to be completely determined by X.
This means that if we need the estimator to be robust to jumps when estimating integrated variance, we need to use the pre-averaging approach. Other than this, the differences between the pre-averaging and kernel-based approach are subtle in reality.

6.5 Noise robust estimators

We will use the pre-averaging approach to develop the theory for detecting bursts in presence of microstructure noise and construct an expanded version of the T-estimator.

6.5.1 Previous literature on pre-averaging

In literature the pre-averaging has been forward-looking - such that it uses future observations for estimating the current volatility. These are either defined as in Jacod et al. (2009) with

$$\Delta \bar{Y}_{i,n} = \sum_{j=1}^{k_n - 1} g\left(\frac{j}{k_n}\right) \Delta Y_{i+j+1} \qquad \text{where} \qquad \Delta Y_{i+j+1} = Y_{i+j+1} - Y_{i+j}$$

or in the less general case as in Christensen et al. (2012) for some even k_n

$$r^*_{i,n} = \frac{1}{k_n} \left(\sum_{j=k_n/2}^{k_n - 1} Y_{i+j} - \sum_{j=0}^{k_n/2 - 1} Y_{i+j} \right) \tag{23}$$

It is already shown that these are equivalent when k_n is even and $g(x)$ is given as in (20).

We introduce the following constants that will become handy for the future results.

$$\psi_1 = \int_0^1 g(x)dx, \qquad \psi_2 = \int_0^1 g(x)^2 dx \quad \text{and} \quad \psi_3 = \int_0^1 g'(x)^2 dx$$

The most used estimators with the pre-averaging approach let $k_n = \theta\sqrt{n}$. This is due to the fact that choosing $k_n \propto \sqrt{n}$ will leave some noise bias, which can be subtracted later to improve the convergence of the estimator. We will adopt this for our estimator.

Since drift estimation has been neglected in the past, the pre-averaging approach has primarily been used to estimate integrated variance. One of the estimators for the integrated variance is the realized variance with pre-averaging (RV^*), which with small-sample corrections, can be written as

$$RV^* = \frac{n}{(n-k_n+2)k_n\psi_2} \sum_{i=0}^{n-k_n+1} (\Delta\bar{Y}_{i,n})^2 - \frac{\psi_3}{\theta^2\psi_2}\hat{\omega}^2.$$

It is already known (e.g. from Appendix A in Christensen et al. (2012)) that if the noise process is i.i.d. and independent of X, then

$$RV^* \xrightarrow{p} \int_0^T \sigma_s^2 ds + \sum_{i=1}^N J_i.$$

Hence it is not robust to jumps, but in the absence of jumps, it would be a consistent estimator of the integrated variance.

Another estimator is the bi-power variation

$$BV^* = \frac{n\pi}{2(n-2k_n+2)k_n\psi_2} \sum_{i=0}^{n-2k_n+1} |\Delta\bar{Y}_{i,n}||\Delta\bar{Y}_{i+k_n,n}| - \frac{\psi_3}{\theta^2\psi_2}\hat{\omega}^2 \tag{24}$$

which has the advantage over the realized variance, that

$$BV^* \xrightarrow{p} \int_0^T \sigma_s^2 ds$$

so it is robust to jumps.

6.5.2 From integrated to instantaneous variance

The previous literature has primarily studied integrated variance and transformed this into volatility estimates by piece-wise constant approximations. However, in the burst analysis, we want to detect drift and volatility bursts, hence we want the estimates to be as instantaneous as possible. To do this, we will introduce a left sided kernel with a bandwidth converging to zero, just as in the case without the noise, and we choose to do this in the pre-averaging part. We also want to detect it *in real time*. Therefore it is most convenient to change the indices pre-averaging such that it is backward looking - index i takes values of the past $k_n - 1$ indices and not the future.
For the burst analysis we redefine the pre-averaging to

$$\Delta\bar{Y}_i = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_i - t}{h_n}\right)(Y_{i-j} - Y_{i-j-1})$$

6.5.3 Unscaled estimators

First we re-define a pair of drift and volatility estimators to be

$$\tilde{\mu}_t^n = \frac{1}{k_n h_n} \sum_{i=k_n}^{n+1} \Delta \bar{Y}_i \tag{25}$$

and

$$(\tilde{\sigma}_t^n)^2 = \frac{1}{h_n k_n} \sum_{i=k_n}^{n+1} (\Delta \bar{Y}_i)^2. \tag{26}$$

These are denoted with a $\tilde{}$ symbol (because they will later be re-scaled). Note that our volatility estimator is similar to RV^* and not BV^*. This is chosen on purpose. It is very important that it is NOT consistent in the presence of jumps. This point will will be elaborated later.

With the estimators defined above we have the following results:

Theorem 7. *Let $\tilde{\mu}_t^n$ be defined as in (25). We then have at a time point without jump that*

$$\sqrt{h_n}(\tilde{\mu}_t^n - \mu_{t-}) \xrightarrow{d} N(0, \psi_1^2 K_2 \sigma_{t-}^2) \tag{27}$$

stably in law as $n \to \infty$, $h_n \to 0$ and $\sqrt{n} h_n \to \infty$.

Proof. See Appendix 1. □

Note that this is again a new result. To our knowledge no-one has estimated the drift with this type of pre-averaging method before. An important difference between this result and the convergence result in Theorem 1 is that we now need $\sqrt{n} h_n \to \infty$, where it was previously only necessary that $n h_n \to \infty$. Because of this h_n cannot converge to zero as fast as previously and the convergence is therefore be much slower.

For the volatility estimator, we have the following result:

Theorem 8. *Let $(\tilde{\sigma}_t^n)^2$ be given by (26) and let the noise process be iid. We then have at a time point without jump that*

$$(\tilde{\sigma}_t^n)^2 \xrightarrow{p} \psi_2 K_2 \sigma_{t-}^2 + \frac{\psi_3 K_2}{\theta^2 t} \omega^2 \tag{28}$$

as $n \to \infty$, $h_n \to 0$ and $\sqrt{n} h_n \to \infty$.

Proof. See Appendix 1. □

Note that this result is very similar to the result in Jacod et al. (2009), with the only difference, that the left-sided kernel K is included in the convergence result. The result can be expanded to dependent noise with finite sum of covariance - this is what ω^2 would represents - however this is beyond the scope of the thesis.

6.5.4 Scaled estimators

Using the previous results, we can scale the estimators, and define the new scaled estimators as

$$\hat{\mu}_t^n = \frac{1}{\psi_1 k_n h_n} \sum_{i=k_n}^{n+1} \bar{Y}_i \tag{29}$$

$$(\hat{\sigma}_t^n)^2 = \frac{1}{\psi_2 h_n k_n} \sum_{i=k_n}^{n+1} \bar{Y}_i^2 - \frac{\psi_3 K_2}{\psi_2 \theta^2 t} \hat{\omega}^2. \tag{30}$$

where $\hat{\omega}^2$ is an estimator with the property that $\hat{\omega}^2 \xrightarrow{p} \omega^2$.

Note, that we know have $(\hat{\sigma}_t^n)^2 \xrightarrow{p} K_2 \sigma_{t-}^2$, where as we in the chapter without noise had $(\hat{\sigma}_t^n)^2 \xrightarrow{p} \sigma_{t-}^2$. The reason for this is that the leftsided kernel appears in the pre-averaging part, before the process is squared. It thus effectively appears with squared weight in the estimator.
Because of this we do not need to rescale the T-estimator with K_2. We let T_t^n be given by

$$T_t^n = \frac{\sqrt{h_n}\hat{\mu}_t^n}{\sqrt{(\hat{\sigma}_t^n)^2}}. \tag{31}$$

With these scaled estimators, we can now study the behavior T_t^n both in the normal case, with a jump and in presence of bursts.

Proposition 2. *Let T_t^n be defined by (31) with $\hat{\mu}_t^n$ and $\hat{\sigma}_t^n$ from (29) and (30). Then in absence of jumps and bursts, it holds that*

$$T_t^n \xrightarrow{d} N(0,1) \tag{32}$$

as $n \to \infty$, $h_n \to 0$ and $\sqrt{n}h_n \to \infty$.

Proof. This follows immediately from Slutsky's (for stable convergence) and Theorem 7 and 8. □

We see that in absence of jumps and bursts, it converges to a standard normal distribution similarly to the scenario without microstructure noise.
The jump and bursts are more difficult to handle in the pre-average case. The reason is that when we pre-average, the jump component in the T-estimators is included in every $\Delta\bar{Y}_i$ for $i \in \{\tau, \tau+1, ..., \tau+k_n-2\}$. In other words, the jump is *smoothed out*. We need to show that the estimator is bounded from the beginning of the jump to the peak of the weight of the jump component in the estimators such that we can make sure that the T-estimator does not diverge. It turns out that the estimator also converges in probability k_n observations after the jump and that the same probability bound holds as in the pre-averaging case.

The drift estimator diverges in presence of jump (with same reasoning as in absence of noise). For the estimator to be consistent, the volatility estimator also needs to diverge in case of a jump. This is why the estimator was chosen deliberately to be based on the RV^* estimator and not the BV^* estimator. We have the following results:

Theorem 9. *Let X follow the dynamics from (1) and define τ as a time point with a jump with size J. In addition assume we only observe Y as defined in (16). It then holds that*

$$T_{\tau+k_n-1}^n \xrightarrow{p} sign(J) \tag{33}$$

as $n \to \infty$, $h_n \to 0$ and $\sqrt{n}h_n \to \infty$.

Furthermore for all jump sizes and at all time points after the jump - the probability bound on the T-estimator from Theorem 1 is still valid for this case.

Proof. See appendix 1. □

Theorem 10. *Let X follow the dynamics from (3) and assume that there is a drift- and volatility burst with a peak at time τ. In addition assume we only observe Y as defined in (16). It then holds that*

$$|T^n_{\tau+k_n-1}| \xrightarrow{p} \infty \quad \text{for } \beta + 1/2 < \alpha \tag{34}$$

$$T^n_{\tau+k_n-1} \xrightarrow{d} N(0,1) \quad \text{for } \beta + 1/2 > \alpha \tag{35}$$

as $n \to \infty$, $h_n \to 0$ and $\sqrt{n}h_n \to \infty$.

Proof. See Appendix 1. □

Note that when the volatility burst dominates, the limiting distribution is a standard normal, whereas before pre-averaging (Theorem 5) the asymptotic variance was $2^{2\beta}$. This is a consequence of the fact that the left sided kernel K only appears once in the volatility estimator without pre-averaging but twice in the estimator with pre-averaging.

The three theorems above show that if we pre-average and redefine the drift- and volatility estimators, we should still be able to detect drift bursts in the presence of microstructure noise.

6.5.5 Estimating noise variance

A necessity for the convergence of $\hat{\sigma}^2_t$ is that we can create a consistent estimator of ω^2. This estimator should not diverge neither in jump nor burst.

One possible estimator is the estimator

$$\hat{\omega}^2 = -\frac{1}{n-1}\sum_{i=2}^{n}(\Delta Y_i)(\Delta Y_{i-1}) \tag{36}$$

as is used in Christensen et al. (2012).

Proposition 3. *Let the noise process $(\epsilon_i)_{i\in\mathbb{N}}$ be iid and independent of X. Let $\hat{\omega}^2$ be given as in (36). Then*

$$\hat{\omega} \xrightarrow{p} \omega^2$$

both in presence of finite number of jumps and bursts.

Proof. In absence of bursts or jumps we have

$$\hat{\omega}^2 = -\frac{1}{n-1}\sum_{i=2}^{n}(\Delta Y_i)(\Delta Y_{i-1}) = -\frac{1}{n-1}\sum_{i=2}^{n}(\Delta X_i + \epsilon_i - \epsilon_{i-1})(\Delta X_{i-1} + \epsilon_{i-1} - \epsilon_{i-2})$$

and since the process ΔY_i is mixing (we have $\text{cov}(\Delta Y_i, \Delta Y_j) = 0$ for $|i-j| \geq 2$) the strong law of large numbers yields

$$\hat{\omega}^2 \xrightarrow{p} E\left[-(\Delta Y_i)(\Delta Y_{i-1})\right]$$

and we have

$$E\left[-(\Delta Y_i)(\Delta Y_{i-1})\right] = E\left[-(\Delta X_i + \epsilon_i - \epsilon_{i-1})(\Delta X_{i-1} + \epsilon_{i-1} - \epsilon_{i-2})\right] = E(\epsilon_{i-1}^2) = \omega^2.$$

Thus we have $\hat{\omega} \xrightarrow{p} \omega^2$ in absence of jumps and bursts. In presence of a jump with size J at time τ, we have

$$\hat{\omega}^2 = -\frac{1}{n-1}\sum_{i=2}^{n}(\Delta Y_i + 1_{\{i=\tau\}}J)(\Delta Y_{i-1} + 1_{\{i-1=\tau\}}J) = -\frac{1}{n-1}J(\Delta Y_{\tau-1} + \Delta Y_{\tau+1}) - \frac{1}{n-1}\sum_{i=2}^{n}(\Delta Y_i)(\Delta Y_{i-1})$$

and since the first term converges in probability to zero, the convergence follows from Slutsky's.

Regarding the bursts, we make the conventions

$$D_{i,n} = \int_{t_{i-1}}^{t_i} \frac{1}{(\tau - s)^\alpha} ds$$

and

$$S_{i,n} = \int_{t_{i-1}}^{t_i} \frac{1}{(\tau - s)^\beta} dW_s$$

which are respectively the drift and volatility bursts from time point t_{i-1} to time point t_i for a burst with peak at time τ.

We can then write the estimator as

$$-\frac{1}{n-1}\sum_{i=2}^{\tau}(\Delta Y_i + S_{i,n} + D_{i,n})(\Delta Y_{i-1} + S_{i-1,n} + D_{i-1,n}) - \frac{1}{n-1}\sum_{i=\tau+1}^{n}(\Delta Y_i)(\Delta Y_{i-1}).$$

Investigating the terms, we have for the drift burst

$$\sum_{i=2}^{\tau} D_{i,n} = \int_{t_1}^{\tau} \frac{1}{(\tau - s)^\alpha} ds < \infty$$

and thus

$$\frac{1}{n-1}\sum_{i=2}^{\tau} D_{i,n}D_{i-1,n} \leq \frac{1}{n-1}\max_i D_{i,n}\sum_{i=2}^{\tau} D_{i,n} \to 0.$$

Regarding the volatility burst, we have

$$E\left[\frac{1}{n-1}\sum_{i=2}^{\tau} S_{i,n}S_{i-1,n}\right] = 0$$

and

$$\begin{aligned}
\mathrm{Var}\left[\frac{1}{n-1}\sum_{i=2}^{\tau} S_{i,n}S_{i-1,n}\right] &= \frac{1}{(n-1)^2}\sum_{i=2}^{\tau}\sum_{j=2}^{\tau}\mathrm{cov}\left(S_{i,n}S_{i-1,n}, S_{j,n}S_{j-1,n}\right) \\
&= \frac{1}{(n-1)^2}\sum_{i=2}^{\tau}\mathrm{Var}(S_{i,n}S_{i-1,n}) \\
&= \frac{2\tau}{(n-1)^2}\mathrm{Var}(S_{i,n}) \to 0
\end{aligned}$$

since $\mathrm{Var}(S_{i,n}) \to 0$. From this follows that

$$\frac{1}{n-1}\sum_{i=2}^{\tau} S_{i,n}S_{i-1,n} \xrightarrow{p} 0.$$

With similar calculations for the product terms, we see that these also converge in probability to zero. Then using Slutsky's, we get

$$-\frac{1}{n-1}\sum_{i=2}^{\tau}(\Delta Y_i + S_{i,n} + D_{i,n})(\Delta Y_{i-1} + S_{i-1,n} + D_{i-1,n}) - \frac{1}{n-1}\sum_{i=\tau+1}^{n}(\Delta Y_i)(\Delta Y_{i-1}) \xrightarrow{p} \omega^2.$$

The estimator is thus consistent to both jumps and bursts. □

The drawback with this estimator is that the noise process has to be stationary with constant variance for this to be consistent.

A way to accommodate for a slowly changing variance of the noise process is with a kernel based estimator

$$\hat{\omega}^2 = -\frac{1}{h_n}\sum_{i=2}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta Y_i)(\Delta Y_{i-1}).$$

This estimator is consistent if $h_n\sqrt{n} \to \infty$. This follows from the proof of convergence of the T-estimator in presence of jumps and bursts.

The kernel based estimator has a much lower bias for a noise with non-constant variance. However it has a much higher variance, because it effectively uses fewer observations. It also converges slower than the estimator from (36). Hence we will in subsequent chapters use the first estimator.

6.5.6 Comments regarding the kernel-based approach

Though not shown theoretically, we believe the same results can be achieved with an expansion of the kernel-based volatility estimation from Barndorff-Nielsen et al. (2011).

What was done in the pre-averaging above was to expand the RV^* estimator, which estimates the quadratic variance. Therefore it had the properties

$$RV^* \xrightarrow{p} \int_0^T \sigma_s^2 ds + \sum_{i=1}^{n_J} J_i^2$$

with a left sided kernel. It was important here that the estimator estimated the quadratic variance and thus was inconsistent in presence of jumps.

The estimator from Barndorff-Nielsen et al. (2011) in eq. (22) also estimates the quadratic variance. In the article, they refer to the heuristically argument in Section 5.6 of Barndorff-Nielsen et al. (2008), which shows that

$$\hat{\Sigma} \xrightarrow{p} \int_0^T \sigma_s^2 ds + \sum_{i=1}^{n_J} J_i^2$$

similarly to the RV^* estimator. As previously mentioned Jacod et al. (2009) also shows how the RV^* estimator is very similar to this estimator, and this suggests that we can achieve the same results by expanding the kernel-based estimator with a left-sided kernel.

This could be done with

$$\hat{\Sigma}_t = \gamma_0 + 2\sum_{l=1}^{L(T)} w(l, L(T))\gamma_l \qquad \text{with} \qquad \gamma_l = \frac{1}{h_n}\sum_{i=1+l}^{T} K\left(\frac{t_i - t}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-l} - t}{h_n}\right)\Delta Y_{i-l} \tag{37}$$

where we still have that Y_0 and Y_T are an average of m observations before and after the end-points to eliminate bias. Then if the bandwidth just converges *sufficiently slowly* to zero, it seems reasonable, that the same results hold. We will be more specific about what we mean by *sufficiently slowly* later.

6.6 Simulation study

To test the validity of Proposition 2, Theorem 9 and Theorem 10, we conduct a simulation study. We do this by adding microstructure noise to the simulated paths from Section 5.4. To obtain a realistic and useful simulation study, we first estimate the variance of the noise that we expect to encounter in the data.

6.6.1 Noise parameters

We estimate the microstructure noise on the data from the S&P-500 index. We do this with the estimator from eq. (36) on the full year of 2014. This yields $\hat{\omega}^2 = 2.64 \cdot 10^{-10}$. Hence in the simulation study, we let the microstructure noise be i.i.d. with distribution $\epsilon_i \sim N(0, 2.64 \cdot 10^{-10})$.

6.6.2 Pre-average and bandwidth parameters

For the previous theorems to be true we must have $k_n = \theta_1\sqrt{n}$, $h_n \to 0$ and $\sqrt{n}h_n \to \infty$.

The optimal choice of θ_1 depends on the variance of the noise. The larger θ_1, the smaller the bias from the noise. Unfortunately, the larger θ_1, the slower is the convergence/divergence of the estimators. Hence it is a trade-off between bias and convergence.
Because the variance of the noise is relatively small, we choose $\theta_1 = \frac{1}{10}$. With this choice, we pre-average over very few observations.

We let the bandwidth parameter be given by $h_n = T\theta_2 n^{-\gamma}$ with $0 < \gamma < 1/2$, since this fulfills the conditions $h_n \to 0$ and $\sqrt{n}h_n \to \infty$. The maturity is included in the bandwidth because the size of the bandwidth should depend on the time-scale.

In the choice of θ_2, we notice that a smaller bandwidth yields faster divergence in case of a burst, but slower convergence in the normal case and a larger bias from the noise. Because the divergence is the bottleneck and bias very small, we want a very low bandwidth. Therefore we choose $\theta_2 = 10^{-1}$ and $\gamma = 1/3$.

6.6.3 Convergence of T-estimator

We test the convergence of the T-estimator for the same 4 processes as without noise, i.e.

1) Heston with microstructure noise
2) Heston with microstructure noise + volatility burst
3) Heston with microstructure noise + jump
4) Heston with microstructure noise + volatility burst + drift burst

We use the same simulated paths as in Figure 8 and 9 but add a microstructure noise with the distribution from Section 6.6.1.

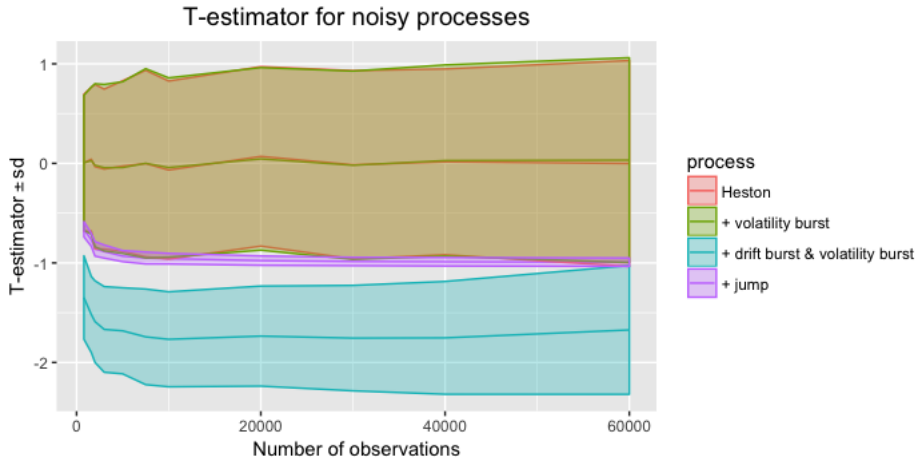


Figure 14: Convergence of the pre-averaged $T-$estimator at peak of event time ($\tau = 3.25$hours) for different processes (with the parameters from Figure 8 added an iid noise with distribution $\epsilon_i \sim N(0, 2.64 \cdot 10^{-10})$). The ribbon is the standard deviation of the estimator. The estimator under *Heston* and *Heston + volatility burst* are almost indistinguishable from each other and are both in the brown ribbon. Note that the drift burst converges because $\beta + 1/2 > \alpha$, but this is extremely slowly.

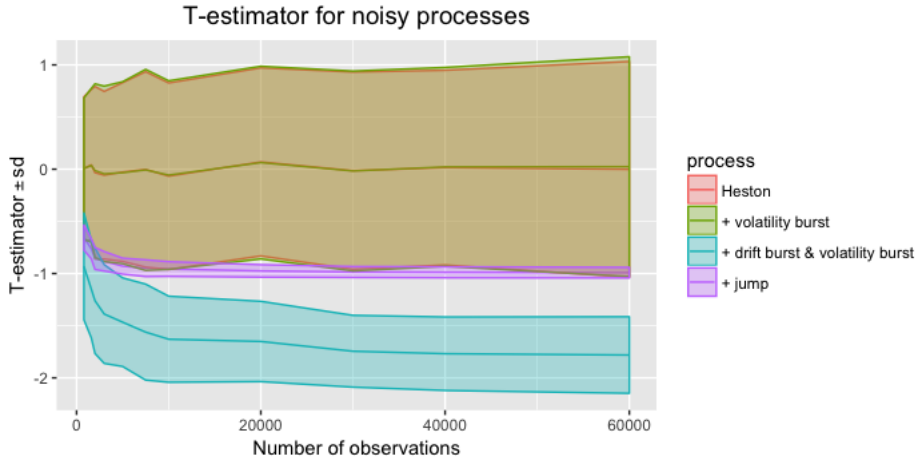


Figure 15: Convergence of the pre-averaged $T-$estimator at peak of event time ($\tau = 3.25$hours) for different processes (with the parameters from Figure 9 added an iid noise with distribution $\epsilon_i \sim N(0, 2.64 \cdot 10^{-10})$). The ribbon is the standard deviation of the estimator. The estimator under *Heston* and *Heston + volatility burst* are almost indistinguishable from each other and are both in the brown ribbon. Note that the drift burst diverges slowly because $\beta + 1/2 < \alpha$, but this is extremely slowly.

The convergence and divergence of the T-estimator can be seen in Figure 10 and 11.

We see that all theorems seems to hold in practice, and we have the convergences and divergence that we expected. Namely, all of the old results - with the exception of a variance of 1 in case of a volatility burst instead of $2^{2\beta}$. However the convergence of the T-estimator in the presence of a drift burst in Figure 14 is much slower similar to the divergence of the estimator in presence of drift burst in Figure 15 which is also much slower. We would of course expect it to be slower, since the bandwidth decreases much slower due to the fact that $\sqrt{n}h_n \to \infty$ is necessary for convergence. However the convergence and divergence is so slow that it might be difficult to use this estimator in practice. We will therefore need to improve the convergence and this will be the focus in the proceeding chapters.

One of the primary reasons for the slow divergence in case of a burst is that the last k_n observations do not have full weight (see Figure 7). Hence we only detect the burst some time after the burst. By then the weight is reduced due to the left sided kernel. This is also the reason why the time point for the jump and burst in Theorem 9 and 10 is $k_n - 1$ points after the event. This *fading out* in the edges had to be done in order to get rid of the noise bias in the end-points, but the question is, whether this is necessary in practice, when the variance of the noise is so small. We will therefore investigate how *damaging* it would be to keep the noise bias in the end in exchange for a faster convergence/divergence and live detection.

7 Deliberate bias creation[7]

One way to keep the noise bias in the edges is by not fading out these in the pre-averaging approach. This could be done by changing the weights of the individual delta terms, such that they are given as in Figure 16 (still with the left sided kernel).
For the kernel-based approach, the bias comes naturally by not averaging over X_n in the estimator in eq. (22).

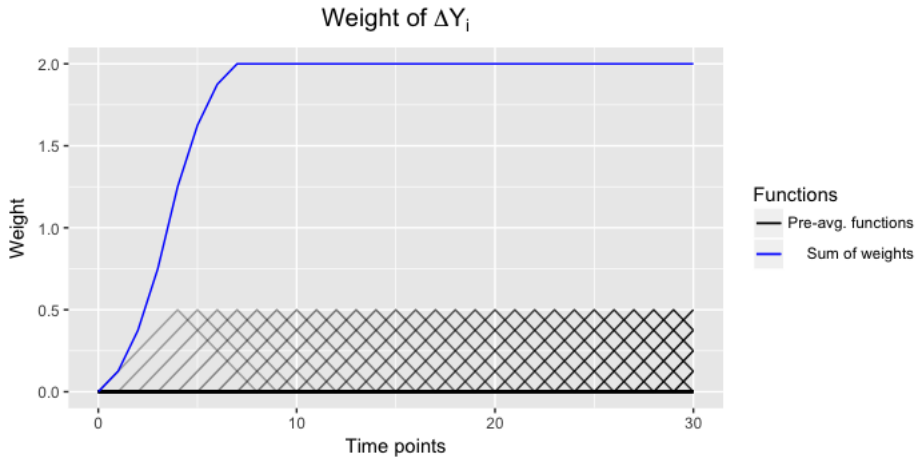


Figure 16: An illustration of how to change the weights on the ΔY_i process to create a bias in the right end, plotted along with the sum of weights. The parameters are the same as in Figure 7.

Note that it does not matter for neither of the estimators, whether we make the correction in the beginning as well or not - the left sided kernel will ensure that the oldest observations are assigned no weight (it follows from the proof of Proposition 4 below). For easier calculations and implementations we will thus choose to introduce the bias in both ends.

7.1 The bias of the drift estimator

For the drift estimator, the terms are simply summed up. This means that the terms all have the same weight and changing the pre-averaging in the ends is actually just a fancy way of multiplying every ΔY and then re-scaling it again with ψ_1/k_n. No change is therefore necessary for the drift estimator.

Proposition 4. *For the biased estimator from (4) we have*

$$\sqrt{h_n}\hat{\mu}_t^n - \frac{1}{\sqrt{h_n}}\epsilon_n \xrightarrow{d} N(0, K_2\sigma_{t-}^2)$$

for $h_n \to 0$ *and* $h_n\sqrt{n} \to 0$.

[7]Responsible: Sebastian

Proof. See Appendix 1. □

This means that in contrary to before, the estimator now has a bias, which is increasing as $h_n \to 0$. Note that we need $h_n\sqrt{n} \to 0$ to avoid additional asymptotic bias from the other noise terms as well.

To get an intuition of the relative bias that this introduces, we first rewrite the convergence as

$$\sqrt{h_n}\hat{\mu}_t^n = N(0, K_2\sigma_{t-}^2) + \frac{1}{\sqrt{h_n}}\epsilon_n + o_p(1)$$

which is just a reformulation of the theorem above. These are all stochastic, so we can write it as

$$\sqrt{h_n}\hat{\mu}_t^n = X_n + \eta_n + \zeta_n \tag{38}$$

with $X_n \sim N(0, K_2\sigma_{t-}^2)$, $\eta_n = \frac{1}{\sqrt{h_n}}\epsilon_n$ and ζ_n describing the error term.
We wish to achieve an intuitive understanding of the size of the bias. We assume the noise process is iid and independent of the X process. We also assume it is normally distributed as in the simulation study. We can then collapse the normal distributions to obtain

$$\sqrt{h_n}\hat{\mu}_t^n = N\left(0, K_2\sigma_{t-}^2 + \frac{1}{h_n}\omega^2\right) + o_p(1)$$

For a clearer overview, we define the *noise ratio*

$$\gamma_t = \frac{\omega_t}{\sigma_{t-}\sqrt{dt}}$$

where $dt = T/n$ and rewrite the variance of the normal distribution (omitting the error term) as

$$K_2\sigma_{t-}^2 + \frac{1}{h_n}\omega^2 = K_2\sigma_{t-}^2 + \frac{\sigma_{t-}^2\gamma_t^2 dt}{h_n}$$

and to obtain the relative difference, we normalize it by $K_2\sigma_{t-}^2$. This yields

$$R_{bias} = 1 + \frac{\gamma_t^2 dt}{h_n K_2}.$$

This means that as long as the term $\frac{\gamma_t^2 dt}{h_n K_2}$ is sufficiently small in practice, the bias from the last noise term is relatively insignificant.
In Christensen et al. (2016) it is mentioned, that the literature suggests a signal-to-noise ratio around 0.5. In our simulation studies we have a ratio of 0.46. The relative bias is small if dt/h_n is small - i.e. we have enough observations per bandwidth. This indicates that the bias is quite small in real data where most days contain more than 100,000 observations, as long as the bandwidth is not on a scale comparable with 1/100,000th of a day.

7.1.1 Simulation study of the bias of the drift estimator

We wish to thoroughly study how significant this bias is for different choices of bandwidths. To avoid unnecessary complexity, the bias is studied in a setting with a constant volatility in the underlying process X. Therefore we let the process in this section have Black-Scholes dynamics with variance equivalent to the long term mean θ of the Heston model. This results in the dynamics of the log-price process as

$$dX = \sigma dW_t$$

with $\sigma^2 = 0.0457$.

As in the previous sections we let the microstructure noise be i.i.d. with $\epsilon_i \sim N(0, 2.64 \cdot 10^{-10})$.

We choose the number of observations to be $n = 23,400$ which is equivalent to one observation per second during the 6.5 hour trading day. This is similar to the simulations in Figure 8 and 9. This is a much lower number of observations than what we encounter in the data. We choose this number of observations to ensure that our conclusions are valid even on scarce data sets.

To investigate how much the individual terms in (38) contribute to the total variance of the drift estimator we will estimate the following:

$$\text{Var}(X_n), \quad \text{Var}(X_n + \eta_n) \quad \text{and} \quad \text{Var}(X_n + \eta_n + \zeta_n).$$

Then we will use the differences in the three variances to enlighten us on how much the individual term contributes to the total variance of the drift estimator. We first normalize the terms, and define $\tilde{X}_n = \frac{X_n}{\sqrt{K_2 \sigma_{t-}^2}}$, $\tilde{\eta}_n = \frac{\eta_n}{\sqrt{K_2 \sigma_{t-}^2}}$ and $\tilde{\zeta}_n = \frac{\zeta_n}{\sqrt{K_2 \sigma_{t-}^2}}$.
This gives us

$$\text{Var}(\tilde{X}_n), \quad \text{Var}(\tilde{X}_n + \tilde{\eta}_n) \quad \text{and} \quad \text{Var}(\tilde{X}_n + \tilde{\eta}_n + \tilde{\zeta}_n). \tag{39}$$

which is equivalent to

$$1, \quad 1 + \frac{\omega^2}{h_n K_2 \sigma_{t-}^2} \quad \text{and} \quad \text{Var}\left(\frac{\sqrt{h_n} \hat{\mu}_t^n}{K_2 \sigma_{t-}} \right).$$

as the first is the normalization constant, the second is the normalization constant plus relative noise contribution and lastly the sum of all terms, which is simply the normalized drift estimator. As we do not know the variance of the error term ζ_n, we estimate $\text{Var}(\tilde{X}_n + \tilde{\eta}_n + \tilde{\zeta}_n)$ by simulating 1000 paths and calculating the empirical variance for the bandwidths bandwidths $h_n = \{30\text{sec}, 60\text{sec}, 90\text{sec}, ..., 600\text{sec}\}$.

Figure 17 shows the three expressions from (39) for various bandwidths. From the difference between the red and blue line, we see, that the noise adds very little to the variance, and makes up a very small relative part of the total variance. On the other hand, we see that for the expression with the error term where ζ_n is included, the variance is up to 0.3 higher than for the expression where it is excluded. From this it follows that the variance of the error term adds 30% to the variance of the normal distribution when the bandwidth is half a minute. Hence it adds a significant bias for small bandwidths.

To validate the theoretical result, Figure 18 shows the same plot but with the variance of the noise increased with a factor 25 and the variance of the X process decreased by the same factor. Now we see from the difference between the blue and red line that the noise term contributes significantly to the variance, which is up to 8 times higher than what is explained by the normal distribution $N(0, K_2 \sigma_{t-}^2)$.
The error term however does not have a large relative influence on the variance (the contribution is the difference between the blue and green line). Hence we clearly see that the bias of the estimator for low bandwidth primarily stems mainly from the last noise term which validates the theoretical result.

To conclude, this means that the bias of the $\hat{\mu}$-estimator induced by the microstructure noise in practice is generally very small.

7.2 The bias of the volatility estimator

In Section 6, the convergence theorems of the T-estimator were shown using the pre-average approach. We have along the way drawn comparisons to the kernel-based estimators from Barndorff-Nielsen et al. (2008)

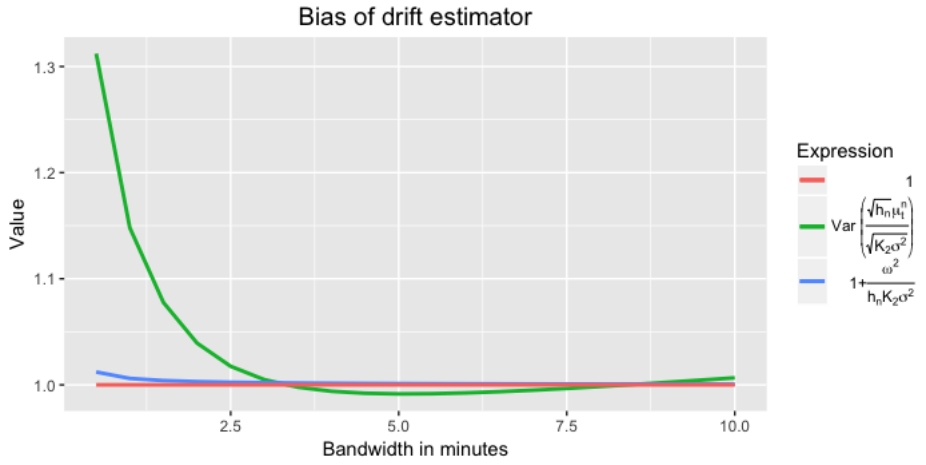


Figure 17: The value of the three terms in eq. (39) for different bandwidths. The parameters are $(\sigma^2, \omega^2, n) = (0.0457, 2.64 \cdot 10^{-10}, 23400)$. This corresponds to the signal-to-noise ratio $\gamma \approx 0.43$. The relative contribution of the last noise ϵ_n on the variance of the drift estimator is the difference between the blue and red line, and we see, that it is very small.

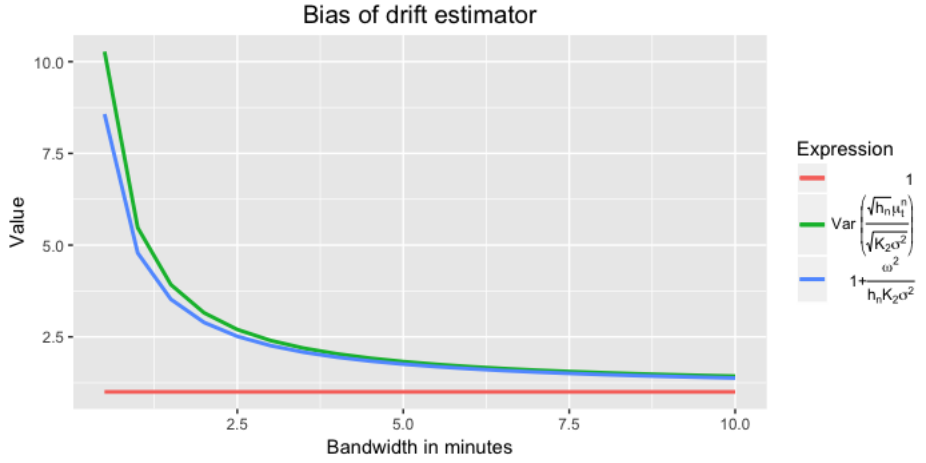


Figure 18: The value of the three terms in eq. (39) for different bandwidths. The parameters are $(\sigma^2, \omega^2, n) = (25 \cdot 0.0457, \frac{1}{25} \cdot 2.64 \cdot 10^{-10}, 23400)$. This corresponds to the quite extreme signal-to-noise ratio $\gamma \approx 10.66$. The relative contribution of the last noise ϵ_n on the variance of the drift estimator is the difference between the blue and red line, and we see, that it is extremely significant. This validates the theoretical result.

and Barndorff-Nielsen et al. (2011), and explained how these are fundamentally very similar and how all results are very plausible to hold for kernel-based estimators as well once a left-sided kernel has been added. In Christensen et al. (2016) the volatility estimator used is a biased version of the estimator from Barndorff-Nielsen et al. (2011) with an added left-sided kernel. To be able to compare our results with the results from Christensen et al. (2016), we will therefore leave the pre-averaging approach, which we used to build the mathematical foundation, and instead switch to the kernel-based approach. This means that although we believe the same convergence results to be true for the kernel-based approach, due to the similarities, they are not proven formally.

We start by summarizing and explaining the estimator in Barndorff-Nielsen et al. (2011) in greater detail. They advocate for the following estimator:

$$\hat{\Sigma} = \gamma_0 + 2\sum_{l=1}^{L(T)} w(l, L(T))\gamma_l \quad \text{with} \quad \gamma_l = \sum_{i=1+l}^{T} \Delta X_i \Delta X_{i-l}.$$

They show in Lemma 1 that if $\frac{L}{\sqrt{n}} \to \infty$, the estimator consistently estimates the quadratic variation given that the kernel is sufficiently nice - namely that it satisfies assumption K in the article. Note the similarity between this convergence rate and the result for the slow converging pre-average estimators which needed $\frac{k_n}{\sqrt{n}} \to \infty$.
Regarding the kernel, they advocate for the Parzen kernel as this fulfills the assumptions needed for consistency, is guaranteed to yield a non-negative estimates, is relatively quick to compute and converges fast. They suggest to let $w(l, L(T)) = w(\frac{l}{L+1})$ with

$$w(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{for } 0 \leq x \leq 1/2 \\ 2(1-x)^3 & \text{for } 1/2 < x \leq 1 \end{cases}$$

This choice is adopted by Christensen et al. (2016) and we will use the same. The weights of the Parzen-kernel for $L = 10$ can be seen in Figure 19.

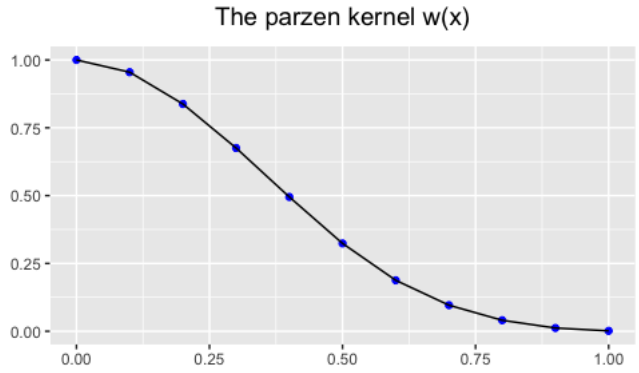


Figure 19: An illustration of the parzen kernel $w(\frac{l}{L+1})$ for $l = \{1, 2, ..., L+1\}$ and $L = 10$.

In Barndorff-Nielsen et al. (2011) the observations X_0 and X_n were also an average of the previous and following observations, but we will neglect this to purposely induce the bias and be able to detect more bursts and in real time as well. As was mentioned before, it follows from Barndorff-Nielsen et al. (2009) that without pre-averaging in the edges, we have

$$\hat{\Sigma}_t^n \xrightarrow{p} \epsilon_0^2 + \epsilon_n^2 + \int_0^T \sigma_s^2 ds.$$

To estimate the instantaneous variance instead of the integrated, we add the left sided kernel (as described earlier), so the estimator is modified to

$$\hat{\Sigma}_t^n = \gamma_0 + 2\sum_{l=1}^{L(T)} w(\frac{l}{L+1})\gamma_l \tag{40}$$

$$\gamma_l = \frac{1}{h_n}\sum_{i=l+1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right)\Delta Y_{i,n} \cdot K\left(\frac{t_{i-l-1} - t}{h_n}\right)\Delta Y_{i-l,n}.$$

That way the kernel is multiplied on the delta sequence exactly as with the pre-average approach, and we therefore expect it to have the same properties (i.e. converging towards $K_2\sigma_{t-}^2$). As mentioned before it was necessary that $L/\sqrt{n} \to \infty$ and for h_n to decrease *slowly*. In the pre-average approach this meant that

$\frac{k_n \cdot}{nh_n} \to 0$, which because of our choice of $k_n \propto n$ was equivalent to $\sqrt{n}h_n \to 0$). Thus, we expect that it is necessary that $\frac{L}{nh_n} \to 0$. This gives us with the following conjecture:

Conjecture 1. *Let $\hat{\Sigma}_t^n$ be given as in (40). We then have*

$$\hat{\Sigma}_t^n - \frac{1}{h_n}\epsilon_n^2 \xrightarrow{p} K_2\sigma_{t-}^2 \tag{41}$$

as $n \to \infty$, $\sqrt{n}/L \to 0$, $h_n \to 0$ and $\frac{L}{nh_n} \to 0$.

Because K_2 is included in the convergence of the volatility estimator we do not need to re-scale the T-estimator with K_2 as previously. Hence we define the T-estimator as

$$T_t^n = \frac{\sqrt{h_n}\hat{\mu}_t^n}{\sqrt{\hat{\Sigma}_t^n}} \tag{42}$$

exactly as in the pre-averaging theory.

We can rewrite the conjecture above as

$$\hat{\Sigma}_t^n = K_2\sigma_{t-}^2 + \frac{1}{h_n}\epsilon_n^2 + o_p(1)$$

and see the similarity between this expression and the expressions for the drift estimator. The conjecture is that this estimator has same bias expressions as the variance of the drift estimator. While this fact might seem of little importance because of the small size of the noise, it is in fact extremely vital. The reason is that although the term $\frac{1}{h_n}\epsilon_n$ might be small relatively to $K_2\sigma_{t-}^2$, the bias is not irrelevant in the jump case. Naturally in the latest value of ΔY, it is impossible to distinguish between which part of the change came from a microstructure noise, and which part came from the jump. Theoretically the jump also appears in the exactly same way in the proofs as ϵ_n at jump time, hence in the jump case, the last noise ϵ_n is replaced by $\epsilon_n + J_\tau$, and this term is most likely to be highly significant! It is therefore important that the bias in the drift- and volatility estimator cancels out such that the jumps will not be detected.

7.2.1 Lag selection

Before we begin the simulation study of the bias, we need to determine the number of lags to use. Finding the optimal number of lags is a comprehensive procedure. Others have already investigated the subject with variance estimators and have come up with well-founded lag selection algorithms. Among these are Newey & West (1994) and Barndorff-Nielsen et al. (2008).

There is a major difference between these estimators and the estimator in (40) as our estimator contains the extra left sided kernel K in the variance estimator, and therefore estimates the instantaneous variance and not the quadratic variation. The lag selection algorithms cannot be directly transformed to our scenario as the kernel changes the effective number of observations and has an effect on the convergence.
Finding the optimal number of lags is beyond the scope of this thesis. Christensen et al. (2016) achieve good results with the number of lags in the range 7-12 and they choose $L = 10$ in the practical study, which is why we will also use $L = 10$ in the following - unless stated othewise.

7.2.2 Simulation study of bias in the new T-estimator

To see how well the volatility estimator and new T-estimator mimics the conjecture, we will create a similar study as for the drift estimator.

To investigate what influence the noise has on the T-estimator, we simulate the same paths as in Figure 17 and 18.

As with the drift estimator, we can write

$$\hat{\Sigma}_t^n = E(X_n^2) + \eta_n^2 + \vartheta_n$$

with $X_n = N(0, K_2\sigma_{t-}^2)$ and $\eta_n = \frac{1}{\sqrt{h_n}}\epsilon_n$ as in the drift estimator, and ϑ_n is the error term. We again normalize it with $K_2\sigma_{t-}^2$ to get

$$\frac{\hat{\Sigma}_t^n}{K_2\sigma_{t-}^2} = E(\tilde{X}_n^2) + \tilde{\eta}_n^2 + \tilde{\vartheta}_n.$$

Note that the first term is 1 as we normalized with this, which exactly coincides with the variance of the first term in the drift estimator. The second term is also exactly the square of the second term in the drift. The only thing we do not know the true size of is the error term $\tilde{\vartheta}_n$.
We will plot

$$E(\tilde{X}_n^2), \quad E(\tilde{X}_n^2 + \tilde{\eta}_n^2) \quad \text{and} \quad E(\tilde{X}_n^2 + \tilde{\eta}_n^2 + \tilde{\vartheta}_n) \tag{43}$$

equivalent to

$$1, \quad 1 + \frac{\omega^2}{h_n K_2 \sigma_{t-}^2} \quad \text{and} \quad E\left(\frac{\hat{\Sigma}_t^n}{K_2\sigma_{t-}^2}\right)$$

to investigate how much the volatility estimator increases when the individual terms are added. This allows us some insight into the contribution of the individual terms on the volatility estimator. Notice that the first to terms are equivalent to the first to terms in the drift analysis.

Figure 20 shows the expressions from (43) in the normal case with the simulations from Section 5.4. It also contains the variance of the drift estimator - the third variance in (39) and the T-estimator.
There are several important things to notice. We see from the difference of the brown and red line in Figure 20 that the relative contribution in the volatility estimator from the noise term is insignificant. This is natural as it is the same as for the variance of the drift estimator.
The relative contribution of the error term is the difference between the blue and brown line and clearly significant.
A very interesting fact is that the error term's contribution to the volatility estimator is extremely similar to the contribution in the variance for the error term of the drift estimator. This might seem like a lucky coincidence, but the error terms consist of the same parts (namely the product terms between the noise and X-process - the other noise terms besides ϵ_n etc.). This is evident from the proofs. This leads to the T-estimator to still have a variance close to one - actually slightly below as seen in the Figure. This bodes very well for the estimator.

Figure 21 shows the same, but as in the drift study, the noise variance is increased by 25 and the variance of the process decreased with the same factor. This (unrealistic) picture shows that the conjecture seems to be valid for the volatility estimator. In fact the bias from the drift- and volatility estimator are closely aligned such that the line from the drift estimator cannot be seen. The differences are on the 2nd decimal.
In this study with the increased variance of the noise, we have increased the number of lags to $L = 100$, as the bias from the other noise observations would otherwise also have been increasingly significant. This is seen as with k_n in the pre-averaging approach, we need to average over more observations if the noise has higher variance.

7.3 Convergence of estimator

We have validated that both the theorem regarding the bias of the drift estimator and conjecture regarding the bias of the volatility estimator hold. We have seen that the noise has a very limited contribution in

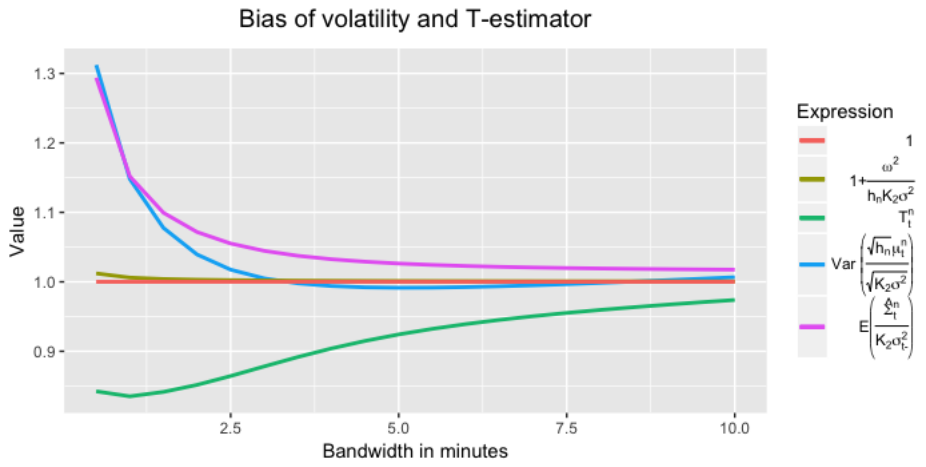


Figure 20: The parameters and three terms from Figure 17 with the addition of the last term in eq. (43) and the variance of the T-estimator. We see, that contribution of the error term in the volatility estimator (the difference between the brown and pink line) is very closely related to the variance contribution of the error on the drift estimator (the difference between the brown and blue line).

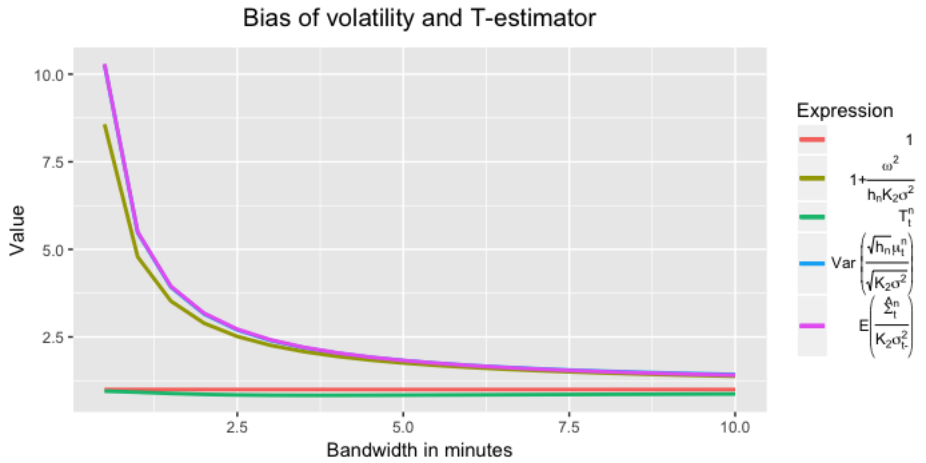


Figure 21: The parameters and three terms from Figure 18 with the addition of the last term in eq. (43) and the variance of the T-estimator. The average of the volatility estimator (the pink line) is so closely aligned with the variance of the drift estimator (the blue line), that the latter is unnoticeable. We see that the error contributes the most to the variance estimator for small bandwidths (the difference between the brown and red line). This strengthens the conjecture.

practice and that the estimator behaves as expected with deviations from the convergence results canceling out.
Next we will investigate the how close the convergence results in jump and burst case are in regards to the noise bias as we have not proven these for the kernel-based estimator.
Figure 22 and 23 show the convergence and divergence of the T-estimator for the same paths as used in the previous chapters. It indicates that we obtain the exact same convergence results as in the pre-averaging case. It also shows that the value of the T-estimator in presence of bursts is higher than in the pre-averaging study but lower than in the simulation without microstructure noise.

Though the convergence result seem identical to the results with pre-averaging, we must remember that they in fact are not due to the bias. As $h_n \to 0$ we will in all cases have $T_t^n \xrightarrow{p} \text{sign}(\epsilon_n)$, however this is so far out

in the tail that it is irrelevant in practice.

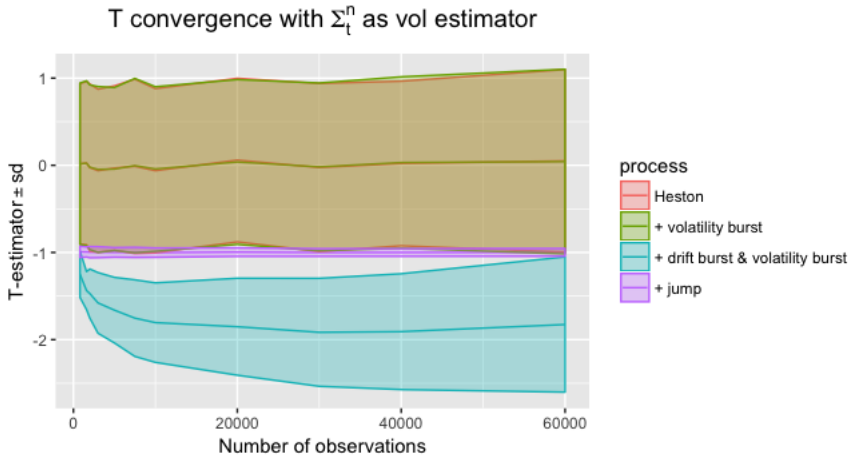


Figure 22: Convergence/divergence of the estimator from (40) with the same simulated paths as in 14. We see that the finite values for the drift burst are higher than when pre-averaging but below the values without microstructure noise.

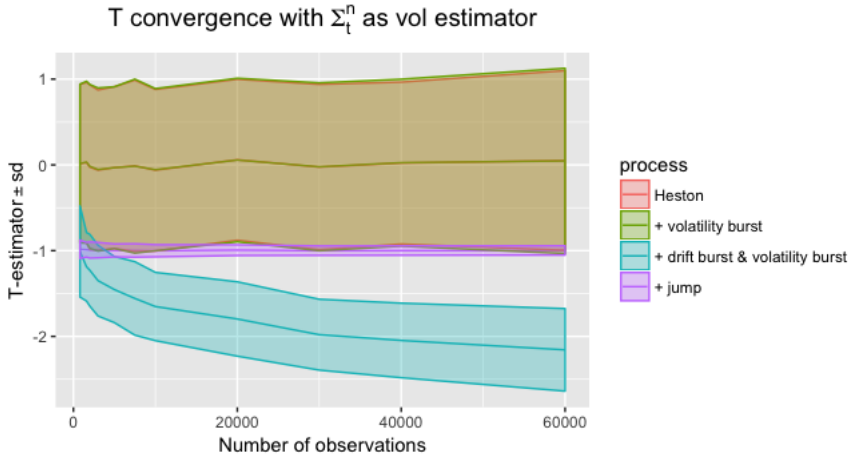


Figure 23: Convergence/divergence of the estimator from (40) with the same paths as in Figure 15. We see that the finite values for the drift burst are higher than when pre-averaging but below the values without microstructure noise.

Figure 24 shows a QQ-plot of T_t^n for three different bandwidths. We see that it is very close to being normally distributed as hoped. It is slightly conservative, which is preferred over too heavy tails.
We conclude that the estimator behaves just as expected with a little faster divergence in the drift burst case. Besides it has the advantage that the bursts are detected *live* when we accept the noise bias and we thus keep this setting with the biased estimators in the further analysis.

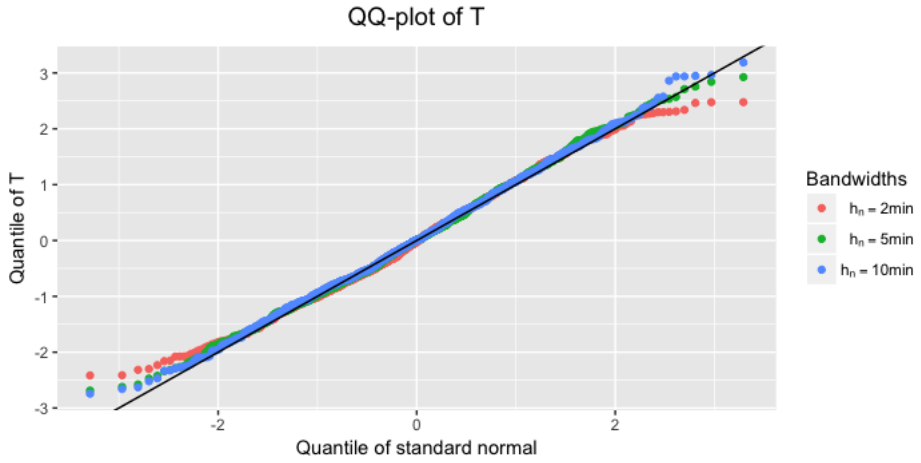


Figure 24: QQ-plot of the T-estimator from eq. (42), with the simulated Heston paths from Section 6. It is very close to be normally distributed with slightly thinner tails.

8 Multiple testing problem[8]

We have now seen that the bias in the T-estimator is very small and the T-estimator is very close to being normally distributed for various bandwidths. We will use this result to create a quantile such that we, with a certain probability, can say when we detect a burst. However there is a problem with the burst detection which is yet unaccounted for. We will most likely have very few bursts in the market, but if we set the threshold for a burst at the 95% quantile in the normal distribution, we will on average falsely detect a drift burst every 20^{th} time. As we want to calculate T at least once per minute, we will therefore on average falsely detect a drift burst more often than every half hour. However we do not expect to find a burst that often. We have directly from from Bayes formula that the conditional probability of a burst conditioned on $T > q_{0.95}$ will still be extremely small. This is very undesirable as the 'burst' we *detect* are mostly false detection.

To overcome this, we will set the threshold such that the probability that there is not *any* T above the threshold during the entire day is 95%. With this we will only expect to have false detections every 20^{th} day, which is much more desirable. This is equivalent to setting the quantile q such that

$$P\left(\max_{t_i \in \text{day}} |T^n_{t_i}| = q\right) \geq 95\%$$

We can find this quantile as Appendix B in Christensen et al. (2016) by defining

$$T^*_m = \max_{t_i} |T^n_{t_i}|, \quad i = 1, ..., m.$$

We then have as $m \to \infty$ that

$$(T^*_m - b_m)a_m \xrightarrow{d} \xi$$

where

$$a_m = \sqrt{2\ln(m)}, \quad b_m = a_m - \frac{1}{2}\frac{\ln(\pi \ln(m))}{a_m}$$

and ξ is the standard Gumbel distribution.

[8]Responsible: Sebastian

For a given number of calculations of T per day, we can with the Gumbel distribution find the asymptotic quantile such that the probability that all T-estimations are below the quantile during the day is 95% under the null hypothesis of a Heston model.

This means that it is possible to control the number of false detections by using the Gumbel distribution with a specific m. We will in this thesis choose m as the number of T-calculations per day. We will do this, because this gives us a 5% probability of having a T value above the quantile during the day. We see that the event $T^* > q$ will occur every 20^{th} day due to a false detection.

8.1 Issue with finite sample and high correlation

According to Christensen et al. (2016) this distributional result holds for covariance-stationary normal random variables with weak serial dependence. However in finite samples the asymptotic result might not be close to the actual quantile when there is a significant correlation between the variables - which is indeed the case with the T-estimator

Christensen et al. (2016) argues that the auto-correlation function decays very similarly to an AR(1) process with positive auto-regressive coefficient, and that for small samples or when there is a significant auto-correlation is much better to compare the estimator against the quantiles for the distribution of Z_i with

$$Z_i = \rho Z_{i-1} + \epsilon_i, \quad i = 1, ..., m$$

where $|p| < 1$ and $\epsilon_i \sim N(0, 1-\rho^2)$. Note that this has the stationary solution $Z_i \sim N(0,1)$.
To account for dependencies in T, it is better to estimate ρ in the AR(1) process, and use the quantile from this process instead. Because the process Z_i is also standard normal and mixing, $Z_m^* = \max_i |Z_i|$ will also have the limiting distribution $(Z_m^* - b_m)a_m \xrightarrow{d} \xi$. Hence the distribution also converges to the Gumbel and in the limit there is no difference between evaluating T^* against the Gumbel or Z^*.

In the estimation of ρ, we choose to diverge from the path taken by Christensen et al. (2016). The reason is that in the statistical model defined by the family of probability measures $(P_\rho)_{\rho\in\Theta}$ with P_ρ given by $Z_i = \rho Z_{i-1} + \epsilon_i$, where $\epsilon_i \sim N(0, 1-\rho^2)$ and i.i.d., the maximum likelihood estimator is not given by the OLS. This comes from the fact that the variance of the noise is determined by ρ and not independent of the correlation. We will therefore derive the MLE below.

The likelihood function in an AR(1) process is given by

$$\log L(\rho, \sigma^2) = \log\left(\prod_{t=1}^{T} f(X_t|X_{t-1})\right) =$$

$$-\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(X_t - \rho X_{t-1})^2.$$

Replacing σ^2 with $1-\rho^2$ gives the log-likelihood

$$\log L(\rho) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\sum_{t=1}^{T}(X_t - \rho X_{t-1})^2$$

and differentiating this equation and setting it to zero gives the third order equation

$$-\frac{T}{2}\frac{1}{1-\rho^2}\cdot(-2\rho) - \frac{1}{2(1-\rho^2)}\sum_{t=1}^{T}2(-X_{t-1})(X_t - \rho X_{t-1}) +$$

$$\frac{1}{4(1-\rho^2)^2} \cdot (-4\rho) \sum_{t=1}^{T} (X_t - \rho X_{t-1})^2 = 0.$$

We can rewrite this as

$$(1-\rho^2)T\rho + (1-\rho^2) \sum_{t=1}^{T} X_{t-1}(X_t - \rho X_{t-1}) - \rho \sum_{t=1}^{T} (X_t - \rho X_{t-1})^2 = 0.$$

Looking at the first two terms, these converge to zero as $\rho \to -1$ and $\rho \to 1$, however the last term converges to a positive and negative number respectively. Therefore the third degree polynomial always has a solution in the interval $-1 < \rho < 1$, and since the polynomial is positive in the left end and negative in the right, the solution corresponds to a maximum of the likelihood function inside the interval.
The solutions for the third degree polynomial is found analytically and there are two possible outcomes. The first one is that there is only one real solution inside the interval $[-1, 1]$, in which case this i obviously the optimum.
The other possible outcome is three real solutions in the interval $[-1, 1]$, in which case the maximum is chosen.

Summarizing, the procedure for detecting bursts in the subsequent sections will thus be:

1) Calculate T at every time point.
2) Estimate $\hat{\rho}$ based on the values of T.
3) Use Monte Carlo to calculate multiple values of $Z^* = \max(Z_i)$ where Z_i is AR(1) with correlation $\hat{\rho}$.
4) Approximate the daily quantile of $q_{0.95}$ with the empirical quantile $P(Z^* > q_{0.95})$.
5) Use the value of $q_{0.95}$ as threshold for T^* in burst detection.

9 Bandwidth investigations

So far the bandwidth parameter that is used in the drift- and volatility estimator has not been discussed. As we will see in this section, this is a key parameter in the estimator, and tuning this is essential for optimal burst detection.

9.1 Theory for separate bandwidths

Throughout the thesis, the bandwidth in the drift- and volatility estimator has been the same. This is not a strictly necessary. Christensen et al. (2016) choose a bandwidth for the volatility estimator which is 5 times higher than the bandwidth for the drift estimator. We will discover the consequences of this later.
Changing the bandwidth parameter is not harmless as some convergence results changes, which is important to take into account.

We denote the ratio between the bandwidth $C = \frac{h_\Sigma}{h_\mu}$ with h_Σ being the bandwidth of the volatility estimator and h_μ the bandwidth of the drift estimator. With this notation, we have the following updated results:

Heston
For Proposition 2 we have the same result under the pure Heston model that

$$T_t^n \xrightarrow{d} N(0,1).$$

Jump
In the presence of a jump with size J at time t_τ, we have the updated result

$$T_\tau^n \xrightarrow{p} \sqrt{C} \cdot \text{sign}(J)$$

Drift and/or volatility burst
In the presence of a burst with peak at time t_{db}, the following holds

$$\begin{aligned} |T_{\tau_{db}}| &\xrightarrow{p} \infty \qquad \text{for } \beta + 1/2 < \alpha \\ T_{\tau_{db}} &\xrightarrow{d} N(0, C) \qquad \text{for } \beta + 1/2 > \alpha \end{aligned}$$

and finite values in presence of a drift burst is a factor of $\approx \sqrt{C}$ higher.

All of the results above follows from the proofs for identical bandwidth by replacing the volatility bandwidth to $h_\Sigma = Ch_\mu$.
The consequence of these results is that increasing the ratio C results in a higher detection rate of both drift bursts, volatility bursts and jumps. However since the probability of falsely detecting a jump as a burst is extremely small, it is possible to increase the ratio C to detect more busts, while still maintaining the probability of detecting a jump as a burst relatively low as long as $\sqrt{C}$ is sufficiently below the threshold (we need to watch the upper bound of the jump). In next section we will give an intuition of what consequences different bandwidth will have on burst detection

9.2 The effects of separate bandwidths

To investigate the influence that the use of separate bandwidths has on bursts, we simulate a wide variety of different drift- and volatility bursts. We will try to replicate the setup in Table 1 in Christensen et al. (2016) as close as possible such that the results can be compared.

The X-process, time span, burst duration and microstructure noise are all similar to that in Section 5.4, and the burst parameters examined are $\alpha \in \{0.55, 0.65, 0.75\}$ and $\beta \in \{0.1, 0.2, 0.3, 0.4\}$. The constant c_1 is set such that the integrated drift bursts for the various parameter choices for α are respectively 0.5%, 1% and 1.5%. The constant c_2 is set such that the volatility bursts correspond to an increase in the average standard deviation over the burst period of respectively 25%, 50%, 75% and 100%. For the volatility burst, the average standard deviation over the interval is calculated as

$$\sigma_{avg} = \left(\int_{\tau_{db} - 10\text{min}}^{\tau_{db}} \theta dt \right)^{1/2}$$

where θ is the long term mean of the volatility in the Heston model from Section 5.4. This gives 0.093%.
Note that though the burst sizes are similar to the ones in Table 1 in Christensen et al. (2016), the functional form seem to be a bit different as the constants c_1 and c_2 do not match.

We calculate T every 5^{th} second and calculate T^* as the maximum over the whole day. This choice of approach is chosen such that the study mimics the approach that we intend to use on the real data.
Again to compare the results to Christensen et al. (2016), we choose to do the study for three different bandwidths for the drift estimator - 2min, 5min and 10min.

9.2.1 Same bandwidth for the estimators

We first investigate the detection of bursts, when the volatility bandwidth is the same as the drift bandwidth. The detection rates in percentage can be seen in Table 3. We find that we get almost zero false detection in the pure Heston model (respectively 0, 0.4% and 2% for bandwidth 2min, 5min and 10min). We also have less than 5% detection for all pure volatility bursts - this will be justified in next section.

Regarding the bursts, we detect most bursts for a high bandwidth. In the examination of the bias, we saw that variance of both the drift- and volatility estimator was significant for low bandwidths, but that this was not due to the bias from the last noise variable. It comes from the fact that for a too low bandwidth, the left sided kernel results in too few observations with significant weight. With this, the terms are too far from the convergence results and e.g. terms that from the proofs converge in probability to zero still have a significant effect in finite sample. It has the same effect on the drift and volatility estimator, and this is why they are similar (in the extreme case with only a single observations, we would have $T = 1$). This is why a higher bandwidth detects more as the distributions have converged more. It must not be too high however, because then we would average over too many observations and the burst will therefore not be detected. In other words, the higher bandwidth the less instantaneous estimates.
Notice how we detect very few bursts, much fewer than in Christensen et al. (2016). As they in the practical study of SPY only detects a burst about every second week in the financial markets with far better simulation results, we would expect to find close to zero bursts in the market with these settings.

Table 3: $P(T^* > q_{95})$ for various burst settings (using the algorithm from Section 8) with same bandwidth for the drift- and volatility estimator. The parameters for the drift bursts are $(c_1, \alpha) = \{(0.299, 0.55), (0.157, 0.65), (0.057, 0.75)\}$, and for the volatility bursts $(c_2, \beta) = \{(0.016, 0.1), (0.009, 0.2), (0.004, 0.3), (0.001, 0.4)\}$.

			Drift burst		
Bandwidths	β	**Heston**	$\alpha = 0.55$	$\alpha = 0.65$	$\alpha = 0.75$
$\mu = 2$min, $\sigma = 2$min	0	0	7.7	31	37.3
	0.1	0	1	11.9	17.5
	0.2	0	0	3.5	6.3
	0.3	0	0	0.8	2.1
	0.4	0	0	0.3	1.4
$\mu = 5$min, $\sigma = 5$min	0	0.4	51.7	89.3	94.4
	0.1	0.4	30.1	77.3	86.8
	0.2	0.4	14.4	59.5	72.9
	0.3	0.4	8	41.4	59.1
	0.4	0.4	5.4	35.6	52.5
$\mu =10$min, $\sigma = 10$min	0	2	64.7	94.3	97.1
	0.1	2	45.5	87.8	93.5
	0.2	1.9	26.4	77.3	86.5
	0.3	1.9	17.6	65.3	78
	0.4	1.9	14.7	60.1	74.8

9.2.2 Bandwidth ratio of 5

We subsequently do the exact same study with a bandwidth ratio of 5 (similar to the one used by Christensen et al. (2016)). We have argued that with a bandwidth ratio above one, we could not be sure not to detect pure volatility bursts. This is not visible from Table 1 in Christensen et al. (2016), and we will here show why this is so.
With a bandwidth ratio of 5, we get the results seen in Table 4. We see that we detect almost all drift bursts but no volatility bursts, i.e. same conclusion as in Christensen et al. (2016). We actually detect even more bursts than in Christensen et al. (2016), but this is most likely due to a different functional form of the bursts.

In this simulation study the volatility is increased up to 100% more than the *average volatility over the period*, and though this seems high, it corresponds to a mere total standard deviation of $\sigma_{avg} = 0.186\%$ over the of 10min interval compared to the usual 0.093%. This is overshadowed by even the tinniest of the drift bursts, where the price increases with 0.5% during the same period. Hence the drift and volatility component seem to be disproportional, which distorts the conclusion.
If we instead set c_2 such that the volatility contributes 8 times more to the average volatility over the period,

Table 4: $P(T^* > q_{95})$ for various burst settings (using the algorithm from Section 8) when the volatility estimator has a 5 times higher bandwidth than the drift estimator. The burst parameters are as in Table 3.

			Drift burst		
Bandwidths	β	**Heston**	$\alpha = 0.55$	$\alpha = 0.65$	$\alpha = 0.75$
$\mu = 2$min, $\sigma = 10$min	0	0.1	75	99.4	100
	0.1	0.2	56.7	98.1	99.8
	0.2	0.5	36.2	95.5	99.4
	0.3	0.5	23.4	91	98.3
	0.4	0.5	20.2	85.8	97.3
$\mu = 5$min, $\sigma = 25$min	0	1.3	87.2	99.7	100
	0.1	1.5	77.6	99.5	99.9
	0.2	2.3	66.4	99.3	99.9
	0.3	3	58.2	98.2	99.7
	0.4	3.2	55.8	97.5	99.7
$\mu =10$min, $\sigma = 50$min	0	2	83.8	99.7	100
	0.1	2.3	76.8	99.5	100
	0.2	3.2	71	99.3	99.9
	0.3	4.4	64.7	98.2	99.7
	0.4	4.7	62.2	98.2	99.7

we get $\sigma_{interval} \in [0.279\%, 0.837\%]$. We find this to be more comparable to the drift bursts. With these, the standard deviation in the smallest and largest volatility bursts are close to half of the total burst in the smallest and largest drift burst. This means that the volatility burst will overshadow the drift burst $\approx 5\%$ of the times - the total change from the drift burst will still dominate 95% of the times but not always.

With these parameters we get the results illustrated in Table 5. Here it is evident that the estimator also detects volatility bursts, which is in line with the theoretical results. Hence a large ratio can inflate the value of the T-estimator such that we detect both volatility- and drift bursts.

This shows, that it is not possible to differentiate a drift burst from a pure volatility burst, unless the bandwidth ratio is 1.

9.3 The beginning of the day

If we abandon the hope of separating the drift bursts and volatility bursts, we can investigate which bandwidth ratio yields the highest detection rate of bursts while still not falsely detecting jumps (as this convergence was affected by the bandwidth ratio).
In order to do this analysis, we need to handle an issue with the bias in the beginning of the day.

The convergence results for the distribution of both the $\hat{\mu}$-estimator and $\hat{\Sigma}$-estimator are based on the following convergence result

$$\frac{dt}{h_n}\sum_{i=1}^{n} K\left(\frac{t_i - t_n}{h_n}\right)^2 \to \int_{-\infty}^{0} K(x)^2 dx$$

as $nh_n \to \infty$. This is simply the Riemann approximation of the integral

The $-\infty$ comes from the idea that we have observations *infinitely in the past*. Although this is not the case in practice, it does not matter much as the value of $K(x)^2$ decays very fast, depending on the bandwidth.

Table 5: $P(T^* > q_{95})$ for the same setup as in Table 4 but where standard deviation over the interval coming from the volatility burst is increased by a factor 8. This corresponds to the volatility burst parameters $(c_2, \beta) = \{(0.13, 0.1), (0.075, 0.2), (0.031, 0.3), (0.010, 0.4)\}$

			Drift burst		
Bandwidths	$\boldsymbol{\beta}$	**Heston**	$\alpha = 0.55$	$\alpha = 0.65$	$\alpha = 0.75$
$\mu = 2$min, $\sigma = 10$min	0	0	73.9	99.7	99.9
	0.1	0.6	3.1	33.4	71.5
	0.2	1.6	2.2	6.6	16.8
	0.3	2	2.3	4.6	8.6
	0.4	2.1	2.4	4.9	7.5
$\mu = 5$min, $\sigma = 25$min	0	1.3	87.4	99.9	100
	0.1	6.2	23.6	76.5	96.2
	0.2	11.6	18.6	38.6	58.6
	0.3	15.6	19.5	30.5	42.6
	0.4	16.5	20.5	31.6	40
$\mu =10$min, $\sigma = 50$min	0	3.1	84.9	99.8	100
	0.1	11.6	35.6	84.1	97.3
	0.2	19.2	28.8	52.2	72.3
	0.3	24.3	30.5	43.6	55.1
	0.4	25.1	32.1	44.5	53.7

However in the beginning of the day, the lack of observations in the past becomes apparent - especially with longer bandwidths. It is important to notice the difference between the number of observations, which we have many of, and the time of observations, which is very much limited by the time of the day. Ten minutes into the trading day, we can only have observations at a maximum of ten minutes in the past - which is not enough for the kernel to sufficiently decay with our choices of bandwidths. To account for this, better approximation of the distribution are

$$\sqrt{h_n}\hat{\mu}_t^n \approx N\left(0, \sigma_{t-}^2 \cdot \int_{\frac{t_{\text{beginning}}-t}{h_n}}^{0} K(x)^2 dx\right)$$

and

$$\hat{\Sigma}_t^n \approx \sigma_{t-}^2 \cdot \int_{\frac{t_{\text{beginning}}-t}{h_n}}^{0} K(x)^2 dx$$

Note that for $h_n \to 0$ we have $\int_{\frac{t_{\text{beginning}}-t}{h_n}}^{0} K(x)^2 dx \to \int_{-\infty}^{0} K(x)^2 dx$, and then by Slutsky's, we have the same convergence results. But for time points with $\frac{t_{\text{beginning}}-t}{h_n}$ too close to zero (time points close to the beginning of the day), the approximation is very poor.
One solution to this is not to calculate it in the beginning of the day, where the kernel has not sufficiently decayed. This solution is what is chosen in Christensen et al. (2016). They wait a full volatility bandwidth before estimating T. With a bandwidth of 5 and a drift bandwidth of 5min, this corresponds to waiting 25min. This is undesirable, especially on SPY, because there is a high trading volume and a lot of activity in the beginning of the day. It would be sub-optimal not to be able to test for bursts before 25min into the day, and even worse, if we want to have a higher bandwidth ratio.

We have come up with a new way to accommodate the bias in the beginning of the day. Instead of discarding the estimations in the beginning of the day, they will be re-scaled such that

$$(\hat{\mu}_t)_{\text{new}} = \frac{\sqrt{\int_{-\infty}^{0} K(x)^2 dx}}{\sqrt{\int_{\frac{t_{\text{beginning}}-t}{h_n}}^{0} K(x)^2 dx}} \cdot (\hat{\mu}_t)_{\text{old}}$$

and

$$(\hat{\Sigma}_t)_{\text{new}} = \frac{\int_{-\infty}^{0} K(x)^2 dx}{\int_{\frac{t_{\text{beginning}}-t}{h_n}}^{0} K(x)^2 dx} \cdot (\hat{\Sigma}_t)_{\text{old}}.$$

With this re-scaling, the T-estimator can be calculated earlier on the day. Instead of waiting until a bandwidth has passed, we now only need to wait a sufficient number of observations such that the convergence results are good approximations. This is much more desirable as we have a lot of observations in the beginning of the day.

Figure 25 shows $P(|T| > 95\%)$ during the day both with and without re-scaling for a bandwidth ratio of 5. We see that without re-scaling the bias is significant up to more than half an hour into the trading, whereas the re-scaling removes the bias extremely well, and the bias is actually slightly on the conservative side in the very beginning of the day, which is to prefer. In the figure, only the first minute of estimates are burned. Note how the graphs align as the re-scaling multiplier converges to one.

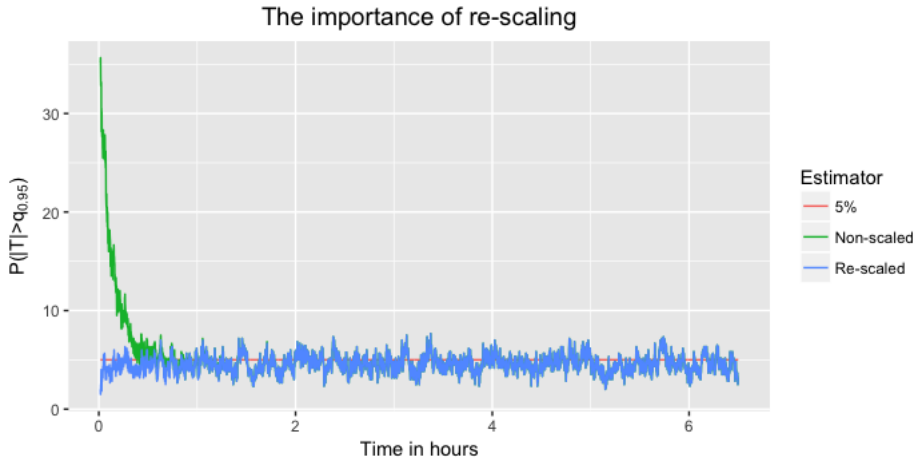


Figure 25: The percentage $P(|T| > q_{0.95})$ during the day. We see, that the non-scaled estimator has a large bias in the beginning, whereas the re-scaled estimator does not but is slightly conservative. The bandwidth parameters are $h_\mu = 5$min and $h_\Sigma = 25$min as in Christensen et al. (2016).

We conclude that with re-scaling we do not have to worry about the bias in the beginning of the day and burning only 60 observations seems to be enough.

9.3.1 How to choose the optimal ratio

The theory suggested that both the bursts and jumps are more often detected when the bandwidth ratio is increased.
Figure 26 shows $P(T^* > q_{0.95})$ for various processes at different bandwidth ratios. As expected, the probability increases as the bandwidth ratio increases. This shows that it is desirable to increase the bandwidth ratio as much as possible to enhance the detection of bursts while keeping the risk of falsely detecting jumps as bursts at a tolerated level.

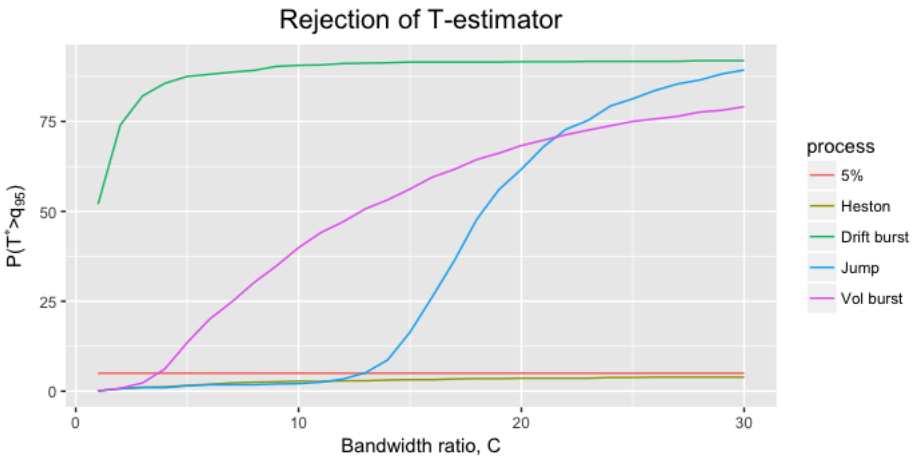


Figure 26: $P(T^* > q_{0.95})$ for different bandwidth ratios. The bandwidth of the drift estimator is $h_\mu = 5$min, and the T-estimator is calculated every 5 second throughout the day. The drift burst parameters are $(c_1, \alpha) = (0.299, 0.55)$ corresponding to the total size of 0.5%, and the jump is of the same size. The volatility burst parameters are $(c_2, \beta = (0.75, 0.2)$ corresponding to the volatility burst in Table 5.

It turns out that the maximum bandwidth ratio possible, without falsely detecting jumps as burst, is dependent on the bandwidth of the drift estimator. This is due to the fact that the finite sample distribution of the estimator in the normal case is slightly influenced by the bandwidth ratio. It is therefore necessary to optimize the bandwidth ratio for every bandwidth of the drift individually.

9.4 The optimal drift bandwidth and bandwidth ratio

We find the maximum bandwidth ratio possible (without falsely detecting jumps) for the bandwidths 2min, 5min and 10min as these were the bandwidths investigated in Christensen et al. (2016). Figure 27 shows the rejection percentages for three different jump sizes (0.5%, 0.75% and 1%) and the three different bandwidths. We find that with a bandwidth of 2min, it is possible to use a bandwidth ratio of 15 without falsely detecting jumps, with 5min a ratio of 12 and with 10min a ratio of 10.

Now to find the best of these combinations, we create a table with the same burst sizes as in Section 9.2, but with the increased volatility bursts, and investigate which bandwidth detects the most bursts. This can be seen in Table 6.
We see from the Table that there is very little difference between the individual choices. We might detect slightly more combinations of volatility and drift burst with the parameter tuple $(h_\mu, h_\Sigma) = (5min, 60min)$. We therefore choose these parameters for the estimator.

9.5 Jump detection

We found in the previous section that the optimal bandwidth for the drift estimator was 5min and the optimal bandwidth ratio is 12, and saw that with this, we do not detect any false jumps with size either 0.5%, 0.75% or 1%. This is not enough to rule out jump detection. As we found out in Section 5.1.1 that there exists both jumps sizes, and time points after a jump with an increased probability of false detection (in the direction of the jump).
As a last check for jump robustness, we need verify two things:

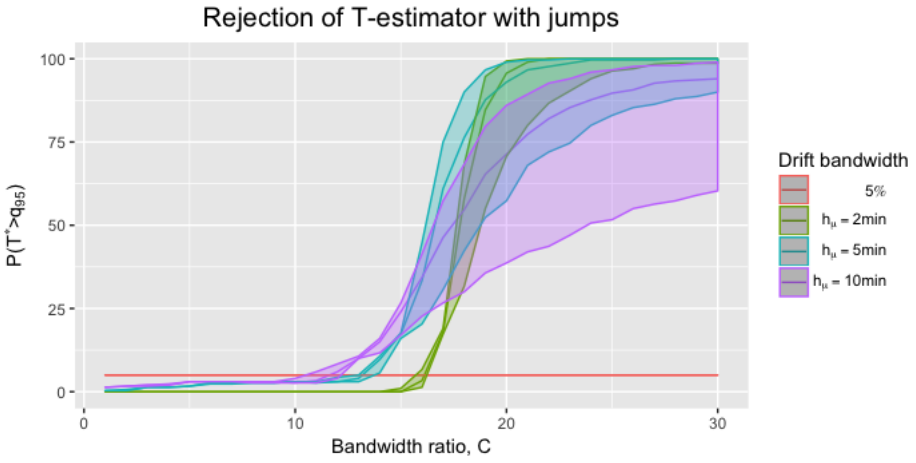


Figure 27: $P(T^* > q_{0.95})$ for different bandwidth ratios, jump sizes and bandwidths of the drift estimator. The ribbon is the rejection percentage for a jump of respectively 0.5%, 0.75% and 1%. We see, that we do not falsely detect any jumps as bursts before the bandwidth of the volatility estimator is more than 10 times as high as the bandwidth of the drift estimator.

1) We do not detect jumps of any sizes.
2) We do not detect any of the jumps in the period following the jump time.

The first question is not obvious, because there are factors pointing in both directions. As we saw in Section 5.1.1, there exists jump sizes, where we detect more false bursts than 2.5% in one direction (which is quite unfortunate). On the other hand, we have seen both on QQ-plots and in the tables that the tails for the T-estimator is slightly thinner than that of the normal distribution, which might compensate for the higher probability of detection following the jump.

Figure 28 shows $P(T^* > q_{0.95})$ for jump sizes from 0 to 1%. We see the same tendency as in Section 5.1.1, however we never detect above 5%. This shows, that the thinner tails have enough influence to decreases the detection percentage below 5% for all jumps. We are therefore pretty confident that we do not detect jumps more than the usual 5%, no matter the size of the jump.

Figure 29 shows the detection of both a drift burst (with $\alpha = 0.55$), a volatility burst (with $\beta = 0.3$) and a jump (same size as drift burst) on different times of the day spanning from beginning to middle of the day. It supports the theory that we do not detect any jumps no matter what time of the day they occur. The estimator is conservative in the beginning because of the re-scaling, but we still detect some volatility-burst, which verifies that the re-scaling is useful.

Following the above results, we conclude, that jumps will not be detected with a drift bandwidth of 5min and a volatility bandwidth of 60min (ratio of 12). As this choice of bandwidth and ratio yielded the best burst detection, these parameters choices will be used in investigation on real data.

10 Simulation with non-equidistant time steps[9]

Up to now, all theory and simulations have been based on equidistant time steps, i.e. we have chosen the size of the time steps as $\Delta_t = \frac{\mathrm{T}}{n}$, where n denotes the number of observations in our simulation and T denotes

[9]Responsible: Frederik

Table 6: $P(T^* > q_{0.95})$ for the drift bandwidths $h_\mu = (2\text{min}, 5\text{min}, 10\text{min}$ and the bandwidth ratios $C = (15, 12, 10)$. The bandwidth ratios are based on the findings in Section 9.4 and Figure 27. The other parameters are as in Table 5.

			Drift burst		
Bandwidths	β	**Heston**	$\alpha = 0.55$	$\alpha = 0.65$	$\alpha = 0.75$
$\mu = 2\text{min}$, $\sigma = 30\text{min}$	0	1.4	89.1	100	100
	0.1	30.7	51.4	92.7	99.6
	0.2	53.3	59.1	76	91.2
	0.3	62.1	65.5	75	82.4
	0.4	66.6	68.7	76.6	82.7
$\mu = 5\text{min}$, $\sigma = 60\text{min}$	0	2.8	91.1	100	100
	0.1	31.1	59.2	94.3	99.6
	0.2	51.7	61.6	81.3	93
	0.3	61.2	66.5	78.6	85.9
	0.4	64	69	79.8	85.6
$\mu = 10\text{min}$, $\sigma = 100\text{min}$	0	3.8	88.1	100	100
	0.1	26.1	56.2	92.9	99.2
	0.2	45.5	56	78.4	90.9
	0.3	55.1	61.7	74.2	83.9
	0.4	58.4	64	75.4	83.1

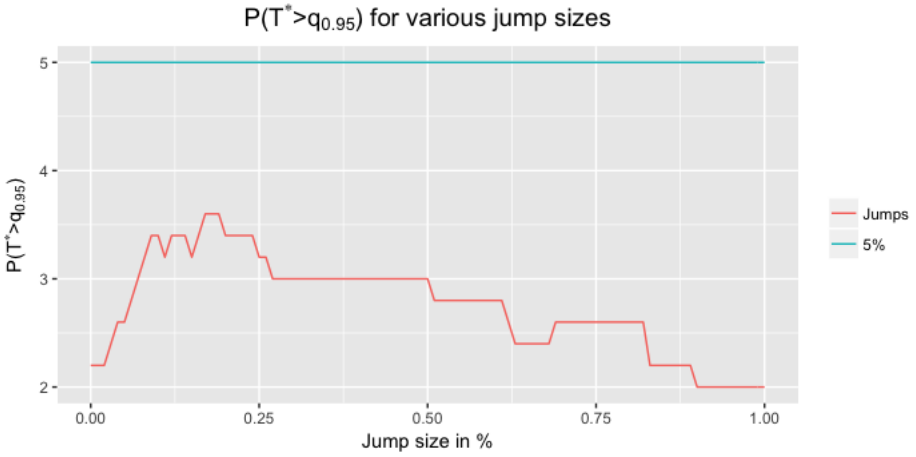


Figure 28: $P(T^* > q_{0.95})$ for various jump sizes.This shows, that though Theorem 1 states that it is theoretically possible to have above 5% detection rate for a jump, this is not the case no matter what the size of the jump is.

the maturity. Although simplistic, it leaves us with a constant time interval between observations. This does not reflect a realistic scenario, as Figure 4 showed a significant intraday seasonality in the frequency of trades on SPY. As the time of the observations is such an integral part of the estimators, we must investigate the performance of our estimators in a setting, where the time of our observations mimics that of the real-data before being able to derive any conclusions on data. This allows us to verify that the optimal estimation values from the simplistic simulations also apply to less simple cases.

In this thesis we will not dive into the theoretical foundation behind non-equidistant time steps as is done e.g. in Mattiussi & Renò (2009). Instead of deriving even more theoretical convergence results, we will here take a practical approach and investigate how the estimator behaves in a finite sample case with a distribution of time steps similar to what we see in data.

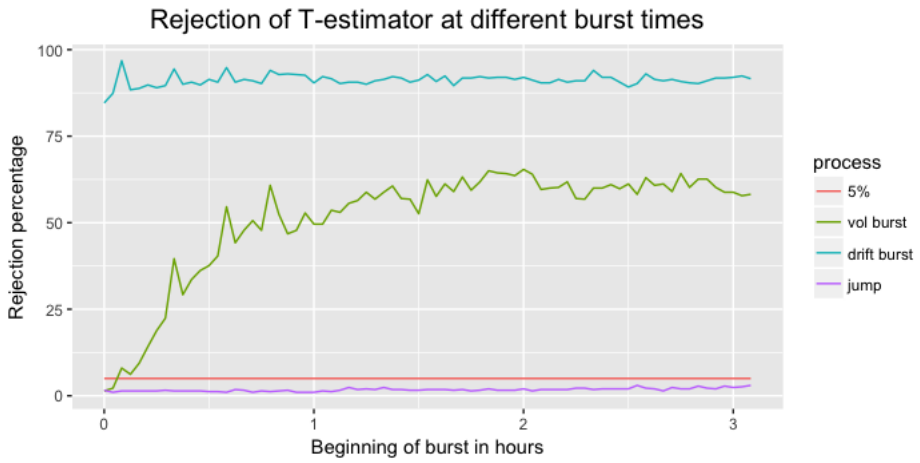


Figure 29: $P(T^* > q_{0.95})$ for bursts and jumps at different times of the day. We see, that we do not falsely detect jumps independent of the time on the day, and also that re-scaling works as we detect bursts at the beginning of the day, though the detection percentage in the beginning is conservative.

Firstly we attempt to mimic the observation frequency of the real data by incorporating the seasonality component that is seen in the SPY data. This is done by partitioning the trade day into smaller intervals, and then matching the share of trades that is seen in real data for each partition. This ensures that our simulated data displays roughly the same intraday seasonality.
Secondly we attempt to avoid equidistant time steps within each of the intraday partitions. This is done very simplistic by distributing the trades within each partition uniformly.

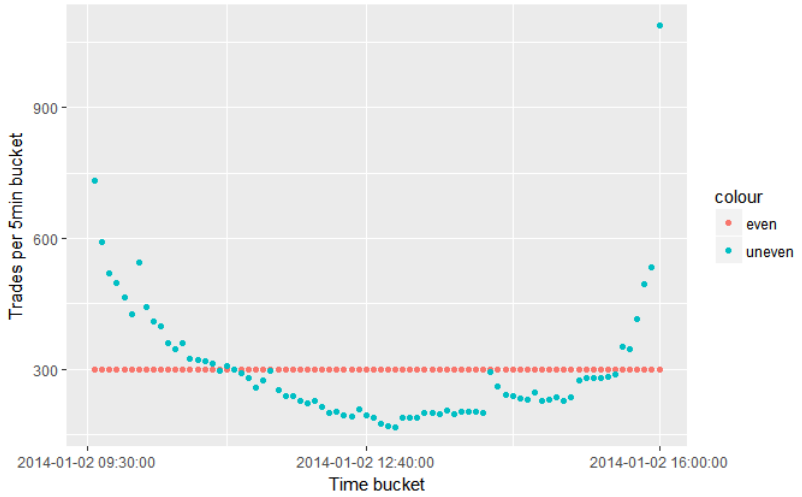


Figure 30: The amount of trades for each 5-minute bucket for the simplistic and the uneven simulation in a setting with a total number of 23,400 trades per day.

With this partitioning of trades, we might see a systematic seasonality behaviour in the T-estimator due to the seasonality of the underlying trades distribution. We investigate this by simulating 250 days of 23,400 trades each. Then T is estimated on data with T-estimations at every 5 second throughout every single day (as in the bandwidth studies).
We estimate the empirical variance of the values of the T-estimations over every day such that the first estimate is the empirical variance of the T-estimations at the first point of estimation on each day, the second

empirical variance of the of T-estimations of the second estimation and so on, that is

$$\text{Var}\left(T_{\text{average}}\right)_i = \frac{1}{N_{\text{days}}} \sum_{j=1}^{N_{\text{days}}} (T_{i,j} - \bar{T}_i)^2, \qquad i \in (5\text{sec}, 10\text{sec}, 15\text{sec}..., 6.5\text{hours})$$

The result of this simulation can be seen in Figure 31. It appears that the non-equidistant simulation scheme is not particularly affected by the intraday seasonality of trades. This shows that the estimator is very robust to the skewed distribution of trades seen in data. We thus deduce that the conclusions from previous simulation studies still hold - despite the simplistic simulation of time.

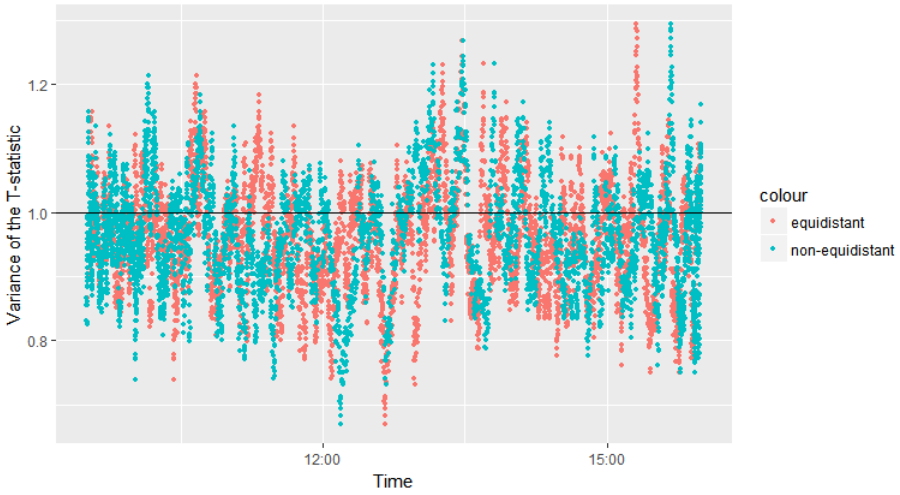


Figure 31: Intraday seasonality of the variance of the T-statistic in simulated data.

As a final note: it may appear as if there is a systematic offset between the two simulation schemes. Both of these have been simulated using the same seed such that the realizations of the brownian motions are equivalent. The offset is simply the result of different values of $\sqrt{dt}$.

11 Need for speed[10]

Due to vastness of our simulation studies, some of the simulation studies in previous chapters would have taken more than a month to calculate with naïve implementations of the drift- and volatility estimators. It has therefore been crucial to the studies that the computational performance of the estimators could be improved.
This section will thus deal with computational optimization of the estimators. The optimization will also ease the data exploration and live computation of the estimator.

The first part of this section will focus the mathematical foundation behind optimization, and the second part will focus on the computational aspect of the optimization.

11.1 Optimization

We will investigate the possibility of enhancing the performance of the computation of the estimator in a setting, where we calculate the estimators in a series of time points that are chronologically ordered by

[10]Responsible: Frederik

exploiting the repetitive form of the estimator and the well-behavior of the left-sided exponential kernel function.

We assume that we have estimated the value of the μ-estimator at time t_n using all of our n observations available. Now imagine we observe an additional observation at a new time point $t_n + \Delta_n$, and wish to calculate the estimator at this new time point using all our now $n+1$ observations:

$$\begin{aligned}\hat{\mu}^{n+1}_{t_n+\Delta_n} &= \frac{1}{h_n}\sum_{i=1}^{n+1} K\left(\frac{t_{i-1}-(t_n+\Delta_n)}{h_n}\right)\Delta Y_i \\ &= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-(t_n+\Delta_n)}{h_n}\right)\Delta Y_i + \frac{1}{h_n}K\left(\frac{t_{n+1-1}-(t_n+\Delta_n)}{h_n}\right)\Delta Y_{n+1}\end{aligned}$$

In the naïve implementation of the estimator, we calculate the expression above. However, it is clear that the first term is very similar to that of the estimator evaluated previously, $\hat{\mu}^n_t$.
We can exploit the well-behaving nature of the exponential function in the our kernel to show that the first term is just a scaling of the previously calculated estimator.

$$\begin{aligned}\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-(t_n+\Delta_n)}{h_n}\right)\Delta Y_i &= \frac{1}{h_n}\sum_{i=1}^{n} \exp\left(\frac{t_{i-1}-(t_n+\Delta_n)}{h_n}\right)\Delta Y_i \\ &= \frac{1}{h_n}\sum_{i=1}^{n} \exp\left(\frac{t_{i-1}-t_n}{h_n} - \frac{\Delta_n}{h_n}\right)\Delta Y_i \\ &= \frac{1}{h_n}\sum_{i=1}^{n} \exp\left(\frac{t_{i-1}-t_n}{h_n}\right)\Delta Y_i \exp\left(\frac{-\Delta_n}{h_n}\right) \\ &= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t_n}{h_n}\right)\Delta Y_i \exp\left(\frac{-\Delta_n}{h_n}\right) \\ &= \hat{\mu}^n_{t_n} \exp\left(\frac{-\Delta_n}{h_n}\right)\end{aligned}$$

The full expression for our estimator is then given as:

$$\hat{\mu}^{n+1}_{t_n+\Delta_n} = \hat{\mu}^n_{t_n} \exp\left(\frac{-\Delta_n}{h_n}\right) + \frac{1}{h_n}K\left(\frac{t_{n+1-1}-(t_n+\Delta_n)}{h_n}\right)\Delta Y_{n+1}$$

It is possible to generalize this such that we assume not to just observe a single new observation but m new observations up to time point $t_{n+m} = t_n + \Delta_{n+m}$. The estimator is then given as:

$$\begin{aligned}\hat{\mu}^{n+m}_{t_n+\Delta_{n+m}} &= \frac{1}{h_n}\sum_{i=1}^{n+m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \\ &= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i + \frac{1}{h_n}\sum_{i=n+1}^{m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \\ &= \hat{\mu}^n_{t_n} \exp\left(\frac{-\Delta_{n+m}}{h_n}\right) + \frac{1}{h_n}\sum_{i=n+1}^{m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i\end{aligned}$$

From this expression it is clear that the benefit of this is higher the more frequently we calculate the estimator.

The variance estimator is slightly more complicated to calculate. It is both the most difficult estimator to update and the most rewarding estimator to optimize. Luckily, the structure of each of the realized l'th order autocovariances, $\hat{\gamma}(l)$, is just as repetitive as that of the drift estimator above.

Following the same scenario as described in the case of the drift estimator, we rewrite the autocovariances:

$$\begin{aligned}\hat{\gamma}^{n+m}_{t_n+\Delta_{n+m}}(l) &= \sum_{i=|l|+1}^{n+m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|} \\ &= \sum_{i=|l|+1}^{n} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|} \\ &\quad + \sum_{i=n+1}^{m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|}\end{aligned}$$

The first sum can be rewritten à la the way we did in the drift estimator:

$$\begin{aligned}&\sum_{i=|l|+1}^{n} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|} \\ &= \exp\left(-\frac{\Delta_{n+m}}{h_n}\right)^2 \hat{\gamma}^n_{t_n}(l)\end{aligned}$$

This yields the following expression for the autocovariance estimator:

$$\begin{aligned}\hat{\gamma}^{n+m}_{t_n+\Delta_{n+m}}(l) &= \exp\left(-\frac{\Delta_{n+m}}{h_n}\right)^2 \hat{\gamma}^n_{t_n}(l) \\ &\quad + \sum_{i=n+1}^{m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|}\end{aligned}$$

As seen above, it is possible to re-use the previously calculated autocovariances very similarly to how we did so with the drift estimator.
The entire variance estimator is given as:

$$\hat{\Sigma}^n_t = \hat{\gamma}^n_t(0) + 2\sum_{l=1}^{L} w_L(l)\hat{\gamma}^n_t(l)$$

We derive the expression of the estimator $\hat{\Sigma}^{n+m}_{t_n+\Delta_{n+m}}$, assuming that the lag-length L, are identical:

$$\begin{aligned}\hat{\Sigma}^{n+m}_{t_n+\Delta_{n+m}} &= \hat{\gamma}^{n+m}_{t_n+\Delta_{n+m}}(0) + 2\sum_{l=1}^{L} w_L(l)\hat{\gamma}^{n+m}_{t_n+\Delta_{n+m}}(l) \\ &= \exp\left(-\frac{\Delta_{n+m}}{h_n}\right)^2 \left(\hat{\gamma}^n_{t_n}(0) + 2\sum_{l=1}^{L} w_L(l)\hat{\gamma}^n_{t_n}(l)\right) + \left(K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i\cdot\right)^2 \\ &\quad + 2\sum_{l=1}^{L} w_L(l) \sum_{i=n+1}^{m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|} \\ &= \exp\left(-\frac{\Delta_{n+m}}{h_n}\right)^2 \hat{\Sigma}^n_{t_n} + \left(K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i\cdot\right)^2 \\ &\quad + 2\sum_{l=1}^{L} w_L(l) \sum_{i=n+1}^{m} K\left(\frac{t_{i-1}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_i \cdot K\left(\frac{t_{i-1-|l|}-(t_n+\Delta_{n+m})}{h_n}\right)\Delta Y_{i-|l|}\end{aligned}$$

The new expressions for the drift and variance makes it clear that we can calculate our estimators in '*chunks*' or '*blocks*', and add those together.

This opens up amazing possibilities from a computational point of view, and means that we can split up our data into a series of *blocks* that honour the chronological ordering of our data, and then spread the *blocks* out between several threads such that each thread calculates the estimators on each *block*, and eventually assemble them back together in the right chronological order.
In most of our simulations, the time points are identical across every path. This means that we in practice only need to evaluate the kernel expression once for every step - irregardless of the number of paths in the simulation. This also improve the computation time

In order to show the power of the recursive form, we apply both of the estimators on a simulated path of 23,400 time steps and compute the estimator for every 1000 time step. We allow the recursive form to use the estimator at the previously calculated point. It is quite obvious from the form that the more frequently we calculate the estimator, the more effective the recursive form is.

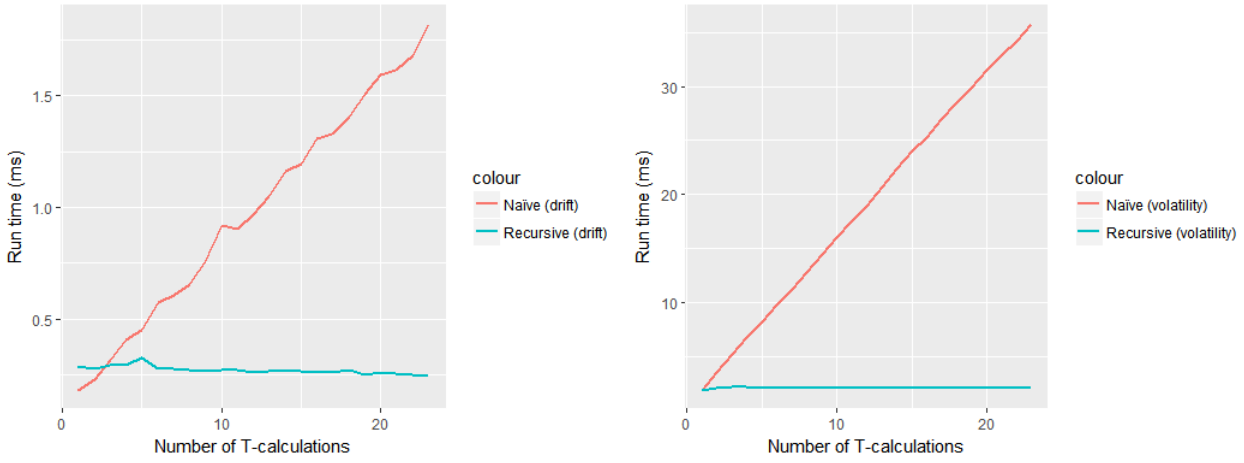


Figure 32: Computation time of the estimators after a certain number of T-calculations that occur every 1000th observation.

In Figure 32, we see how the computation time of the naïve estimator increases with the number of T values that need to be calculated throughout the day, whereas the computation time of the next estimation, given the previous estimation, is linear as long as the number of observations between two points of estimation is constant. Hence by the 20th computation of the T-estimation, the naïve estimator is significantly slower than the recursive. It has to calculate on the basis of 20 000 observations, whereas the recursive only calculates on the basis of 1000 observations and the latest computed T-estimate.

11.2 Implementation

The estimators are implemented in R which is a high-level programming language. While R has a lot of advantages such as fast implementation, a vast array of add-in libraries and line-by-line execution, it also has disadvantages. The main disadvantage of R is that it is an interpreted language and runs slower than compiled languages such as C++. Fortunately, this disadvantage can be minimized by implementing the R functions such that they rely on built-in functions that have been implemented in a compiled language to achieve the best performance.

The recursive form of the estimator in the previous section is tricky to implement without using a 'for-loop' in R. The test in last section was done with a for-loop, and though it is much faster than the non-recursive estimator, it is more than 100 times slower than if the for-loop was implemented in C++.

A way around using the for-loop in R is to first calculate the cumulative sum of ΔY_i and the rescale it with the correct kernel $K(\frac{t_i - t}{h_n})$ afterwards (and exploiting the properties of the exponential kernel). The code in R looks as below:

```
mu_estimator <- function(dy_vector, K_function, h_mu, time_points){

  t_end <- time_points[length(time_points)]
  kernels <- K_function(
             (time_points[1:(length(time_points)-1)]-t_end)/h_mu)
  sum_terms <- kernels*dy_vector
  mu_non_scaled <- 1/h_mu * cumsum(sum_terms)
  mu_estimates <- mu_non_scaled/kernels

  return(mu_estimates)
}
```

With this method, the calculations are several times faster than with a simple R for-loop. The reason for this is the built-in function *cumsum* function in R, as the built-in function is vectorized and is implemented as a single for-loop in a pre-compiled low language code. Thus it can calculate all the estimations without looping in R.

It can be quite limiting to restrict oneself to only using built-in functions for fast computation in R. One alternative is to implement the function in C++ and source it into R. The R package "Rcpp" allows intuitive and effortless interoperation between R and C++ and handles all of the complexity *under the hood*. Unfortunately, this also means that we are more restricted when it comes to including libraries, which makes it much more difficult to implement more complex features - such as multi-threading. All-in-all the package enables us to draw from the high flexibility of R while enjoying the lightning-speed of C++.

We benchmark the execution time of these various implementations with one path with 23,400 steps, estimating at every 5th time step, using a lag length of 10. This gives the result seen in Table 7.
We see that despite being fully implemented in R, the *cumsum*-version is almost as fast as the fully C++ implemented, which is implemented in "Rcpp". It also shows how huge an impact replacing a for-loop in R with the built-in function *cumsum* has on the performance. After all of the optimization, the estimator is now computed roughly 10,000 times faster. As a result the estimation of T is (although still quite time consuming in some simulation studies) no longer unmanageable in neither our the simulations nor in real-data exploration.

Table 7: Comparison of computation time across implementations

	Computation time (ms)
Naïve	91,000
Recursive	897
Recursive (*cumsum*)	21
Recursive (fully built in C++)	8

11.3 Extra optimization

If the lag-kernel used in the volatility-estimator is sufficiently smooth, the calculation of $\hat{\Sigma}_t^n$ can also be carried out almost independently of the number of lags (i.e. we can avoid looping over lags).

In this thesis the parzen-kernel is used. This is made up from two third degree polynomials, and by using the fact that

$$f(x) = 4f(x-k) - 6f(x-2k) + 4f(x-3k) - f(x-4k)$$

for a third degree polynomial, we can avoid looping over lags. However because it is split into two third degree polynomials, it is important to keep track of which lags to evaluate in which polynomial. Because of this, using a recursive formula for the parzen-kernel is only really a significant advantage if we have $L \geq 20$. An implementation in R of the volatility estimator the recursive calculations of the parzen weights can be seen in Appendix 2.

12 The hunt for bursts in financial data[11]

We are now ready to test the estimator in practice. The burst estimators will be used on the tick-by-tick SPY and Bitcoin data, which was described in Section 2.
The procedure and set-up is kept similar to what was done in Table 6. That means we compute the biased T-estimator from (40)every 5^{th} second throughout every trading day. The bandwidth and bandwidth-ratio is respectively 5 minutes and 12, as this was found to be optimal in the bandwidth investigation.
For volatility estimation, we choose a constant lag size of 10 on the basis of the results of our simulations and will make our results comparable with Christensen et al. (2016).
The estimators are re-scaled such that the initial burn-in is kept to a minimum despite the wide bandwidth used in the Σ-estimator. Due to the large amount of observations in the beginning of the day, we choose a burn in time of merely 1 minute. This allows us to examine and detect bursts within the busy early hours of the day.
Since Bitcoin trades occur around-the-clock, we will not burn every single day in this market, but solely on the very first day. Table 8 summarizes the parameter choices.

Table 8: Estimation set-up

T estimation frequency	Every 5 seconds
Bandwidth (μ)	5 minutes
Bandwidth (Σ)	60 minutes
Lags	10
Initial burn	1 minute

Table 9: Summary of T-estimations

	Mean	Standard deviation	Kurtosis
SPY	0.01	0.92	4.36
Bitfinex	-0.02	1.58	9.08
BitMEX	-0.01	1.28	9.04
Kraken	-0.01	1.06	6.68

The summary of the T-estimations shows that the kurtosis is much higher than that of a standard normal distribution. This suggests that the tail are heavier and supports the existence of bursts in financial data. The kurtosis is extremely high in the Bitcoin market and from Figure 33, we can see how the heavier tails of the Bitcoin exchanges produce more extreme events. We should therefore expect to see a substantial amount of bursts on these exchanges compared to the SPY.

12.1 Detecting bursts

We calculate the maximum T-statistic every day and evaluate it against an AR-process with an autocorrelation matching the empirical autocorrelation of the T-estimator in accordance with the algorithm that is found in section 8. We use this to construct the 95% quantile for the estimator. The number of days where

[11] Responsible: Frederik

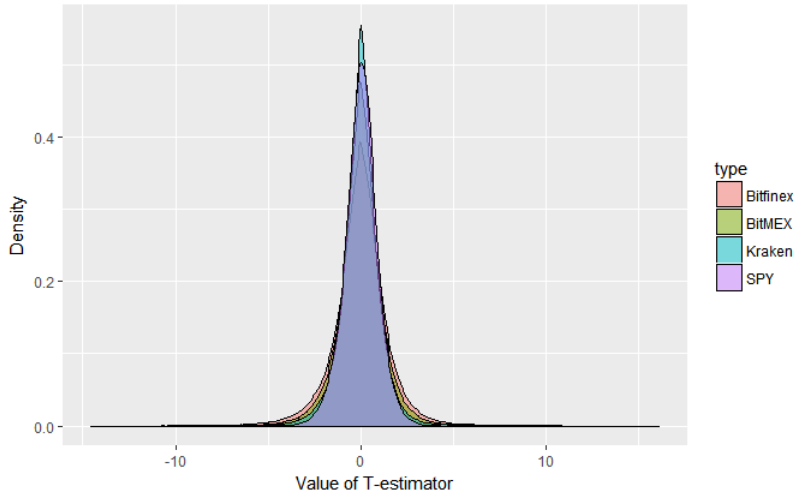


Figure 33: Density plot demonstrating the distribution of T-estimations across the four assets.

the T-estimator is above this interval along with other thresholds can be seen Table 10. These results will commented in the following subsections.

Table 10: Summary of daily T^*_{day}

	Days	$T^*_{\text{day}} \geq q_{95}$	$T^*_{\text{day}} > 5$	> 6	> 7	> 10
SPY	249	84	23	14	7	0
Bitfinex	80	80	80	80	80	56
BitMEX	80	80	80	80	78	43
Kraken	78	78	74	42	13	0

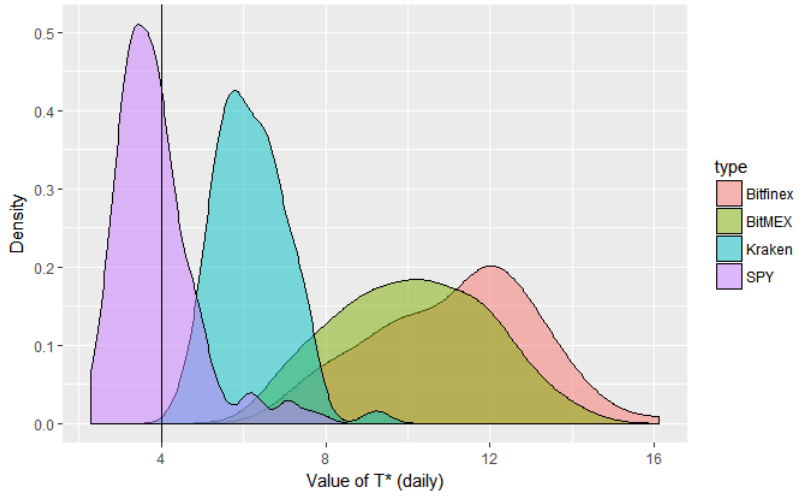


Figure 34: The distribution of daily measured T^*_{day} values. The vertical line at 4 corresponds roughly to the 95% quantile used for T^*. This generally varies from 4 to 4.5 depending on the number of T-estimations and auto-correlation.

12.1.1 Expected false detections

With the 95% quantile, we should expect a false-positive for every 20^{th} evaluation of T^*, if the T-estimator is truly normally distributed.

From the simulations in Table 6 we saw that with the Heston setup, we had $P(T^* > q_{95}) = 2.4\%$ which corresponds to a false detection even more infrequent than every 20^{th} day. This implies that we, depending on the quality of the convergence, should expect the estimator to be above q_{95} between 2.4% to 5% of the days - or equivalently every 20^{th} to 40^{th} day in the absence of bursts.

We detect bursts much more often than every 20^{th}, which bodes well for the hypothesis of the existence of bursts. In the next two sections we will examine the results in further detail and *zoom in* on specific bursts that have been found in the data.

12.2 SPY

In the SPY data, we detect burst(s) in one third of the days. This is more often than Christensen et al. (2016), which is due to the increased bandwidth ratio as this allows for increased detection of volatility bursts. The amount of bursts found in the SPY is quite astounding, when considering that the SPY is an ETF that tracks the entire S&P 500 - hence the price process can therefore be quite robust. Although the stocks within the S&P 500 index may contain bursts (e.g. at a publication of an annual report of a firm), they may not be sufficiently large to translate into bursts on the entire index.

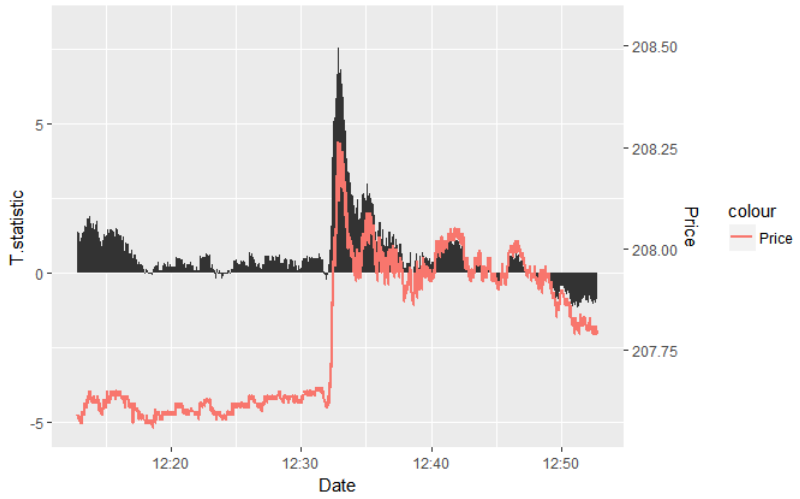


Figure 35: A sharp upwards burst in SPY on Thursday the 4th of December 2014

This means that for an event to spark an index-wide burst, it would have to either be of a very large caliber on a single stock, or affect several stocks within the index. While the first scenario is very much possible, one could expect that market agents that sell a stock, causing a down-wards burst, would very likely place the money in another stock to maintain the fraction of stocks, bonds and capital in their portfolio. If the market agents should choose to purchase stocks on the same index as the previous sold, the overall price level of the index, as a whole, would remain close to unchanged.
In the latter scenario, bursts that affect several stocks within the index are more likely to cause market agents to relocate their capital away from the index entirely. This could be caused by events that alter the balance between the S&P and other markets. Such events could be an interest rate hike, causing market agents to

allocate a larger fraction of their portfolio into bonds, or events on foreign exchange market, causing agents to reduce holdings of US dollar based holdings.

12.2.1 Event-driven bursts in SPY

We will attempt to connect a handful of bursts found in the SPY with events throughout 2014. It is important to remark that we do not intend to claim that there is a causality. Instead we merely consider the events a plausible explanation for the bursts.

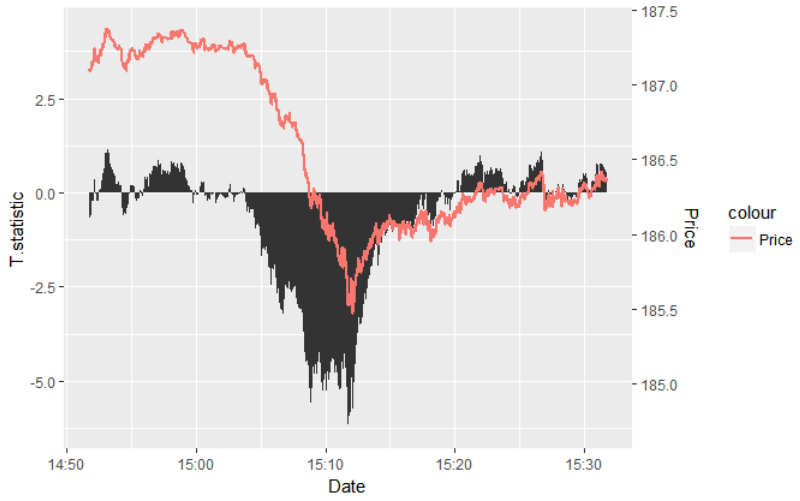


Figure 36: Downwards burst on the 19th of March 2014 following the Federal Reserve's announcement of increased interest rates

The SPY experiences a downwards burst as seen in Figure 36 following chair of the Federal Reserve Janet Yellen's announcement[12] that the Federal Reserve will *probably* end its bond buying program in the fall of 2014, and *could* start raising the interest rates roughly six months later. As explained before, this increases incitement to relocate capital from the stock market to the bond market, causing an overall decline in demand for stock holdings.

As the FED has announced the *possibility* of an increase in interest rate in March 2014, one can imagine that market agents withdraw market orders in the hours up to the FED meeting, reducing liquidity and adding volatility due to the uncertainty of the announcement.

Another burst is found on the 30th of July and portrayed in Figure 38. Similar to the burst on the 30th of April, it appears that the price process moves in less of an unanimous direction, hinting that the two bursts might be more strongly driven by volatility than drift.

The last example of a burst related to the FED decision is the burst that is found on the 28th of October 2014 that follows the FED announcement of no interest rate hike[13]. As seen in Figure 39, the market responded positively with an upwards burst. This announcement was roughly six months after 19th of March 2014, where the FED announced the *possibility* of an interest rate increase in six months. As an increase in interest rate will lower the demand for stocks, the market reacted in a positive manner, because the expected increase did not happen. The price process moves in a more unanimous direction than the two previous bursts. This hints that this burst might contain a drift component that is stronger than the two previous bursts - yet whether the drift term dominates the volatility term is unknown.

All in all, it seems plausible that nation-wide money policy could have the ability to spark bursts in the entire index.

[12] https://www.reuters.com/article/us-usa-fed/fed-may-raise-rates-as-soon-as-next-spring-yellen-suggests-idUSBREA2I07520140320

[13] http://money.cnn.com/2015/10/28/news/economy/federal-reserve-october-meeting-no-rate-hike/index.html

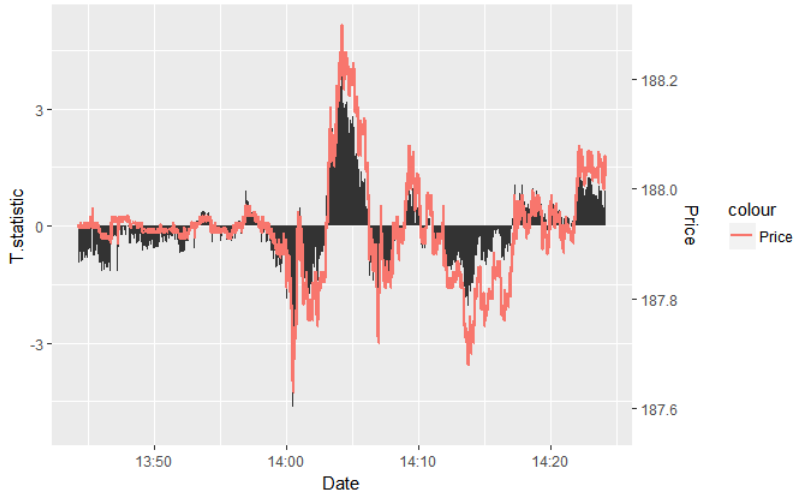


Figure 37: A burst on the day of the FED meeting on the 30th of April 2014.

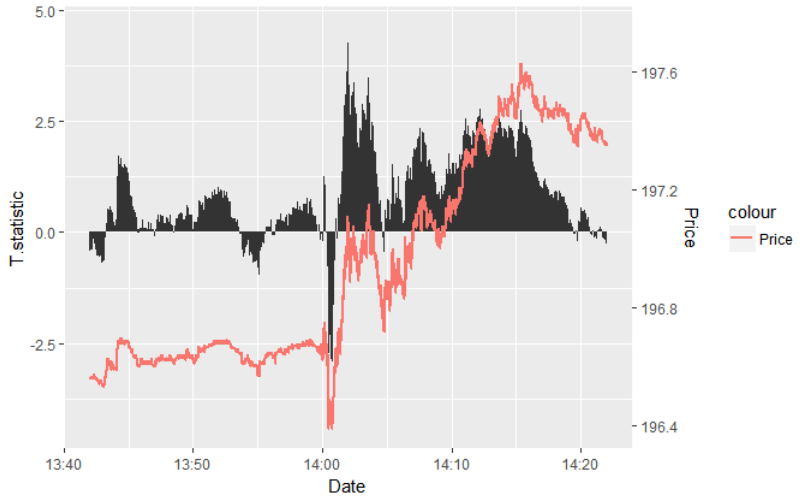


Figure 38: An upwards, yet volatile, burst on the day of the FED meeting of 30th of July 2014.

12.2.2 Re-scaling in action

In one of the previous sections, we introduced a re-scaling of our estimators that would allow for conservative T-estimation in the beginning of the day.
Figure 40 shows that the re-scaling of our estimators also works in practice, and not just within the controlled environment of simulation studies. Here the estimator detects a burst 20 minutes after the market opens.

12.3 The Bitcoin market

The Bitcoin is different from the S&P. It is not robust like an index and has a much higher volatility. Besides a high volatility in general, bursts could also be much more likely in this market. One explanation for bursts is that the market is less regulated which means that, although the exchanges might have rules for trading and price manipulation, "pump-and-dump" schemes and other market manipulation are much more present here.

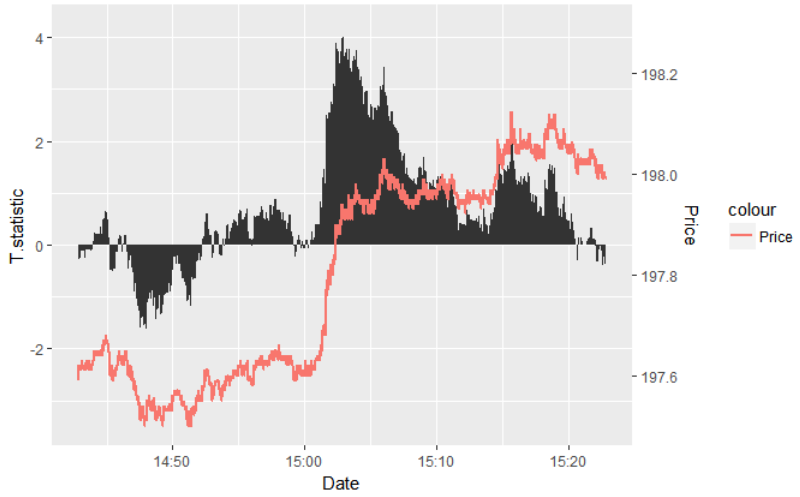


Figure 39: An upwards burst on 28th of October 2014 following FED announcement of no interest rate hike.

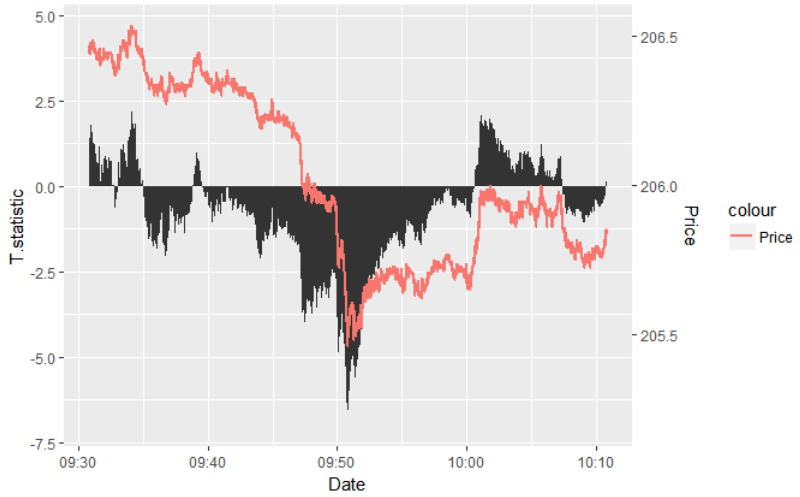


Figure 40: Burst in the SPY process in the morning of Monday the 1st of December 2014, reassuring that rescaling works.

Besides this, the general behaviour is more aggressive, and many of the possible explanations for bursts mentioned in Christensen et al. (2016) such as predatory trading, forced liquidation because of a high leverage, or agents with tournament type preferences and an aversion to missing out on trends are most likely much more applicable in this market. This is also why we choose to include a burst analysis on this market as the existence of burst seem much more probable in this setting.

The following two figures allows for a glimpse of how volatile the Bitcoin market is. In Figure 41 we see Bitcoin price on BitMEX along with the T-statistics during the Bitcoin Consensus 2018. A three day long event from May 14th to May 16th. The price is so volatile that when viewed on a full day the process seems to contain several bursts or jumps of more than 1%! In Figure 42 we see the burst in Bitfinex that produced the largest T-statistic. The price process moves almost unanimously upwards throughout the ten minute period from 7:20 to 7:30. The size of the burst is much larger than any burst seen on the SPY data, and it is even bigger than the largest drift bursts in the simulation study of 1.5% - this indicates that bursts indeed are present.

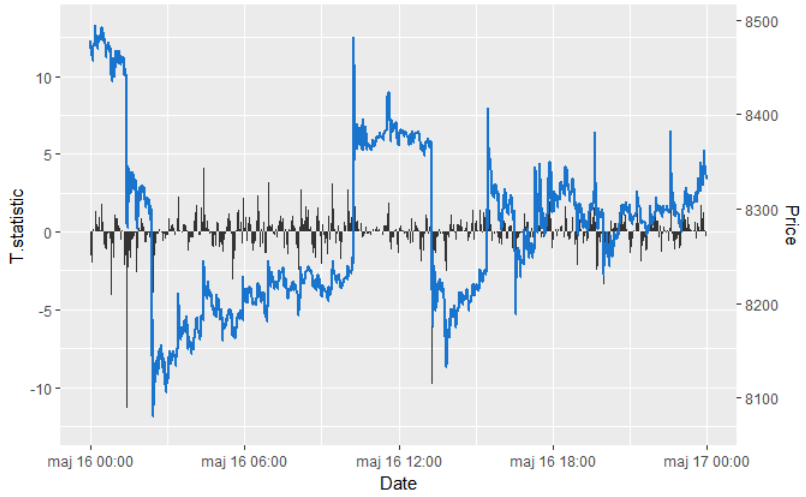


Figure 41: Bitcoin price on BitMEX on the last day of the Bitcoin Consensus 2018.

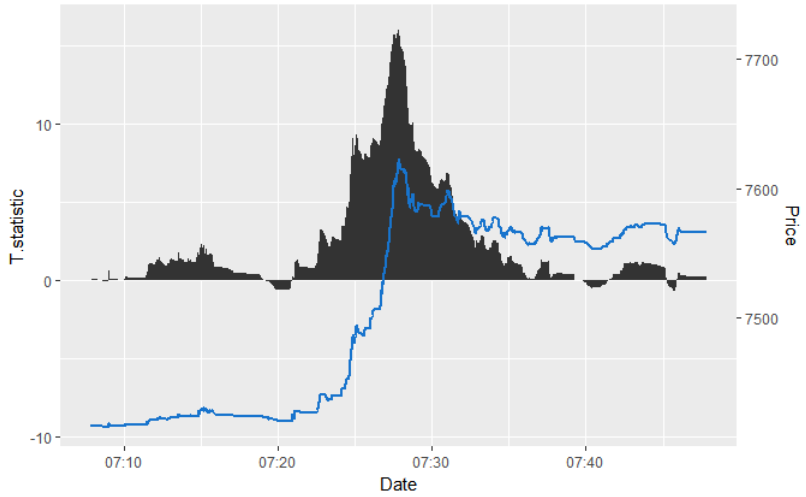


Figure 42: A massive upwards burst on Bitfinex the 26th of May 2018

On Bitcoin we find a bursts every day on Bitfinex and BitMEX, with values of the T-estimator far above the threshold, and on Kraken almost every day. Although the reasoning above also argue for a higher frequency of bursts in this market, there is also an issue, which is very likely to bias the findings towards too many detections in this market.

While the price process on SPY is based on trades, the mid-quotes are sampled on the Bitcoin exchanges. This has an influence on the noise process, and Hansen & Lunde (2006) show (in *Fact I* in the article) that the noise process is negative correlated with the efficient returns. They find that this is more evident for mid-quotes than for transaction prices.

This is very likely affecting the number of bursts detected in the Bitcoin market. In Section 19.2 it is explained, why high-frequency frictions on BitMEX and Bitfinex may inflate the detection rate with a microstructure noise that induce *sticky prices*, which is in line with a negative autocorrelation between the price process and the noise as found in Hansen & Lunde (2006). It will not be elaborated here, however it is kept in mind that this might falsely inflate the detection rate on these two platforms.

12.3.1 Event-driven bursts in Bitcoin

An example of an event-driven burst is seen in Figure 43. It portrays the Bitcoin price traded on the Bitfinex exchange on Sunday the 10th of June 2018. On that day, one of the major South Korean cryptocurrency exchanges was hacked[14], causing a large drop in Bitcoin price on all three exchanges. The large drop may reflect that the market agents have underestimated the security risk associated with Bitcoin exchanges, and wish to reduce the holdings of Bitcoin in order to reduce the exposure to security breaches. Another explanation could be that the market agents anticipate that the stolen Bitcoin will be sold within the near future, increasing the supply and decreasing the price.

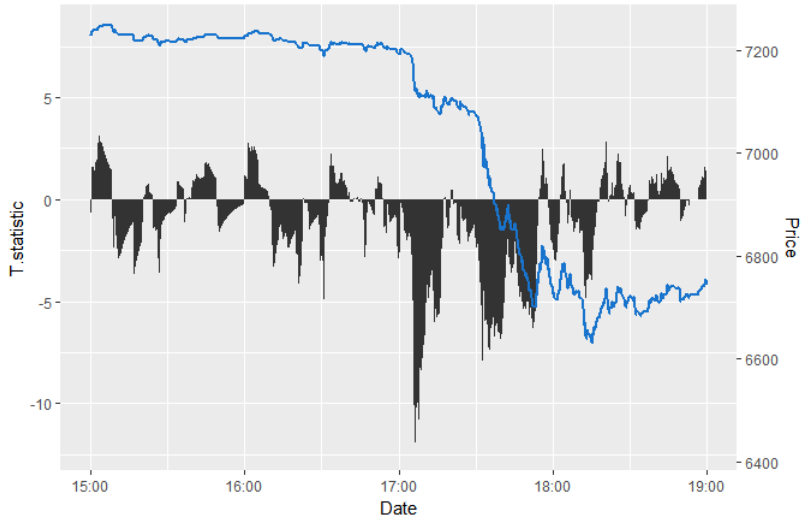


Figure 43: Bitcoin (Bitfinex) after a South Korean cryptocurrency exchange Coinrail is hacked on 10th of June 2018

12.4 Potentially grave issue with bandwidth ratio

In previous sections, we argued for an optimal bandwidth ratio of 12. Based on the findings in our simulation study, we deemed it worth sacrificing the distinction between drift and volatility bursts for better detection. However the well-behavedness of the simulation study, both with and without non-equidistant observations, blinded us to an other potential consequence that arises when using a high bandwidth ratio: the diminishing robustness to changes in volatility.
When estimating with a high bandwidth ratio, the estimated volatility is *smoothed out* more than the estimated drift. This means that the drift estimator adapts quicker to a change in volatility than the volatility estimator. Therefore the distribution of the T-estimator might be skewed as a result of the choice of bandwidth ratio if the volatility changes sufficiently fast - such that the volatility estimator is a poor approximation of the instantaneous volatility.
This was not an issue in the simulation study as the used Heston dynamics are based on annualized volatility parameters of the S&P market. This results in a very slow change of volatility that is not sufficiently large to skew the distribution of the T-estimator. Hence the conclusions that were drawn from the simulation study are only sure to be valid if the intraday changes of volatility in data is not too violent compared to the changes based on the volatility of volatility from the simulation study.

The Bitcoin market is, as mentioned, much more volatile than SPY, but we do not know how the volatility of volatility behaves. If this is high as well, the volatility estimator is to a much higher degree a poor

[14] https://www.cnbc.com/2018/06/10/bitcoin-tumbles-10-percent-after-news-of-south-korea-crypto-exchange-hack.html

approximation of the instantaneous volatility. This would lead to the variance of the T-estimator being more than 1 throughout periods of increasing volatility and vice versa. With a variance slightly above 1, the probability of an outcome above a certain threshold increases significantly as a consequence of the heavier tails than the standard normal distribution. The largest bursts detected in the Bitcoin data are generally detected following a period of very low volatility. This is portrayed in both Figure 42 and 43.

Although the SPY data does not appear to exhibit the same sudden change in volatility, it is a well-known fact in literature that the S&P market exhibits a certain degree of seasonality in intraday volatility. Because of this the volatility changes more during the day than our simulation suggests, as it is a component that is not included in our simulation study so far. This more rapid change in intraday volatility might have a significant effect on our T-estimator.

Luckily, since we already expect some degrees of intraday seasonality in volatility in SPY, we can do a preliminary investigation into whether this smoothing effect is present, or if it is only of theoretical concern.

12.4.1 Preliminary investigation of smoothing and intraday seasonality

Similar to the approach in the non-equidistant simulation section, we will calculate the variance of the T-statistic every 5^{th} second throughout the day:

$$\text{Var}\,(T_i) = \frac{1}{N_{\text{days}} - 1} \sum_{j=1}^{N_{\text{days}}} (T_{i,j} - \bar{T}_i)^2, \qquad i \in (\text{09:31:05, 09:31:10,..., 16:00:00})$$

To investigate, we plot $\text{Var}(\sqrt{h_n}\mu_t^n)$ and $E[\sqrt{\Sigma_t^n}]$ for every T-statistic, similar to the procedure in study of bias investigation, which is seen in in Figure 44. By the intuition above, it is argued that the drift estimator adapts more quickly than the volatility estimator. Hence when the true volatility decreases, the volatility estimate is above the variance of the drift estimator - and vice versa. Assume then that the volatility decreases from the start of the day towards the middle of the day, and then increases again towards the end of the day. This intraday volatility pattern is commonly found in literature. We should then see that the variance of the drift estimate should decrease faster than the mean of the variance estimate that is used to normalize the distribution of T. This intuition matches the pattern found in Figure 44. This can be formulated in terms of the T estimates, as is seen in Figure 45.
While the intraday seasonality of the volatility could be a plausible explanation of the pattern seen in the intraday T-statistic, it could also be explained by other factors - such as more bursts in the afternoon than in the morning. The only thing that we can conclude at this moment, is that it is not caused by the changing number of trades throughout the day. This effect was tested in our simulation with non-equidistant observations and proved to have no effect. To further validate the suspicion that intraday volatility is the cause, the same empirical estimations can be computed for a bandwidth ratio of 1. In this case, the effect should not be present as the degree of smoothing should be the same for both estimators. This is seen in Figures 46 and 47. This time the values in Figure 46 are of similar levels as before, i.e. likely to still follow the underlying intraday volatility, but unlike previously, the two estimates are much more in sync with each other. This is naturally also reflected in Figure 47, where there is no longer found the same change of level throughout the day.

Since changing to bandwidth ratio 1 potentially resolved the issues, the earlier analysis is repeated with the new bandwidth ratio.

12.4.2 An attempt with bandwidth ratio of 1

From our simulation study with same bandwidths, we found in Table 3 that a bandwidth of 10 minutes had the best detection rate, and this parameter choice will therefore be used.

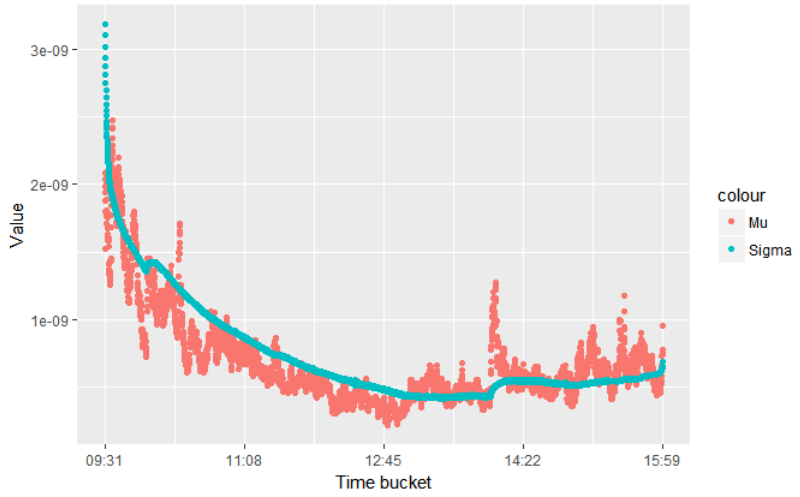


Figure 44: Emperical estimation of $\mathrm{Var}(\sqrt{h_n}\mu_t^n)$ and $E[\Sigma_t^n]$ across all days on SPY.

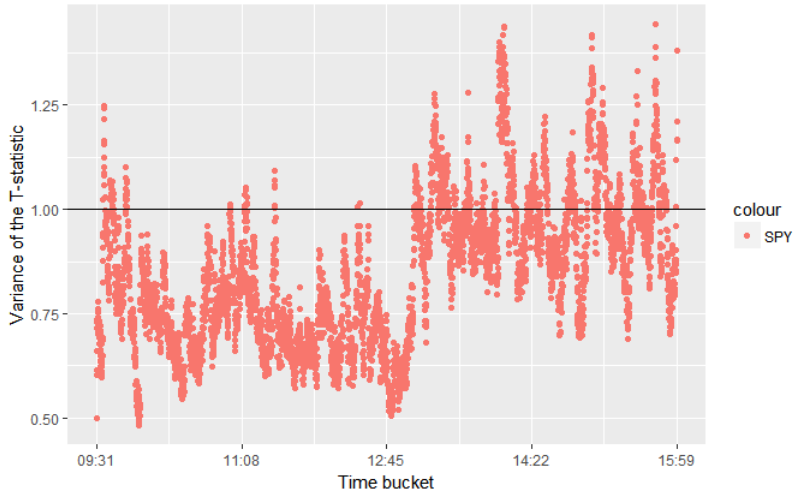


Figure 45: The intraday average of the magnitude of the T-statistic on SPY.

In Figure 48 we plot the distributions of T^*_{day} in the alternative setting with a bandwidth of 10 minutes and ratio of 1. In contrast to our conclusions in the bandwidth ratio study, we are still able to detect many bursts in the Bitcoin data when using a bandwidth ratio of 1. Since pure volatility bursts are not detected with a bandwidth ratio of 1 which was both shown theoretically and in Table 3, this suggests that the drift bursts are indeed present in the Bitcoin market - though the conclusions are blurred due to the high-frequency feature on BitMEX and Bitfinex that will be explained in Section 19.2. This also makes the existence of bursts, where the drift term is dominant in SPY doubtful, though unclear. We see that we only detect bursts 2% of the days in this market, which matches well with the expected amount of false positives in the simulation study.

While this study avoids the potential issues related to using a bandwidth ratio different from 1, it does not detect any bursts on SPY. This does not necessarily lead to the conclusion that there are no drift bursts in SPY, as we in Section 9.2.1 found that the T-estimator performed poorly compared to the case with larger bandwidth ratio. This could mean that the absence of detected bursts with dominant drift term in SPY is in fact due to a sub-optimal detection capabilities, and there might occur bursts in SPY that are simply not detected.

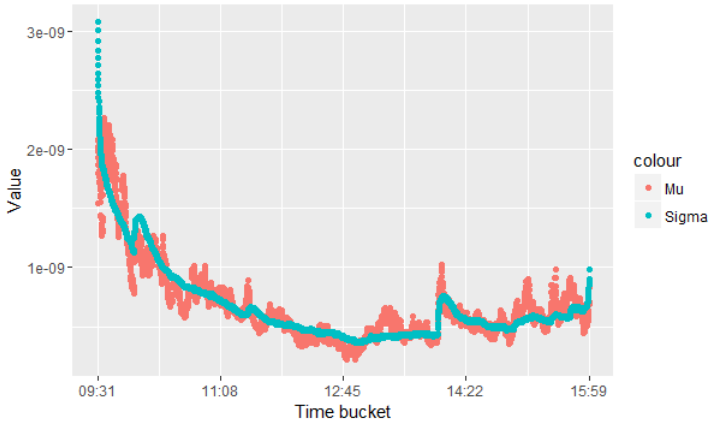


Figure 46: The intraday average of $\mathrm{Var}(\sqrt{h_n}\mu)$ and $\sqrt{\Sigma}$ with bandwidth ratio 1.

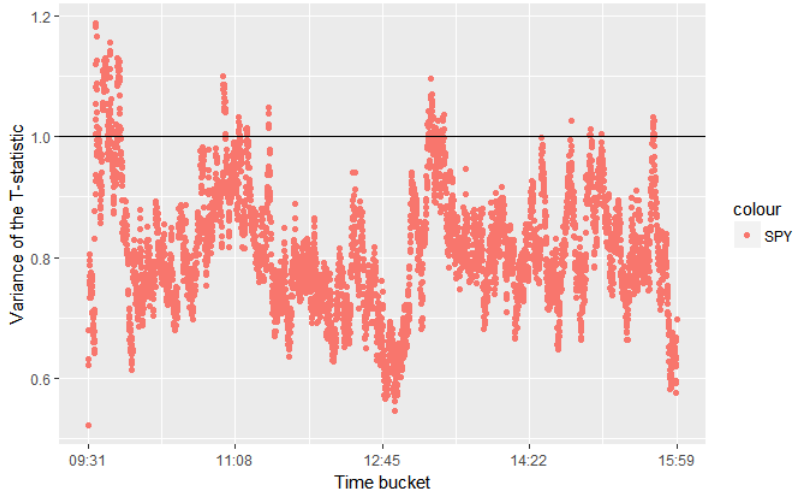


Figure 47: The intraday average of the magnitude of the T-statistic on SPY.

Whether bursts are present in SPY is therefore still uncertain, as the estimator with ratio 1 might be too conservative, and the estimator with different bandwidths might be invalid due to a poor modelling of the intraday volatility. In order obtain more conclusive results, a better understanding of the interplay between the volatility process and the T-estimator is needed.
Hence we will analyze the volatility much more in-depth than previously. Eventually we will introduce a volatility model that is much more suited for intraday volatility trends. This model will serve as an alternative to the Heston dynamics that our simulations have been based on.

Table 11: Estimation set-up with ratio 1

T estimation frequency	Every 5 seconds
Bandwidth (μ)	10 minutes
Bandwidth (Σ)	10 minutes
Lags	10
Initial burn	1 minute

Table 12: Summary of daily T^*_{day}

	Days	$T^*_{\text{day}} \geq q_{95}$	$T^*_{\text{day}} > 5$	> 6	> 7	> 10
SPY	249	5	0	0	0	0
Bitfinex	80	79	64	17	0	0
BitMEX	80	43	3	0	0	0
Kraken	78	1	0	0	0	0

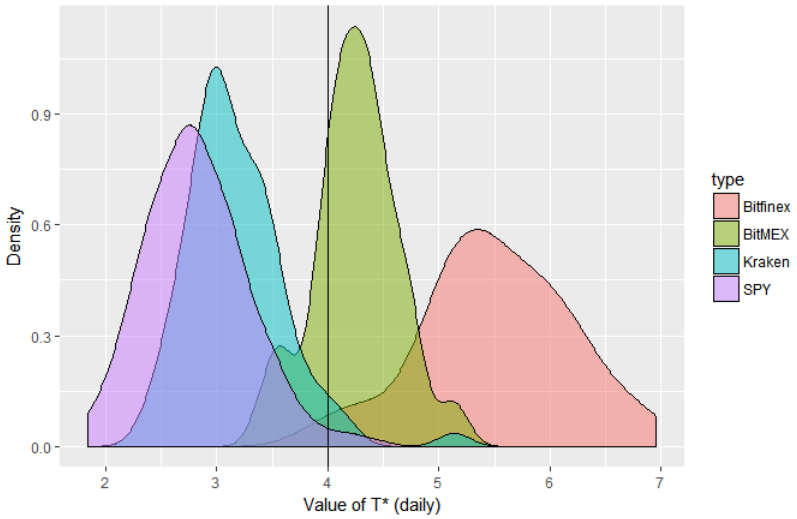


Figure 48: Distribution of T^*_{day} with $h_n = 10$ minutes and bandwidth ratio of 1.

13 Framework for volatility estimation[15]

In the following work the bursts are completely disregarded and volatility is attempted estimated in the usual setting used in literature. The framework for the estimation is already described, as we consider a Itô process in a setting with jumps and microstructure noise as in equations (1) and (16), with a diffusion term generically given as $\sigma_t dW_t$, and a drift term being disregarded. In order to do volatility estimation in this setting, we utilize the pre-averging as found in Christensen et al. (2012) defined as in equation (23). This allows for the use of the BV^* estimator as defined by equation (24). This yields an estimator of integrated variance. This leads to the following approximation for some time $0 < t < T$:

$$\sigma_t^2 \approx \bar{\Delta}^{-1} \int_{t-\bar{\Delta}}^{t} \sigma_s^2 ds$$

for $0 < \bar{\Delta} < t$. This approximation associates the interval $[t - \bar{\Delta}, t[$ with one estimate of variance or, by taking square root, volatility. For small values of $\bar{\Delta}$, this is considered to be somewhat accurate. Now, since we can approximate the integral with the BV^*, and denote the estimator of the appropriate interval by

[15]Responsible: Mathias

$BV_t^{\bar{\Delta}*}$, we get the following estimator:

$$\hat{\sigma}_t = \sqrt{\bar{\Delta}^{-1} BV_t^{\bar{\Delta}*}}$$

With this estimator in hand, we briefly reiterate from above sections that the volatility estimator used above would not be satisfactory here as it is not a consistent estimator in the presence of jumps.

14 Analysis of empirical volatility[16]

The analysis above has been kept notationally simple as it all relates to one continuous period of trading. However, additional notation will be need once data consists of observations spread across several days - each separated by a night with no observations.
With this in mind, let $d \in \{1, ..., D\}$ denote a continuity period, e.g. a day, with D denoting the number of continuity periods. Let $s \in \{1, ..., S\}$ divide each period d in equally-size sub-periods, e.g. intraday periods, with s denoting a specific sub-period. Often, these sub-periods will be referred to as buckets, and their length, corresponding to the $\bar{\Delta}$ above will be denoted the bucket width.
For example, if we want to split up a trading day into 5 minute buckets, we would specify the number of sub-periods in the following manner

$$S = \frac{6.5\text{hours} \cdot 60 \frac{\text{minutes}}{\text{hours}}}{5\text{minutes}} = 78$$

Further the general notion of a timepoint t is now understood as having $0 \leq t \leq T$ such that t specifies a point within an interval given by some pair (d_t, s_t), simply saying that t will always denote a point in time where trading can occur. Hence (d_t, s_t) will be used to denote the period corresponding to a specific value of t.

Finally exact understanding of T will be made clear in the context for which it is needed.
When investigating data, many results will depend greatly on the choice of the number of intraday periods S. In order to both follow Bennedsen et al. (2016) and keep the analysis enlightening, we impose some limitations. Most importantly, the attempt is to analyze the intraday behaviour of volatility, meaning that $S > 1$ and preferably S as large as possible to increase granularity. However, investigating how large S can be while still yielding reasonable results is a task that will be conducted later on. This means that, initially, a large portion of the analysis will be based on a heuristically chosen values of S.

14.1 Seasonality

Following the ideas presented in Bennedsen et al. (2016), it is assumed that the volatility can be *decomposed* in the following manner:

$$\sigma_t = \tilde{\sigma}_t \bar{\sigma}_{(d_t, s_t)}$$

where $\bar{\sigma}_{(d,s)}$ denotes a deterministic function describing the intraday seasonality of σ_t and $\tilde{\sigma}_t$ is an adapted stochastic process, which will be referred to as the deseasonalized volatility process.
With this specification we wish to estimate the deterministic function under the assumption that it does not depend on d in any way. In other words, seasonality does not change from one continuity period to another - day to day. Therefore we write $\bar{\sigma}_{(s)} := \bar{\sigma}_{(d,s)}$.
Hence we can write the decomposed log-volatility as

$$\log \sigma_t = \log \bar{\sigma}_{(s_t)} + \log \tilde{\sigma}_t$$

We use two different approaches to fit the deterministic intraday seasonality. The first approach, which we will refer to as simple the simple approach, is the following regression:

$$\log \hat{\sigma}_{(d,s)} = \alpha_s + \epsilon_{(d,s)}$$

[16] Responsible: Mathias

Where $\hat{\sigma}_{(d,s)}$ denotes estimates resulting from the BV^* estimator corresponding to the choice of S, and $\epsilon_{(d,s)}$ denotes a sequence of iid normally distributed random variables with mean 0 and variance σ_ϵ^2. This yields the estimator

$$\widehat{\log \bar{\sigma}}_{(s)} = \frac{1}{D}\sum_{d=1}^{D} \log \hat{\sigma}_{(d,s)} \tag{44}$$

i.e. simply for each sub-period, taking the average across all continuity period.
The other approach is the flexible Fourier form which is inspired by Andersen & Bollerslev (1998) with some slight modifications. Again σ_t is replaced with $\hat{\sigma}_{(d_t,s_t)}$, but this time it is assumed that $\log \bar{\sigma}_{(s)}$ can estimated by a flexible Fourier form and $\log \tilde{\sigma}_t$ is replaced by a single parameter - specifically

$$\log \hat{\sigma}_{(d,s)} = \underbrace{\alpha + \delta_1 s + \delta_2 s^2 + \sum_{p=1}^{P}\left(\beta_p \sin\left(p\frac{2\pi s}{S}\right) + \gamma_p \sin\left(p\frac{2\pi s}{S}\right)\right)}_{\widehat{\log \bar{\sigma}}_{(s)}} + \epsilon_{(d,s)} \tag{45}$$

for some $P \in \mathbb{N}$. In practice, choosing $P < S/2$ for an even value of S is useful as the terms of $p = S/2$ corresponds to specifying linearly dependent column in the design matrix.
Naturally this term could be excluded, as well as any other of the form $p = \frac{m}{2}S$ for $m \in \mathbb{N}$.
The regression specified by (45) requires a well-founded choice for P - the number of sine and cosine terms. We will base this choice on the P that minimizes the AICc:

$$AICc = 2k - 2\log \hat{L} + \frac{2k^2 + k}{N - k - 1}$$

where k denotes the number of parameters estimated, $\hat{L}$ the value of the minimized log-likelihood for the regression, and finally N denotes the total number of observations - in this case N is at most $D \cdot S$.

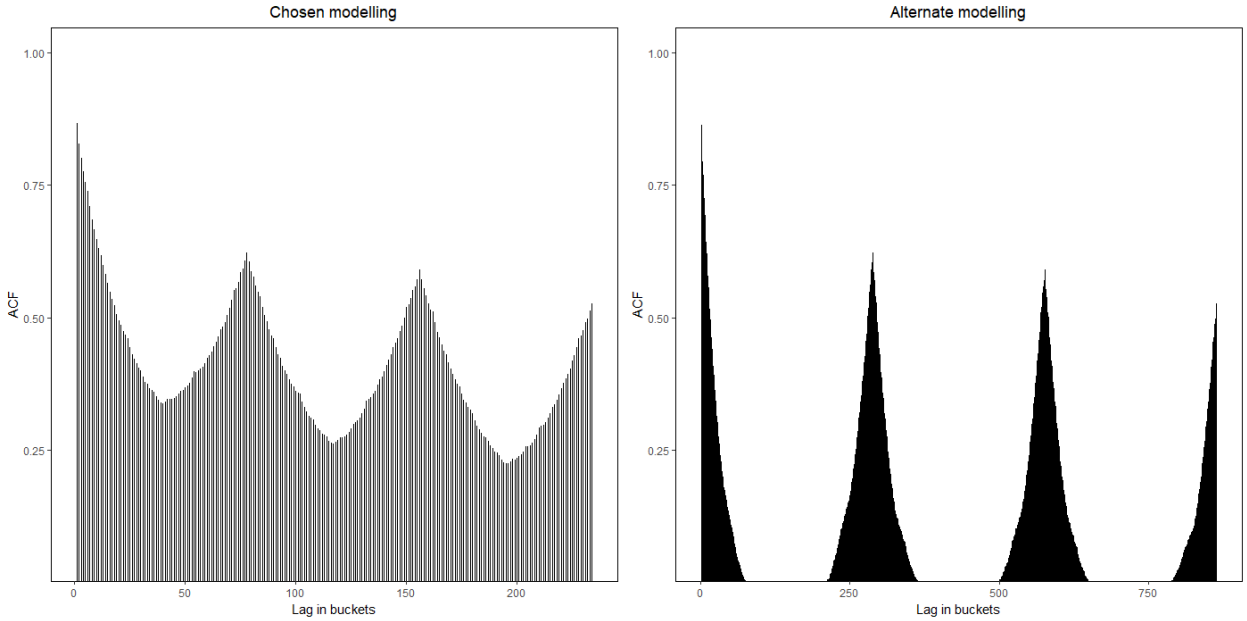
Both of the described methods suffers from some of the same drawbacks. They both replace $\log \tilde{\sigma}_t$ by the sequence $\epsilon_{(d,s)}$. Effectively implying that the deseasonalized log volatility can be described by iid normally distributed random variables with mean zero. However, later on the distribution of $\log \tilde{\sigma}_t$ will be proposed to be very different that.

This means the estimates themselves are interesting, they cannot be expected to maximize the joint likelihood presented in the later chapters. The development of more thorough estimation procedures would be interesting. The lack of interaction between the continuity period and the sub-period could be a major simplification. In Andersen & Bollerslev (1998) a significant level of interaction is found in an equity market. However allowing for such interaction will severely impact the method of prediction presented in Bennedsen et al. (2016).
Furthermore, this regression is not identical to the one in found in Andersen & Bollerslev (1998). A reason for this is the required estimation of $E(X_t)$ for t in some interval of interest. Computing any such estimate is simple, however as seen in earlier sections, most simple calculations will not result in both microstructure noise and jump robust estimators, and while it is likely possible to derive such an estimator, its practical value in this context is uncertain. Ultimately the flexible Fourier form from Andersen & Bollerslev (1998) is used as a sufficient estimate of the intraday volatility.
Finally, one could consider a regression containing both models. This is easily achieved. However, this would highly defeat the purpose of the flexible Fourier approach, as it is exactly a way of smoothing a periodic fit from data - whereas the first easily can produce quite rigid results.

Once one of the above estimation methods have been applied, the deseasonalized log volatility can be approximated as

$$\log \tilde{\sigma}_t \approx \widehat{\log \bar{\sigma}}_{(d_t,s_t)} := \log \hat{\sigma}_{(d_t,s_t)} - \widehat{\log \bar{\sigma}}_{(s_t)}$$

This deseasonalized log volatility process will be of major interest in regards to modelling volatility.



ACF of the estimated deseasonalized log volatility. Estimates based on 5-minute buckets. Left panel showing the effect of 0 hour night modelling, and right panel showing the effect of 17.5 hour night modelling. Large gaps are seen in the ACF in the right panel, due to the lack of observations outside trading hours.

Figure 49: ACF of $\widehat{\log \tilde{\sigma}_t}$ for different models of time

14.1.1 Modelling time and volatility

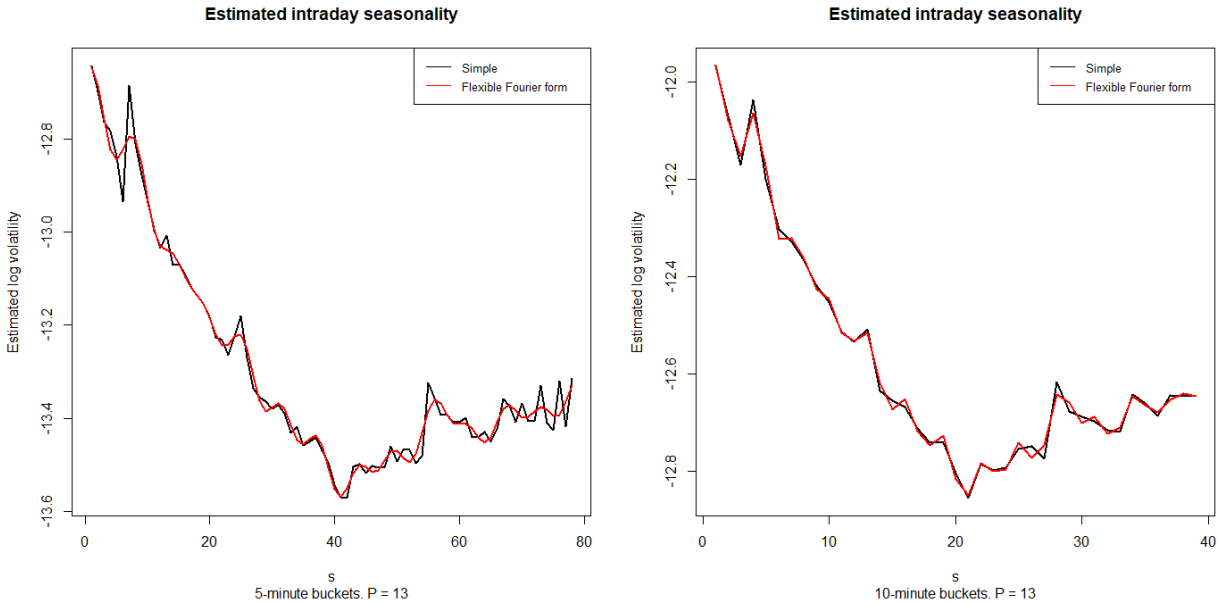
So far the analysis of intraday volatility has been carried out with any references to the value of t - only values of (d_t, s_t) were needed. Going forward however, this will not be sufficient, and in particular a understanding of modelling the change of continuity period is important. In order to match the method of Bennedsen et al. (2016) the following model is used: Data is cleaned such that after S is chosen, intraday-equidistant observations are available for all continuity periods in the data set. This means that intraday time can be measured by $\Delta := \frac{T}{D \cdot S}$. Further, the change of continuity period time is chosen to be modelled also as 0. This assumption yields a time series with $D \cdot S$ equidistant observations, as the first observation of every continuity is not at the start, but rather after Δ time has passed, and is a crucial assumption when computing e.g. autocorrelation.

As an example, consider figure 49. Here the effect of the above assumption is shown on the SPY data. Given the 6.5 hour trading day, acting as a continuity period, the change of continuity period could also be modelled as a change of time of size $24 - 6.5 = 17.5$ hours. Clearly this is also a simplification, but somewhat closer to reality. The point, however, is that the time series with the assumption above allows for a much nicer analysis, as illustrated in the figure.

Ultimately, while this assumptions yields a time series potentially quite different from other naïve modelling choices, the results it produces are both nice and in any case applicable. One could argue that the continuous model set up for X_t means that any model for σ_t should be extended, i.e. interpolated, between continuity period, however one could also conjecture that the true process X_t behaves differently in-between continuity period, and thereby that any extension would be a very inaccurate extrapolation.

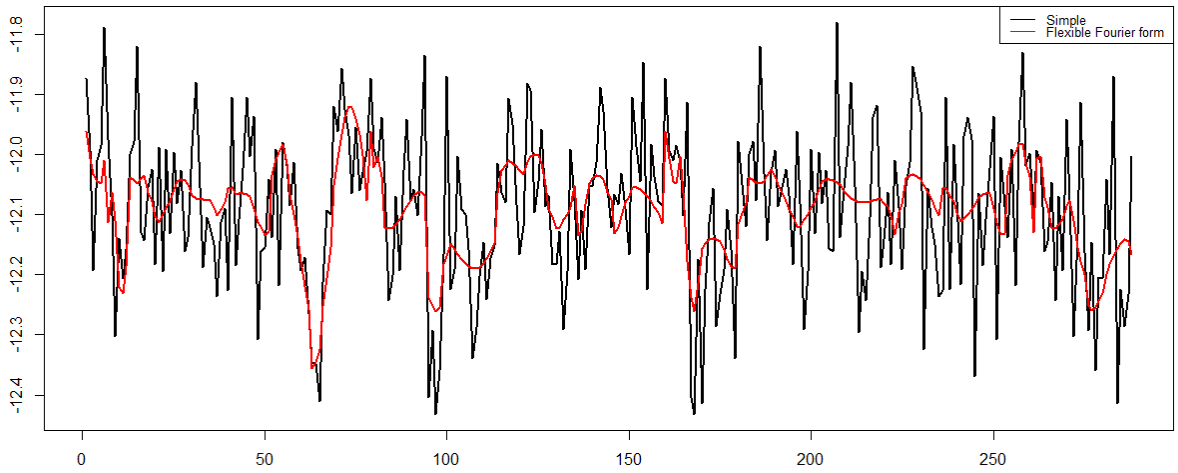
SPY

The results of regressions (44) and (45) can be seen in figure 50. When considering 10-minute buckets, there seem to not be too big of a difference between the two methods, however, in the 5-minute case the flexible



Estimates of the deterministic intraday seasonality component of log volatility on SPY data. Left panel showing estimates using 5-minute bucket, right panel showing 10-minute buckets. In both cases the AICc criteria was used to pick the number of sinusoidal terms in the regression for the flexible Fourier form, incidentally resulting in $P = 13$ terms.

Figure 50: Estimated intraday seasonality, i.e $\widehat{\log \bar{\sigma}}_{(s)}$



Estimates of the deterministic intraday seasonality component of log volatility on Bitcoin data modelled with 5-minute buckets.

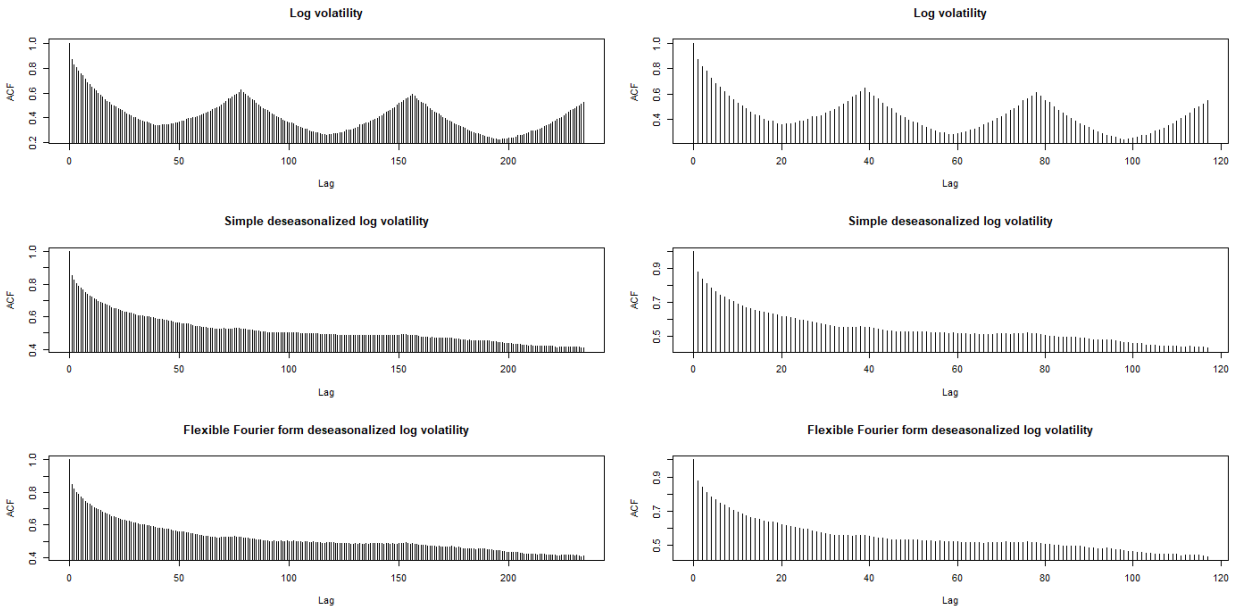
Figure 51: Estimated intraday seasonality, i.e $\widehat{\log \bar{\sigma}}_{(s)}$ on Bitcoin data

Fourier form results in a more smooth fit, although the difference is still not too large due to the quite high number of sinusoidal terms included. The overall resemblance is quite promising as it shows that there seem to be a quite distinct intraday seasonal pattern.
Figure 52 shows the autocorrelation before and after the two methods of deseasonalization. Clearly, both methods accomplish the goal of removing seasonality from data. One reason for preferring the flexible Fourier form is the concern of overfitting. In particular, one might be concerned about the rigid estimation for large values of s in the case of 5-minute buckets in figure 50 for the simple estimation method, as it could be a potential sign of overfitting. However, results have proven to be extremely robust to the choice of method, hence only results corresponding to deseasonalization using the flexible Fourier form will be stated. As a side note, the found pattern does coincide with the intuition described in section 12.4, and thereby could serve as an explanation of the intraday behaviour found in the T estimator as theorized. However rather than attempting to find validation for our theory, we continue expanding on the empirical behaviour below, in order to determine the behavior of the T estimator under the most realistic assumptions possible.

Bitcoin
When modelling volatility on Bitcoin, only data from Bitfinex will be used, as there are by far the most observations. However, this data contrasts the SPY data heavily in several ways. Firstly, since the exchange never closes, there is no immediate question of how to model time when the price is not observed. Secondly, the available data is missing quite a few observations. This means that data will still be structured in the same manner of having D and S, with different values when compared to SPY, but the process will not be defined for every combination (d, s). This is a point that will be returned to on many occasions.
Even with the missing observations, the autocorrelation can be computed. It is done simply by computing the variance using the empirical variance of all observations and the autocovariance based on observations where the lag of interest exists. Further, it should be noted that the bucket size mean something quite different in the case of missing observations. Here, a larger bucket-size does not necessarily lead to more observations, therefore another parameter is introduced that represents a minimum number of observations in a bucket for is to be included. If the requirement is not met, the bucket is treated as if there were no observations at all. A short discussion of this parameter can be found in the section regarding small bucket widths.
Comparable to figures 50 and 52, we have figure 51 and 54 for Bitcoin. Here we see no immediate seasonality, which is also reflected in the fact that the autocorrelation function seem close to invariant to the deseasonalization. It is not obvious that this feature would be present, but ultimately not surprising, as the seasonality in the SPY data seems to be closely related to the closing hours on that market. Like for the SPY data, we choose to use the more smooth modelling of intraday seasonality, especially so since the lack of pattern could potentially indicate that with enough observations, the difference throughout the day would even out complete.

14.2 Stationarity

SPY
To discuss stationarity consider figure 53. This figure displays ACF for 5-minute buckets, meaning a total of 19344 observations. In the left panel the ACF corresponds quite well to a stationary process with a high degree of persistence as discussed below, where the autocorrelation eventually becomes insignificant. On the right panel, this feature becomes much more doubtable, since the autocorrelation starts to fluctuate - never quite seems to stay insignificant which required by a stationary process. Note that while the left panel uses a high number of lags, around 1500, it is still a small portion of the total number of observations.
A discussion of this phenomenon can be found in Mikosch (n.d.), where, among other things, it is suggested to split data into smaller samples to attempt to detect stationarity within some structural periods. However, this is not the approach taken in Bennedsen et al. (2016), and while undoubtedly more thorough it quickly diverges into a different analysis than the one attempted in this thesis. Rather, much less ambitiously we first conduct the Pearson-Phillips test. Since the underlying estimator comes from Newey & West (1987), the lag selection for this method is $c_{pp}N^{1/4}$ for some $c_{pp} > 0$. Secondly, the augmented Dickey-Fuller test,



ACF of the estimated deseasonalized log volatility. Left panel showing results for 5-minute bucket, and right panel showing results for 10-minute buckets, without any notable difference.

Figure 52: ACF of $\widehat{\log \tilde{\sigma}}_{(d,s)}$

Bucket Width	Test type	Lag	P-value
5-minute	PP	44	< 0.01
	ADF	167	< 0.01
10-minute	PP	37	< 0.01
	ADF	37	< 0.01
15-minute	PP	34	< 0.01
	ADF	52	< 0.01
30-minute	PP	28	< 0.01
	ADF	16	< 0.01

Table 13: Tests for stationarity

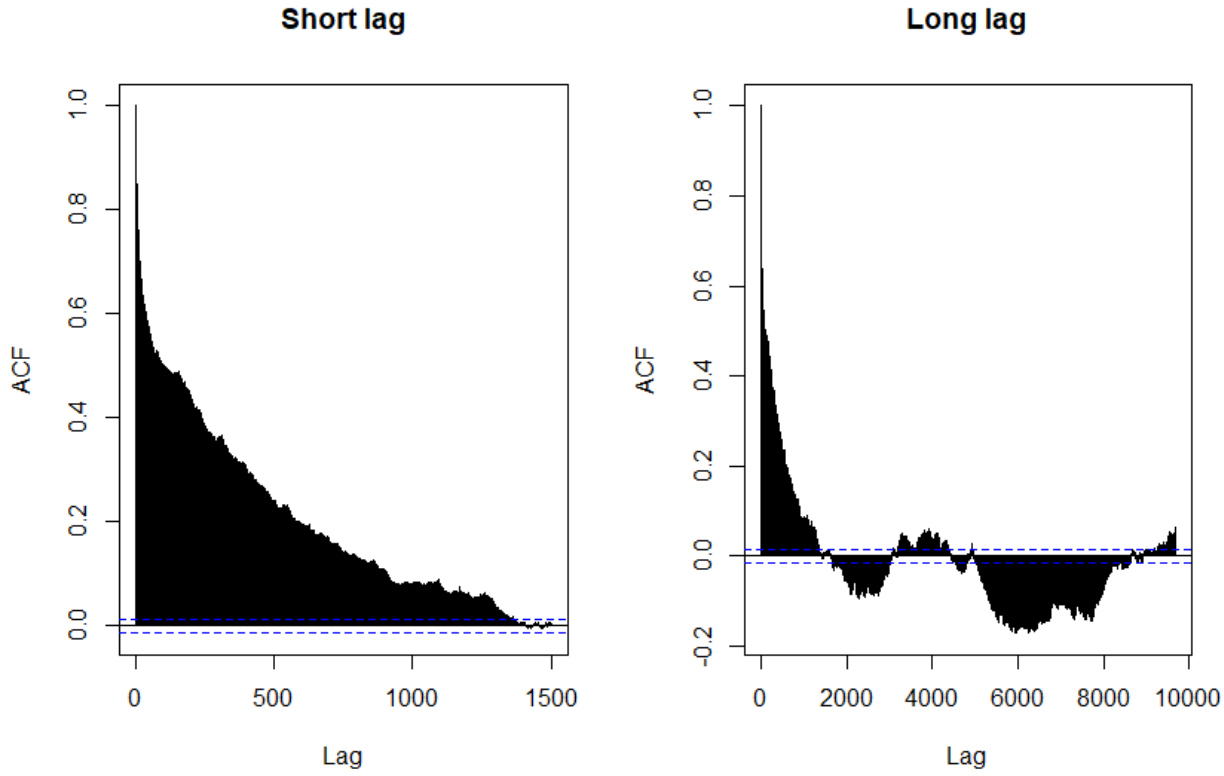
without trend or constant, is computed. Lag selection here is done by finding the optimal fit of a form

$$\Delta\widehat{\log \tilde{\sigma}}_n = \delta\widehat{\log \tilde{\sigma}}_n + \delta_1\Delta\widehat{\log \tilde{\sigma}}_{n-1} + ... + \delta_{P_{DF}}\Delta\widehat{\log \tilde{\sigma}}_{n-P_{DF}+1} \tag{46}$$

for some coefficients $\delta, \delta_1, ..., \delta_{P_{DF}}$, where $(\widehat{\log \tilde{\sigma}}_n)_{n=1,...,N}$ is used to denote the sequence $\widehat{\log \tilde{\sigma}}_{(d,s)}$ for index-wise convenience, and Δ denotes usual time series differencing. Optimality is again determined by the AICc. Table 13 shows results of the tests which are in clear favour of stationarity, which will therefore be assumed from here on out.

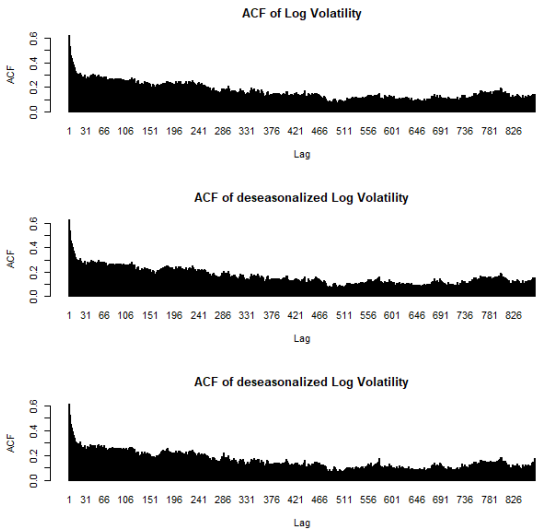
Bitcoin

Once again, we attempt the same analysis for the Bitfinex data and run into quite some difficulties. In figure 55, we see the same features of autocorrelation going negative and making it unclear if stationarity is present not. While there are much fewer observations available, the general shape does seem similar to the one in the SPY case. Now consider equation (46) with missing observations. The way the fit is determined is to only consider the complete cases, i.e. the values of n where there exists observations such that the left hand side



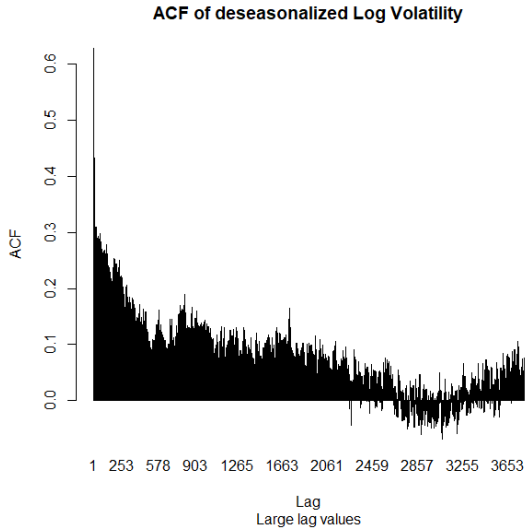
ACF of the estimated deseasonalized log volatility. Left panel showing results for 5-minute bucket and shorter maximal lag, and right panel showing results for 5-minute buckets but very large maximal lag.

Figure 53: ACF of $\widehat{\log \tilde{\sigma}}_{(d,s)}$ large lag values on SPY data



ACF of the estimated deseasonalized log volatility on Bitfinex data, with 5-minute buckets.

Figure 54: ACF of $\widehat{\log \tilde{\sigma}}_{(d,s)}$ for Bitcoin



ACF of the estimated deseasonalized log volatility on Bitfinex data, with 5-minute buckets for large lag values.

Figure 55: ACF of $\widehat{\log \tilde{\sigma}}_{(d,s)}$ for large lag for Bitcoin

can be calculated. Let a complete case denote such a n. An interesting feature occurs, since increasing the bucket width reduces the potential number of complete cases, as there simply are fewer possible observations, it is entirely possible to achieve more complete cases than with more and slim buckets. This is closely related to the fact that buckets with sufficiently few observations are discarded. This feature explains how wider buckets can result in necessity of higher number of lags, as what we are really fitting to is the number of complete cases. Due to these complications, we only attempt the ADF test and not the PP test.
Since the fit is now based solely on complete cases, i.e. different data, it is not possible to use the AICc, nor any other criteria involving the likelihood. Rather, we analyze the p-value across many different possible lag values. For 15-minute bucket, there are no issues, as all lag values results in more complete cases than parameters. We find p-values < 0.01. However, the 5-, 10- and 30-minute bucket yields unclear conclusions. For 30-minute bucket, we seem to lack a criteria against overfitting. Here we can continue to add more and more terms, and where for the SPY the AICc, equivalently the AIC or BIC, would stop us in doing so, we are now left without much option. Since the 30-minute bucket will not play a large role in the following modelling we choose to exclude it here. In the case of the 5- and 10-minute buckets, the number of complete cases can become very small, for the 5-minute bucket all the way down to around 50 at lag length around 25, which means that almost all of the data at this point has been discarded as there are 7606 observations in total. Here we find a general tendency of when increasing lags, thereby decreasing the number of complete cases, the p-value increases - even up to quite large values. If we only include lags where there are at least 300 complete cases, we find that for the 10-minute bucket the p-value < 0.01, but for the 5-minute bucket, the p-value is barely smaller than 0.10. This makes it extremely difficult to determine whether the process is stationary or not, but seeing as the small p-values were only found when there are only very few available complete cases, especially combined with the similarities found in the autocorrelation function when compared to the SPY data, the process is deemed stationary.

14.3 Persistence and roughness

In this section we will discuss the closely related concepts of persistence and roughness of the deseasonalized log volatility. However, these concepts can be defined for general continuous time series as follows.
Let $(V_t)_{t\geq 0}$ denote an adapted stochastic process, and let $\rho_V(t,h)$ denote the autocorrelation function of V, i.e.

$$\rho_V(t,h) := \text{cor}(V_t, V_{t+h})$$

with $h > -t$.
If V_t is stationary then

$$\rho_V(t,h) = \rho_V(t+c_0,h)$$

for any c_0, h such that both expressions are well-defined. This means that ρ_V only depends on the value of h, and we simply write $\rho_V(h) := \rho_V(t,h)$.

Definition 1. ***Roughness***
A process stationary V_t said to have short term memory, if

$$1 - \rho_V(h) \sim c|h|^a, \quad |h| \to 0$$

for some constant $c > 0$, and $a \in (0,\infty)$. Further, $\alpha := \frac{a-1}{2}$ will be referred to as the roughness index.

where the notation " $\sim$ " means that the ratio of the left-hand side and right-hand side tends to 1.

Definition 2. ***Persistence***
A process stationary V_t said to have long term memory, if

$$\rho_V(h) \sim c|h|^{-\beta}, \quad |h| \to \infty$$

for some constant $c > 0$, and $\beta \in (0,1)$.

Notably, any process with long term memory will have non-integrable autocorrelation, i.e.

$$\int_0^\infty |\rho_V(h)|dh = \infty$$

These definitions have clear and natural equivalents for discrete time series, in particular the integral corresponds to an appropriate sum, and we will not distinguish between the two. It is however worth noting that these definitions allow for more heuristic discussions of the persistence and roughness to be formulated in terms of parameters a, or α, and β, or even more generally rates of decay of the autocorrelation function.

One final tool is needed - the second order variogram for a process, V_t:

$$\gamma_2(h) := E(V_{t+h} - V_t)^2, \quad t, h > 0$$

for a stationary process, one can utilize $\text{Var}V_{t+h} = \text{Var}V_t = EV_t^2 - (EV_t)^2$ to rewrite the above

$$\begin{aligned}\gamma_2(h) = E(V_{t+h} - V_t)^2 &= E\left(V_{t+h}^2 + V_t^2 - 2V_{t+h}V_t\right) \\ &= E\left(V_{t+h}^2 + V_t^2 - 2V_{t+h}V_t + 2(EV_t)^2 - 2(EV_t)^2\right) \\ &= 2\text{Var}V_t(1 - \rho_V(h))\end{aligned} \tag{47}$$

utilizing the definition of correlation in the final equality. The variogram will not in itself be of interest, but instead be useful for estimation. The empirical estimate of the variogram is calculated as:

$$\hat{\gamma_2}(h) = \frac{1}{N}\sum_{i=1}^{N-h}(v_{i+h} - v_i)^2 \tag{48}$$

where $\{v_1, ..., v_N\}$ denotes a set of observations of the process V_t.
A key observation has to be made in regards to the variogram. Assume observations of V_t are not available, and only noise versions of the following form are:

$$\hat{V}_t = V_t + \nu_t$$

for some process ν_t independent of V_t. This yields the following variogram calculations:

$$\begin{aligned}E\left(\hat{V}_{t+h} - \hat{V}_t\right)^2 &= E\left(V_{t+h} - V_t + \nu_{t+h} - \nu_t\right)^2 \\ &\overset{\text{independence}}{=} E\left(V_{t+h} - V_t\right)^2 + E\left(\nu_{t+h} - \nu_t\right)^2 \\ &= \gamma_2(h) + E\left(\nu_{t+h} - \nu_t\right)^2\end{aligned} \tag{49}$$

Note that if the last term is non-negative, means that the variogram of the noise observations $\hat{V}_t$ is a biased version of the variogram of V_t. Similarly, a calculation of the correlation from noise observations can be made:

$$\rho_{\hat{V}}(\hat{V}_{t+h}, \hat{V}_t) = \frac{\text{cov}\left(V_{t+h} + \nu_{t+h}, V_t + \nu_t\right)}{\text{Var}V_t + \text{Var}(\nu_t)} = \frac{\rho_V(h) + \frac{\text{cov}(\nu_{t+h}, \nu_t)}{\text{Var}V_t}}{1 + \frac{\text{Var}(\nu_t)}{\text{Var}V_t}}$$

If one further assumes that $\rho_\nu(t, h) = 0$ for $h > 0$, this reduces to

$$\rho_{\hat{V}}(\hat{V}_{t+h}, \hat{V}_t) = \frac{\rho_V(h)}{1 + \frac{\text{Var}(\nu_t)}{\text{Var}V_t}} \tag{50}$$

These calculation on noisy observations are made to explain the effects of not directly observing the volatility, nor even the integrated variance. More strict assumptions on ν_t, such as iid when V_t describes a discrete process, is often done with noise processes, however no investigation into the actual behaviour of ν_t will be made.

14.3.1 Estimating roughness

Consider the stationary discrete process $\widehat{\log \tilde{\sigma}}_{(d,s)}$ and let $\rho(h)$ denote the autocorrelation of this process, and $\gamma_2(h)$ the variogram. This process satisfies the assumptions regarding modelling time made in 14.1.1, and recall in particular that all observations in such a process are equidistantly spaced by some distance $\Delta > 0$.
The combination of definition 1 and (47) inspires the following ordinary least squares, OLS, regression:

$$\log \hat{\gamma}_2(h) = c_1 + a \log h + \epsilon_h, \quad h = \Delta, 2\Delta, 3\Delta, ...m\Delta \tag{51}$$

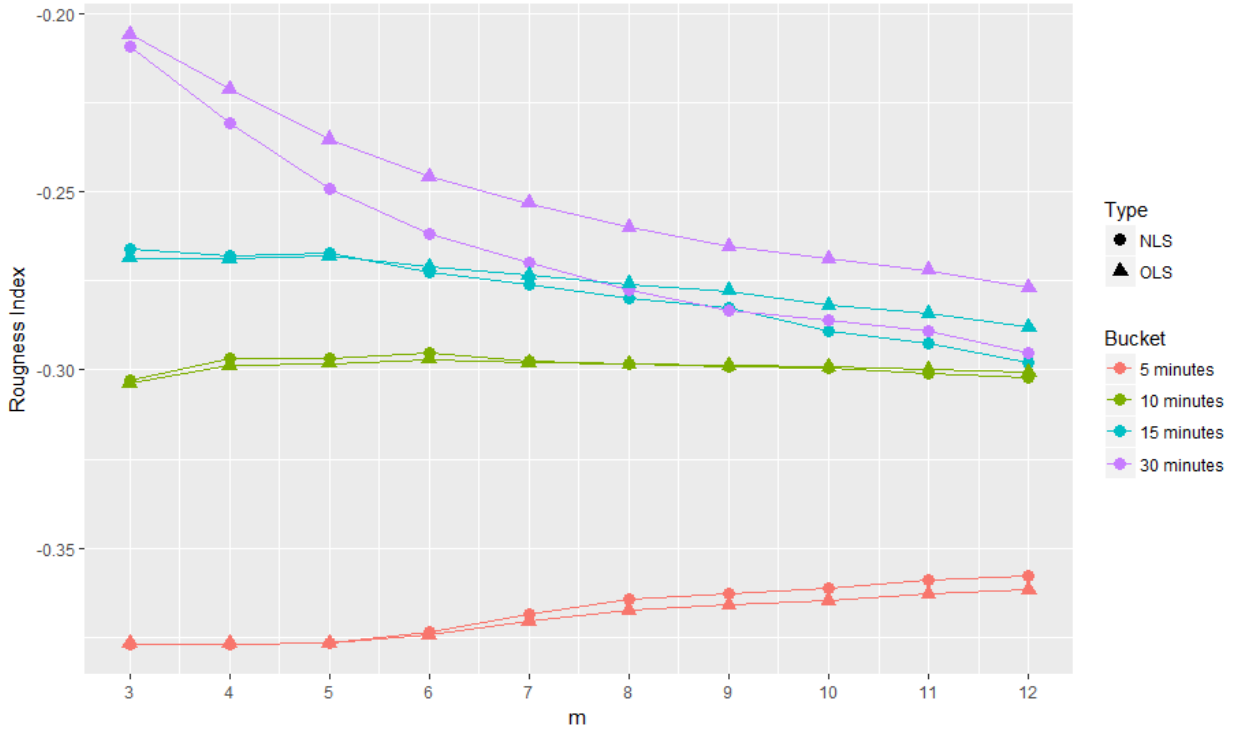
with coefficients $c_1, a \in \mathbb{R}$, for an iid normally distributed sequence ϵ_h and some fixed $m \in \mathbb{N}$. This yields an estimate $\hat{a}$ of the rate of decay of $1 - \rho(h)$, and $\hat{\alpha} = \frac{\hat{a}-1}{2}$ of the roughness index, and will reveal whether the process has short term memory or not.
Note that the constant c_1 in (51) does not correspond to the coefficient in definition 1, but rather one including 2 times the variance of the process in question. However, this estimate is of no interest and will be discarded.
Further, the choice of m_{γ_2} is somewhat tricky, as it should be so small that the asymptotics of definition 1 holds, but also so large that the regression in (51) is reasonable.
Equation 49 indicates that this estimate might be biased. In Bennedsen et al. (2016) theorem 2.1, it is stated that a non-linear least squares procedure, as specified in (52), can provide more robust estimates. However, the assumptions needed for the theorem to hold are very restrictive and will generally not be satisfied.
Ultimately, due to this theorem the following non-linear least squares, NLS, is also considered

$$\hat{\gamma}_2(h) = c_2 h^{2\alpha+1} + \epsilon_h, \quad h = \Delta, 2\Delta, 3\Delta, ...m\Delta \tag{52}$$

for some coefficient $c_2 > 0$.

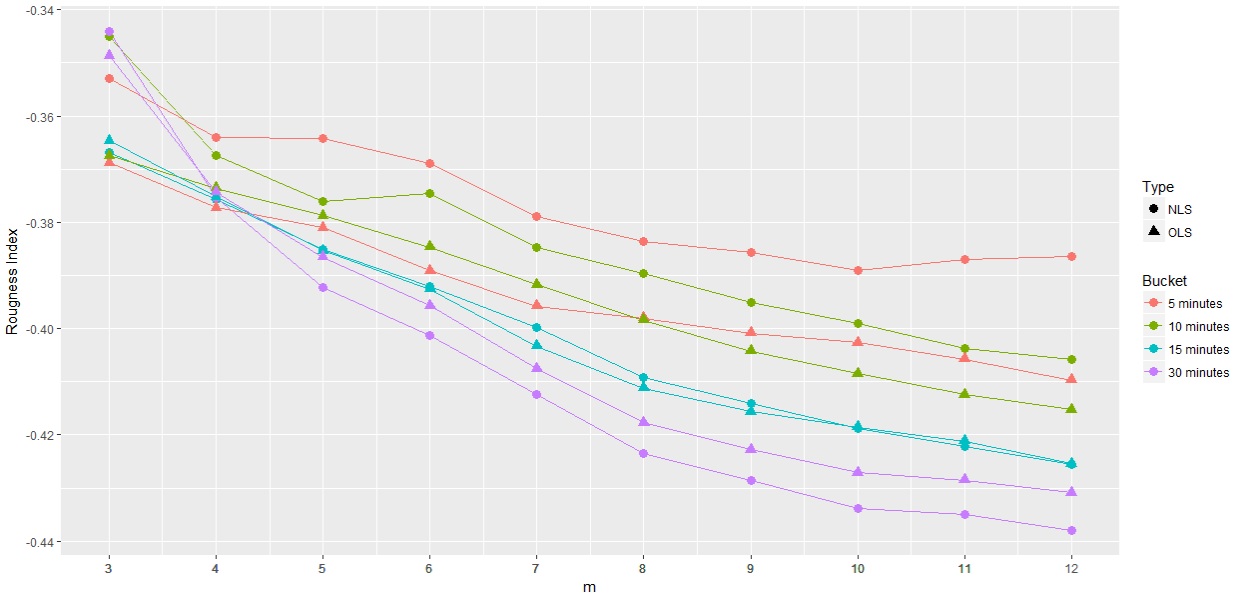
SPY

Estimates of α are shown in figure 56, and there are two major comparisons to be made. First, the estimates should be rather robust to the choice of m, which seems partially true. As argued, robustness for small values of m is the concern and seems to be present in all cases but the 30 minute bucket. Here there actually appears to be quite a lack of robustness, which raises questions of the reasonability of the estimate. Ultimately, we wish to focus attention on shorter buckets, in which case there seems to be a high degrees of robustness for m up to around 6 which corresponds to the findings in Bennedsen et al. (2016). Secondly, there seems to be a high degree of consistency between the OLS and NLS estimates - again with the exception of the 30 minute bucket. Although this indicates result that are not completely in line with Bennedsen et al. (2016). Further investigation into the higher bucket sizes are not carried out, however this estimation has also been carried out at smaller bucket sizes, which as stated earlier, will be discussed separately later.
The overall tendency of a decreasing roughness index, i.e. rougher paths, as bucket widths decrease, while seeming immediately intuitively reasonable does raise questions. Since these values describe different processes, i.e. the autocorrelation functions of different $\widehat{\log \tilde{\sigma}}_{(d,s)}$ with d taking values in the same space but s taking values in different spaces, they are all estimates of the same underlying continuous process $\log \tilde{\sigma}_t$. Robustness across different bucket sizes would have perhaps been even more intuitive than the differences seen. Due to this, one could wonder if perhaps the values m depicted are simply too small to be reasonable estimates, and whether a higher number would reveal a higher degree of agreement across different bucket. This is shown in figure 76 in the appendix, and while no clear agreement presents itself, values greater than -0.35 do seem to generally prevail for bucket of sizes 5 minutes and above. However, this could simply reflect that the autocorrelation functions corresponding to the different bucket sizes are similar, not in the limit as lag tends to 0, but for larger lags.
Ultimately, the differences will be treated as reasonable, and the values will play a key role in modelling $\widehat{\log \tilde{\sigma}}_{(d,s)}$, although this ignores the deeper issues of drawing inference on the underlying model.



Estimated values of α on SPY data, based on estimation procedures (51) and (52) for large ranges of values of m.

Figure 56: Estimates of roughness index α on SPY data



Estimated values of α on Bitcoin data, based on estimation procedures (51) and (52) for large ranges of values of m.

Figure 57: Estimates of roughness index α on Bitcoin data

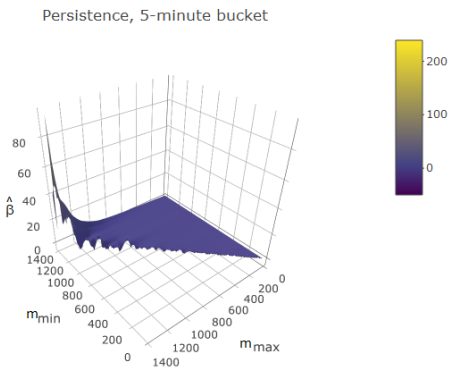


Figure 58: Results of naively estimating roughness for various values of $m_{\min}$ and $m_{\max}$ on SPY data

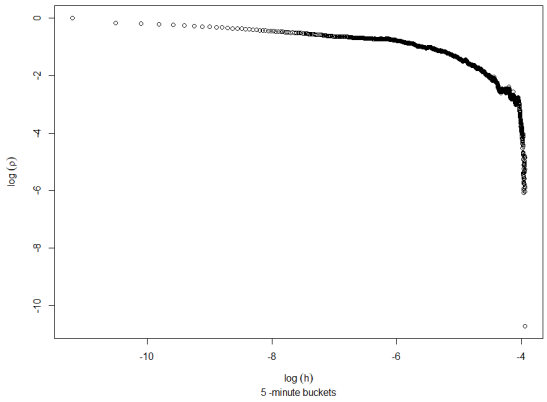


Figure 59: Underlying values for estimating persistence. The persistence is estimated as the slope, as calculate between $m_{\min}$ and $m_{\max}$ on SPY data

Bitcoin
Similarly to calculating the autocorrelation, we only use the observations pairs of observations that are available for a given lag when computing the variogram. Based on this procedure, estimates of α are shown in figure 57. In this case there is a slightly larger difference between the two methods, but the overall level seems to be agreed upon. In this case the effect of increasing m seems quite consistently to reducing the value of α. In figure 77 in the appendix, it can be seen that this trend continues until values of m at least around 36. Since the goal is to estimate the behaviour as m tends to 0. It is not obvious whether the more negative values for larger m actually represent the roughness of the process better. Despite this, the overall level appears to be lower than for the SPY data, indicating that the paths of the price on Bitfinex are more rough than that of the SPY data. This matches with the general intuition of Bitcoin prices as being more violent when compared to SPY prices.

14.3.2 Estimating persistence

Similar to the estimation of roughness, one is inspired by definition 2 to perform the following regression:

$$\log \hat{\rho}(h) = c_2 + b \log h + \epsilon_h, \quad h = m_{\min}\Delta, (m_{\min}+1)\Delta, ..., m_{\max}\Delta, \tag{53}$$

for some fixed $m_{\min}, m_{\max} \in \mathbb{N}$, $m_{\min} < m_{\max}$, and where $\hat{\rho}$ denotes the empirical autocorrelation function with the interpretation $\hat{\beta} = -\hat{b}$. Again the estimate of c_2 is of no interest and will be discarded.
This time both $m_{\min}$ and $m_{\max}$ have to be chosen sufficiently large such that the asymptotics of definition 2 are valid.
Luckily, (50) shows that there is no issues when using noisy observations of the autocorrelation, since after applying the logarithm function, the noise will only affect the estimate of c_2, and not in $\hat{\beta}$.

SPY
In the case of the SPY data, the method specified in (53) is quite troublesome. Firstly, the method is only well-defined for positive autocorrelation, but as seen in figure 53, the autocorrelation is not necessarily positive for large numbers of lags. This approach is therefore carried out with $m_{\max} < N_0$ for $N_0 \in \mathbb{N}$ denoting the first time the autocorrelation is negative. The results of using this method naively can be seen in figure 58. These results are highly unstable, and while most values range from 0 to 4, others end up at several hundred. One could discard these unstable results as a consequence of $m_{\min}$ and $m_{\max}$ too close to each

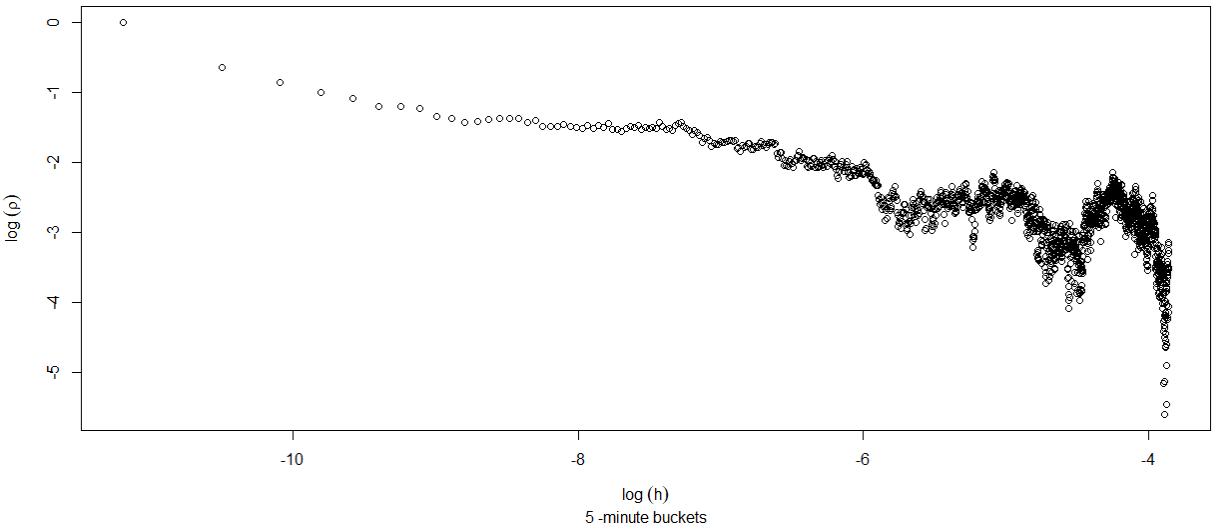
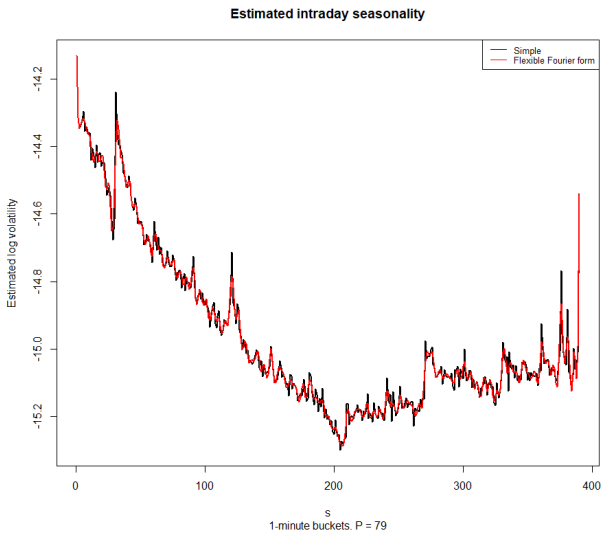


Figure 60: Underlying values for estimating persistence. The persistence is estimated as the slope, as calculate between $m_{\min}$ and $m_{\max}$ on Bitcoin data

other to produce accurate results. However this does not tell the full story. Consider figure 59. It portrays the data on which the regression of (53) is performed. The choice of $m_{\min}$ and $m_{\max}$ is equivalent to choosing a section on which to perform the regression. If chosen small, estimates yields $\hat{\beta} \approx 0.17$. If chosen large, estimates easily increases to $\hat{\beta} \approx 50$ - even after some results have been discarded as being due to $m_{\min}$ and $m_{\max}$ being too close. In Bennedsen et al. (2016) they advocate values $N^{\frac{1}{4}}$ and $N^{\frac{1}{3}}$, however with the SPY data and 5-minute buckets. This yields $m_{\min} = 11$ and $m_{\max} = 26$. Considering the asymptotics of definition 2 these values seem very small - however the results $\hat{\beta} \approx 0.17$ does correspond decently well with the results found in Bennedsen et al. (2016) of around 0.1.

Ultimately, it is unclear which values to deem reasonable and this choice is clearly an extremely important one. It is difficult to argue for a reasonable choice from a theoretical point of view due to the occurrence of negative estimates of autocorrelation. Instead, we will try multiple values whenever needed to. Any result and conclusion based on the choice of value will be analyzed with this choice in mind.
It is worth noting that figure59 does not exhibit a straight line. This suggests that the model suggested in (53) is not well suited, and that the persistence should instead be dependent on, and in particular decreasing in, h. That is that the rate of decay should be increasing in h. It is imaginable that this could yield a model which could initially look quite persistent, but ultimately not have long term memory.

Bitcoin
From figure 60, we see the same issues as with the SPY data. The problems are even more pronounced as the slope is not monotone in h in any way. Beyond this, there is nothing further to add. All the conclusions about the SPY data seem to also hold for the Bitcoin data as well. In particular, while the Bitcoin seems to be more rough, the estimated persistence appears to be of similar level in both SPY and Bitcoin.
This could indicate that to capture both properties of both types of data, a model which can have different levels of roughness while retaining the same level of persistence would be needed.



Constructed analogously to figure 50

Figure 61: Estimated intraday seasonality, i.e. $\widehat{\log \bar{\sigma}_{(s)}}$, for 1-minute buckets

14.4 Discussion of small bucket widths

Thus far the results have been stated for bucket widths of at least 5 minutes wide. Intuitively, there is motivation to use bucket widths as small as possible. Small bucket widths allows for more instantaneous volatility - spot volatility, as opposed to local averages. However, on the other hand the estimate of IV itself also needs to be accurate and this compels us to use wide buckets. Hence we balance the trade-off between accuracy of the estimates and granularity of the estimates.

SPY

On the SPY data, estimates of IV are computed for bucket widths of 1, 2 and 3 minutes. Unfortunately, the results are not reliable as both RV^* and BV^* yield negative estimates. For BV^* it happens respective 242, 5 and 1 time(s). The reason for this is closely connected to the number of observations available after pre-averaging. In all cases there were less than 246 observations available.
It should be noted that for 1-minute buckets, in almost half of the 96720 buckets the number of observations is less than 246. Since the autocorrelation is only negative 242 times out of 96720, it tells us that 246 observations is not necessarily insufficient.
This makes it hard to determine to correct bucket size to use. Even if estimates are positive, they might be so inaccurate that estimates of volatility should not be trusted. The choice of bucket widths of 5 minutes, which is generally used, yields bucket where the lowest number of observations is 237, at the edge of the highest observed negative. While this seems like a promising feature, a check of quality could potentially be carried out by investigating the estimates of IV across different bucket widths to see if there are general levels of agreement across different choices. However this investigation is not carried out in this thesis.

The above deals with the accuracy of the IV estimates, whereas the loss of granularity of wide buckets is also worth a short investigation. Consider figure61 which corresponds to figure 50. One major difference between them pops out - namely the the smaller buckets capture a sudden peak at the final minute of the day. When compared to the distribution of trades throughout the day, this is much more in sync with what was originally found in figure 50. This an example of where some conclusions can be lost by prioritizing accuracy over granularity. Ultimately, there are arguments both for and against using smaller bucket widths. We choose to settle for a bucket width of 5 minutes. Any smaller than that is considered too inaccurate, and any conclusions based on the added granularity is overshadowed by the inaccuracy of our estimation.

Bitcoin
As indicated earlier, the issue of missing observations, cannot naively be resolved by increasing the bucket width. In fact, increasing the bucket width might not increase the number of available meaningful observations. This is the case in our data, since observations happen to occur in such a way that some intervals are almost void of observations. This means that without any addition modelling, negative estimates of IV occur, specifically in the case when using the BV^* estimator - even with buckets as wide as 10-minutes! This is solved by discarding any interval wherever the number of observations in the interval is below a certain threshold.

Choosing this threshold is non-trivial, but also nearly impossible to test the correctness of. As seen in the SPY data, the number of observations to produce reasonable estimates of IV depends on how the observations occur, not necessarily the sheer number. In order to retain as much data as possible, we choose the threshold of 300 observations. It is a decent amount above the approximately 250 observations that proved insufficient in the case of the SPY data. This choice is debatable, but it results in positive estimates of IV for bucket widths down to 5 minutes, which combined with the analysis of the SPY data indicates that the threshold is somewhat solid.

The natural next step is to investigate volatility models that encompasses the features found in this section.

15 Models for volatility[17]

We propose a model at timepoint t of the form

$$\log \bar{\sigma}_t = \log c_\sigma + \xi_t \tag{54}$$

for some constant c_σ such that ξ_t models the deseasonalized log volatility up to a drift of $\log c_\sigma$. Before stating our choice of model, we will briefly consider the criteria for such a model.

15.1 Motivation and fractional Brownian motion

In the above analysis a flexible Fourier form was proposed as a approach model of intraday seasonality. The deseasonalized log volatility process that remains to be modelled showed signs of stationarity, roughness and persistence. That is why many common models such as the Heston model used in section 5.4 is insufficient, they simply don't produce the needed degree of roughness. Although no clear conclusions were drawn regarding the long term memory property, it certainly can not be excluded yet, and even some literature suggests that volatility should be modelled with long term memory such as Andersen & Labys (2003).
One way to model a process with stationary increments is to use Mandelbrot and Van Ness style fractional Brownian motions as defined in Marinucci & Robinson (1999). Such a process can be characterized by a parameter, H, the Hurst exponent, which is related to the persistence parameter β from definition 2 by the equation $H = 1 - \frac{\beta}{2}$. This means that the model has long term memory for $H > \frac{1}{2}$. However, H is also related to the roughness index α from definition 1 as $H = \alpha + \frac{1}{2}$, which unfortunately means that the increments cannot not be rough and persistent at the same time. Models utilizing such a fractional Brownian motion with $H > \frac{1}{2}$ has been termed a fractional stochastic volatility model, FSV, and can be found e.g. in Comte & Renault (1996). However, more recently in Gatheral et al. (2014) it is suggested to use $H < \frac{1}{2}$ which has then been termed a RFSV, rough FSV model.
Alternatively, it is suggested in Christian Bayer (2016) to use a Riemann-Liouville fractional style Brownian motion, which is still characterized by a Hurst exponent, H, ultimately yield the so-called rBergomi model. In this article, the focus is on the roughness, that is $H < \frac{1}{2}$, and is motivated by option pricing and modelling implied volatility. In addition to not being persistent it is not stationary. This is unfortunate as stationarity

[17] Responsible: Mathias

is important when modelling the volatility across longer periods of time.
Due to the above shortcomings of the above models we completely abandon the concept of a fractional Brownian motion, and follow the modelling in Bennedsen et al. (2016), which uses a so-called Brownian semi-stationary process, henceforth referred to as $\mathcal{BSS}$ process as described in Barndorff-Nielsen & Schmiegel (2009). This is a quite general model formulation and allows for specifications that satisfy all the desired criteria of stationarity, roughness and a more flexible degree of persistence, potentially allowing for long term memory.
The way these properties are achieved in both models is through the use of two parameters: one determining the roughness and one determining the persistence.

15.2 The $\mathcal{BSS}$ framework

In order to use the $\mathcal{BSS}$ model as defined in Barndorff-Nielsen & Schmiegel (2009), the overall framework needs to change slightly. Assume therefore $t \in \mathbb{R}$ rather than $t > 0$. This has theoretical implications, but does not change any estimation or other procedures described earlier as observations will still fall at times denoted by values of $t > 0$. However it does yield a new space measurable space $(\Omega, \mathcal{F}, P)$.

The $\mathcal{BSS}$ model is then given by

$$\xi_t = c_\xi + \int_{-\infty}^{t} g(t-s)\nu_s dW_s + \int_{-\infty}^{t} q(t-s)a_s ds$$

for some constant $c_\xi \in \mathbb{R}$, functions g, q and processes ν_s, a_s. In Bennedsen et al. (2016) it is suggested to drop the last term and the constant, i.e have $c_\xi, q(\cdot) = 0$. We adopt this approach, which leaves us with a model

$$\xi_t = \int_{-\infty}^{t} g(t-s)\nu_s dW_s \tag{55}$$

where $g : (0, \infty) \to \mathbb{R}$, g is non-negative and square-integrable, and $(\nu_t)_{t\in\mathbb{R}}$ denotes a adapted stationary process. Further, we will assume that ν_t independent of $(W_t)_{t\in\mathbb{R}}$. Many of the following results hold for more general processes ν_s, as long as they yield a well-defined process, ξ_t and are stationary. However we leave these out of scope of this thesis.
It is stated in Bennedsen et al. (2016) that

$$\xi_t|(\nu_s)_{s<t} \sim N\left(0, \int_{-\infty}^{t} g^2(t-s)\nu_s^2 ds\right) = N\left(0, \int_0^{\infty} g^2(u)\nu_{t-u}^2 du\right) \tag{56}$$

which is in accordance with usual conclusion drawn from Björk (2009). An important result that follows is formulated in the proposition below, which is also stated in Bennedsen et al. (2016).

Proposition 5 (Autocorrelation and stationarity of ξ_t)**.**
Let W_t denote a Brownian motion, let ν_t be a stationary process independent of the Brownian motion, and let ξ_t be given as in (55)*. Then ξ_t is stationary with the covariance given as*

$$cov(\xi_t, \xi_{t+h}) = E[\nu_0^2] \int_0^{\infty} g(u+h)g(u)du$$

for $h > 0$, and autocorrelation $\rho_\xi(h)$ as

$$\rho_\xi(h) = \frac{cov(\xi_t, \xi_{t+h})}{V\xi_t} = \frac{\int_0^{\infty} g(u+h)g(u)du}{\int_0^{\infty} g^2(u)du}.$$

Note that the autocorrelation does not depend on ν_t.

Sketch of proof.
Rather than doing the full proof, we will indicate how the proof would run in a simpler setting, and indicate the possibility of an extension.
Rather than considering the integral in (55), consider the following definition:

$$\xi_t = \int_{-b}^{t} g(t-s)\nu_s dW_s \tag{57}$$

with $b \in [0, \infty]$. Note that $b = \infty$ should be understood as the appropriate limit of the integral corresponding exactly to the integral in (55).

Let now initially $b = 0$. This is referred to as a truncated $\mathcal{BSS}$ model in Bennedsen et al. (2015) and is in fact part of the rBergomi model from Christian Bayer (2016).
This case is simple with a rich literature describing the properties of the stochastic integral. The result can then be computed as shown below.
Let $\mathcal{F}_t^X$ denote σ-algebra $\sigma(X_s : -b \leq s \leq t)$ for any stochastic process X_t. Then by Proposition 4.4 in Björk (2009) we have ξ_t measurable with respect to $\mathcal{F}_t^W$ and $E\xi_t = 0$. Note that a truncated version of the result stated in (56) then follows from by Lemma 4.15 in Björk (2009).

Now we turn to the covariance:

$$\begin{aligned}
\operatorname{cov}(\xi_t, \xi_{t+h}) &= E[\xi_{t+h}\xi_t] - E[\xi_{t+h}]E[\xi_t] \\
&= E[\xi_{t+h}\xi_t] - 0 \\
&= E\left(\int_{-b}^{t+h} g(t+h-s)\nu_s dW_s \int_{-b}^{t} g(t-s)\nu_s dW_s\right) \\
&= E\left(\left(\int_{t}^{t+h} g(t+h-s)\nu_s dW_s + \int_{-b}^{t} g(t+h-s)\nu_s dW_s\right)\int_{-b}^{t} g(t-s)\nu_s dW_s\right)
\end{aligned}$$

where the second equality follows from $E\xi_t = 0$, as argued above.
The above can be split into two parts. However, before continuing, we utilize the following observation, for $a, c \in \mathbb{R}$

$$(a+c)^2 - (a-c)^2 = a^2 + c^2 + 2ac - (a^2 + c^2 - 2ac) = 4ac \tag{58}$$

so $ac = \frac{1}{4}((a+c)^2 - (a-c)^2)$.
Now we want to compute the product of integrals of processes ϕ_t and ψ_t with respect to W_t for two predictable and integrable stochastic processes ϕ_t and ψ_t using Itô's isometry, e.g. corollary 3.1.7 Øksendal (2000):

$$\begin{aligned}
E\int_{-b}^{t}\phi_t dW_s \int_{-b}^{t}\psi_t dW_s &\overset{(58)}{=} E\frac{1}{4}\left(\left(\int_{-b}^{t}\phi_t dW_s + \int_{-b}^{t}\psi_t dW_s\right)^2 - \left(\int_{-b}^{t}\phi_t dW_s - \int_{-b}^{t}\psi_t dW_s\right)^2\right) \\
&= E\frac{1}{4}\left(\left(\int_{-b}^{t}\phi_t + \psi_t dW_s\right)^2 - \left(\int_{-b}^{t}\phi_t - \psi_t dW_s\right)^2\right) \\
&\overset{\text{Itô's isometry}}{=} E\frac{1}{4}\left(\int_{-b}^{t}(\phi_t + \psi_t)^2 ds - \int_{-b}^{t}(\phi_t - \psi_t)^2 ds\right) \\
&= E\left(\int_{-b}^{t}\frac{1}{4}((\phi_t + \psi_t)^2 - (\phi_t - \psi_t)^2) ds\right) \\
&\overset{(58)}{=} E\int_{-b}^{t}\phi_t\psi_t ds
\end{aligned} \tag{59}$$

Utilizing this, we return to the last term in the covariance calculation:

$$\begin{aligned} E\int_{-b}^{t} g(t+h-s)\nu_s dW_s \int_{-b}^{t} g(t-s)\nu_s dW_s &\overset{(59)}{=} E\int_{-b}^{t} g(t+h-s)g(t-s)\nu_s^2 ds \\ &= \int_{-b}^{t} E(g(t+h-s)g(t-s)\nu_s^2) ds \\ &= \int_{-b}^{t} g(t+h-s)g(t-s)E(\nu_0^2) ds \\ &= E(\nu_0^2)\int_{-b}^{t} g(t+h-s)g(t-s) ds \\ &= E(\nu_0^2)\int_{0}^{b+t} g(u+h)g(u) du \end{aligned} \tag{60}$$

the second equality follows by Fubini's theorem, e.g. Hansen (2009). The third follows from the stationarity of ν_s and the last by substitution with $s = t - u$.
Continuing with the first part of the covariance calculation:

$$\begin{aligned} &E\int_{t}^{t+h} g(t+h-s)\nu_s dW_s \int_{b}^{t} g(t-s)\nu_s dW_s \\ &\quad = E\left(E\left(\int_{t}^{t+h} g(t+h-s)\nu_s dW_s \int_{b}^{t} g(t-s)\nu_s dW_s \middle| \mathcal{F}_t^W\right)\right) \\ &\quad = E\left(E\left(\int_{t}^{t+h} g(t+h-s)\nu_s dW_s \middle| \sigma(W_t)\right)\int_{b}^{t} g(t-s)\nu_s dW_s\right) \\ &\quad = E\left(0\int_{b}^{t} g(t-s)\nu_s dW_s\right) = 0 \end{aligned} \tag{61}$$

The first equality follows from the tower property, the second because fact $\int_b^t g(t-s)\nu_s dW_s$ is F_t^W-measurable from Proposition 4.4 in Björk (2009) regarding measurability and the third from proposition 4.7.

Combining (60) and (61) shows that the covariance depends only on ν_t through $E\nu_0^2$. However with $b = 0$, or rather any $b < \infty$, the process is not stationary, which is unwanted but ultimately also uninteresting.

In order for the proof to be complete, the above calculations need to hold in the case where $b = \infty$. We will not attempt to argue why this holds - i.e. that the results regarding measurability, integrability and that both $E\xi_t = 0$ and Itô's isometry still holds. This will simply be assumed.
Let $b = \infty$ and if all of the above holds, the covariance reads:

$$\operatorname{cov}(\xi_t, \xi_{t+h}) = E\int_{-\infty}^{t} g(t+h-s)\nu_s dW_s \int_{-\infty}^{t} g(t-s)\nu_s dW_s + 0 = E\nu_0^2 \int_{0}^{\infty} g(u+h)g(u) du$$

In this case, since the variance of ξ_t is a special case of the autocovariance with $h = 0$, the autocorrelation is determined only by h and not by t. Hence ξ_t is recognized as stationary, exactly as desired, since it has constant, 0 mean, constant variance and a covariance not depending on t, only h. We note that, since $E\nu_0^2$ enters both the variance and the covariance, the autocorrelation does not depend on ν_t in any way, and is therefore solely determined by the function g. □

15.2.1 Some theory regarding $\mathcal{BSS}$ processes

We present some assumptions and results of Bennedsen et al. (2016). The focus will be on the uses of these results rather than the derivations, these can be found in Bennedsen et al. (2016).
For the results in Bennedsen et al. (2016) to hold, we need the following three assumptions:

(A1) For some $\alpha_g \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$,

$$g(x) = x^{\alpha_g} L_0(x), \quad x \in (0, 1]$$

where $L_0 : (0, \infty) \to (0, \infty)$ denotes a slowly varying function at 0 with L_0 being bounded away from 0 and $L_0 \in C^1$. Further $|L_0'(x)| \leq c_L(1 + x)^{-1}$ for $x \in (0, 1]$ and some constant $c_L > 0$.

(A2) $g \in C^1$ and g' is ultimately monotonic and satisfies $\int_1^\infty g'(x)^2 dx < \infty$

(A3) For some $\lambda \geq 0$ and some $\gamma_g \in \mathbb{R}$, such that $\gamma_g > \frac{1}{2}$ when $\lambda = 0$,

$$g(x) = e^{-\lambda x} x^{-\gamma_g} L_1(x), \quad x \in (1, \infty)$$

where L_1 is slowly varying at infinity and bounded away from 0 and ∞ on any finite interval.

(A4) That $\int_0^1 g(x)dx < \infty$

Note, (A4) is added for ease of reference as it is a less strict requirement than (A1). This adds clarity as to which assumptions are needed for which propositions. Further, it is argued in Bennedsen et al. (2016) that (A1) and (A3) together will specify a square integrable function g.

We have the following proposition:

Proposition 6. *Let ξ_t be distributed as in* (55) *and assume that g satisfies (A1) and (A2) then for $h > 0$,*

$$1 - \rho_\xi(h) \sim c_g L_0(h)^2 h^{2\alpha_g + 1}, \quad h \to 0$$

for a constant $c_g > 0$ dependent on g that can be found in Bennedsen et al. (2016).

This can be an important tool when determining the roughness of a $\mathcal{BSS}$ process.
The next two propositions involve persistence.

Proposition 7. *Let ξ_t be distributed as in* (55)*, and assume that g satisfies (A4) and (A3) with $\lambda = 0$, then for $h > 0$,*

(i) If $\gamma_g \in (1, \infty)$ then

$$\rho_\xi(h) \sim c_g L_1(h) h^{-\gamma_g}, \quad h \to \infty$$

for a constant $c_g > 0$ dependent on g that can be found in Bennedsen et al. (2016).

(ii) If $\gamma_g \in (\frac{1}{2}, 1)$ then

$$\rho_\xi(h) \sim c_g L_1(h)^2 h^{1 - 2\gamma_g}, \quad h \to \infty$$

for a constant $c_g > 0$ dependent on g that can be found in Bennedsen et al. (2016).

and the corresponding results for $\lambda > 0$.

Proposition 8. *Let ξ_t be distributed as in* (55)*, and assumes that g satisfies (A4) and (A3) with $\lambda > 0$ and $\gamma_g \in \mathbb{R}$, then for $h > 0$,*

$$\rho_\xi(h) \sim c_g L_1(h) h^{-\gamma_g} e^{-\lambda h}, \quad h \to \infty$$

for a constant $c_g > 0$ dependent on g that can be found in Bennedsen et al. (2016).

All of the above proposition are going to be used to determine the short and long term memory properties of $\mathcal{BSS}$ processes, and are particularly practical when combined with the following simple observation:

Lemma 1. *Let $f_1, f_2, f_3 : (0,\infty) \to (0,\infty)$ denote some generic functions and assume*

$$f_1(x) \sim c_{f,1} f_2(x) f_3(x), \quad x \geq 0$$

for some constant $c_{f,1} > 0$.
If $f_3(x) \to c_{f_3}$ for $x \to \infty$ and some constant $c_{f_3} > 0$. Then

$$f_1(x) \sim c_{f,2} f_2(x)$$

for another constant $c_{f,2} > 0$.

Proof. It is seen that

$$\frac{c_{f,1} f_2(x)}{f_1(x)} = \frac{1}{f_3(x)} \frac{c_{f,1} f_2(x) f_3(x)}{f_1(x)} \to \frac{1}{c_{f_3}} \quad \text{for } x \to \infty$$

So letting $c_{f,2} = c_{f,1} c_{f_3}$ concludes the proof. □

Naturally, this also holds for $x \to 0$, assuming the appropriate limit exists.

This result while very simple, it is quite useful as convergent functions are slowly varying, meaning that for some, but not all, L_0 and L_1, this proposition can be used in conjunction with the above propositions.

15.3 Examples of $\mathcal{BSS}$ process

In this thesis, we will restrict ourselves to two processes, i.e. to two different choices of g - namely the Power-$\mathcal{BSS}$ and the Gamma-$\mathcal{BSS}$ process. Each with their own distinct features.

15.3.1 Power-$\mathcal{BSS}$

Let

$$g(x) = x^{\alpha_g} (1+x)^{-(\gamma_g + \alpha_g)}$$

for $x > 0$, $\alpha_g \in (-\frac{1}{2}, \frac{1}{2}) \backslash \{0\}$ and $\gamma_g > \frac{1}{2}$.
It is shown in Bennedsen et al. (2015) that g satisfies (A1)-(A3) with $L_0(x) = (1+x)^{-(\gamma_g+\alpha_g)}$ and $L_1(x) = (1+x^{-1})^{-(\gamma_g+\alpha_g)}$.
From proposition 5 we recognize in particular the variance of ξ_t for any t,

$$V\xi_t = E\nu_0^2 \int_0^\infty x^{2\alpha_g} (1+x)^{-2(\gamma_g+\alpha_g)} dx = E\nu_0^2 B(2\alpha_g + 1, 2\gamma_g - 1)$$

where B denotes the beta function. The autocovariance of ξ_t can be calculated up to the factor $E\nu_0^2$ by numerical integration from proposition 5.
Note that both $L_0(x) \to 1$ for $x \to 0$ and $L_1(x) \to 1$ for $x \to \infty$. This means that it is possible to combine proposition 1 with both proposition 6 and 7, where the role of f_3 being played respectively by L_0 and L_1. This yields:

$$1 - \rho_\xi(h) \sim c_{g,1} h^{2\alpha_g + 1} \quad h \to 0$$

for some $c_{g,1} > 0$. By definition 1 this means that α_g denotes the roughness index of the Power-$\mathcal{BSS}$. Also if $\gamma_g > 1$ then

$$\rho_\xi(h) \sim c_{g,2} h^{-\gamma_g} \quad h \to \infty$$

for some $c_{g,2} > 0$. In this case ρ_ξ is integrable and we say that ξ_t does not have long term memory according to definition 2 whereas if $\gamma_g < 1$ then

$$\rho_\xi(h) \sim c_{g,2} h^{1-2\gamma_g} \quad h \to \infty$$

again regarding definition 2, this results in $\beta \in (0,1)$ and ξ_t then has long term memory.

15.3.2 Gamma-$\mathcal{BSS}$

Let instead g be given as:

$$g(x) = x^{\alpha_g} e^{-\lambda x}$$

for $x > 0$, $\alpha_g \in (-\frac{1}{2}, \frac{1}{2}) \backslash \{0\}$ and $\lambda > 0$.
It is shown in Bennedsen et al. (2015) that g satisfies (A1)-(A2) with $L_0(x) = e^{-\lambda x}$. Clearly, (A3) holds with $\gamma_g = -\alpha_g$ and $L_1(x) = 1$.
Again, by proposition 5 the variance can be recognized as

$$V\xi_t = E\nu_0^2 \int_0^\infty x^{2\alpha_g} e^{-2\lambda x} dx = E\nu_0^2 (2\lambda)^{-(2\alpha_g+1)} \Gamma(2\alpha_g + 1)$$

where Γ denotes the gamma function. The last equality follows directly from Gradshteyn & Ryzhik (2007) formula 8.312.2.
The covariance can also be found as

$$\operatorname{cov}(\xi_t, \xi_{t+h}) = E\nu_0^2 e^{-\lambda h} \int_0^\infty (u+h)^{\alpha_g} u^{\alpha_g} e^{-\lambda 2u} du = E\nu_0^2 \frac{\Gamma(\alpha_g + 1)}{\sqrt{\pi}} \left(\frac{h}{2\lambda}\right)^{\alpha_g + \frac{1}{2}} K^{(Bessel)}_{\alpha_g + \frac{1}{2}}(\lambda h)$$

where $K^{(Bessel)}_y(x)$ denotes Bessel's function of the third kind, and the last equality follows directly from Gradshteyn & Ryzhik (2007) formula 3.383.8.
Similarly to the Power-$\mathcal{BSS}$, $L_0(x) \to 1$ for $x \to 0$, proposition 1 and 6 can be combined to yield:

$$1 - \rho_\xi(h) \sim c_{g,1} h^{2\alpha_g + 1} \quad h \to 0$$

for some $c_{g,1} > 0$. By definition 1, this once again means that α_g denotes the roughness index.
Further proposition 8 can be used directly to yield:

$$\rho_\xi(h) \sim c_{g,2} h^{-\alpha_g} e^{-\lambda h}, \quad h \to \infty$$

for some $c_{g,2} > 0$. In terms of persistence, this specifies exponential rate of decay, meaning integrability of the autocorrelation and in particular no long term memory. Notably, the value λ, while important in terms of rate of decay, does not correspond the value of β in definition 2.

The next step is to fit these models on data.

15.4 Fitting the Power-$\mathcal{BSS}$ and Gamma-$\mathcal{BSS}$

We present the method of fitting the Power- and Gamma-$\mathcal{BSS}$ as described in Bennedsen et al. (2016).
For this estimation recall that $\widehat{\log \tilde{\sigma}}_n$ denotes the deseasonalized estimates of log volatility. Eventually, the interest is not of this process, but rather the volatility process σ_t. However, some point should be noted first.
Firstly, we can fit the parameter α_g for either process utilizing the OLS method described by (51).
Next, the parameter λ of the Gamma-$\mathcal{BSS}$ cannot easily be estimated so the following method is applied:
The estimated $\hat{\alpha_g}$ is plugged into the theoretically derived autocorrelation. Next we numerically find the value of λ, or equivalently γ_g, that minimizes the squared errors between the resulting autocorrelation function and the empirically observed - once again using non-linear least square. Letting $\bar{\rho}_{\hat{\alpha}_g, \lambda}(h)$ denote the autocorrelation resulting from plugging in $\hat{\alpha}_g$, we can phrase the estimation as:

$$\hat{\lambda} := \arg\min_\lambda \sum_{i=1}^{m_{\max}} \left(\bar{\rho}_{\hat{\alpha}_g, \lambda}(i\Delta) - \hat{\rho}_\xi(i\Delta)\right)^2 \tag{62}$$

where Δ and $m_{\max}$ is to be understood as in equation (53). The same method can be applied to achieve estimates of γ_g. Compared to the persistence estimation described above, we only have the one parameter $m_{\max}$. It is once again important that $m_{\max}$ is chosen sufficiently large, whereas there is no choice of a minimum lag. This is due to the fact that we should fit the full autocorrelation, since we in this method also include the α_g, which controls the autocorrelation for small lag values. As discussed in earlier sections, choosing $m_{\max}$ is not trivial and the performance of different values will be investigated.
Next, from (54) we have

$$\log \tilde{\sigma}_t = \log c_\sigma + \xi_t$$

with $E\xi_t = 0$, which follows from (56). This means that the sample mean of $\widehat{\log \tilde{\sigma}}_n$ will be an unbiased estimator of $\log c_\sigma$ up to the unbiasedness of the estimates of $\widehat{\log \tilde{\sigma}}_n$ themselves.
Finally, since ξ_t has constant variance and mean 0, by proposition 5 and (56), let $\widehat{V\xi_0}$ denote the sample second moment of $\widehat{\log \tilde{\sigma}}_n - \widehat{\log c_\sigma}$. This could be referred to as the *de-meaned* version of $\widehat{\log \tilde{\sigma}}_n$. Now, $\widehat{V\xi_0}$ is used as an estimator of the variance of ξ_t. Then by proposition 5, this is sufficient to calculate $E\nu_0^2$, however as remains to be seen, this value will be not be needed for predictions to be made.

To summarize, *the process* will in the following refer to $\log \tilde{\sigma}_t$, i.e. the deseasonalised log volatility:

1. Fit α_g by estimation of the roughness of the process.
2. Fit λ or γ_g by non-linear least squares, minimizing the square difference between the theoretical autocorrelation with the α_g from step 1 plugged in and the observed autocorrelation as described in (62).
3. Estimate $\log c_\sigma$ by sample mean.
4. Estimate the variance of ξ_t by the sample second moment of the de-meaned process, utilizing the estimate of $\log c_\sigma$ achieved in step 3.

15.4.1 Critique of the method

From a mathematical point of view, there are several points to be made about this procedure. First, regarding step 1 and 2, no argument has been presented as to why this stepwise procedure should be convergent. In particular, the non-linear least squares is not argued to converge, and it is unanswered whether estimating both α_g and λ or γ_g separately would be a better or worse method. Let this be denoted as an alternative method which will be tested as well.
Second, regarding 3 and 4, once again there has not been made enough assumptions to justify the method. Clearly the estimator in step 3 is unbiased, but neither estimator is guaranteed to be consistent. To make a consistent estimator in step 4, one would need a version of the law of large numbers. To make such a version applicable, one would could assume ergodicity of ξ_t, however that will be a strong additional assumption. The assumption of ergodicity on ξ_t would be sufficient to yield a consistent estimator in step 3 and 4, and in both cases one could apply corollary 2.3.13 of Sokol & Rønn-Nielsen (2015).

15.4.2 The parameter dimension

The above specifies a method utilizing a number of parameters. Isolating the parameters used after estimates of $\log \tilde{\sigma}_t$ are available, we are left with:
One m from step 1 and another similar value $m_{\max}$ from step 2. With the alternative method of collapsing step 1 and 2, only the parameter $m_{\max}$ is needed. However, what has been left out so far, is which data to use for estimation.
As time passes, it is possible to fit the autocorrelation on more and more data, eventually fitting to the entire dataset. This would be in line with the mathematical model described, however, if the model does not in fact follow a model with constant parameters, but rather one with changes over time, one could potentially achieve better results using a backwards horizon that is shorter as a shorter backwards horizon relies less on

the assumption of stationarity.
Note that the stationarity test is an attempt to rule out some changes over time. The conclusion of this test was in favor of stationarity - therefore it may be reasonable to use all data. This is a point to return to when evaluating the model later on.
Let $(t_n)_{n=1,...N}$ refer to time points corresponding to observations of the log volatility process. Consider a procedure of consisting of fitting at every t_n for $n \geq n_0$ for some $1 < n_0 < N$. Let the number of observations from time t_1 to t_n be denoted N_n and let b denote a backward horizon for each t_n. The thought here is to only use observations within the backwards horizon for steps 1, 3 and 4 to fit the model. As stated, one might consider using the full horizon $b_n = N_n$ with a sequence of b_n rather than a fixed value, however as argued some $b < N_n$ could potentially produce better local fits.
In general, let also $m_{\max}$ depend on n and be denote $m_{\max}$. It is naturally understood that $m < b$ and $m_{\max} < N_n$.
Further, in step 2 the memory properties are estimated which requires a large amount of observations to be estimated accurately. Due to this, $\hat{\rho}_\xi$ from in (62) will always be estimated with all the n_0 observations. This means that we use all available data at the first fit, but roll the window of observations from there on out.
In total we are left with the additional parameter b and less importantly n_0.

15.5 Final assumptions and prediction in Gaussian models

As suggested in the title of the this section, the final assumption is that ξ_t is Gaussian distributed. Note that there exists a valid choice of ν_t to ensure Gaussianity - namely $\nu_t = c_\nu$ for some constant $c_\nu > 0$ as the distribution of ξ_t is then given in (56). We will neither prove nor disprove the existence of another ν_t that would yield a Gaussian distribution as that would not aid our cause. Further, the Gaussian assumption does not necessarily yield ergodicity. This can for instance be as seen in section 5.10 in Grenander (1950) stating that a Gaussian stationary process with continuous autocorrelation function is ergodic if and only if it has continuous spectral distribution function.
From equation (2.24) in Glasserman (2004) we have for a partitioned Gaussian vector (Z_1, Z_2):

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

with Σ denoting the covariance matrix, the conditional distribution:

$$(Z_1 | Z_2 = z_2) \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(z_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

which can be rewritten in terms of correlation matrix since if $\mu_1, \mu_2 = 0$ and Z_1 is a real 1-dimensional random variable

$$(Z_1 | Z_2 = z_2) \sim N\left(\Gamma_{12}\Gamma_{22}^{-1} z_2, VZ_1\left(\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}\right)\right) \tag{63}$$

trivially by utilizing that for a real constant, c, and matrix, A, it holds that $(cA)^{-1} = c^{-1}A^{-1}$.
Considering the backwards horizon b_n, let ξ_{t^*}, for some $t^* \in (t_n)_{n\in\{2,...,N\}}$ with $t_n < t_*$ and $n \in \{1, ..., N-1\}$, play the role of Z_1 in (63) and $(\xi_{t_{n-b_n}}, ..., \xi_{t_n})$ play the role of Z_2.
Note that $V\xi_{t^*} = V\xi_{t_n}$, which means that the estimate of the future variance is reasonable. Further, this shows that only the correlation is needed, not $E\nu_0$. Only the conditional mean is needed to predict the future ξ_t, meaning that the variance estimate is not needed.
Note that the autocorrelation considered in this case is the theoretical autocorrelation implied by the parameters estimated in step 1 and 2 of the section regarding fitting the model.
Since the ultimate goal is to estimate $\exp(\xi_{t^*})$, it is suggested to view this as log-normally distributed and hence use the estimate

$$\widehat{\exp(\xi_{t^*})} = E\left(\exp\xi_{t^*} | (\xi_{t_{n-b_n}}, ..., \xi_{t_n})\right) = \exp\left(E\left(\xi_{t^*} | (\xi_{t_{n-b_n}}, ..., \xi_{t_n})\right) + \frac{1}{2}V\left(\xi_{t^*} | (\xi_{t_{n-b_n}}, ..., \xi_{t_n})\right)\right) \tag{64}$$

to avoid the bias described by Jensen's inequality, e.g. Hansen (2009), when transformed by a convex function - i.e. $x \mapsto \exp(x)$
This leaves another unanswered question:
Since the original estimates of ξ_t come from taking the log of estimates of volatility, the estimates $\widehat{\log \tilde{\sigma}_n}$ are already biased. So if the conditional mean, calculable from (63), does actually yield an accurate estimate of $\log \tilde{\sigma}_{n+1}$, then the bias from the convex function exp might be desired and a better estimate would be

$$\widehat{\exp(\xi_{t^*})} = E\left(\exp \xi_{t^*} | (\xi_{t_{n-b_n}}, ..., \xi_{t_n})\right) = \exp E\left(\xi_{t^*} | (\xi_{t_{n-b_n}}, ..., \xi_{t_n}\right) \tag{65}$$

This approach also has the advantage that the variance $V\xi_t$ is not needed, and hence the issues with step 3 and 4 of fitting in section 15.4, relieving an assumption such as ergodicity.
Both methods (64) and (65) will be tested.

16 Predicting volatility[18]

Most importantly, since the volatility is not observable, the attempt will not be to estimate volatility directly. Rather, we will attempt to estimate future BV^* estimates, which in terms of accuracy is the equivalent to estimate the estimates of volatility that are generated by the BV^* estimator.
To fully clarify the predict setup, consider the following setup:
Take initially some data and a Δ. This yields equidistant observations at timepoints $(t_n)_{n=1,...,N}$. Now consider a subsequence with $Q < N$, $(s_q)_{q=1,...,Q} \subset (t_n)_{n=1,...,N}$, a backward horizon sequence $(b_q)_{q=1,...,Q}$ and a prediction horizon h_p such that in the subsequence $s_1 = t_{n_0}$ for some $n_0 \geq b_1$ and such that $Q = N - h_p + 1$ and for $1 < q < Q$, $s_q = t_{n_0+q-1}$. In words, the subsequence starts at the n_0'th value and *hits* every point until the $N - c_q$'th at which point it stops. In practice n_0 needs to be sufficiently large that the estimate of $\hat{\rho}_\xi(\Delta i)$ is reasonable for all $i = 1, ..., m_{\max_1}$.
With this construction, at every s_q a model will be fitted and used to predict the BV^* at the s_{q+h_p}'th timepoint. If the model is a $\mathcal{BSS}$ model, observations used to fit the model will be the observations in the interval $[s_{q-b_q}, s_q]$.

The final transformation to volatility is given by:

$$\widehat{\sigma}_t = \exp(\widehat{\log c_\sigma}) \widehat{\exp(\xi_t)} \exp(\widehat{\log \tilde{\sigma}_t}) \tag{66}$$

the last term simply adds the estimated deseasonality factor from earlier sections and the first term adds the estimated sample mean of the log-process before taking the exponential function.

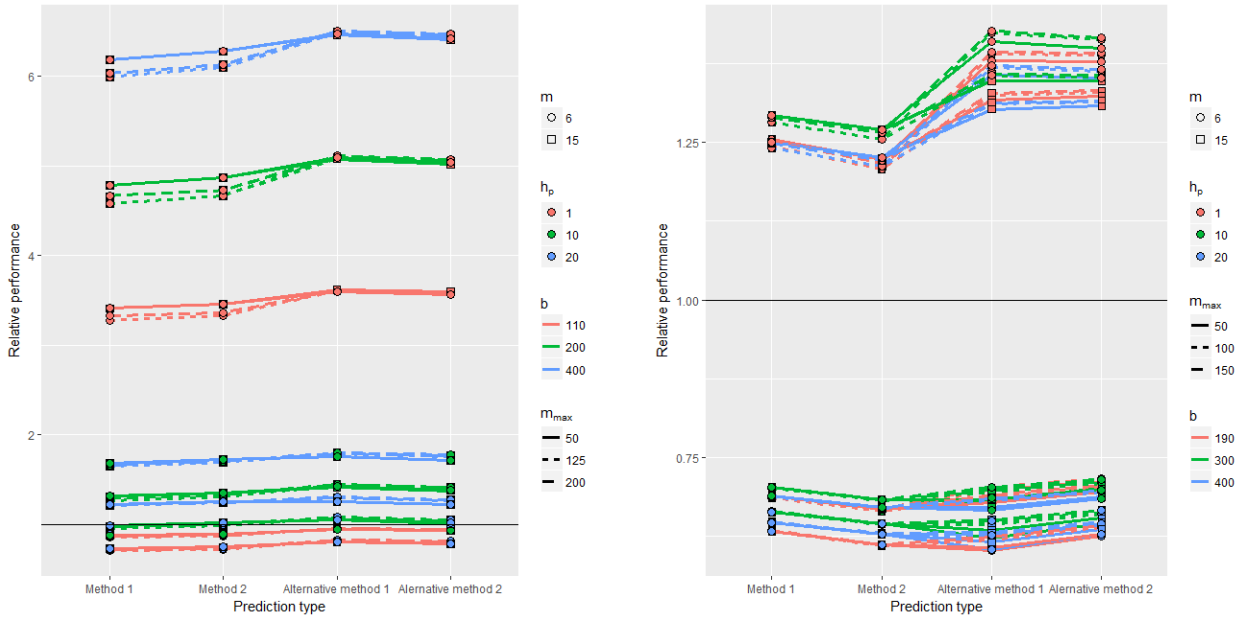
16.1 Initial results of prediction

To evaluate the effectiveness of the predictions generated, we consider the mean squared error:

$$MSE = \frac{1}{Q} \sum_{i=1}^{Q} \left(\hat{\sigma}_{n_0+i-1} - \sigma_{n_0+i-1}\right)^2$$

here $\hat{\sigma}_n$ denotes an estimate as described above, and σ_n denotes the volatility estimates associated with the BV^* estimated. Note that the usual decomposition of MSE into variance and bias squared is not sensible when comparing timeseries. Further, in order to investigate the quality of the prediction, we also compute the simple estimate:
For a given horizon h_p let $\hat{\sigma}_{n+h_p} = \sigma_n$ i.e. a static expectation.

[18] Responsible: Mathias



On SPY data

On bitcoin data

Figure 62: Relative performance of the Gamma-$\mathcal{BSS}$ model

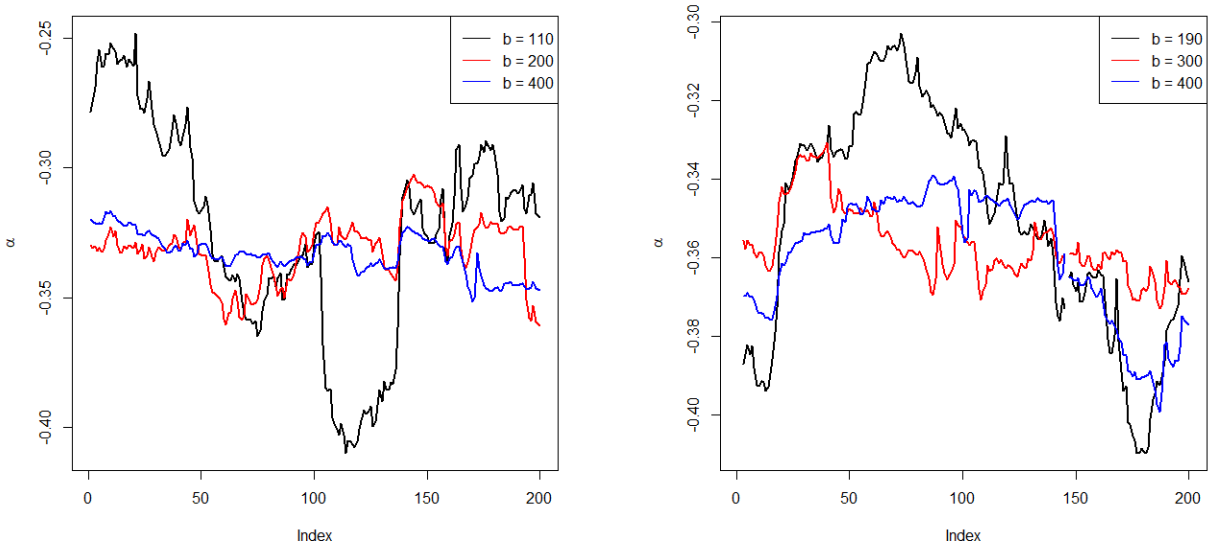
The way to evaluate the fit is to take MSE of the predictions made from a $\mathcal{BSS}$ model normalized by the MSE of the static expectation. This will be referred to as the relative performance below, and since it is a measure of the error, smaller numbers correspond to better performance.
Consider first the Gamma-$\mathcal{BSS}$ on SPY data, for 10 minute buckets, yielding $N = 9672$, With parameters:

1. $n_0 = 4000$
2. $h_p \in \{1, 10, 20\}$
3. $b \in \{110, 200, 400\}$
4. $m_{\max} \in \{50, 125\}$
5. $m \in \{6, 15\}$

Similarly, Gamma-$\mathcal{BSS}$ on bitcoin data, for 10 minute buckets, yielding $N = 7367$, with parameters:

1. $n_0 = 3000$
2. $h_p \in \{1, 10, 20\}$
3. $b \in \{190, 300, 400\}$
4. $m_{\max} \in \{50, 150\}$
5. $m \in \{6, 15\}$

In total there are four predictions for each combination of the above parameters which are all illustrated in figure 62. Several clear conclusions can be drawn from this:
Firstly, there seem to be no significant difference between the four methods. If anything, the primary method which utilizes the OLS estimator of roughness performs better than the purely numerically fitted estimate. Next, the value of m seems completely irrelevant, which was somewhat expected as estimation of α seemed robust to the choice. In terms of relative performance, $m_{\max}$ is surprisingly irrelevant. This begs the question of whether estimation of persistence is robust to this parameter, or whether the persistence is irrelevant to



On SPY data

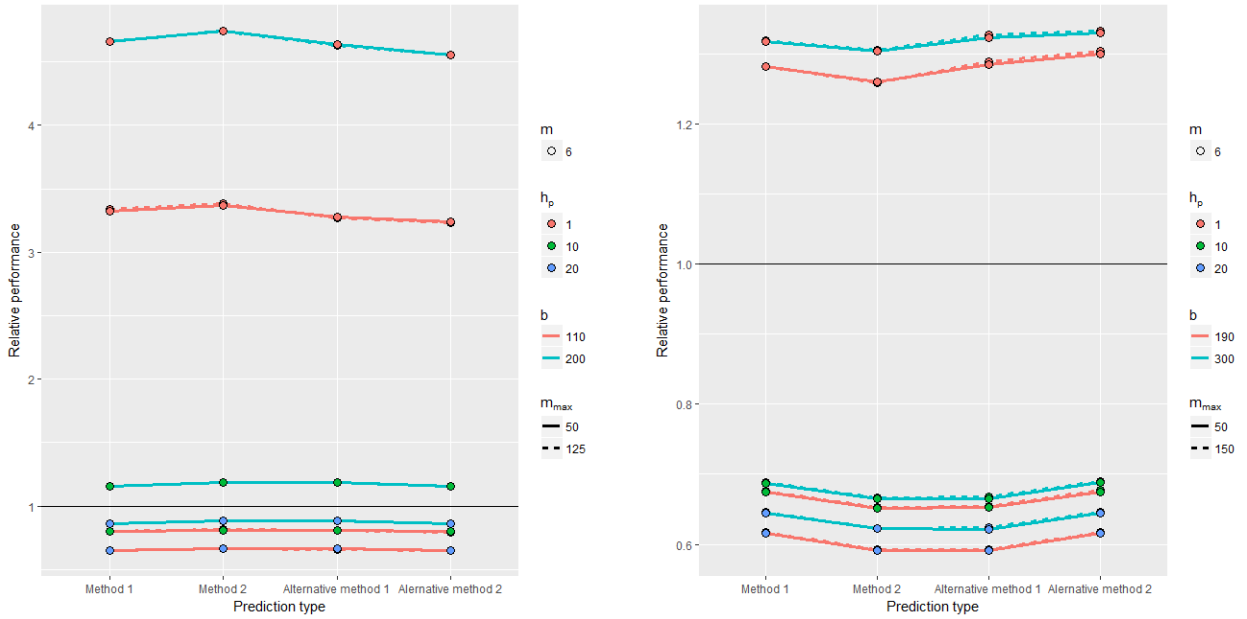
On bitcoin data

Figure 63: The first 200 roughness estimates of the Gamma-$\mathcal{BSS}$ models, with $h_p = 10$ and $m = 6$

the prediction. This will be investigated below. Finally, the values of b and h_p are by far the most important to consider. For h_p the larger values the performance by the Gamma-$\mathcal{BSS}$ model is better. The performance at $h_p = 1$ is particularly poor. Regarding the value of b, smaller values yield better performance. This is clearly evident in the SPY data and slightly less so in the bitcoin data.
Ultimately, the model performs surprisingly poorly and beats the static expectation only for the longest choice of h_p and lowest choice of b. This indicates that some local features are perhaps better captured by a shorter backwards horizon b. Since the main use of the backwards horizon is to construct the variogram used to estimate α, the results would indicate a time dependent α, i.e. that roughness changes over time.

16.2 Discussion of roughness in predictions

Firstly, a brief investigation into the time dependence of parameter estimated from b reveals that they change much more rapidly for smaller values of b than higher. This is seen in figure 63, which in turn yield much better predictions as seen above. There is however a limit to how small b can be as the estimate becomes unreliable for too small values. This is particularly important for the bitcoin estimates, as reflected in the much larger smallest value of b, i.e. 190 rather than 110 which is used for SPY, since the roughness is much closer to the lower bound of -0.5. The fact that we in any case see roughness values significantly smaller than 0 indicates quite clearly that a model with the ability to generate rough paths is needed. This excludes the CIR process used in the Heston model.
As as final note, it is worth noting that the parameter α describes the behaviour of the autocorrelation for short time intervals. Hence, it is possible that the importance of α compared to the memory parameter can simply be explained by the short prediction horizon. However, since the goal was to model intraday volatility, a short prediction horizon is necessary.



On SPY data

On bitcoin data

Figure 64: Relative performance of the Power-$\mathcal{BSS}$ model

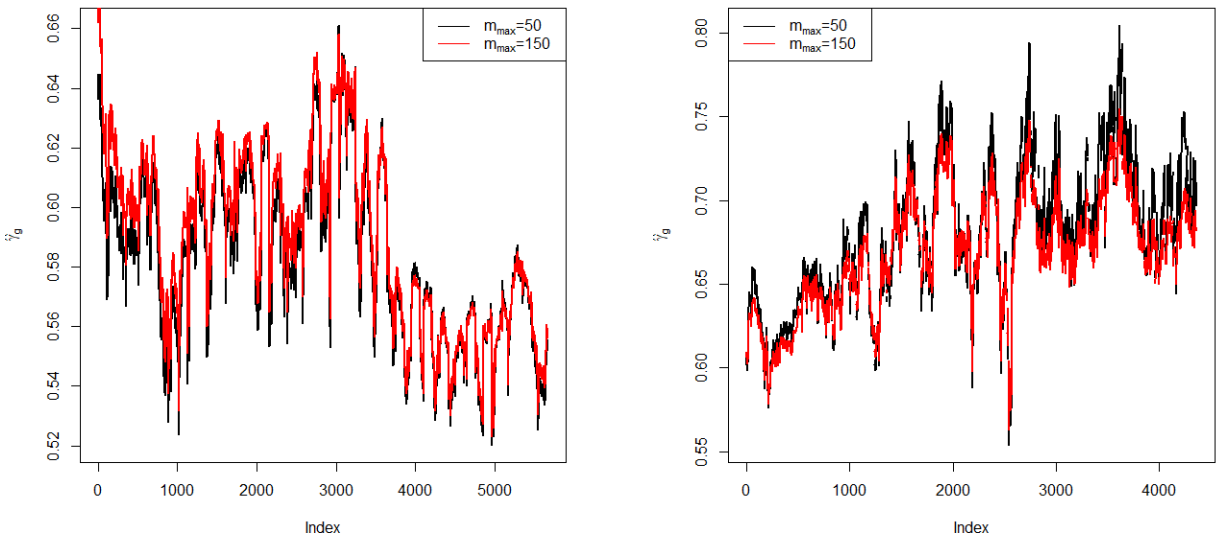
16.3 Discussion of persistence in predictions

Completely analogously to the predictions for the Gamma-$\mathcal{BSS}$ models, predictions for the Power-$\mathcal{BSS}$ models are computed. Some points should be taken into considerations immediately. Firstly, that this requires many times more computational effort due to needed numerical integration of function g compared to the usage of highly optimized evaluations of Γ and Bessel functions. This is particularly important as the autocorrelation function is fitted numerically, effectively by performing a large number of evaluations of the autocorrelation function. Secondly, a fewer number of parameters are interesting. The difference between prediction method 1 & 2, should be expected to be the same. Similarly, the effect of changing m will be the exact same as above, as it solely changes the estimate of α, and does so in a negligible way, hence it is not tested again. Ultimately, the following parameters are tested for SPY:

1. $n_0 = 4000$
2. $h_p \in \{1, 10, 20\}$
3. $b \in \{110, 200\}$
4. $m_{\max} \in \{50, 200\}$
5. $m \in \{6, 15\}$

and for bitcoin:

1. $n_0 = 300$
2. $h_p \in \{1, 10, 20\}$
3. $b \in \{190, 300\}$
4. $m_{\max} \in \{50, 100, 150\}$
5. $m \in \{6, 15\}$



On SPY data

On bitcoin data

Figure 65: Persistence parameter estimates of the Power-$\mathcal{BSS}$ models, with $h_p = 20$ and $b = 50$

In Figure 64, all the same characteristic from Figure 62 are found. As expected, the type of prediction method is largely irrelevant, although the performance of the alternate methods on bitcoin is somewhat better compared to the Gamma-$\mathcal{BSS}$. Once again, we observe a good performance for large prediction horizons and small values of b, indicating again that changes to roughness over time is an important factor. This leaves the importance of persistence as an open question.
There is no question that the memory parameter changes over time. This is demonstrated in Figure 65 but holds generally, and is not surprising as changes in roughness in all likelihood would imply changes in memory parameter. As is seen in Figure 65 for all predictions in the Power-$\mathcal{BSS}$ models, $\hat{\gamma}_g < 1$, which means that the model incorporates long term memory. Notably, the estimates of $\hat{\lambda}$ in the Gamma-$\mathcal{BSS}$ were all of an order of magnitude of 10^{-2}, corresponding to a highly persistent model, although as argued above without the presence of long term memory.

Finally, comparing the best configurations yield relative performance values on SPY of: 0.647 and 0.706 for the Power- and Gamma-$\mathcal{BSS}$ respectively, and on bitcoin of: 0.591 an 0.602 for the Power- and Gamma-$\mathcal{BSS}$ respectively. While this shows a slight favor of the Power-$\mathcal{BSS}$ model, and thereby the long term memory, the difference is very small, particularly so in for bitcoin, meaning no clear conclusions can be drawn. Ultimately, these differences are not deemed sufficient enough to determine which model is more correct. However, there is still more to be learned about the errors found in this study.

16.4 Examples of pathwise prediction errors

So far, all analysis has been based on the performance of predictions relative to the static expectation. To expand on this, consider figure 66, where the squared errors are shown across time. There are several periods with massive prediction errors. To examine the cause, we plot the price and T-statistic as seen in section 12 around the time of the largest squared errors shown in figure 67. This shows that the large prediction error occur around the same time as a quick shift in price, notably around the same time as the T-statistic is significant with 95% confidence, i.e. when a burst occurs. Now, one possible interpretation is simply to dismiss this as the break between two structural periods resulting in prediction error. However, in the light

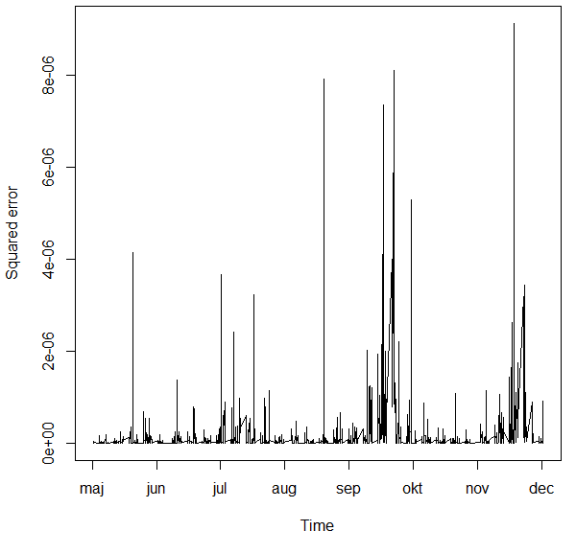


Figure 66: Squared prediction error for the Gamma $\mathcal{BSS}$ model, with $h_p = 10$, $b = 110$, $m_{max} = 50$ and $m = 6$. The largest error is found for the prediction of the interval 14:00-14:10, 14 December 2014

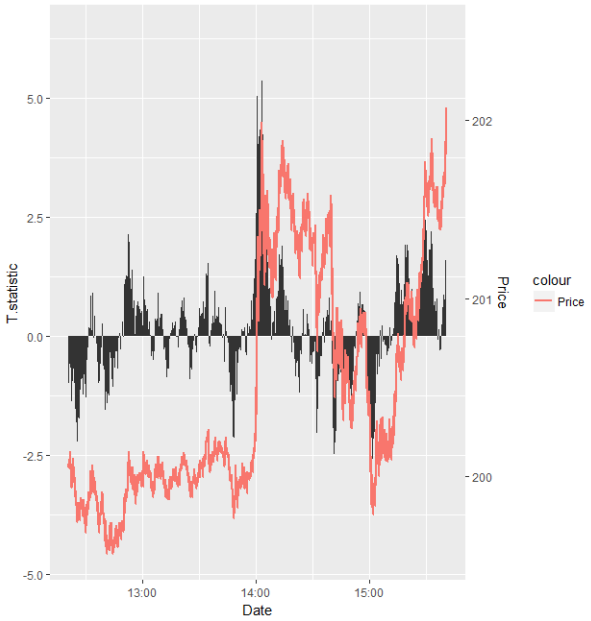


Figure 67: Price and T-statistic for the interval around 14:00-14:10 on 14 December 2014, where the largest burst of the day is detected.

On SPY data

of the earlier analysis, bursts might serve at least partly as an explanatory factor. In fact, this could indicate that incorporation of bursts in one form or another is exactly what might be missing in order to capture some of the remaining large prediction errors.

Similarly, we investigate the prediction errors on Bitcoin data for chosen parameters in figure 68. The clustering is not quite as apparent. However, there are some again large spikes in error. Again, we plot the price process and T-statistic around the time of the error. This is seen in figure 69 where significant T-statistics are found once again. This further supports the suspicion that bursts are a contributing factor to the prediction errors. It is however far from clear whether the prediction error is due to the significant T-statistics or not. Mainly because the time of largest prediction error on this day does not match the time of the largest magnitude of the T-statistics. It is clear that the shorter the backwards horizon is, the more affected is our predictions in case of a burst.

The above are just two examples, and although such obervations can be found in both the Power- and Gamma-$\mathcal{BSS}$ models, it cannot be expected to find significant T-statistic at every large prediction error. Still, given the above evidence, further investigation on the *interplay* between the $\mathcal{BSS}$ models and the burst framework is needed. Notably, one could question both the quality of the BV^* estimator during bursts as an explanation to both poor fits but also inaccuracies in the estimated volatility, i.e. the valued forecasted. Similarly, it sparks interest of whether the observed significant T-statistics are endogenous to the $\mathcal{BSS}$ model or an indication of something missing. Something potentially modelled through the deterministic integral as determined by the function $q(\cdot)$ and the stochastic process a_s that was initially discarded from the model, however while such extensions are possible, they are not attempted in this thesis.

Instead we will return to the burst framework to improve the simulation studies with an alternative volatility process to that of the Heston.

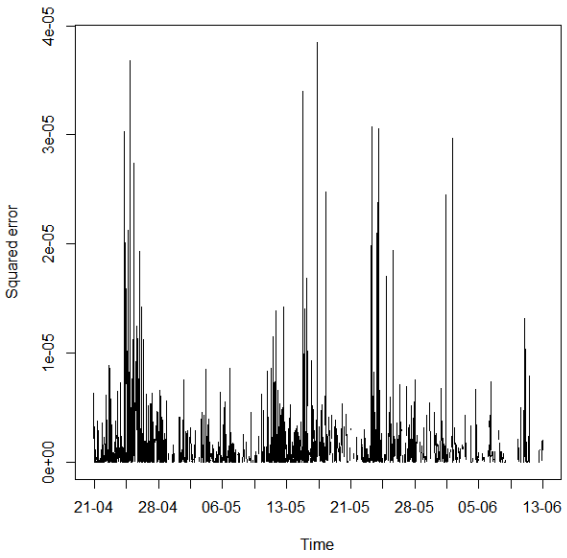


Figure 68: Squared prediction error for the Gamma $\mathcal{BSS}$ model, with $h_p = 20$, $b = 300$, $m_{max} = 100$ and $m = 6$. The largest error is found for the prediction of the interval 07:50-08:00, 17 April 2018

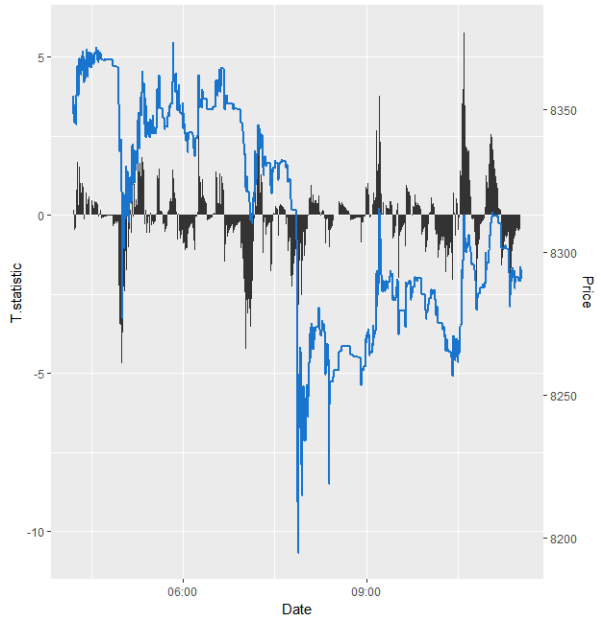


Figure 69: Price and T-statistic for the interval around the interval 07:50-08:00 on 17 April 2018, where a large absolute value of the T-statistic occurs.

On bitcoin data

17 Performance of the T-estimator in $\mathcal{BSS}$ framework

We wish to simulate under the $\mathcal{BSS}$-framework to examine how the the T-estimator behaves under this model, and whether the theoretical problem discussed in Section 12.4 seems to be a serious issue in practice and thereby invalidating the results found with different bandwidths.

17.1 Simulation study using $\mathcal{BSS}$

In previous sections a model for volatility has been proposed, and can be summed up in equation 66, which consists of two important component. The first is the intraday seasonal effects, which was modelled by dividing time into buckets. However, utilizing the flexible Fourier form as described in equation 45, the seasonal component is available for general time points, simply by letting the value of s be non-integer. The second component is the $\mathcal{BSS}$-component that models the deseasonalized volatility which was described in Section 15.2. In order to fully specify a $\mathcal{BSS}$ model such that is can be simulated from, one needs both a function g and a process ν_t. So far, the process ν_t has not been fully specified, but we will in this section assume it simply constant. This in turn makes the $\mathcal{BSS}$ model Gaussian. Fitting such a model is already described in Section 15.4 with estimation of $\nu_t = \nu \in \mathbb{R}$ effectively included in step 4 of the described estimation procedure.
Let $\sigma_t^{\mathcal{BSS}}$ denote a process determined by equation 66 with constant ν, some function g and some piecewise continuous intraday seasonality component and consider the following model:

$$\begin{aligned} d\ln X_t &= \mu_t dt + \sigma_t dW_t \\ \sigma_t &= \sigma_t^{\mathcal{BSS}} \end{aligned} \tag{67}$$

with some $X_0 > 0$ and where $\sigma_t^{\mathcal{BSS}}$ and W_t are assumed independent. With this formulation, the dynamics of X_t can easily be found using Itô's Lemma, however, in the study, we will remain in the world of simulating

log-processes in line with the procedure in the previous burst analysis. Further, μ_t will be assumed to be 0 unless otherwise specified. Let $0 = t_0 < ... < t_N$ for some $t_N > 0$ denote a time discretization, then one can recognize the corresponding euler scheme as

$$Y_{t_{n+1}} = Y_{t_n} + \mu_n \Delta t_{n+1} + \sigma_n^{\mathcal{BSS}} \sqrt{\Delta t_{n+1}} Z_{n+1}$$

with $\Delta t_{n+1} = (t_{n+1} - t_n)$, μ_n and $\sigma_n^{\mathcal{BSS}}$ denoting respectively μ_{t_n} and $\sigma_{t_n}^{\mathcal{BSS}}$ and $(Z_n)_{n=1,...,N}$ are independent and $Z_n \sim N(0,1)$ for all n. Here the process Y_n is the process simulated, and is an approximation of $\ln X_t$.
The simulation of $\sigma_n^{\mathcal{BSS}}$ is simple, and done by calculating the Cholesky decomposition of the covariance matrix of the vector $(\xi_1, ..., \xi_N)^T$ as determined by proposition 5, which utilizes that (Z_n) and (ξ_n) are assumed independent.

Due to memory limitations (16 GB), we restrain ourselves to a simulation with 20,000 trades per day instead of 23,400. This is such that the covariance matrix is of dimension $20,000 \times 20,000$ and can be used in calculations in memory. The conclusions from this study are still meaningful, as this is less than a 15% decrease in the number of observations.

17.1.1 Intraday pattern of T-estimator

In Section 12.4 a potential issue regarding the T-estimator with different bandwidths was discussed. It was shown, that there was an intraday pattern in the drift and volatility estimator one that the Heston model could not describe, and proclaimed that this might invalidate the conclusions regarding burst detection on SPY.

We conduct a simulation study similar to the seasonality study in Section 12.4.1 to examine the behavior of the estimators under this model. Figure 70 shows the intraday pattern of the drift and volatility estimator under the $\mathcal{BSS}$ simulation described above. We see that it exhibits the almost exact same intraday behavior as the pattern found in the data in Figure 44. Thus the model contains factors that could explain the results we see in data.
Figure 71 shows what this implies for the T-estimator. It is very evident that the seasonality pattern for the estimator found in Figure 45 is very well portrayed by the $\mathcal{BSS}$-simulation - and very poorly by the Heston model.
These results indicate that by studying burst detection in the $\mathcal{BSS}$ framework, we will achieve a very good indication of whether the theoretical issue with from Section 12.4 is of practical importance on SPY as the intraday behavior of the estimator in the data and $\mathcal{BSS}$-model are closely aligned.

17.1.2 Burst analysis

For the validity of the results from the burst analysis, it is important that the estimator under the base model was normally distributed.
The QQ-plot of the T-estimator calculated on the simulated paths from the $\mathcal{BSS}$-model can be seen in Figure 72. We see that despite the introduction of intraday seasonality of volatility, persistence and roughness, the average distribution of the estimator is very well-behaved. The tails seem to be closer aligned with a normal distribution than in the Heston model, where the tails were slightly thinner. This suggests that the quantile that we use to as a threshold for the burst detection might still be valid.
To examine the behavior of the T-estimator in the burst framework with the $\mathcal{BSS}$-model, we repeat the simulation study from Table 6 but where the underlying process is exchanged from Heston to $\mathcal{BSS}$. This leads to the results in Table 14.
The most crucial outcome of Table 14 is that when the drift estimator has a bandwidth of 5 minutes and the bandwidth ratio is 12, the false detection rate in the pure $\mathcal{BSS}$ model is 7.3%. This is higher than in

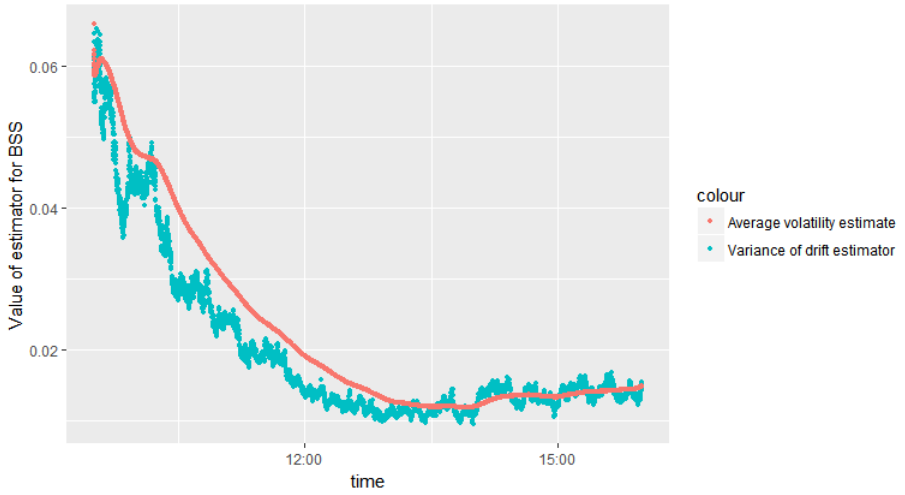


Figure 70: Values $\mathrm{Var}(\sqrt{h_n}\mu_t^n$ and $E[\Sigma_t^n]$ across all days in the $\mathcal{BSS}$-simulation.

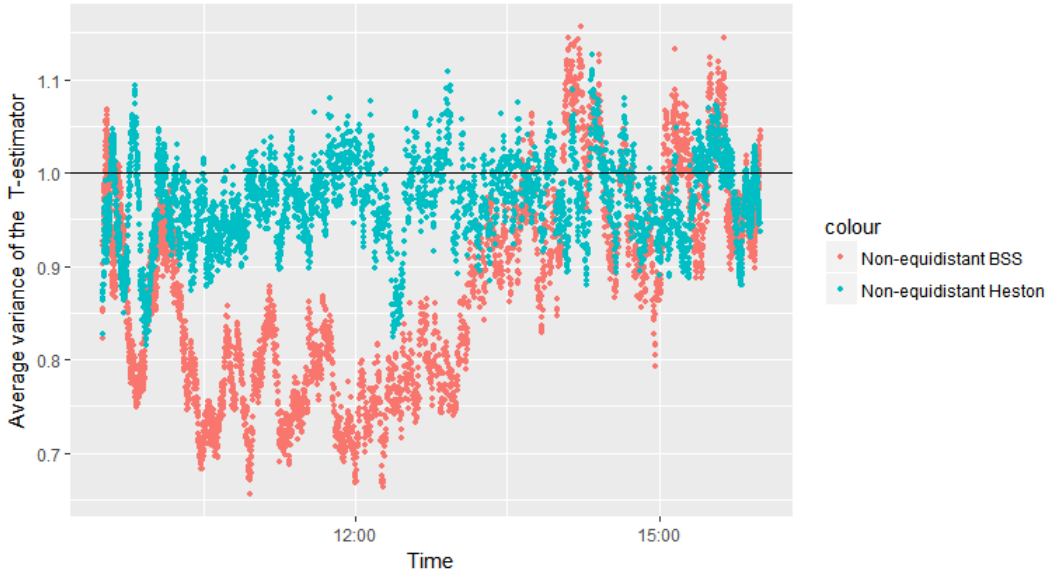


Figure 71: Intraday pattern of variance of T-statistic in simulated data.

the Heston model, and slightly higher than the theoretical 5% for the standard normal distribution. This suggests that we, in the $\mathcal{BSS}$-framework, would expect false detection approximately every 14^{th} day. We found bursts every 3^{rd} day on SPY, which means, that the potential issue with different bandwidth seems to be of little concern as it explains less than 25% of the frequency of bursts we see in the data.
It suggests that the bursts detected with bandwidths $(h_\mu, h_\sigma) = (5\text{min}, 60\text{min})$ are most likely bursts and not a bi-product of the intraday seasonality.
Additionally, it shows that our $\mathcal{BSS}$-framework does not induce bursts in volatility endogenously.

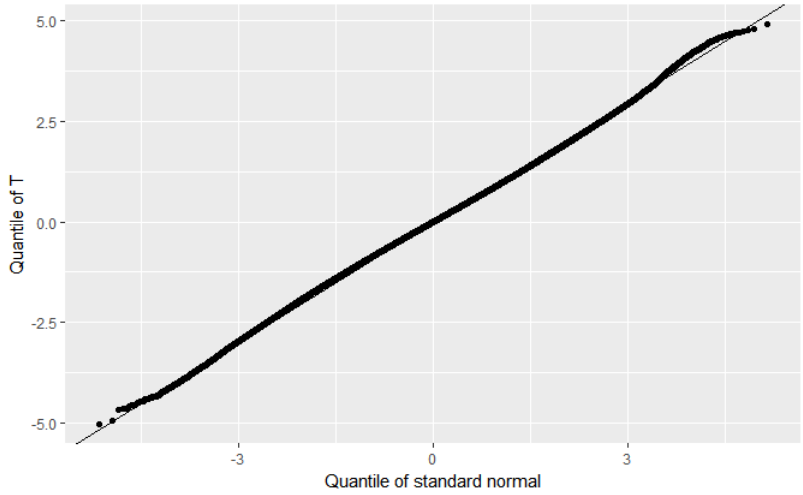


Figure 72: QQ-plot for the T-estimations under the $\mathcal{BSS}$-framework with bandwidth of 5 minutes and bandwidth ratio of 12.

Bandwidths	Beta	$\mathcal{BSS}$	Drift burst a = 0.55	Drift burst a = 0.65	Drift burst a = 0.75
$\mu = 2$min, $\sigma = 30$ min	0	5.3	96.0	99.9	100.0
	0.1	32.3	68.0	96.5	99.4
	0.2	53.7	68.2	85.6	95.2
	0.3	62.9	72.2	82.9	89.7
	0.4	67.3	74.1	83.8	89.1
$\mu = 5$min, $\sigma = 60$ min	0	7.3	96.8	99.9	100.0
	0.1	31.5	75.1	97.3	99.6
	0.2	50.3	70.7	88.9	96.3
	0.3	61.9	72.0	84.2	92.1
	0.4	64.4	74.5	84.9	91.0
$\mu = 10$min, $\sigma = 100$ min	0	5.7	93.8	99.9	100.0
	0.1	23.3	70.5	96.1	99.6
	0.2	41.8	64.5	86.4	94.4
	0.3	51.6	65.5	80.4	89.5
	0.4	56.5	68.2	81.3	87.9

Table 14: Performance of T-estimator of bursts over various bursts under the $\mathcal{BSS}$ framework

18 The robustness of BV^* to bursts[19]

If we accept the hypothesis of existence of bursts in the market, it is very interesting to investigate, how well the BV^*-estimator is at estimating the integrated variance during bursts to see whether values of integrated variance are trustworthy. We will therefore examine the performance of the estimator during bursts. When fitting a volatility model based on a lengthy time horizon, the probability of including periods of bursts increases. Therefore it makes sense to examine the impact of bursts on our volatility estimator. The study presented here will follow along the lines of the simulation study found in the article Fact or Friction by Christensen et al. (2012).

We simulate 10 000 price processes X that follow the dynamics of the Heston-model with annualized pa-

[19]Responsible: Frederik, Mathias

rameters $(\theta, \kappa, \xi, c, \rho, T) = (0.0457, 5.07, 0.48, 1.11, -0.767, 6.5 \text{ hours})$. This is similar to previous simulation studies. Each price process is simulated with 100 000 time steps. The large number of time steps is crucial to the quality of this study. This is also the reason that we base the simulation on Heston and not the $\mathcal{BSS}$-framework. We add microstructure noise similar to that used in most of the burst chapters $\omega^2 = 2.64 \cdot 10^{-10}$. For the pre-average horizon we use $\theta = 1$ unless stated otherwise.

$$Y_t = X_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \omega^2)$$

On top of the Heston-dynamics, we add a jump, volatility burst, two different drift bursts and mix of these. We use similar burst settings to those in section 5.4. The volatility burst is given by $(\beta, c_2) = (0.45, 0.0005)$ and the drift burst are given by $(\tilde{\alpha}, \tilde{c_1}, \bar{\alpha}, \bar{c_1}) = (0.55, 0.2992, 0.8, 0.0088)$. These burst settings produce two burst of identical size and can be seen in figure 8.

The performance of the estimators are compared to the integrated variance (IV) by normalizing the estimates by IV such that:

$$\text{Performance}_{BV^*} = \frac{BV^*}{IV}$$

The integrated variance (IV) is approximated by the sum of the volatilities:

$$IV = \int_0^T \sigma_t^2 dt \approx \sum_{i=1}^{N} \sigma_t^2 \Delta_t$$

The σ_t used for IV are extracted from the simulation scheme and correspond to the real volatilities. Due to the sum approximation, the number of time steps used in the simulation is very important to avoid any bias.

Table 15: Estimator results over various bursts

			Drift burst		
Value	**Beta**	**Heston**	a = 0.55	a = 0.65	a = 0.75
$IV \cdot 10^5$	0	3.369	3.369	3.369	3.369
	0.1	3.374	3.374	3.374	3.374
	0.2	3.390	3.390	3.390	3.390
	0.3	3.415	3.415	3.415	3.415
	0.4	3.439	3.439	3.439	3.439
$BV^* \cdot 10^5$	0	3.367	3.671	4.572	5.736
	0.1	3.412	3.704	4.592	5.753
	0.2	3.463	3.742	4.618	5.775
	0.3	3.514	3.779	4.641	5.793
	0.4	3.537	3.792	4.644	5.795
$\frac{BV^*}{IV}$	0	1.000	1.175	1.686	2.340
	0.1	1.014	1.181	1.682	2.328
	0.2	1.028	1.182	1.663	2.291
	0.3	1.036	1.177	1.634	2.238
	0.4	1.033	1.165	1.604	2.191

Table 15 shows both the value of integrated variance (IV), the estimator of integrated variance (BV^*) and the ratio between the two. The performance of the BV^* estimator appears robust to volatility bursts but not to drift bursts. The key property of a drift burst is that the drift prevails over the volatility, which the BV^* estimator is not suited to handle. Therefore it is not too surprising that it cannot estimate well during drift bursts.

This means that any outcome of our volatility prediction models will not produce reliable results and conclusions based on the volatility estimation are not necessarily valid when applied to periods that include drift bursts. The model that is fit on periods containing drift bursts will be strongly inflated depending on the amount and size of the drift bursts.

19 Discussions and future work

There are several analyses, where extensions to the study would be worth investigating. Some of these have been omitted in this thesis simply because they were beyond the scope of this thesis, while others because solutions to the problem proved difficult to create or discover. This section will present some of the studies that have been *left out*.

19.1 Reversal of bursts

A major topic in Christensen et al. (2016) is a study and discussion of the reversing nature of drift bursts. In this section we will first summarize the approach in Christensen et al. (2016) and afterwards how those result can match with the results found in this thesis.

In this thesis we have kept the focus of the study on how many days with drift bursts we see and did not go into investigating the total number of drift bursts. There were two specific reasons for this.
The first is that if we began investigating the number of drift bursts, we would have to find some method to quantify the length of a burst - such that we can distinguish one lengthy burst from two smaller consecutive bursts. As the estimator is extremely correlated close in time, it is natural that the estimator is above the threshold several observations after the peak of the burst.
In Christensen et al. (2016) each day is divided into 5-minute intervals, and whenever the T-estimator is above the threshold it is considered a burst, but only one is allowed in every 5-minute interval. This however has the disadvantage that a burst that occurs on the edge between two intervals will count double. If the burst is long enough, it might even be included in several intervals.
The other reason is that it is more difficult to assess the expected number of false positives. We know we expect false bursts on average every 20^{th} day with our threshold but it does not tell us anything about how many false positives to expect per day.
Instead of conducting this study ourselves, we will merely summarize the study in Christensen et al. (2016) and comment on the findings.

19.1.1 The study

The study is carried out by examining the log-return before and after a 5 minute bucket (t_j) with the occurrence of a drift burst at time:

$$R^-_{t_j} = X_{t_j} - X_{t_j - 5\text{min}} \qquad R^+_{t_j} = X_{t_j + 5\text{min}} - X_{t_j}$$

and set up the model

$$R^+_{t_j} = \alpha + \beta R^-_{t_j} + \epsilon_j$$

to examine the relationship between R^- and R^+. The study concludes that β is significantly negative, and we thus often have returns of opposite signs before and after a burst.

The biggest concern with this study is that t_j is defined as the point in time where the T-estimator *peaks*. This implies information of the size of the T-estimator succeeding the burst, thus also having information

about the movements of the price after the timepoint t_j, as we know the process does not result in a higher T-estimator afterwards. Knowing that the T-estimator decreases in absolute value afterwards could very well have an influence of the expected return of Y due to the close connection between the price process and estimator.

To clarify the point above, we show a simulation of a Heston price process with burst parameters ($\alpha = 0.65,\ \beta = 0.1,\ c_1 = -0.157,\ c_2 = 0.001$) similar to the bursts from the bandwidth study. The burst is set to last 19.5 minutes and ends at the middle of the trading day. The price process is seen in Figure 73. The burst ends at the vertical black line, and we see, that we have $R^- < 0$ due to the negative drift burst, but also that the 5 minutes following the burst has a negative return, i.e. $R^+ < 0$ by mere coincidence from the volatility of the underlying process. If we denote the time point as the peak of the T-estimator however, it is where the blue vertical line is. This would give us $R^- < 0$ and $R^+ > 0$ as the process has a positive return in 5 minutes following this time point. This biases the result towards a negative correlation between R^- and R^+.

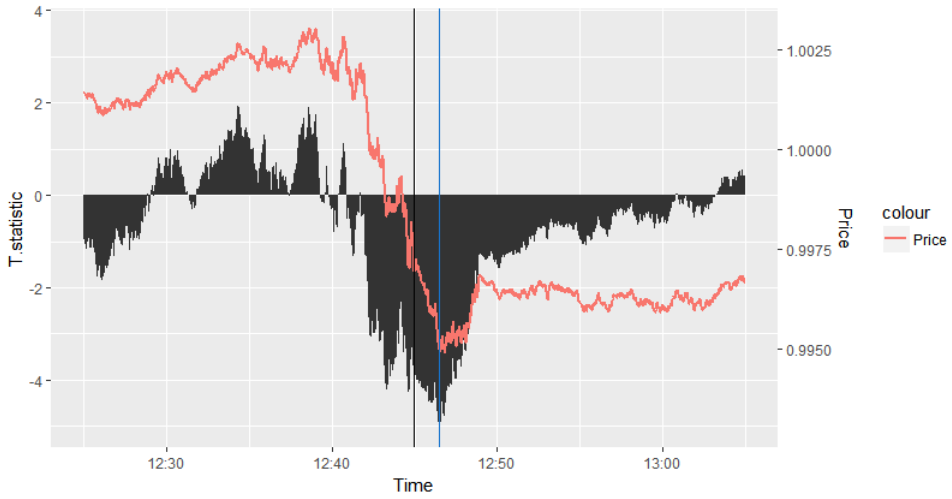


Figure 73: A 20-minute window of the price process of 23,400 time steps with an added burst. The black vertical line represents the end of the simulated burst, and the blue line represents the maximum T^*-statistic.

19.1.2 Reversal in volatility burst

In this section we will present a possible rationale behind why bursts could tend to reverse even if they are purely given by volatility bursts.
We assume that the market agents are risk-averse that select a portfolio such that for a fixed return, they wish to minimize the variance of the portfolio. An example of the behaviour of such a risk-averse agent can be described as:

$$\min_w \ \frac{1}{2} w^T \Sigma w \ \text{ subject to } \ w^T \mu^e = \mu_P^e$$

Where w represents the weight of each asset, Σ the covariance matrix of the assets, μ^e the excess return of the assets and μ_P^e represents the fixed excess return that the agent wish to obtain. A result of this behaviour is that agents wish to maximize the Sharpe-ratio of their portfolio:

$$\text{Sharpe-ratio}_P = \frac{\mu_P^e}{\sigma_P}$$

To keep it simple, we assume that all agents have the same degree of risk-aversion and that there only exists two assets, named asset 1 and asset 2. Hence the Sharpe-ratio of a market agent is given by:

$$\text{Sharpe-ratio}_P = w_1 \cdot \text{Sharpe-ratio}_1 + w_2 \cdot \text{Sharpe-ratio}_2$$

We assume that the two assets start with the same Sharpe-ratio such that the any combination of the two assets still yields the highest Sharpe-ratio.
If the volatility of asset 1 suddenly increases (e.g. up to an upcoming event such as the publication of an annual report), the Sharpe-ratio of the current portfolio is decreased. The risk-averse agents will reallocate their portfolio by selling asset 1 and buying asset 2. However, as agents sell asset 1, the price drops which increases the expected return of that asset. This balances the Sharpe-ratio such that the price drops until the Sharpe-ratio of asset 1 is again equivalent to that of asset 2.
Eventually, after the event occurs, the volatility decreases again, and the opposite happens. The Sharpe-ratio is increased, and demand for asset 1 increases, which causes an increase in its price and a decrease in expected return until the two Sharpe-ratios are again equivalent.
This explains, why a volatility burst, that resulted in a negative price movement, and thus a T-estimator below the negative threshold tends have a positive expected following the burst.

In Table 5 in Christensen et al. (2016), we see that the tendency for the price change to switch sign after a burst is more pronounced when the burst had a negative price movement than a positive. This supports the rationale above, as the rationale only supports, bursts with negative expected return prior to the peak and positive expected return post-burst.

19.1.3 Conclusion of reversal study

The inability to distinguish volatility and drift bursts with ours and Christensen et al. (2016) choices of bandwidth means that we cannot test the reversing properties of drift bursts and volatility bursts separately. This is rather unfortunate as this could shed some light on why negative bursts are much more likely to reverse.
It would be very interesting to replicate a similar study on the Bitcoin exchanges, and compare the behaviour of bursts across the platforms. This is especially interesting because of the fact that the ability to detect bursts with a bandwidth ratio of 1 in the Bitcoin data makes it easier to detect bursts that we expect to include a drift burst.
Before doing this however, one would need to solve the issue with separating bursts and find a way to choose the appropriate time span before and after the burst to measure R^- and R^+. This is especially important for the Bitcoin market as we saw several bursts within a short time period.

19.2 Analysis of exchange specific effects

The high-frequency behavior of the Bitcoin price is very much affected by the fee structure on the specific platform. The fees on the three platforms studied is based on a maker/taker structure. This means that placing a limit order has different cost than taking a limit order.
The limit order is usually cheaper on the platforms, because placing a limit order provides liquidity in the market, and this can be an issue on some platforms due to few traders and low volume.
Table 16 shows the fees and 24hour volume on BTC/USD[20], and how this affects the price can be seen in Figure 74. It is evident that the quotes on Kraken are less steady than the other platforms. This is mostly due to the smaller trading volume and less liquidity. This results in a gap bid and ask prices which is varies a lot. This is due to two reasons, the first being, that a single transaction can change the mid-price because of low volume of the limit orders, the best offer might change. The second is, that bots continually place and cancel quotes in the gap between the best bid and best ask price and thereby changing the mid-quote.

For the other platforms, the best bid price and best ask price are often without gap in between. Because the limit order has a fee of 0.1% and -0.025%, the limit orders are much cheaper than market orders, and this results in massive walls of orders being build up on both buy and sell side. Because of the trading fees, it is

[20]The fees on Bitfinex and Kraken depend on he traded volume. Here the fees for the lowest traded volume are taken. The volume is taken from https://coinmarketcap.com/exchanges the 2^{nd} of August.

not profitable to trade, as long as the underlying process price is less than the trading fee from the price on the exchange. This leads to the price being quite steady, when the underlying process has slow fluctuations. This results in negative correlations between the underlying process and price process, as also found very evident for mid-quotes in Hansen & Lunde (2006), and has leads to the price processes on BitMEX and Bitfinex looking more as if they are stable with successive jumps compared to the price process on Kraken.

Table 16: Fee structure and 24hour volume on the three Bitcoin exchanges.

	Taker fee	Maker fee	24hour volume
Bitfinex	0.2%	0.1%	163,519,189 $
BitMEX	0.075%	-0.025%	3,835,526,019 $
Kraken	0.26%	0.16%	30,047,961 $

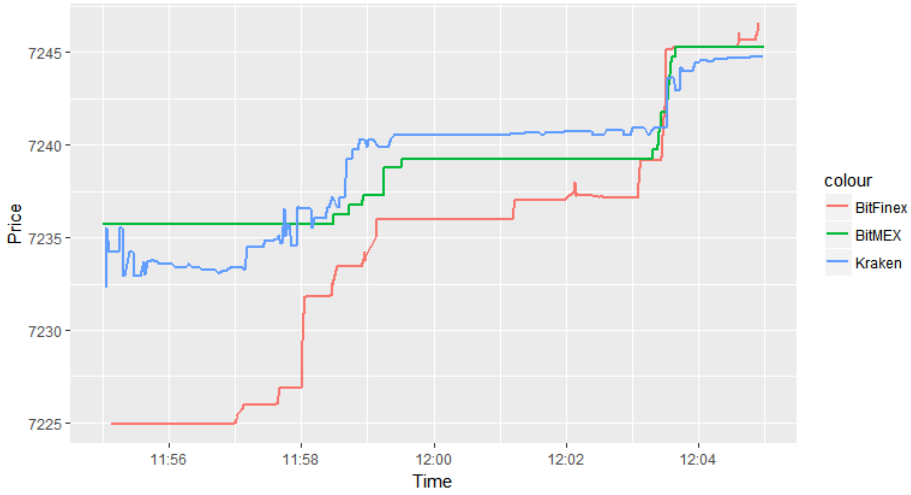


Figure 74: Bitcoin prices on the three different exchanges over a period of 20 minutes.

These different behaviors on very short time scale definitely effects the results, and can be one of the primary explanations of the significant differences in the distribution of the T-estimator between Kraken on one side and Bitfinex and BitMEX on the other.

Investigating how these effect the T-estimator is important, but it is not obvious how this should be done. It is difficult to construct a mathematical process with behavior similar to e.g. BitMEX on short scale that has these *sticky prices*, especially because the underlying process is unknown, so the exact behavior of this kind of noise is very difficult to determine. Based on the reasoning above the true underlying process is most likely a mixture of Kraken and BitMEX/Bitfinex, but it is difficult to tell for sure. It is obvious however that these frictions has an effect on the detection of bursts. Since we detect significantly more bursts on BitMEX and Bitfinex than on Kraken, it is most likely as a result of these exchange specific behaviors.
It would be very interesting to develop a model with *sticky prices* as seen on Bitfinex and BitMEX and study the behavior of the T-estimator under this model.

The "true" amount of bursts is possibly a mixture of the frequency observed on Kraken and the frequency on the others as the volatility of the underlying process X based on the reasoning above is overestimated on Kraken and underestimated on the others.

19.3 Parameter estimation on Bitcoin prices

The simulation studies that laid the foundation for the choice of bandwidth and bandwidth ratio were all based on the Heston dynamics with estimated parameters for the S&P 500 data. Knowing what we do now, we believe that the underlying simulations do not properly reflect the behaviour seen in the Bitcoin market. Hence, it is tailor-fitting the choice of bandwidth and ratio to the different exchanges instead of using those for the SPY market.
Ultimately, we believe that the choice of a bandwidth ratio of 12 might not be suitable for the Bitcoin market, and the conclusions of the bursts found in the Bitcoin should keep this in mind.

19.4 Compensation of seasonality

It was discovered in Section 12.4 that the T-estimator exhibits an intraday pattern in data. Though we with $\mathcal{BSS}$-simulations showed that this intraday pattern can be present without invalidating the results, an even better solution to the problem would be to take the intraday seasonality into account when estimating T. In fact, by re-scaling the estimators to remove the bias coming from intraday seasonality, we may minimize the impact that the seasonality has on the T-estimator with a high bandwidth. This is very similar to the correction of the bias in the beginning of the day.
One way to do this could be to re-scale every observation in the delta-process ΔY_i with the average standard deviation at that time of the day (given by the calibrated Fourier-transform), i.e. transitioning to working with the deseasonalized volatility.
Let $\log(\bar{\sigma}_t)$ denote the estimated seasonality of the log-volatility at time t. The seasonality compensating estimator is then given as:

$$\hat{\mu}_t = \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left(\Delta Y_i \frac{1}{\bar{\sigma}_i}\right)$$

$$\hat{\gamma}(l)_t = \frac{1}{h_n} \sum_{i=|l|+1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) \left(\Delta Y_i \frac{1}{\bar{\sigma}_i}\right) K\left(\frac{t_{i-|l|-1} - t}{h_n}\right) \left(\Delta Y_{i-|l|} \frac{1}{\bar{\sigma}_{i-|l|}}\right)$$

We apply these estimators on the $\mathcal{BSS}$-simulations from Figure 71. The results of the T-estimations on two identical $\mathcal{BSS}$-simulations of 20,000 time steps and 1000 paths can be seen in Figure 75. Here we see that the above estimators greatly reduce the intraday seasonality. Unfortunately, it is not straight-forward to apply

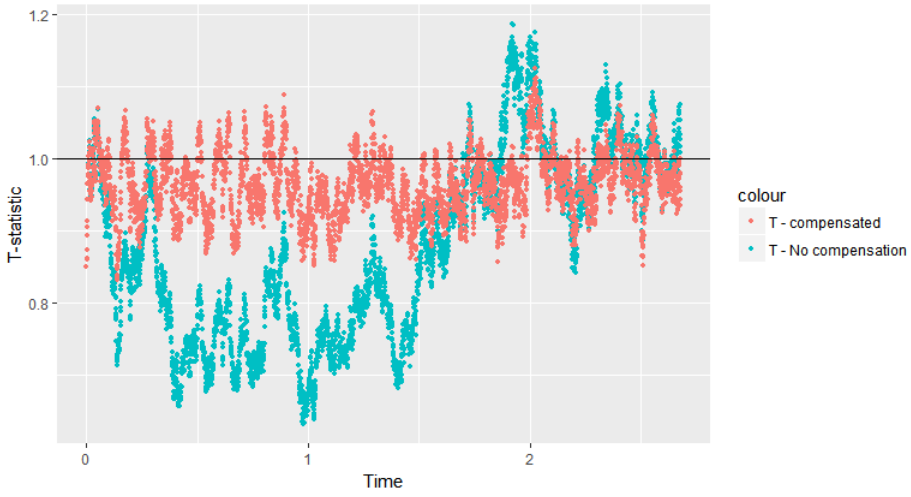


Figure 75: Intraday pattern of variance of T-statistic in simulated data. The estimation has been carried out both with and without the compensation of seasonality.

this method on real data. The benefit of the above simulation is that the intraday seasonality is equivalent across all days. This is NOT the case in real data. Hence, a naïve attempt like this on real data may distort the results far too much. Instead we leave this method as an inspiration for future work.

19.5 Bandwidth selection algorithm

In this thesis the bandwidth was chosen based on a simulation study and was kept constant. It would however be very valuable if the convergence results shown in this thesis were generalized such that the rate of convergence was also known. By showing a convergence to a mixed normal distribution with variance going towards zero as is done e.g. in Barndorff-Nielsen et al. (2011), we would be able to construct bandwidth selection algorithms (along with lag selection algorithms), which minimizes the expected error.
This would also result in a varying bandwidths, which adopts to the frequency of the observations, as a longer bandwidth is optimal at times with a low frequency and vice versa. This could possibly enhance the burst detection significantly.

19.6 Creating a model

After illustrating specific properties of the market that many of the current models fall short of describing, a natural next step is to create a new model containing bursts. These bursts could e.g. be added to an existing model as an extra component in the same way we added jumps to the Heston model.
Christensen et al. (2016) suggest a model, which can create bursts using feedback effects, i.e. where the change of the price can lead to further price change in a spiral and thus re-enforcing the directional price move.
While both are extremely interesting choices, accurate detection of bursts is needed, and potentially even estimation of size and duration of bursts could be required for calibration.

20 Conclusion

In this thesis we further develop the mathematical foundation behind the ideas of short-lived explosive trends in the financial markets, first presented formally and parametric in Christensen et al. (2016). We show new convergence results of the estimators used in Christensen et al. (2016), when the financial price exhibits a short-lived explosion in either volatility or drift known as a *burst*. The theory is expanded to incorporate microstructure noise, and we show that if we wish to be able to improve the detection of the *bursts* and be able to detect the short-lived explosive periods as they occur, we need to accept a small bias from in the noise, which is found to be insignificant in practice. In addition to the theory, we make new finite sample corrections, which allows us to investigate the existence of bursts on almost the full spectrum of the day.
To create thorough simulation studies, we construct fast computational algorithms, utilizing specific properties of the estimator. These allow us to tune parameters in the estimator, and we especially find that the parameter choices which determine the weights of the past observations in the estimator are vital for the detection of bursts.

The *burst* detection is applied in practice on S&P-500, where we find short-lived explosive periods every 3^{rd} day. However the study also shows that the market behavior is very different from the simulation studies under the Heston model with jumps. In practice, intraday behaviour of the financial markets has a profound effect on the quality of detection. This means that the conclusions from previous simulation studies may not hold in practice. Hence we investigate the behavior of the volatility in the market, and develop an alternative volatility model to that of the Heston.

We utilize a flexible Fourier form to fit the intraday seasonality of volatility and estimate roughness and persistence on the deseasonalized log volatility process of both the S&P 500 and the Bitcoin market. We find that both the deseasonalized log volatilites are stationary and exhibit a higher degree of roughness than usually produced by a Brownian motion, estimates suggests slightly more roughness in the Bitcoin market compared to S&P 500. Estimation of persistence is attempted, but we highlight difficulties of estimating the values accurately, however the possibility of high persistence cannot be ruled out. Given these characteristics, we turn to two variations of the $\mathcal{BSS}$-framework, namely the Gamma-$\mathcal{BSS}$ and the Power-$\mathcal{BSS}$. We fit these models on our data and evaluate their performance in a volatility prediction scheme. From this we are able to conclude that correlation structures in data are best matched by highly persistent models, potentially featuring long term memory. Further, we find indications that fitting these models with taking the possibility of bursts into account, could hurt their predictive, or more generally their descriptive, power of the volatility processes observed in the markets.

Utilizing simulation study featuring the $\mathcal{BSS}$-models, that imitates the intraday market behavior far better than the Heston model, we conclude that our original findings in both the simulation studies and in practice are still valid. In addition, we conclude that our $\mathcal{BSS}$-model is not capable of explaining the short-lived intraday explosive periods in the market.

This significantly strengthens the validity of the data study, and the findings of a short-lived explosive periods every 3^{rd} day on S&P-500. This is much more frequent than Christensen et al. (2016) which is most likely due to the finite sample corrections and extra tuning of parameters in the estimator used for detection. We find it questionable however, whether drift bursts are present or if all bursts are in fact explained by an explosion of the volatility.
We also apply the theory on Bitcoin and give a justification for both the existence of drift bursts and volatility bursts on this market with much higher frequency than on S&P 500. These conclusions are shown to be somewhat dubious as we believe that exchange specific frictions on the Bitcoin platforms complicates the investigation in this market.

21 APPENDIX 1 - Proofs[21]

21.1 Notations

In the proofs below we use the notation $a \sim b$ to denote that the ratio a/b converges to a constant different from zero, and if one or to of them are random variables, we use $a \sim_p b$ to denote that the ratio a/b converges in probability to a constant different from zero.

The notation in the proofs will follow the notation in Christensen et al. (2016) close. In the burst proofs, we will therefore use the notation

$$\Delta X_{i,n}^b = \Delta X_{i,n} + D_{i,n} + S_{i,n}$$

where

$$\Delta X_{i,n} = \int_{t_{i-1}}^{t_i} \mu_s ds + \int_{t_{i-1}}^{t_i} \sigma_s dW_s$$

is the process without bursts, and

$$D_{i,n} = \int_{t_{i-1}}^{t_i} \frac{1}{(1-s)^\alpha} ds, \quad \text{and} \quad S_{i,n} = \int_{t_{i-1}}^{t_i} \frac{1}{(1-s)^\beta} dW_s.$$

21.2 Results used

We will repeatedly make use of several results from Appendix A in Christensen et al. (2016). We refer to that paper for the proofs and just state the results below

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right) D_{i,n}^2 = O_p(\Delta_n h_n^{-2\alpha}) \tag{68}$$

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right) E\left(S_{i,n}^2\right) = O_p(\Delta_n h_n^{-2\beta}) \tag{69}$$

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) E[S_{i,n}^4] = O_p(h_n^{-4\beta}) \tag{70}$$

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \mu_s ds - \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) \mu_s ds = O_p\left(\frac{1}{nh_n}\right) \tag{71}$$

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right) \int_{t_{i-1}}^{t_i} \sigma_s^2 ds - \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) \sigma_s^2 ds = O_p\left(\frac{1}{nh_n}\right) \tag{72}$$

$$\frac{1}{dt \cdot h_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{h_n}\right) \left(\int_{t_{i-1}}^{t_i} \mu_s ds\right)^2 - \int_0^T \frac{1}{h_n} K\left(\frac{s - t}{h_n}\right) \mu_s^2 ds = O_p\left(\frac{1}{nh_n}\right) \tag{73}$$

[21] Responsible: Sebastian

We will also make use of the lemma below:

Lemma 2. *If X_i is a stationary process with*

$$\sum_{j=1}^{\infty} |cov(X_{i,n}, X_{i+j,n})| < C\, Var(X_{i,n}) \quad \forall j, n$$

then

$$\sum_{i=1}^{n} Var(X_i) \to 0 \quad for \;\; n \to \infty$$

implies that

$$Var\left(\sum_{i=1}^{n} X_i\right) \to 0 \quad for \;\; n \to \infty.$$

Proof. This follows by the fact that

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(X_i, X_j) \leq \sum_{i=1}^{n} C\mathrm{Var}(X_i) = C\sum_{i=1}^{n} \mathrm{Var}(X_i)$$

□

21.3 Proof of Proposition 1

We want to proof that when

$$T \sim \frac{1}{\sqrt{k^2+\sigma^2}} N(c \cdot k, \sigma^2)$$

then for all $k > 0$ we have

$$P(T > q) \leq P(X > \sqrt{q^2 - c^2})$$

where $X \sim N(0, 1)$.

We know that

$$T \sim N\left(\frac{c \cdot k}{\sqrt{k^2+\sigma^2}}, \frac{\sigma^2}{k^2+\sigma^2}\right) = \frac{c \cdot k}{\sqrt{k^2+\sigma^2}} + \frac{\sigma}{\sqrt{k^2+\sigma^2}} \cdot N(0, 1).$$

This means that if we let x and q have the relationship

$$\frac{c \cdot k}{\sqrt{k^2+\sigma^2}} + x \cdot \frac{\sigma}{\sqrt{k^2+\sigma^2}} = q$$

then

$$P(T > q) = P(X > x)$$

where $X \sim N(0, 1)$.

Given a quantile q we want to find the k that maximizes the probability $P(X > x)$. This is equivalent to minimizing x wrt. k.

We rewrite

$$x = \frac{1}{\sigma}\left(q\sqrt{k^2+\sigma^2} - ck\right).$$

Differentiating wrt. k and setting equal to zero gives

$$q\frac{1}{2\sqrt{k^2+\sigma^2}}2k - c = 0 \Leftrightarrow$$

$$qk = c\sqrt{k^2+\sigma^2} \Leftrightarrow$$

$$k = \sqrt{\frac{c^2\sigma^2}{q^2-c^2}}.$$

We note, that this is indeed a minimum as the derivative is decreasing for lower values and increasing for higher values. The expression for k is inserted in the expression for x. First it is noted that

$$\sqrt{k^2+\sigma^2} = \sqrt{\frac{q^2\sigma^2}{q^2-c^2}}$$

so

$$x = q\sqrt{\frac{q^2}{q^2-c^2}} - c\sqrt{\frac{c^2}{q^2-c^2}} =$$

$$\frac{q^2-c^2}{\sqrt{q^2-c^2}} = \sqrt{q^2-c^2}.$$

We thus have

$$P(T > q) \leq P(X > \sqrt{q^2-c^2})$$

and it holds with equality when $k = \sqrt{\frac{c^2\sigma^2}{q^2-c^2}}$.

21.4 Proof of Theorem 5

We investigate the estimator

$$T_t^n = \sqrt{\frac{h_n}{K_2}}\frac{\hat{\mu}_t^n}{\sqrt{(\hat{\sigma}_t^n)^2}} = \sqrt{\frac{h_n}{K_2}}\frac{\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_{i,n} + D_{i,n} + S_{i,n})}{\left(\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_{i,n} + D_{i,n} + S_{i,n})^2\right)^{1/2}} \tag{74}$$

and look at the convergence one term at a time.

The numerator
In the numerator, we know from the drift estimator in absence of noise that

$$\frac{1}{\sqrt{h_n}}\sum_{i=1}^n K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta X_{i,n} \xrightarrow{d} N(0, K_2\sigma_{t-}^2)$$

as $h_n \to 0$ and $nh_n \to \infty$.

Regarding the drift burst, we have

$$\begin{aligned}\frac{1}{h_n}\sum_{i=1}^n K\left(\frac{t_{i-1}-t_{db}}{h_n}\right)D_{i,n} &= \frac{1}{h_n}\int_0^{t_{db}} K\left(\frac{s-t_{db}}{h_n}\right)\frac{1}{(t_{db}-s)^\alpha}ds + O\left(\frac{1}{nh_n}\right) \\ &= \int_{-\frac{1}{h_n}}^0 K(x)|x|^{-\alpha}h_n^{-\alpha}dx + O\left(\frac{1}{nh_n}\right) \sim h_n^{-\alpha}\Gamma(1-\alpha)\end{aligned}$$

where the first equation follows from (71), the second from integration substituting $x = \frac{s-t_{db}}{h_n}$ and the last from the fact that $\int_{-\infty}^{0} K(x)|x|^{\alpha}dx = \Gamma(1-\alpha)$ for the exponential kernel. From this follows that

$$h_n^{\alpha-1/2}\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right) D_{i,n} \to \Gamma(1-\alpha)$$

Regarding the volatility burst we first define

$$G_{i,n} = \frac{1}{\sqrt{h_n}} K\left(\frac{t_{i-1}-t}{h_n}\right) S_{i,n}.$$

We have

$$\sum_{i=1}^{n} E[G_{i,n}] = \frac{1}{\sqrt{h_n}}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right) E[S_{i,n}] = 0$$

and

$$\begin{aligned}
\sum_{i=1}^{n} E[G_{i,n}^2] &= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)^2 E[S_{i,n}^2] \\
&= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)^2 E\left[\left(\int_{t_{i-1}}^{t_i}\frac{1}{(1-s)^{\beta}}dW_s\right)^2\right] \\
&= \frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)^2 \int_{t_{i-1}}^{t_i}\frac{1}{(1-s)^{2\beta}}ds \\
&= \frac{1}{h_n}\int_0^t K\left(\frac{s-t}{h_n}\right)^2 \frac{1}{(t-s)^{2\beta}}ds + O\left(\frac{1}{nh_n}\right) \\
&= \int_{-\frac{1}{h_n}}^{0} K(x)^2|x|^{-2\beta}h_n^{-2\beta}dx + O\left(\frac{1}{nh_n}\right) \sim h_n^{-2\beta}2^{2\beta-1}\Gamma(1-2\beta)
\end{aligned}$$

where the fourth equation follows from (72) and the "$\sim$" result follows from the fact that $\int_{-\infty}^{0} e^{2x}|x|^{-2\beta}dx = 2^{2\beta-1}\Gamma(1-2\beta)$. Note that the integral finite as $\beta < 1/2$.

Collecting the three terms in the numerator, we have from independence (with slight abuse of notation) that

$$h_n^{\beta}\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_{i,n} + D_{i,n} + S_{i,n}) \xrightarrow{d} \tag{75}$$

$$N\left(h_n^{-\alpha+1/2+\beta}\Gamma(1-\alpha), 2^{2\beta-1}\Gamma(1-2\beta)\right) \tag{76}$$

so when $\alpha > \beta + 1/2$ it diverges and when $\alpha < \beta + 1/2$ it converges to a normal distribution.

The denominator

We first wish to show that if $\alpha < \beta + 1/2$ have

$$h_n^{2\beta}\frac{1}{h_n}\sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_{i,n} + D_{i,n} + S_{i,n})^2 \xrightarrow{p} \Gamma(1-2\beta) \tag{77}$$

because then from the fact that

$$T_t^n = \frac{1}{\sqrt{K_2}} \frac{h_n^\beta \sqrt{h_n} \hat{\mu}_t^n}{\sqrt{h_n^{2\beta} (\hat{\sigma}_t^n)^2}}$$

we can by Slutsky's get that

$$T_{\tau_{db}} \xrightarrow{d} N(0, 2^{2\beta}) \qquad \text{for } \beta + 1/2 > \alpha$$

which we want.

We naturally have

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) (\Delta X_{i,n})^2 \xrightarrow{p} \sigma_{t-}$$

since this is the volatility estimator. Multiplying the term with $h_n^{2\beta}$, it converges to zero.

From (68) we have

$$\frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) D_{i,n}^2 = O_p\left(\frac{1}{n h_n^{2\alpha}}\right) = o_p(h_n^{1-2\alpha})$$

and from this follows

$$h_n^{2\beta} \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) D_{i,n}^2 = o_p(h_n^{1-2\alpha+2\beta})$$

and since $1 - 2\alpha + 2\beta = \beta + 1/2 - \alpha > 0$ it converges to zero.

The last squared term has the expectation

$$E\left[h_n^{2\beta} \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) (S_{i,n})^2\right] = h_n^{2\beta} \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) E[S_{i,n}^2]$$

$$\to \Gamma(1 - 2\beta)$$

where the convergence result follows in the same way as in the calculations for the numerator (just without having the kernel squared this time).

The variance of the last squared term is

$$\text{Var}\left(h_n^{2\beta} \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) (S_{i,n})^2\right) = h_n^{2\beta} \frac{1}{h_n} \sum_{i=1}^{n} \text{Var}\left(K\left(\frac{t_{i-1} - t}{h_n}\right) (S_{i,n})^2\right) \leq$$

$$\frac{h_n^{4\beta}}{h_n} \cdot \frac{1}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right)^2 E[S_{i,n}^4] = \frac{h_n^{4\beta}}{h_n} O\left(\frac{1}{n h_n^{4\beta}}\right) = O\left(\frac{1}{n h_n}\right)$$

where the second equation follows from (70). The variance thus converges to zero as $nh_n \to \infty$ so the expression converges in probability to the mean.

For all the product terms coming from expanding the parenthesis on the left-hand side in (77) the expectation converges to zero and variance as well (they all following with similar reasoning as calculations previously done, i.e. mean is zero and variance converging to zero) - so these converge in probability to zero. Collecting the terms and using Slutsky repeatedly we get the convergence in (77) as we wanted. This proofs the

convergence for $\alpha < \beta + 1/2$.

Now we wish to show, that if $\alpha > \beta + 1/2$, then

$$h_n^{2\beta} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) (\Delta X_{i,n} + D_{i,n} + S_{i,n})^2 = o_p(h_n^{1+2\beta-2\alpha})$$

because then it follows, that the mean of the estimator increases.

The only difference between this and before is, that now the term

$$\frac{h_n^{2\beta}}{h_n} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) D_{i,n}^2$$

isn't nessesarily decreasing, however it is still $o_p(h_n^{1+2\beta-2\alpha})$ following the same calculations, and thus

$$h_n^{2\beta} \sum_{i=1}^{n} K\left(\frac{t_{i-1} - t}{h_n}\right) (\Delta X_{i,n} + D_{i,n} + S_{i,n})^2 = o_p(h_n^{1+2\beta-2\alpha}).$$

From that follows, that if $\alpha > \beta + 1/2$ we have

$$T_{\tau_{db}} \xrightarrow{p} \infty \quad \text{for} \quad \beta + 1/2 < \alpha$$

which we wanted.

This gives us the full result

$$\begin{aligned} |T_{\tau_b}| &\xrightarrow{p} \infty && \text{for} \quad \beta + 1/2 < \alpha \\ T_{\tau_b} &\xrightarrow{d} N(0, 2^{2\beta}) && \text{for} \quad \beta + 1/2 > \alpha \end{aligned}$$

as desired.

21.5 Proof of Theorem 7

We want to show that

$$\sqrt{h_n}(\tilde{\mu}_t^n - \mu_{t-}) \xrightarrow{d} N(0, \psi_1^2 K_2 \sigma_t^2)$$

stably in law as $n \to \infty$, $h_n \to 0$ and $\sqrt{n} h_n \to \infty$.
We have

$$\tilde{\mu}_t^n = \frac{1}{k_n h_n} \sum_{i=k_n}^{n+1} K\left(\frac{t_i - t}{h_n}\right) \Delta \bar{Y}_i$$

with

$$\Delta \bar{Y}_i = \sum_{j=1}^{k_n - 1} g\left(\frac{j}{k_n}\right) (Y_{i-j} - Y_{i-j-1}).$$

It can be split into the X-process and noise process and multiplying with $\sqrt{h_n}$ gives

$$\sqrt{h_n} \tilde{\mu}_t^n = \frac{1}{k_n \sqrt{h_n}} \sum_{i=k_n}^{n+1} K\left(\frac{t_i - t}{h_n}\right) (\Delta \bar{X}_i + \Delta \bar{\epsilon}_i).$$

The convergence of the X-process and noise process will be investigated separately.

21.5.1 The X-process

Referring to the graphical representation in Figure 7, we can rewrite

$$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1}\Delta\bar{X}_i = \frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n-1}K\left(\frac{t_{i-j}-t}{h_n}\right)g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})\right)$$

and split it into the first k_n-1 observations, the middle $n-2k_n+2$ and the last k_n-1 observations:

$$= \frac{1}{k_n\sqrt{h_n}}\sum_{i=1}^{k_n-2}\left(\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right)\right)K\left(\frac{t_i-t}{h_n}\right)\Delta X_i \tag{78}$$

$$+ \frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n-1}^{n-(k_n-2)}\left(\sum_{j=1}^{k_n-1}g\left(\frac{k_n-j}{k_n}\right)\right)K\left(\frac{t_i-t}{h_n}\right)\Delta X_i \tag{79}$$

$$+ \frac{1}{k_n\sqrt{h_n}}\sum_{i=n-(k_n-3)}^{n}\left(\sum_{j=1}^{n+1-i}g\left(\frac{j}{k_n}\right)\right)K\left(\frac{t_i-t}{h_n}\right)\Delta X_i. \tag{80}$$

We inspect the terms one at a time.

The first term
For the expression in (78) we have an expectation of zero. We can also create the inequality

$$\frac{1}{k_n}\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right) \leq \frac{1}{k_n}\sum_{j=1}^{k_n-1}g\left(\frac{k_n-j}{k_n}\right) \leq c \quad \forall i \in \{1,2,...,k_n-2\}.$$

The last inequality follows from the fact that $\frac{1}{k_n}\sum_{j=1}^{k_n-1}g\left(\frac{k_n-j}{k_n}\right) \to \int_0^1 g(x)dx$ and when it converges it must be bounded by a constant. With this in mind we have

$$\mathrm{Var}\left(\frac{1}{k_n\sqrt{h_n}}\sum_{i=1}^{k_n-2}\left(\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right)\right)K\left(\frac{t_i-t}{h_n}\right)\Delta X_i\right) =$$

$$\mathrm{Var}\left(\frac{1}{\sqrt{h_n}}\sum_{i=1}^{k_n-2}\left(\frac{1}{k_n}\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right)\right)K\left(\frac{t_i-t}{h_n}\right)\Delta X_i\right) =$$

$$\frac{1}{h_n}\sum_{i=1}^{k_n-2}\left(\frac{1}{k_n}\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right)\right)^2 K\left(\frac{t_i-t}{h_n}\right)^2\mathrm{Var}(\Delta X_i) \leq$$

$$\frac{c^2}{h_n}\sum_{i=1}^{k_n}\mathrm{Var}(\Delta X_i) = \frac{c^2}{h_n}k_n\frac{t}{n} \propto \frac{k_n}{h_n n}$$

where the first inequality follows from the fact that $\mathrm{cov}(\Delta X_i, \Delta X_j) = 0$ for $i \neq j$. The last term naturally converges to zero when $k_n \propto \sqrt{n}$ and $\sqrt{n}h_n \to \infty$.

From this follows that the whole expression in (78) converges in probability to zero.

The third term
For the expression in (80), we also have an expectation of zero. The term is similar to the first (just in the

other end), and by exactly same line of reasoning this also converges in probability to zero.

The second term
Regarding the expression in (79) we have

$$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n-1}^{n-(k_n-2)}\left(\sum_{j=1}^{k_n-1}g\left(\frac{k_n-j}{k_n}\right)\right)K\left(\frac{t_i-t}{h_n}\right)\Delta X_i =$$

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n-1}g\left(\frac{k_n-j}{k_n}\right)\right)\frac{1}{\sqrt{h_n}}\sum_{i=k_n-1}^{n-(k_n-2)}K\left(\frac{t_i-t}{h_n}\right)\Delta X_i.$$

It holds that

$$\frac{1}{k_n}\sum_{j=1}^{k_n-1}g\left(\frac{k_n-j}{k_n}\right)\to\int_0^1 g(x)dx=\psi_1.$$

For the second part, we rewrite

$$\frac{1}{\sqrt{h_n}}\sum_{i=k_n-1}^{n-(k_n-2)}K\left(\frac{t_i-t}{h_n}\right)\Delta X_i =$$

$$\begin{aligned}&\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_i-t}{h_n}\right)\Delta X_i\\&-\frac{1}{\sqrt{h_n}}\sum_{i=1}^{k_n-2}K\left(\frac{t_i-t}{h_n}\right)\Delta X_i\\&-\frac{1}{\sqrt{h_n}}\sum_{i=n-(k_n-3)}^{n}K\left(\frac{t_i-t}{h_n}\right)\Delta X_i.\end{aligned}$$

The first term is the pure drift estimator in absence of noise, and from Christensen et al. (2016) it follows that this term converges stably in law to $N(0,K_2\sigma^2_{t-})$.

The second and third term converges in probability to zero (it follows with same argumentation as the expressions from (78) and (80)).

Using the equivalence of Slutsky's for stable convergence, we can collect the terms to conclude that

$$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1}\Delta\bar{X}_i\xrightarrow{d}N(0,\psi_1^2K_2\sigma^2_{t-})$$

and the convergence is stably in law.

21.5.2 The noise part

As with the X process, the noise part can be rewritten as

$$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1} K\left(\frac{t_i-t}{h_n}\right)\Delta\bar{\epsilon}_i = \frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right)(\epsilon_{i-j}-\epsilon_{i-j-1})\right)$$
$$= \frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1} K\left(\frac{t_i-t}{h_n}\right)\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)\epsilon_{i-j}\right)$$

$$= \frac{1}{k_n\sqrt{h_n}}\sum_{i=1}^{k_n-2}\left(\sum_{j=1}^{i} g\left(\frac{k_n-j}{k_n}\right)\right) K\left(\frac{t_i-t}{h_n}\right)\Delta\epsilon_i$$
$$+ \frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n-1}^{n-(k_n-2)}\left(\sum_{j=1}^{k_n-1} g\left(\frac{k_n-j}{k_n}\right)\right) K\left(\frac{t_i-t}{h_n}\right)\Delta\epsilon_i$$
$$+ \frac{1}{k_n\sqrt{h_n}}\sum_{i=n-(k_n-3)}^{n}\left(\sum_{j=1}^{n+1-i} g\left(\frac{j}{k_n}\right)\right) K\left(\frac{t_i-t}{h_n}\right)\Delta\epsilon_i$$

For the noise examinations, the $\Delta\epsilon_i$ terms are re-written to ϵ_i giving

$$= \frac{1}{k_n\sqrt{h_n}}\sum_{i=0}^{k_n-2}\left(\left(K\left(\frac{t_i-t}{h_n}\right)-K\left(\frac{t_{i+1}-t}{h_n}\right)\right)\left(\sum_{j=1}^{i} g\left(\frac{k_n-j}{k_n}\right)\right)-K\left(\frac{t_{i+1}-t}{h_n}\right) g\left(\frac{k_n-(i+1)}{k_n}\right)\right)\epsilon_i$$
$$+\left(\frac{1}{k_n}\sum_{j=1}^{k_n-1} g\left(\frac{k_n-j}{k_n}\right)\right)\frac{1}{\sqrt{h_n}}\sum_{i=k_n-1}^{n-(k_n-1)}\left(K\left(\frac{t_i-t}{h_n}\right)-K\left(\frac{t_{i+1}-t}{h_n}\right)\right)\epsilon_i$$
$$+\frac{1}{k_n\sqrt{h_n}}\sum_{i=n-(k_n-2)}^{n}\left(\left(K\left(\frac{t_i-t}{h_n}\right)-K\left(\frac{t_{i+1}-t}{h_n}\right)\right)\sum_{j=1}^{n-i} g\left(\frac{j}{k_n}\right)+K\left(\frac{t_i-t}{h_n}\right) g\left(\frac{n+1-i}{k_n}\right)\right)\epsilon_i.$$

We wish to show the convergence one term at a time. Before showing the above, it will be useful with a sub-result:

Lemma 3.

$$\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}\left(K\left(\frac{t_i-t}{h_n}\right)-K\left(\frac{t_{i+1}-t}{h_n}\right)\right)\epsilon_i \xrightarrow{p} 0 \tag{81}$$

for $k_n \propto \sqrt{n}$, $h_n \to 0$ and $\sqrt{n}h_n \to \infty$.

Proof. We make the convention $\Delta K_i = K\left(\frac{t_{i-1}-t}{h_n}\right) - K\left(\frac{t_i-t}{h_n}\right)$.

We naturally have an expectation of zero due to linearity of expectations. For the variance we have

$$\mathrm{Var}\left(\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n-1}\left(K\left(\frac{t_{i-1}-t}{h_n}\right)-K\left(\frac{t_i-t}{h_n}\right)\right)\epsilon_i\right) = \frac{1}{h_n}\mathrm{Var}\left(\sum_{i=1}^{n-1}\Delta K_i\epsilon_i\right)$$

and we want to show, that this converges to zero. If the assumptions in Lemma 2 are fulfilled for the noise (which is assumed), it is enough to show that

$$\frac{1}{h_n}\sum_{i=1}^{n-1}\mathrm{Var}(\Delta K_i\epsilon_i) = \frac{1}{h_n}\omega^2\sum_{i=1}^{n-1}(\Delta K_i)^2$$

converges to zero. Since ΔK_i is a telescope sum, we have $\sum_{i=1}^{n-1}\Delta K_i = K(0)-K(n) < K(0) = 1$ so we have

$$\frac{1}{h_n}\omega^2\sum_{i=1}^{n-1}(\Delta K_i)^2 \leq \frac{1}{h_n}\omega^2\sum_{i=1}^{n-1}(\max_i|\Delta K_i|)\Delta K_i \leq \frac{1}{h_n}\omega^2\max_i|\Delta K_i|.$$

Because $K'(x) > 0$, the difference is biggest at zero, so

$$\max_i \Delta K_i \leq e^0 - e^{0-\frac{dt}{h_n}} = 1 - e^{\frac{t}{nh_n}}.$$

If we and again let $h_n \propto n^{-\alpha}$ we have

$$\omega^2 n^\alpha\left(1-e^{-tn^{\alpha-1}}\right) \propto \frac{1-e^{-tn^{\alpha-1}}}{n^{-\alpha}}.$$

Using L'Hôpital gives

$$\frac{t(\alpha-1)n^{\alpha-2}e^{-tn^{\alpha-1}}}{(-\alpha)n^{-\alpha-1}} \propto n^{2\alpha-1}e^{-tn^{\alpha-1}}.$$

This converges to a constant $\alpha = 1/2$, meaning that as long as $\sqrt{n}h_n \to \infty$, the variance converges to zero. Then the expression must converge in probability to zero. □

With this lemma we can go back to showing the convergence of the three terms in (81).

<u>The first term</u>
The first term can be split into two (expanding the parenthesis into two parts. We have

$$|\frac{1}{k_n\sqrt{h_n}}\sum_{i=0}^{k_n-2}K\left(\frac{t_{i+1}-t}{h_n}\right)g\left(\frac{k_n-(i+1)}{k_n}\right)\epsilon_i| \leq$$

$$\left(\frac{1}{\sqrt{k_nh_n}}\right)\frac{1}{\sqrt{k_n}}\sum_{i=0}^{k_n-2}|K\left(\frac{t_{i+1}-t}{h_n}\right)g\left(\frac{k_n-(i+1)}{k_n}\right)\epsilon_i| \leq$$

$$\left(\frac{1}{\sqrt{k_nh_n}}\right)\frac{1}{\sqrt{k_n}}\sum_{i=0}^{k_n-2}|\epsilon_i|.$$

We have that $\frac{1}{\sqrt{k_nh_n}} \to 0$ because $\sqrt{n}h_n \to \infty$. The other term converges to a normal distribution (simple CLT), so by Slutsky it all converges in probability to zero.

For the other part of the first term, we have

$$|\frac{1}{k_n\sqrt{h_n}}\sum_{i=0}^{k_n-2}\Delta K\left(\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right)\right)\epsilon_i| =$$

$$|\frac{1}{\sqrt{h_n}}\sum_{i=0}^{k_n-2}\Delta K\left(\frac{1}{k_n}\sum_{j=1}^{i}g\left(\frac{k_n-j}{k_n}\right)\right)\epsilon_i| \leq$$

$$|\frac{1}{\sqrt{h_n}}\sum_{i=0}^{k_n-2}\Delta KC\epsilon_i|$$

where the last inequality follows by the fact that

$$\frac{1}{k_n}\sum_{j=1}^{i} g\left(\frac{k_n-j}{k_n}\right) \leq \frac{1}{k_n}\sum_{j=1}^{k_n-1} g\left(\frac{k_n-j}{k_n}\right) \to \int_0^1 g(x)dx$$

and because it is bounded by a term converging in probability it must be bounded by some constant C.

Rewriting

$$|\frac{1}{\sqrt{h_n}}\sum_{i=0}^{k_n-2}\Delta KC\epsilon_i| = \left(C\frac{\Delta K}{\sqrt{k_n h_n}}\right)\frac{1}{\sqrt{k_n}}\sum_{i=0}^{k_n-2}|\epsilon_i|$$

we get by CLT that the last part converges to a normal distribution and we only need $\frac{\Delta K}{\sqrt{h_n/k_n}} \to 0$ for the expression to converge in probability.

It was already seen in the proof of Lemma 3 that $\frac{\max(\Delta K)}{\sqrt{n}} \to k$ for some constant k, and since $\sqrt{h_n/k_n}\sqrt{n} = \sqrt{\frac{h_n n}{k_n}} \propto \sqrt{h_n\sqrt{n}} \to \infty$, we must have $\frac{\Delta K}{\sqrt{h_n/k_n}} \to 0$.

From this it follows that the first term in (81) converges in probability to zero.

The second term
The convergence of the second term follows from the single fact that the first part is bounded and the second part of the term converges in probability to zero from Lemma 3.

The third term
The convergence of the third term follows exactly as the convergence of the first term.

We conclude by Slutskys that we in fact have
$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1} K\left(\frac{t_i-t}{h_n}\right)\Delta\bar{\epsilon}_i \xrightarrow{p} 0$.

21.5.3 Conclusion

Since it is now shown that

$$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1} K\left(\frac{t_i-t}{h_n}\right)\Delta\bar{X}_i \xrightarrow{d} N(0,\psi_1^2 K_2\sigma_{t-}^2)$$

stably in law and

$$\frac{1}{k_n\sqrt{h_n}}\sum_{i=k_n}^{n+1} K\left(\frac{t_i-t}{h_n}\right)\Delta\bar{\epsilon}_i \xrightarrow{p} 0$$

and by Slutsky's for stable convergence it follows that

$$\sqrt{h_n}\hat{\mu}_t^n \xrightarrow{d} N(0,\psi_1^2 K_2\sigma_{t-}^2)$$

stably in law, as we wanted to show.

21.6 Proof of Theorem 8

We want to investigate the convergence of

$$\hat{\sigma}_t^2 = \frac{1}{k_n h_n} \sum_{i=1}^{n} (\Delta \bar{Y}_i)^2 = \frac{1}{k_n h_n} \sum_{i=1}^{n} \left((\Delta \bar{X}_i)^2 + (\Delta \bar{\epsilon}_i)^2 + 2(\Delta \bar{X}_i)(\Delta \bar{\epsilon}_i) \right)$$

when the noise process is iid. We split it up into the X-process, noise process and joint product.

21.6.1 The X-process

The process is rewritten to

$$\frac{1}{k_n h_n} \sum_{i=k_n}^{n+1} (\Delta \bar{X}_i)^2 = \frac{1}{k_n h_n} \sum_{i=k_n}^{n+1} \left(\sum_{j=1}^{k_n - 1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j} - t}{h_n}\right) (X_{i-j} - X_{i-j-1}) \right)^2$$

$$= \frac{1}{k_n h_n} \sum_{i=k_n}^{n+1} \left(\sum_{j=1}^{k_n - 1} g\left(\frac{j}{k_n}\right)^2 K\left(\frac{t_{i-j} - t}{h_n}\right)^2 (X_{i-j} - X_{i-j-1})^2 \right) \tag{82}$$

$$+ \frac{1}{k_n h_n} \sum_{i=k_n}^{n+1} \left(\sum_{j=1}^{k_n - 1} \sum_{k=1, k \neq j}^{k_n - 1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j} - t}{h_n}\right) (X_{i-j} - X_{i-j-1}) g\left(\frac{k}{k_n}\right) K\left(\frac{t_{i-k} - t}{h_n}\right) (X_{i-k} - X_{i-k-1}) \right). \tag{83}$$

We first show the convergence of the first term. We can follow the exact same steps as in the convergence of the drift estimator. By splitting it into three parts, we get

$$= \frac{1}{k_n h_n} \sum_{i=1}^{k_n - 2} \left(\sum_{j=1}^{i} g\left(\frac{k_n - j}{k_n}\right)^2 \right) K\left(\frac{t_i - t}{h_n}\right)^2 (\Delta X_i)^2 \tag{84}$$

$$+ \frac{1}{k_n h_n} \sum_{i=k_n - 1}^{n - (k_n - 2)} \left(\sum_{j=1}^{k_n - 1} g\left(\frac{k_n - j}{k_n}\right)^2 \right) K\left(\frac{t_i - t}{h_n}\right)^2 (\Delta X_i)^2 \tag{85}$$

$$+ \frac{1}{k_n h_n} \sum_{i=n - (k_n - 3)}^{n} \left(\sum_{j=1}^{n+1-i} g\left(\frac{j}{k_n}\right)^2 \right) K\left(\frac{t_i - t}{h_n}\right)^2 (\Delta X_i)^2. \tag{86}$$

We first investigate the first term. Similar to earlier methods, we can create an upper limit

$$\frac{1}{k_n h_n} \sum_{i=1}^{k_n - 2} \left(\sum_{j=1}^{i} g\left(\frac{k_n - j}{k_n}\right)^2 \right) K\left(\frac{t_i - t}{h_n}\right)^2 (\Delta X_i)^2 \leq$$

$$\frac{1}{k_n h_n} \sum_{i=1}^{k_n - 2} \left(\sum_{j=1}^{i} g\left(\frac{k_n - j}{k_n}\right)^2 \right) (\Delta X_i)^2 \leq$$

$$\frac{c}{h_n} \sum_{i=1}^{k_n - 2} (\Delta X_i)^2.$$

We have $(\Delta X_i)^2 = O_p(\frac{1}{n})$ so $n(\Delta X_i)^2 = O_p(1)$. This means that

$$\frac{c_1}{h_n}\frac{1}{n}\sum_{i=1}^{k_n} n(\Delta X_i)^2 = \frac{c_2}{h_n k_n}\left(\frac{1}{k_n}\sum_{i=1}^{k_n} n(\Delta X_i)^2\right).$$

The expression in the parenthesis is bounded as $n(\Delta X_i)^2 = O_p(1)$ and because $\frac{1}{h_n k_n} \to 0$ it converges in probability to zero.

The third term also converges in probability to zero by the same reasoning.

The second term can be rewritten to

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n-1} g\left(\frac{k_n - j}{k_n}\right)^2\right)\frac{1}{h_n}\sum_{i=k_n-1}^{n-(k_n-2)} K\left(\frac{t_i - t}{h_n}\right)^2 (\Delta X_i)^2.$$

The first part converges to $\int_0^1 g(x)^2 dx = \psi_2$, and from eq. (45) in Christensen et al. (2016) the second part converges in probability to $K_2\sigma^2_{t-}$. Thus the whole part converges in probability to $\psi_2 K_2 \sigma^2_{t-}$.

For the expression in (83) we have

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} Y_{i,k_n}$$

where Y_{i,k_n} represents the expression in the big parenthesis in (83). We have that

$$\mathrm{Var}\left(\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} Y_{i,k_n}\right) = \frac{1}{k_n^2 h_n^2}\mathrm{Var}\left(\sum_{i=k_n}^{n+1} Y_{i,k_n}\right) =$$

$$\frac{1}{k_n^2 h_n^2}\sum_{i=k_n}^{n+1}\sum_{j=k_n}^{n+1}\mathrm{cov}(Y_{i,k_n}, Y_{j,k_n}) = \frac{1}{k_n h_n^2}\sum_{i=k_n}^{n+1}\frac{1}{k_n}\sum_{j=k_n}^{n+1}\mathrm{cov}(Y_{i,k_n}, Y_{j,k_n}).$$

Because Y_{i,k_n} only consists of k_n delta-sequences, it only has covariance with maximum $2k_n$ of the Y_{j,k_n}'s. These covariances are decreasing (and less than the variance), so there exists a constant c (by similar reasoning as in Lemma 2) such that

$$\frac{1}{k_n h_n^2}\sum_{i=k_n}^{n+1}\frac{1}{k_n}\sum_{j=k_n}^{n+1}\mathrm{cov}(Y_{i,k_n}, Y_{j,k_n}) \leq \frac{1}{k_n h_n^2}\sum_{i=k_n}^{n+1}\frac{1}{k_n}\sum_{j=i-k_n}^{i+k_n}\mathrm{cov}(Y_{i,k_n}, Y_{j,k_n}) \leq \frac{1}{k_n h_n^2}\sum_{i=k_n}^{n+1} c\cdot\mathrm{Var}(Y_{i,k_n}).$$

Regarding the variance of Y_{i,k_n} we have

$$\mathrm{Var}\left(\sum_{j=1}^{k_n-1}\sum_{k=1,k\neq j}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(X_{i-j}-X_{i-j-1})g\left(\frac{k}{k_n}\right)K\left(\frac{t_{i-k}-t}{h_n}\right)(X_{i-k}-X_{i-k-1})\right).$$

Because K is monotone, there exists an $\alpha \in [0,1]$ such that

$$\mathrm{Var}\left(\sum_{j=1}^{k_n-1}\sum_{k=1,k\neq j}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(X_{i-j}-X_{i-j-1})g\left(\frac{k}{k_n}\right)K\left(\frac{t_{i-k}-t}{h_n}\right)(X_{i-k}-X_{i-k-1})\right) =$$

$$\text{Var}\left(K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^2\sum_{j=1}^{k_n-1}\sum_{k=1,k\neq j}^{k_n-1}g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})g\left(\frac{k}{k_n}\right)(X_{i-k}-X_{i-k-1})\right)=$$

$$K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4\text{Var}\left(\sum_{j=1}^{k_n-1}\sum_{k=1,k\neq j}^{k_n-1}g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})g\left(\frac{k}{k_n}\right)(X_{i-k}-X_{i-k-1})\right).$$

Because ΔX is stationary and independent, we have from CLT that

$$\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\Delta X\xrightarrow{d}N\left(0,\text{Var}(\Delta X)\right)$$

which means that

$$\left(\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\Delta X\right)^2\xrightarrow{d}\chi^2_{\text{Var}(\Delta X)}$$

and thus

$$\text{Var}\left[\left(\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\Delta X\right)^2\right]\xrightarrow{p}2\text{Var}(\Delta X).$$

From this follows that the variance is bounded by a constant, so

$$\text{Var}\left[\left(\sum_{i=1}^{k_n}\Delta X\right)^2\right]\leq c\cdot k_n^2\text{Var}(\Delta X).$$

This means that

$$K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4\text{Var}\left(\sum_{j=1}^{k_n-1}\sum_{k=1,k\neq j}^{k_n-1}g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})g\left(\frac{k}{k_n}\right)(X_{i-k}-X_{i-k-1})\right)\leq$$

$$K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4ck_n^2\max_{j,k}\text{Var}\left(g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})g\left(\frac{k}{k_n}\right)(X_{i-k}-X_{i-k-1})\right)\leq$$

$$K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4ck_n^2\max_{j,k}\text{Var}\left((X_{i-j}-X_{i-j-1})(X_{i-k}-X_{i-k-1})\right)\leq$$

$$ck_n^2K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4\text{Var}\left(\Delta X_{i-j}\cdot\Delta X_{j-k}h_n\right)=ck_n^2K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4\text{Var}(\Delta X_{i-j})\text{Var}(\Delta X_{i-k})$$

where the last equality follows because of independence and an expectation of zero. Thus

$$\text{Var}\left(\frac{1}{k_nh_n}\sum_{i=k_n}^{n+1}Y_{i,k_n}\right)\leq\frac{c_1}{k_nh_n^2}\sum_{i=k_n}^{n+1}\text{Var}(Y_{i,k_n})\leq$$

$$\frac{c_1}{k_nh_n^2}\sum_{i=k_n}^{n+1}c_2k_n^2K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4\text{Var}(\Delta X_{i-j})^2=$$

$$\left(\frac{c_3k_n}{h_nn}\right)\frac{1}{nh_n}\sum_{i=k_n}^{n+1}K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4n^2\text{Var}(\Delta X_{i-j})^2.$$

The first part converges in probability to zero because $k_n = \theta\sqrt{n}$ and $\sqrt{n}h_n \to \infty$. For the rest of the expression we have

$$\frac{t}{nh_n}\sum_{i=1}^{n} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 \to \int_{-\infty}^{0} K(x)^4 dx.$$

and also $n^2\mathrm{Var}(\Delta X_{i-j})^2 = O_p(1)$. From that it follows that the whole expression converges in probability to zero.

Putting it all together gives

$$\frac{1}{k_n h_n}\sum_{i=1}^{n}(\Delta\bar{X}_i)^2 \xrightarrow{p} \psi_2 K_2 \sigma_{t-}^2$$

21.6.2 The noise part

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}(\Delta\bar{\epsilon}_i)^2 = \frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right)(\epsilon_{i-j}-\epsilon_{i-j-1})\right)^2 =$$

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}\left(g\left(\frac{1}{k_n}\right) K\left(\frac{t_{i-1}-t}{h_n}\right)\epsilon_{i-1} - g\left(\frac{k_n-1}{k_n}\right) K\left(\frac{t_{i-(k_n-1)}-t}{h_n}\right)\epsilon_{i-k_n} + \sum_{j=2}^{k_n-1}\left(g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) - g\left(\frac{j-1}{k_n}\right) K\left(\frac{t_{i-(j-1)}-t}{h_n}\right)\right)\epsilon_{i-j}\right)^2 =$$

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) - g\left(\frac{j-1}{k_n}\right) K\left(\frac{t_{i-(j-1)}-t}{h_n}\right)\right)\epsilon_{i-j}\right)^2$$

exploiting that $g(0) = g(1) = 0$.

We expand the outer parenthesis and (as with the drift part) we split it into the squared terms and products

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) - g\left(\frac{j-1}{k_n}\right) K\left(\frac{t_{i-(j-1)}-t}{h_n}\right)\right)^2 \epsilon_{i-j}^2 + \sum_{j=1}^{k_n}\sum_{k=1,k\neq j}^{k_n}\left(g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) - g\left(\frac{j-1}{k_n}\right) K\left(\frac{t_{i-(j-1)}-t}{h_n}\right)\right)\epsilon_{i-j} \cdot \left(g\left(\frac{k}{k_n}\right) K\left(\frac{t_{i-k}-t}{h_n}\right) - g\left(\frac{k-1}{k_n}\right) K\left(\frac{t_{i-(k-1)}-t}{h_n}\right)\right)\epsilon_{i-k}\right)$$

The squared part is investigated first. It is first noted, that because K is monotone then for all j there exists an $\alpha \in [0,1]$ such that

$$\left(g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) - g\left(\frac{j-1}{k_n}\right) K\left(\frac{t_{i-(j-1)}-t}{h_n}\right)\right)^2 = \left(\left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right) K\left(\frac{t_{i-(j-\alpha_j)}-t}{h_n}\right)\right)^2.$$

For the squared terms we can create a sum of ϵ's split into three parts as is done in previous proofs:

$$\frac{1}{k_n h_n}\sum_{i=1}^{k_n-1}(...)\cdot\epsilon_{i-1}^2+$$
$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n-k_n+1}\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)^2 K\left(\frac{t_{i-(j-\alpha_j)}-t}{h_n}\right)^2\right)\cdot\epsilon_{i-1}^2+$$
$$\frac{1}{k_n h_n}\sum_{i=n-k_n+2}^{n+1}(...)\cdot\epsilon_{i-1}^2+$$

The first and last part converges in probability to zero, which will not be shown this time (the approach is similar to the other times where the sum was split into three). It is intuitive from the fact that it only sums over k_n observations whereas the term in the middle sums over $\sim n$ which will be shown to converge in probability.

For the middle part we can once again use the fact, that there exists $\alpha\in[0,1]$ such that

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n-k_n+1}\left(\sum_{j=1}^{k_n}(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right))^2 K\left(\frac{t_{i-(j-\alpha_j)}-t}{h_n}\right)^2\right)\cdot\epsilon_{i-1}^2=$$
$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n-k_n+1}\left(\sum_{j=1}^{k_n}(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right))^2\right)\cdot K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^2\epsilon_{i-1}^2.$$

Because K is the exponential kernel, we have

$$K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^2=K\left(\frac{t_{i-k_n\alpha}-t_i}{h_n}\right)^2\cdot K\left(\frac{t_i-t}{h_n}\right)^2=e^{-\frac{k_n\alpha t}{nh_n}}K\left(\frac{t_i-t}{h_n}\right)^2$$

and we can thus rewrite

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n-k_n+1}\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)^2\right)\cdot K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^2\epsilon_{i-1}^2=$$
$$\left(e^{-\frac{k_n\alpha t}{nh_n}}k_n\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)^2\right)\frac{1}{k_n^2 h_n}\sum_{i=k_n}^{n-k_n+1}K\left(\frac{t_i-t}{h_n}\right)^2\epsilon_{i-1}^2.$$

The expression in the parenthesis converges to $\psi_3=\int_0^1 g'(x)^2dx$ because

$$e^{-\frac{k_n\alpha t}{nh_n}}\to 1$$

and

$$k_n\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)^2=\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\frac{g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)}{1/k_n}\right)^2\to\int_0^1 g'(x)^2dx.$$

For the right part of the expression we exploit that $\frac{t}{nh_n}\sum_{i=1}^{n} K\left(\frac{t_i-t}{h_n}\right)^2 \to K_2$ and $k_n^2 = \theta \cdot n$ to see that

$$E\left[\frac{1}{k_n^2 h_n}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^2 \epsilon_{i-1}^2\right] =$$
$$\frac{1}{k_n^2 h_n}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^2 E\left[\epsilon_{i-1}^2\right] =$$
$$\frac{1}{\theta^2 n h_n}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^2 \omega^2 \to \frac{K_2}{\theta^2 t}\omega^2$$

With restriction on the sum of covariance of the noise, we also have that the variance converges in probability to zero. In the independent case, we have

$$\mathrm{Var}\left(\frac{1}{k_n^2 h_n}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^2 \epsilon_{i-1}^2\right) =$$
$$\frac{1}{k_n^4 h_n^2}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^4 \mathrm{Var}(\epsilon_{i-1}^2) =$$
$$\frac{1}{k_n^4 h_n^2}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^4 \mathrm{Var}(\epsilon_{i-1}^2) =$$
$$\left(\frac{1}{k_n^2 h_n}\right)\frac{1}{k_n^2 h_n}\sum_{i=k_n}^{n-k_n+1} K\left(\frac{t_i-t}{h_n}\right)^4 \omega^2 \to 0$$

as the term in the parenthesis converges to zero and the right part converges to a constant.

For the product part we first let Y_{i,k_n} be defined as

$$Y_{i,k_n} = \sum_{j=1}^{k_n}\sum_{k=1,k\neq j}^{k_n}\left(g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right) - g\left(\frac{j-1}{k_n}\right)K\left(\frac{t_{i-(j-1)}-t}{h_n}\right)\right)\epsilon_{i-j}$$
$$\cdot\left(g\left(\frac{k}{k_n}\right)K\left(\frac{t_{i-k}-t}{h_n}\right) - g\left(\frac{k-1}{k_n}\right)K\left(\frac{t_{i-(k-1)}-t}{h_n}\right)\right)\epsilon_{i-k}$$

$$= K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 \sum_{j=1}^{k_n}\sum_{k=1,k\neq j}^{k_n}\left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right)\epsilon_{i-j}$$
$$\cdot\left(g\left(\frac{k}{k_n}\right) - g\left(\frac{k-1}{k_n}\right)\right)\epsilon_{i-k}$$

With exactly same reasoning as the X part, we get

$$\mathrm{Var}\left(\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} Y_{i,k_n}\right) \leq \frac{c}{k_n h_n^2}\sum_{i=k_n}^{n+1}\mathrm{Var}(Y_{i,k_n}).$$

We also have

$$\mathrm{Var}(Y_{i,k_n}) \leq$$

$$ck_n^2 K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 \max_{j,k} \mathrm{Var}\Bigg(\left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right)\epsilon_{i-j} \cdot \left(g\left(\frac{k}{k_n}\right) - g\left(\frac{k-1}{k_n}\right)\right)\epsilon_{i-k}\Bigg)$$

with the same argumentation as in the proof of the convergence of the X-process above (because the noise is iid). We can again introduce $\alpha \in [0,1]$ to get the kernel K outside the variance expression to get

$$= ck_n^2 K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 \max_{j,k} \mathrm{Var}\left(\Delta g_j \epsilon_{i,j} \cdot \Delta g_k \epsilon_{j,k}\right) \leq c\frac{1}{k_n^4} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 k_n^2 \mathrm{Var}\left(\epsilon_{i,j}\cdot\epsilon_{j,k}\right)$$
$$\leq c\frac{1}{k_n^2} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 \max\left(\mathrm{Var}(\epsilon_{i,j}^2), \mathrm{Var}(\epsilon_{i,k}^2)\right) \leq c\frac{E(\epsilon^4)}{k_n^2} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4.$$

Then

$$\frac{c_1}{k_n h_n^2}\sum_{i=k_n}^{n+1} \mathrm{Var}(Y_{i,k_n}) \leq \frac{c_1}{k_n h_n^2}\sum_{i=k_n}^{n+1} c_2 \frac{E(\epsilon^4)}{k_n^2} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 =$$
$$\frac{c_3 E(\epsilon^4)}{k_n h_n}\frac{1}{\theta^2 n}\frac{1}{h_n}\sum_{i=k_n}^{n+1} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4.$$

and because $k_n h_n \to \infty$ and

$$\frac{t}{nh_n}\sum_{i=1}^{n} K\left(\frac{t_{i-k_n\alpha}-t}{h_n}\right)^4 \to \int_{-\infty}^{0} K(x)^4 dx$$

the variance converges to zero. Thus the product part converges in probability to zero.

From that follows that

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} (\Delta\bar{\epsilon}_i)^2 \xrightarrow{p} \frac{\psi_3 K_2}{\theta^2 t}\omega^2$$

21.6.3 The product part

We investigate the variance

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} (\Delta\bar{X}_i)(\Delta\bar{\epsilon}_i).$$

The expectation is zero because of independence between noise and X and that $E(\Delta\bar{X}_i) = 0$.

For the variance we rewrite

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right)(\epsilon_{i-j}-\epsilon_{i-j-1})\right)\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right)(X_{i-j}-X_{i-j-1})\right) =$$

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)\epsilon_{i-j}\right)\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(X_{i-j}-X_{i-j-1})\right)=$$

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} K\left(\frac{t_{i-k_n\alpha_{i_1}}-t}{h_n}\right)K\left(\frac{t_{i-k_n\alpha_{i_2}}-t}{h_n}\right)\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)\epsilon_{i-j}\right)\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})\right)$$

which follows from the same argumentation as for the noise part and with introducing $\alpha \in [0,1]$ as previously done.

We can again make the upper limit

$$\mathrm{Var}\left(\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} Y_{i,k_n}\right) \leq \frac{c}{k_n h_n^2}\sum_{i=k_n}^{n+1}\mathrm{Var}(Y_{i,k_n})$$

and

$$\begin{aligned}\mathrm{Var}(Y_{i,k_n}) =& K\left(\frac{t_{i-k_n\alpha_{i_1}}-t}{h_n}\right)^2 K\left(\frac{t_{i-k_n\alpha_{i_2}}-t}{h_n}\right)^2 \\ &\cdot \mathrm{Var}\left(\left(\sum_{j=1}^{k_n}\left(g\left(\frac{j}{k_n}\right)-g\left(\frac{j-1}{k_n}\right)\right)\epsilon_{i-j}\right)\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)(X_{i-j}-X_{i-j-1})\right)\right) \leq\end{aligned}$$

$$K\left(\frac{t_{i-k_n\alpha_{i_1}}-t}{h_n}\right)^2 K\left(\frac{t_{i-k_n\alpha_{i_2}}-t}{h_n}\right)^2 c k_n^2 \max_j \mathrm{Var}\left(\Delta g_j \epsilon_j g\left(\frac{j}{k_n}\right)\Delta X\right).$$

From the same CLT-argumentation made previously. We have

$$\max_j \mathrm{Var}\left(\Delta g_j \epsilon_j g\left(\frac{j}{k_n}\right)\Delta X\right) \leq \max_j \mathrm{Var}\left(\frac{1}{k_n}\left(\frac{\Delta g_j}{1/k_n}\right)\epsilon_j g\left(\frac{j}{k_n}\right)\frac{1}{\sqrt{n}}\left(\sqrt{n}\Delta X\right)\right) \leq$$

$$\frac{1}{k_n^2 n^2}\max_j \mathrm{Var}\left(g'(j)(\sqrt{n}\Delta X)\right) \leq \frac{c}{k_n^2 n^2}.$$

Then

$$\mathrm{Var}\left(\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1} Y_{i,k_n}\right) \leq \frac{c}{k_n h_n^2}\sum_{i=k_n}^{n+1}\mathrm{Var}(Y_{i,k_n}) \leq$$

$$\frac{c_1}{k_n h_n^2}\sum_{i=k_n}^{n+1} K\left(\frac{t_{i-k_n\alpha_{i_1}}-t}{h_n}\right)^2 K\left(\frac{t_{i-k_n\alpha_{i_2}}-t}{h_n}\right)^2 k_n^2 \frac{c_2}{k_n^2 n^2} =$$

$$\frac{c_3}{k_n h_n}\frac{1}{h_n n}\sum_{i=k_n}^{n+1} K\left(\frac{t_{i-k_n\alpha_{i_1}}-t}{h_n}\right)^2 K\left(\frac{t_{i-k_n\alpha_{i_2}}-t}{h_n}\right)^2.$$

We have $K\left(\frac{t_{i-k_n\alpha_i}-t}{h_n}\right)^2 \to K\left(\frac{t_i-t}{h_n}\right)^2$, $\frac{t}{h_n n}\sum_{i=k_n}^{n+1} K\left(\frac{t_i-t}{h_n}\right)^4 \to \int_{-\infty}^{0} K(x)^4 dx$ and $k_n h_n \to \infty$. Thus the whole expression converges to zero, and we end up with

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}(\Delta\bar{X}_i)(\Delta\bar{\epsilon}_i) \xrightarrow{p} 0.$$

21.6.4 Collecting everything

By using Slutsky's equation we end up with

$$\frac{1}{k_n h_n}\sum_{i=k_n}^{n+1}(\Delta\bar{Y}_i)^2 \xrightarrow{p} \psi_2 K_2 \sigma_{t-}^2 + \frac{\psi_3 K_2}{\theta^2 t}\omega^2$$

which is exactly as we wanted.

21.7 Proof of Theorem 9

We wish to investigate how a jump contributes to the T-estimator. We denote the jump time τ, and the number of time points after a jump as n_p. We can then write the drift-estimator at any time point following a jump as

$$\begin{aligned}
\mu_{\tau+n_p} &= \left(\frac{1}{\psi_1 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\Delta\bar{Y}_i\right) \\
&= \left(\frac{1}{\psi_1 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(Y_{i-j}-Y_{i-j-1}+J_{i-j})\right)\right) \\
&= \left(\frac{1}{\psi_1 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\Delta\bar{Y}_i\right) + \left(\frac{1}{\psi_1 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)J_{i-j}\right)\right) \\
&= \mu_{\tau+n_p}^{no\ jump} + \left(\frac{1}{\psi_1 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)J_{i-j}\right)\right).
\end{aligned}$$

We note that $J_{i=j}=0$ for $i-j\neq\tau$ and $J_{i=j}=J$ for $i-j=\tau$.

For the volatility-estimator, we can rewrite

$$\begin{aligned}
(\sigma_{\tau+n_p})^2 &= \left(\frac{1}{\psi_2 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}(\Delta\bar{Y}_i)^2\right) - \frac{\psi_3}{\psi_2\theta^2}\hat{\omega}^2 \\
&= \left(\frac{1}{\psi_2 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(Y_{i-j}-Y_{i-j-1}+J_{i-j})\right)^2\right) - \frac{\psi_3}{\psi_2\theta^2}\hat{\omega}^2 \\
&= \left(\frac{1}{\psi_2 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(Y_{i-j}-Y_{i-j-1})\right)+\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)J_{i-j}\right)\right)^2\right) - \\
&= \left(\frac{1}{\psi_2 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}(\Delta\bar{Y}_i)^2 - \frac{\psi_3}{\psi_2\theta^2}\hat{\omega}^2\right) + \left(\frac{1}{\psi_2 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)J_{i-j}\right)^2\right) \\
&\quad + \left(\frac{2}{\psi_2 k_n h_n}\sum_{i=k_n}^{\tau+n_p+1}\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)(Y_{i-j}-Y_{i-j-1})\right)\left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-j}-t}{h_n}\right)J_{i-j}\right)\right)
\end{aligned}$$

We now split the investigation of convergence into three parts:

1) The time-points before $\tau + k_n - 1$ (i.e. where $n_p < k_n - 1$).
2) Time-point $\tau + k_n - 1$ (i.e. $n_p = k_n - 1$).
3) The time-points after $\tau + k_n - 1$ (i.e. $n_p > k_n - 1$).

The first part
Assuming $n_p < k_n - 1$ we have

$$\mu_{\tau+n_p} = \mu_{\tau+n_p}^{no\ jump} + \left(\frac{1}{k_n h_n} K\left(\frac{t_{i-j} - t}{h_n}\right) J_\tau \sum_{i=\tau+1}^{\tau+n_p+1} g\left(\frac{i-\tau}{k_n}\right)\right) =$$

$$\mu_{\tau+n_p} = \mu_{\tau+n_p}^{no\ jump} + \left(\frac{1}{k_n h_n} K\left(\frac{t_{\tau-1} - t_{\tau+n_p}}{h_n}\right) J_\tau \sum_{i=\tau+1}^{\tau+n_p} g\left(\frac{i-\tau}{k_n}\right)\right) =$$

$$\mu_{\tau+n_p} = \mu_{\tau+n_p}^{no\ jump} + \left(\frac{1}{k_n h_n} K\left(\frac{t_{\tau-1} - t_{\tau+n_p}}{h_n}\right) J_\tau \sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)\right).$$

The σ estimator can be written as

$$\sigma^2_{\tau+n_p} = \left(\frac{1}{\psi_2 k_n h_n} \sum_{i=k_n}^{\tau+n_p+1} (\Delta \bar{Y}_i)^2 - \frac{\psi_3}{\psi_2 \theta^2} \hat{\omega}^2\right) + \left(\frac{1}{\psi_2 k_n h_n} K\left(\frac{t_\tau - t_{\tau+n_p}}{h_n}\right)^2 J_\tau^2 \sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2\right)$$
$$+ \frac{2}{\psi_2 k_n h_n} \sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right) K\left(\frac{t_\tau - t_{\tau+n_p}}{h_n}\right) J_\tau \cdot \Delta \bar{Y}_{\tau+i}.$$

The first expression is the old estimator, which converges in probability to the true volatility.

From eq. (80) in the proof of the drift estimator, we had that

$$\frac{1}{k_n \sqrt{h_n}} \sum_{j=1}^{k_n - 1} g\left(\frac{j}{k_n}\right) \Delta \bar{Y}_{\tau+j} \xrightarrow{p} 0$$

and from that follows, that the last expression in the volatility-estimator is $O_p(\frac{1}{\sqrt{h_n}})$.

Regarding the second term, we have that $\frac{1}{k_n} \sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2 \to \int_0^{n_p/k_n} g(x)^2 dx$, so the expression is $O\left(\frac{1}{h_n}\right)$. From that follows that as $h_n \to 0$, the second term dominates (in the sence that the third term divided with the second converges to zero).

Since $\psi_1 = \frac{1}{k_n} \sum_{i=1}^{k_n - 1} g\left(\frac{i}{k_n}\right) + o_p(1)$ and $\psi_2 = \frac{1}{k_n} \sum_{i=1}^{k_n - 1} g\left(\frac{i}{k_n}\right)^2 + o(1)$, we have

$$\frac{\psi_2}{\psi_1} = \frac{\frac{1}{k_n} \sum_{i=1}^{k_n - 1} g\left(\frac{i}{k_n}\right)^2}{\frac{1}{k_n} \sum_{i=1}^{k_n - 1} g\left(\frac{i}{k_n}\right)} + o(1).$$

As $n_p < k_n - 1$ and $g(x) \geq 0$ we also have

$$\frac{\frac{1}{k_n}\sum_{i=1}^{k_n-1} g\left(\frac{i}{k_n}\right)^2}{\left(\frac{1}{k_n}\sum_{i=1}^{k_n-1} g\left(\frac{i}{k_n}\right)\right)^2} \leq \frac{\frac{1}{k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2}{\left(\frac{1}{k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)\right)^2}$$

and from this follows that

$$\frac{\psi_2}{\psi_1^2} \leq \frac{\frac{1}{k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2}{\left(\frac{1}{k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)\right)^2} \quad \Rightarrow$$

$$\frac{1}{\psi_1 k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right) \leq \sqrt{\frac{1}{\psi_2 k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2} + o(1).$$

This entails that the square root of the third term in the volatility estimator is greater than the second term in the drift estimator. Hence we can write the estimator as

$$T_{\tau+n_p} = \frac{\sqrt{h_n}\mu_{\tau+n_p}}{\sqrt{\sigma^2_{\tau+n_p}}} = \frac{N(0, K_2\sigma^2_{t-}) + k_1}{\sqrt{K_2\sigma^2_{t-} + k_2}} + o_p(1).$$

where $\sqrt{k_2} \geq k_1$.

This means that

$$P(T_{\tau+n_p} > q) \leq P(X > q)$$

with $X \sim \frac{1}{\sqrt{\sigma^2+k^2}}N(k, \sigma^2)$, and we find that the bound from Theorem 1 holds.

The second part

Assuming $n_p = k_n - 1$, we have

$$T_{\tau+k_n-1} = \frac{\sqrt{h_n}\mu_{\tau+k_n-1}}{\sqrt{\sigma^2_{\tau+k_n-1}}} = \frac{N(0, K_2\sigma^2_{t-}) + \left(\frac{1}{\psi_1 k_n\sqrt{h_n}}K\left(\frac{t_{\tau-1}-t_{\tau+n_p}}{h_n}\right)J_\tau\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)\right) + o_p(1)}{\sqrt{K_2\sigma^2_{t-} + \frac{1}{\psi_2 k_n h_n}K\left(\frac{t_{\tau-1}-t_{\tau+n_p}}{h_n}\right)^2 J_\tau^2\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2 + o_p\left(\frac{1}{\sqrt{h_n}}\right)}}$$

and multiplying with $\sqrt{h_n}$ in numerator and denominator gives

$$\begin{aligned}
&= \frac{K\left(\frac{t_{\tau-1}-t_{\tau+n_p}}{h_n}\right)J_\tau\frac{1}{\psi_1 k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right) + O_p(\sqrt{h_n})}{\sqrt{K\left(\frac{t_{\tau-1}-t_{\tau+n_p}}{h_n}\right)^2 J_\tau^2\frac{1}{\psi_2 k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2 + o_p(\sqrt{h_n})}} \\
&= \text{sign}(J)\cdot\frac{\frac{1}{\psi_1 k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right) + O_p(\sqrt{h_n})}{\frac{1}{\psi_2 k_n}\sum_{i=1}^{n_p} g\left(\frac{i}{k_n}\right)^2 + o_p(\sqrt{h_n})} \\
&\xrightarrow{p} \text{sign}(J).
\end{aligned}$$

The third part

For the third part, we just multiply the jump contribution with $K\left(\frac{t_{n_p}-t_{k_n-1}}{h_n}\right)$ compared to the second part. This means, that we still have

$$T_{\tau+n_p} = \frac{\sqrt{h_n}\mu_{\tau+n_p}}{\sqrt{\sigma^2_{\tau+n_p}}} = \frac{N(0, K_2\sigma^2_{t-}) + k_1}{\sqrt{K_2\sigma^2_{t-} + k_2}} + o_p(1).$$

with $\sqrt{k_2} \geq k_1$ and therefore the bound from Theorem 1 still holds.

21.8 Proof of Theorem 10

With pre-averaging, we can write

$$T_{\tau+(k_n-1)} = \frac{\sqrt{h_n}\mu_{\tau+(k_n-1)}}{\sqrt{\hat{\sigma}^2_{\tau+(k_n-1)}}} = \frac{\frac{1}{\psi_1 k_n\sqrt{h_n}}\sum_{i=1}^{\tau+k_n}\Delta\bar{Y} + \bar{D}_{i,n} + \bar{S}_{i,n}}{\left(\frac{1}{\psi_2 k_n h_n}\sum_{i=1}^{n}(\Delta\bar{Y} + \bar{D}_{i,n} + \bar{S}_{i,n})^2\right)^{1/2}}$$

where

$$\bar{D}_{i,n} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) D_{i-j,n} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) \int_{t_{i-j-1}}^{t_{i-j}} \frac{1}{(1-s)^\alpha} ds$$

and

$$\bar{S}_{i,n} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) S_{i-j,n} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) K\left(\frac{t_{i-j}-t}{h_n}\right) \int_{t_{i-j-1}}^{t_{i-j}} \frac{1}{(1-s)^\beta} dW_s.$$

At time-point $\tau + (k_n - 1)$ the pre-averaging gives every difference from point τ and before the weight $\sum_{i=1}^{k_n-1} g\left(\frac{i}{k_n}\right)$. Thus the numerator can be rewritten as

$$\sqrt{h_n}\mu_{\tau+(k_n-1)} = \left(\frac{1}{\psi_1 k_n\sqrt{h_n}}\sum_{i=1}^{\tau+k_n}\Delta\bar{Y}\right) + \left(\left(\frac{1}{\psi_1 k_n}\sum_{i=1}^{k_n-1} g\left(\frac{i}{k_n}\right)\right)\frac{1}{\sqrt{h_n}}\sum_{i=1}^{\tau} K\left(\frac{t_{i-j}-t}{h_n}\right)(D_{i,n} + S_{i,n})\right) + o_p(1).$$

We recognize the first part as the estimator without bursts. For the second part we first recall that $\frac{1}{\psi_1 k_n}\sum_{i=1}^{k_n-1} g\left(\frac{i}{k_n}\right) \to 0$. From the proof of the estimator in presence of bursts without microstructure noise it is known that

$$h^\beta \sum_{i=1}^{n} K\left(\frac{t_{i-1}-t}{h_n}\right)(D_{i,n} + S_{i,n}) \xrightarrow{d} N\left(h_n^{-\alpha+1/2+\beta}\Gamma(1-\alpha), 2^{2\beta-1}\Gamma(1-2\beta)\right)$$

from which it then follows, that

$$h_n^\beta\sqrt{h_n}\mu_{\tau+(k_n-1)} = N\left(h_n^{-\alpha+1/2+\beta}\Gamma(1-\alpha), 2^{2\beta-1}\Gamma(1-2\beta)\right) + o_p(1).$$

Regarding the denominator, we first investigate the products. We have

$$(\bar{D}_{i,n})^2 \leq \left(\sum_{j=1}^{k_n-1} K\left(\frac{t_{i-j}-t}{h_n}\right) D_{i-j,n}\right)^2 \leq k_n^2 K\left(\frac{t_{i-1}-t}{h_n}\right)^2 D^2_{i-1,n}$$

since the last time-point has highest drift. From this follows that

$$\frac{1}{\psi_2 k_n h_n}\sum_{i=1}^{n}(\bar{D}_{i,n})^2 \leq \frac{k_n}{\psi_2}\frac{1}{h_n}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)^2 D_{i-1,n}^2 = k_n O_p\left(\frac{1}{nh_n^{-2\alpha}}\right) = O_p\left(\frac{1}{n^{1/2}h_n^{-2\alpha}}\right) = o_p\left(\frac{1}{h_n^{1-2\alpha}}\right)$$

where the first equality follows (68) , and the last from the fact that $h_n\sqrt{n} \to \infty$.

For the volatility part in the denominator, we have

$$E\left[\frac{1}{\psi_2 k_n h_n}\sum_{i=1}^{n}(\bar{S}_{i,n})^2\right] = \frac{1}{\psi_2 k_n h_n}\sum_{i=1}^{n}\left(\sum_{j=1}^{k_n-1}g\left(\frac{j}{k_n}\right)K\left(\frac{t_{i-1}-t}{h_n}\right)^2 E\left[S_{i,n}^2\right]\right)$$

which follows from independence of $S_{i,n}$. Then from the proof without microstructure noise we have

$$\left(\frac{1}{\psi_2 k_n}\sum_{j=1}^{k_n-1}g\left(\frac{j}{k_n}\right)\right)\left(\frac{1}{h_n}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)^2 E\left[S_{i,n}^2\right]^2\right) \sim \frac{1}{h_n^{2\beta}}2^{2\beta-1}\Gamma(1-2\beta)$$

because the first part of the product converges to one. The variance converges to zero because of independence (following the same calculations as the variance of the drift estimator without microstructure noise).

We see that these results are very similar to the case without microstructure noise. The only important difference is that the kernel K is squared in the denominator, which changes the variance in the case where $1/2+\beta > 1/2$. Otherwise the results are the same in both numerator and denominator and we get the final result

$$\begin{aligned} |T_{\tau+k_n-1}^n| &\xrightarrow{p} \infty \quad \text{for } \beta + 1/2 < \alpha \\ T_{\tau+k_n-1}^n &\xrightarrow{d} N(0,1) \quad \text{for } \beta + 1/2 > \alpha \end{aligned}$$

which was exactly as we wanted.

21.9 Proof of Proposition 4

We want to investigate the convergence of

$$\sqrt{h_n}\hat{\mu}_t^n - \frac{1}{\sqrt{h_n}}\epsilon_n = \frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)(\Delta X_{i,n} + \Delta\epsilon_{i,n}) - \frac{1}{\sqrt{h_n}}\epsilon_n.$$

This can be split into the X part and noise part.

It follows directly from the convergence of the drift-estimator without noise that

$$\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta X_{i,n} \xrightarrow{d} N(0, K_2\sigma_{t-}^2).$$

For the noise part we rewrite

$$\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta\epsilon_{i,n} =$$

$$-\frac{1}{\sqrt{h_n}}K\left(\frac{t_0-t}{h_n}\right)\epsilon_0+\frac{1}{\sqrt{h_n}}K\left(\frac{t_{n-1}-t}{h_n}\right)\epsilon_n+\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}\left(K\left(\frac{t_i-t}{h_n}\right)-K\left(\frac{t_{i+1}-t}{h_n}\right)\right)\epsilon_i.$$

The first term converges in probability to zero as $\frac{1}{\sqrt{h_n}}K\left(\frac{t_0-t}{h_n}\right)\to 0$. For the third term we have directly from Lemma 3 that it converges in probability to zero. Then only the second term has a contribution. This means that

$$\frac{1}{\sqrt{h_n}}\sum_{i=1}^{n}K\left(\frac{t_{i-1}-t}{h_n}\right)\Delta\epsilon_{i,n}-\frac{1}{\sqrt{h_n}}\epsilon_n\xrightarrow{p}0$$

and from Slutsky's we then get the desired result that

$$\sqrt{h_n}\hat{\mu}_t^n-\frac{1}{\sqrt{h_n}}\epsilon_n\xrightarrow{d}N(0,K_2\sigma_{t-}^2).$$

22 APPENDIX 2 - volatility estimator without looping over lags[22]

The calculation of the volatility estimator is split into two functions. The outer function is written in R and uses the inner function written in RCPP (where the optimization over lags happens).

The R-function is

```
sigma_estimator <- function(dY_vector, h_sigma, K_func,
                            time_points,lags=10){


  t_end <- time_points[length(time_points)]
  kernels <- K_func(
              (time_points[1:(length(time_points)-1)]-t_end)/h_sigma)
  products <- kernels*dy_vector

  #Using the 'sigmas_cpp' function
  sigmas_non_scaled <- sigmas_cpp(KdY = products,lags = lags)
  sigmas <- 1/hv * sigmas_non_scaled/kernels^2
  return(sigmas)
}
```

and the Rcpp part (which is called in the third to last equation in the R function above) looks as below:

```
#include <Rcpp.h>
#include <math.h>

Rcpp::NumericVector parzen_func(int lags) {

  Rcpp::NumericVector parzen_weights(lags);
  double x;

  for (int i=0; i < lags+1; i++) {

    x = double(i)/(lags+1.0);
    if (x < 0.5) {
      parzen_weights[i] = 2.0*(1-6*pow(x,2) + 6*pow(x,3));
    }
    else {
      parzen_weights[i] = 2.0*2*pow(1-x,3);
    }
  }

  return parzen_weights;
}


// [[Rcpp::export]]
Rcpp::NumericVector sigmas_cpp(Rcpp::NumericVector KdY, int lags) {
  //Assuming parzen kernel and
  //      exp kernel and that number of lags is even

  //The runtime of this algorithm is almost independent of
  //      the number of lags and calculates sigma at EVERY point

  // size of dy should be greater than lags+4

  int n = KdY.size();

  Rcpp::NumericVector sigmas(n); //The result vector
  Rcpp::NumericVector weights(n); //The weights for the loop
  Rcpp::NumericVector products(n);

  //Calculate parzen-weights with function above
```

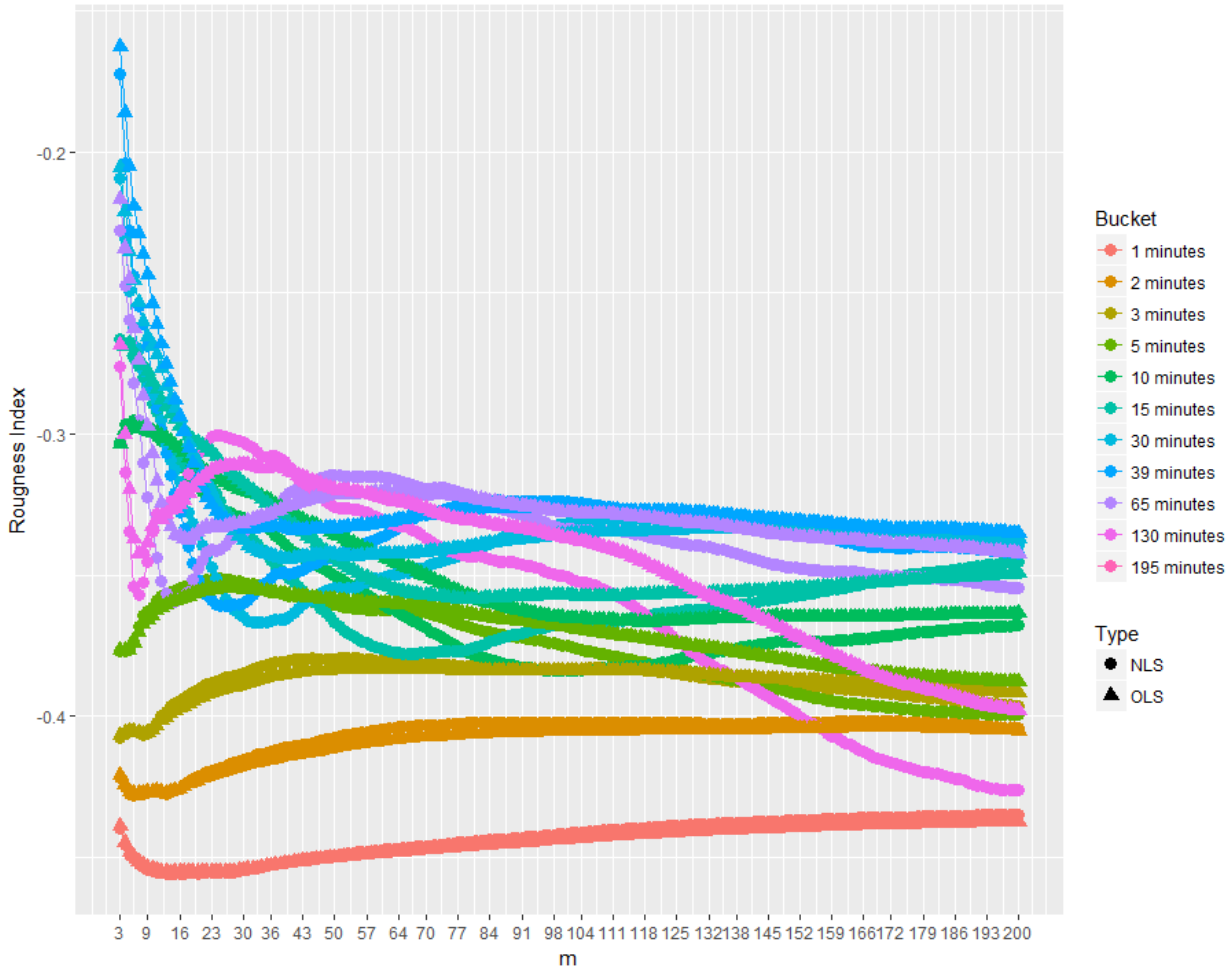
[22] Responsible: Sebastian

```
Rcpp::NumericVector parzen_weights = parzen_func(lags);

Rcpp::NumericVector first_weights(4);
Rcpp::NumericVector last_weights(4);
Rcpp::NumericVector middle_weights(4);

//First weights
first_weights[0] = parzen_weights[lags];
first_weights[1] = -4*parzen_weights[lags] +
  parzen_weights[lags-1];
first_weights[2] = 6*parzen_weights[lags] -
  4*parzen_weights[lags-1] + parzen_weights[lags-2];
first_weights[3] = -4*parzen_weights[lags] +
  6*parzen_weights[lags-1] - 4*parzen_weights[lags-2] +
  parzen_weights[lags-3];

//Middle weights - assuming L is even
//The correct weights
double lags_db = double(lags);
double temp1 = 2.0*2*pow(1-(lags_db/2.0) / (lags_db+1),3);
double temp2 = 2.0*2*pow(1-(lags_db/2.0-1.0) / (lags_db+1),3);
double temp3 = 2.0*2*pow(1-(lags_db/2.0-2.0) / (lags_db+1),3);
double temp4 = 2.0*2*pow(1-(lags_db/2.0-3.0) / (lags_db+1),3);

//Extract the wrong side, add the right
middle_weights[0] = parzen_weights[lags/2] - temp1;
middle_weights[1] = -4*parzen_weights[lags/2] +
  4*temp1+parzen_weights[lags/2-1]-temp2;
middle_weights[2] = 6*parzen_weights[lags/2]-6*temp1 -
  4*parzen_weights[lags/2-1] +
  4*temp2+parzen_weights[lags/2-2]-temp3;
middle_weights[3] = -4*parzen_weights[lags/2]+4*temp1 +
  6*parzen_weights[lags/2-1] - 6*temp2 -
  4*parzen_weights[lags/2-2] +
  4*temp3+parzen_weights[lags/2-3]-temp4;


//Last weights
last_weights[3] = 2;
last_weights[2] = parzen_weights[1] - 4*2;
last_weights[1] = parzen_weights[2] - 4*parzen_weights[1] + 6*2;
last_weights[0] = parzen_weights[3] - 4*parzen_weights[2] +
  6*parzen_weights[1] - 4*2;


// Calculate weights(L+1,L+2,L+3,L+4), and sigmas(L+1,L+2,L+3)
//First sigma (divided to avoid if statement)
for (int k=0; k<lags+1; k++) {
  //includes lag = 0 (with weight 2, so will be subtracted later)
  weights[lags] += KdY[k]*parzen_weights[lags-k];
  sigmas[lags] += KdY[k]*KdY[k];
}

long double temp;
for (int lag =1; lag < lags+1;lag++) {
  temp = 0;
  for (int k = 0; k < lags+1-lag; k++){
    temp += KdY[k]*KdY[k+lag];
  }
  sigmas[lags] += temp*parzen_weights[lag];
}


//Next 3 weights and sigmas
for (int i=1; i < 4; i++) {
  for (int k=0; k<lags+1; k++) {
    weights[lags+i] += KdY[k+i]*parzen_weights[lags-k];
  }
```

```
    sigmas[lags+i] = sigmas[lags+i-1] +
      (weights[lags+i] - KdY[lags+i])*KdY[lags+i];
  }

  //For third degree polynomial:
  //          f(x) = 4*f(x-k)-6*f(x-2k)+4*f(x-3k)-f(x-4k)

  //Using the equation above
  //Then corrects the first-, middle- and last points
  for (int i = lags+4; i < n; i++) {
    weights[i] = (4*weights[i-1] - 6*weights[i-2] +
                  4*weights[i-3]-weights[i-4]) +
      first_weights[0]*KdY[i-lags-4] +
      first_weights[1]*KdY[i-lags-3] +
      first_weights[2]*KdY[i-lags-2] +
      first_weights[3]*KdY[i-lags-1] +
      middle_weights[0]*KdY[i-lags/2-4] +
      middle_weights[1]*KdY[i-lags/2-3] +
      middle_weights[2]*KdY[i-lags/2-2] +
      middle_weights[3]*KdY[i-lags/2-1] +
      last_weights[0]*KdY[i-3] + last_weights[1]*KdY[i-2] +
      last_weights[2]*KdY[i-1] + last_weights[3]*KdY[i];

    sigmas[i] = sigmas[i-1] + (weights[i]-KdY[i])*KdY[i];
  }

  return sigmas;
}
```

23 APPENDIX 3 - volatility[23]

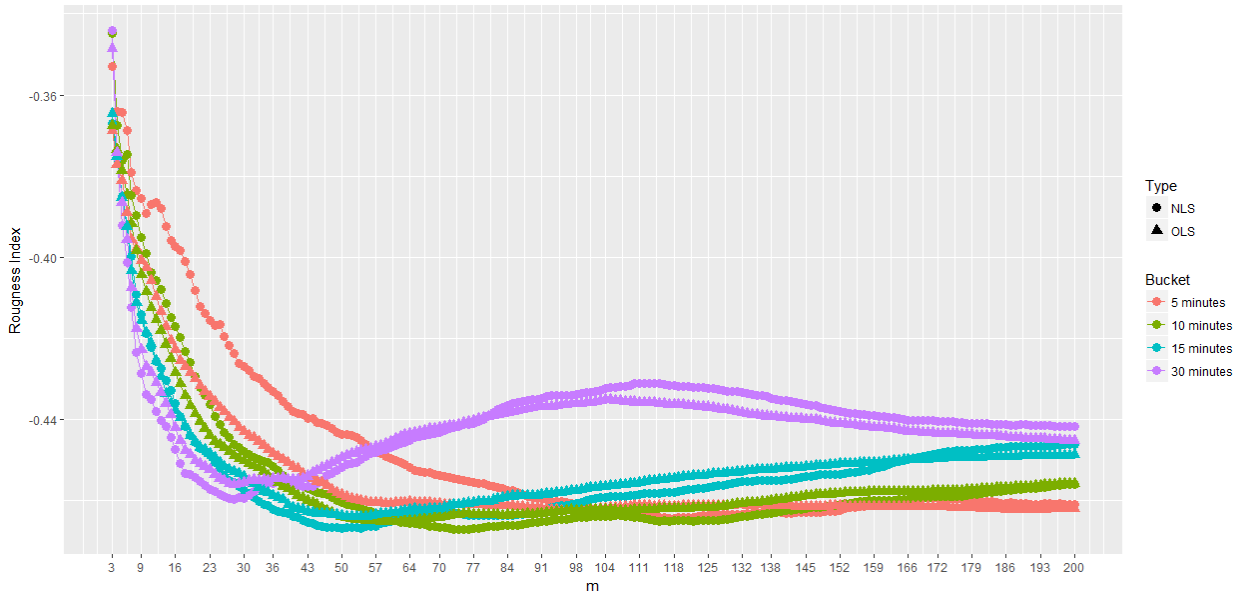
23.1 Volatility Charts



Estimated values of α, based on estimation procedures (51) and (52) for various values of m.

Figure 76: Estimates of roughness index α on SPY data

[23]Responsible: Mathias



Estimated values of α, based on estimation procedures (51) and (52) for various values of m.

Figure 77: Estimates of roughness index α on bitcoin data

References

Andersen, T. G. & Bollerslev, T. (1998), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**(2-3), 115–158.

Andersen, T. G., T. B. F. X. D. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Aït-Sahalia, Y. & Kimmel, R. (2007), 'Maximum likelihood estimation of stochastic volatility models', *Journal of Financial Economics* **83**(1), 413–452.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008), 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica* **76**(6), 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realised kernels in practice: Trades and quotes', *Econometrics Journal* **12**(3), 1–33.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2011), 'Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

Barndorff-Nielsen, O. E. & Schmiegel, J. (2009), Brownian semistationary processes and volatility/ intermittency, *in* 'In Advanced financial modelling', Vol. 8 of *Radon Series on Computational and Applied Mathematics*, Berlin: Walter de Gruyter, pp. 1–25.

Bennedsen, M., Lunde, A. & Pakkanen, M. S. (2015), 'Hybrid scheme for Brownian semistationary processes', *ArXiv e-prints* .

Bennedsen, M., Lunde, A. & Pakkanen, M. S. (2016), 'Decoupling the short- and long-term behavior of stochastic volatility', *ArXiv e-prints* .

Björk, T. (2009), *Arbitrage Theory in Continuous Time*, 3 edn, Oxford University Press.

Christensen, K., Kinnebrock, S. & Podolskij, M. (2010), 'Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data', *Journal of Econometrics* **159**(1), 116–133.

Christensen, K., Oomen, R. & Podolskij, M. (2012), 'Fact or friction: Jumps at ultra high frequency'.

Christensen, K., Oomen, R. & Renò, R. (2016), 'The drift burst hypothesis'.

Christian Bayer, P. F. . J. G. (2016), 'Pricing under rough volatility', *Quantitative Finance* **16**(6), 887–904.

Comte, F. & Renault, E. (1996), 'Long memory continuous time models', *Journal of Econometrics* **73**(1), 101–149.

Gatheral, J., Jaisson, T. & Rosenbaum, M. (2014), 'Volatility is rough', *ArXiv e-prints* .

Glasserman, P. (2004), *Monte Carlo methods in financial engineering*, 1 edn, Springer.

Gradshteyn, I. S. & Ryzhik, I. M. (2007), *Table of integrals, series, and products*, 7 edn, Academic Press.

Grenander, U. (1950), 'Stochastic processes and statistical inference', *Arkiv För Matematik* **1**, 195–275.

Hansen, E. (2009), *Measure theory*, 4 edn, University of Copenhagen.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**(2), 127–161.

Jacod, J., Lib, Y., Myklandb, P. A., Podolskijc, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: The pre-averaging approach', *Stochastic Processes and their Applications* **119**(7), 2249–2276.

Marinucci, D. & Robinson, P. M. (1999), 'Alternative forms of fractional brownian motion', *Journal of Statistical Planning and Inference* **80**(1-2), 111–122.

Mattiussi, V. & Renò, R. (2009), 'Spot volatility estimation using delta sequences'.

Mikosch, T. (n.d.), 'Lecture notes for stat Øk 2'.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Newey, W. K. & West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies* **61**, 631–653.

Sokol, A. & Rønn-Nielsen, A. (2015), *Advanced Probability*, 3 edn, University of Copenhagen.

Øksendal, B. (2000), *Stochastic Differential Equations*, 5 edn, Springer-Verlag.