

SIMULTANEOUS CONFIDENCE STATEMENTS  
ABOUT THE DIFFUSION COEFFICIENT OF AN ITÔ-PROCESS  
WITH APPLICATION TO SPOT VOLATILITY ESTIMATION

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades  
Doctor rerum naturalium (Dr. rer. nat.)  
der Georg-August-Universität Göttingen

im Promotionsprogramm

PhD School of Mathematical Sciences (SMS)  
der Georg-August University School of Science (GAUSS)

vorgelegt von

Till Sabel  
aus Oldenburg (Oldb)

Göttingen, Juni 2014

---

## BETREUUNGSAUSSCHUSS

Prof. Dr. Axel Munk, Institut für mathematische Stochastik, Georg-August-Universität Göttingen

Prof. Dr. Lutz Dümbgen, Institut für mathematische Statistik und Versicherungslehre, Universität Bern

## MITGLIEDER DER PRÜFUNGSKOMMISSION

### REFERENT:

Prof. Dr. Axel Munk, Institut für mathematische Stochastik, Georg-August-Universität Göttingen

### KORREFERENT:

Prof. Dr. Lutz Dümbgen, Institut für mathematische Statistik und Versicherungslehre, Universität Bern

### WEITERE MITGLIEDER:

Prof. Dr. Dorothea Bahns, Mathematisches Institut, Georg-August-Universität Göttingen

JProf. Dr. Andrea Krajina, Institut für mathematische Stochastik, Georg-August-Universität Göttingen

Prof. Dr. Gerlind Plonka-Hoch, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Prof. Dr. Dominic Schuhmacher, Institut für mathematische Stochastik, Georg-August-Universität Göttingen

TAG DER MÜNDLICHEN PRÜFUNG: 16.07.2014

---

---

---

---

---

---

## ACKNOWLEDGMENTS

An dieser Stelle möchte ich all denen danken, die mich auf meinem bisherigen Lebensweg unterstützt haben und meine Promotion überhaupt erst ermöglicht haben:

Zunächst bedanke ich mich herzlich bei meinem Betreuer Axel Munk, der mich immer bestmöglich unterstützt und gefördert hat. Sein Enthusiasmus und seine hervorragende Intuition für mathematische Statistik und insbesondere für das interessante Thema Volatilitätsschätzen, das mich auf die eine oder andere Weise seit meiner Bachelorarbeit begleitet hat, färbte stets auf mich ab und war eine große Motivation für mich. Besonders angenehm fand und finde ich, dass ich jederzeit mit Problemen zu ihm kommen konnte und er sich immer die Zeit genommen hat, einen Lösungsweg zu finden.

Desweiteren gilt mein großer Dank Lutz Dümbgen, von dem ich im Rahmen von zwei wunderbaren Besuchen in Bern und mehreren Diskussionen am Rande von Konferenzen sehr viel im Bereich der Multiskalenmethoden gelernt habe. Es freut mich sehr, dass er sich bereit erklärt hat, mich als Zweitbetreuer durch die Promotion zu begleiten.

Ein besonderer Dank gilt Johannes Schmidt-Hieber. Unter anderem während zahlreicher sehr schöner (und sehr arbeitsreicher) Besuche in Amsterdam, Paris und Leiden hatte ich das große Vergnügen mit ihm zusammenzuarbeiten und insbesondere viel von ihm zu lernen. Er ist seit Jahren ein Mentor für mich. Es hat mir sehr viel Spaß gemacht, Johannes!

I am grateful to all my colleagues at the IMS. The old-fashioned coffee breaks and the new tradition of playing table soccer have always been fun. Mein spezieller Dank geht hierbei an diejenigen, die es mit mir im Büro ausgehalten haben: Rebekka, Hannes, Ina, Johannes und Philipp, vielen Dank für die nette Atmosphäre, es war immer eine Freude als Erster ins Büro zu kommen!

Außerdem möchte ich allen meinen Freunden in Göttingen danken, die diese Stadt in den vergangenen acht Jahren zu einem Zuhause für mich gemacht haben. Ein ganz spezieller Dank gilt Rebecca, meinen Eltern und Jan, die mich zu jeder Zeit mit vollem Einsatz unterstützt haben, und ohne die weder mein Studium noch meine Promotion möglich gewesen wären.

---

---

---

## SUMMARY

In this PhD thesis, we address the problem of giving simultaneous confidence statements about local features of the diffusion of an Itô process. To this end, we construct a multiscale test based on weighted quadratic variation and prove that the test statistic can be strongly approximated by a sequence of Gaussian martingales which are distribution-free. Further, we give optimality results and present different visualization methods.

In the second part of the thesis, we extend the approach to data corrupted by additive noise to cover applications from high-frequency finance. Additionally, we show which difficulties arise from real data and apply our method exemplarily to prices of Euro-Bund-Futures (FGBL).

As an outlook for future work, we present ideas of generalizing the method to inference on the local covariance and point out some interesting applications from finance.

---

---



---

# CONTENTS

---

1	INTRODUCTION	1
1.1	Some History of Volatility Estimation . . . . .	1
1.2	Confidence Statements about the Diffusion . . . . .	3
1.3	Main Results of this Thesis . . . . .	4
1.3.1	The Low-Frequency Setting . . . . .	4
1.3.2	The High-Frequency Setting . . . . .	6
2	METHODOLOGY	9
2.1	Some Preliminaries from Martingale Theory . . . . .	9
2.2	Further Definitions and Notation . . . . .	11
3	THEORY	13
3.1	Modeling . . . . .	13
3.2	Assumptions . . . . .	13
3.3	Results . . . . .	15
3.4	Choice of the Spot Volatility Estimator . . . . .	16
4	INFERENCE ON SPOT VOLATILITY	19
4.1	Application of Theorem 3.4 . . . . .	19
4.2	Detection Rates . . . . .	22
4.3	Visualization and First Data Examples . . . . .	26

## CONTENTS

---

5	EXTENSION TO HIGH-FREQUENCY DATA	33
5.1	Motivation . . . . .	33
5.2	Modeling . . . . .	33
5.3	Results . . . . .	35
6	APPLICATION TO FINANCIAL DATA	39
6.1	Model Discretization . . . . .	39
6.2	Parameter Optimization . . . . .	39
6.3	Difficulties in Practice . . . . .	42
6.3.1	Model Violations . . . . .	43
6.3.2	Time Schemes . . . . .	47
6.4	Real Data Example . . . . .	49
7	OUTLOOK	55
7.1	Multidimensional Extension . . . . .	55
7.2	Investigation of the Leverage Effect . . . . .	56
	BIBLIOGRAPHY	57
	APPENDICES	65
A	A PROOF OF THEOREM 3.4	65
B	A PROOF OF THEOREM 5.5	77
C	TECHNICAL AND AUXILIARY RESULTS	83
	CURRICULUM VITAE	XI

---

# CHAPTER 1

## INTRODUCTION

---

### 1.1 SOME HISTORY OF VOLATILITY ESTIMATION

In many fields of science, data can be approximated by a heterogeneous random walk. Among these are examples from physics, biology, and maybe most prominently, mathematical finance. Here, some arbitrage-free log-price  $X$  is often modeled as an Itô process (cf. for example [Delbaen and Schachermayer \(1994, 1998\)](#)), that is

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad t \in [0, 1], \quad (1.1)$$

where  $W$  is a Brownian motion, the drift  $b$  and the diffusion  $\sigma^2$  (referred to as “spot volatility“ in the financial literature and playing the role of a local variance) are predictable and integrable processes. For identifiability reasons, we restrict  $\sigma$  to be non-negative.

Both in financial application and in theory, the diffusion  $\sigma^2$  and surrogates of it (such as integrated volatility  $\int \sigma_s^2 ds$ ) are of interest for modeling and pricing (cf. for example [Black and Scholes \(1973\)](#), or more recently [Heston \(1993\)](#)). Furthermore, looking on small scales only (for example in intraday trading), the martingale part of (1.1) dominates the comparably smooth integrated drift part, such that the latter is asymptotically negligible. Typically, integrated volatility is considered since it is well-known that it is the limit of realized quadratic variation, that is

$$\sum_{i=1}^n (X_{i/n} - X_{(i-1)/n})^2,$$

when  $n$  tends to infinity. However, pathwise estimation of  $\sigma^2$  is more challenging. An approach based on realized quadratic variation is given in [Fan and Wang \(2008\)](#).

Due to technical progress however, the frequency of observing data increased more and more during the past decades. As it turned out in financial practice, Model (1.1) was

no longer appropriate to describe these high-frequency observations, since there are so-called microstructure effects occurring due to market frictions such as bid-ask-spreads, discreteness of prices, or trading costs. Overviews of these frictions are given in [Hasbrouck \(1993\)](#) and [Madhavan \(2000\)](#). In the literature, these effects are often modeled as additive noise, resulting in an observed log-price

$$Y = X + \epsilon \tag{1.2}$$

as the sum of the latent log-price  $X$  and some noise process  $\epsilon$  (cf. for example [Zhou \(1996\)](#)). In these models, the well-established quadratic variation methods fail if the noise process is rougher than the semimartingale part. Especially, realized quadratic variation does not converge to integrated volatility any longer but tends to infinity (cf. [Bandi and Russell \(2008\)](#)). This fact is also displayed in Figure 1.1, which is also known as a signature plot: Sub-sampling the data at frequency  $k$  ( $x$ -axis) and computing realized quadratic variation of each sub-sample ( $y$ -axis) results in an tremendous increase if  $k$  gets small. This singularity cannot be explained by any continuous semimartingale model.

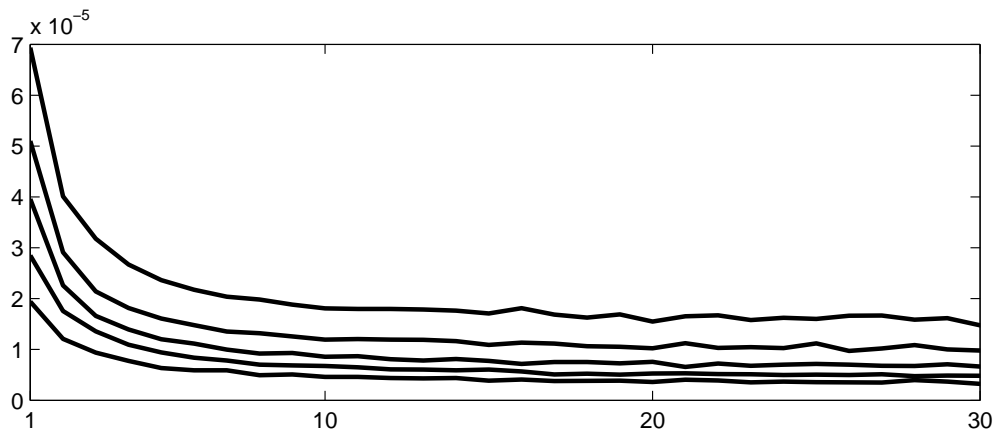


FIGURE 1.1: Realized volatilities of FGBL data from June 4th, 2007 to June 8th, 2007 for different sub-sampling frequencies.<sup>1</sup>

At this point, two different strands can be found in the literature: On the one hand, [Gloter and Jacod \(2001a,b\)](#) targeted the problem of estimating volatility in a parametric family. Proving a local asymptotic normality (LAN) property, they found the asymptotic minimax lower bound  $8\tau\sigma^3n^{-1/2}$  for the mean squared error (MSE). An easier proof of this sharp bound is given in [Cai et al. \(2010\)](#). Further, authors have investigated asymptotic minimax properties in a parametric version of Model (1.2), where the driving Brownian motion  $W$  in

(1.1) is replaced by more general Gaussian processes, including fractional Brownian motion with Hurst index  $H \in (0, 1)$ . Optimal rates for the fractional Brownian motion case are given in [Gloter and Hoffmann \(2004\)](#), optimal constants follow from the results in [Sabel and Schmidt-Hieber \(2014a\)](#).

On the other hand, researchers targeted nonparametric estimation of the volatility and its surrogates. Starting with the seminal work of [Ait-Sahalia et al. \(2005\)](#), many sophisticated regularization methods for estimation of integrated volatility have been proposed in the literature. Among those are the papers of [Zhang \(2006\)](#), [Barndorff-Nielsen et al. \(2008\)](#), and the pre-averaging approach of [Podolskij and Vetter \(2009\)](#) and [Jacod et al. \(2009\)](#). A huge advance in understanding volatility was achieved by [Reiß \(2011\)](#), who proved that under quite strong assumptions, the microstructure noise model is asymptotically equivalent in Le Cam's sense to a regression setting. At the same time, a first approach to spot volatility estimation, that is nonparametric pathwise reconstruction of  $\sigma^2$ , was introduced by [Munk and Schmidt-Hieber \(2010\)](#), where the authors proposed a rate-optimal Fourier series estimator. However, this estimator lacks in adapting to the unknown smoothness. This problem was solved in [Hoffmann et al. \(2012\)](#) by proposing an estimator based on a wavelet decomposition which is adaptive and rate-optimal over Besov classes. In [Sabel et al. \(2014\)](#), this estimator is further adjusted and fine tuned to overcome difficulties occurring in practice. A Matlab based implementation can be found in the Spotvol Toolbox by [Sabel and Schmidt-Hieber \(2014b\)](#).

## 1.2 CONFIDENCE STATEMENTS ABOUT THE DIFFUSION

Despite of the large scientific and practical impact of all the methods mentioned above, practitioners are often interested in confidence statements about the diffusion coefficient. Surprisingly, there are very few references dealing with this topic in either the pure semimartingale or the high-frequency model. [Hoffmann et al. \(2012\)](#) provide a thresholding rule (given in Theorem 3.3 in their work) for each wavelet coefficient  $\int \psi_{j,k} \sigma^2$ , which can be viewed as a confidence interval for the respective coefficient. Nevertheless, if we consider a more general set of test functions  $\psi_{t,h}$  with support in  $[t, t+h]$  for some non-dyadic  $t$  and  $h$ , it is not clear at all if their statement holds simultaneously over all combinations  $(t, h)$ . A different approach was chosen in [Spokoiny \(2009\)](#): Here, it is assumed that volatility is locally constant, allowing the author to rewrite the problem as a change point problem.

Under this assumption, different results including confidence regions and forecasting methods are presented. However, this assumption seems to be quite restrictive compared to the semimartingale model. [Jacod et al. \(2009\)](#) proved asymptotic normality of their estimator of integrated volatility allowing the construction of asymptotic confidence intervals. However, this is only proved for some fixed time interval  $[0, T]$  and therefore lacks in uniformity and localization, that is information about local features of the spot volatility function.

## 1.3 MAIN RESULTS OF THIS THESIS

The main body of the thesis is split into two parts: In the first half, we develop a method to obtain uniform confidence statements about the diffusion in a low-frequency setting, that is directly observing  $X$  on a discrete grid. The second half provides an extension to high-frequency observations and an application to real financial data.

### 1.3.1 THE LOW-FREQUENCY SETTING

Firstly, we consider data from Model [\(1.1\)](#) observed at discrete time points  $i/n, i = 0 \dots, n$ . We concentrate on simultaneous confidence statements for integrals of the type

$$h^{-1/2} \int_t^{t+h} \psi\left(\frac{s-t}{h}\right) \sigma_s^2 ds \tag{1.3}$$

for some test function  $\psi$  with support in  $[0, 1]$  and simultaneously for  $(t, h)$  belonging to some subset of  $[0, 1]^2$ . Our approach has to be regarded as a compromise between inference on spot volatility and integrated volatility. Although the confidence statements do not give us adaptive confidence bands for the function  $s \mapsto \sigma_s^2$  (which is an impossible task at least in nonparametric regression, cf. [Low \(1997\)](#)) but only for weighted averages of  $\sigma^2$  on some interval  $[t, t+h]$ , they contain information about very local features of  $\sigma^2$  if  $h$  becomes small.

Since approaches like this consider multiple “scales” defined by the location  $t$  and the bandwidth  $h$ , they are often termed “multiscale methods” in the literature. Similar approaches focusing on nonparametric regression and density estimation were developed in [Dümbgen and Spokoiny \(2001\)](#), [Dümbgen and Walther \(2008\)](#), and [Schmidt-Hieber et al. \(2013\)](#).

However, there has been no comparable techniques dealing with inference on volatility even for non-high-frequent data so far.

To get confidence intervals for (1.3), we choose a weighted quadratic variation approach: Let  $\mathcal{T}_n \subset [0, 1]^2$  be a set of scales  $(t, h)$  (again using the notation  $t$  for the location and  $h$  for the bandwidth), so that  $[t, t + h] \subset [0, 1]$  and  $l_n < h < u_n$  for some minimal and maximal bandwidths  $l_n$  and  $u_n$ , both tending to zero. Observing discrete data  $X_{i,n}$  from Model (1.1), we consider the test statistic

$$T_{n,t,h}^{(1)} := h^{-1/2} \sum_{i=0}^{n-1} \psi\left(\frac{i-t}{h}\right) (X_{i+1,n} - X_{i,n})^2.$$

We prove that under mild assumptions on  $\psi$ ,  $\sigma^2$ ,  $l_n$ , and  $u_n$ , and for some estimator  $\hat{\sigma}_{t,h}^2$  of  $\sigma_t^2$  which is consistent at a log-rate, there exists a sequence of Brownian motions  $W^{[n]}$ , s.t.

$$\sup_{(t,h) \in \mathcal{T}_n} w_h \left| \frac{T_{n,t,h}^{(1)} - h^{-1/2} \int_t^{t+h} \psi\left(\frac{s-t}{h}\right) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \sqrt{\frac{2}{nh}} \int_t^{t+h} \psi\left(\frac{s-t}{h}\right) dW_s^{[n]} \right| = o(n^{-1/2}) \text{ a.s.}$$

Thus, when approximating (1.3) by the test statistic  $T_{n,t,h}^{(1)}$ , the rescaled approximation error

$$\frac{T_{n,t,h}^{(1)} - h^{-1/2} \int_t^{t+h} \psi\left(\frac{s-t}{h}\right) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2}$$

is uniformly small and does not depend on  $\sigma$  or  $X$  but is a Gaussian process with known distribution (for more details cf. Theorem 3.4).

This construction allows us to simulate quantiles of the supremum of the Gaussian process over  $\mathcal{T}_n$  to obtain quantiles for the supremum of the approximation error. Afterwards, these can be used to construct asymptotic confidence intervals for (1.3) (cf. Proposition 4.1). Finally, we show that the lengths of these intervals are rate-optimal in a certain sense.

More thorough explanations and rigorous results can be found in Chapter 3. The construction of confidence intervals, a discussion of optimality, as well as aspects concerning application such as the choice of the test function  $\psi$  and visualization methods are given in Chapter 4.

### 1.3.2 THE HIGH-FREQUENCY SETTING

As mentioned previously, modern financial data is often sampled at very high-frequencies, so that there are two options: Either one sub-samples the data (resulting in a loss of information, cf. [Ait-Sahalia et al. \(2005\)](#)), or one finds new methods suited to the microstructure noise model given in (1.2). For our purpose, it turns out that we can construct a high-frequency analog of our multiscale test using the pre-averaging technique introduced in [Jacod et al. \(2009\)](#) and refined in [Hoffmann et al. \(2012\)](#):

In a first step, we compute local averages of the noisy data in (1.2). This reduces the effect of the noise term by some argument similar to the law of large numbers, while the continuous martingale term is not affected (up to some small bias). However, the data size is reduced by this method. It turns out that averaging over intervals of length of order  $n^{-1/2}$  balances the negative effects of the microstructure noise and the data size reduction, which corresponds to previous results in the literature. After the pre-averaging procedure, the multiscale approach developed for the low-frequency setting can be transferred almost directly up to some technicalities (cf. Theorem 5.5). More details can be found in Chapter 5.

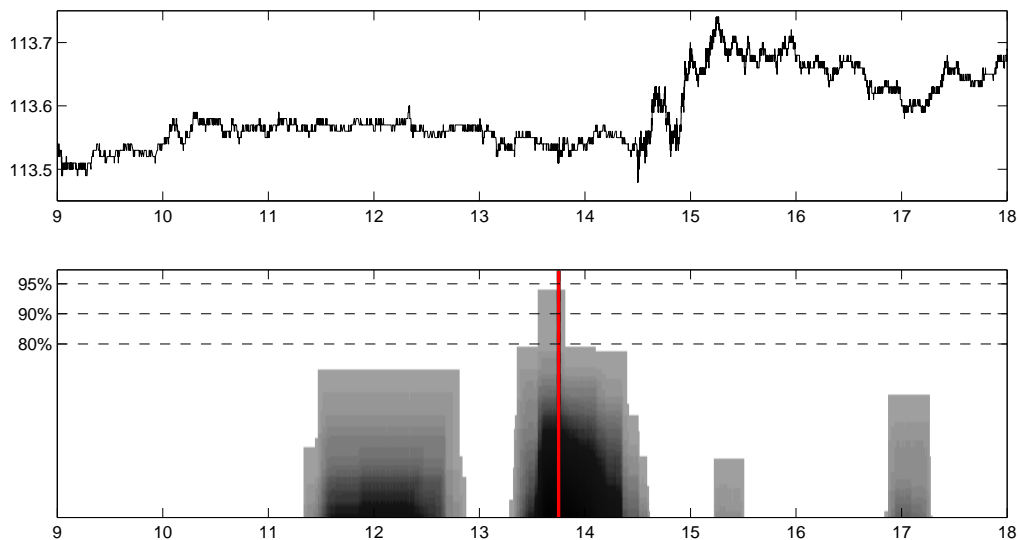


FIGURE 1.2: FGBL price of May 10th, 2007 (panel 1), and areas of significant increase ( $x$ -axis of panel 2) for different levels of significance ( $y$ -axis). The vertical red line at 13.75 refers to the announcement of not changing the key interest rate. Every interval of increase is indicated gray and darker regions only refer to intersections of these intervals.



Chapter 6 is devoted to the application of the method in practice. Since there are different model violations such as jumps in the data or non-equidistant time schemes, we present different solutions to overcome these difficulties. Afterwards, we turn towards data analysis in the last part of the chapter, where we exemplarily investigate the volatility of Euro-Bund-Futures (FGBL) with our method. Here, we find a significant increase of volatility during some of the monthly press conferences of the president of the European Central Bank, where changes of the key interest rate are announced. The results for one of these days (May 10th, 2007) is displayed in Figure 1.2, where regions of significant increase are displayed for different levels of significance. Here, we observe a significant increase (with significance level clearly above 90%) at the time of the announcement (1.45 p.m., indicated by the red vertical line) as well as some less significant increase all over the day. More thoroughly, we investigate at which days in 2007 the spot volatility at 1.45 p.m. exceeds the daily average, that is the integrated volatility, significantly and find that this effect appears more often on days with announcements than on regular trading days.

In Chapter 7, we give an outlook to future work: We motivate an extension to multidimensional volatility estimation (so-called covolatility estimation), which seems to be surprisingly simple. Furthermore, we present an interesting application dealing with testing of the presence of the leverage effect in financial data.

Most of the proofs and further technicalities are postponed to the Appendices A, B, and C.



---

## CHAPTER 2

# METHODOLOGY

---

In this chapter, we will give a short introduction to the theory of martingales and quadratic variation to provide some tools which are useful later on. Furthermore, we will introduce some notation we will frequently use.

### 2.1 SOME PRELIMINARIES FROM MARTINGALE THEORY

Throughout this thesis, we consider some Itô process  $X = (X_t)_{t \in [0,1]}$ , that is  $X$  is a semi-martingale with representation

$$X = (X_0 + \int_0^t b_t dt + \int_0^t \sigma_s dW_s)_{t \in [0,1]},$$

where  $W$  is a Brownian motion and  $b$  and  $\sigma^2$  are predictable and almost surely integrable processes. It will turn out that we can restrict ourselves to the case  $X_0 = b_t = 0$  almost surely for all  $t \in [0, 1]$ , that is  $X$  is the continuous martingale

$$\left(\int_0^t \sigma_s dW_s\right)_{t \in [0,1]}.$$

Here, the predictable quadratic variation  $\langle X \rangle$ , defined as the unique predictable process, such that  $X^2 - \langle X \rangle$  is a martingale (cf. [Jacod and Shiryaev \(2003\)](#), Chapter 1, Theorem 4.2), is of particular interest for practical purposes (cf. [Jacod and Protter \(2011\)](#), p.92). Let us collect some facts about the predictable quadratic variation:

**PROPOSITION 2.1.** *Let  $X$  be a martingale with representation  $X = (\int_0^t \sigma_s dW_s)_{t \in [0,1]}$  for a Brownian motion  $W$  and some predictable, positive, and square-integrable process  $\sigma$ . Then, we obtain the following statements:*

1. *The process  $\langle X \rangle$  is given by  $(\int_0^t \sigma_s^2 ds)_{t \in [0,1]}$  (cf. for example [Jacod and Shiryaev \(2003\)](#), Proposition 4.10).*

2. Let  $T \in (0, 1]$ . For any adapted partition  $\pi = \{0 = t_0 < \dots < t_n = T\}$  with mesh tending to zero (i.e.  $\inf_{1 \leq i \leq n} t_i - t_{i-1} \rightarrow 0$ ), the sum

$$\sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2 \tag{2.1}$$

tends to a limit  $[X]_T$  in probability uniformly in  $T$ . This limit process is called quadratic variation and coincides with the predictable quadratic variation for continuous martingales (cf. [Jacod and Shiryaev \(2003\)](#), Theorems 4.47 and 4.52). For that reason, we will use the term “quadratic variation” for both processes synonymously.

3. In Theorem 3.4 of this thesis, we will prove that under certain assumptions on  $\sigma^2$  and  $\pi$ , the uniform convergence of (2.1) is almost sure. Moreover, we see that for fixed  $T > 0$ , the rate of convergence is  $\sqrt{n}$ , if  $t_i = \frac{i}{n}$ .

4. By Itô’s formula (cf. for example [Jacod and Shiryaev \(2003\)](#), Theorem 4.57), we find the explicit representation

$$X_t^2 - \langle X \rangle_t = 2 \int_0^t \sigma_s \int_0^s \sigma_u dW_u dW_s.$$

Especially the second part of the previous proposition gives us an idea of how to construct estimators for  $\int_a^b g(s) \sigma_s^2 ds$ , for some constants  $a$  and  $b$ , and a real-valued (piecewise smooth) function  $g$ , when we observe  $X$  at time points  $i/n$ ,  $i = 0, \dots, n$ : If  $g$  is piecewise constant, the proposition yields that the estimator

$$\sum_{i=0}^{n-1} g\left(\frac{i}{n}\right) (X_{\frac{i+1}{n}} - X_{\frac{i}{n}})^2 \tag{2.2}$$

is consistent. If  $g$  is (piecewise) sufficiently smooth (for example, if it has finite total variation), this holds as well, as one can easily check with some approximation arguments. In the course of this thesis, it will turn out that the estimator in 2.2 is rate-optimal in a certain sense and that it performs well simultaneous over some class of smooth functions  $g$ .

In the literature on martingale theory, there are various articles concerning probabilistic bounds on martingales, the so-called martingale inequalities. At this place, we like to state the Burkholder-Davis-Gundy inequality, which is probably the most prominent one and will be extensively used in our proofs. A first version was proved in [Burkholder \(1966\)](#).

However, we will use the refined one given in [Barlow and Yor \(1982\)](#), Proposition 4.2:

PROPOSITION 2.2. *Let  $M$  be a continuous martingale with  $M_0 = 0$ . Then, for any  $k \geq 2$ , there exists a constant  $C_B$  which is independent of  $M$  and  $k$ , s.t.*

$$\mathbb{E} \left( \sup_{0 \leq s \leq t} M_t \right)^k \leq (C_B k^{1/2})^k \mathbb{E} \langle M \rangle^{k/2}.$$

## 2.2 FURTHER DEFINITIONS AND NOTATION

In the following, we consider Hölder classes as parameter spaces. These are defined and denoted as follows:

DEFINITION 2.3. *For some  $K + 1 \geq \gamma > K$ ,  $K \in \mathbb{N}$ ,  $L > 0$ ,  $T \subset \mathbb{R}$ , the Hölder class  $\mathcal{H}(\gamma, L, T)$  is defined as the class of all functions  $f : T \rightarrow \mathbb{R}$  which are  $K$ -times continuously differentiable and satisfy*

$$|f^{(K)}(x) - f^{(K)}(x')| \leq L|x - x'|^{\gamma-K}$$

for any points  $x, x' \in T$ . For convenience, we will write  $\mathcal{H}(\gamma, L) = \mathcal{H}(\gamma, L, [0, 1])$ .

For  $\gamma = 1$ ,  $\mathcal{H}(\gamma, L)$  is known as the class of Lipschitz continuous functions. A prominent example for non-Lipschitz but Hölder continuous functions are the paths of Brownian motion or more generally, of Itô processes. For any realization, there exists an  $L$ , so that it is in any class  $\mathcal{H}(\gamma, L)$  for  $\gamma < 1/2$  (cf. for example [Steele \(2001\)](#)). In most examples from finance,  $\sigma^2$  is an Itô process as well, so that we should keep in mind that from a practical point of view, we focus on the case  $\gamma < 1$ .

A classical result from nonparametric statistics is the following minimax bound for nonparametric estimation of a function  $f$  in a Hölder space. We will use it as a benchmark to show rate-optimality of our method.

PROPOSITION 2.4 (cf. for example [Tsybakov \(2009\)](#)). *Consider nonparametric estimation in a Hölder space  $\mathcal{H}(\gamma, L)$  for an equidistant design with Gaussian errors, that is observing  $Z_i = f(\frac{i}{n}) + \epsilon_i$ ,  $i = 1, \dots, n$ , where  $(\epsilon_i)$  are i.i.d. centered Gaussian random variables and*

$f \in \mathcal{H}(\gamma, L)$ . Then, the optimal rate of convergence is given by:

$$n^{-\frac{\gamma}{2\gamma+1}}, \text{ if we consider } L^2\text{-loss, and}$$

$$\left(\frac{\log n}{n}\right)^{-\frac{\gamma}{2\gamma+1}}, \text{ if we consider } L^\infty\text{-loss.}$$

Throughout this thesis, we will make use of the following notation, describing the asymptotic order of two sequences:

DEFINITION 2.5. Let  $(a_n)$  and  $(b_n)$  be two real-valued deterministic sequences. Then, we use the notation

- $a_n = o(b_n)$  or  $a_n \ll b_n$ , if and only if  $a_n/b_n \rightarrow 0$ ,
- $a_n = O(b_n)$  or  $a_n \lesssim b_n$ , if and only if there exists a real number  $C > 0$  with  $-C < a_n/b_n < C$  for all  $n$ ,
- $a_n \asymp b_n$ , if and only if  $a_n = O(b_n)$  and  $b_n = O(a_n)$ .

If  $(a_n)$  or  $(b_n)$  are random, we either write the adjunct "a.s." if the statement holds almost surely, or use the symbols  $O_p$  and  $o_p$  to indicate that the statement holds in probability. Note that in both cases the real number  $C$  is replaced by a real-valued random variable  $C$  which does not depend on  $n$ .

---

# CHAPTER 3

## THEORY

---

### 3.1 MODELING

We consider discrete observations coming from Model (1.1), sampled at time points  $\frac{i}{n}$ ,  $i = 0, \dots, n$ , that is we observe

$$X_{i,n} = \int_0^{\frac{i}{n}} \sigma_s dW_s, \quad i = 0, \dots, n. \quad (3.1)$$

Without loss of generality, we assume that  $X_0 = 0$  and  $b_s = 0$  for all  $s$ , since we will consider increments of the observations. As already mentioned in the introduction, the increments of the drift part  $\int b_s ds$  are of negligible small order compared to those of the martingale part  $\int \sigma_s dW_s$ . Furthermore, it is no restriction to consider  $\sigma > 0$  only, since  $\int_0^t \sigma_s dW_s$  and  $\int_0^t |\sigma_s| dW_s$  are identically distributed. Therefore, we cannot distinguish between  $\sigma$  and  $-\sigma$ . A more detailed discussion of these assumptions is given in [Jacod and Protter \(2011\)](#).

In the following chapters, we will refer to this setting as the “low-frequency” or the “pure-semimartingale” setting, in contrast to the “high-frequency” or “microstructure noise” model, which we will introduce later on (cf. Chapter 5).

### 3.2 ASSUMPTIONS

We make the following assumptions on the volatility  $\sigma^2$ :

**ASSUMPTION 3.1** (Assumptions on  $\sigma^2$ ). *Assume that  $\sigma^2$  is a stochastic process which is adapted to the natural filtration of  $W$  and which is Hölder continuous with index  $\gamma \in (0, 1]$  and Hölder constant  $L_{\sigma^2}$ , that is  $\sigma^2 \in \mathcal{H}(\gamma, L_{\sigma^2})$ . Assume further that  $\sigma^2$  is bounded from*

below and above, that is, there exist  $0 < \underline{\sigma} < \bar{\sigma} < \infty$  with

$$\underline{\sigma}^2 < \inf_{0 \leq t \leq 1} \sigma_t^2 \leq \sup_{0 \leq t \leq 1} \sigma_t^2 < \bar{\sigma}^2.$$

REMARK 3.2. Note that the Hölder continuity assumption is easily met in practice (cf. for example [Heston \(1993\)](#)). Here, we restrict ourselves to  $L_{\sigma^2}$  being deterministic but arbitrary large similarly to the restrictions chosen in [Hoffmann et al. \(2012\)](#).

For  $\gamma > 1$ , all the results are still true, since then,  $\mathcal{H}(\gamma, L_{\sigma^2}) \subset \mathcal{H}(1, L')$  for some  $L'$ . However, we will only use Lipschitz continuity in that case. Thus, there is no efficiency gain from further restricting the parameter space.

Moreover, note the upper bound is a direct consequence of Hölder continuity, while the lower bounds exists as long as there are no periods without market activity. Admittedly, this is the minimal requirement for obtaining economic reasonable statements, and is indeed unavoidable for statistical inference.

In this thesis, we like to give qualitative statements about  $\int_t^{t+h} \psi(\frac{\bullet-t}{h}) \sigma^2$  for some test function  $\psi$  and some  $(t, h) \in [0, 1]^2$ . To obtain these statements, let us introduce a test based on a multiscale approach. To this end, let  $\psi$  be some function with support  $[0, 1]$  and positive and finite  $L^2$  norm. We consider the family of functions

$$\{\psi_{t,h} = \frac{\psi(\frac{\bullet-t}{h})}{\|\psi(\frac{\bullet-t}{h})\|_{L^2_{[0,1]}}, (t, h) \in \mathcal{T}_n\}, \quad (3.2)$$

where  $\|\bullet\|_{L^2_{[0,1]}}$  denotes the  $L^2$  function norm on  $[0, 1]$  and  $\mathcal{T}_n \subset [0, 1]^2$ . Here,  $t$  refers to the location of the considered interval while  $h$  plays the role of a bandwidth.

We require the following assumptions on  $\mathcal{T}_n$  and  $\psi$ :

ASSUMPTION 3.3 (Assumptions on the set of test functions  $\psi_{t,h}$ ). *We assume that  $\psi$  is a test function with  $\text{supp } \psi \subseteq [0, 1]$  and  $0 < \|\psi\|_{L^2_{[0,1]}} < \infty$ . Further, we suppose that the total variation of  $\psi$  is finite.*

Define

$$\mathcal{T}_n = \{(t, h) : 0 \leq t < t + h \leq 1, l_n \leq h \leq u_n\},$$

where  $l_n$  and  $u_n$  are deterministic sequences fulfilling

$$\frac{\log^3(n)}{n} \ll l_n \ll u_n \ll \log^{-\epsilon}(n),$$



for some  $\epsilon > 1/(2\gamma)$ .

Some reasonable examples for  $\psi$  are discussed in Section 4.1.

### 3.3 RESULTS

In the following, we consider the test statistic

$$T_{n,t,h}^{(1)} := \sum_{i=0}^{n-1} \psi_{t,h}\left(\frac{i}{n}\right) (X_{i+1,n} - X_{i,n})^2, \quad (3.3)$$

where  $X_{i,n}$  are observations from Model (3.1), and  $\psi_{t,h}$  is chosen according to Assumption 3.3. Observe that for fixed  $t, h, n$ ,

$$T_{n,t,h}^{(1)} \approx \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds, \quad (3.4)$$

which is the quantity we like to obtain confidence statements about.

Assume that there exists a collection of estimators (or for theoretical considerations even oracles)  $\hat{\sigma}_{t,h}^2$  of  $\sigma_t^2$  which is uniformly consistent in the sense that

$$\sup_{(t,h) \in \mathcal{T}_n} |\hat{\sigma}_{t,h}^2 - \sigma_t^2| = O(s_n) \text{ a.s.}, \quad (3.5)$$

with  $s_n = o\left(\frac{\log \log(1/l_n)}{\log(1/l_n)}\right)$ . We address the problem of finding such an estimator in the next section.

The following theorem shows how good the asymptotic approximation given in (3.4) is uniformly in  $(t, h) \in \mathcal{T}_n$ . In particular, it gives a strong invariance principle, that is the difference between the terms in (3.4) can be a.s. approximated by a sequence of stochastic processes with known distribution, uniformly in  $\mathcal{T}_n$ . It can be used to obtain quantiles for testing purposes (cf. Section 4.1).

For some  $\nu > e$ , let  $w_h = \sqrt{\frac{1}{2} \log \frac{\nu}{h} (\log \log \frac{\nu}{h})^{-1}}$ . This term is needed to calibrate different scales for application of the following theorem.

**THEOREM 3.4.** *Work under Model (3.1), and Assumptions 3.1 and 3.3. Then, there exists*

a sequence of Brownian motions  $W^{[n]}$ , s.t.

$$\sup_{(t,h) \in \mathcal{T}_n} w_h \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \sqrt{\frac{2}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n]} \right| = O(q_n) \text{ a.s.}, \quad (3.6)$$

where

$$q_n = w_{l_n} l_n^{-1/2} \frac{\log(n)}{n} + w_{u_n} u_n^\gamma \left(\frac{\log n}{n}\right)^{1/2} + s_n n^{-1/2} \frac{\log(1/l_n)}{\log \log(1/l_n)}.$$

The proof of this theorem may be found in Appendix A.

REMARK 3.5. Admittedly, the approximation rate  $q_n$  in Theorem 3.4 is very slow. However, if there is prior knowledge of the approximate smoothness of  $\sigma^2$ , one may choose  $l_n$  and  $u_n$  much closer to each other than demanded by Assumption 3.3. Moreover, the required rate of convergence of  $\hat{\sigma}_{t,h}^2$  (cf. (3.5)) is a very weak assumption. Especially, it is sub-polynomial for any choice of  $l_n$  covered by Assumption 3.3. However, this should also be understood as a minimal requirement: The third part of the rate of approximation in Theorem 3.4 is the faster the better the estimator is.

### 3.4 CHOICE OF THE SPOT VOLATILITY ESTIMATOR

The estimators  $\hat{\sigma}_{t,h}^2$  determine the length of the confidence intervals which result from Theorem 3.4, cf. Section 4. We decided to formulate the Theorem for very general estimators, since practitioners might want to choose their own favorite. In particular, any sufficiently good nonparametric estimator  $(\tilde{\sigma}_t^2)_{t \in [0,1]}$  of the spot volatility is allowed, which refers to a collection of estimators which are constant in  $h$ . However, the estimators may depend on  $h$  as well, since this gives us the following very natural collection:

PROPOSITION 3.6. *Let*

$$\hat{\sigma}_{t,h}^2 := \frac{1}{h} \sum_{i=0}^{n-1} \mathbb{I}_{[t,t+h]} \left(\frac{i}{n}\right) (X_{i+1,n} - X_{i,n})^2.$$

Let further  $u_n \ll \log^{-\epsilon}(n)$  for some  $\epsilon > 1/\gamma$ . Then, the assumption in (3.5) is fulfilled.

*Proof.* Note that

$$\sup_{(t,h) \in \mathcal{T}_n} \left| \hat{\sigma}_{t,h}^2 - \sigma_t^2 \right| \leq \sup_{(t,h) \in \mathcal{T}_n} \left| \hat{\sigma}_{t,h}^2 - \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds \right| + \sup_{(t,h) \in \mathcal{T}_n} \left| \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds - \sigma_t^2 \right|. \quad (3.7)$$

For the first summand, apply Theorem 3.4 with  $\psi = \mathbb{I}_{[0,1]}$  and the oracle  $\hat{\sigma}_{t,h}^2 = \sigma_t^2$ . This gives us that almost surely,

$$\begin{aligned} \sup_{(t,h) \in \mathcal{T}_n} \left| w_h h^{1/2} \sigma_t^{-2} \left( \hat{\sigma}_{t,h}^2 - \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds \right) \right| &= \sup_{(t,h) \in \mathcal{T}_n} \left| w_h \sqrt{\frac{2}{hn}} \int_t^{t+h} dW_s^{[n]} \right| \cdot (1 + o(1)) \\ &\lesssim \sup_{(t,h) \in \mathcal{T}_n} n^{-1/2} w_h \log^{1/2}\left(\frac{\nu}{h}\right) \asymp n^{-1/2} w_{l_n} \log^{1/2}\left(\frac{1}{l_n}\right), \end{aligned}$$

where the last asymptotic inequality is due to the almost sure finiteness of the limiting statistic proved in Theorem 1 in [Schmidt-Hieber et al. \(2013\)](#). Hence, we obtain

$$\begin{aligned} \sup_{(t,h) \in \mathcal{T}_n} \left| \hat{\sigma}_{t,h}^2 - \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds \right| &\lesssim n^{-1/2} w_{l_n} \log^{1/2}\left(\frac{1}{l_n}\right) \sup_{(t,h) \in \mathcal{T}_n} w_h^{-1} h^{-1/2} \sigma_t^2 \\ &\leq \bar{\sigma}^2 (l_n n)^{-1/2} \log^{1/2}\left(\frac{1}{l_n}\right) \lesssim \log^{-3/2}(n) \log^{1/2}\left(\frac{1}{l_n}\right), \end{aligned}$$

due to Assumption 3.3. Further, we observe that

$$\log^{-3/2}(n) \log^{1/2}\left(\frac{1}{l_n}\right) \log\left(\frac{1}{l_n}\right) (\log \log\left(\frac{1}{l_n}\right))^{-1} \lesssim (\log \log\left(\frac{1}{l_n}\right))^{-1} = o(1),$$

since  $\log\left(\frac{1}{l_n}\right) \lesssim \log(n)$ .

The second summand in (3.7) is bounded by

$$\left| \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds - \sigma_t^2 \right| \leq \frac{1}{h} \int_t^{t+h} |\sigma_s^2 - \sigma_t^2| ds \lesssim \frac{1}{h} \int_0^h s^\gamma ds = \frac{1}{\gamma+1} h^\gamma,$$

where we used Assumption 3.1. Finally, observe that  $\sup_{(t,h) \in \mathcal{T}_n} h^\gamma = u_n^\gamma \ll \log^{-1}(n)$ . Again, by  $\log\left(\frac{1}{l_n}\right) \lesssim \log(n)$ , the proof is complete.  $\square$



---

# CHAPTER 4

## INFERENCE ON SPOT VOLATILITY

---

In this chapter, we like to indicate how to apply Theorem 3.4 to obtain inferential statements on the diffusion. In the following, we choose  $\hat{\sigma}_{t,h}^2$  as described in Proposition 3.6.

### 4.1 APPLICATION OF THEOREM 3.4

This subsection is strongly connected to the ideas of Dümbgen and Walther (2008) and Schmidt-Hieber et al. (2013), who developed tests for local features in density estimation without and with deconvolution, as well as to Dümbgen and Spokoiny (2001), who introduced such results for nonparametric regression.

For any combination  $(t, h) \in \mathcal{T}_n$ , Theorem 3.4 in combination with the triangle inequality gives us that the a.s. approximation

$$\sup_{(t,h) \in \mathcal{T}_n} w_h \left( \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} \right| - \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \quad (4.1)$$

$$= \sup_{(t,h) \in \mathcal{T}_n} w_h \sqrt{\frac{2}{n}} \left( \left| \int_0^1 \psi_{t,h}(s) dW_s^{[n]} \right| - \sqrt{2 \log \frac{\nu}{h}} \right) \cdot (1 + o(1)) \quad (4.2)$$

holds uniformly over  $\mathcal{T}_n$ . Here, the terms  $w_h$  and  $(2 \log \frac{\nu}{h})^{1/2}$  are chosen to balance the influence of the different scales, so that  $n^{1/2}$  times the right hand side is a.s. finite (cf. Schmidt-Hieber et al. (2013) and Dümbgen and Spokoiny (2001)). Denote the  $(1 - \alpha)$ -quantile of (4.2) by  $\sqrt{\frac{2}{n}} q_{1-\alpha}$ . Then, (4.2) implies the following proposition:

PROPOSITION 4.1. *Asymptotically,  $\int_0^1 \psi_{t,h}(s) \sigma_s^2 ds$  is in the interval*

$$CI_{t,h} = \left[ T_{n,t,h}^{(1)} - \left( \sqrt{\frac{2}{n}} \frac{q_{1-\alpha}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2, T_{n,t,h}^{(1)} + \left( \sqrt{\frac{2}{n}} \frac{q_{1-\alpha}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2 \right] \quad (4.3)$$

with probability  $1 - \alpha$  uniformly in  $\mathcal{T}_n$ .

Since  $\text{supp } \psi_{t,h} \subseteq [t, t+h]$  and  $h < u_n \rightarrow 0$ , Proposition 4.1 allows us to make local statements about  $\sigma^2$ . Here, the choice of  $\psi$  heavily depends on the quantity of interest. For a heuristic explanation, assume for the moment that  $\sigma^2$  is  $m$ -times continuously differentiable. Even though this assumption is not fulfilled in practice, it will yield a good impression of how to choose  $\psi$ .

Let  $\mathfrak{D}$  be a linear differential operator of order  $p$  given by

$$\mathfrak{D}f = \sum_{k=0}^p a_k D^k f,$$

with  $Df(x) = \frac{d}{dx}f(x)$ , for  $f$  which is  $p$  times continuously differentiable, and for some  $k$  times continuously differentiable functions  $a_k$ ,  $k = 0, \dots, p$ , where  $a_p \neq 0$ . Suppose that we are interested in the local behavior of  $\mathfrak{D}\sigma^2$ . Let  $K$  be a non-negative kernel with  $\text{supp } K \subseteq [0, 1]$  and  $K^{(i)}(0) = K^{(i)}(1) = 0$  for all  $i = 0, \dots, p-1$ . Then, we obtain by integration by parts that

$$\int_t^{t+h} K\left(\frac{s-t}{h}\right) (\mathfrak{D}\sigma^2)(s) ds = \int_t^{t+h} (\mathfrak{D}^* K\left(\frac{\bullet-t}{h}\right))(s) \sigma^2(s) ds, \quad (4.4)$$

where  $\mathfrak{D}^*$  is the formal adjoint of  $\mathfrak{D}$ . Note that

$$\mathfrak{D}^* K\left(\frac{\bullet-t}{h}\right)(s) \sim h^{-p} (-1)^p a_p(s) K^{(p)}\left(\frac{s-t}{h}\right),$$

when  $h$  tends to zero (which is the case here, since the upper bound  $u_n$  in Assumption 3.3 tends to zero). Thus, choosing  $\psi = (-1)^p a_p(s) K^{(p)}$  yields simultaneous confidence intervals as given in (4.3), which in turn can be transformed via (4.4) into confidence intervals  $CI_{t,h}$  for  $\int_t^{t+h} K\left(\frac{s-t}{h}\right) (\mathfrak{D}\sigma^2)(s) ds$ . Note that the scaling term  $h^{-p}$  is not relevant here, since  $\psi_{t,h}$  is  $L^2$ -normalized.

Since  $K$  is chosen as a kernel and  $\mathfrak{D}\sigma^2$  is assumed to be continuous, we may conclude that with probability  $1 - \alpha$ , for all  $(t, h) \in \mathcal{T}_n$ , there exists an  $s \in (t, t+h)$  with  $\mathfrak{D}\sigma^2(s) \in CI_{t,h}$ .

The following example will make it more obvious how to use these equations result in practice:

**EXAMPLE 4.2.** *Let  $\mathfrak{D}$  be the first derivative operator, that is  $(\mathfrak{D}f)(s) = \frac{d}{ds}f(s)$ . Further, write  $CI_{t,h} = [a_{t,h}, b_{t,h}]$ . Then, for any  $(t, h)$  with  $a_{t,h} \geq 0$ , we may reject the hypothesis “ $\sigma^2$  is strictly decreasing on  $[t, t+h]$ ” simultaneously at level  $\alpha$ .*

*This connection between testing and confidence sets is addressed more detailed in Section 4.2.*

While a differentiable diffusion  $\sigma^2$  is often of theoretical interest merely, financial spot volatility is often modeled to be rougher in practice. It is often assumed to be not even Hölder continuous with smoothness parameter  $\gamma \geq \frac{1}{2}$  (cf. for example the Heston model in [Heston \(1993\)](#), where  $\sigma^2$  is a continuous semimartingale itself). To understand the results in this situation, let us first discuss how to choose the kernel appropriately to obtain a simple interpretation: Consider the setting of Example 4.2. Here, the triangular kernel  $\psi_\Delta(x) = 4x\mathbb{I}_{[0,1/2]}(x) + (4 - 4x)\mathbb{I}_{(1/2,1]}(x)$  appears suitable, since its derivative is given by  $4\mathbb{I}_{[0,1/2]}(x) - 4\mathbb{I}_{(1/2,1]}(x)$ . Thus, we may interpret the situation “ $a_{t,h} > 0$ ” as the average of  $\sigma^2$  on  $[t, t + h/2]$  being significantly larger than the average on  $[t + h/2, t + h]$ , even if  $\sigma$  is not differentiable. Similar interpretation are possible for higher-order differential operators as well. For instance, we may choose  $K$ , such that its second-derivative is proportional to  $\mathbb{I}_{[0,1/4] \cup [3/4,1]} - \mathbb{I}_{[1/4,3/4]}$ , referring to statements about convexity/concavity, or in practice, changes in the volatility of volatility. This shows that by choosing the kernel in such a way that  $\mathfrak{D}^*K$  is as simple as possible, we find simple interpretations of the results relying on discretized versions of the problem.

**REMARK 4.3.** *The Gaussian approximation in the proof of Theorem 3.4 given in Appendix A (cf. step I in the proof) is useful to obtain the correct penalization for each scale (which is done by subtraction of  $\sqrt{2 \log \frac{v}{h}}$  and multiplication by  $w_h$ ). However, it is dispensable for practical purposes. Instead, it is more accurate to consider quantiles of the supremum of weighted sums of centered and normalized  $\chi^2$  random variables. The proof reveals that these sums are already distribution-free.*

**REMARK 4.4.** *In practice, it is useful to imply further restrictions on  $\mathcal{T}_n$ , such as requiring  $t$  and  $h$  coming from the discrete grid  $\mathcal{X}_n := \{\frac{i}{n} : i \in \mathbb{N}, 0 \leq i \leq n\}$ . To reduce running time, one can also consider only intervals of dyadic length, that is  $h \in \{\frac{2^i}{n} : i \in \mathbb{N}, 0 \leq i \leq \log_2 n\}$ , while  $t$  is in  $\mathcal{X}_n$  again. In any of this cases, Theorem 3.4 holds as well. The only difference (besides the running time) is the finite performance of detecting features. However, we like to emphasize that it is not possible to decide which choice of candidate intervals works better in general, since the performance heavily relies on the respective realization of  $\sigma^2$ . On the one hand, the quantiles of the limiting statistic will be smaller, if we consider some subset of  $\mathcal{T}_n$ . This allows us to more easily detect a feature, whose support is of dyadic length. On the other hand, features of non-dyadic length are possibly not detected at all.*

An even more restrictive assumption would be to consider only those  $t$  that have a representation  $t = hk$  for some  $k \in \mathbb{N}$  and  $h$  on the dyadic grid as in the second case. In this case, each of the considered intervals  $[t, t + h]$  is the support of a basis function used in wavelet decomposition. Thus, Theorem 3.4 gives us a simultaneous hard thresholding rule for wavelet reconstruction.

In Table 4.1, we compare running times and quantiles of (4.2) for the three scenarios described above for the choice  $\psi = \mathbb{I}_{[0,1]}$  and  $n = 10,000$ . Note that the “standard” case refers to  $\mathcal{T}_n$  as in Assumption 3.3 with  $l_n = \log_{10}^3(n)/n = 0.027$  and  $u_n = \log_{10}^{-1}(n) = 1/3$ . We observe that running times differ considerably between the three scenarios while the quantiles appear stable.

## 4.2 DETECTION RATES

At this point, we like to discuss how small a certain feature of the spot volatility may be to keep it detectable. This can also be viewed as a power analysis of the underlying multiscale test, as we will see in the following. Note that all statements have to be understood as asymptotic statements, since our result is also only of this nature.

Let  $\psi$  be some test function satisfying Assumption 3.3. Let us consider one specific  $(t, h) \in \mathcal{T}_n$ , while we still keep in mind that our confidence statements hold simultaneously in  $\mathcal{T}_n$ . Let  $CI_{t,h}$  be the confidence interval given in (4.3), that is

$$CI_{t,h} = \left[ T_{n,t,h}^{(1)} - \left( \sqrt{\frac{2}{n}} \frac{q_{1-\alpha}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2, T_{n,t,h}^{(1)} + \left( \sqrt{\frac{2}{n}} \frac{q_{1-\alpha}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2 \right].$$

Candidate set	$\mathcal{T}_n \cap \mathcal{X}_n^2$	$h$ dyadic	wavelet supports
Running Time (milliseconds)	1,998.67	26.02	1.13
90%-quantile	1.42	1.11	0.33
95%-quantile	1.72	1.36	0.61

TABLE 4.1: Simulated quantiles of (4.2) (without the factor  $(2/n)^{1/2}$ ) for  $\psi = \mathbb{I}_{[0,1]}$  and  $n = 10,000$  based on 10,000 repetitions and the average running time per simulation. The different columns refer to different choices of candidate sets as described in Remark 4.4.



Suppose that we are interested in the test problem

$$H : \int \psi_{t,h} \sigma^2 \in I_{t,h} \text{ vs. } K : \int \psi_{t,h} \sigma^2 \notin I_{t,h},$$

for some interval  $I_{t,h} = [c_{t,h}, d_{t,h}] \subset \mathbb{R}$ , which contains at least one point. Since  $CI_{t,h}$  is a level- $\alpha$  confidence interval, we can reject  $H$  significantly at level  $\alpha$ , if  $I_{t,h} \cap CI_{t,h}$  is empty. Obviously, if the truth  $\int \psi_{t,h} \sigma^2$  is very close to  $I_{t,h}$ , a significant rejection of  $H$  is quite unlikely. Thus, the following question arises: How large is the minimal distance between  $\int \psi_{t,h} \sigma^2$  and  $I_{t,h}$  to guarantee a rejection with probability  $1 - \beta$ ,  $\beta \in [0, 1]$ , that is guaranteeing power  $1 - \beta$ ?

To answer this question, observe that

$$T_{n,t,h}^{(1)} \in \left[ \int \psi_{t,h} \sigma^2 - \left( \sqrt{\frac{2}{n}} \frac{q_{1-\beta}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2, \int \psi_{t,h} \sigma^2 + \left( \sqrt{\frac{2}{n}} \frac{q_{1-\beta}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2 \right]$$

with probability at least  $1 - \beta$ , which can be obtained by Proposition 4.1. Thus, if the distance between  $\int \psi_{t,h} \sigma^2$  and  $I_{t,h}$  is at least

$$\left( \sqrt{\frac{2}{n}} \frac{q_{1-\beta}}{w_h} + \sqrt{\frac{4}{n} \log \frac{\nu}{h}} \right) \hat{\sigma}_{t,h}^2 \asymp n^{-1/2} (-\log h)^{1/2},$$

we can assure power  $1 - \beta$  asymptotically. Note that often  $\int \psi_{t,h} \sigma^2$  is not directly the quantity of interest, since it is not scaled properly for the purpose of interpretation. Rather, one is interested  $h^k \int \psi_{t,h} \sigma^2$ , where  $k$  heavily depends on the choice of  $\psi$ , as we can see in the following two examples. They deal with the important special cases of testing the (weighted) average of  $\sigma^2$  and the variation of  $\sigma^2$  on the interval  $[t, t + h]$ :

**EXAMPLE 4.5** (Testing the (weighted) average of  $\sigma^2$ ). *Suppose that we are interested in the average*

$$\text{av}_{t,h}(\sigma^2) := \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds.$$

*To this end, consider the indicator function  $\psi = \mathbb{I}_{[0,1]}$ , that is  $\psi_{t,h} = h^{-1/2} \mathbb{I}_{[t,t+h]}$ . Then,*

$$\text{av}_{t,h}(\sigma^2) = h^{-1/2} \int \psi_{t,h} \sigma^2,$$

*so that testing whether  $\text{av}_{t,h}(\sigma^2) \in I^* := [c^*, d^*]$  for some  $0 \leq c^* < d^* \leq \infty$  refers to testing whether  $\int \psi_{t,h} \sigma^2 \in [h^{1/2} c^*, h^{1/2} d^*]$ . For detecting  $\text{av}_{t,h}(\sigma^2) \notin I^*$ , it is therefore necessary that the distance between  $\int \psi_{t,h} \sigma^2$  and  $[h^{1/2} c^*, h^{1/2} d^*]$  is at least of order  $n^{-1/2} (-\log h)^{1/2}$ .*

Thus,  $\inf_{u \in I^*} |\text{av}_{t,h}(\sigma^2) - u|$  must be at least of order  $(hn)^{-1/2}(-\log h)^{1/2}$ .

Note that in principle,  $\psi$  can be any kernel fulfilling Assumption 3.3. This would give us statements about the weighted average of  $\sigma^2$ .

Even more, note that  $(hn)^{-1/2}(-\log h)^{1/2}$  is rate-optimal in the following sense:

Suppose that we know the Hölder smoothness  $\gamma \in [0, 1]$  of  $\sigma^2$ . If we want to draw conclusions about  $\sigma_u^2$  for every  $u$  in  $[t, t+h]$ , we must take into account that  $\sigma^2$  may vary in  $[t, t+h]$  by terms of order  $h^\gamma$ . Thus, we need a minimum distance of order  $h^\gamma + (hn)^{-1/2}(-\log h)^{1/2}$  between the boundary of the hypothesis and the truth. Minimizing this is equivalent to balancing the rates  $h^\gamma$  and  $(hn)^{-1/2}(-\log h)^{1/2}$ . This gives us the rate-optimal bandwidth

$$h_* \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{2\gamma+1}},$$

which turns out to be the same as the rate-optimal bandwidth in nonparametric regression w.r.t. sup-norm, cf. for example [Tsybakov \(2009\)](#). Furthermore, we obtain

$$h_*^\gamma + (h_*n)^{-1/2}(-\log h_*)^{1/2} \asymp \left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\gamma+1}},$$

which is in turn the optimal rate of convergence in nonparametric regression w.r.t. sup-norm. Here, sup-norm estimation is the appropriate procedure to compare our test with, since the level of the test is kept simultaneously over  $\mathcal{T}_n$ .

**REMARK 4.6.** *In the previous paragraph, we have seen that a bandwidth of rate  $h_*$  is more informative than bandwidths of other rates. Therefore, our test of  $H$  vs.  $K$  is adaptive over all Hölder classes  $\mathcal{H}(\gamma, L)$  with  $(t, (\frac{\log n}{n})^{\frac{1}{2\gamma+1}}) \in \mathcal{T}_n$  for some  $t \in [0, 1]$ .*

*However, it is not possible to construct adaptive and rate-optimal uniform confidence bands with this technique, that is confidence intervals for  $\sigma_s^2$ , which are valid simultaneously in  $s \in [0, 1]$ . This is indeed a very hard task: According to [Low \(1997\)](#), it is not possible to construct adaptive and honest confidence bands in nonparametric regression. We conjecture that this result also applies in our setting: Note that for deterministic  $\sigma^2$ , we are in a situation close to a regression setting (cf. for example the results on asymptotic equivalence presented in [Reiß \(2011\)](#)). Hence, it is not possible to construct adaptive and honest confidence bands in that simple case, already. In the more general model, that is assuming that  $\sigma^2$  is random and possibly depends on the underlying Brownian motion  $W$ , there is*

no direct analogon to the honesty condition. However, the uniformity over the underlying Hölder space is partly expressed in the probability, which is a function on the sigma-algebra generated by  $(\sigma, W)$ . Therefore, the only hope is that the realizations of  $\sigma^2$ , which disturb the uniformity, lie in a null set, that is they do not occur almost surely.

EXAMPLE 4.7 (Testing for variation of  $\sigma^2$  via differential operators). *This example is a continuation of Example 4.2. Again, assume that  $\sigma^2$  is in  $\mathcal{H}(p+\gamma, L)$ ,  $\gamma \in (0, 1)$ ,  $1 \leq p \in \mathbb{N}$ ,  $L \in \mathbb{R}_+$ , that is  $\sigma^2$  is  $p$ -times continuously differentiable with Hölder( $\gamma$ ) continuous  $p$ -th derivative. Let  $\mathfrak{D}$  be the  $p$ -th derivative operator, that is  $(\mathfrak{D}f)(s) = (\frac{d}{ds})^p f(s)$ . If we are interested in statements about the  $p$ -th derivative of  $\sigma^2$ , we may consider*

$$\text{dev}_{t,h}^p(\sigma^2) = \frac{1}{h} \int_t^{t+h} K\left(\frac{s-t}{h}\right) \mathfrak{D}\sigma_s^2 ds$$

for  $(t, h) \in \mathcal{T}_n$  and some kernel  $K$  fulfilling the conditions in Section 4.1. By integration by parts as described in Section 4.1, we can choose  $\psi$ , so that

$$\text{dev}_{t,h}^p(\sigma^2) = h^{-p-1/2} \int_t^{t+h} \psi_{t,h}(s) \sigma_s^2 ds. \quad (4.5)$$

Thus, we may proceed as in the previous example: Inverting the confidence bands gives us a test of the hypothesis  $H : \text{dev}_{t,h}^p(\sigma^2) \in I$  for some interval  $I$ . Combining (4.5) and the length of the confidence bands in (4.3),  $\inf_{u \in I} |\text{dev}_{t,h}^p(\sigma^2) - u|$  must be at least of order  $(hn)^{-1/2} h^{-p} (-\log h)^{1/2}$  to ensure (correct) rejection of the hypothesis.

Furthermore, since  $\mathfrak{D}\sigma^2$  is Hölder( $\gamma$ ) continuous, pointwise detection of features of  $\mathfrak{D}\sigma^2$  is only possible if they exceed  $h^\gamma + (hn)^{-1/2} h^{-p} (-\log h)^{1/2}$ . This term is minimized by

$$h_* \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{2(\gamma+p)+1}},$$

yielding a detection rate of

$$h_*^\gamma + (h_* n)^{-1/2} h_*^{-p} (-\log h_*)^{1/2} \asymp \left(\frac{\log n}{n}\right)^{\frac{\gamma}{2(\gamma+p)+1}}.$$

These rates coincide with the optimal bandwidth and convergence rates for nonparametric estimation of the derivative of a regression function w.r.t. sup-norm (cf. for example [Donoho \(1994\)](#), corollary of Theorem D).

### 4.3 VISUALIZATION AND FIRST DATA EXAMPLES

So far, we have constructed confidence intervals, which can be used for further computations. However, the set of confidence intervals

$$\{(t, t + h, \inf CI_{t,h}, \sup CI_{t,h}) : (t, h) \in \mathcal{T}_n\}$$

is very complex if the cardinality of  $\mathcal{T}_n$  is large. Therefore, it is crucial to find a good visualization of this set.

In this section, we like to present and compare four different approaches to this aim: (A) The minimal lines approach presented in [Dümbgen and Walther \(2008\)](#) and more detailed the minimal rectangle approach in [Schmidt-Hieber et al. \(2013\)](#), (B) a level set plot coming from the projection of this rectangles to the  $x$ -axis, (C) a visualization of the degree of rejection, and (D) illustration methods using moving pictures.

(A) The probably most intuitive way of visualization is to draw two-dimensional rectangles with vertices

$$\left( (t, \inf CI_{t,h}), (t + h, \inf CI_{t,h}), (t + h, \sup CI_{t,h}), (t, \sup CI_{t,h}) \right),$$

for every  $(t, h) \in \mathcal{T}_n$ . This approach appears to be especially useful for simultaneous testing

$$(\mathfrak{D}f)(s) > 0 \text{ for all } s \in I \in \mathcal{I} \tag{4.6}$$

for some differential operator  $\mathfrak{D}$ , some intervals  $I$ , and a collection of those intervals  $\mathcal{I}$ . In this case, rejection of the hypothesis on a small interval  $I$  already implies rejection on any superinterval of  $I$ , so that the huge collection of rectangles can be reduced to a hopefully small set of minimal rectangles containing all the information regarding the test. This reduction is more generally discussed in [Schmidt-Hieber et al. \(2013\)](#). As a further reduction, it is often enough to draw an indicator line from  $t$  to  $t + h$  implying that the lower bound of the rectangle is larger than zero (that is projecting the rectangle to the  $x$ -axis) instead of drawing the whole rectangle.

However, our procedure (as well as the ones presented in [Dümbgen and Walther \(2008\)](#) and [Schmidt-Hieber et al. \(2013\)](#)) only tests if the (weighted) average of  $\mathfrak{D}f$  on a interval is

### 4.3. VISUALIZATION AND FIRST DATA EXAMPLES

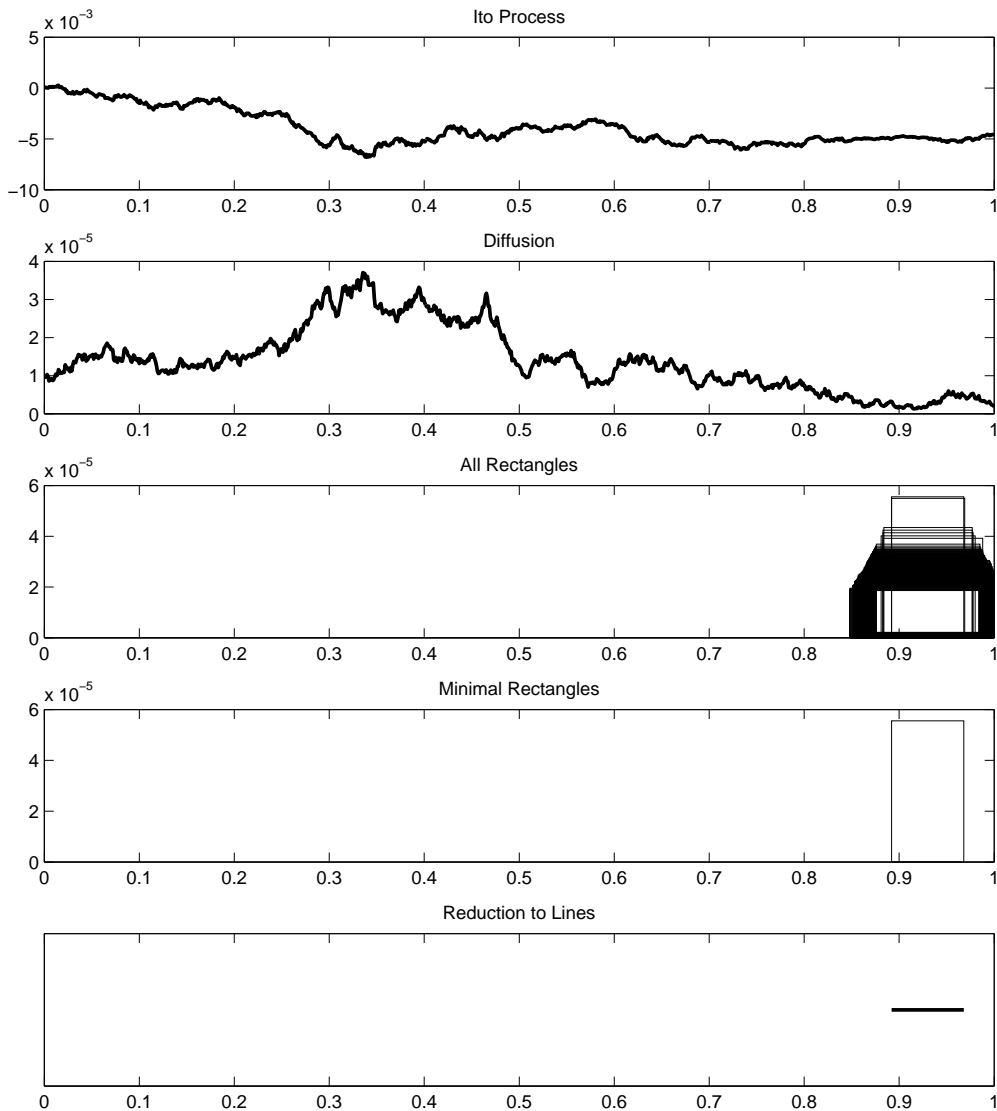


FIGURE 4.1: Observed  $\hat{I}_t$  process (panel 1) and unobserved diffusion (panel 2) following Model (4.7) as well as visualizations (panels 3-5) of intervals where  $\sigma^2$  is significantly increasing in average (at level  $\alpha = 5\%$ ). Here, panel 3 displays all rectangles which are strictly above the  $x$ -axis, panel 4 contains the sufficient reduction for testing (4.6), and in panel 5, each sufficient rectangle is projected to a line from  $t$  to  $t+h$ . Note that in the last case, the  $y$ -values are without any relevance.

positive. Pointwise statements as in (4.6) are obtained afterwards via mean value theorem. For this purpose, it is necessary to assume continuity of  $\mathfrak{D}f$ , which is a common assumption in density estimation but not in financial practice. Here, the expression  $\mathfrak{D}\sigma^2$  does generally not even exist. Therefore, for volatility inference, one is interested in the weighted averages themselves. This thwarts the attempt to reduce the set of rectangles to an informative subset suitable for visualization.

Nevertheless, we depict the method of rectangles and its reductions in Figure 4.1 using simulated data following the Heston model (cf. Heston (1993) and the survey in Chapter 7 of Čížek et al. (2005)), that is we assume that  $\sigma^2$  is a semimartingale itself possibly correlated with the driving Brownian motion  $W$ . More precisely, we consider

$$\begin{aligned} dX_t &= -\frac{1}{2}\sigma_t^2 dt + \sigma_t dW_t, \\ d\sigma_t^2 &= \kappa(\theta - \sigma_t^2) dt + \epsilon\sigma_t d\widetilde{W}_t, \end{aligned} \tag{4.7}$$

where the  $W$  and  $\widetilde{W}$  are Brownian motion with (constant) correlation  $\mathbb{E}(W_1\widetilde{W}_1) = \rho \in (-1, 1)$ . We choose the parameters

$$\rho = -2/3, \theta = 10^{-5}, \kappa = 4, \epsilon = \sqrt{\kappa\theta} = 2 \cdot 10^{-5/2}. \tag{4.8}$$

Further, we choose  $n = 1,000$  and our goal is to identify intervals where  $\sigma^2$  is increasing in average significantly at a simultaneous level of 5%. To this end, we used  $\psi \propto \mathbb{I}_{[0,1/2)} - \mathbb{I}_{[1/2,1]}$ . The 95%-quantile 1.74 is obtained via 10,000 Monte-Carlo simulations, whose empirical distribution function is given in Figure 4.2.

As a result, we see that the reduced set of rectangles yields an easily understandable and quite clear graphic (at the expense of a loss of information). Surprisingly, the apparently large increase in the first half of the data is not declared significant, while the comparably small increase at the end is. This can be explained by the fact that in Theorem 3.4, the approximation error is divided by the estimator  $\hat{\sigma}_{t,h}^2$  to obtain a distribution-free limit. Therefore, our method detects an increase if it is large relatively to the actual value of the diffusion.

(B) As we have seen, reduction may cause a loss of information on large scales, if we are interested in the averages themselves. Therefore, we like to propose a different approach:

---

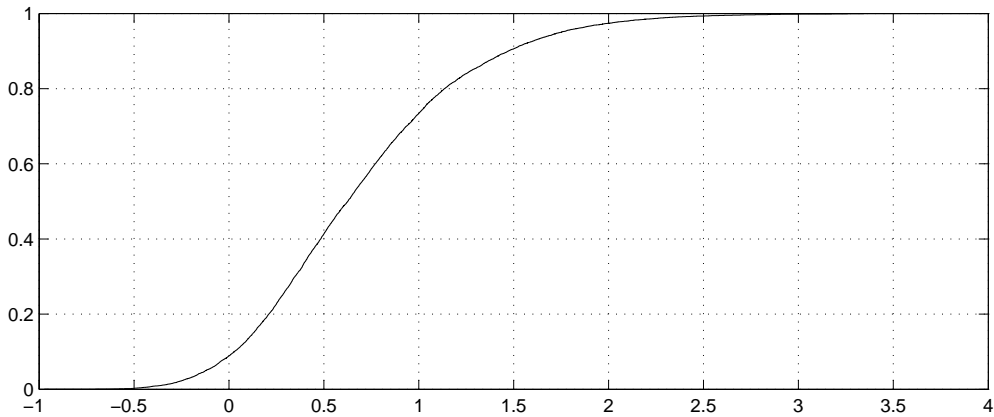


FIGURE 4.2: Empirical distribution function of 10,000 simulations of the supremum over the penalized limit process (cf. (4.2)) with  $\psi \propto \mathbb{I}_{[0,1/2)} - \mathbb{I}_{[1/2,1]}$ .

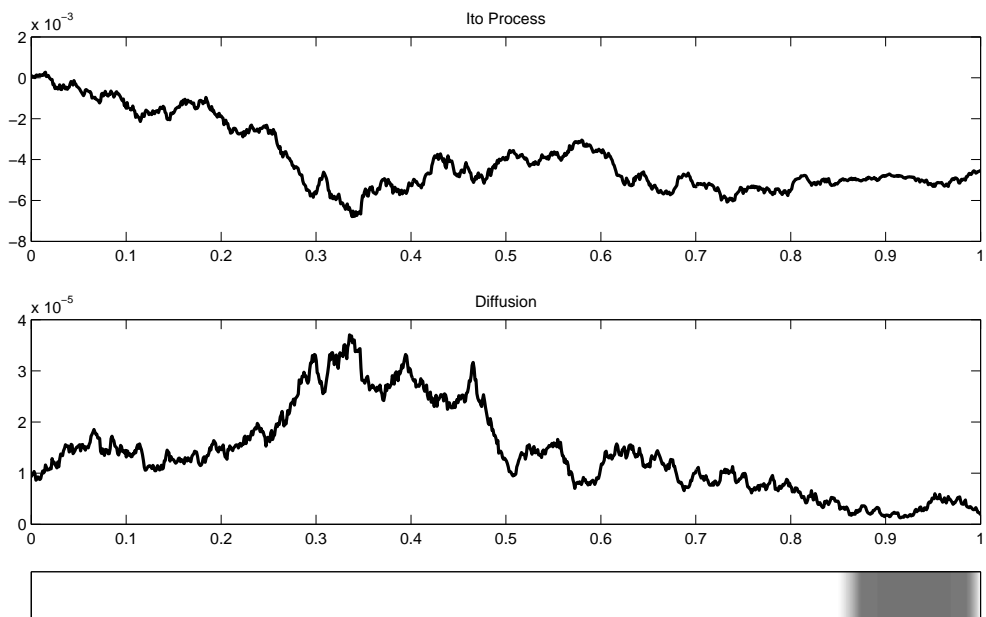


FIGURE 4.3: Observed Itô process (panel 1), unobserved diffusion (panel 2), and level set plot (panel 3).

For each point  $s \in [0, 1]$ , let  $R(s)$  be the number of intervals containing  $s$  on which the hypothesis is rejected. Displaying the level sets of this function as a one-dimensional gray-scale plot gives a reasonable overview over all intervals which are rejected. Figure 4.3 shows the level set plot for the data considered in (A).

This method obviously suffers from a loss of information, since one cannot determine in retrospect on which interval a rejection took place. Moreover, the viewer might get the impression that darker regions respond to heavier increases. Although this implication is often correct, it is not true in general. Therefore, the different gray-scales have to be considered as an indication to distinguish between different intervals and not as an indicator for the size of the increase.

Besides these drawbacks, this method gives a very clear intuition if and where something happened, and therefore, if the user should take a closer look at the original confidence intervals. Especially, large intervals are not omitted as it is often done in the minimal rectangle procedure.

(c) In Method (B), we lose the information how large a certain violation is. If we consider a test problem  $\int \psi_{t,h} \sigma^2 > 0$ , we can think of a 2-dimensional plot, displaying level set plots ( $x$ -axis) for different choices of quantiles ( $y$ -axis). Note that these quantiles can directly be transferred to  $p$ -values (cf. Figure 4.2).

Fortunately, there is a.s. no loss of information (up to the null set that the supremum of the penalized test statistic as given in (4.1) is attained at two different intervals  $(t, t+h)$ ): By finding and subtracting the highest peak consecutively, we can theoretically recover the original confidence intervals. However, this can hardly be done by eye, so that the resulting plot requires a careful interpretation.

The huge advantage of this method is that the user obtains quantitative statements in terms of  $p$ -values, while the methods proposed in (A) and (B) only contains statements for a fixed significance level  $\alpha$ .

The resulting image is given in Figure 4.4. It proves our visual impression that there is some increase in the first half of the data. Here, we see that it can be detected with significance 7% or below.

(d) The main difficulty in visualization is the aim of presenting a lot of information while trying to avoid a loss of information. In our opinion, the best way to solve this problem is to consider multidimensional approaches, such as movies or graphical user interfaces (GUI). These methods offer the chance of splitting information into small, “local” parts, which are understandable for the user.



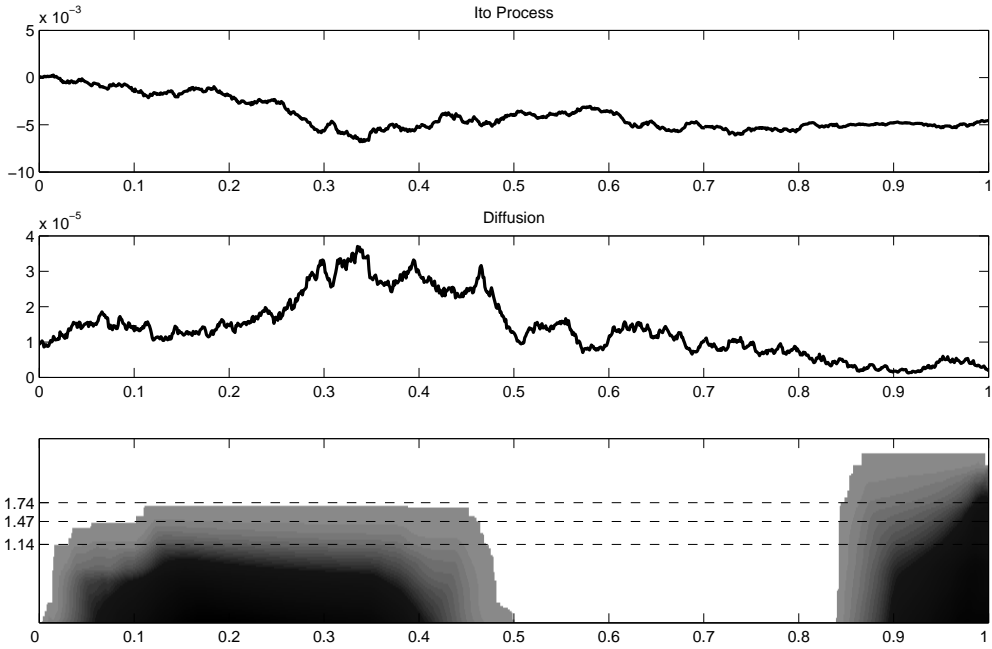


FIGURE 4.4: Observed Itô process (panel 1), unobserved diffusion (panel 2), and intervals of significant increase in average (panel 3). In panel 3, the  $y$ -axis gives different quantiles of the (4.2), that is the supremum over the limiting process. For any of this values, the gray regions indicate intervals on which the test statistic is greater than the respective quantile. The quantiles corresponding to type-I-errors  $\alpha = 5\%, 10\%, 20\%$  are indicated by the dashed lines.

A possible solution would be to localize the level set plots in the following sense: In the construction described in (B) or (C), we replace  $R$  by the local functions

$$R_u(s) = \#\{[t, t + h] : s \in [t, t + h], u \in [t, t + h], \text{ Hypothesis is rejected on } [t, t + h]\},$$

where  $u$  takes values  $[0, 1]$ . In this case,  $R_u$  carries all the information contained in  $R$  which is connected to the process at time  $u$ . Considering  $u$  as the time variable of a movie will give us some clear local information about the diffusion.



---

## CHAPTER 5

# EXTENSION TO HIGH-FREQUENCY DATA

---

### 5.1 MOTIVATION

As we mentioned in the introduction, high-frequently sampled data cannot be explained well enough by a semimartingale model. Before turning to a more general model and to theory within this new model, we like to verify the inadequacy of the simple model on the basis of prices of Euro-Bund-Futures (FGBL).

The FGBL price of August 1st, 2007 is displayed in Figure 5.1. In the magnification in the bottom of the plot, we observe that the process only takes very discrete values and seems to vary a lot on small scales.

Additionally, classic quadratic variation based methods fail for the data (cf. for example Figure 1.1 in the Introduction).

To overcome these problems, [Hasbrouck \(1993\)](#) (among others) suggested to integrate an additive noise term into the model, the so-called microstructure noise.

### 5.2 MODELING

Instead of considering data from Model (1.1) corrupted by additive noise, we assume that we observe the whole process  $(Y_t)_{t \in [0,1]}$  described by

$$dY_t = X_t dt + \tau n^{-1/2} dW_t^*, \quad t \in [0, 1], \quad (5.1)$$

---

<sup>1</sup>Figure 1.1 is taken from [Sabel et al. \(2014\)](#)

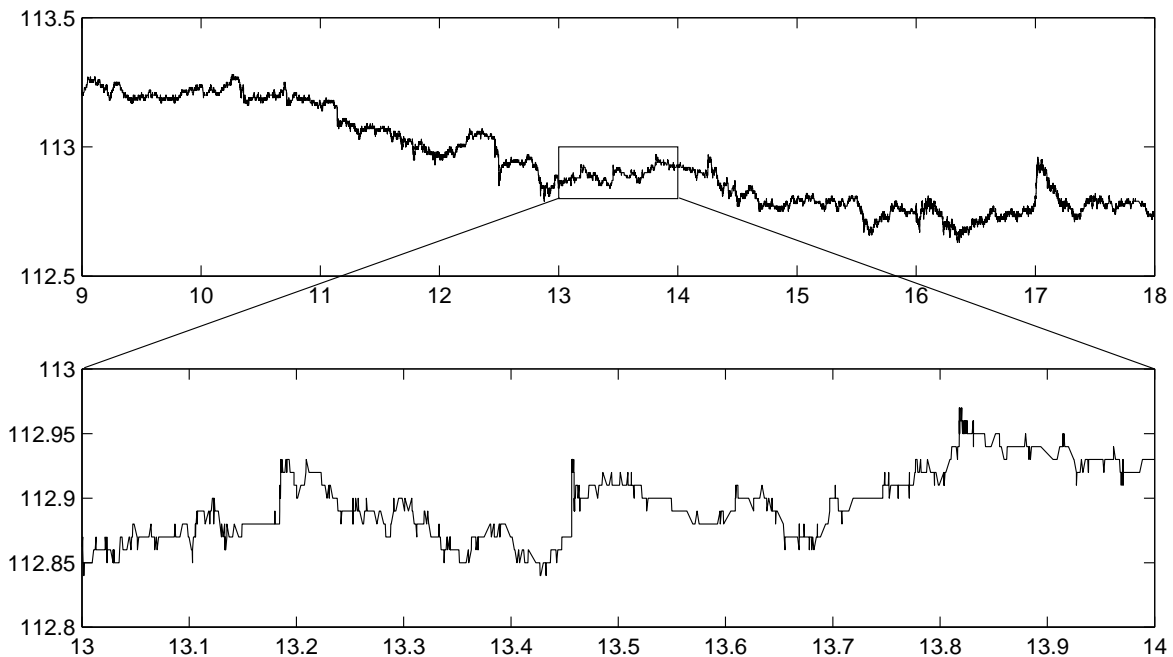


FIGURE 5.1: FGBL data for August 1st, 2007, and magnification of the time interval from 1 p.m. to 2 p.m.

where  $X$  is the process given in (1.1) (w.l.o.g., assume again  $X_0 = 0$  and  $b \equiv 0$ ),  $\tau > 0$  is the standard deviation of the noise and  $W^*$  is Brownian motion, independent of  $X$  and  $\sigma$ .

REMARK 5.1. *Under the assumption that  $\sigma$  is a process, such that  $X$  is conditionally on  $\sigma$  a Gaussian process, it is proved in Reiß (2011) that this model is asymptotically equivalent in Le Cam's sense to observing discrete data*

$$Y_{i,n} = X_{i/n} + \tau\epsilon_i, i = 1, \dots, n, \quad (5.2)$$

where  $\epsilon$  is Gaussian white noise. Roughly speaking, this means that any decision procedure based on Model (5.1) can asymptotically be imitated by a procedure based on Model (5.2) without losing information, and vice versa. A more rigorous definition is given in Le Cam and Yang (1990).

Here, we choose Model (5.1), since it is easier to handle for our purpose since it avoids discretization effects.

REMARK 5.2.  $\tau$  may be random and non-constant, in general. However, one can estimate

it very precisely from the data via quadratic variation (for example, an  $n^{-\beta/(2\beta+1)}$ -rate is achievable if  $\tau$  is in a bounded Sobolev ellipsoid of smoothness  $\beta > 1$ , cf. [Munk and Schmidt-Hieber \(2010\)](#)) and is therefore of minor importance for the scope of this thesis. Thus, we assume that  $\tau$  is a known constant in order to keep the proofs as simple as possible.

## 5.3 RESULTS

In the following, we work under Model (5.1). To handle the additional noise, we will use the pre-averaging approach introduced in [Jacod et al. \(2009\)](#) and refined in [Hoffmann et al. \(2012\)](#). Loosely speaking, we will compute local averages of the data, reducing the impact of the noise process while not affecting the continuous martingale part. More thoroughly, let us recall the definition of a normalized pre-average function:

**DEFINITION 5.3.** *A piecewise Lipschitz continuous function  $\lambda : \mathbb{R} \rightarrow \mathbb{R}$  with support in  $[0, 1]$ , satisfying  $\lambda(t) = -\lambda(1 - t)$  for all  $t \in [0, \frac{1}{2}]$ , and*

$$\int_0^1 \left( \int_0^s \lambda(u) du \right)^2 ds = 1 \quad (5.3)$$

*is called a (normalized) pre-average function.*

For some normalized pre-average function  $\lambda$ , consider its (negative) anti-derivative  $\Lambda(s) = -\int_0^s \lambda(t) dt$ . For some  $m = m(n)$  with  $1 \ll m \ll n$ , let  $\lambda_i = m\lambda(m \bullet - i)$ ,  $i = 0, \dots, m-1$  and  $\Lambda_i = \Lambda(m \bullet - i)$ . We introduce the pre-averaged values

$$\begin{aligned} \bar{Y}_{i,m} &:= \int_0^1 \lambda_i(s) dY_s = \int_0^1 \lambda_i(s) X_s ds + \tau n^{-1/2} \int_0^1 \lambda_i(s) dW_s^* \\ &= \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \sigma_s dW_s + \tau n^{-1/2} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^*, \end{aligned} \quad (5.4)$$

where the third equality is valid due to integration by parts and the symmetry assumption  $\lambda(t) = -\lambda(1 - t)$  (implying in particular that  $\Lambda(1) = 0$ ).

**REMARK 5.4.** *Note that our pre-averaging procedure produces independent pre-averaging values, if  $\sigma$  is deterministic. This relies on the fact that the  $i$ -th pre-averaged value only depends on data from the interval  $[\frac{i}{m}, \frac{i+1}{m}]$ , while in the literature, data from  $[\frac{i}{m}, \frac{i+2}{m}]$  is*

considered, that is the support of the pre-average values overlaps. This further reduces the impact of microstructure effects at the expense of a small negative influence of the dependency. For example, consider estimation of integrated volatility with pre-averaging function  $\lambda(s) = \sqrt{2\pi} \cos(\pi s)$ , when  $\sigma$  is a deterministic constant. In the case of overlapping pre-averaged values and for optimally chosen  $m$ , the respective asymptotic MSE is approximately  $10.21\tau n^{-1/2}\sigma^3$  by Lemma 6 in [Schmidt-Hieber \(2010\)](#). With the pre-averaging procedure described in (5.4) and again choosing  $m$  optimal for this procedure, we obtain an asymptotic MSE of approximately  $19.35n^{-1/2}\tau\sigma^3$  by some calculations analogously to the proof of that lemma. Despite of this drawback in performance, our definition of pre-averaged values allows us to construct a limiting distribution as it is done in Theorem 5.5, since the additional dependency would cause severe problems in several steps of the proof.

In (5.4), it becomes clear why the normalization of the pre-average function is chosen as in Definition 5.3 assuring  $\int \Lambda_i^2 = m^{-1}$ . This allows us to consider  $m\Lambda_i^2$  as a kernel. Therefore, the compensator of  $(\int \Lambda_i \sigma dW)^2$ , which is given by  $\int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i^2(s) \sigma_s^2 ds$ , equals approximately  $\sigma_{i/m}^2/m$ . This gives us reasons to approximate  $\int \psi_{t,h} \sigma^2$  by the weighted sum of the squared pre-average values:

We consider the test statistic

$$T_{n,t,h}^{(2)} := \sum_{i=0}^{m-1} \psi_{t,h}\left(\frac{i}{m}\right) \left( \bar{Y}_{i,m}^2 - \mathfrak{b}(Y, i, m, n) \right),$$

where

$$\mathfrak{b}(Y, i, m, n) := \tau^2 n^{-1} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i^2(s) ds$$

plays the role of an unavoidable bias correcting term compensating the expectation of the squared noise.

Note that

$$T_{n,t,h}^{(2)} = T_{\sigma^2} + 2T_{\sigma\tau} + T_{\tau^2}, \tag{5.5}$$

where

$$T_{\sigma^2} = T_{\sigma^2}(t, h, m) := \sum_{i=0}^{m-1} \psi_{t,h}\left(\frac{i}{m}\right) \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \sigma_s dW_s \right)^2,$$

$$\begin{aligned}
 T_{\sigma\tau} &= T_{\sigma\tau}(t, h, n, m) := \sum_{i=0}^{m-1} \psi_{t,h}\left(\frac{i}{m}\right) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \sigma_s dW_s \tau n^{-1/2} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^*, \\
 T_{\tau^2} &= T_{\tau^2}(t, h, n, m) := \sum_{i=0}^{m-1} \psi_{t,h}\left(\frac{i}{m}\right) \tau^2 n^{-1} \left( \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right)^2 - \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i^2(s) ds \right).
 \end{aligned}$$

Unfortunately, these terms depend on different powers of  $\sigma$ . Therefore, dividing by  $\hat{\sigma}_{t,h}^2$  as in Theorem 3.4 does not result in a limiting distribution not depending on unknown quantities. On the contrary, we have to approximate each of the terms separately, cf. Theorem 5.5. These results can afterwards be combined using Bonferoni's inequality to obtain confidence statements (cf. Corollary 5.9 and Chapter 6).

**THEOREM 5.5.** *Work under Model (5.1), and Assumptions 3.1 and 3.3. If  $\gamma < 1/2$ , also assume that  $u_m \ll m^{-1+2\gamma-\epsilon}$  for some  $\epsilon > 0$ . Further, suppose that there is a collection of estimators  $(\sigma_{t,h}^2)$  which are uniformly and almost surely consistent in the sense of (3.5) with  $n$  replaced by  $m$ .*

*Then, there exist sequences of Brownian motions  $W^{[n],m,i}$ ,  $i = 1, \dots, 3$ , so that we find the almost sure approximations*

$$\begin{aligned}
 a.) \quad & \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \frac{T_{\sigma^2} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \sqrt{\frac{2}{m}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,1} \right| = O(q_1^{n,m}), \\
 b.) \quad & \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \frac{T_{\sigma\tau}}{\hat{\sigma}_{t,h}} - \tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,2} \right| = O(q_2^{n,m}), \\
 c.) \quad & \sup_{(t,h) \in \mathcal{T}_m} w_h \left| T_{\tau^2} - \sqrt{2} \tau^2 \|\lambda\|_{L^2_{[0,1]}}^2 \frac{m^{3/2}}{n} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,3} \right| = O(q_3^{n,m}).
 \end{aligned}$$

Here, the specific approximation rates  $q_i^{n,m}$  tend to zero if  $n \geq m \rightarrow \infty$ , and are given by

$$\begin{aligned}
 q_1^{n,m} &= w_{l_m} l_m^{-1/2} \frac{\log(m)}{m} + w_{u_m} \left( u_m^\gamma \left( \frac{\log(m)}{m} \right)^{1/2} + u_m^{1/2} m^{-\gamma} \right) + s_m m^{-1/2} \frac{\log(1/l_m)}{\log \log(1/l_m)}, \\
 q_2^{n,m} &= w_{l_m} l_m^{-1/2} \frac{\log(m)}{n^{1/2}} + w_{u_m} u_m^\gamma \left( \frac{m \log(m)}{n} \right)^{1/2} + s_m m^{1/2} n^{-1/2} \frac{\log(1/l_m)}{\log \log(1/l_m)}, \\
 q_3^{n,m} &= w_{l_m} l_m^{-1/2} \frac{m \log(m)}{n}.
 \end{aligned}$$

A proof of the Theorem is given in Appendix B. Note that it heavily relies on the proof of Theorem 3.4 given in Appendix A.

**REMARK 5.6.** *The restriction  $u_m \ll m^{-1+2\gamma}$  is necessary for the approximation of  $\int \psi_{t,h} \sigma^2$  by  $\sum_i \int \psi_{t,h} \Lambda_i^2 \sigma^2$ , since the approximation error is given by  $u_m^{1/2} m^{-\gamma}$  up to logarithmic*

terms. Note that this is an unavoidable artifact coming from the pre-averaging procedure. Fortunately,  $m^{-1+2\gamma}$  is larger than the optimal regression bandwidth  $(\log m/m)^{1/(2\gamma+1)}$  for any  $\gamma > 0$  (cf. Remark 4.6). Therefore, if we know the smoothness of  $\sigma$ , the assumption  $u_m \ll m^{-1+2\gamma}$  is not restrictive. Moreover, if we do not know the exact smoothness, we may choose  $u_m$  as  $m^{-1+2\gamma_0}$  for some  $0 < \gamma_0 < 1/2$ . Then, the optimal bandwidth is included in  $\mathcal{T}_m$  for every  $\gamma \in (\gamma_0, (\gamma_0/(1-2\gamma_0)))$ . These intervals are the same as the regions of adaptivity given in Theorem 3.4 in Hoffmann et al. (2012).

REMARK 5.7. As described in various publications regarding high-frequency data, the problem of estimating in this scenario with  $n$  observations is as hard as estimating in the pure semimartingale model without microstructure noise from  $\sqrt{n}$  observations (cf. for example Gloter and Jacod (2001a,b), Jacod et al. (2009), or Reiß (2011)). In Theorem 5.5, this is reflected in the fact that the terms  $q_i^{n,m}$ ,  $i = 1, \dots, 3$ , as well as the terms  $T_{\sigma^2}$ ,  $T_{\sigma\tau}$ , and  $T_{\tau^2}$ , are only balanced if  $m$  is of order  $n^{1/2}$ .

REMARK 5.8. The estimator  $\sigma_{t,h}^2$  can be constructed analogously to Proposition 3.6 if  $m \asymp n^{1/2}$ .

To obtain confidence statements from Theorem 5.5, the following corollary is helpful:

COROLLARY 5.9. Under the Assumptions of Theorem 5.5, we obtain a.s. that

$$\begin{aligned}
 \sup_{(t,h) \in \mathcal{T}_m} w_h & \left| \left( T_{n,t,h}^{(2)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds \right) \right. \\
 & - \hat{\sigma}_{t,h}^2 \sqrt{\frac{2}{m}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,1} \\
 & - 2\hat{\sigma}_{t,h} \tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,2} \\
 & \left. - \sqrt{2} \tau^2 \|\lambda\|_{L^2_{[0,1]}}^2 \frac{m^{3/2}}{n} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,3} \right| = O(q_1^{n,m} + q_2^{n,m} + q_3^{n,m}).
 \end{aligned} \tag{5.6}$$

Again, the proof is postponed to Appendix B.



---

## CHAPTER 6

# APPLICATION TO FINANCIAL DATA

---

In order to apply the presented procedure to real data, parameters have to be chosen adequately to guarantee a good performance in a finite setting, as all results have been of asymptotic nature, so far. Furthermore, different difficulties concerning model violations occur in practice. In this chapter, we will address these problems and show exemplarily how to use our method for high-frequency financial data. Note that parts of Section 6.3 as well as the data description and the overview of existing literature in Section 6.4 coincide with [Sabel et al. \(2014\)](#).

## 6.1 MODEL DISCRETIZATION

In finance, the observable price of an underlying changes only when it is traded. Therefore, it is not possible to observe data in continuous time as it is done in Model (5.1), but only on a discrete grid. As described in Remark 5.1, we consider in the following a discrete version of our theoretical model, that is data  $Y_{i,n}$  from Model (5.2). In this model, the pre-averaged values are given by

$$\bar{Y}_{i,m} := \frac{1}{n} \sum_{j: \frac{j}{n} \in \left[ \frac{i}{m}, \frac{i+1}{m} \right)} \lambda_i\left(\frac{j}{n}\right) Y_{j,n}, \quad i = 0, \dots, m-1.$$

This definition coincides with the definition given in [Hoffmann et al. \(2012\)](#), up to the fact that intervals are overlapping there (cf. Remark 5.4).

## 6.2 PARAMETER OPTIMIZATION

Our procedure requires to choose a pre-averaging function  $\lambda$  and a pre-averaging “bandwidth”  $1/m$ . The specific choices will heavily influence the finite sample performance and

even the asymptotic behavior as we can see in the approximating terms in Theorem 5.5.

Before choosing  $m$  and  $\lambda$ , let us show how to construct confidence intervals using Theorem 5.5 and Corollary 5.9. Afterwards, we will give a reasonable choice for  $m$  based on minimizing the length of these intervals.

For  $z_{m,n,t} = \hat{\sigma}_{t,h}^2 \sqrt{\frac{2}{m}} + 2\hat{\sigma}_{t,h}\tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} + \sqrt{2}\tau^2 \|\lambda\|_{L^2_{[0,1]}}^2 \frac{m^{3/2}}{n}$ , we obtain by Theorem 5.5 and the triangle inequality that simultaneously over  $(t, h) \in \mathcal{T}_m$  and a.s.,

$$\begin{aligned}
 & w_h \left( \left| T_{n,t,h}^{(2)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds \right| - z_{m,n,t} \sqrt{2 \log \frac{\nu}{h}} \right) \\
 &= w_h \left( \left| \hat{\sigma}_{t,h}^2 \sqrt{\frac{2}{m}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,1} + 2\hat{\sigma}_{t,h}\tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,2} \right. \right. \\
 &\quad \left. \left. + \sqrt{2}\tau^2 \|\lambda\|_{L^2_{[0,1]}}^2 \frac{m^{3/2}}{n} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,3} \right| - z_{m,n,t} \sqrt{2 \log \frac{\nu}{h}} \right) (1 + o(1)) \\
 &\leq w_h \hat{\sigma}_{t,h}^2 \sqrt{\frac{2}{m}} \left( \left| \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,1} \right| - \sqrt{2 \log \frac{\nu}{h}} \right) (1 + o(1)) \\
 &\quad + 2w_h \hat{\sigma}_{t,h}\tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} \left( \left| \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,2} \right| - \sqrt{2 \log \frac{\nu}{h}} \right) (1 + o(1)) \\
 &\quad + w_h \sqrt{2}\tau^2 \|\lambda\|_{L^2_{[0,1]}}^2 \frac{m^{3/2}}{n} \left( \left| \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,3} \right| - \sqrt{2 \log \frac{\nu}{h}} \right) (1 + o(1)).
 \end{aligned} \tag{6.1}$$

Denote the  $(1 - \alpha)$ -quantile of

$$\sup_{(t,h) \in \mathcal{T}_m} w_h \left( \left| \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,1} \right| - \sqrt{2 \log \frac{\nu}{h}} \right)$$

by  $q_{1-\alpha}$ . Then, Bonferoni's inequality in combination with (6.1) yields that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \bigcup_{(t,h) \in \mathcal{T}_m} \left\{ w_h \left( \left| T_{n,t,h}^{(2)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds \right| - z_{m,n,t} \sqrt{2 \log \frac{\nu}{h}} \right) \geq z_{m,n,t} q_{1-\frac{\alpha}{3}} \right\} \right) \leq \alpha.$$

Thus, simultaneous level- $\alpha$  confidence intervals for  $\int_0^1 \psi_{t,h} \sigma^2$  are given by

$$\left[ T_{n,t,h}^{(2)} - z_{m,n,t} \left( \frac{q_{1-\frac{\alpha}{3}}}{w_h} + \sqrt{2 \log \frac{\nu}{h}} \right), T_{n,t,h}^{(2)} + z_{m,n,t} \left( \frac{q_{1-\frac{\alpha}{3}}}{w_h} + \sqrt{2 \log \frac{\nu}{h}} \right) \right]. \tag{6.2}$$

Note that the length of the confidence intervals is determined by  $h$  and  $z_{m,n,t}$ . Therefore, we would like to choose  $m$  locally (that is  $m_t(n)$  as a function in  $t$  and  $n$ ) as the largest integer lower than the minimizer of  $z_{m,n,t}$ . This is given by  $m_t(n) = \left[ c \frac{\hat{\sigma}_{t,h}}{\tau} n^{1/2} \right]$ , where  $c = c(\lambda) = \frac{2}{\sqrt{2+\sqrt{14}}} \|\lambda\|_{L^2_{[0,1]}}^{-1}$  and  $[\cdot]$  is the Gauss bracket. As we need a global choice of

$m = m(n)$ , that is one not depending on  $t$ , we define in the following

$$m := \left[ c \left( \int \widehat{\sigma^2} \right)^{1/2} \tau^{-1} n^{1/2} \right],$$

where  $\int \widehat{\sigma^2}$  is any estimator of the integrated volatility. We will use the one presented in [Hoffmann et al. \(2012\)](#), here.

Different choices for  $\lambda$  are discussed in [Schmidt-Hieber \(2010\)](#), and more extensively in [Sabel et al. \(2014\)](#). In our case, minimizing  $z_{m,n,t}$  in  $\lambda$  is equivalent to minimizing  $\|\lambda\|_{L^2_{[0,1]}}$  in the class of all pre-average functions. This can be done via calculus of variation, and we obtain  $\sqrt{2}\pi \cos(\pi\bullet)$  as the optimal pre-average function. In this case,

$$c = \frac{2}{\sqrt{2+\sqrt{14}}}\pi^{-1} \approx 0.12.$$

Let us conclude with some remarks:

REMARK 6.1. *The choice of  $m$  assures a minimal length of the resulting confidence intervals. However, in practice, people might be interested in being robust against microstructure effects by choosing  $m$  smaller than optimal. This decreases the impact of microstructure noise, while it increases the variance of the signal part. In our setting, it can simplify the implementation of the test: Consider for example  $m \ll n^{1/2}$ . Then, the terms  $T_{\sigma\tau}$  and  $T_{\tau^2}$  are negligible, such that the testing procedure can be applied in complete analogy to the non-high-frequent semimartingale case.*

REMARK 6.2. *Dividing  $\alpha$  into three equal parts as done above is not necessary for Bonferroni's inequality. Instead, one can choose  $\alpha_i > 0$ ,  $i = 1, \dots, 3$  with  $\alpha_1 + \alpha_2 + \alpha_3 = \alpha$  to balance the three summands of  $z_{m,n,t}$ . However, there is no closed form for the quantiles, making optimization difficult.*

REMARK 6.3. *The choice  $\sqrt{2}\pi \cos(\pi\bullet)$  as optimal pre-average function is due to asymptotic considerations. Especially for small sample sizes, this function might be outperformed by other ones. In [Sabel et al. \(2014\)](#), it is argued that  $\sqrt{8/3}\pi \sin(2\pi\bullet)$  is a better choice for estimation if one considers FGBL data. However, this might partly be a result of the overlapping pre-average procedure as described in [Remark 5.4](#).*

### 6.3 DIFFICULTIES IN PRACTICE

While the semimartingale model (Model (1.1)) describes non-high-frequent data quite well, the high-frequency model (5.2) can only be viewed as an idealized model which is close to reality in many but not in all purposes.

In order to handle real data, there are some problems to overcome:

1. *Rounding errors:* Due to discreteness of prices, rounding errors are inevitable in practice. Unfortunately, these errors are not independent of the price itself and thereby not covered by the microstructure noise model. However, we show in Table 6.1 that these errors are absorbed by pre-averaging as well.
2. *Discontinuities:* While we assume the latent log-price  $X$  to be continuous in time, in practice, micro- or macroeconomic announcements may cause jumps in the price process. The presence of such jumps is discussed in [Ait-Sahalia and Jacod \(2009\)](#), [Bollerslev and Todorov \(2011\)](#), [Ait-Sahalia et al. \(2012\)](#), and the references therein. Unfortunately, the presence of jumps is indeed a very delicate problem: Figure 6.2 shows that jumps severely change the reconstruction of  $\sigma^2$  in spot volatility estimation using the adaptive spot volatility estimator (ASVE) given in [Sabel et al. \(2014\)](#). As the estimator is based on a Haar wavelet deconvolution, that is on integrals over (dyadic) intervals, it seems reasonable that jumps disturb our aim to find confidence bands for integrals over different intervals as well. In order to eliminate jumps in the price, we apply a pre-processing procedure to the data, described in Subsection 6.3.1. Note that this should be viewed as a data cleaning method, as we attach more importance to practical purposes than to mathematical accuracy. The performance of this procedure is presented in Table 6.1.
3. *Trading times:* We have to deal with data recorded at non-equidistant time points. One possibility to 'convert' data into the equispaced framework of model (5.2) is to sub-sample the process, that is to sample for example every tenth second. However, this clearly results in a loss of information. Therefore, we propose another method by defining different time schemes distinguishing between real time and tick time in Subsection 6.3.2.



FIGURE 6.1: FGBL data of November 2nd, 2007 and magnification of a small time interval around 1.30 p.m., where multiple consecutive jumps occur in the price process.

### 6.3.1 MODEL VIOLATIONS

As discussed above, the continuity assumption in the model is often too strict in reality. The most natural way to include a jump component into the model is to allow for non-continuous semimartingales. Estimation of the integrated volatility under microstructure noise and jumps has been considered for instance in [Podolskij and Vetter \(2009\)](#).

In order to correct for jumps, we adopt a rather practical point of view here. In fact, looking at financial data, relevant jumps seem to occur very irregularly. Occasionally, there are isolated jumps, while quite rarely, jump clusters of very short duration appear (cf. Figure 6.1). Therefore, our aim in this section is a hands-on approach to detect and to remove possible jumps as a pre-processing of the data.

As usual, we model jumps as an additive càdlàg jump process  $(J_t)_{t \in [0,1]}$ . If jumps are present, any estimator ignoring these jumps will reconstruct the pointwise sum of the spot volatility plus the jump process  $t \mapsto (J_t - J_{t-})^2$ , where  $J_{t-}$  denotes the left limit of  $J$  at time point  $t$ . Note that  $(J_t - J_{t-})^2$  is either zero or produces a spike depending on whether there is a jump at time point  $t$  (cf. Figure 6.2, Panel 2). As our results can be seen as confidence bands for integrals of such an estimator, the bumps in the reconstruction, which are results of jumps, transfer directly to these integrals. In order to separate spot volatility and jump part, we apply the following method:

Let  $m_1 = \lfloor n^{3/4} \rfloor$  and  $\lambda$  be a pre-average function. For  $r = \frac{n}{m_1}, \dots, n - \frac{n}{m_1}$ , define

$$Q_r := \sum_{j=r-\frac{n}{m_1}}^{r+\frac{n}{m_1}} \lambda\left(\frac{1}{2} + (j-r)\frac{m_1}{2n}\right) Y_{j,n}. \quad (6.3)$$

If there is no jump in  $[r - \frac{n}{m_1}, r + \frac{n}{m_1}]$ , then  $Q_r = o_P(1)$ . Under the alternative, that is there is a jump with height  $\Delta_r$  at  $r/n$ , we obtain  $Q_r = O_P(\Delta_r)$ . Note that by some CLT argument,  $Q_r$  is approximately Gaussian distributed. Therefore, we may apply a procedure mimicking a local  $t$ -test:

1. We partition the set  $\{Q_r : r = \frac{n}{m_1}, \dots, n - \frac{n}{m_1}\}$  into blocks of length  $n^{1/2}$ .
2. For each of these blocks, we compute the mean  $\hat{\mu}$  and the standard deviation  $\widehat{sd}$ .
3. For each  $Q_r$  in a block, we compare  $(Q_r - \hat{\mu})/\widehat{sd}$  with a fixed threshold  $t$ , rejecting values larger than  $t$ . Here, simulations show that  $t = 3.5$  performs well for FGBL data (cf. Table 6.1).

Afterwards, we reject those pre-averaged value  $\bar{Y}_{i,m}$ , whose supports intersect the support of a  $Q_r$  which is rejected by the procedure. The rejected pre-averaged values are replaced by the average of the nearest neighbors which are not rejected.

This procedure ensures that isolated jumps are detected. However, we sometimes observe in real data that there are consecutive jumps within a short time period (cf. FGBL data of November 2nd, 2007 in Figure 6.1 as an example). This may result in acceptance of the hypothesis that there is no jump, since every single jump might be not high enough in comparison to the estimated variance of  $Q_r$ . However, it is high enough to disrupt any inference procedure severely. To overcome this problem, we introduce a second data cleaning procedure which directly compares increments of the observations and is more suitable to detect jump clusters.

From our data sets, we find that the level of the microstructure noise, that is  $\tau$ , remains almost constant over a day. Thus, to explain the test, we might assume that  $\tau$  is constant. Then,

$$Y_{i,n} - Y_{i-1,n} = \tau(\epsilon_{i/n} - \epsilon_{(i-1)/n}) + O_P(n^{-1/2}) \approx \tau(\epsilon_{i/n} - \epsilon_{(i-1)/n}),$$

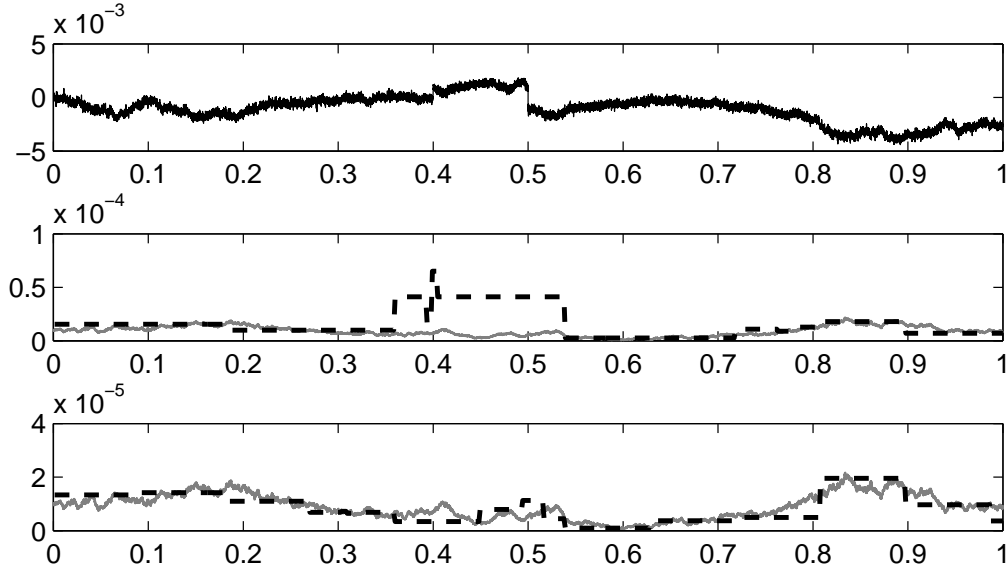


FIGURE 6.2: Simulated data (Panel 1) coming from the Heston model with parameter choices given in (4.8) for  $n = 15,000$  and true signal-to-noise-ratio  $\approx 15 - 20$  with two artificial jumps at 0.4 and 0.5, the true spot volatility function (gray, solid line, Panel 2 and 3) and the adaptive spot volatility estimator (ASVE) from [Sabel et al. \(2014\)](#) neglecting the presence of jumps (dashed line, Panel 2) and automatically finding and correcting the jumps (dashed line, Panel 3).

if there is no jump. Additionally, we observe that the distribution of the noise is well-concentrated around zero, justifying the assumption of Gaussian (or even sub-Gaussian) noise. Using Corollary 2.1 in [Li and Shao \(2002\)](#), we find the following behavior regarding extreme values:

$$\lim_{n \rightarrow \infty} \mathbb{P}(\max_{i=2, \dots, n} (\epsilon_{i/n} - \epsilon_{(i-1)/n})^2 \leq 4\tau^2 \log n) = 1.$$

Consequently, we identify the difference  $Y_{i,n} - Y_{(i-1),n}$  as due to a jump, if the squared increment exceeds  $4\tau^2 \log n$ . Note that the latter procedure is much less powerful for isolated jumps than the first one, since it cannot detect jumps of size  $o_P(\log n)$ . However, application proves that this data cleaning is very helpful to obtain robust results.

Figure 6.2 exemplarily displays the effects of jumps and the success of the procedure for one simulated price, following the Heston model (cf. (4.7)). Here, we want to emphasize again that the results directly transfer from estimation to confidence intervals, so that Figure 6.2 should be seen as a visualization. More thoroughly, the performance is displayed in Figure 6.3. Here, we simulated high-frequency data with constant volatility  $\sigma^2 = 10^{-5}$ ,

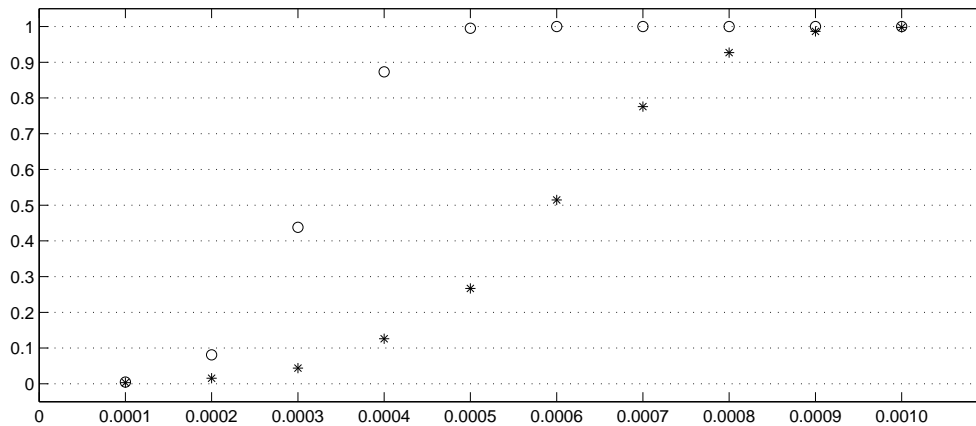


FIGURE 6.3: Success of the jump detection and impact of jumps: Simulation study based on 1,000 repetitions of Model (5.1) with constant volatility  $\sigma^2 = 10^{-5}$ ,  $n = 10,000$ ,  $\tau = 5 \cdot 10^{-5}$ , and one jump added to the log-price at 0.5. The jump height varies along the  $x$ -axis. The circles correspond to the estimated probabilities at which a jump of certain height is detected. Neglecting the presence of jumps, the asterisks represent the estimated probability of falsely rejecting (at level  $\alpha = 10\%$ ) the hypothesis that the volatility is constant, conditionally on the event that the hypothesis would not have been rejected without the added jump.

Percentage of rejection	pure price	rounded price	price with jump
with jump detection	0.119	0.119	0.125
without jump detection	0.112	0.112	0.217

TABLE 6.1: Simulated actual levels (for nominal level  $\alpha = 10\%$ ) based on 1,000 repetitions of Model (5.1) with constant volatility  $\sigma^2 = 10^{-5}$ ,  $n = 10,000$ , and  $\tau = 5 \cdot 10^{-5}$ . In the fourth column a jump of height 0.0004 is added to the log-price at 0.5.

$n = 10,000$ , and noise level  $\tau = 5 \cdot 10^{-5}$ . The values are chosen to mimic the signal-to-noise ratio of FGBL data, which we use later on. Afterwards we add a jump at 0.5 to the data. The different heights of the jump are displayed at the  $x$ -axis, while the corresponding circles show how often the jump was found by the procedure. Moreover, we provide some information about the influence of an ignored jump: Skipping the jump detection, we investigate the percentage of false rejections (at level  $\alpha = 10\%$ ) of the hypotheses that the volatility is constant, conditionally on the event that the hypothesis would not have been rejected without the jump. The estimated values are given by the asterisks in Figure 6.3. As a result, we observe that our procedure performs well whenever the jump size is large enough to severely inflate the level of the testing procedure.



The impact of the jump detection procedure is analyzed in Table 6.1. Again, we choose a constant volatility  $\sigma^2 = 10^{-5}$ ,  $n = 10,000$ , and noise level  $\tau = 5 \cdot 10^{-5}$ , and set  $\alpha = 10\%$ . As a first result, we observe that the multiscale test without jump detection and in the absence of jumps performs pretty well in keeping the nominal level. The small discrepancy can be explained by the comparably small sample size ( $m \approx 750$  in this setting). Furthermore, we observe that the confidence procedure is not influenced too badly by the jump detection compared to the impact of ignored jumps given in the fourth column, where a jump at 0.5 of height 0.0004 is added to the data.

Furthermore, we investigated the effect of rounding errors here. To this end, the simulated log-prices  $X$  are transformed to prices  $100 \log(X)$ . These are rounded to two decimal digits and are transformed back afterwards. Again, this transformation is chosen to mimic FGBL data. Regarding the values in Table 6.1 we may conclude that rounding effects are negligible in this setting.

### 6.3.2 TIME SCHEMES

It has been noticed in the econometric literature that an increase in volatility might be due to different reasons. One explanation would be that there are larger price changes. Alternatively, the volatility of course also increases as price changes are of the same size and only the number of trades per time interval goes up (cf. for example [Ederington and Lee \(1995\)](#), Section IV.B). Disentangling the different explanations is quite difficult without an underlying mathematical concept. Nevertheless, determining the source of an increase in volatility is clearly of importance.

A more rigorous investigation of this problem leads to the definition of different notions of time (for instance in [Dahlhaus and Neddermeyer \(2013\)](#)). In this thesis, we discuss the most prominent examples: real time and tick time (sometimes also referred to as clock time and transaction time, respectively).

Volatility in real time is appealing as it seems very intuitive. In tick time successive ticks are treated as one time unit. By definition, this time scheme does not depend on the speed at which successive trades occur. Consequently, volatility in tick time is independent of the trading intensity and hence measures the volatility of the price changes only. As the trading speed can be estimated directly from the ticks, we argue in this section that tick time volatility is the more natural object. A drawback of tick times is that there is no

straightforward extension of the concept to multivariate processes, where price changes might occur non-synchronously.

Let us clarify the connection between both time schemes in more detail. Denote by  $t_i$ ,  $i = 0, \dots, n$  the ordered ( $t_0 < t_1 < t_2 < \dots < t_n$ ) sample of trading times. Then, for  $i < j$  the time between  $t_i$  and  $t_j$  equals  $\frac{j-i}{n}$  time units in tick time and  $t_j - t_i$  time units in real time. With this notation, the tick time model is given by

$$Y_{i,n}^T = X_{t_i} + \epsilon_{i,n}, \quad i = 1, \dots, n. \quad (6.4)$$

Inspired by the classical high-frequency framework, we think about the trading times as an array, that is  $t_i = t_{i,n}$ , where the sampling rate gets finer for increasing  $n$ . Define the trading intensity  $\theta$  at time  $t$  as

$$\theta(t) = \lim_{n \rightarrow \infty} \frac{\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{[t-\delta_n, t+\delta_n]}(t_i)}{2\delta_n} (t_n - t_0), \quad (6.5)$$

provided this limit exists and is unique for any sequence  $\delta_n \rightarrow 0$  and  $\delta_n n \rightarrow \infty$ .

As an example, consider the following toy model: Assume that  $\sigma$  is deterministic and there exists a deterministic, differentiable, and strictly monotone function  $h : [0, 1] \rightarrow [0, 1]$  with  $h(i/n) = t_{i,n}$ . Note that in this setting,  $\theta$  is deterministic as well and given by the derivative of  $h^{-1}$ .

Let  $\sigma_{RT}$  denote the original (real time) volatility. Recall that under tick time, we consider successive trading times as equidistant. Therefore, the tick time volatility  $\sigma_{TT}$  satisfies for all  $i = 1, \dots, n$

$$\int_0^{i/n} \sigma_{TT}(h(s)) dW_s = \int_0^{h(i/n)} \sigma_{RT}(s) dW_s =_{\mathcal{L}} \int_0^{i/n} \sqrt{h'(s)} \sigma_{RT}(h(s)) dW_s$$

in law. Thus, the first and the latter integrand are (roughly) equal, that is  $\sigma_{TT}^2(h(s)) = h'(s) \sigma_{RT}^2(h(s))$ . Rewriting this, we obtain

$$\theta \sigma_{TT}^2 = \sigma_{RT}^2, \quad (6.6)$$

cf. also [Dahlhaus and Neddermeyer \(2013\)](#), Section 4. This formula clarifies the connection between tick time and real time volatility.

In order to make inference about the real time volatility directly from tick data, we have

to construct artificial observations by recording the price each tenth second, for example. This method leads to a loss of information if there are many ticks in one time interval. Additionally, nonparametric multiscale inference about the trading intensity  $\theta$  can be done using the results in [Dümbgen and Walther \(2008\)](#). Since there is no additional ill-posedness in this problem, detection rates are comparably fast here. Thus, in view of (6.6), it seems more informative to consider real time spot volatility as the product of  $\sigma_{TT}^2$  and  $\theta$ .

Furthermore tick time volatility seems to be smoother than its real time counterpart making inference easier: To analyze volatilities of Euro-Bund-Futures on all days in 2007 (for a description of the data, cf. also Section 6.4), we estimate tick time volatility using the estimator presented in [Sabel et al. \(2014\)](#) and the real time counterpart via its product representation. We use Haar wavelets and hence obtain piecewise constant reconstructions. As a measure for the oscillation behavior of the volatility, we take the sum of squared jump sizes of the reconstructions for every of these days. In average, for tick time spot volatility, this gives  $9.68 \cdot 10^{-11}$  per day, while for real time volatility, the corresponding value is  $1.98 \cdot 10^{-10}$ .

To summarize, the tick time volatility is the quantity measuring the volatility of the price changes. Furthermore, it seems to be smoother and therefore easier to handle. Moreover, we can transfer results about tick time volatility via (6.6) to results about real time volatility. For these reasons, we restrict ourselves throughout the following to spot volatility in tick time.

## 6.4 REAL DATA EXAMPLE

We analyze the spot volatility of Euro-Bund-Futures (FGBL) using tick data from EUREX database in 2007. The underlying is an 100,000 Euro debt security of the German Federal Government with coupon rate 6% and maturity 8.5 – 10.5 years. The price is given in percentage of the par value. The tick times are recorded with precision of 10 milliseconds. The minimum price change is 0.01% (one basis point), corresponding to 10 Euro. For more information on FGBL, cf. [Eurex Frankfurt AG \(2005\)](#).

In our datasets, the number of trades per day varies among 10,000 and 30,000. Observations which are not due to trading are removed from the sample. If there are different FGBL contracts at a time referring to different expiration days, we only consider these

belonging to the next possible date. Trading takes place from 8:00 a.m. until 7:00 p.m. Central European Time (CET). Here, we restrict ourselves to observations between 9 a.m. and 6 p.m. CET. Outside this period, trading intensity is often too low to make use of a high-frequency setting.

During business hours, FGBL prices fit well as an example for high-frequency data. On the one hand, trading is very liquid due to low transaction costs and high trading volume. In average, the holding period is less than two days (cf. [Dorflleitner \(2004\)](#), Figure 4). On the other hand, microstructure effects are present and simple quadratic variation techniques fail as indicated in Figures [5.1](#) and [1.1](#).

When macroeconomic announcements or events occur, one expects an increase in volatility due to the uncertainty of the market. There has been a large body of literature in economics devoted to this subject. Nevertheless, up to now, there seems to be no general consensus quantifying how much the volatility is affected by public announcements. [Ederington and Lee \(1993, 1995\)](#) claim that volatility is substantially higher for a few minutes after the announcement and is still visible in the data for several hours. They also find evidence that volatility is slightly elevated for some minutes before an announcement. They conclude that macroeconomic announcements are the driving force for volatility. In contrast, in the seminal paper [Andersen and Bollerslev \(1998\)](#) daily volatility patterns are found to explain most of the spot volatility behavior, while public announcements have only a secondary effect on overall volatility. In a recent study, [Lunde and Zebedee \(2009\)](#) focus on the effects of US monetary policy events on volatility of US equity prices. In accordance with previous work, they conclude that there are spikes in the volatility around macroeconomic announcements, lasting for approximately 15 minutes. In [Jansen and de Haan \(2006\)](#) effects of certain European Central Bank (ECB) announcements on price changes and volatility are studied. Although these papers deal with volatility on relatively short time intervals, none of them accounts for microstructure effects.

As an example, we investigate the volatility within those days in 2007 when the president of the European Central Bank, Jean-Claude Trichet, announced possible changes of the key interest rates. We suppose that there is a strong connection between FGBL prices and these events since government bonds are closely related to this rate. However, it is not clear, if and how much the volatility is effected as well.

The announcements take place regularly once per month and consist of the announcement itself at 1.45 p.m. followed by a press conference at 2.30 p.m. which lasts exactly one hour.

In advance, between 20 and 62 financial experts were interviewed by the news and data agency Bloomberg L.P. about their prediction of the change. In the following, we will use the sample standard deviation of their answers as an indicator for market uncertainty.

In a first step, let us exemplarily turn to one of these days, May 10th, 2007: In Figure 6.4, the FGBL log-price and the corresponding pre-averaged values are displayed. Furthermore, the last panel of the plot indicates regions of significant increase of the volatility (in the spirit of Section 4). Here, we set  $l_m = 20/m \approx 0.0225$  and  $u_m = 50/m \approx 0.0562$  referring to time intervals of approximately 6 to 15 minutes in average because we are interested in very local statements. The quantiles of the limit distribution are again obtain via 10,000 Monte Carlo simulations. We observe that although market uncertainty was 0 for that day, there is a highly significant increase at the time of the announcement, where the hypothesis of non-increase can be rejected at level 7%.

Day	Market uncertainty	Change of the key interest rate in basis points	Minimal level $\alpha$ , s.t. there is a rejection
January 11th, 2007	0	0	-
February 8th, 2007	0	0	0.017
March 8th, 2007	0	+0.25	-
April 12th, 2007	0	0	-
May 10th, 2007	0	0	0.211
June 6th, 2007	0	+0.25	-
July 5th, 2007	0	0	0.362
August 2nd, 2007	0.05	0	0.066
September 6th, 2007	0.1	0	-
October 4th, 2007	0.03	0	-
November 8th, 2007	0	0	-
December 6th, 2007	0	0	0.4314

TABLE 6.2: For those days in 2007 where a ECB press conference took place, minimal levels allowing to detect a significant overshoot of the integrated volatility on some interval containing 1.45 p.m. over the daily integrated volatility. “-” indicates that the respective value is above 50%. Note that the levels has to be understood as levels which are kept simultaneously over one day and must not be confused with levels for a simultaneous test for all days together.

More thoroughly, we investigated these days in Table 6.2 and compare them to all days in 2007 in Table 6.3: Here, we are interested in the question if there are any intervals containing 1.45 p.m. such that the integrated volatility on the interval is significantly larger than the integrated volatility over the whole day. As a benchmark, the same procedure is applied to all days in 2007. For each day, we applied our procedure separately. This gives

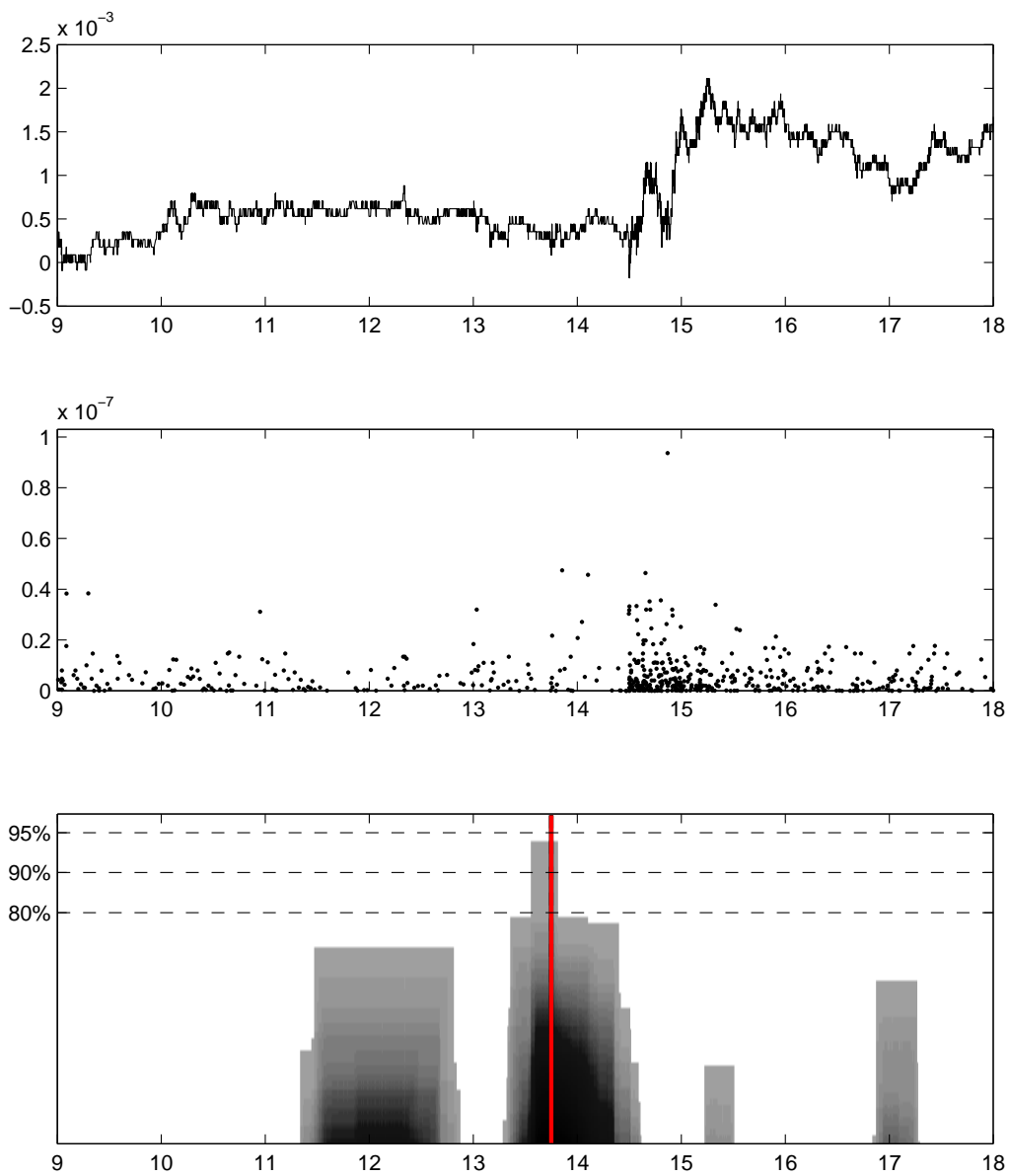


FIGURE 6.4: Logarithm of the FGBL price of May 10th, 2007 (panel 1), jump-corrected pre-averaged values (panel 2), and areas of significant increase ( $x$ -axis of panel 3) for different levels of significance ( $y$ -axis). The vertical red line at 13.75 refers to the announcement of not changing the key interest rate. Again, every interval of increase is indicated gray and darker regions only refer to intersections of these intervals. In this example,  $n = 16,232$ ,  $m = 890$ , and we chose  $l_m = 20/m$  and  $u_m = 50/m$ .

us confidence statements which hold simultaneously over one day but not simultaneously over all days.

We observe that at least at some of the considered days, there is a significant peak in the volatility, which cannot be explained by a trend occurring on an average day. However, we find no reliable connection between market uncertainty and appearance of these peaks.

---

Level $\alpha$	Proportion regarding all days in 2007	Proportion regarding all days with key interest rate announcement
5%	0.012	0.083
10%	0.021	0.167
15%	0.029	0.167
20%	0.045	0.167

---

TABLE 6.3: Proportion of those days in 2007, on which integrated volatility is significantly higher at some interval containing 1.45 p.m. than the average of the day. Again, significance has to be understood in the sense of confidence bounds being kept simultaneously over one day.





---

# CHAPTER 7

# OUTLOOK

---

## 7.1 MULTIDIMENSIONAL EXTENSION

In practice, one is often interested in the key figures of different time series, which are correlated with each other. This situation appears for example in portfolio management, where often a large number of shares is bundled and investors are not only interested in the respective volatilities but also in the local covariance between the prices, the so-called “covolatility”. Estimation of integrated covolatility or even the whole integrated covariance matrix has recently drawn a lot of attention (cf. the publications of [Hayashi and Yoshida \(2005\)](#), [Ait-Sahalia et al. \(2010\)](#), [Christensen et al. \(2010\)](#), [Barndorff-Nielsen et al. \(2011\)](#), [Zhang \(2011\)](#), [Bibinger \(2011\)](#), and [Bibinger et al. \(2014\)](#), among others). Moreover, in [Bibinger and Reiß \(2014\)](#) and [Sabel et al. \(2014\)](#), pointwise estimators of the covolatility are presented.

In this chapter, we like to give a short heuristic explanation of how to extend the multiscale test from this thesis to inference on covolatility. Here, we restrict ourselves to the two-dimensional case and to the continuous-time model (5.1). These assumptions are quite common in literature, cf. for example [Bibinger and Reiß \(2014\)](#).

Suppose that we observe two different time series  $Y^{(1)}$  and  $Y^{(2)}$  in the continuous-time model (5.1) with diffusions  $\sigma^{(1)}$  and  $\sigma^{(2)}$ . We assume that the driving Brownian motions are correlated with time-dependent correlation  $\rho_t$  and that both noise processes are independent of each other and of both signal parts.

We are interested in the local behavior of  $t \mapsto \sigma_t^{(1)} \sigma_t^{(2)} \rho_t$ . To this end, we compute the pre-averaged vectors  $\bar{Y}_{i,m}^{(1)}$  and  $\bar{Y}_{i,m}^{(2)}$  in a first step analogously to Chapter 5. Then, we consider the test statistic

$$T_{n,t,h}^{(3)} = \sum_{i=0}^{m-1} \psi_{t,h}(\frac{i}{m}) \bar{Y}_{i,m}^{(1)} \bar{Y}_{i,m}^{(2)}$$

where  $\psi$  is again any regular test function designed for the specific problem and  $(t, h) \in \mathcal{T}_m$ . Note that in this case, no bias correction is necessary, since both noise terms are assumed to be independent.

This test statistic approximates the integral

$$\int_t^{t+h} \psi_{t,h}(s) \sigma_s^{(1)} \sigma_s^{(2)} \rho_s ds.$$

Moreover, by the same arguments as given in the proof of Theorem 5.5, one can prove that the approximation error divided by some pre-estimator for  $(\sigma_t^{(1)} \sigma_t^{(2)} \rho_t)$  again behaves like a properly scaled Brownian integral uniformly in  $\mathcal{T}_m$ .

Proceeding as in Chapter 4, this result can be used to obtain confidence intervals for features of the covolatility of these prices.

## 7.2 INVESTIGATION OF THE LEVERAGE EFFECT

The (financial) leverage effect provides important tools to control key figures of a company such as return on equity (that is earnings per invested monetary unit): If the costs of further liabilities are smaller than the costs of equity, one can raise the return on equity by borrowing additional money and investing it.

In the economic literature, it is argued that this effect directly implies that the correlation between the price and the volatility is negative: A decrease of the price results in an increase of the debt-to-equity ratio which will not be compensated because it is favorable due to the leverage effect. However, this also increases the risk of the company and therefore the volatility of the share (cf. for example Yu (2005) and the references therein).

Thus, a non-negative correlation implies the absence of the leverage effect. Starting with the papers of Black (1976) and Christie (1982), there have been many empirical studies testing this correlation. However, these are limited from a statistical point of view, since to the best of our knowledge there are no simultaneous and localized tests in a fairly general model like Model (5.1).

Using the methods introduced in this thesis, one can test simultaneously on the size of local in- and decreases of the volatility. By similar (and even easier) arguments as in Theorem

5.5, one can approximate the limit distribution of

$$\sup_{(t,h) \in \mathcal{T}_m} w_h \hat{\sigma}_{t,h}^{-1} \left| \sum_{i=0}^{m-1} \psi_{t,h}(\frac{i}{m}) \bar{Y}_{i,m} \right|,$$

under the hypothesis that the unobserved log-price  $X$  is constant. Here,  $\psi$  is again the  $L^2$ -normalized derivative of a piecewise differentiable kernel, for example  $\psi = \mathbb{I}_{[0,1/2]} - \mathbb{I}_{[1/2,1]}$ .

This multiscale test provides confidence intervals for the size of in- and decreases in the price process. Combining both test (for example by Bonferoni's inequality), we obtain significant statements about the sign of the correlation between volatility and price.



---

## BIBLIOGRAPHY

---

- Y. Ait-Sahalia and J. Jacod. Testing for jumps in a discretely observed process. *Ann. Statist.*, 37:184–222, 2009.
- Y. Ait-Sahalia, P. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18:351–416, 2005.
- Y. Ait-Sahalia, J. Fan, and D. Xiu. High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.*, 105:1504–1517, 2010.
- Y. Ait-Sahalia, J. Jacod, and J. Li. Testing for jumps in noisy high frequency data. *J. Econometrics*, 168:207–222, 2012.
- T. G. Andersen and T. Bollerslev. Deutsche Mark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *The Journal of Finance*, 53:219–265, 1998.
- F. Bandi and J. Russell. Microstructure noise, realized variance, and optimal sampling. *Rev. Econom. Stud.*, 75:339–369, 2008.
- M.T. Barlow and M. Yor. Semi-martingale inequalities via the Garsia-Rodemich-Rumsey lemma and application to local times. *J. Funct. Anal.*, 49:198–229, 1982.
- O. Barndorff-Nielsen, P. Hansen, A. Lunde, and N. Shephard. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.
- O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162:149–169, 2011.

## BIBLIOGRAPHY

---

- M. Bibinger. Efficient covariance estimation for asynchronous noisy high-frequency data. *Scand J. Statist.*, 38:23–45, 2011.
- M. Bibinger and M. Reiß. Spectral estimation of covolatility from noisy observations using local weights. *Scand J. Statist.*, 41(1):23–50, 2014.
- M. Bibinger, N. Hautsch, P. Malec, and M. Reiß. Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. *Ann. Statist.*, to appear, 2014.
- F. Black. Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pages 177–181, 1976.
- F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.
- T. Bollerslev and V. Todorov. Estimation of jump tails. *Econometrica*, 79:1727–1783, 2011.
- D.L. Burkholder. Martingale transforms. *Ann. Math. Stat.*, 33:887–893, 1966.
- T. Cai, A. Munk, and J. Schmidt-Hieber. Sharp minimax estimation of the variance of Brownian motion corrupted with Gaussian noise. *Statist. Sinica*, 20:1011–1024, 2010.
- K. Christensen, S. Kinnebrock, and M. Podolskij. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159:116–133, 2010.
- A.A. Christie. The stochastic behavior of common stock variances. *Journal of Financial Economics*, 10:407–432, 1982.
- P. Čížek, W. Härdle, and R. Weron. *Statistical Tools for Finance and Insurance*. Springer-Verlag, 2005.
- R. Dahlhaus and J.C. Neddermeyer. On-line spot volatility-estimation and decomposition with nonlinear market microstructure noise models. *ArXiv e-prints*, 2013. arXiv:1006.1860v4.
- F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300:463–520, 1994.

- F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Math. Ann.*, 312:215–250, 1998.
- D.L. Donoho. Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probability Theory and Related Fields*, 99(2):145–170, 1994.
- G. Dorfleitner. How short-termed is the trading behaviour in Eurex futures markets? *Applied Financial Economics*, 14:1269–1279, 2004.
- L. Dümbgen and V.G. Spokoiny. Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29:124–152, 2001.
- L. Dümbgen and G. Walther. Multiscale inference about a density. *Ann. Statist.*, 36:1758–1785, 2008.
- L.H. Ederington and J.H. Lee. How markets process information: new releases and volatility. *The Journal of Finance*, 48:1161–1191, 1993.
- L.H. Ederington and J.H. Lee. The short-run dynamics of the price adjustment to new information. *Journal of Financial and Quantitative Analysis*, 30:117–134, 1995.
- Eurex Frankfurt AG. *Interest Rate Derivatives - Complete Your Picture in Fixed Income Investment Management*, 2005.
- J. Fan and Y. Wang. Spot volatility estimation for high-frequency data. *Statistics and its Interface*, 1:279–288, 2008.
- E. Giné, V. Koltchinskii, and L. Sakhanenko. Kernel density estimators: convergence in distribution for weighted sup-norms. *Probab. Theory Relat. Fields*, 130:167–198, 2004.
- A. Gloter and M. Hoffmann. Stochastic volatility and fractional Brownian motion. *Stochastic Process. Appl.*, 113:143–172, 2004.
- A. Gloter and J. Jacod. Diffusions with measurement errors. i. Local asymptotic normality. *ESAIM Probab. Stat.*, 5:225–242, 2001a.
- A. Gloter and J. Jacod. Diffusions with measurement errors. ii. Optimal estimators. *ESAIM Probab. Stat.*, 5:243–260, 2001b.
- J. Hasbrouck. Assessing the quality of a security market: a new approach to transaction-cost measurement. *The Review of Financial Studies*, 6:191–212, 1993.

## BIBLIOGRAPHY

---

- T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11:359–379, 2005.
- S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- M. Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79:135–163, 1999.
- M. Hoffmann, A. Munk, and J. Schmidt-Hieber. Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Ann. Inst. H. Poincaré Probab. Stat.*, 48(4):1186–1216, 2012.
- J. Jacod and P. Protter. *Discretization of Processes*. Springer-Verlag, 2011.
- J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, 2nd edition, 2003.
- J. Jacod, Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Process. Appl.*, 119(7):2249–2276, 2009.
- D. Jansen and J. de Haan. Look who’s talking: ECB communication during the first years of EMU. *International Journal of Finance and Economics*, 11:219–228, 2006.
- V.I. Koltchinskii. Komlos-Major-Tusnady approximation for the general empirical process and Haar expansions of classes of functions. *Journal of Theoretical Probability*, 7:73–118, 1994.
- J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv’s, and the sample df. i. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 32:111–131, 1975.
- J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv’s, and the sample df. ii. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 34:33–58, 1976.
- L. Le Cam and L. O. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag, 1990.



- W.V. Li and Q.M. Shao. A normal comparison inequality and its applications. *Probab. Theory Relat. Fields*, 122:494–508, 2002.
- M. Low. On nonparametric confidence intervals. *Ann. Statist.*, 24(6):2273–2717, 1997.
- A. Lunde and A. A. Zebedee. Intraday volatility responses to monetary policy events. *Financial Markets and Portfolio Management*, 23(4):383–299, 2009.
- A. Madahavan. Market microstructure: a survey. *Journal of Financial Markets*, 3:205–258, 2000.
- A. Munk and J. Schmidt-Hieber. Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electron. J. Stat.*, 4:781–821, 2010.
- M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15:634–658, 2009.
- M. Reiß. Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.*, 39:772–802, 2011.
- D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, 1999.
- T. Sabel and J. Schmidt-Hieber. Asymptotically efficient estimation of a scale parameter in Gaussian time series and closed-form expressions for the Fisher information. *Bernoulli*, 20(2):747–774, 2014a.
- T. Sabel and J. Schmidt-Hieber. Matlab toolbox spotvol. <http://www.stochastik.math.uni-goettingen.de/SpotvolToolbox>, 2014b.
- T. Sabel, J. Schmidt-Hieber, and A. Munk. Spot volatility estimation for high-frequency data: adaptive estimation in practice. *Springer Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension*, to appear, 2014.
- J. Schmidt-Hieber. *Nonparametric methods in spot volatility estimation*. PhD thesis, Georg-August-Universität Göttingen, 2010.
- J. Schmidt-Hieber, A. Munk, and L. Dümbgen. Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.*, 41(3): 1299–1328, 06 2013.

## BIBLIOGRAPHY

---

- V. Spokoiny. Multiscale local change point detection with application to value-at-risk. *Ann. Statist.*, 37:1405–1436, 2009.
- M. Steele. *Stochastic Calculus and Financial Applications*. Springer-Verlag, 2001.
- A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer-Verlag, 2009.
- A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer-Verlag, 1996.
- J. Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127:165–178, 2005.
- L. Zhang. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12:1019–1043, 2006.
- L. Zhang. Estimating covariation: Epps effect and microstructure noise. *Journal of Econometrics*, 160:33–47, 2011.
- B. Zhou. High-frequency data and volatility in foreign-exchange rates. *J. Business Econom. Statist.*, 14:45–52, 1996.

---

## APPENDIX A

# A PROOF OF THEOREM 3.4

---

The generality of the choice of the test function  $\psi$  heavily relies on the following theorem, which is motivated by Theorem 11.4 in [Koltchinskii \(1994\)](#). Although it relies only on surprisingly basic arguments, it simplifies the proofs of Theorems 3.4 and 5.5 tremendously.

**THEOREM A.1.** *For some uniformly bounded class  $\mathcal{F}$  of functions on  $[0, 1]$ , let  $\mathcal{H}_s$  denote the symmetric convex hull of  $\mathcal{F}$ , that is  $\mathcal{H}_s$  contains all functions  $g$ , s.t.  $g = \sum_{i=1}^N a_i f_i$ , where  $N \in \mathbb{N}$ ,  $f_i \in \mathcal{F}$ ,  $a_i \in [-1, 1]$  for  $i = 1, \dots, N$ , and  $\sum_{i=1}^N |a_i| \leq 1$ . Let  $\overline{\mathcal{H}}_s$  denote the sequential closure of  $\mathcal{H}_s$ , that is  $\overline{\mathcal{H}}_s$  is the smallest closed class containing all functions which are pointwise limits of sequences of functions in  $\mathcal{H}_s$ .*

Let  $J$  be some sublinear, real-valued (possibly random) functional on  $\overline{\mathcal{H}}_s$ . Assume further that  $J$  is continuous in the sense that for any sequence  $(g^N) \subset \mathcal{H}_s$ ,

$$g^N \rightarrow g \text{ a.s. pointwise} \Rightarrow J(g^N) \rightarrow J(g) \text{ in probability.}$$

Then, the almost surely,

$$\sup_{g \in \overline{\mathcal{H}}_s} |J(g)| = \sup_{f \in \mathcal{F}} |J(f)|. \tag{A.1}$$

*Proof.* Note that due to continuity of  $J$ , both expressions in (A.1) are finite.

Let  $g = \lim_{N \rightarrow \infty} \sum_{i=1}^N a_{i,N} f_{i,N} \in \overline{\mathcal{H}}_s$  for some  $f_{i,N} \in \mathcal{F}$  and  $a_{i,N} \in [-1, 1]$  with  $\sum_{i=1}^N |a_{i,N}| \leq 1$ . Then, we obtain by sublinearity and continuity that

$$J(g) = J\left(\lim_{N \rightarrow \infty} \sum_{i=1}^N a_{i,N} f_{i,N}\right) = \lim_{N \rightarrow \infty} J\left(\sum_{i=1}^N a_{i,N} f_{i,N}\right) \leq \lim_{N \rightarrow \infty} \sum_{i=1}^N a_{i,N} J(f_{i,N}) \leq \sup_{f \in \mathcal{F}} |J(f)|.$$

Note that the limit after the second equality is only in probability. However, taking appropriate subsequences yields almost sure convergence.

Furthermore, the relation  $\mathcal{F} \subset \overline{\mathcal{H}}_s$  provides equality in (A.1). □

---

## APPENDIX A. A PROOF OF THEOREM 3.4

---

REMARK A.2. Recall that  $\psi_{t,h}$  is  $L^2$ -normalized and  $\text{TV}(\psi) < \infty$  by Assumption 3.3. Thus, we obtain that

$$\text{TV}(\psi_{t,h}) = \text{TV}(\psi) \|\psi\|_{L^2_{[0,1]}}^{-1} h^{-1/2} = C_1 h^{-1/2}$$

for any  $(t, h) \in \mathcal{T}_n$ , where  $C_1$  is a finite, positive constant independent of  $(t, h)$ . Therefore, the total variation of  $h^{1/2}\psi_{t,h}/C_1$  is bounded by 1, so that we can write  $h^{1/2}\psi_{t,h}/C_1 = g_{t,h}$ , where  $g_{t,h}$  is in the sequential closure (in the sense of pointwise convergence) of the symmetric convex hull of the class of all indicators  $\mathcal{F} = \{\mathbb{I}_{[t,t+h]}, (t, h) \in \mathcal{T}\}$ , where  $\mathcal{T} = \{(t, h) : h \in (0, 1), [t, t+h] \subset [0, 1], h \leq u_n\}$  (cf. [Giné et al. \(2004\)](#), p.172).

This allows us to apply Theorem A.1, whenever a term depends sublinearly on  $\psi_{t,h}$ . Thus, it suffices to consider indicators  $\mathbb{I}_{[t,t+h]}$  instead of  $h^{1/2}\psi_{t,h}$  in many steps of the following proof. However, the supremum has to be taken over  $\mathcal{T}$  instead of  $\mathcal{T}_n$  in that case, that is the lower bound on  $h$  is omitted.

It is very natural to split the proof into three parts. To this end, observe that by the triangle inequality,

$$\sup_{(t,h) \in \mathcal{T}_n} w_h \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \sqrt{\frac{2}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n]} \right| \leq I + II + III,$$

where

$$\begin{aligned} I &:= \sup_{(t,h) \in \mathcal{T}_n} w_h \left| \sum_{i=0}^{n-1} \psi_{t,h}\left(\frac{i}{n}\right) \left( (W_{\frac{i+1}{n}} - W_{\frac{i}{n}})^2 - n^{-1} \right) - \sqrt{\frac{2}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n]} \right|, \\ II &:= \sup_{(t,h) \in \mathcal{T}_n} w_h \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\sigma_t^2} - \sum_{i=0}^{n-1} \psi_{t,h}\left(\frac{i}{n}\right) \left( (W_{\frac{i+1}{n}} - W_{\frac{i}{n}})^2 - \frac{1}{n} \right) \right|, \\ III &:= \sup_{(t,h) \in \mathcal{T}_n} w_h \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\sigma_t^2} \right|. \end{aligned}$$

Each of the three parts deals with a different approximation problem and we will therefore use a different proving technique in each step:

- The first approximation ( $I$ ) may be seen as a strong invariance principle for  $\chi^2$  distributed random variables, that is a uniform approximation of the sum of these random variables with Gaussian ones. For the proof, classic large deviation results are used, including Theorem 1 in [Komlós et al. \(1976\)](#) (termed “KMT-Theorem” in

---

the literature, cf. also Theorem C.2 in this thesis) and Lévy's modulus of continuity for Brownian motion.

- *II* states that in this specific situation, canceling  $\sigma_t$  and  $\sigma_s$  (for  $s \in [t, t+h]$ ) is correct up to some small error, as  $h$  tends to zero. To prove this, we use martingale inequalities (especially the Burkholder-Davis-Gundy inequality (cf. Proposition 2.2) and a result from Hoffmann (1999) (Lemma C.4)) combined with techniques from empirical process theory based on chaining.
- The last term reflects that there is almost no difference between using the true value  $\sigma_t^2$  or the estimator  $\hat{\sigma}_{t,h}^2$  in the denominator. This is proved via basic computations.

AD *I*. In a first step, let us uniformly and almost surely approximate a sum of independent, centered and normalized  $\chi^2$ -distributed random variables  $(\xi_i)_{i \in \mathbb{N}}$  by a sum of standard Gaussian ones, denoted by  $(\eta_i)_{i \in \mathbb{N}}$ . Observe that for  $t < 2^{-1/2}$ , the moment generating function  $M(t) = \mathbb{E} \exp(tX)$  of a centered and normalized  $\chi_1^2$  distributed random variable  $X$  is finite. Thus, we may apply Theorem 1 in Komlós et al. (1976) (cf. Theorem C.2, known as KMT-Theorem): Using the fact that

$$\max_{0 \leq j \leq k \leq n-1} \left| \sum_{i=j}^k (\xi_i - \eta_i) \right| \leq \max_{0 \leq j \leq k \leq n-1} \left( \left| \sum_{i=0}^k (\xi_i - \eta_i) \right| + \left| \sum_{i=0}^{j-1} (\xi_i - \eta_i) \right| \right) \leq 2 \max_{0 \leq k \leq n-1} \left| \sum_{i=0}^k (\xi_i - \eta_i) \right|,$$

their theorem yields that for any  $n \in \mathbb{N}$ , there exists a sequence  $(\eta_i)_{i \in \mathbb{N}}$  of i.i.d. standard Gaussian random variables, so that

$$\begin{aligned} & \mathbb{P} \left( \max_{0 \leq j \leq k \leq n-1} \left| \sum_{i=j}^k (\xi_i - \eta_i) \right| > C \log n + x \right) \leq \mathbb{P} \left( \max_{k \leq n-1} \left| \sum_{i=0}^k (\xi_i - \eta_i) \right| > (C \log n + x)/2 \right) \\ & \leq \mathbb{P} \left( \max_{k \leq n-1} \left| \sum_{i=0}^k (\xi_i - \eta_i) \right| > C' \log n + x/2 \right) < K \exp(-\lambda x/2) = K \exp(-\lambda' x), \end{aligned}$$

for all  $x > 0$ . Here,  $C', K, \lambda'$  are global constants independent of  $x$  and  $n$ . Furthermore, we may bound

$$\begin{aligned} \max_{0 \leq j \leq k \leq n-1} \left| \sum_{i=j}^k (\xi_i - \eta_i) \right| & \geq \sup_{(t,h) \in \mathcal{T}} \left| \sum_{i=0}^{n-1} \mathbb{I}_{[t,t+h]} \left( \frac{i}{n} \right) (\xi_i - \eta_i) \right| \\ & \geq \sup_{(t,h) \in \mathcal{T}_n} \left| \sum_{i=0}^{n-1} C_1^{-1} h^{1/2} \psi_{t,h} \left( \frac{i}{n} \right) (\xi_i - \eta_i) \right|, \end{aligned}$$

due to Remark A.2. This results in

$$\mathbb{P}\left(\sup_{(t,h)\in\mathcal{T}_n}\left|\sum_{i=0}^{n-1}h^{1/2}\psi_{t,h}\left(\frac{i}{n}\right)(\xi_i-\eta_i)\right|>C_1(C'\log n+x)\right)<K\exp(-\lambda'x). \quad (\text{A.2})$$

Note that this means that  $\sup_{(t,h)\in\mathcal{T}_n}\left|\sum_{i=0}^{n-1}h^{1/2}\psi_{t,h}\left(\frac{i}{n}\right)(\xi_i-\eta_i)\right|=O(\log n)$  a.s., since for  $x>2\log(n)/\lambda'$ , (A.2) is summable in  $n$ . By Lemma C.5, we find for  $n$  large enough that  $w_h h^{-1/2}$  is strictly decreasing in  $h$ . Thus,

$$\begin{aligned} \frac{\sqrt{2}}{n}\sup_{(t,h)\in\mathcal{T}_n}w_h\left|\sum_{i=0}^{n-1}\psi_{t,h}\left(\frac{i}{n}\right)(\xi_i-\eta_i)\right| &\leq\frac{\sqrt{2}}{n}\sup_{(t,h)\in\mathcal{T}_n}w_h h^{-1/2}\sup_{(t,h)\in\mathcal{T}_n}\left|\sum_{i=0}^{n-1}h^{1/2}\psi_{t,h}\left(\frac{i}{n}\right)(\xi_i-\eta_i)\right| \\ &=O(n^{-1}w_n l_n^{-1/2}\log(n)) \quad \text{a.s.} \end{aligned} \quad (\text{A.3})$$

In expression  $I$ , the squared increments  $((W_{\frac{i+1}{n}}-W_{\frac{i}{n}})^2-1/n)$ ,  $i=0,\dots,n-1$  are i.i.d. random variables with the same distribution as  $\sqrt{2}n^{-1}\xi_0$ . Thus by (A.3), we can approximate them by Gaussian random variables.

As a last step of the first part of the proof, we construct a sequence of Brownian motions  $W^{[n]}$ , so that the weighted partial sum process  $\frac{1}{n}\sum_{i=0}^{n-1}w_h\psi_{t,h}\left(\frac{i}{n}\right)\eta_i$  can be approximated by integrals with respects to  $W^{[n]}$  uniformly in  $(t,h)\in\mathcal{T}_n$ :

Let  $(B^{n,j})_{0\leq j\leq n-1}$  be a triangular scheme of independent Brownian bridges on  $[0,1]$ . Here and in the following, we assume that all random variables are defined on the same probability space to overcome measurability problems. Further let  $\eta_0,\dots,\eta_{n-1}$  be defined as before as i.i.d standard Gaussian random variables which additionally are assumed to be independent of  $(B^{n,j})$ . Define

$$W_s^{[n]}:=n^{-1/2}\left(\sum_{i=0}^{[ns]-1}\eta_i+\eta_{[ns]}(ns-[ns])+B_{ns-[ns]}^{n,[ns]}\right),$$

where  $[x]$  denotes the Gauss bracket. This is by construction a Brownian motion and

$$\begin{aligned} &\sup_{(t,h)\in\mathcal{T}_n}w_h\left|\sum_{i=0}^{n-1}n^{-1}\psi_{t,h}\left(\frac{i}{n}\right)\eta_i-n^{-1/2}\int_0^1\psi_{t,h}(s)dW_s^{[n]}\right| \\ &= \sup_{(t,h)\in\mathcal{T}_n}w_h n^{-1/2}\left|\sum_{i=0}^{n-1}\int_{\frac{i}{n}}^{\frac{i+1}{n}}\left(\psi_{t,h}\left(\frac{i}{n}\right)-\psi_{t,h}(s)\right)dW_s^{[n]}\right| \\ &\leq C_1 n^{-1/2}\sup_{(t,h)\in\mathcal{T}_n}w_h h^{-1/2}\sup_{(t,h)\in\mathcal{T}}\left|\sum_{i=0}^{n-1}\int_{\frac{i}{n}}^{\frac{i+1}{n}}\mathbb{I}_{[t,t+h]}\left(\frac{i}{n}\right)-\mathbb{I}_{[t,t+h]}(s)dW_s^{[n]}\right|, \end{aligned} \quad (\text{A.4})$$

applying again Theorem A.1 as described in Remark A.2. This requires continuity of the operator. For that purpose, recall that for any bounded sequence of measurable functions  $(g_j)$  with pointwise limit  $g$  and any locally square-integrable martingale  $Z$ ,  $\sup_{0 \leq u \leq 1} \left| \int_0^u g_j(s) - g(s) dZ_s \right| \rightarrow 0$  in probability (cf. for example Theorem 4.40 in [Jacod and Shiryaev \(2003\)](#)). Since we need a.s. continuity here, this is not enough. However, from the proof of Theorem A.1, it becomes clear that continuity of the functional  $J$  can be replaced by the condition  $J(\lim_{N \rightarrow \infty} g_N) = \lim_{N \rightarrow \infty} J(g_N)$  a.s. for the very specific sequence  $(g_j)$  used for the approximation of  $g$ . Thus, we may consider an appropriate subsequence  $(g_{j'})$  of  $(g_j)$  to make the convergence a.s.

In the last sum in (A.4), for each  $(t, h)$ , at most two summands are non-zero. Therefore, we may bound (A.4) by

$$\begin{aligned} & 2C_1 n^{-1/2} \sup_{(t,h) \in \mathcal{T}_n} w_h h^{-1/2} \sup_{s,u \in [0,1]: |s-u| \leq n^{-1}} |W_s^{[n]} - W_u^{[n]}| \\ &= O(w_{l_n} l_n^{-1/2} n^{-1} \sqrt{\log(n)}) = o(w_{l_n} l_n^{-1/2} n^{-1} \log(n)) \quad \text{a.s.}, \end{aligned}$$

using Lévy's modulus of continuity (cf. [Revuz and Yor \(1999\)](#)) and again Lemma C.5 for the monotonicity of  $w_h h^{-1/2}$ .

AD II. Observe that  $II \leq A + B$ , where

$$\begin{aligned} A &:= \sup_{(t,h) \in \mathcal{T}_n} w_h \left| \sum_{i=0}^{n-1} \psi_{t,h} \left( \frac{i}{n} \right) \left( \frac{\left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sigma_s dW_s \right)^2 - \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sigma_s^2 ds}{\sigma_t^2} - \left( (W_{\frac{i}{n}} - W_{\frac{i-1}{n}})^2 - \frac{1}{n} \right) \right) \right|, \\ B &:= \sup_{(t,h) \in \mathcal{T}_n} \frac{w_h}{\sigma_t^2} \left| \sum_{i=0}^{n-1} \psi_{t,h} \left( \frac{i}{n} \right) \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sigma_s^2 ds - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds \right|. \end{aligned}$$

To bound the second term, apply again Theorem A.1 as described in Remark A.2:

$$\begin{aligned} B &\leq \sup_{(t,h) \in \mathcal{T}_n} w_h h^{-1/2} \sigma_t^{-2} \sup_{(t,h) \in \mathcal{T}_n} h^{1/2} \left| \sum_{i=0}^{n-1} \psi_{t,h} \left( \frac{i}{n} \right) \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sigma_s^2 ds - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds \right| \\ &\leq \sup_{(t,h) \in \mathcal{T}_n} w_h h^{-1/2} \sigma_t^{-2} \sup_{(t,h) \in \mathcal{T}} \left| \sum_{i=0}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sigma_s^2 (\mathbb{I}_{[t,t+h]} \left( \frac{i}{n} \right) - \mathbb{I}_{[t,t+h]}(s)) ds \right| \tag{A.5} \\ &\leq C_1 \frac{\bar{\sigma}^2}{\underline{\sigma}^2} \sup_{(t,h) \in \mathcal{T}_n} w_h h^{-1/2} 2n^{-1} = O(w_{l_n} l_n^{-1/2} n^{-1}) = o(w_{l_n} l_n^{-1/2} n^{-1} \log(n)). \end{aligned}$$

Note that Theorem A.1 is not directly applicable to  $A$ , since  $A$  is not linear in  $\sigma_t^{-2}$ , while Theorem A.1 requires linearity in all terms depending on  $t$  and  $h$ . To obtain a similar result, note that by Lemmata C.1 and C.5,  $w_h C_1^{-1} \psi_{t,h}$ ,  $(t, h) \in \mathcal{T}_n$ , is in the sequential closure of the symmetric convex hull  $\overline{\mathcal{H}}_s$  of  $\{w_{h \vee l_n} (h \vee l_n)^{-1/2} \mathbb{I}_{[t, t+h]}, (t, h) \in \mathcal{T}\}$ . Further, observe that  $(\underline{\sigma}_{\sigma_{t_i, N}})^2 (\overline{\sigma}_{\sigma_t})^{-2} \leq 1$ . Therefore, we obtain by Lemma C.1 that the scaled function  $w_h C_1^{-1} \underline{\sigma}^2 (\overline{\sigma}_{\sigma_t})^{-2} \psi_{t,h}$  is in the sequential closure of the symmetric convex hull of  $\{w_{h \vee l_n} (h \vee l_n)^{-1/2} \mathbb{I}_{[t, t+h]} / \sigma_t^2 : (t, h) \in \mathcal{T}\}$ . Since the indicators used for both the approximation of  $w_h C_1^{-1} \psi_{t,h}$  and of  $w_h C_1^{-1} \underline{\sigma}^2 (\overline{\sigma}_{\sigma_t})^{-2} \psi_{t,h}$  are the same and only the coefficients  $a_{i,N}$  differ (cf. the proof of Lemma C.1), we can still use the linearity arguments in the proof of Theorem A.1 in this situation. This allows us to proceed as described in Remark A.2.

To bound the first term, we obtain by Itô's formula and by the triangle inequality:

$$\begin{aligned}
 A &= 2 \sup_{(t,h) \in \mathcal{T}_n} w_h \left| \sum_{i=0}^{n-1} \psi_{t,h} \left( \frac{i}{n} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \frac{\sigma_s}{\sigma_t} \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u - \int_{\frac{i}{n}}^s dW_u dW_s \right) \right| \\
 &\leq C_2 \sup_{(t,h) \in \mathcal{T}} q(h, n) \left| \sum_{i=0}^{n-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{n} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \frac{\sigma_s}{\sigma_t} \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u - \int_{\frac{i}{n}}^s dW_u dW_s \right) \right| \\
 &\leq C_2 (A_1 + A_2),
 \end{aligned} \tag{A.6}$$

where

$$\begin{aligned}
 A_1 &:= \sup_{(t,h) \in \mathcal{T}} q(h, n) \left| \sum_{i=0}^{n-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{n} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left( \frac{\sigma_s}{\sigma_t} - 1 \right) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u dW_s \right) \right|, \\
 A_2 &:= \sup_{(t,h) \in \mathcal{T}} q(h, n) \left| \sum_{i=0}^{n-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{n} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \int_{\frac{i}{n}}^s \left( \frac{\sigma_u}{\sigma_t} - 1 \right) dW_u dW_s \right) \right|,
 \end{aligned}$$

and  $C_2$  is a global constant and  $q(h, n) = w_{h \vee l_n} (h \vee l_n)^{-1/2}$ . Define further for  $j = 1, \dots, n$ ,

$$\begin{aligned}
 A_1^{t,h}(j) &:= q(h, n) \sum_{i=0}^{j-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{n} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left( \frac{\sigma_s}{\sigma_t} - 1 \right) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u dW_s \right), \\
 A_2^{t,h}(j) &:= q(h, n) \sum_{i=0}^{j-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{n} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \int_{\frac{i}{n}}^s \left( \frac{\sigma_u}{\sigma_t} - 1 \right) dW_u dW_s \right),
 \end{aligned}$$

and  $A_p^{t,h}(0) := 0$ ,  $p = 1, 2$ .

For fixed  $t, h, n$ ,  $A_1^{t,h}(j)$  and  $A_2^{t,h}(j)$  are locally square-integrable martingales in  $j$  w.r.t. the filtration  $\mathcal{F} = (\mathcal{F}_j)_{j \in \mathbb{N}}$ , where  $\mathcal{F}_j := \sigma(\{W_t : t \leq \frac{j}{n}\})$ .



Moreover, for  $j = 1, \dots, n$ ,  $k \geq 2$ ,

$$\begin{aligned}
& \mathbb{E} \left( \left| A_1^{t,h}(j+1) - A_1^{t,h}(j) \right|^k \middle| \mathcal{F}_j \right) \\
& \leq \left( \frac{C_B k^{1/2} q(h, n)}{\underline{\sigma}} \sup_{s \in [t, t+h+\frac{1}{n}]} |\sigma_s - \sigma_t| \right)^k \mathbb{E} \left( \left| \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left( \int_{\frac{j}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u \right)^2 ds \right|^{k/2} \middle| \mathcal{F}_j \right) \\
& \leq \left( \frac{C_B k^{1/2} q(h, n) L_{\sigma^2} (h + \frac{1}{n})^\gamma}{\underline{\sigma}} \right)^k n^{-k/2} \sup_{s \in [\frac{j}{n}, \frac{j+1}{n}]} \mathbb{E} \left( \left| \int_{\frac{j}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u \right|^k \middle| \mathcal{F}_j \right) \\
& \leq k^k \left( \frac{C_B^2 q(h, n) \bar{\sigma} L_{\sigma^2} (2(h \vee l_n))^\gamma}{n \underline{\sigma}^2} \right)^k,
\end{aligned}$$

where we used the Burkholder-Davis-Gundy inequality with factor  $C_B k^{1/2}$  (cf. Proposition 2.2) for the first and third inequality and Assumption 3.1 for the second one.

Choose  $C_3 = \frac{C_B^2 \bar{\sigma} L_{\sigma^2} 2^{2\gamma}}{\underline{\sigma}^2}$ . Note that  $A_1^{t,h}(j+1) - A_1^{t,h}(j) = 0$  if  $\frac{j}{n} \notin [t, t+h]$ , providing that there are at most  $hn + 1$  non-zero summands. Thus, we can apply Lemma C.4 with  $\ell = 2(h \vee l_n)n \geq hn + 1$ . This yields for every  $(t, h) \in \mathcal{T}_n$  and  $n$  large enough,

$$\begin{aligned}
& \mathbb{P}(|A_1^{t,h}(n)| \geq r_{n,h}) \\
& \leq 2 \exp \left( \frac{-\frac{1}{2} r_{n,h}^2}{e C_3 q(h, n) (h \vee l_n)^\gamma n^{-1} (2 C_3 q(h, n) (h \vee l_n)^\gamma 2(h \vee l_n) + r_{n,h})} \right) \quad (\text{A.7}) \\
& \leq 2 \exp \left( \frac{-\frac{1}{2} r_{n,h}^2}{e C_3 w_{h \vee l_n} (h \vee l_n)^{\gamma-1/2} n^{-1} (4 C_3 w_{h \vee l_n} (h \vee l_n)^{\gamma+1/2} + r_{n,h})} \right).
\end{aligned}$$

Choose  $r_{n,h} = C_4 w_{h \vee l_n} (h \vee l_n)^\gamma (\frac{\log n}{n})^{1/2}$  for  $C_4 = 8e^{1/2} C_3$ . Then,  $4 C_3 w_{h \vee l_n} (h \vee l_n)^{\gamma+1/2} > r_{n,h}$  for  $n$  large enough, and therefore, we may bound (A.7) by

$$2 \exp \left( \frac{-\frac{1}{2} r_{n,h}^2}{8e C_3^2 w_{h \vee l_n}^2 (h \vee l_n)^{2\gamma} n^{-1}} \right) = 2 \exp \left( \frac{-\frac{1}{2} C_4^2 \log n}{8e C_3^2} \right) = 2n^{-4}. \quad (\text{A.8})$$

Let  $r_n = \sup_{(t,h) \in \mathcal{T}} r_{n,h} = C_4 w_{u_n} u_n^\gamma (\frac{\log n}{n})^{1/2}$ . Further, denote the discrete grid with constant mesh  $n^{-1}$  on  $[0, 1]$  by  $\mathcal{X}_n := \{i/n : i \in \mathbb{N}, 0 \leq i \leq n\}$ , and the closed  $\epsilon$ -neighborhood (w.r.t. sup-norm) of  $x \in \mathbb{R}^2$  by  $\mathcal{B}_\epsilon(x)$ . Then, we may bound

$$A_1 = \sup_{(t,h) \in \mathcal{T}} |A_1^{t,h}(n)| \leq \sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} \left[ |A_1^{t,h}(n)| + \sup_{(t',h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t,h)} |A_1^{t',h'}(n) - A_1^{t,h}(n)| \right] \quad (\text{A.9})$$

Note that by (A.7) and (A.8),

$$\mathbb{P}\left(\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} |A_1^{t,h}(n)| \geq r_n\right) \leq n^2 \sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} \mathbb{P}(|A_1^{t,h}(n)| \geq r_n) \leq 2n^{-2}. \quad (\text{A.10})$$

This is summable and therefore,  $\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} |A_1^{t,h}(n)| = O(r_n)$  a.s.

The second summand in (A.9) is bounded by

$$\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} \sup_{(t',h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t,h)} \left[ |A_{1,1}| + |A_{1,2}| + |A_{1,3}| \right], \quad (\text{A.11})$$

where

$$\begin{aligned} A_{1,1} &:= (q(h, n) - q(h', n)) \sum_{i=0}^{n-1} \mathbb{I}_{[t, t+h]}(\frac{i}{n}) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} (\frac{\sigma_s}{\sigma_t} - 1) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u dW_s \right), \\ A_{1,2} &:= q(h', n) \sum_{i=0}^{n-1} (\mathbb{I}_{[t, t+h]} - \mathbb{I}_{[t', t'+h']})(\frac{i}{n}) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} (\frac{\sigma_s}{\sigma_t} - 1) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u dW_s \right), \\ A_{1,3} &:= q(h', n) \sum_{i=0}^{n-1} \mathbb{I}_{[t', t'+h']}(i/n) \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left[ (\frac{\sigma_s}{\sigma_t} - 1) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u \right. \right. \\ &\quad \left. \left. - (\frac{\sigma_s}{\sigma_{t'}} - 1) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_{t'}} dW_u \right] dW_s \right). \end{aligned}$$

Note that it suffices to cover  $\mathcal{T}$  by the set of upper halves of the  $\frac{1}{n}$ -neighborhoods around the points in  $\mathcal{X}_n^2$ , that is to assume that  $t' \geq t$ . This guarantees measurability (and therefore integrability) of the integrands in  $A_{1,2}$  and  $A_{1,3}$ , since we must ensure that  $t \leq i/n$  for integration of  $\sigma_t^{-2}$ .

In the following, let  $(t, h) \in \mathcal{T} \cap \mathcal{X}_n^2$  be arbitrary but fixed. Note that

$$\sup_{|h-h'| \leq \frac{1}{n}} \frac{|q(h, n) - q(h', n)|}{q(h, n)} = \frac{q(l_n, n) - q(l_n + \frac{1}{n}, n)}{q(l_n, n)} \rightarrow 0, n \rightarrow \infty.$$

Therefore,  $A_{1,1} \leq A_1^{t,h}(n)$  uniformly in  $(t', h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t, h)$ , showing that

$$\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} \sup_{(t',h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t,h)} A_{1,1} = O(r_n) \text{ a.s.}$$

---

For  $A_{1,2}$ , we approximate  $q(h', n)$  by  $q(h, n)$ . As we have seen in the preceding paragraph, this is possible up to an uniform  $o(1)$ -error. Further, observe that (again for fixed  $(t, h)$ ) all but at most 6 summands in  $A_{1,2}$  are zero for all  $(t', h')$  considered, so that the supremum over  $(t', h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t, h)$  reduces to a maximum over finitely many outcomes, all already covered by  $\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} |A_1^{t,h}(n)|$ . This shows that  $\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} \sup_{(t',h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t,h)} A_{1,2} = O(r_n)$  a.s.

Instead of  $A_{1,3}$ , it suffices by the same arguments as above to consider

$$\begin{aligned} \tilde{A}_{1,3} = \tilde{A}_{1,3}(t') := q(h, n) \sum_{i=0}^{n-1} \mathbb{I}_{[t, t+h]}(\frac{i}{n}) & \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left[ \left( \frac{\sigma_s}{\sigma_t} - 1 \right) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_t} dW_u \right. \right. \\ & \left. \left. - \left( \frac{\sigma_s}{\sigma_{t'}} - 1 \right) \int_{\frac{i}{n}}^s \frac{\sigma_u}{\sigma_{t'}} dW_u \right] dW_s \right), \end{aligned}$$

where again for measurability reasons, we now may assume w.l.o.g. that  $t' \leq t$ . To bound this term, we like to apply the following proposition, given as Corollary 2.2.5 in [van der Vaart and Wellner \(1996\)](#):

**PROPOSITION A.3.** *Let  $\Psi$  be a convex, nondecreasing, nonzero function with  $\Psi(0) = 0$  and  $\limsup_{x,y \rightarrow \infty} \Psi(x)\Psi(y)/\Psi(cxy) < \infty$  for some constant  $c > 0$ . Let  $(Z_t)_{t \in T}$  be a separable stochastic process with*

$$\|Z_s - Z_t\|_{\Psi} \leq Cd(s, t), \text{ for every } s, t \in T,$$

for some semimetric  $d$  on  $T$ , a constant  $C$ , and the Orlicz norm

$$\|Q\|_{\Psi} := \inf\{R > 0 : \mathbb{E}(\Psi(\frac{|Q|}{R})) \leq 1\}.$$

Then, there exists a constant  $K$ , such that

$$\left\| \sup_{s,t} |Z_s - Z_t| \right\|_{\Psi} \leq K \int_0^{\text{diam } T} \Psi^{-1}(D(\epsilon, d)) d\epsilon,$$

where  $D(\epsilon, d)$  is the packing number of  $T$  defined as the maximal number of  $\epsilon$ -separated points in  $T$ .

Choose  $\Psi(\bullet) = \exp(\bullet) - 1$ . Observe that for any  $C > 0$  for  $t_1, t_2 \in [t - \frac{1}{n}, t]$ , we may bound

$$\begin{aligned}
 & \mathbb{E} \Psi \left( \frac{|\tilde{A}_{1,3}(t_1) - \tilde{A}_{1,3}(t_2)|}{C} \right) + 1 \\
 &= \sum_{k=0}^{\infty} \frac{\mathbb{E} \left| \sum_{i=0}^{n-1} \mathbb{I}_{[t, t+h]} \binom{i}{n} \left( \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left[ \left( \frac{\sigma_s - \sigma_{t_1}}{\sigma_{t_1}^2} \right) - \left( \frac{\sigma_s - \sigma_{t_2}}{\sigma_{t_2}^2} \right) \right] \int_{\frac{i}{n}}^s \sigma_u dW_u \right) \right|^k}{q^{-k}(h, n) C^k k!} \\
 &\leq \sum_{k=0}^{\infty} \frac{(C_5 k q(h, n))^{\frac{1}{n}} (hn + 1)^{1/2} |t_1 - t_2|^\gamma)^k}{C^k k!} \\
 &\leq \sum_{k=0}^{\infty} \frac{(e C_5 q(h, n))^{\frac{1}{n}} (hn + 1)^{1/2} |t_1 - t_2|^\gamma)^k}{C^k},
 \end{aligned}$$

for some global constant  $C_5$ . Here, the second inequality comes from applying the Burkholder-Davis-Gundy (cf. Proposition 2.2) inequality twice (similarly as before), and from

$$\left| \frac{\sigma_s - \sigma_{t_1}}{\sigma_{t_1}^2} - \frac{\sigma_s - \sigma_{t_2}}{\sigma_{t_2}^2} \right| \leq C_6 |t_1 - t_2|^\gamma,$$

due to Hölder continuity of  $\sigma$ . The third inequality is due to  $k! > (k/e)^k$ .

Thus, for  $C > 2(e C_5 q(h, n))^{\frac{1}{n}} (hn + 1)^{1/2} |t_1 - t_2|^\gamma$ , we obtain

$$\mathbb{E} \Psi \left( \frac{|\tilde{A}_{1,3}(t_1) - \tilde{A}_{1,3}(t_2)|}{C} \right) \leq 1,$$

and therefore,

$$\|\tilde{A}_{1,3}(t_1) - \tilde{A}_{1,3}(t_2)\|_\Psi \leq 2e C_5 q(h, n)^{\frac{1}{n}} (hn + 1)^{1/2} |t_1 - t_2|^\gamma \leq C_7 w_{l_n} n^{-1/2} |t_1 - t_2|^\gamma,$$

for all  $(t, h) \in \mathcal{T}$ ,  $t_1, t_2 \in [t - \frac{1}{n}, t]$ . Thus, we may choose the semimetric  $d$  as

$$d(t_1, t_2) = w_{l_n} n^{-1/2} |t_1 - t_2|^\gamma.$$

Define the packing number  $D(\epsilon, d)$  as the largest number of points pairwise having  $d$ -distance strictly larger than  $\epsilon$ . It follows that  $D(\epsilon, d) \leq 2\epsilon^{-1/\gamma} n^{1/(2\gamma)} w_{l_n}^{1/\gamma}$ .

Note that  $\tilde{A}_{1,3}(t) = 0$ . Therefore, by Proposition A.3,

$$A_1^\Psi := \left\| \sup_{t' \in [t - \frac{1}{n}, t]} \tilde{A}_{1,3}(t') \right\|_\Psi \leq C_8 \int_0^{w_{l_n} n^{-\gamma-1/2}} \log(2\epsilon^{-1/\gamma} n^{1/(2\gamma)} w_{l_n}^{1/\gamma} + 1) d\epsilon$$

---


$$\leq C_8 \int_0^{w_{i_n} n^{-\gamma-1/2}} \log(4\epsilon^{-1/\gamma} n^{1/(2\gamma)} w_{i_n}^{1/\gamma}) d\epsilon \leq C_9 w_{i_n} n^{-\gamma-1/2} \log(n),$$

for some global constants  $C_8, C_9 > 0$ . We may conclude that for any  $(t, h) \in \mathcal{T}$ ,

$$\begin{aligned} \mathbb{P}\left(\sup_{t' \in [t-\frac{1}{n}, t]} \tilde{A}_{1,3}(t') \geq r_n\right) &= \mathbb{P}\left(\Psi\left(\sup_{t' \in [t-\frac{1}{n}, t]} \tilde{A}_{1,3}(t')/A_1^\Psi\right) + 1 \geq \Psi(r_n/A_1^\Psi) + 1\right) \\ &\leq \frac{\mathbb{E}\left(\Psi\left(\sup_{t' \in [t-\frac{1}{n}, t]} \tilde{A}_{1,3}(t')/A_1^\Psi\right) + 1\right)}{\Psi(r_n/A_1^\Psi) + 1} \leq 2 \exp(-r_n/A_1^\Psi) \\ &\leq 2 \exp(-C_4/C_9 w_{u_n} w_{i_n}^{-1} (u_n n)^\gamma \log^{-1/2} n) \leq \exp(-n^\beta), \end{aligned}$$

for some global  $\beta \in (0, 1)$  and  $n$  large enough, since  $u_n n$  is assumed to increase polynomially. Thus, we may proceed in analogy to (A.9), which yields

$$\sup_{(t,h) \in \mathcal{T} \cap \mathcal{X}_n^2} \sup_{(t',h') \in \mathcal{T} \cap \mathcal{B}_{\frac{1}{n}}(t,h)} A_{1,3} = O(r_n) \text{ a.s.},$$

which in turn proves that  $A_1 = O(r_n)$  a.s.

Note that all the arguments above concerning  $A_1$  are true for  $A_2$  as well. Actually, only the constants  $C_3$  and  $C_5$  are smaller by a factor  $\bar{\sigma}/\underline{\sigma}$  in that case. Therefore,  $II = O(r_n)$  a.s.

AD *III*. Since the convergence of  $\hat{\sigma}_{t,h}^2$  is uniform and almost sure, we may work on the event  $\{\inf_{t \in [0,1]} \hat{\sigma}_{t,h}^2 > \frac{1}{2}\underline{\sigma}^2\}$ . In the following, we use the notation  $\|\hat{\sigma}^2 - \sigma^2\|_\infty := \sup_{(t,h) \in \mathcal{T}_n} |\hat{\sigma}_{t,h}^2 - \sigma_t^2|$ . Observe

$$III \leq 2 \frac{\|\hat{\sigma}^2 - \sigma^2\|_\infty}{\underline{\sigma}^2} \sup_{(t,h) \in \mathcal{T}_n} w_h \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\sigma_t^2} \right|.$$

By Theorem 1 in [Schmidt-Hieber et al. \(2013\)](#) in combination with the bounds on  $I$  and  $II$ ,

$$\sup_{(t,h) \in \mathcal{T}_n} \left[ n^{1/2} w_h \left| \frac{T_{n,t,h}^{(1)} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\sigma_t^2} \right| - \frac{\sqrt{2} \log \frac{\nu}{h}}{\log \log \frac{\nu}{h}} \right] < \infty \text{ a.s.}$$

## APPENDIX A. A PROOF OF THEOREM 3.4

---

Since  $\|\hat{\sigma}^2 - \sigma^2\|_\infty = O(s_n)$  a.s.,

$$III = O(\|\hat{\sigma}^2 - \sigma^2\|_\infty \sup_{(t,h) \in \mathcal{T}_n} \frac{\log \frac{\nu}{h}}{n^{1/2} \log \log \frac{\nu}{h}}) = O(s_n n^{-1/2} \frac{\log(1/l_n)}{\log \log(1/l_n)}) \text{ a.s.},$$

which completes the proof. □

---

## APPENDIX B

# A PROOF OF THEOREM 5.5

---

Since the proof of Theorem 5.5 is massively based on the proof of Theorem 3.4 in structure, notation, and techniques, we strongly advise the reader take a look at Appendix A before reading this one.

*Proof of a.* In analogy to the proof of the semimartingale case, we decompose

$$\sup_{(t,h) \in \mathcal{T}_m} w_h \left| \frac{T_{\sigma^2} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \sqrt{\frac{2}{m}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,1} \right| \leq I + II + III,$$

where

$$\begin{aligned} I &:= \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \sum_{i=0}^{m-1} \psi_{t,h}\left(\frac{i}{m}\right) \left( \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) dW_s \right)^2 - m^{-1} \right) \right. \\ &\quad \left. - \sqrt{\frac{2}{m}} \int_0^1 \psi_{t,h}(s) dW_s^{[n]} \right|, \\ II &:= \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \frac{T_{\sigma^2} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\sigma_t^2} \right. \\ &\quad \left. - \sum_{i=0}^{m-1} \psi_{t,h}\left(\frac{i}{m}\right) \left( \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) dW_s \right)^2 - m^{-1} \right) \right|, \\ III &:= \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \frac{T_{\sigma^2} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\hat{\sigma}_{t,h}^2} - \frac{T_{\sigma^2} - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds}{\sigma_t^2} \right|. \end{aligned}$$

Again, we will bound these terms separately:

AD *I*. Note that  $m \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) dW_s \right)^2$  is standard  $\chi^2$ -distributed due to the normalization of the pre-average function. Therefore, the proof is the same as in part *I* of the proof of Theorem 3.4 (with  $n$  replaced by  $m$ ), yielding

$$I = O(w_{l_m} l_m^{-1/2} m^{-1} \log(m)) \text{ a.s.}$$

AD II. Again, similar to the proof in the semimartingale case, observe that  $II \leq A + B_1 + B_2$ , with

$$A := \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \sum_{i=0}^{m-1} \psi_{t,h} \left( \frac{i}{m} \right) \left[ \frac{\left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \sigma_s dW_s \right)^2 - \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i^2(s) \sigma_s^2 ds}{\sigma_t^2} - \left( \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) dW_s \right)^2 - m^{-1} \right) \right] \right|,$$

$$B_1 := \sup_{(t,h) \in \mathcal{T}_m} \frac{w_h}{\sigma_t^2} \left| \sum_{i=0}^{m-1} \psi_{t,h} \left( \frac{i}{m} \right) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \sigma_s^2 - \Lambda_i^2(s) \sigma_s^2 ds \right|,$$

$$B_2 := \sup_{(t,h) \in \mathcal{T}_m} \frac{w_h}{\sigma_t^2} \left| \sum_{i=0}^{m-1} \psi_{t,h} \left( \frac{i}{m} \right) \int_{\frac{i}{n}}^{\frac{i+1}{m}} \sigma_s^2 ds - \int_0^1 \psi_{t,h}(s) \sigma_s^2 ds \right|.$$

$B_2$  is the same as  $B$  in the proof of Theorem 3.4 (cf. (A.5)), while for  $B_1$ , note that  $\sigma_{i/m}^2 \int_{\frac{i}{m}}^{\frac{i+1}{m}} (1 - \Lambda_i^2(s)) ds = 0$ . Thus,

$$\begin{aligned} B_1 &= \sup_{(t,h) \in \mathcal{T}_m} \frac{w_h}{\sigma_t^2} \left| \sum_{i=0}^{m-1} \psi_{t,h} \left( \frac{i}{m} \right) \int_{\frac{i}{m}}^{\frac{i+1}{m}} (\sigma_s^2 - \sigma_{i/m}^2) (1 - \Lambda_i^2(s)) ds \right| \\ &= O(m^{-\gamma}) \sup_{(t,h) \in \mathcal{T}_m} w_h \sum_{i=0}^{m-1} |\psi_{t,h} \left( \frac{i}{m} \right)| \int_{\frac{i}{m}}^{\frac{i+1}{m}} |1 - \Lambda_i^2(s)| ds = O(m^{-\gamma} w_{u_m} u_m^{1/2}), \end{aligned}$$

where we used the Hölder continuity of  $\sigma$  (and therefore of  $\sigma^2$ ), and that  $\psi_{t,h}$  is  $L^2$ -normalized resulting in  $\sum_{i=0}^{m-1} |\psi_{t,h} \left( \frac{i}{m} \right)| = O(mh^{1/2})$ .

For the first term, observe again that by Itô's formula,

$$A = 2 \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \sum_{i=0}^{m-1} \psi_{t,h} \left( \frac{i}{m} \right) \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \left( \frac{\sigma_s}{\sigma_t} \int_{\frac{i}{m}}^s \Lambda_i(u) \frac{\sigma_u}{\sigma_t} dW_u - \int_{\frac{i}{m}}^s \Lambda_i(u) dW_u \right) dW_s \right) \right| \leq C_2(A_1 + A_2),$$

where

$$\begin{aligned} A_1 &:= \sup_{(t,h) \in \mathcal{T}} q(h, m) \left| \sum_{i=0}^{m-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{m} \right) \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \left( \frac{\sigma_s - \sigma_t}{\sigma_t} \int_{\frac{i}{m}}^s \Lambda_i(u) \frac{\sigma_u}{\sigma_t} dW_u dW_s \right) \right) \right|, \\ A_2 &:= \sup_{(t,h) \in \mathcal{T}} q(h, m) \left| \sum_{i=0}^{m-1} \mathbb{I}_{[t, t+h]} \left( \frac{i}{m} \right) \left( \int_{\frac{i}{n}}^{\frac{i+1}{m}} \Lambda_i(s) \int_{\frac{i}{m}}^s \Lambda_i(u) \left( \frac{\sigma_u - \sigma_t}{\sigma_t} \right) dW_u dW_s \right) \right|, \end{aligned}$$

in analogy to (A.6). Therefore, only the additional factor  $\sup_{s \in [0,1]} \Lambda^2(s) < \infty$  occurs in



---

the Burkholder-Davis-Gundy inequalities in that part of the proof and therefore in all the calculations after (A.6). This in combination with the bound on  $B_1$  gives us the bound on  $II$ , that is

$$II = O(m^{-\gamma} w_{u_m} u_m^{1/2} + w_{u_m} u_m^\gamma (\frac{\log(m)}{m})^{1/2}), \text{ a.s.}$$

AD *III*. This is again completely analog to part *III* of the previous proof.  $\square$

*Proof of b.* Again, we decompose

$$\sup_{(t,h) \in \mathcal{T}_m} w_h \left| \frac{T_{\sigma\tau}}{\hat{\sigma}_{t,h}} - \tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,2} \right| \leq I + II + III,$$

where

$$\begin{aligned} I &:= \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \sum_{i=0}^{m-1} \psi_{t,h}(\frac{i}{m}) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) dW_s \frac{\tau}{\sqrt{n}} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right. \\ &\quad \left. - \tau \|\lambda\|_{L^2_{[0,1]}} \sqrt{\frac{m}{n}} \int_0^1 \psi_{t,h}(s) dW_s^{[n],m,2} \right|, \\ II &:= \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \sum_{i=0}^{m-1} \psi_{t,h}(\frac{i}{m}) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) (1 - \frac{\sigma_s}{\sigma_t}) dW_s \frac{\tau}{\sqrt{n}} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right|, \\ III &:= \sup_{(t,h) \in \mathcal{T}_m} w_h \left| (\frac{1}{\sigma_t} - \frac{1}{\hat{\sigma}_{t,h}}) \sum_{i=0}^{m-1} \psi_{t,h}(\frac{i}{m}) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) \sigma_s dW_s \frac{\tau}{\sqrt{n}} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right|. \end{aligned}$$

AD *I*. Let  $(\xi_i^{(1)})_{i \in \mathbb{N}}$  and  $(\xi_i^{(2)})_{i \in \mathbb{N}}$  be two sequences of i.i.d. standard Gaussian random variables, independent of each other. Let  $(\xi_i)_{i \in \mathbb{N}} = (\xi_i^{(1)} \xi_i^{(2)})_{i \in \mathbb{N}}$  be their component-wise product. Then, the moment-generating function of  $\xi_1$  is finite in  $(-1, 1)$ . Thus, we may again apply the result in Komlós et al. (1976) analogously to the proof in the semimartingale case. To obtain the bound on  $I$ , observe that  $m^{1/2} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) dW_s$  and  $m^{-1/2} \|\lambda\|_{L^2_{[0,2]}}^{-1} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^*$ ,  $i = 0, \dots, m-1$ , are i.i.d standard Gaussian random variables, independent of each other. This shows that the scaling  $\tau \|\lambda\|_{L^2_{[0,2]}} \sqrt{\frac{m}{n}}$  is chosen correctly, and therefore

$$I = O(w_{l_m} l_m^{-1/2} \log(m) n^{-1/2}) \text{ a.s.}$$

Note that the a.s. approximation of the sum of standard Gaussian random variables by an Itô integral is again negligible, since it causes an error of order  $O(w_{l_m} l_m^{-1/2} \log^{1/2}(m) n^{-1/2})$ .

AD *II*. Observe that

$$II \leq \max_{0 \leq i \leq m-1} \left| \frac{\tau}{\sqrt{n}} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right| \sup_{(t,h) \in \mathcal{T}_m} w_h \left| \sum_{i=0}^{m-1} \psi_{t,h}(\frac{i}{m}) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) (1 - \frac{\sigma_s}{\sigma_t}) dW_s \right|.$$

The first term can be a.s. bounded by

$$\max_{0 \leq i \leq m-1} \left| \frac{\tau}{\sqrt{n}} \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right| = O((m/n)^{1/2} \log^{1/2} m), \quad (\text{B.1})$$

due to Lévy's modulus of continuity. By Theorem A.1 (cf. also Remark A.2), we may bound the second term by

$$\sup_{(t,h) \in \mathcal{T}} q(h, m) \left| \sum_{i=0}^{m-1} \mathbb{I}_{[t, t+h]}(\frac{i}{m}) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) (1 - \frac{\sigma_s}{\sigma_t}) dW_s \right|.$$

Here,  $q(h, m) = w_{h \vee l_m} (h \vee l_m)^{-1/2}$  as in Part *II* of the proof of Theorem 3.4. Define

$$A_{t,h}(j) := q(h, m) \sum_{i=0}^{j-1} \mathbb{I}_{[t, t+h]}(\frac{i}{m}) \int_{\frac{i}{m}}^{\frac{i+1}{m}} \Lambda_i(s) (1 - \frac{\sigma_s}{\sigma_t}) dW_s,$$

which is a locally square-integrable martingale in  $j$  w.r.t. the filtration  $\mathcal{F} = (\mathcal{F}_j)_{j \in \mathbb{N}}$ , where  $\mathcal{F}_j := \sigma(\{W_t : t \leq \frac{j}{n}\})$ . We obtain by the Burkholder-Davis-Gundy inequality

$$\begin{aligned} \mathbb{E} \left( \left| A^{t,h}(j+1) - A^{t,h}(j) \right|^k \middle| \mathcal{F}_j \right) &\leq (Cq(h, m)(k/m)^{1/2} (h \vee l_m)^\gamma)^k \\ &\leq k^k (Cq(h, m)m^{-1/2} (h \vee l_m)^\gamma)^k, \end{aligned}$$

for some constant  $C$ . Thus, the proof of Theorem 3.4 is valid with  $r_m = w_{u_m} u_m^\gamma$ , showing that the second term is  $O(r_m)$  a.s. Combining this with (B.1), we obtain that

$$II = O(w_{u_m} u_m^\gamma (m/n)^{1/2} \log^{1/2} m) \text{ a.s.}$$

AD *III*. This part of proof goes again along the lines of part *III* in Appendix A.  $\square$

*Proof of c.* Note that for this part, the steps *II* and *III* in the proof of Theorem 3.4 are not necessary, since the terms in c.) do not depend on  $\sigma$ . Thus, we only have to prove the

---

strong approximation previously denoted by  $I$ . To this end, observe that

$$\frac{\tau^2}{n} \left( \left( \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i(s) dW_s^* \right)^2 - \int_{\frac{i}{m}}^{\frac{i+1}{m}} \lambda_i^2(s) ds \right), \quad i = 0, \dots, m-1 \quad (\text{B.2})$$

are i.i.d. scaled and centered  $\chi^2$ -distributed random variables. Therefore, the proof is completely analog to  $I$  in Appendix A up to the different scaling. Here, the factor  $\sqrt{2}\tau^2 \frac{m^{3/2}}{n} \|\lambda\|_{L^2_{[0,2]}}^2$  can directly be verified by calculation of the standard deviation of (B.2).  $\square$

*Proof of Corollary 5.9.* Summing up, we obtain (5.6) by the decomposition in (5.5), the triangle inequality, and the upper bound on  $\hat{\sigma}_{t,h}^2$ . The latter is given by the almost sure and uniform convergence of  $\hat{\sigma}_{t,h}^2$  and the bound on  $\sigma^2$ .  $\square$



---

## APPENDIX C

# TECHNICAL AND AUXILIARY RESULTS

---

Let us state a technical implication concerning convex hulls of functions:

LEMMA C.1. *Denote the sequential closure of the symmetric convex hull of the class of all indicators*

$$\mathcal{F} = \{\mathbb{I}_{[t,t+h]} : (t, h) \in \mathcal{T} = \{(t, h) : t \in [0, 1], t + h \in [0, 1], 0 < h < h_{max}\}\}$$

by  $\overline{\mathcal{H}}_s$ . Let  $g$  be a function with support in  $[0, 1]$  and  $g(\frac{\bullet-t^*}{h^*}) \in \overline{\mathcal{H}}_s$  for some  $(t^*, h^*)$ . Then,  $p(t^*, h^*)g(\frac{\bullet-t^*}{h^*})$  is in the sequential closure of the symmetric convex hull of

$$\{(p(t, h)\mathbb{I}_{[t,t+h]} : (t, h) \in \mathcal{T}\},$$

if  $|p(t^*, h^*)/p(t, h)| \leq 1$  for all  $t, h$  with  $[t, t+h] \subseteq [t^*, t^* + h^*]$ .

*Proof.* Consider a sequence  $g_{t^*, h^*}^{[N]} = \sum_{i=0}^N a_{i,N} \mathbb{I}_{[t_{i,N}, t_{i,N} + h_{i,N}]}$  approximating  $g(\frac{\bullet-t^*}{h^*})$  pointwise with  $a_{i,N} \in \mathbb{R}$ ,  $\sum_{i=0}^N |a_{i,N}| \leq 1$ , and  $[t_{i,N}, t_{i,N} + h_{i,N}] \subseteq [t^*, t^* + h^*]$ . Such a sequence exists due to the definition of the sequential closure of the symmetric convex hull and due to the fact, that  $\text{supp } g(\frac{\bullet-t^*}{h^*}) \subset [t^*, t^* + h^*]$ . Note that  $p(t^*, h^*)g(\frac{\bullet-t^*}{h^*})$  can be approximated pointwise by

$$p(t^*, h^*)g_{t^*, h^*}^{[N]} = \sum_{i=0}^N a_{i,N} \frac{p(t^*, h^*)}{p(t_{i,N}, h_{i,N})} p(t_{i,N}, h_{i,N}) \mathbb{I}_{[t_{i,N}, t_{i,N} + h_{i,N}]}.$$

Since  $\sum_{i=0}^N |a_{i,N} \frac{p(t^*, h^*)}{p(t_{i,N}, h_{i,N})}| \leq 1$  by assumption, the proof is complete.  $\square$

The following result from [Komlós et al. \(1976\)](#) is essential for step I in our proofs. In [Komlós et al. \(1975\)](#), the same result was proved under stronger conditions. However, the main idea of the proof is more obvious in the first paper.

---

APPENDIX C. TECHNICAL AND AUXILIARY RESULTS

---

THEOREM C.2. *Let  $F$  be a distribution function with moment generating function  $M$ , which is finite in some neighborhood of 0, that is for  $X \sim F$  and some  $t_0 > 0$ :*

$$\mathbb{E}(\exp(tX)) < \infty \text{ for all } |t| < t_0.$$

*Further assume that  $\mathbb{E}X = 0$  and  $\mathbb{E}X^2 = 1$ , if  $X \sim F$ . Then, one can construct a sequence of i.i.d. random variables  $(\xi_i)$  with distribution  $F$  and a sequence of i.i.d. standard Gaussian random variables  $(\eta_i)$  on the same probability space, such that for all  $n$  and all  $x > 0$*

$$\mathbb{P}(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k (\xi_i - \eta_i) \right| > C \log n + x) \leq K \exp(-\lambda x),$$

*for some constants  $C, K, \lambda$  not depending on  $n$  or  $x$ .*

*Thus,  $\max_{1 \leq k \leq n} \left| \sum_{i=1}^k (\xi_i - \eta_i) \right| = O(\log n)$  a.s.*

REMARK C.3. *In the proof of this theorem, the authors construct  $(\xi_i)$  based on a given standard Gaussian sequence  $(\eta_i)$ . Note that in the proof of Theorem 3.4, we like to go the other way around, that is constructing standard Gaussian random variables from given (normalized)  $\chi^2$  ones. However, the proofs in Komlós et al. (1975, 1976) reveal that this is possible as well (provided that the probability space is large enough).*

Next, we state a lemma about large deviation of martingales, which can be found in Hoffmann (1999), Lemma 3.

LEMMA C.4. *Let  $(M_\ell)_{\ell \geq 0}$  be an  $(\mathcal{F}_\ell)$ -martingale with  $M_0 = 0$  such that for  $k = 2, 3, \dots$ ,  $\mathbb{E}(|M_{i+1} - M_i|^k | \mathcal{F}_i) \leq c^k k^k$  for some constant  $c > 0$ . Then, for all  $t \geq 0$ :*

$$\mathbb{P}(|M_\ell| > t) \leq 2 \exp\left(-\frac{1}{2} \frac{t^2}{ce(2c\ell + t)}\right).$$

In the proofs of Theorems 3.4 and 5.5, it is crucial to know the behavior of  $w_h h^z$  for  $h \rightarrow 0$ . To this end, the following lemma is useful.

LEMMA C.5. *Let  $w_h$  be as in Theorem 3.4. Then for any  $z > 0$  ( $z \leq 0$ ), there exists an  $\epsilon > 0$ , s.t.  $f(h) = w_h h^z$  is increasing (decreasing) on  $(0, \epsilon)$ .*

*Proof.* The result can be obtained by differentiation w.r.t.  $h$ . □

---

Let us collect some facts about total variation:

LEMMA C.6. *Let  $f, g : [a, b] \rightarrow \mathbb{R}$  be bounded functions with finite total variation. Then,*

1.  $\text{TV}(f + g) \leq \text{TV}(f) + \text{TV}(g)$ ,
2.  $\text{TV}(fg) \leq \|f\|_\infty \text{TV}(g) + \|g\|_\infty \text{TV}(f)$ .

*Proof.* The first inequality is true due to the triangle inequality. For the second one, observe that for any partition  $a = x_0 \leq \dots \leq x_n = b$ , we find

$$\begin{aligned} & \sum_{i=1}^n \left| f(x_i)g(x_i) - f(x_{i-1})g(x_{i-1}) \right| && \text{(C.1)} \\ & \leq \sum_{i=1}^n \left| f(x_i)g(x_i) - f(x_i)g(x_{i-1}) \right| + \sum_{i=1}^n \left| f(x_i)g(x_{i-1}) - f(x_{i-1})g(x_{i-1}) \right| \\ & \leq \|f\|_\infty \text{TV}(g) + \|g\|_\infty \text{TV}(f). \end{aligned}$$

Since  $\text{TV}(fg)$  is defined as the supremum of sums of type (C.1), the statement is proved.  $\square$





---

# CURRICULUM VITAE

---

- February 3rd, 1986    born in Oldenburg (Oldb)
- June, 2005            Abitur,  
                          Gymnasium Cäcilienchule Oldenburg
- September, 2005    Community Service,  
– August, 2006      Sportjugend Niedersachsen
- September, 2009    B.Sc. in Mathematics,  
                          Georg-August-Universität Göttingen
- March, 2011         M.Sc. in Econometrics,  
                          Georg-August-Universität Göttingen
- April, 2011          Research Assistant (funded by DFG/SNF research group  
– March, 2014      FOR 916) and Ph.D. Student in Mathematics,  
                          Georg-August-Universität Göttingen and Universität Bern