Rough Continuous-Time Processes: Theory and Applications
ROUGH CONTINUOUS-TIME PROCESSES: THEORY AND APPLICATIONS

By Mikkel Bennedsen

A dissertation submitted to
School of Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of
the PhD degree in
Economics and Business Economics

August, 2016
This dissertation comprises parts of the research I have undertaken during my graduate studies at the Department of Economics and Business Economics and at the Center for Research in Time Series Analysis (CREATES), Aarhus University, where I enrolled in September, 2012. I would like to express my gratitude to the Department and to CREATES and its director, Professor Niels Haldrup, for creating a wonderful research environment. A special thank you goes to Solveig Sørensen for her patience and help with all matters, large and small.

A number of people have had a substantial influence on this thesis and they all deserve mention. First and foremost, I would like to thank my main supervisor, Professor Asger Lunde, for countless stimulating and thought-provoking conversations over the years. I am grateful for your guidance and your patience and I hope we can continue to work together in the years to come. Secondly, I would like to thank my co-supervisor, Professor Ole Barndorff-Nielsen, who inspired me to go into research. Without Ole, there would have been no dissertation, and I feel extremely privileged to have had the chance to work and converse with one of the true legends of our field. Ole's power of mind and his ability to create ideas have left a profound impression on me and I will never forget it. Next, I would like to thank my "de-facto" co-co-supervisor, Dr. Mikko S. Pakkanen. Most of my research has been in close collaboration with Mikko, and I think it is fair to say that it is from Mikko that I have learned most of the "craft" that research also is. Thank you, Mikko, I am forever grateful. Hopefully we can continue our work in the years to come. I would also like to extend a special thanks to Professor Neil Shephard for sponsoring my research period abroad. Neil graciously hosted me at Harvard University in Cambridge, MA, from August to December, 2014, and this period stands as one of the great moments of my life. I am grateful to Neil for welcoming me and in particular for the many stimulating discussions during my stay. It was a great honor and an inspiration to see you work, Neil. Lastly, I would like to thank my other co-authors and close collaborators: Dr. Ulrich Hounyo and Dr. Almut Veraart.

While the research environment at Aarhus University has been fantastic, the social environment has been even better, and I would like to thank all my friends who made these four years some of the most wonderful years of my life. First, I would like to thank Anne and Strange. I can not begin to express my gratitude for what
you have done for me. You invited me into your office and into your lives – and the rest is history, as I think they say. I am certain that we will stay close for the rest of our lives. A special thanks also goes to your significant others, Signe and Rasmus, and your wonderful children, Lærke and Emil. Another special shout-out goes to the not-so-sad Prof. Jonas – I always enjoy your company and I hope we will continue to see each other often. Next, thank you to my new office mates, Kasper and Mads, and thank you to all the crazy kids who always made lunch, coffee breaks, and the many parties and Friday bars fun and unforgettable: Anne, Sanni, Silvana, Stine, Tine, Alex, Anders, Anders, Bo, Carsten, Christian, Husted, Jakob, Johan, Juan Carlos, Jonas, Kasper, Magnus, Orimar, Palle, Rasmus, Rune, Simon, and Søren.

My friends outside Academia also deserve my gratitude for putting up with me over the years, as I got increasingly out of touch with life outside the Ivory Tower. First, a special thanks to a special friend, Big R. Also thank you to my old study buddies Big K and Stevens. Thank you to the best roommates in the world, Louise and Villum. Thank you to my childhood friends, “Jungs”: Tina, Asser, Baggi, Bobo, Caller, Chag, Chef, Clausi, Flæske, Jepf, Julli, Mike, Jøgge, Polle, Pommel, and Togge. If it were not for all of you reminding me that there is a life outside research, I would never have kept my sanity. Thank you!

I would like to thank my family: Mor and Far, Lise, Mille, Dorthe, Henrik, Torben, Maja, Malte, and Simon. Without your love and support I would never have completed my studies, never mind this dissertation. Lastly, I want to thank my girlfriend, Fie, for love and support in the good times, as well as the tough times. I am forever grateful for your kindness and encouragement. Thank you.

Mikkel Bennedsen
Aarhus, August 2016
My predefense took place on October 27, 2016. The assessment committee consists of Dr. Almut Veraart, Imperial College London, Professor Roberto Renò, University of Verona, and Associate Professor Anders Bredahl Kock, Aarhus University. I am thankful to the members of the committee for their careful reading of my dissertation and for their many constructive and insightful comments and suggestions, many of which have been incorporated into the final version of the dissertation. The public defense is scheduled to take place on February 3, 2017.

Mikkel Bennedsen
Aarhus, January 2017
# CONTENTS

Summary vii

Danish summary xi

1 Hybrid scheme for Brownian semistationary processes 1
   1.1 Introduction ................................................. 2
   1.2 The model and theoretical results ......................... 4
   1.3 Implementation and numerical experiments ............... 15
   1.4 Proofs .................................................. 23
   1.5 References ............................................. 33

2 Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data 37
   2.1 Introduction ................................................ 38
   2.2 Setup ..................................................... 41
   2.3 Semiparametric estimation and bootstrapping of the fractal index . 43
   2.4 Extensions ................................................ 49
   2.5 Simulation study .......................................... 53
   2.6 Empirical application: testing the semimartingale assumption on financial time series ........................................... 57
   2.7 Conclusion ................................................ 64
   2.8 References ............................................. 65
   Appendix ...................................................... 69
   A.1 Proofs .................................................. 69
   A.2 Deriving an expression for the Λ matrix of Theorem 2.3.1 .... 72
   A.3 Simulation setup ......................................... 73

3 Decoupling the short- and long-term behavior of stochastic volatility 77
   3.1 Introduction ................................................ 78
   3.2 Empirical behavior of realized volatility ................. 80
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Models of stochastic volatility that decouple short- and long-term behavior</td>
<td>92</td>
</tr>
<tr>
<td>3.4</td>
<td>Estimating the models</td>
<td>100</td>
</tr>
<tr>
<td>3.5</td>
<td>Application to individual equities</td>
<td>102</td>
</tr>
<tr>
<td>3.6</td>
<td>Application to volatility forecasting</td>
<td>106</td>
</tr>
<tr>
<td>3.7</td>
<td>Conclusions and further research</td>
<td>113</td>
</tr>
<tr>
<td>3.8</td>
<td>References</td>
<td>118</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
<td>123</td>
</tr>
<tr>
<td>A.1</td>
<td>Evaluation of the RFSV integral</td>
<td>123</td>
</tr>
<tr>
<td>A.2</td>
<td>Proofs</td>
<td>123</td>
</tr>
</tbody>
</table>
This dissertation contains three independent chapters in the form of three original research manuscripts. The common theme that underlies the dissertation is the concept of a rough continuous-time process. In other words, the mathematical object of interest will be a particular class of stochastic processes: the processes we have in mind are formulated in continuous-time; they are real-valued; and they are continuous. In fact, for some $\alpha \in (-1/2, 1/2)$, they are $\phi$-Hölder continuous for all $\phi \in (0, \alpha + 1/2)$. We call the parameter $\alpha$ the roughness index as it controls the degree of roughness of the trajectories of the stochastic process under study. In particular, when $\alpha$ is negative, the stochastic process will have sample paths which are Hölder continuous of a lower order than those of the Brownian motion. The mathematical challenge of working with these processes comes mainly from the fact that for $\alpha \neq 0$, they are non-semimartingales. Although many results in this dissertation will hold for processes with roughness index $\alpha$ in the whole range $(-1/2, 1/2)$, we will primarily be interested in the rough case, i.e. $\alpha \in (-1/2, 0)$.

As hinted by the title of the dissertation, the research presented here will be concerned both with theory of rough processes as well as their application to real world data. The first two chapters are primarily theoretical in nature, while the third chapter illustrates the application of rough processes to stochastic volatility modelling. Not included in this dissertation, but intimately related, are the papers Bennedsen (2015) and Bennedsen, Hounyo, Lunde, and Pakkanen (2016a). In the former of these, a rough process is employed to arrive at a novel mathematical model for electricity spot prices; in the latter, we derive a new resampling method, the Local Fractional Bootstrap, applicable to power variations of a certain class of rough processes.

Chapter 1 (Bennedsen, Lunde, and Pakkanen, 2015) develops a fast and accurate simulation scheme for the Brownian semistationary (BS) process (Barndorff-Nielsen and Schmiegel, 2007, 2009), which is a large and flexible class of rough processes. The scheme is valid more broadly, and we show how to extend it beyond the BS framework; for instance, it is an effective way to simulate the rough Bergomi model of Bayer, Friz, and Gatheral (2016). Simulation schemes for rough processes should not be taken for granted, since such processes generally are non-Markovian and non-semimartingales making conventional iterative simulation schemes void. Additionally, our approach is applicable to non-Gaussian BS processes (obtained
through volatility modulation) as well making its scope even wider. For these reasons, we believe that this research paper should be of great general interest.

Chapter 2 (Bennedsen, 2016) is concerned with the statistical properties of a widely used semiparametric estimator of the roughness index $\alpha$. Using theory on (conditionally) Gaussian processes with stationary increments developed in Barndorff-Nielsen, Corcuera, and Podolskij (2009, 2011), we present a flexible framework for conducting estimation and inference on $\alpha$. Particular focus is given to constructing confidence intervals and to testing hypotheses of the form $H_0 : \alpha = \alpha_0$ for some $\alpha_0 \in (-1/2, 1/2)$, and it is shown how this can be used to test whether a stochastic process is a non-semimartingale. The paper also considers two empirical applications of the methods to financial data; most importantly, we find evidence that the stochastic volatility of financial assets is rough, which supports a conjecture made in Gatheral, Jaisson, and Rosenbaum (2014). The novel contribution of this paper is to show and emphasize that, under suitable assumptions, all rough processes give rise to the same asymptotics of the semiparametric estimator of $\alpha$. That is, it is not necessary to know, or assume, which data generating process underlies the data before performing inference. We believe these results — together with the extension to non-Gaussian processes — will make the paper an important resource for practitioners and researchers alike.

Chapter 3 (Bennedsen, Lunde, and Pakkanen, 2016b) is concerned with the volatility of financial assets; it is half-part empirical investigations and half-part mathematical modelling. Using ultra-high-frequency data of the extremely liquid E-mini S&P 500 futures contract, we first investigate the empirical properties of intraday volatility. We find evidence that the stochastic process driving (log) volatility is very rough and very persistent. We then study daily volatility of more than 5 000 U.S. equities from the Trades and Quotes database and confirm — without exception — the findings from the E-mini data. That is, roughness and persistence seem to be universal features of volatility. As existing mathematical models of stochastic volatility are unable to parsimoniously capture both roughness and persistence, we suggest new mathematical models to close this gap in the literature. We argue that the $BIS$ process is a particularly promising candidate for such a model and we prove a number of theoretical properties of this process. Finally, we illustrate the usefulness of our new models in a forecasting exercise, where it is found that our proposed models of volatility outperform state-of-the-art benchmarks in volatility forecasting, especially for intraday horizons. The modelling of stochastic volatility by rough processes is currently a vibrant and active area of research, (e.g., Gatheral et al., 2014; Guennoun, Jacquier, and Roome, 2014; Fukasawa, 2015; Bayer et al., 2016; Euch, Fukasawa, and Rosenbaum, 2016; Jaisson and Rosenbaum, 2016), and we expect our work to be an important contribution to this growing field. In particular, we think that the thorough empirical work, as well as the modelling contribution, will make lasting impact and serve as catalyst for further research.
References


Danish summary

Denne afhandling består af tre uafhængige kapitler i form af tre originale forskningsprojekter. Det fælles tema for afhandlingen er konceptet om en ru proces i kontinueret tid. Med andre ord, så er det matematiske objekt, som vi er interesserede i, en bestemt klasse af stokastiske processor: processorne er formulert i kontinueret tid; de tager værdier i de reelle tal; og de er kontinuerte. Faktisk, for et $\alpha \in (-1/2, 1/2)$, så er de $\phi$-Hölder kontinuerte for alle $\phi \in (0, \alpha + 1/2)$. Vi kalder parametren $\alpha$ for ruhedsindekset ("roughness index"), da det kontrollerer graden af ruhed af udfaldsstierne af den stokastiske proces. Når $\alpha$ er negativ, vil den stokastiske proces have udfaldsstier som er Hölder kontinuerte af en lavere orden end udfaldsstierne af en Brownsk bevægelse. Se fx Figur 2.1. i Kapitel 2 for illustrationer af ru processor. Den matematiske udfordring ved at arbejde med disse processor kommer hovedsageligt af det faktum, at når $\alpha \neq 0$, vil processorne være ikke-semimartingaler. Selvom mange af resultaterne i denne afhandling gælder for ruhedsindeks i hele $\alpha \in (-1/2, 1/2)$, er vi primært interesserede i det ru tilfælde, altså i $\alpha \in (-1/2, 0)$.


bred klasse af ikke-Gaussiske processer (defineret via volatilitetsmodulering). Af disse grunde mener vi, at forskningen i dette kapitel vil være af bred og generel interesse.

Kapitel 2 (Bennedsen, 2016) beskæftiger sig med de statistiske egenskaber af en meget anvendt semiparametrisk estimator af ruhedsindekset $\alpha$. Ved brug af teori ved-rørende Gaussiske processer med stationære tilvækster udviklet i Barndorff-Nielsen et al. (2009, 2011), præsenterer vi et fleksibelt setup til at estimere og udføre inferens på $\alpha$. Særlig fokus giver vi til at konstruere konfidensintervaller og til at teste hypoteser såsom $H_0 : \alpha = \alpha_0$ for et $\alpha_0 \in (-1/2, 1/2)$, og vi viser, hvordan dette kan bruges til at teste om en stokastisk proces er en ikke-semimartingal. Papiret betrægter også to empiriske applikationer af metoderne til finansiel data; vigtigst, så finder vi tegn på, at den stokastiske volatilitet underliggende finansielle aktiver er ru, hvilket understøtter en hypotese fremført i Gatheral et al. (2014). Det nye bidrag i denne artikel er at vise, og lægge vægt på, at under passinge antagelser giver alle ru processer anledning til den samme asymptotiske fordeling af den semiparametriske estimator af $\alpha$. Det vil sige, at det ikke er nødvendigt at vide, eller antage, hvilken data genererende proces, som underligger observationerne fra et eksperiment, før der udføres inferens. Vi mener, at disse resultater — sammen med en udvidelse til ikke-Gaussiske processer — vil gøre artiklen til en vigtig ressource for både praktiserende og teoretiske forskere.

Kapitel 3 (Bennedsen et al., 2016b) beskæftiger sig med stokastisk volatilitet af finansielle aktiver; artiklen består halvt af empiriske undersøgelser og halvt af matematisk modellering. Ved brug af ultra-højfrekvent data fra den meget likvide E-mini S&P 500 futures kontrakt undersøger vi først de empiriske egenskaber af intradag volatilitet. Vi finder beviser for, at den stokastiske proces, som driver (log) volatiliteten er meget ru og meget persistent. Vi studerer derefter daglig volatilitet fra mere end 5 000 amerikanske aktier fra Trades and Quotes databasen, og vi bekræfter — uden undtagelse — konklusionerne fra E-mini datasættet. Med andre ord ruhed og persistens synes at være universelle egenskaber af volatilitet. Eksisterende matematiske modeller af stokastisk volatilitet er ude af stand til samtidigt at modellere både ruhed og persistens, og vi foreslår derfor nye matematiske modeller for at lukke dette hul i litteraturen. Vi argumenterer for, at $\mathcal{B}\mathcal{F}$ processen er en særdeles lovende kandidat for sådan en model og vi udleder en række teoretiske resultater vedrørende denne proces. Til sidst illustrerer vi brugbarheden af vores nye modeller i en forecastingsøvelse, hvor vi finder, at vores foreslåede modeller for volatiliteten udkonkurrerer state-of-the-art benchmarkmodeller i volatilitetsforecasting, især for intradaglige foreacasthorisonter. Modellering af stokastisk volatilitet af ru processer er i øjeblikket et meget aktivt forskningsområde (fx Gatheral et al., 2014; Guennoun et al., 2014; Fukasawa, 2015; Bayer et al., 2016; Euch et al., 2016; Jaisson and Rosenbaum, 2016), og vi forventer, at vores arbejde vil være et vigtigt bidrag til dette voksende felt. I særdeleshed tror vi, at vores grundige empiriske arbejde, såvel som vores modelbidrag, vil have en vedvarende indvirkning på, og fungere som katalysator for, fremtidig forskning.
Litteratur


Abstract

We introduce a simulation scheme for Brownian semistationary processes, which is based on discretizing the stochastic integral representation of the process in the time domain. We assume that the kernel function of the process is regularly varying at zero. The novel feature of the scheme is to approximate the kernel function by a power function near zero and by a step function elsewhere. The resulting approximation of the process is a combination of Wiener integrals of the power function and a Riemann sum, which is why we call this method a hybrid scheme. Our main theoretical result describes the asymptotics of the mean square error of the hybrid scheme and we observe that the scheme leads to a substantial improvement of accuracy compared to the ordinary forward Riemann-sum scheme, while having the same computational complexity. We exemplify the use of the hybrid scheme by two numerical experiments, where we examine the finite-sample properties of an estimator of the roughness parameter of a Brownian semistationary process and study Monte Carlo option
pricing in the rough Bergomi model of Bayer, Friz, and Gatheral (2015), respectively.

**Keywords**: Stochastic simulation; discretization; Brownian semistationary process; stochastic volatility; regular variation; estimation; option pricing; rough volatility; volatility smile.

**JEL Classification**: C22, G13, C13

**MSC 2010 Classification**: 60G12, 60G22, 65C20, 91G60, 62M09

### 1.1 Introduction

We study simulation methods for Brownian semistationary (BSP) processes, first introduced by Barndorff-Nielsen and Schmiegel (2007, 2009), which form a flexible class of stochastic processes that are able to capture some common features of empirical time series, such as stochastic volatility (intermittency), roughness, stationarity and strong dependence. By now these processes have been applied in various contexts, most notably in the study of turbulence in physics, see, e.g., Corcuera, Hedevang, Pakkanen, and Podolskij (2013) and Barndorff-Nielsen, Pakkanen, and Schmiegel (2014), and in finance as models of energy prices, see, e.g., Barndorff-Nielsen, Benth, and Veraart (2013a) and Bennedsen (2015). A BSP process $X$ is defined via the integral representation

$$X(t) = \int_{-\infty}^{t} g(t-s) \sigma(s) dW(s), \quad (1.1.1)$$

where $W$ is a two-sided Brownian motion providing the fundamental noise innovations, the amplitude of which is modulated by a stochastic volatility (intermittency) process $\sigma$ that may depend on $W$. This driving noise is then convolved with a deterministic kernel function $g$ that specifies the dependence structure of $X$. The process $X$ can also be viewed as a moving average of volatility-modulated Brownian noise and setting $\sigma(s) = 1$, we see that stationary Brownian moving averages are nested in this class of processes.

In the applications mentioned above, the case where $X$ is not a semimartingale is particularly relevant. This situation arises when the kernel function $g$ behaves like a power-law near zero; more specifically, when for some $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$,

$$g(x) \propto x^\alpha \quad \text{for small } x > 0. \quad (1.1.2)$$

The case $\alpha = -\frac{1}{6}$ is important in statistical modeling of turbulence (Corcuera et al., 2013) as it gives rise to processes that are compatible with Kolmogorov’s scaling law for ideal turbulence. Moreover, processes of similar type with $\alpha \approx -0.4$ have been recently used in the context of option pricing as models of rough volatility (Bayer et al., 2015; Gatheral, Jaiison, and Rosenbaum, 2014), see Sections 1.2 and 1.3 below.
The case $\alpha = 0$ would (roughly speaking) lead to a process that is a semimartingale, which is thus excluded. We formulate the relation (1.1.2) below rigorously using the theory of regular variation (Bingham, Goldie, and Teugels, 1989), which plays a significant role in our subsequent arguments.

Under (1.1.2), the trajectories of $X$ behave locally like the trajectories of a fractional Brownian motion with Hurst index $H = \alpha + \frac{1}{2} \in (0, 1) \setminus \{\frac{1}{2}\}$. While the local behavior and roughness, measured in terms of Hölder regularity, of $X$ are determined by the parameter $\alpha$, the global behavior of $X$ (e.g., whether the process has long or short memory) depends on the behavior of $g(x)$ as $x \to \infty$, which can be specified independently of $\alpha$. This should be contrasted with fractional Brownian motion and related self-similar models, which necessarily must conform to a restrictive affine relationship between their Hölder regularity (local behavior and roughness) and Hurst index (global behavior), see Gneiting and Schlather (2004). Indeed, in the realm of $\mathcal{BIS}$ processes, local and global behavior are conveniently decoupled, which underlines the flexibility of these processes as a modeling framework.

In connection with practical applications, it is important to be able to simulate the process $X$. If the volatility process $\sigma$ is deterministic and constant in time, then $X$ will be strictly stationary and Gaussian. This makes $X$ amenable to exact simulation, e.g., using the Cholesky factorization or circulant embeddings (Asmussen and Glynn, 2007, Chapter XI). However, it seems difficult, if not impossible, to develop an exact method that is applicable with a stochastic $\sigma$, as the process $X$ is then neither Markovian nor Gaussian. Thus, in the general case one must resort to approximative methods. To this end, Benth, Eyjolfsson, and Veraart (2014) have recently proposed a Fourier-based method of simulating $\mathcal{BIS}$ processes, and more general Lévy semistationary (LSS) processes, which relies on approximating the kernel function $g$ in the frequency domain.

In this paper, we introduce a new discretization scheme for $\mathcal{BIS}$ processes based on approximating the kernel function $g$ in the time domain. Our starting point is the Riemann-sum discretization of (1.1.1). The Riemann-sum scheme builds on an approximation of $g$ using step functions, which has the pitfall of failing to capture appropriately the steepness of $g$ near zero. In particular, this becomes a serious defect under (1.1.2) when $\alpha \in \left(-\frac{1}{2}, 0\right)$. In our new scheme, we mitigate this problem by approximating $g$ using an appropriate power function near zero and a step function elsewhere. The resulting discretization scheme can be realized as a linear combination of Wiener integrals with respect to the driving Brownian motion $W$ and a Riemann sum, which is why we call it a hybrid scheme. The hybrid scheme is only slightly more demanding to implement than the Riemann-sum scheme and the schemes have the same computational complexity as the number of discretization cells tends to infinity.

Our main theoretical result describes the exact asymptotic behavior of the mean square error (MSE) of the hybrid scheme and, as a special case, that of the Riemann-
sum scheme. We observe that switching from the Riemann-sum scheme to the hybrid scheme reduces the asymptotic root mean square error (RMSE) substantially. Using merely the simplest variant the of hybrid scheme, where a power function is used in a single discretization cell, the reduction is at least 50% for all \( \alpha \in \left(0, \frac{1}{2}\right) \) and at least 80% for all \( \alpha \in \left(- \frac{1}{2}, 0\right) \). The reduction in RMSE is close to 100% as \( \alpha \) approaches \(-\frac{1}{2}\), which indicates that the hybrid scheme indeed resolves the problem of poor precision that affects the Riemann-sum scheme.

To assess the accuracy of the hybrid scheme in practice, we perform two numerical experiments. Firstly, we examine the finite-sample performance of an estimator of the roughness index \( \alpha \), introduced by Barndorff-Nielsen, Corcuera, and Podolskij (2013b) and Corcuera et al. (2013). This experiment enables us to assess how faithfully the hybrid scheme approximates the fine properties of the BS S process \( X \). Secondly, we study Monte Carlo option pricing in the rough Bergomi stochastic volatility model of Bayer et al. (2015). We use the hybrid scheme to simulate the volatility process in this model and we find that the resulting implied volatility smiles are indistinguishable from those simulated using a method that involves exact simulation of the volatility process. Thus we are able propose a solution to the problem of finding an efficient simulation scheme for the rough Bergomi model, left open in the paper Bayer et al. (2015).

The rest of this paper is organized as follows. In Section 3.3 we recall the rigorous definition of a BS S process and introduce our assumptions. We also introduce the hybrid scheme, state our main theoretical result concerning the asymptotics of the mean square error and discuss an extension of the scheme to a class of truncated BS S processes. Section 1.3 briefly discusses the implementation of the discretization scheme and presents the numerical experiments mentioned above. Finally, Section 1.4 contains the proofs of the theoretical results given in the paper.

1.2 The model and theoretical results

Brownian semistationary process

Let \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}}, \mathbb{P})\) be a filtered probability space, satisfying the usual conditions, supporting a (two-sided) standard Brownian motion \( W = \{W(t)\}_{t \in \mathbb{R}} \). We consider a Brownian semistationary process

\[
X(t) = \int_{-\infty}^t g(t-s)\sigma(s)dW(s), \quad t \in \mathbb{R},
\]

where \( \sigma = \{\sigma(t)\}_{t \in \mathbb{R}} \) is an \( \{\mathcal{F}_t\}_{t \in \mathbb{R}} \)-predictable process with locally bounded trajectories, which captures the stochastic volatility (intermittency) of \( X \), and \( g : (0, \infty) \to [0, \infty) \) is a Borel measurable kernel function.

To ensure that the integral (1.2.1) is well-defined, we assume that the kernel function \( g \) is square integrable, that is, \( \int_0^\infty g(x)^2\,dx < \infty \). In fact, we will shortly
introduce some more specific assumptions on $g$ that will imply its square integrability. Throughout the paper, we also assume that the process $\sigma$ has finite second moments, $\mathbb{E}[\sigma(t)^2] < \infty$ for all $t \in \mathbb{R}$, and that the process is covariance stationary, namely,

$$
\mathbb{E}[\sigma(s)] = \mathbb{E}[\sigma(t)], \quad \text{Cov}(\sigma(s), \sigma(t)) = \text{Cov}(\sigma(0), \sigma(|s-t|)), \quad s, t \in \mathbb{R}.
$$

These assumptions imply that also $X$ is covariance stationary, that is,

$$
\mathbb{E}[X(t)] = 0, \quad \text{Cov}(X(s), X(t)) = \mathbb{E}[\sigma(0)^2] \int_0^\infty g(x)g(x+|s-t|)dx, \quad s, t \in \mathbb{R}.
$$

However, the process $X$ need not be strictly stationary as the dependence between the volatility process $\sigma$ and the driving Brownian motion $W$ may be time-varying.

**Kernel function**

As mentioned above, we consider a kernel function that satisfies $g(x) \propto x^\alpha$ for some $\alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ when $x > 0$ is near zero. To make this idea rigorous and to allow for additional flexibility, we formulate our assumptions on $g$ using the theory of regular variation (Bingham et al., 1989) and, more specifically, slowly varying functions.

To this end, recall that a measurable function $L : (0, 1] \to [0, \infty)$ is *slowly varying* at 0 if for any $t > 0$,

$$
\lim_{x \to 0} \frac{L(tx)}{L(x)} = 1.
$$

Moreover, a function $f(x) = x^\beta L(x)$, $x \in (0, 1]$, where $\beta \in \mathbb{R}$ and $L$ is slowly varying at 0, is said to be *regularly varying* at 0, with $\beta$ being the *index of regular variation*.

**Remark 1.2.1.** Conventionally, slow and regular variation are defined at $\infty$ (Bingham et al., 1989, pp. 6, 17–18). However, $L$ is slowly varying (resp. regularly varying) at 0 if and only if $x \to L(1/x)$ is slowly varying (resp. regularly varying) at $\infty$.

A key feature of slowly varying functions, which will be very important in the sequel, is that they can be sandwiched between polynomial functions as follows. If $\delta > 0$ and $L$ is slowly varying at 0 and bounded away from 0 and $\infty$ on any interval $(u, 1], u \in (0, 1)$, then there exist constants $C_\delta, C_{-\delta} > 0$ such that

$$
C_\delta x^\delta \leq L(x) \leq C_{-\delta} x^{-\delta}, \quad x \in (0, 1]. \quad (1.2.2)
$$

The inequalities above are an immediate consequence of the so-called *Potter bounds* for slowly varying functions (Bingham et al., 1989, Theorem 1.5.6(ii)). Making $\delta$ very small therein, we see that slowly varying functions are asymptotically negligible in comparison with polynomially growing/decaying functions. Thus, by multiplying power functions and slowly varying functions, regular variation provides a flexible framework to construct functions that behave asymptotically like power functions.

Our assumptions concerning the kernel function $g$ are as follows:
(A1) For some $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$,
\[ g(x) = x^\alpha L_g(x), \quad x \in (0, 1], \]
where $L_g : (0, 1] \to [0, \infty)$ is continuously differentiable, slowly varying at 0 and bounded away from 0. Moreover, there exists a constant $C > 0$ such that the derivative $L_g'$ of $L_g$ satisfies
\[ |L_g'(x)| \leq C(1 + x^{-1}), \quad x \in (0, 1]. \]

(A2) The function $g$ is continuously differentiable on $(0, \infty)$, so that its derivative $g'$ is ultimately monotonic and satisfies $\int_1^\infty g'(x)^2 \, dx < \infty$.

(A3) For some $\beta \in (-\infty, -\frac{1}{2})$,
\[ g(x) = O(x^\beta), \quad x \to \infty. \]

(We use $f(x) = O(h(x)), \; x \to a$, to indicate that $\limsup_{x \to a} \frac{|f(x)|}{h(x)} < \infty$. Additionally, analogous notation is later used for sequences and computational complexity.) In view of the bound (1.2.2), these assumptions ensure that $g$ is square integrable. It is worth pointing out that (A1) accommodates functions $L_g$ with $\lim_{x \to 0} L_g(x) = \infty$, e.g., $L_g(x) = 1 - \log x$.

The assumption (A1) influences the short-term behavior and roughness of the process $X$. A simple way to assess the roughness of $X$ is to study the behavior of its variogram (also called the second-order structure function in turbulence literature)
\[ V_X(h) := E[|X(h) - X(0)|^2], \quad h \geq 0, \]
as $h \to 0$. Note that, by covariance stationarity,
\[ V_X(|s - t|) = E[|X(s) - X(t)|^2], \quad s, t \in \mathbb{R}. \]

Under our assumptions, we have the following characterization of the behavior of $V_X$ near zero, which generalizes a result of Barndorff-Nielsen (2012, p. 9) and implies that $X$ has a H"{o}lder continuous modification. Its proof is carried out in Section 1.4.

**Proposition 1.2.1** (Local behavior and continuity). *Suppose that (A1), (A2) and (A3) hold.*

(i) The variogram of $X$ satisfies
\[ V_X(h) \sim E[\sigma(0)^2] \left( \frac{1}{2\alpha + 1} + \int_0^\infty (y + 1)^\alpha - y^\alpha \, dy \right) h^{2\alpha + 1} L_g(h)^2, \quad h \to 0, \]
which implies that $V_X$ is regularly varying at zero with index $2\alpha + 1$. 
(ii) The process $X$ has a modification with locally $\phi$-Hölder continuous trajectories for any $\phi \in \left(0, \alpha + \frac{1}{2}\right)$.

Motivated by Proposition 1.2.1, we call $\alpha$ the roughness index of the process $X$. Ignoring the slowly varying factor $L_g(h)^2$ in (1.2.1), we see that the variogram $V(h)$ behaves like $h^{2\alpha+1}$ for small values of $h$, which is reminiscent of the scaling property of the increments of a fractional Brownian motion (fBm) with Hurst index $H = \alpha + \frac{1}{2}$. Thus, the process $X$ behaves locally like such an fBm, at least when it comes to second order structure and roughness. (Moreover, the factor $\frac{1}{2\alpha+1} + \int_0^\infty (y + 1)^\alpha - y^\alpha)^2 dy$ coincides with the normalization coefficient that appears in the Mandelbrot–Van Ness representation (Mishura, 2008, Theorem 1.3.1) of an fBm with $H = \alpha + \frac{1}{2}$.)

Let us now look at two examples of a kernel function $g$ that satisfies our assumptions.

**Example 1.2.1** (The gamma kernel). The so-called gamma kernel

$$g(x) = x^\alpha e^{-\lambda x}, \quad x \in (0, \infty),$$

with parameters $\alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ and $\lambda > 0$, has been used extensively in the literature on $\mathcal{BSP}$ processes. It is particularly important in connection with statistical modeling of turbulence (see Corcuera et al., 2013), but it also provides a way to construct generalizations of Ornstein–Uhlenbeck (OU) processes with roughness that differs from the usual semimartingale case $\alpha = 0$, while mimicking the long-term behavior of an OU process. Moreover, $\mathcal{BSP}$ and $\mathcal{LPSP}$ processes defined using the gamma kernel have interesting probabilistic properties, see Pedersen and Sauri (2015). An in-depth study of the gamma kernel can be found in Barndorff-Nielsen (2012). Setting $L_g(x) := e^{-\lambda x}$, which is slowly varying at 0 since $\lim_{x \to 0} L_g(x) = 1$, it is evident that (A1) holds. Since $g(x)$ decays exponentially fast to 0 as $x \to \infty$, it is clear that also (A3) holds. To verify (A2), note that $g$ satisfies

$$g'(x) = \left(\frac{\alpha}{x} - \lambda\right) g(x), \quad g''(x) = \left(\left(\frac{\alpha}{x} - \lambda\right)^2 - \frac{\alpha}{x^2}\right) g(x), \quad x \in (0, \infty),$$

where $\lim_{x \to \infty} ((\frac{\alpha}{x} - \lambda)^2 - \frac{\alpha}{x^2}) = \lambda^2 > 0$, so $g'$ is ultimately increasing with

$$g'(x) \leq (|\alpha| + \lambda)^2 g(x)^2, \quad x \in [1, \infty).$$

Thus, $\int_1^\infty g'(x)^2 dx < \infty$ since $g$ is square integrable.

**Example 1.2.2** (Power-law kernel). Consider the kernel function

$$g(x) = x^\alpha (1 + x)^{\beta-\alpha}, \quad x \in (0, \infty),$$

with parameters $\alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ and $\beta \in (-\infty, -\frac{1}{2})$. The behavior of this kernel function near zero is similar to that of the gamma kernel, but $g(x)$ decays to zero polynomially as $x \to \infty$, so it can be used to model long memory. In fact, it can be shown
that if $\beta \in (-1, -\frac{1}{2})$, then the autocorrelation function of $X$ is not integrable. Clearly, (A1) holds with $L_g(x) := (1 + x)^{\beta - \alpha}$, which is slowly varying at 0 since $\lim_{x \to 0} L_g(x) = 1$. Moreover, note that we can write
\[ g(x) = x^\beta K_g(x), \quad x \in (0, \infty), \]
where $K_g(x) := (1 + x^{-1})^{\beta - \alpha}$ satisfies $\lim_{x \to \infty} K_g(x) = 1$. Thus, also (A3) holds. We can check (A2) by computing
\[ g'(x) = \left(\frac{\alpha + \beta x}{x(1 + x)}\right) g(x), \quad g''(x) = \left(\frac{\alpha + \beta x}{x(1 + x)}\right)^2 + \frac{-\alpha - 2\alpha x - \beta x^2}{x^2(1 + x)^2} g(x), \quad x \in (0, \infty), \]
where $-\alpha - 2\alpha x - \beta x^2 \to \infty$ when $x \to \infty$ (as $\beta < -\frac{1}{2}$), so $g'$ is ultimately increasing. Additionally, we note that
\[ g'(x)^2 \leq (|\alpha| + |\beta|)^2 g(x)^2, \quad x \in [1, \infty), \]
implies $\int_1^\infty g'(x)^2 \, dx < \infty$ since $g$ is square integrable.

**Hybrid scheme**

Let $t \in \mathbb{R}$ and consider discretizing $X(t)$ based on its integral representation (1.2.1) on the grid $\mathcal{G}_n(t) := \{t, t - \frac{1}{n}, t - \frac{2}{n}, \ldots\}$ for $n \in \mathbb{N}$. To derive our discretization scheme, let us first note that if the volatility process $\sigma$ does not vary too much, then it is reasonable to use the approximation
\[ X(t) = \sum_{k=1}^\infty \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} g(t - s) \sigma(s) \, dW(s) \approx \sum_{k=1}^\infty \sigma\left(t - \frac{k}{n}\right) \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} g(t - s) \, dW(s), \quad (1.2.3) \]
that is, we keep $\sigma$ constant in each discretization cell. If $k$ is “small”, then due to (A1) we may approximate
\[ g(t - s) \approx (t - s)^{\alpha} L_g\left(\frac{k}{n}\right), \quad t - s \in \left[\frac{k - 1}{n}, \frac{k}{n}\right] \setminus \{0\}, \quad (1.2.4) \]
as the slowly varying function $L_g$ varies “less” than the power function $y \to y^\alpha$ near zero, cf. (1.2.2). If $k$ is “large”, or at least $k \geq 2$, then choosing $b_k \in [k - 1, k]$ provides an adequate approximation
\[ g(t - s) \approx g\left(\frac{b_k}{n}\right), \quad t - s \in \left[\frac{k - 1}{n}, \frac{k}{n}\right], \quad (1.2.5) \]
by (A2). Applying (1.2.4) to the first $\kappa$ terms, where $\kappa = 1, 2, \ldots$, and (1.2.5) to the remaining terms in the approximating series in (1.2.3) yields
\[
\sum_{k=1}^\infty \sigma\left(t - \frac{k}{n}\right) \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} g(t - s) \, dW(s) \approx \sum_{k=1}^\kappa L_g\left(\frac{k}{n}\right) \sigma\left(t - \frac{k}{n}\right) \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} (t - s)^{\alpha} \, dW(s) + \sum_{k=\kappa+1}^\infty g\left(\frac{b_k}{n}\right) \sigma\left(t - \frac{k}{n}\right) \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} dW(s),
\]
(1.2.6)
For completeness, we also allow for $\kappa = 0$, in which case we require that $b_1 \in (0, 1)$ and interpret the first sum on the right-hand side of (1.2.6) as zero. To make numerical implementation feasible, we truncate the second sum on the right-hand side of (1.2.6) so that both sums have $N_n \geq \kappa + 1$ terms in total. Thus, we arrive at a discretization scheme for $X(t)$, which we call a *hybrid scheme*, given by

$$X_n(t) := \check{X}_n(t) + \hat{X}_n(t),$$

(1.2.7)

where

$$\check{X}_n(t) := \sum_{k=1}^{\kappa} L g \left( \frac{k}{n} \right) \sigma \left( t - \frac{k}{n} \right) \int_{t - \frac{k}{n}}^{t - \frac{k - 1}{n}} (t - s)^{\alpha} dW(s),$$

(1.2.8)

$$\hat{X}_n(t) := \sum_{k=\kappa+1}^{N_n} g \left( \frac{b_k}{n} \right) \sigma \left( t - \frac{k}{n} \right) \left( W \left( t - \frac{k}{n} + \frac{1}{n} \right) - W \left( t - \frac{k}{n} \right) \right),$$

(1.2.9)

and $b := \{b_k \}_{k=\kappa+1}^{\infty}$ is a sequence of real numbers, evaluation points, that must satisfy $b_k \in [k - 1, k] \setminus \{0\}$ for each $k \geq \kappa + 1$, but otherwise can be chosen freely.

As it stands, the discretization grid $G_n(t)$ depends on the time $t$, which may seem cumbersome with regard to sampling $X_n(t)$ simultaneously for different times $t$. However, note that whenever times $t$ and $t'$ are separated by a multiple of $\frac{1}{n}$, the corresponding grids $G_n(t)$ and $G_n(t')$ will intersect. In fact the hybrid scheme defined by (1.2.8) and (1.2.9) can be implemented efficiently, as we shall see in Section 1.3, below. Since

$$g \left( \frac{b_k}{n} \right) = g \left( t - \left( t - \frac{b_k}{n} \right) \right),$$

the degenerate case $\kappa = 0$ with $b_k = k$ for all $k \geq 1$ corresponds to the usual Riemann-sum discretization scheme of $X(t)$ with (Itô type) forward sums from (1.2.9). Henceforth, we denote the associated sequence $\{k\}_{k=\kappa+1}^{\infty}$ by $b_{\text{FWD}}$, where the subscript “FWD” refers to forward sums. However, including terms involving Wiener integrals of a power function given by (1.2.8), that is having $\kappa \geq 1$, improves the accuracy of the discretization considerably, as we shall see. Having the leeway to select $b_k$ within the interval $[k - 1, k] \setminus \{0\}$, so that the function $g(t - \cdot)$ is evaluated at a point that does not necessarily belong to $G_n(t)$, leads additionally to a moderate improvement.

The truncation in the sum (1.2.9) entails that the stochastic integral (1.2.1) defining $X$ is truncated at $t - \frac{N_n}{n}$. In practice, the value of the parameter $N_n$ should be large enough to mitigate the effect of truncation. To ensure that the truncation point $t - \frac{N_n}{n}$ tends to $-\infty$ as $n \to \infty$ in our asymptotic results, we introduce the following assumption:

(A4) For some $\gamma > 0$,

$$N_n \sim n^{\gamma + 1}, \; n \to \infty.$$
Asymptotic behavior of mean square error

We are now ready to state our main theoretical result, which gives a sharp description of the asymptotic behavior of the mean square error (MSE) of the hybrid scheme as \( n \to \infty \). We defer the proof of this result to Section 1.4.

**Theorem 1.2.1** (Asymptotics of mean square error). Suppose that (A1), (A2), (A3) and (A4) hold, so that

\[
\gamma > -\frac{2\alpha + 1}{2\beta + 1},
\]

and that for some \( \delta > 0 \),

\[
\mathbb{E}[|\sigma(s) - \sigma(0)|^2] = \mathcal{O}(s^{2\alpha+1+\delta}), \quad s \downarrow 0.
\]

Then for all \( t \in \mathbb{R} \),

\[
\mathbb{E}[|X(t) - X_n(t)|^2] \sim J(\alpha, \kappa, b)\mathbb{E}[\sigma(0)^2] n^{-(2\alpha+1)} L_g(1/n)^2, \quad n \to \infty,
\]

where

\[
J(\alpha, \kappa, b) := \sum_{k=\kappa+1}^{\infty} \int_{k-1}^{k} (y^{\alpha} - b_k^{\alpha})^2 dy < \infty.
\]

**Remark 1.2.2.** Note that if \( \alpha \in \left(-\frac{1}{2}, 0\right) \), then having

\[
\mathbb{E}[|\sigma(s) - \sigma(0)|^2] = \mathcal{O}(s^{\theta}), \quad s \downarrow 0,
\]

for all \( \theta \in (0, 1) \), ensures that (1.2.11) holds. (Take, say, \( \delta := \frac{1}{2}(1 - (2\alpha + 1)) > 0 \) and \( \theta := 2\alpha + 1 + \delta = \alpha + 1 \in (0, 1) \).)

In Theorem 1.2.1, the asymptotics of the MSE (1.2.12) are determined by the behavior of the kernel function \( g \) near zero, as specified in (A1). The condition (1.2.10) ensures that error from approximating \( g \) near zero is asymptotically larger than the error induced by the truncation of the stochastic integral (1.2.1) at \( t - \frac{N_p}{n} \). In fact, different kind of asymptotics of the MSE, where truncation error becomes dominant, could be derived when (1.2.10) does not hold, under some additional assumptions, but we do not pursue this direction in the present paper.

While the rate of convergence in (1.2.12) is fully determined by the roughness index \( \alpha \), which may seem discouraging at first, it turns out that the quantity \( J(\alpha, \kappa, b) \), which we shall call the asymptotic MSE, can vary a lot, depending on how we choose \( \kappa \) and \( b \), and can have a substantial impact on the precision of the approximation of \( X \). It is immediate from (1.2.13) that increasing \( \kappa \) will decrease \( J(\alpha, \kappa, b) \). Moreover, for given \( \alpha \) and \( \kappa \), it is straightforward to choose \( b \) so that \( J(\alpha, \kappa, b) \) is minimized, as shown in the following result.
Proposition 1.2.2 (Optimal discretization). Let $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ and $\kappa \geq 0$. Among all sequences $b = \{b_k\}_{k=\kappa+1}^\infty$ with $b_k \in [k-1, k] \setminus \{0\}$ for $k \geq \kappa + 1$, the function $J(\alpha, \kappa, b)$, and consequently the asymptotic MSE induced by the discretization, is minimized by the sequence $b^*$ given by

$$b^*_k = \left(\frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{\alpha + 1}\right)^{1/\alpha}, \quad k \geq \kappa + 1.$$

Proof. Clearly, a sequence $b = \{b_k\}_{k=\kappa+1}^\infty$ minimizes the function $J(\alpha, \kappa, b)$ if and only if $b_k$ minimizes $\int_{k-1}^k (y^\alpha - b_k^\alpha)^2 \, dy$ for any $k \geq \kappa + 1$. By standard $L^2$-space theory, $c \in \mathbb{R}$ minimizes the integral $\int_{k-1}^k (y^\alpha - c)^2 \, dy$ if and only if the function $y \mapsto y^\alpha - c$ is orthogonal in $L^2$ to all constant functions. This is tantamount to

$$\int_{k-1}^k (y^\alpha - c) \, dy = 0,$$

and computing the integral and solving for $c$ yields

$$c = \frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{\alpha + 1}.$$

Setting $b^*_k := c^{1/\alpha} \in (k-1, k)$ completes the proof. \qed

To understand how much increasing $\kappa$ and using the optimal sequence $b^*$ from Proposition 1.2.2 improves the approximation, we study numerically the asymptotic root mean square error (RMSE) $\sqrt{J(\alpha, \kappa, b)}$. In particular, we assess how much the asymptotic RMSE decreases relative to RMSE of the forward Riemann-sum scheme ($\kappa = 0$ and $b = b_{\text{FWD}}$) using the quantity

$$\text{reduction in asymptotic RMSE} = \frac{\sqrt{J(\alpha, \kappa, b)} - \sqrt{J(\alpha, 0, b_{\text{FWD}})}}{\sqrt{J(\alpha, 0, b_{\text{FWD}})}} \cdot 100\%. \quad (1.2.14)$$

The results are presented in Figure 1.1. We find that employing the hybrid scheme with $\kappa \geq 1$ leads to a substantial reduction in the asymptotic RMSE relative to the forward Riemann-sum scheme when $\alpha \in (-\frac{1}{2}, 0)$. Indeed, when $\kappa \geq 1$, the asymptotic RMSE, as a function of $\alpha$, does not blow up as $\alpha \to -\frac{1}{2}$, while with $\kappa = 0$ it does. This explains why the reduction in the asymptotic RMSE approaches 100% as $\alpha \to -\frac{1}{2}$. When $\alpha \in (0, \frac{1}{2})$, the improvement achieved using the hybrid scheme is more modest, but still considerable. Figure 1.1 also highlights the importance of using the optimal sequence $b^*$, instead of $b_{\text{FWD}}$, as evaluation points in the scheme, in particular when $\alpha \in (0, \frac{1}{2})$. Finally, we observe that increasing $\kappa$ beyond 2 does not appear to lead to a significant further reduction. Indeed, in our numerical experiments, reported in Section 1.3 and 1.3 below, we observe that using $\kappa = 1, 2$ already leads to good results.
Remark 1.2.3. It is non-trivial to evaluate the quantity $J(\alpha, \kappa, b)$ numerically. Computing the integral in (1.2.13) explicitly, we can approximate $J(\alpha, \kappa, b)$ by

$$J_N(\alpha, \kappa, b) := \sum_{k=\kappa+1}^{\infty} \left( \frac{k^{2\alpha+1} - (k-1)^{2\alpha+1}}{2\alpha + 1} - \frac{2b_k^{\alpha}(k^{\alpha+1} - (k-1)^{\alpha+1})}{\alpha + 1} + b_k^{2\alpha} \right)$$

with some large $N \in \mathbb{N}$. This approximation is adequate when $\alpha \in (-\frac{1}{2}, 0)$, but its accuracy deteriorates when $\alpha \to \frac{1}{2}$. In particular, the singularity of the function $\alpha \to J(\alpha, \kappa, b)$ at $\frac{1}{2}$ is difficult to capture using $J_N(\alpha, \kappa, b)$ with numerically feasible values of $N$. To overcome this numerical problem, we introduce a correction term in the case $\alpha \in (0, \frac{1}{2})$. The correction term can be derived informally as follows. By the mean value theorem, and since $b_k^{*} \approx k - \frac{1}{2}$ for large $k$, we have

$$(y^\alpha - b_k^{\alpha})^2 = \alpha^2 \xi^{2\alpha-2}(y - b_k)^2 \approx \begin{cases} \alpha^2 k^{2\alpha-2}(y - k)^2, & b = b_{\text{FWD}}, \\ \alpha^2 k^{2\alpha-2}(y - k + \frac{1}{2})^2, & b = b^{*}, \end{cases}$$

where $\xi = \xi(y, b_k) \in [k - 1, k]$, for large $k$. Thus, for large $N$, we obtain

$$(y^\alpha - b_k^{\alpha})^2 \approx \begin{cases} \alpha^2 \sum_{k=N+1}^{\infty} k^{2\alpha-2} \int_{k-1}^{k} (y - k)^2 dy, & b = b_{\text{FWD}}, \\ \alpha^2 \sum_{k=N+1}^{\infty} k^{2\alpha-2} \int_{k-1}^{k} (y - k + \frac{1}{2})^2 dy, & b = b^{*}, \end{cases}$$

and

$$(y^\alpha - b_k^{\alpha})^2 \approx \begin{cases} \frac{\alpha^2}{3} \xi(2 - 2\alpha, N+1), & b = b_{\text{FWD}}, \\ \frac{\alpha^2}{12} \xi(2 - 2\alpha, N+1), & b = b^{*}, \end{cases}$$

for large $N$. In all computations, we have used the approximations outlined in Remark 1.2.3 with $N = 100000$. 

Figure 1.1. Left: The asymptotic RMSE given by $\sqrt{J(\alpha, \kappa, b)}$ as a function of $\alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ for $\kappa = 0, 1, 2, 3$ using $b = b^{*}$ of Proposition 1.2.2 (solid lines) and $b = b_{\text{FWD}}$ (dashed lines). Right: Reduction in the asymptotic RMSE relative to the forward Riemann-sum scheme ($\kappa = 0$ and $b = b_{\text{FWD}}$) given by the formula (1.2.14), plotted as a function of $\alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ for $\kappa = 0, 1, 2, 3$ using $b = b^{*}$ (solid lines) and for $\kappa = 1, 2, 3$ using $b = b_{\text{FWD}}$ (dashed lines). In all computations, we have used the approximations outlined in Remark 1.2.3 with $N = 100000$. 


where \( \zeta(x, s) := \sum_{k=0}^{\infty} \frac{1}{(k+s)^x}, \ x > 1, \ s > 0, \) is the *Hurwitz zeta function*, which can be evaluated using accurate numerical algorithms.

**Remark 1.2.4.** Unlike the Fourier-based method of Benth et al. (2014), the hybrid scheme does not require truncating the singularity of the kernel function \( g \) when \( \alpha \in (-\frac{1}{2}, 0) \), which is beneficial to maintaining the accuracy of the scheme when \( \alpha \) is near \(-\frac{1}{2}\). Let us briefly analyze the effect of truncating the singularity of \( g \) on the approximation error, cf. Benth et al. (2014, pp. 75–76). Consider, for any \( \varepsilon > 0 \), the modified BS process \( \tilde{X}_\varepsilon(t) := \int_{-\infty}^{t} g_\varepsilon(t-s)\sigma(s)\,dW(s), \ t \in \mathbb{R}, \)

defined using the truncated kernel function

\[
g_\varepsilon(x) := \begin{cases} 
g(\varepsilon), & x \in (0, \varepsilon], \\
g(x), & x \in (\varepsilon, \infty). \end{cases}
\]

Adapting the proof of Theorem 1.2.1 in a straightforward manner, it is possible to show that, under (A1) and (A3),

\[
\mathbb{E}[\left|X(t) - \tilde{X}_\varepsilon(t)\right|^2] = \mathbb{E}[\sigma(0)^2] \int_0^\varepsilon (g(s) - g(\varepsilon))^2 \, ds \\
\sim \left( \frac{1}{2\alpha + 1} - \frac{2}{\alpha + 1} + 1 \right) \mathbb{E}[\sigma(0)^2] \varepsilon^{2\alpha+1} L_g(\varepsilon)^2, \ \varepsilon \downarrow 0,
\]

for any \( t \in \mathbb{R} \). While the rate of convergence, as \( \varepsilon \downarrow 0 \), of the MSE that arises from replacing \( g \) with \( g_\varepsilon \) is analogous to the rate of convergence of the hybrid scheme, it is important to note that the factor \( \tilde{J}(\alpha) \) blows up as \( \alpha \downarrow -\frac{1}{2} \). In fact, \( \tilde{J}(\alpha) \) is equal to the first term in the series that defines \( J(\alpha, 0, b_{\text{FWD}}) \) and

\[
\tilde{J}(\alpha) \sim J(\alpha, 0, b_{\text{FWD}}), \ \alpha \downarrow -\frac{1}{2},
\]

which indicates that the effect of truncating the singularity, in terms of MSE, is similar to the effect of using the forward Riemann–sum scheme to discretize the process when \( \alpha \) is near \(-\frac{1}{2}\). In particular, the truncation threshold \( \varepsilon \) would then have to be very small in order to keep the truncation error in check.

**Extension to truncated Brownian semistationary processes**

It is useful to extend the hybrid scheme to a class of non-stationary processes that are closely related to \( \mathcal{B}\mathcal{S}\mathcal{P} \) processes. This extension is important in connection with an application to the so-called rough Bergomi model, which we discuss in Section 1.3, below. More precisely, we consider processes of the form

\[
Y(t) = \int_0^t g(t-s)\sigma(s)\,dW(s), \ t \geq 0,
\]
where the kernel function $g$, volatility process $\sigma$ and driving Brownian motion $W$ are as before. We call $Y$ a truncated Brownian semistationary ($\mathcal{TBS}$) process, as $Y$ is obtained from the $\mathcal{BS}$ process $X$ by truncating the stochastic integral in (1.2.1) at 0. Of the preceding assumptions, only (A1) and (A2) are needed to ensure that the stochastic integral in (1.2.15) exists — in fact, of (A2), only the requirement that $g$ is differentiable on $(0,\infty)$ comes into play.

The $\mathcal{TBS}$ process $Y$ does not have covariance stationary increments, so we define its (time-dependent) variogram as

$$V_Y(h,t) := \mathbb{E}[|Y(t+h) - Y(t)|^2], \quad h, t \geq 0.$$ 

Extending Proposition 1.2.1, we can describe the behavior of $h \to V_Y(h,t)$ near zero as follows. The existence of a Hölder continuous modification is then a straightforward consequence. We omit the proof of this result, as it would be straightforward adaptation of the proof of Proposition 1.2.1.

**Proposition 1.2.3** (Local behavior and continuity). Suppose that (A1) and (A2) hold.

(i) The variogram of $Y$ satisfies for any $t \geq 0$,

$$V_Y(h,t) \sim \mathbb{E}[\sigma(0)^2] \left( \frac{1}{2\alpha+1} + 1_{(0,\infty)}(t) \int_0^\infty (y+1)^\alpha - y^\alpha d\gamma \right) h^{2\alpha+1} L_g(h)^2,$$

as $h \to 0$, which implies that $h \to V_Y(h,t)$ is regularly varying at zero with index $2\alpha+1$.

(ii) The process $Y$ has a modification with locally $\phi$-Hölder continuous trajectories for any $\phi \in \left(0,\alpha + \frac{1}{2}\right)$.

Note that while the increments of $Y$ are not covariance stationary, the asymptotic behavior of $V_Y(h,t)$ is the same as that of $V_X(h)$ as $h \to 0$ (cf. Proposition 1.2.1) for any $t > 0$. Thus, the increments of $Y$ (apart from increments starting at time 0) are locally like the increments of $X$.

We define the hybrid scheme to discretize $Y(t)$, for any $t \geq 0$, as

$$Y_n(t) := \bar{Y}_n(t) + \hat{Y}_n(t),$$

where

$$\bar{Y}_n(t) := \sum_{k=1}^{\lfloor |nt| \rfloor} L_g \left( \frac{k}{n} \right) \sigma \left( t - \frac{k}{n} \right) \int_{t - \frac{k}{n}}^{t - \frac{k}{n} + \frac{1}{n}} (t-s)^\alpha dW(s),$$

$$\hat{Y}_n(t) := \sum_{k=\lfloor k \rfloor + 1}^{\lfloor nt \rfloor} g \left( \frac{b_k}{n} \right) \sigma \left( t - \frac{k}{n} \right) \left( W\left( t - \frac{k}{n} + \frac{1}{n} \right) - W\left( t - \frac{k}{n} \right) \right).$$

In effect, we simply drop the summands in (1.2.8) and (1.2.9) that correspond to integrals and increments on the negative real line. We make remarks on the implementation of this scheme in Section 1.3, below.
The MSE of hybrid scheme for the TBS process \( Y \) has the following asymptotic behavior as \( n \to \infty \), which is, in fact, identical to the asymptotic behavior of the MSE of the hybrid scheme for BS processes. We omit the proof of this result, which would be a simple modification of the proof of Theorem 1.2.1.

**Theorem 1.2.2** (Asymptotics of mean square error). Suppose that (A1) and (A2) hold, and that for some \( \delta > 0 \),

\[
\mathbb{E}[|\sigma(s) - \sigma(0)|^2] = \Theta(s^{2\alpha+1+\delta}), \quad s \downarrow 0.
\]

Then for all \( t > 0 \),

\[
\mathbb{E}[|Y(t) - Y_n(t)|^2] \sim J(\alpha, \kappa, b)\mathbb{E}[\sigma(0)^2]n^{-(2\alpha+1)}Lg(1/n)^2, \quad n \to \infty,
\]

where \( J(\alpha, \kappa, b) \) is as in Theorem 1.2.1.

### 1.3 Implementation and numerical experiments

**Practical implementation**

Simulating the BS process \( X \) on the equidistant grid \( \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{|nT|}{n}\} \) for some \( T > 0 \) using the hybrid scheme entails generating

\[
X_n\left(\frac{i}{n}\right), \quad i = 0, 1, \ldots, |nT|.
\]

Provided that we can simulate the random variables

\[
W_{i,j}^n := \int_{\frac{i}{n}}^{\frac{i+j}{n}} \left(\frac{i+j}{n} - s\right)^\alpha dW(s), \quad i = -N_n, -N_n + 1, \ldots, |nT| - 1, \quad j = 1, \ldots, \kappa,
\]

\[
W_i^n := \int_{\frac{i}{n}}^{\frac{i+1}{n}} dW(s), \quad i = -N_n, -N_n + 1, \ldots, |nT| - 1,
\]

\[
\sigma_i^n := \sigma\left(\frac{i}{n}\right), \quad i = -N_n, -N_n + 1, \ldots, |nT| - 1,
\]

we can compute (1.3.1) via the formula

\[
X_n\left(\frac{i}{n}\right) = \sum_{k=1}^\kappa Lg\left(\frac{k}{n}\right)\sigma_i^n \left(\frac{k-1}{n}\right) + \sum_{k=\kappa+1}^N g\left(\frac{b_k^*}{n}\right)\sigma_i^n \left(\frac{k-1}{n}\right)W_i^n.
\]

In order to simulate (1.3.2) and (1.3.3), it is instrumental to note that the \( \kappa + 1 \)-dimensional random vectors

\[
W_i^n := \{W_i^n, W_i^n, \ldots, W_i^n\}, \quad i = -N_n, -N_n + 1, \ldots, |nT| - 1,
\]
are i.i.d. according to a multivariate Gaussian distribution with mean zero and covariance matrix $\Sigma$ given by

$$
\Sigma_{1,1} := \frac{1}{n}, \quad \Sigma_{1,j} = \Sigma_{j,1} := \frac{(j-1)\alpha+1 - (j-1)\alpha+1}{(\alpha+1)n^{\alpha+1}}, \quad \Sigma_{j,j} := \frac{(j-1)^{2\alpha+1} - (j-2)^{2\alpha+1}}{(2\alpha+1)n^{2\alpha+1}},
$$

$$
\Sigma_{j,k} := \frac{1}{(j-k)(\alpha+1)n^{2\alpha+1}} \left( (j-1)(k-1)^{\alpha+1} \right)_{2F1} \left( 1, 2(\alpha+1), \alpha+2, \frac{k-1}{k-j} \right) - (j-2)(k-2)^{\alpha+1} \right)_{2F1} \left( 1, 2(\alpha+1), \alpha+2, \frac{k-2}{k-j} \right),
$$

for $j, k = 2, \ldots, \kappa + 1$ such that $j \neq k$, where $2F1$ stands for the Gauss hypergeometric function.

Thus, $(W^n_i)_{i=-N_n}^{nT-1}$ can be generated by taking independent draws from the multivariate Gaussian distribution $N_{\kappa+1}(0, \Sigma)$. If the volatility process $\sigma$ is independent of $W$, then $(\sigma^n_i)_{i=-N_n}^{nT-1}$ can be generated separately, possibly using exact methods. (Exact methods are available, e.g., for Gaussian processes, as mentioned in the introduction, and diffusions, see Beskos and Roberts (2005).) In the case where $\sigma$ depends on $W$, simulating $(W^n_i)_{i=-N_n}^{nT-1}$ and $(\sigma^n_i)_{i=-N_n}^{nT-1}$ is less straightforward. That said, if $\sigma$ is driven by a standard Brownian motion $Z$, correlated with $W$, say, one could rely on a factor decomposition

$$
Z(t) := \rho W(t) + \sqrt{1-\rho^2} W_\perp(t), \quad t \in \mathbb{R},
$$

where $\rho \in [-1, 1]$ is the correlation parameter and $(W_\perp(t))_{t \in [0,T]}$ is a standard Brownian motion independent of $W$. Then one would first generate $(W^n_i)_{i=-N_n}^{nT-1}$, use (1.3.5) to generate $\{Z(\frac{i+1}{n}) - Z(\frac{i}{n})\}_{i=-N_n}^{nT-1}$ and employ some approximate appropriate method to produce $[\sigma^n_i]_{i=-N_n}^{nT-1}$ thereafter. This approach has, however, the caveat that it induces an additional approximation error, not quantified in Theorem 1.2.1.

**Remark 1.3.1.** In the case of the $\mathcal{TBF}$ process $Y$, introduced in Section 1.2, the observations $Y_n(\frac{i}{n})$, $i = 0, 1, \ldots, [nT]$, given by the hybrid scheme (1.2.16) can be computed via

$$
Y_n(\frac{i}{n}) = \min[i, \kappa] \sum_{k=1}^{\min[i, \kappa]} Lg \left( \frac{k}{n} \right) \sigma^n_{i-k} W^n_{i-k},
$$

using the random vectors $(W^n_i)_{i=0}^{nT-1}$ and random variables $[\sigma^n_i]_{i=0}^{nT-1}$.

In the hybrid scheme, it typically suffices to take $\kappa$ to be at most 3. Thus, in (1.3.4), the first sum $\hat{X}_n(\frac{i}{n})$ requires only a negligible computational effort. By contrast, the number of terms in the second sum $\tilde{X}_n(\frac{i}{n})$ increases as $n \to \infty$. It is then useful to note that

$$
\hat{X}_n(\frac{i}{n}) = \sum_{k=1}^{N_n} \Gamma_k \xi_{i-k} = (\Gamma \star \xi)_i,
$$
where

\[ \Gamma_k := \begin{cases} 0, & k = 1, \ldots, \kappa, \\ g\left(\frac{b^*}{n}\right), & k = \kappa + 1, \kappa + 2, \ldots, N_n, \end{cases} \]

\[ \Xi_k := \sigma^n_k W^n_k, \quad k = -N_n, -N_n + 1, \ldots, \lfloor nT \rfloor - 1. \]

and \( \Gamma \star \Xi \) stands for the discrete convolution of the sequences \( \Gamma \) and \( \Xi \). It is well-known that the discrete convolution can be evaluated efficiently using a fast Fourier transform (FFT). The computational complexity of simultaneously evaluating \( (\Gamma \star \Xi)_i \) for all \( i = 0, 1, \ldots, \lfloor nT \rfloor \) using an FFT is \( O(N_n \log N_n) \), see Mallat (2009, pp. 79–80), which under (A4) translates to \( O(n^{\gamma+1} \log n) \). The computational complexity of the entire hybrid scheme is then \( O(n^{\gamma+1} \log n) \), provided that \( \{\sigma^n_i\}_{i=-N_n}^{\lfloor nT \rfloor-1} \) is generated using a scheme with complexity not exceeding \( O(n^{\gamma+1} \log n) \). As a comparison, we mention that the complexity of an exact simulation of a stationary Gaussian process using circulant embeddings is \( O(n \log n) \) (Asmussen and Glynn, 2007, p. 316), whereas the complexity of the Cholesky factorization is \( O(n^3) \) (Asmussen and Glynn, 2007, p. 312).

**Remark 1.3.2.** With \( \mathcal{B} \mathcal{I} \mathcal{S} \) processes, the computational complexity of the hybrid scheme via (1.3.6) is \( O(n \log n) \).

Figure 2.1 presents examples of trajectories of the \( \mathcal{B} \mathcal{I} \mathcal{S} \) process \( X \) using the hybrid scheme of \( \kappa = 1, 2 \) and \( b = b^* \). We choose the kernel function \( g \) to be the gamma kernel (Example 3.3.2) with \( \lambda = 1 \). We also discretize \( X \) using the Riemann-sum scheme, \( \kappa = 0 \) with \( b \in \{b_{\text{FWD}}, b^*\} \) (that is, the forward Riemann-sum scheme and its counterpart with optimally chosen evaluation points). We can make two observations: Firstly, we see how the roughness parameter \( \alpha \) controls the regularity properties of the trajectories of \( X \) — as we decrease \( \alpha \), the trajectories of \( X \) become increasingly rough. Secondly, and more importantly, we see how the simulated trajectories coming from the Riemann-sum and hybrid schemes can be rather different, even though we use the same innovations for the driving Brownian motion. In fact, the two variants of the hybrid scheme (\( \kappa = 1, 2 \)) yield almost identical trajectories, while the Riemann-sum scheme (\( \kappa = 0 \)) produces trajectories that are comparatively smoother, this difference becoming more apparent as \( \alpha \) approaches \(-\frac{1}{2}\). Indeed, in the extreme case with \( \alpha = -0.499 \), both variants of the Riemann-sum scheme break down and yield anomalous trajectories with very little variation, while the hybrid scheme continues to produce accurate results. The fact that the hybrid scheme is able to reproduce the fine properties of rough \( \mathcal{B} \mathcal{I} \mathcal{S} \) processes, even for values of \( \alpha \) very close to \(-\frac{1}{2}\), is backed up by a further experiment reported in the following section.
Chapter 1. Hybrid scheme for Brownian semistationary processes

Figure 1.2. Discretized trajectories of a $\mathcal{BFS}$ process, where $g$ is the gamma kernel (Example 3.3.2), $\lambda = 1$ and $\sigma(t) = 1$ for all $t \in \mathbb{R}$. Trajectories consisting of $n = 50$ observations on $[0,1]$ were generated with the hybrid scheme ($\kappa = 1, 2$ and $b = b^*$) and Riemann-sum scheme ($\kappa = 0$ and $b = b^*$ (solid lines), $b = b_{\text{FWD}}$ (dashed lines)), using the same innovations for the driving Brownian motion in all cases and $N_n = \lfloor 50^{1.5} \rfloor = 353$. The simulated processes were normalized to have unit (stationary) variance.
Estimation of the roughness parameter

Suppose that we have observations \( X \left( \frac{i}{m} \right), i = 0, 1, \ldots, m \), of the \( \mathcal{B} \mathcal{S} \mathcal{S} \) process \( X \), given by (1.2.1), for some \( m \in \mathbb{N} \). Barndorff-Nielsen et al. (2013b) and Corcuera et al. (2013) discuss how the roughness index \( \alpha \) can be estimated consistently as \( m \to \infty \).

The method is based on the change-of-frequency (COF) statistics
\[
\text{COF}(X, m) = \frac{\sum_{k=5}^{m} \left[ X(\frac{k}{m}) - 2X(\frac{k-2}{m}) + X(\frac{k-4}{m}) \right]^2}{\sum_{k=3}^{m} \left[ X(\frac{k}{m}) - 2X(\frac{k-1}{m}) + X(\frac{k-2}{m}) \right]^2}, \quad m \geq 5,
\]
which compare the realized quadratic variations of \( X \), using second-order increments, with two different lag lengths. Corcuera et al. (2013) have shown that under some assumptions on the process \( X \), which are similar to (A1), (A2) and (A3), it holds that
\[
\hat{\alpha}(X, m) := \frac{\log \left( \text{COF}(X, m) \right)}{2 \log 2} - \frac{1}{2} \to \alpha, \quad m \to \infty.
\]

This result is derived in (1.3.7).

An in-depth study of the finite sample performance of this COF estimator can be found in Bennedsen, Lunde, and Pakkanen (2014).

To examine how well the hybrid scheme reproduces the fine properties of the \( \mathcal{B} \mathcal{S} \mathcal{S} \) process in terms of regularity/roughness, we apply the COF estimator to discretized trajectories of \( X \), where the kernel function \( g \) is again the gamma kernel (Example 3.3.2) with \( \lambda = 1 \), generated using the hybrid scheme with \( \kappa = 1, 2, 3 \) and \( b = b^* \). We consider the case where the volatility process satisfies \( \sigma(t) = 1 \), that is, the process \( X \) is Gaussian. This allows us to quantify and control for the intrinsic bias and noisiness, measured in terms of standard deviation, of the estimation method itself, by initially applying the estimator to trajectories that have been simulated using an exact method based on the Cholesky factorization. We then study the behavior of the estimator when applied to a discretized trajectory, while decreasing the step size of the discretization scheme. More precisely, we simulate \( \hat{\alpha}(X_n, m) \), where \( m = 500 \) and \( X_n \) is the hybrid scheme for \( X \) with \( n = ms \) and \( s \in \{1, 2, 5\} \). This means that we compute \( \hat{\alpha}(X_n, m) \) using \( m \) observations obtained by subsampling every \( s \)-th observation in the sequence \( X_n \left( \frac{i}{n} \right), i = 0, 1, \ldots, n \). As a comparison, we repeat these simulations substituting the hybrid scheme with the Riemann-sum scheme, using \( \kappa = 0 \) with \( b \in \{b_{\text{FWD}}, b^*\} \).

The results are presented in Figure 1.3. We observe that the intrinsic bias of the estimator with \( m = 500 \) observations is negligible and hence the bias of the estimates computed from discretized trajectories is then attributable to an approximation error arising from the respective discretization scheme, where positive (resp. negative) bias indicates that the simulated trajectories are smoother (resp. rougher) than those of the process \( X \). Concentrating first on the baseline case \( s = 1 \), we note that the hybrid scheme produces essentially unbiased results when \( \alpha \in \left(-\frac{1}{2}, 0\right] \), while there is a moderate bias when \( \alpha \in \left(0, \frac{1}{2}\right] \), which disappears when passing from \( \kappa = 1 \) to \( \kappa = 3 \), even for values of \( \alpha \) very close to \( \frac{1}{2} \). (The largest value of \( \alpha \) considered in our simulations is
Figure 1.3. Bias and standard deviation of the COF estimator (1.3.7) of the roughness index $\alpha$, when applied to discretized trajectories of a $BF$ process with the gamma kernel (Example 3.3.2), $\lambda = 1$ and $\sigma(t) = 1$ for all $t \in \mathbb{R}$. Trajectories were generated using an exact method based on the Cholesky factorization, the hybrid scheme ($\kappa = 1, 2, 3$ and $b = b^*$) and Riemann-sum scheme ($\kappa = 0$ and $b = b^*$ (solid lines), $b = b_{FWD}$ (dashed lines)). In the experiment, $n = ms$ observations were generated, where $m = 500$ and $s \in \{1, 2, 5\}$, on $[0, 1]$ using $N_n = \lfloor n^{1.5} \rfloor$. Every $s$-th observation was then subsampled, resulting in $m = 500$ observations that were used to compute the estimate $\hat{\alpha}(X_n, m)$ of the roughness index $\alpha$. Number of Monte Carlo replications: 10000.
α = 0.49; one would expect the performance to weaken as α approaches $\frac{1}{2}$, cf. Figure 1.1, but this range of parameter values seems to be of limited practical interest.) The standard deviations exhibit a similar pattern. The corresponding results for the Riemann-sum scheme are clearly inferior, exhibiting a significant bias, while using optimal evaluation points ($b = b^*$) improves the situation slightly. In particular, the bias in the case $α \in (-\frac{1}{2}, 0)$ is positive, indicating too smooth discretized trajectories, which is connected with the failure of the Riemann-sum scheme with $α$ near $-\frac{1}{2}$, illustrated in Figure 2.1. With $s = 2$ and $s = 5$, the results improve with both schemes. Notably, in the case $s = 5$, the performance of the hybrid scheme even with $κ = 1$ is on a par with the exact method. However, the improvements with the Riemann-sum scheme are more meager, as a considerable bias persists when $α$ is near $-\frac{1}{2}$.

**Option pricing under rough volatility**

As another experiment, we study Monte Carlo option pricing in the rough Bergomi (rBergomi) model of Bayer et al. (2015). In the rBergomi model, the logarithmic spot variance of the price of the underlying is modelled by a rough Gaussian process, which is a special case of (1.2.15). By virtue of the rough volatility process, the model fits well to observed implied volatility smiles (Bayer et al., 2015, pp. 15–19).

More precisely, the price of the underlying in the rBergomi model with time horizon $T > 0$ is defined, under an equivalent martingale measure identified with $P$, as

$$S(t) := S(0) \exp \left( \int_0^t \sqrt{v(s)} dZ(s) - \frac{1}{2} \int_0^t v(s) ds \right), \quad t \in [0, T],$$

using the spot variance process

$$v(t) := \xi_0(t) \exp \left( \eta \sqrt{2α + 1} \int_0^t (t - s)^α dW(s) - \frac{η^2}{2} t^{2α+1} \right), \quad t \in [0, T].$$

Above, $S(0) > 0$, $η > 0$ and $α \in (-\frac{1}{2}, 0)$ are deterministic parameters, and $Z$ is a standard Brownian motion given by

$$Z(t) := \rho W(t) + \sqrt{1 - \rho^2} W_\perp(t), \quad t \in [0, T], \quad (1.3.8)$$

where $ρ \in (-1, 1)$ is the correlation parameter and $\{W_\perp(t)\}_{t \in [0, T]}$ is a standard Brownian motion independent of $W$. The process $\{ξ_0(t)\}_{t \in [0, T]}$ is the so-called forward variance curve (Bayer et al., 2015, p. 11), which we assume here to be flat, $ξ_0(t) = ξ > 0$ for all $t \in [0, T]$.

We aim to compute using Monte Carlo simulation the price of a European call option struck at $K > 0$ with maturity $T$, which is given by

$$C(S(0), K, T) := E[(S_T - K)^+]. \quad (1.3.9)$$
<table>
<thead>
<tr>
<th>$S(0)$</th>
<th>$\xi$</th>
<th>$\eta$</th>
<th>$\alpha$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.235</td>
<td>1.9</td>
<td>−0.43</td>
<td>−0.9</td>
</tr>
</tbody>
</table>

Table 1.1. Parameter values used in the rBergomi model.

The approach suggested by Bayer et al. (2015) involves sampling the Gaussian processes $Z$ and $Y$ on a discrete time grid using exact simulation and then approximating $S$ and $\nu$ using Euler discretization. We modify this approach by using the hybrid scheme to simulate $Y$, instead of the computationally more costly exact simulation. As the hybrid scheme involves simulating increments of the Brownian motion $W$ driving $Y$, we can conveniently simulate the increments of $Z$, needed for the Euler discretization of $S$, using the representation (1.3.8).

We map the option price $C(S(0), K, T)$ to the corresponding Black–Scholes implied volatility $\text{IV}(S(0), K, T)$, see, e.g., Gatheral (2006). Reparameterizing the implied volatility using the log-strike $k := \log(K/S_0)$ allows us to drop the dependence on the initial price, so we will abuse notation slightly and write $\text{IV}(k, T)$ for the corresponding implied volatility. Figure 1.4 displays implied volatility smiles obtained from the rBergomi model using the hybrid and Riemann-sum schemes to simulate $Y$, as discussed above, and compares these to the smiles obtained using an exact simulation of $Y$ via Cholesky factorization. The parameter values are given in Table 1.1. They have been adopted from Bayer et al. (2015), who demonstrate that they result in realistic volatility smiles. We consider two different maturities: “short”, $T = 0.041$, and “long”, $T = 1$.

We observe that the Riemann-sum scheme ($\kappa = 0, b \in \{b_{\text{FWD}}, b^*\}$) is able capture
the shape of the implied volatility smile, but not its level. Alas, the method even breaks down with more extreme log-strikes (the prices are so low that the root-finding algorithm used to compute the implied volatility would return zero). In contrast, the hybrid scheme with \(\kappa = 1, 2\) and \(b = b^*\) yields implied volatility smiles that are indistinguishable from the benchmark smiles obtained using exact simulation. Further, there is no discernible difference between the smiles obtained using \(\kappa = 1\) and \(\kappa = 2\).

As in the previous section, we observe that the hybrid scheme is indeed capable of producing very accurate trajectories of \(\mathcal{F}\mathcal{B}\mathcal{I}\mathcal{F}\) processes, in particular in the case \(\alpha \in (-\frac{1}{2}, 0)\), even when \(\kappa = 1.4\) Proofs

Throughout the proofs below, we rely on two useful inequalities. The first one is the Potter bound for slow variation at 0, which follows immediately from the corresponding result for slow variation at \(\infty\) (Bingham et al., 1989, Theorem 1.5.6). Namely, if \(L : (0, 1] \to (0, \infty)\) is slowly varying at 0 and bounded away from 0 and \(\infty\) on any interval \((u, 1]\), \(u \in (0, 1)\), then for any \(\delta > 0\) there exists a constant \(C_\delta > 0\) such that

\[
\frac{L(x)}{L(y)} \leq C_\delta \max \left\{ \left( \frac{x}{y} \right)^\delta, \left( \frac{x}{y} \right)^{-\delta} \right\}, \quad x, y \in (0, 1]. \tag{1.4.1}
\]

The second one is the elementary inequality

\[
|x^\alpha - y^\alpha| \leq |\alpha| (\min\{x, y\})^{\alpha-1}|x - y|, \quad x, y \in (0, \infty), \quad \alpha \in (-\infty, 1), \tag{1.4.2}
\]

which can be easily shown using the mean value theorem. Additionally, we use the following variant of Karamata’s theorem for regular variation at 0. Its proof is similar to the one of the usual Karamata’s theorem for regular variation at \(\infty\) (Bingham et al., 1989, Proposition 1.5.10).

**Lemma 1.4.1** (Karamata’s theorem). If \(\alpha \in (-1, \infty)\) and \(L : (0, 1] \to [0, \infty)\) is slowly varying at 0, then

\[
\int_0^y x^\alpha L(x) \, dx \sim \frac{1}{\alpha + 1} y^{\alpha + 1} L(y), \quad y \to 0.
\]

**Proof of Proposition 1.2.1**

Proof of Proposition 1.2.1. (i) By the covariance stationarity of the volatility process \(\sigma\), we may express the variogram \(V(h)\) for any \(h \geq 0\) as

\[
V(h) = \mathbb{E}[|X(h) - X(0)|^2] = \int_{-\infty}^h (g(h - u) - g(-u)1_{[\infty, 0]}(u))^2 \mathbb{E}[\sigma(u)^2] \, du \\
= \mathbb{E}[\sigma(0)^2] \left( \int_0^h g(x)^2 \, dx + \int_0^\infty (g(x + h) - g(x))^2 \, dx \right). \tag{1.4.3}
\]
Invoking (A1) and Lemma 1.4.1, we find that
\[ \int_0^h g(x)^2 \, dx \sim \frac{1}{2\alpha + 1} h^{2\alpha + 1} L_g(h)^2, \quad h \to 0. \] (1.4.4)

We may clearly assume that \( h < 1 \), which allows us to work with the decomposition
\[ \int_0^\infty (g(x + h) - g(x))^2 \, dx = A_h + A'_h, \]
where
\[ A_h := \int_0^{1-h} (g(x + h) - g(x))^2 \, dx, \quad A'_h := \int_{1-h}^\infty (g(x + h) - g(x))^2 \, dx. \]

According to (A2), there exists \( M > 1 \) such that \( x \mapsto |g'(x)| \) is non-increasing on \([M, \infty)\). Thus, using the mean value theorem, we deduce that
\[ |g(x + h) - g(x)| = |g'(\xi)| h \leq \begin{cases} \sup_{y \in (1-h,M]} |g'(y)| h, & x \in (1-h,M), \\ |g'(x)| h, & x \in [M, \infty). \end{cases} \]

where \( \xi = \xi(x, h) \in [x, x + h] \). It follows then that
\[ \limsup_{h \to 0} \frac{A'_h}{h^2} \leq (M - 1) \sup_{y \in [1,M]} g'(y)^2 + \int_1^\infty g'(x)^2 \, dx < \infty, \]
which in turn implies that
\[ A'_h = \mathcal{O}(h^2), \quad h \to 0. \] (1.4.5)

Making a substitution \( y = \frac{x}{h} \), we obtain
\[ A_h = \int_0^{1-h} (g(x + h) - g(x))^2 \, dx = h \int_0^{1/h-1} (g(h(y + 1)) - g(hy))^2 \, dy \]
\[ = h^{2\alpha + 1} L_g(h)^2 \int_0^\infty G_h(y) \, dy, \]
where
\[ G_h(y) := \left( (y + 1)^a \frac{L_g(h(y + 1))}{L_g(h)} - y^a \frac{L_g(hy)}{L_g(h)} \right)^2 1_{[0,1/h-1]}(y), \quad y \in (0, \infty). \]

By the definition of slow variation at 0,
\[ \lim_{h \to 0} G_h(y) = \left( (y + 1)^a - y^a \right)^2, \quad y \in (0, \infty). \]

We shall show below that the functions \( G_h, h \in (0, 1) \), have an integrable dominant. Thus, by the dominated convergence theorem,
\[ A_h \sim h^{2\alpha + 1} L_g(h)^2 \int_0^\infty \left( (y + 1)^a - y^a \right)^2 \, dy, \quad h \to 0. \] (1.4.6)
Since \( \alpha < \frac{1}{2} \), we have \( \lim_{h \to 0} \frac{\Delta_h'}{h^{2\alpha + 1} L_g(h)^2} = 0 \) by (1.2.2) and (1.4.5), so we get from (1.4.4) and (1.4.6)

\[
\int_0^h g(x)^2 \, dx + \int_0^\infty (g(x + h) - g(x))^2 \, dx \leq \frac{1}{2\alpha + 1} + \int_0^\infty ((y + 1)^\alpha - y^\alpha)^2 \, dy \bigg( h^{2\alpha + 1} L_g(h)^2 \bigg), \quad h \to 0,
\]

which, together with (1.4.3), implies the assertion.

It remains to justify the use of the dominated convergence theorem to deduce (1.4.6). For any \( y \in (0, 1] \), we have by the Potter bound (1.4.1) and the elementary inequality \( (u + v)^2 \leq 2u^2 + 2v^2 \),

\[
G_h(y) \leq 2(y + 1)^2 \left( \frac{L_g(h(y + 1))}{L_g(h)} \right)^2 + 2y^2 \left( \frac{L_g(hy)}{L_g(h)} \right)^2 
\leq 2C_\delta^2 \left( (y + 1)^{2(\alpha + \delta_1)} + y^{2(\alpha - \delta_1)} \right),
\]

where we choose \( \delta_1 \in (0, \alpha + \frac{1}{2}) \) to ensure that \( 2(\alpha - \delta_1) > -1 \). Consider then \( y \in [1, \infty) \).

By adding and substracting the term \( (y + 1)^\alpha \frac{L_g(hy)}{L_g(h)} \) and using again the inequality \( (u + v)^2 \leq 2u^2 + 2v^2 \), we get

\[
G_h(y) = \left( (y + 1)^\alpha \frac{L_g(h(y + 1))}{L_g(h)} - (y + 1)^\alpha \frac{L_g(hy)}{L_g(h)} + (y + 1)^\alpha \frac{L_g(hy)}{L_g(h)} - y^\alpha \frac{L_g(hy)}{L_g(h)} \right)^2 1_{[\frac{1}{2}, 1)}(y) 
\leq 2(y + 1)^2 \left( \frac{L_g(h(y + 1)) - L_g(hy)}{L_g(h)} \right)^2 
\leq 2\left( (y + 1)^\alpha - y^\alpha \right)^2 \left( \frac{L_g(hy)}{L_g(h)} \right)^2 1_{[1, \infty)}(y).
\]

We recall that \( L_g := \inf_{x \in [0, 1]} L_g(x) > 0 \) by (A1), so

\[
\left| \frac{L_g(h(y + 1)) - L_g(hy)}{L_g(h)} \right| 1_{[0, 1/h-1)}(y) \leq \frac{1}{L_g} |L_g(h(y + 1)) - L_g(hy)| 1_{[0, 1/h-1)}(y).
\]

Using the mean value theorem and the bound for the derivative of \( L_g \) from (A1), we observe that

\[
|L_g(h(y + 1)) - L_g(hy)| = |L_g'(\xi)||h(y + 1) - hy| \leq hC \left( 1 + \frac{1}{\xi} \right) \leq C \left( h + \frac{1}{y} \right),
\]

where \( \xi \) is the derivative of \( L_g \) at some point between \( y + 1 \) and \( hy \).
where \(\xi = \xi(y, h) \in [h y, h(y + 1)]\). Noting that the constraint \(y < \frac{1}{h} - 1\) is equivalent to \(h < \frac{1}{y+1}\), we obtain further

\[
\left| \frac{L_g(h(y + 1)) - L_g(hy)}{L_g(h)} \mathbf{1}_{(0,1/h-1)}(y) \right| \leq \frac{C}{L_g} \left( \frac{y + 1}{h} \right) \mathbf{1}_{(0,1/h-1)}(y) \\
\leq \frac{C}{L_g} \left( \frac{1}{y+1} + \frac{1}{y} \right) \\
\leq \frac{C}{L_g} \left( \frac{3}{y+1} \right),
\]

as \(y \geq 1\), which we then use to deduce that

\[
2(y+1)^{2}\alpha \left( \frac{L_g(h(y + 1)) - L_g(hy)}{L_g(h)} \mathbf{1}_{(0,1/h-1)}(y) \right)^{2} \leq \frac{18C^2}{L_g} (y + 1)^{2(\alpha-1)}.
\]

Additionally, we observe that, by (1.4.1) and (1.4.2),

\[
2(1 + y)^{\alpha - y} \left( \frac{L_g(hy)}{L_g(h)} \mathbf{1}_{(0,1/h-1)}(y) \right)^{2} \leq 2C^2_{\delta_2} a^2 y^{2(\alpha-1+\delta_2)},
\]

where we choose \(\delta_2 \in (0, \frac{1}{2} - \alpha)\), ensuring that \(2(\alpha - 1 + \delta_2) < -1\). We may finally define a function

\[
G(y) := \begin{cases} 
2C^2_{\delta_1} ((y + 1)^{2(\alpha+\delta_1)} + y^{2(\alpha-\delta_1)}), & y \in (0, 1), \\
\frac{18C^2}{L_g} (y + 1)^{2(\alpha-1)} + 2C^2_{\delta_2} a^2 y^{2(\alpha-1+\delta_2)}, & y \in (1, \infty),
\end{cases}
\]

which satisfies \(0 \leq G_1(y) \leq G(y)\) for any \(y \in (0, \infty)\) and \(h \in (0, 1)\), and is integrable on \((0, \infty)\) with the aforementioned choices of \(\delta_1\) and \(\delta_2\).

(ii) To show existence of the modification, we need a localization procedure that involves an ancillary process

\[
F(t) := \int_1^{\infty} g'(s)^2 \sigma(t - s)^2 \, ds, \quad t \in \mathbb{R}.
\]

We check first that \(F\) is locally bounded under (A1) and (A2), which is essential for localization. To this end, let \(T \in (0, \infty)\), and write for any \(t \in [-T, T]\),

\[
F(t) = F^\circ(t) + F^\bar{\circ}(t),
\]

where

\[
F^\circ(t) := \int_1^{M+2T} g'(s)^2 \sigma(t - s)^2 \, ds, \quad F^\bar{\circ}(t) := \int_{M+2T}^{\infty} g'(s)^2 \sigma(t - s)^2 \, ds,
\]

and \(M > 1\) is such that \(x \mapsto |g'(x)|\) is non-increasing on \([M, \infty)\), as in the proof of (i).
Since \( g' \) is continuous on \((0, \infty)\) and \( \sigma \) locally bounded, we have for any \( t \in [-T, T] \),

\[
0 \leq F^g(t) = (M + 2T - 1) \sup_{y \in [1, M+2T]} g'(y)^2 \sup_{u \in [-M+2T, T-1]} \sigma(u)^2 < \infty.
\]

Further, when \( t \in [-T, T] \),

\[
F^g(t) = \int_{-\infty}^{t-(M+2T)} g'(t-u)^2 \sigma(u)^2 \, du,
\]

where \( g'(t-u)^2 \leq g'(-T-u)^2 \) since the arguments satisfy

\[
t - u \geq -T - u \geq -T - (t - (M+2T)) \geq M.
\]

Thus,

\[
0 \leq F^g(t) \leq \int_{-\infty}^{-(M+2T)} g'(-T-u)^2 \sigma(u)^2 \, du \leq \int_{1}^{\infty} g'(s)^2 \sigma(-T-s)^2 \, ds < \infty
\]

for any \( t \in [-T, T] \) almost surely, as we have

\[
\mathbb{E} \left[ \int_{1}^{\infty} g'(s)^2 \sigma(-T+s)^2 \, ds \right] = \int_{1}^{\infty} g'(s)^2 \mathbb{E}[\sigma(-T-s)^2] \, ds = \mathbb{E}[\sigma(0)^2] \int_{1}^{\infty} g'(s)^2 \, ds < \infty,
\]

where we change the order of expectation and integration relying on Tonelli’s theorem and where the final equality follows from the covariance stationarity of \( \sigma \). So we can conclude that \( F \) is indeed locally bounded.

Let now \( m \in \mathbb{N} \) and, for localization, define a sequence of stopping times

\[
\tau_{m,n} := \inf \{ t \in [-m, \infty) : F_t \leq n, |\sigma_t| \leq n \}, \quad n \in \mathbb{N},
\]

that satisfies \( \tau_{m,n} \uparrow \infty \) almost surely as \( n \to \infty \) since both \( F \) and \( \sigma \) are locally bounded. (We follow the usual convention that \( \inf \emptyset = \infty \).) Consider now the modified \( \mathcal{B} \mathcal{F} \mathcal{F} \) process

\[
X_{m,n}^\dagger(t) := \int_{-\infty}^{t} g(t-s)\sigma(\min(s, \tau_{m,n})) \, dW(s), \quad t \in [-m, \infty),
\]

that coincides with \( X \) on the stochastic interval \([-m, \tau_{m,n}]\). The process \( X_{m,n}^\dagger \) satisfies the assumptions of Barndorff-Nielsen, Corcuera, and Podolskij (2011, Lemma 1), so for any \( p > 0 \) there exists a constant \( \tilde{C}_p > 0 \) such that

\[
\mathbb{E}[|X_{m,n}^\dagger(s) - X_{m,n}^\dagger(t)|^p] \leq \tilde{C}_p V(|s-t|)^{p/2}, \quad s, t \in [-m, \infty).
\]

Using the upper bound in (1.2.2), we can deduce from (i) that for any \( \delta > 0 \) there are constants \( \tilde{C}_\delta > 0 \) and \( \tilde{h}_\delta > 0 \) such that

\[
V(h) \leq \tilde{C}_\delta h^{2\alpha+1-\delta}, \quad h \in (0, \tilde{h}_\delta).
\]
Applying (1.4.8) to (1.4.7), we get
\[
\mathbb{E}[|X_{m,n}^t(s) - X_{m,n}^t(t)|^p] \leq C_p C_{\delta}^{p/2} |s - t|^{1 + p(\alpha + \frac{1}{2} - \frac{\delta}{2} - \frac{1}{p})}, \quad s, t \in [-m, \infty), \quad |s - t| < h_\delta.
\]
We may note that \( p(\alpha + \frac{1}{2} - \frac{\delta}{2} - \frac{1}{p}) > 0 \) for small enough \( \delta \) and large enough \( p \) and, in particular,
\[
\frac{p(\alpha + \frac{1}{2} - \frac{\delta}{2} - \frac{1}{p})}{p} \uparrow \alpha + \frac{1}{2},
\]
as \( \delta \downarrow 0 \) and \( p \uparrow \infty \). Thus it follows from the Kolmogorov–Chentsov theorem (Kallenberg, 2002, Theorem 3.22) that \( X_{m,n}^t \) has a modification with \( \phi \)-Hölder continuous trajectories for any \( \phi \in (0, \alpha + \frac{1}{2}) \). A modification of \( X \) on \( \mathbb{R} \), having \( \phi \)-Hölder continuous trajectories for any \( \phi \in (0, \alpha + \frac{1}{2}) \), can then by constructed from these modifications of \( X_{m,n}^t, m \in \mathbb{N}, n \in \mathbb{N} \), by letting first \( n \to \infty \) and then \( m \to \infty \).

**Proof of Theorem 1.2.1**

As a preparation, we shall first establish an auxiliary result that deals with the asymptotic behavior of certain integrals of regularly varying functions.

**Lemma 1.4.2.** Suppose that \( L : (0, 1) \to [0, \infty) \) is bounded away from 0 and \( \infty \) on any set of the form \((u, 1], u \in (0, 1)\), and slowly varying at 0. Moreover, let \( \alpha \in (-\frac{1}{2}, \infty) \) and \( k \geq 1 \). If \( b \in [k - 1, k] \setminus \{0\} \), then

\[
(i) \quad \lim_{n \to \infty} \int_{k-1}^k \left( x^{2\alpha} \frac{L(x/n)}{L(1/n)} - b^{2\alpha} \frac{L(b/n)}{L(1/n)} \right)^2 dx = \int_{k-1}^k (x^{2\alpha} - b^{2\alpha})^2 dx < \infty,
\]

\[
(ii) \quad \lim_{n \to \infty} \int_{k-1}^k x^{2\alpha} \left( \frac{L(x/n)}{L(1/n)} - \frac{L(b/n)}{L(1/n)} \right)^2 dx = 0.
\]

**Proof.** We only prove (i) as (ii) can be shown similarly. By the definition of slow variation at 0, the function
\[
f_n(x) := \left( x^{2\alpha} \frac{L(x/n)}{L(1/n)} - b^{2\alpha} \frac{L(b/n)}{L(1/n)} \right)^2, \quad x \in [k-1, k] \setminus \{0\},
\]
satisfies \( \lim_{n \to \infty} f_n(x) = (x^{2\alpha} - b^{2\alpha})^2 \) for any \( x \in [k-1, k] \setminus \{0\} \). In view of the dominated convergence theorem, it suffices to find an integrable dominant for the functions \( f_n \), \( n \in \mathbb{N} \). The construction of the dominant is quite similar to the one seen in the proof of Proposition 1.2.1, but we provide the details for the convenience of the reader.

Using the Potter bound (1.4.1) and the inequality \((u + v)^2 \leq 2u^2 + 2v^2\), we find that for any \( x \in [k-1, k] \setminus \{0\} \),
\[
0 \leq f_n(x) \leq 2x^{2\alpha} \left( \frac{L(x/n)}{L(1/n)} \right)^2 + 2b^{2\alpha} \left( \frac{L(b/n)}{L(1/n)} \right)^2 
\leq 2C_{\delta}^2 \left( x^{2\alpha} \max\{x^\delta, x^{-\delta}\}^2 + b^{2\alpha} \max\{b^\delta, b^{-\delta}\}^2 \right) =: f(x),
\]
where we choose $\delta \in \left(0, \alpha + \frac{1}{2}\right)$. When $k \geq 2$, we have $x \geq 1$ and $b \geq 1$, so

$$f(x) = 2C_2^2 \left(x^{2(\alpha + \delta)} + b^{2(\alpha + \delta)}\right)$$

is a bounded function of $x$ on $[k-1, k]$. When $k = 1$, we have $x \leq 1$ and $b \leq 1$, implying that

$$f(x) = 2C_2^2 \left(x^{2(\alpha - \delta)} + b^{2(\alpha - \delta)}\right),$$

where $2(\alpha - \delta) > -1$ with our choice of $\delta$, so $f$ is an integrable function on $(0, 1]$. □

**Proof of Theorem 1.2.1.** Let $t \in \mathbb{R}$ be fixed. It will be convenient to write $X_n(t)$ as

$$X_n(t) = \sum_{k=1}^{K} \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} \left(\frac{k}{n}\right)^{\alpha} \left(\frac{t - s}{n}\right)^{\alpha} dW(s) + \sum_{k=x+1}^{N_n} \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} g\left(\frac{b_k}{n}\right) \left(\frac{t - s}{n}\right)^{\alpha} dW(s).$$

Moreover, we introduce an ancillary approximation of $X(t)$, namely,

$$X_n'(t) = \sum_{k=1}^{N_n} \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} g(t) \left(\frac{t - s}{n}\right)^{\alpha} dW(s) + \int_{-\infty}^{t - \frac{N_n}{n}} g(t) \left(\frac{s}{n}\right)^{\alpha} dW(s).$$

By Minkowski’s inequality, we have

$$\mathbb{E}\left[|X_n(t) - X(t)|^2\right]^{\frac{1}{2}} \geq \mathbb{E}\left[|X_n'(t) - X(t)|^2\right]^{\frac{1}{2}} - \mathbb{E}\left[|X_n(t) - X_n'(t)|^2\right]^{\frac{1}{2}},$$

$$\mathbb{E}\left[|X_n(t) - X(t)|^2\right]^{\frac{1}{2}} \leq \mathbb{E}\left[|X_n'(t) - X(t)|^2\right]^{\frac{1}{2}} + \mathbb{E}\left[|X_n(t) - X_n'(t)|^2\right]^{\frac{1}{2}},$$

which together, after taking squares, imply that

$$E_n \left(1 - 2 \sqrt{\frac{E_n'}{E_n}} \right) \leq \mathbb{E}\left[|X_n(t) - X(t)|^2\right] \leq E_n \left(1 + 2 \sqrt{\frac{E_n'}{E_n}} \right),$$

(1.4.9)

where

$$E_n := \mathbb{E}\left[|X_n(t) - X_n'(t)|^2\right], \quad E_n' := \mathbb{E}\left[|X(t) - X_n'(t)|^2\right].$$

Using the Itô isometry, and recalling that $\sigma$ is covariance stationary, we obtain

$$E_n' = \sum_{k=1}^{N_n} \int_{t - \frac{k-1}{n}}^{t - \frac{k}{n}} g(t) \left(\frac{t - s}{n}\right)^{\alpha} \mathbb{E}\left[\left(\frac{t - s}{n} - \sigma(s)\right)^2\right] ds$$

$$\leq \sup_{u \in \left(0, \frac{1}{n}\right]} \mathbb{E}\left[\left(\sigma(u) - \sigma(0)\right)^2\right] \int_{0}^{\infty} g(s)^2 ds.$$
and

\[
E_n = \sum_{k=1}^{\xi} \int_{t-k/n}^{t-k+1/n} \left( (t-s)^{a} L_{g} \left( \frac{k}{n} \right) - g(t-s) \right)^{2} \sigma \left( t - \frac{k}{n} \right) ds + \sum_{k=K+1}^{n} \int_{t-k/n}^{t-k+1/n} \left( g \left( \frac{b_k}{n} \right) - g(t-s) \right)^{2} \sigma \left( t - \frac{k}{n} \right) ds + \sum_{k=n+1}^{n+1} \int_{t-k/n}^{t-k+1/n} \left( g \left( \frac{b_k}{n} \right) - g(t-s) \right)^{2} \sigma \left( t - \frac{k}{n} \right) ds + \int_{-\infty}^{t} \int_{t-k/n}^{t-k+1/n} g(t-s)^{2} \sigma \left( t - \frac{k}{n} \right) ds \]

= \mathbb{E}[\sigma(0)^{2}](D_{n} + D'_{n} + D''_{n} + D'''_{n}),

where

\[
D_{n} := \sum_{k=1}^{\xi} \int_{-\infty}^{t-k/n} \left( s^{a} L_{g} \left( \frac{k}{n} \right) - g(s) \right)^{2} ds, \quad D'_{n} := \sum_{k=K+1}^{n} \int_{t-k/n}^{t-k+1/n} \left( g \left( \frac{b_k}{n} \right) - g(s) \right)^{2} ds, \quad D''_{n} := \sum_{k=n+1}^{n+1} \int_{t-k/n}^{t-k+1/n} \left( g \left( \frac{b_k}{n} \right) - g(s) \right)^{2} ds, \quad D'''_{n} := \int_{-\infty}^{t} \int_{t-k/n}^{t-k+1/n} g^{2}(s)^{2} ds.
\]

(We may assume without loss of generality that \( N_{n} > n > \kappa \), as this will be the case for large enough \( n \).) In what follows, we study the asymptotic behavior of the terms \( D_{n}, D'_{n}, D''_{n}, \) and \( D'''_{n} \) separately, showing that \( D_{n} \) and \( D''_{n} \) are negligible as compared to \( D'_{n} \), and that \( D'_{n} \) and \( D'''_{n} \) give rise to the convergence rate given in Theorem 1.2.1.

Let us analyze the terms \( D''_{n}, D'''_{n} \) and \( D_{n} \) first. By (A3) and (A4), we have

\[
D''_{n} = \Theta \left( \left( \frac{N_{n}}{n} \right)^{2\beta+1} \right) = \Theta (n^{y(2\beta+1)}), \quad n \to \infty. \tag{1.4.10}
\]

Regarding the term \( D'''_{n} \), recall that by (A2) there is \( M > 1 \) such that \( x \to |g'(x)| \) is non-increasing on \([M, \infty)\). So, we have by the mean value theorem,

\[
\left| g \left( \frac{b_k}{n} \right) - g(s) \right| = |g'(\xi)| \left| \frac{b_k}{n} - s \right| \leq \begin{cases} \frac{1}{n} \sup_{y \in [1,M]} |g'(y)|, & \frac{k-1}{n} < M, \\ \frac{1}{n} |g'(\xi)|, & \frac{k-1}{n} \geq M, \end{cases}
\]

where \( \xi = \xi \left( \frac{b_k}{n}, s \right) \in \left[ \frac{k-1}{n}, \frac{k}{n} \right] \). Thus,

\[
\limsup_{n \to \infty} n^{2} D'''_{n} \leq (M - 1) \sup_{y \in [1,M]} |g'(y)|^{2} + \int_{1}^{M} g'(s)^{2} ds < \infty,
\]

which implies

\[
D'''_{n} = \Theta (n^{-2}), \quad n \to \infty. \tag{1.4.11}
\]
To analyze the behavior of $D_n$, we substitute $y = ns$ and invoke (A1), yielding

$$D_n = \sum_{k=1}^{\kappa} \int_{k-1}^{k} \left( \left( \frac{y}{n} \right)^\alpha L_g \left( \frac{k}{n} \right) - g \left( \frac{y}{n} \right) \right)^2 \frac{dy}{n}$$

$$= n^{-(2\alpha+1)} L_g(1/n)^2 \sum_{k=1}^{\kappa} \int_{k-1}^{k} y^{2\alpha} \left( \frac{L_g(k/n)}{L_g(1/n)} - \frac{L_g(y/n)}{L_g(1/n)} \right)^2 dy,$$

where, by Lemma 1.4.2(ii), we have

$$\lim_{n \to \infty} \int_{k-1}^{k} y^{2\alpha} \left( \frac{L_g(k/n)}{L_g(1/n)} - \frac{L_g(y/n)}{L_g(1/n)} \right)^2 dy = 0$$

for any $k = 1, \ldots, \kappa$. Thus, we find that

$$\lim_{n \to \infty} \frac{D_n}{n^{-(2\alpha+1)} L_g(1/n)^2} = 0. \quad (1.4.12)$$

The asymptotic behavior of the term $D_n'$ is more delicate to analyze. By (A1), and substituting $y = ns$, we can write

$$D_n' = \sum_{k=\kappa+1}^{n} \int_{k-1}^{k} \left( g \left( \frac{b_k}{n} \right) - g \left( \frac{y}{n} \right) \right)^2 \frac{dy}{n}$$

$$= n^{-(2\alpha+1)} \sum_{k=\kappa+1}^{n} \int_{k-1}^{k} \left( b_k^\alpha L_g \left( \frac{b_k}{n} \right) - y^\alpha L_g \left( \frac{y}{n} \right) \right)^2 dy$$

$$= n^{-(2\alpha+1)} L_g(1/n)^2 \sum_{k=\kappa+1}^{n} A_{n,k},$$

where

$$A_{n,k} := \int_{k-1}^{k} \left( y^\alpha \frac{L_g(y/n)}{L_g(1/n)} - b_k^\alpha \frac{L_g(b_k/n)}{L_g(1/n)} \right)^2 dy.$$

Let us study the asymptotic behavior of the sum $\sum_{k=\kappa+1}^{n} A_{n,k}$ as $n \to \infty$. By Lemma 1.4.2, we have for any $k \in \mathbb{N}$,

$$\lim_{n \to \infty} A_{n,k} = \int_{k-1}^{k} (y^\alpha - b_k^\alpha)^2 dy < \infty.$$ 

To be able to then deduce, using the dominated convergence theorem, that

$$\lim_{n \to \infty} \sum_{k=\kappa+1}^{n} A_{n,k} = \sum_{k=1}^{\infty} \int_{k-1}^{k} (y^\alpha - b_k^\alpha)^2 dy = J(\alpha, \kappa, b) < \infty, \quad (1.4.13)$$

we seek a sequence $\{A_k\}_{k=\kappa+1}^{\infty} \subset [0, \infty)$ such that

$$0 \leq A_{n,k} \leq A_k, \quad k = \kappa + 1, \ldots, n, \quad n \in \mathbb{N}.$$
and that \( \sum_{k=1}^{\infty} A_k < \infty \). Let us assume, without loss of generality, that \( \kappa = 0 \). Clearly, we may set \( A_1 := \sup_{n \in \mathbb{N}} A_n < \infty \). Consider now \( k \geq 2 \). The construction of \( A_k \) in this case parallels some arguments seen in the proof of Proposition 1.2.1, but we provide the details for the sake of clarity. By adding and subtracting \( b_k \frac{L_g(y/n)}{L_g(1/n)} \) and using the inequality \( (u + v)^2 \leq 2u^2 + 2v^2 \), we get

\[
A_{n,k} = \left( y_n \frac{L_g(y/n)}{L_g(1/n)} - b_k \frac{L_g(y/n)}{L_g(1/n)} \right)^2 + b_k \frac{L_g(b_k/n)}{L_g(1/n)} - b_k \frac{L_g(b_k/n)}{L_g(1/n)} \right)^2 dy
\]

\[
=: I_{n,k} + I'_{n,k}.
\]

(1.4.14)

Recall that \( L_g := \inf_{\xi \in (0,1)} L_g(\xi) > 0 \), so by the estimates \( b_k^{2\alpha} \leq \max(k^{2\alpha}, (k-1)^{2\alpha}) \leq 2(k-1)^{2\alpha} \), valid when \( \alpha < \frac{1}{2} \), we obtain

\[
I'_{n,k} \leq \frac{4}{L_g^2} (k-1)^{2\alpha} \int_{k-1}^{k} \left( L_g(y/n) - L_g(b_k/n) \right)^2 dy.
\]

Note that, thanks to (A1) and the mean value theorem,

\[
|L_g(y/n) - L_g(b_k/n)| = |L'_g(\xi)| \left| \frac{y - b_k}{n} \right| \leq \frac{C(1 + \xi^{-1})}{n} \leq \frac{C}{n} + \frac{C}{k-1} \leq \frac{2C}{k-1},
\]

where \( \xi = \xi(y/n, b_k/n) \in \left[ \frac{k-1}{n}, \frac{k}{n} \right] \) and where the final inequality follows since \( k - 1 < n \). Thus,

\[
I'_{n,k} \leq \frac{16C^2}{L_g^2} (k-1)^{2(\alpha-1)}.
\]

(1.4.15)

Moreover, the Potter bound (1.4.1) and inequality (1.2) imply

\[
I_{n,k} \leq 2\alpha^2 C_0^2 \int_{k-1}^{k} \left( \min(y, b_k) \right)^{2(\alpha-1)} y^{2\delta} dy \leq 2^{1+2\delta} \alpha^2 C_0^2 (k-1)^{2(\alpha-1+\delta)},
\]

(1.4.16)

where we choose \( \delta \in \left( 0, \frac{1}{2} - \alpha \right) \). Applying the bounds (1.4.15) and (1.4.16) to (1.4.14) shows that

\[
A_{n,k} \leq 2^{1+2\delta} \alpha^2 C_0^2 (k-1)^{2(\alpha-1+\delta)} + \frac{16C^2}{L_g^2} (k-1)^{2(\alpha-1)} =: A_k,
\]

where \( 2(\alpha - 1) < -1 \) and \( 2(\alpha - 1 + \delta) < -1 \) with our choice of \( \delta \), so that \( \sum_{k=1}^{\infty} A_k < \infty \). Thus, we have shown (1.4.13), which in turn implies that

\[
D'_n \sim I(\alpha, \kappa, \{b_k\}_{k=K+1}^{\infty}) n^{-2(\alpha+1)} L_g(1/n)^2, \quad n \to \infty.
\]

(1.4.17)
We will now use the obtained asymptotic relations, (1.4.10), (1.4.11), (1.4.12) and (1.4.17), to complete the proof. To this end, it will be convenient to introduce a relation \( x_n \gg y_n \) for any sequences \( \{x_n\}_{n=1}^{\infty} \) and \( \{y_n\}_{n=1}^{\infty} \) of positive real numbers that satisfy \( \lim_{n \to \infty} \frac{x_n}{y_n} = \infty \). By (1.4.12), we have \( D'_n \gg D_n \). Since \( 2\alpha + 1 < 2 \), we find that also \( D'_n \gg D''_n \), in view of (1.4.11). The assumption \( \gamma > -\frac{2\alpha + 1}{2\beta + 1} \) is equivalent to \(-\left(2\alpha + 1\right) > \gamma(2\beta + 1)\), so by the estimate (1.2.2) for slowly varying functions, we have \( D'_n \gg D''_n \). It then follows that \( E_n \sim E[\sigma(0)^2] D'_n \) as \( n \to \infty \). Further, the condition (1.2.11) implies that \( E_n \gg E'_n \). In view of (1.4.9), we finally find that \( \mathbb{E}\left[ |X_n(t) - X(t)|^2 \right] \sim \mathbb{E}[\sigma(0)^2] D'_n \) as \( n \to \infty \), which completes the proof.

**Acknowledgements**

We would like to thank Heidar Eyjolfsson and Emil Hedevang for useful discussions regarding simulation of \( \mathcal{BSF} \) processes and Ulises Márquez for assistance with symbolic computation. Our research has been supported by CREATES (DNRF78), funded by the Danish National Research Foundation, by Aarhus University Research Foundation (project “Stochastic and Econometric Analysis of Commodity Markets”) and by the Academy of Finland (project 258042).

**1.5 References**


Abstract

Using theory on (conditionally) Gaussian processes with stationary increments developed in Barndorff-Nielsen et al. (2009a, 2011), this paper presents a general semiparametric approach to conducting inference on the fractal index, \( \alpha \), of a time series. Our setup encompasses a large class of Gaussian processes and we show how to extend it to a large class of non-Gaussian models as well. It is proved that the asymptotic distribution of the estimator of \( \alpha \) does not depend on the specifics of the data generating process for the observations, but only on the value of \( \alpha \) and a “heteroskedasticity” factor. Using this, we propose a simulation-based approach to inference, which is easily implemented and is valid more generally than asymptotic analysis. We detail how the methods can be applied to test whether a stochastic process is a non-semimartingale. Finally, the methods are illustrated in two empirical applications motivated from finance. We study time series of log-prices and log-volatility from 29 individual US stocks; no evidence of non-semimartingality in asset prices is found, but we do find evidence of non-semimartingality in volatility. This confirms a recently proposed conjecture that stochastic volatility processes of financial assets are rough (Gatheral
Keywords: Fractal index; Monte Carlo simulation; roughness; semimartingality; fractional Brownian motion; stochastic volatility.

JEL Classification: C12, C22, C63, G12

MSC 2010 Classification: 60G10, 60G15, 60G17, 60G22, 62M07, 62M09, 65C05

2.1 Introduction

This paper is concerned with conducting statistical inference on the so-called fractal index of a time series. Formally, we consider zero-mean (conditionally) Gaussian stochastic processes \(X\), with stationary increments satisfying

\[
E[|X_{t+h} - X_t|^p] \sim C_p|h|^{(2\alpha+1)p/2}, \quad h \downarrow 0,
\]

for \(p > 0\), \(C_p > 0\), and \(\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)\). The parameter \(\alpha\) is termed the fractal index because it, under mild assumptions, is related to the fractal dimension \(D = \frac{3}{2} - \alpha\) of the sample paths of the process \(X\) (Falconer, 1990; Gneiting, Sevcikova, and Percival, 2012). The parameter \(\alpha\) is also called the roughness parameter as it reflects itself in the pathwise properties of \(X\), as will be made precise in Proposition 2.2.1 below. Informally, \(\alpha < 0\) implies rough sample paths of \(X\), while \(\alpha > 0\) implies smooth sample paths of the process. After a log-transformation, the scaling relationship (2.1.1) suggests a straightforward estimation procedure of \(\alpha\) based on OLS regression. As we shall see, the OLS estimator, \(\hat{\alpha}\), of \(\alpha\) comes with a central limit theorem (CLT) of the type

\[
\tau_n(\hat{\alpha} - \alpha) \overset{d}{\to} Z_p \cdot S_p, \quad n \to \infty,
\]

where \(n\) is the number of observations, \(\tau_n\) is a rate sequence (e.g. \(\tau_n = \sqrt{n}\)), and \(Z_p\) is a random variable. \(S_p > 0\) is a heteroscedasticity factor, which is equal to one when \(X\) is Gaussian and different from one otherwise. It is well known that the random variable \(Z_p\) is Gaussian when \(\alpha \in (-1/2, 1/4]\) (e.g. Breuer and Major, 1983). However, when \(\alpha > 1/4\), Taqqu (1975) showed that the distribution of \(Z_p\) is given by the Rosenblatt distribution in a quite complicated way. What is more, although the distribution of \(Z_p\) is known in theory, it is generally infeasible to calculate it explicitly when \(p \neq 2\).

The contribution of this paper is to suggest bootstrap/Monte Carlo procedures to the problem of testing statistical hypotheses regarding the fractal index, as well as computing confidence intervals around the estimated value. Specifically, we prove that (2.1.2) holds for a large class of fractal processes, where the distribution of the random variable \(Z_p\) is invariant to the actual data generating process (DGP) of \(X\).
That is, the distribution of $Z_p$ depends only on the value of $\alpha$, i.e. of the roughness properties of the data. We suggest a simple estimator of $S_p$ and then to approximate the distribution of $Z_p$ by Monte Carlo simulation. The key to the Monte Carlo approach is the invariance of the distribution of $Z_p$, enabling one to approximate it by any (Gaussian) fractal process; the obvious choice for such an auxiliary process is the canonical fractal process, the fractional Brownian motion (fBm, Mandelbrot and Van Ness, 1968). The upshot is that the methods presented in this paper are valid for all $p > 0$ and $\alpha \in (-1/2, 1/2)$. Indeed, since the auxiliary simulations directly approximate the distribution of $Z_p$, it is not necessary to distinguish between the Gaussian case ($\alpha \leq 1/4$) and the Rosenblatt case ($\alpha > 1/4$), nor of taking into account the actual convergence rate, $\tau_n$.

Another advantage of the simulation-based approach, as compared to asymptotic analysis, is that the OLS estimator of $\alpha$ requires a choice of bandwidth $m \in \mathbb{N}$, denoting the number of data points used in the regression. The distribution of $Z_p$ will in turn depend on this parameter and this distribution gets increasingly complicated with increasing $m$. Our approach trivially allows for any $m \geq 2$ and we will exploit this to give some guidelines on how to select $m$. In summary, simulations show that choosing $m$ slightly larger than 2 results in a good trade-off between small bias (obtained by having $m$ small) and low variance (obtained by larger $m$). We recommend setting $m = 3$ and will do so in our applications. This is slightly at odds with conventional wisdom which recommends $m = 2$ (e.g. Davies and Hall, 1999).

Most previous work on inference for the fractal index has been done assuming Gaussianity of $X$; notable exceptions being Chan and Wood (2004) and Achard and Coeurjolly (2010). Indeed, in a survey of the asymptotic theory, Gneiting et al. (2012) section 3.1., reports that "a general non-Gaussian theory remains lacking". Below, we will show how to extend the theory to a large class of non-Gaussian processes obtained by volatility modulation of an otherwise Gaussian process. When $X$ is a so-called volatility modulated Brownian semistationary process (Barndorff-Nielsen and Schmiegel, 2007, 2009) a CLT, which obeys (2.1.2), is derived in Theorem 2.4.1. When $\alpha \in (-1/2, 1/4)$, $Z_p$ will be Gaussian with an asymptotic distribution that is feasible to calculate for $p = 2$. That is, this CLT suffers from the same drawbacks as other asymptotic approaches, as discussed above. The bootstrap approach of this paper is still valid generally, i.e. for all $\alpha \in (-1/2, 1/2)$ and $p > 0$, however.

Two different versions of the bootstrap procedure are suggested. The first method is specifically tailored to test null hypotheses such as $H_0 : \alpha = \alpha_0$ for $\alpha_0 \in (-1/2, 1/2)$; of special interest is $\alpha_0 = 0$ which means $X$ has the same roughness as the Brownian motion and is thus, more or less, a semimartingale. For this method the auxiliary fBms are generated with fractal index $\alpha_0$, i.e. the fractal index that $X$ has under the null. The second version is tailored to make confidence intervals around the estimated value $\hat{\alpha}$: here $\alpha$ is first estimated from the observations of $X$, and then the auxiliary fBms are generated such that they have fractal index equal to $\hat{\alpha}$. 
In some respects, the simulation-based approach to inference of this paper is similar to the semiparametric bootstrap method of Hall, Härdle, Kleinow, and Schmidt (2000). Here the authors consider bootstrapping the Hurst index $H$ of a process using the classical rescaled/range (RS) estimator of Mandelbrot and Van Ness (1968). It is well known that the fractal index of the fBm— and indeed of any self-similar process—is related to the Hurst index of that process through the identity $H = \alpha + \frac{1}{2}$ (e.g. Gneiting and Schlather, 2004). So, when the underlying DGP of $X$ is self-similar (e.g. when $X$ is an fBm), the approach of Hall et al. (2000) can be used to estimate, and make inference on, the fractal index $\alpha$ of $X$ by exploiting the identity $\alpha = H - \frac{1}{2}$. However, when $X$ is not self-similar, there is generally no simple relationship between $H$ and $\alpha$, and therefore RS will be inadequate to estimate $\alpha$. Evidently, in this important case, the approach of Hall et al. (2000) will therefore not be applicable to perform inference on $\alpha$, which is the goal in this paper. For other bootstrap approaches related to the fractal index $\alpha$ or the Hurst index $H$, see Davies and Hall (1999), Grau-Carles (2005), Kim and Nordman (2013), and Bennedsen et al. (2016a).

This paper does not restrict attention to self-similar models. Indeed, most of the DGPs considered below will either be stationary or allowing $\alpha$ and $H$ to vary independently of each other, or both. Time series having these features are generically not self-similar and the methods of Hall et al. (2000) do therefore not apply for such DGPs when performing inference on the fractal index for the reasons explained above. Note that especially the 'rough' case $\alpha < 0$ seems to be relevant in empirical applications. For instance, the celebrated scaling laws of Kolmogorov (Kolmogorov, 1941) predict time series of the main velocity component of turbulence with $\alpha = -1/6$, see e.g. Corcuera, Hedevang, Pakkanen, and Podolskij (2013). In finance, Barndorff-Nielsen, Benth, and Veraart (2013a) and Bennedsen (2015) model electricity spot prices using fractal processes with $\alpha < 0$, and recently estimates around $\alpha \approx -0.40$ have been found for time series of volatility (Gatheral et al., 2014; Bayer et al., 2015; Bennedsen et al., 2016b). Note that the DGPs underlying all these empirical examples are, besides fractal, generally believed to be stationary and therefore not self-similar. For this reason, an inference approach that can accommodate this—e.g. the one presented in this paper—is crucial for such applications. Additionally, the time series underlying these examples are likely non-Gaussian as well, and the present approach also allows for this, as discussed above and shown below.

The rest of the paper is structured as follows. Section 2.2 presents the mathematical setup and assumptions and gives an example of the kind of processes we have in mind. Section 2.3 presents the semiparametric estimator of the fractal index and then sets forth the two bootstrap methods. Section 2.4 presents some extensions to the basic setup. Section 2.5 contains extensive simulation evidence that the bootstrap methods work well for a wide range of DGPs, with and without the presence of stochastic volatility. Finally, Section 3.2 presents some empirical applications motivated from finance. Section 3.7 concludes. Proofs of technical results, some
mathematical derivations, and details about the DGPs from the simulation study are given in the Appendix.

### 2.2 Setup

Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space supporting \( X \), a one-dimensional, zero-mean, stochastic process with stationary increments. Define the \( p \)'th order variogram of \( X \):

\[
\gamma_p(h) := \mathbb{E} \left[ |X_{t+h} - X_t|^p \right], \quad h \in \mathbb{R}.
\]

As we intend to make use of the theory developed in Barndorff-Nielsen et al. (2009a, 2011) we adopt the assumptions of those papers. The assumptions are standard in the literature on fractal processes and are as follows.

(A1) For some \( \alpha \in \left( -\frac{1}{2}, \frac{1}{2} \right) \),

\[
\gamma_2(x) = x^{2\alpha + 1} L(x), \quad x \in (0, \infty),
\]

where \( L : (0, \infty) \to [0, \infty) \) is continuously differentiable and bounded away from zero in a neighborhood of \( x = 0 \). The function \( L \) is assumed to be slowly varying at zero, in the sense that \( \lim_{x \to 0} \frac{L(tx)}{L(x)} = 1 \) for all \( t > 0 \).

(A2) \[ \frac{d^2}{dx^2} \gamma_2(x) = x^{2\alpha - 1} L_2(x) \] for some slowly varying (at zero) function \( L_2 \), which is continuous on \( (0, \infty) \).

(A3) There exists \( b \in (0, 1) \) with

\[
\limsup_{x \to 0} \sup_{y \in [x, xb]} \left| \frac{L_2(y)}{L(x)} \right| < \infty.
\]

**Remark 2.2.1.** The technical assumption (A3) can be replaced by the weaker assumption

\[
\left| \frac{\gamma_2((j+1)/n) - 2\gamma_2(j/n) + \gamma_2((j-1)/n)}{2\gamma_2(1/n)} \right| \leq r(j), \quad \frac{1}{n} \sum_{j=1}^{n} r(j)^2 \to 0, \quad n \to \infty,
\]

for some sequence \( r(j) \).

The parameter \( \alpha \in (-1/2, 1/2) \) is called the **fractal index** or **roughness index** of \( X \), since the value of \( \alpha \) reflects itself in the **pathwise properties** of \( X \), as the following result formalizes.

---

\(^1\)Following Bennedsen et al. (2016a) this assumption is replaced by the following in the case \( \alpha = 0 \): 
(A2') \[ \frac{d^2}{dx^2} \gamma_2(x) = f(x)L_2(x), \] where \( L_2 \) is as in (A2), and the function \( f \) is such that \( |f(x)| \leq Cx^{-\beta} \) for some constants \( C > 0 \) and \( \beta > 1/2 \).
**Proposition 2.2.1.** Let \( X \) be a Gaussian process with stationary increments satisfying (A1) with fractal index \( \alpha \in (-1/2, 1/2) \). Then there exists a modification of \( X \) which has locally Hölder continuous trajectories of order \( \phi \) for all \( \phi \in \left(0, \alpha + \frac{1}{2}\right) \).

**Remark 2.2.2.** Proposition 2.2.1, as well as most of the asymptotic theory developed in the statistical literature on the fractal index, takes its departure point in (stationary) Gaussian processes. However, the theory, and also the bootstrap method in this paper, is valid more generally. For instance, the scaling relationship (2.2.1) and its implications for the pathwise properties of \( X \) as expounded in Proposition 2.2.1 is not only applicable to Gaussian processes, see Constantine and Hall (1994) and the discussion in Davies and Hall (1999). We explicitly consider the extension to non-Gaussian processes in Section 2.4.

Proposition 2.2.1 shows that \( \alpha \) controls the degree of (Hölder) continuity of \( X \). In particular, negative values of \( \alpha \) corresponds to \( X \) having very rough paths, while positive values of \( \alpha \) corresponds to smooth paths. It is well known that the Brownian motion has \( \alpha = 0 \). Indeed, \( \alpha = 0 \) is a necessary condition for \( X \) to be a semimartingale, as shown in Section 2.2. An example of the kind of processes we have in mind is the Matérn process.

**Example 2.2.1** (Matérn process, Matérn (1960); Handcock and Stein (1993)). Let \( X \) be the zero-mean, unit variance, stationary Gaussian process with the Matérn correlation function, i.e. with correlation function

\[
\rho(h) = \frac{2^{-\alpha+1/2}}{\Gamma(\alpha + 1/2)}(\beta h)^{\alpha+1/2} K_{\alpha+1/2}(\beta h), \quad h \in \mathbb{R},
\]

with \( \alpha \in (-1/2, 1/2) \), \( \beta > 0 \), and where \( K_\nu \) is the (modified) Bessel function of the third kind with index \( \nu \) (e.g., Gradshteyn and Ryzhik, 2007, sections 8.4 and 8.5). We will call \( X \) a Matérn process; this process is widely used in diverse statistical fields, such as geostatistics, machine learning, and econometrics. Using properties of the Bessel functions, see e.g. Barndorff-Nielsen (2012), one can show that the Matérn process fulfills assumptions (A1)–(A3).

To get an intuitive understanding of how the trajectories of the fractal processes look, and in particular how the value of \( \alpha \) reflects itself in the roughness of the paths, Figure 2.1 plots three simulated trajectories of the Matérn process. It is evident how negative values of \( \alpha \) correspond to very rough paths, while the paths become smoother as \( \alpha \) increases.

**Testing for non-semimartingality**

Fractal processes with \( \alpha \neq 0 \) are not semimartingales, which implies that the null hypothesis \( H_0 : \alpha = 0 \) can be used to test whether a process is a non-semimartingale. In Appendix A.2 the following basic fact is shown.
2.3. Semiparametric estimation and bootstrapping of the fractal index

Consider \( n \) equidistant observations \( X_{1_1/n}, X_{2_1/n}, \ldots, X_1 \) of the stochastic process \( X \), observed over a fixed time interval, which we without loss of generality take to be the unit interval, so that the time between observations is \( \Delta_n := \frac{1}{n} \). As \( n \to \infty \), this gives rise to the so-called \textit{in-fill asymptotics}. In what follows, suppose that the process \( X \) satisfies the assumptions (A1)–(A3).

When \( X \) is Gaussian, it holds, by standard properties on the (absolute) moments of the Gaussian distribution and (2.2.1), that

\[
\gamma_p(h) = C_p |h|^{(2\alpha+1)p/2} L_p(h), \quad h \in \mathbb{R},
\]

where \( p > 0 \), \( L_p \) is slowly varying at zero and \( C_p > 0 \) is a constant. This motivates the regression

\[
\log \hat{\gamma}_p(h) = c_p + a \log |h| + U_h + \epsilon_h, \quad h = \frac{1}{n}, \frac{2}{n}, \ldots, \frac{m}{n},
\]

where \( m \in \mathbb{N} \) is a bandwidth parameter,

\[
c_p = \log C_p, \quad a = \frac{(2\alpha + 1)p}{2}, \quad U_h = \log \left( \frac{\hat{\gamma}_p(h)}{\gamma_p(h)} \right), \quad \text{and} \quad \epsilon_h = \log L_p(h).
\]
The variogram $\gamma_p$ is estimated straightforwardly as
\[
\hat{\gamma}_p(k/n) := \frac{1}{n-k} \sum_{i=1}^{n-k} |X_{i+k} - X_i|^p, \quad k \geq 1,
\]  
(2.3.3)
and, given an estimate $\hat{\alpha}$ of $\alpha$ from (2.3.2), the estimate of the fractal index is $\hat{\alpha} = \frac{\hat{\alpha}}{p} - \frac{1}{2}$. This estimator is well known and well tested in the literature, e.g. Gneiting and Schlather (2004); Bennedsen (2015). The following proposition shows the consistency of the OLS estimator of $\alpha$.

**Proposition 2.3.1.** Suppose $X$ is Gaussian and satisfies assumptions (A1)–(A3) with fractal index $\alpha \in (-1/2, 1/2)$. Fix $p > 0$, $m \in \mathbb{N}$ and let $\hat{\alpha} = \hat{\alpha}_{p,m}$ be the OLS estimator of $\alpha$ from (2.3.2). Now, as $n \to \infty$,
\[
\hat{\alpha} \xrightarrow{\mathbb{P}} \alpha,
\]
where "$\xrightarrow{\mathbb{P}}$" refers to convergence in $\mathbb{P}$-probability.

A number of studies have considered the asymptotic properties of the OLS estimates coming from the regression (2.3.2), e.g. Constantine and Hall (1994), Davies and Hall (1999), and Coeurjolly (2001, 2008). For a brief summary of this literature, see Gneiting et al. (2012), Section 3.1. The following theorem presents the details when $X$ is a Gaussian process satisfying (A1)–(A3) plus a further technical assumption:

(A4) For the slowly varying function $L$ of assumption (A1), it holds that there exists a constant $C > 0$ such that the derivative $L'$ of $L$ satisfies
\[
|L'(x)| \leq C \left(1 + x^{-\delta}\right), \quad x \in (0, 1],
\]
for some $\delta \in (0, 1/2)$.

**Remark 2.3.1.** By Barndorff-Nielsen (2012), page 8, the Matérn process of Example 2.2.1 fulfills assumption (A4) when $\alpha \in (-1/2, 1/4)$.

**Theorem 2.3.1.** Suppose $X$ is Gaussian and satisfies assumptions (A1)–(A4) with fractal index $\alpha \in (-1/2, 1/4)$. Fix $m \in \mathbb{N}$ and let $\hat{\alpha} = \hat{\alpha}_{p,m}$ be the OLS estimator of $\alpha$ from (2.3.2). Now, as $n \to \infty$,
\[
\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} Z_p, \quad Z_p \sim N\left(0, \sigma^2_{m,p}\right),
\]
where
\[
\sigma^2_{m,p} = \frac{x_m^T \Lambda_p x_m}{(x_m^T x_m)^2 p^2},
\]
with \( T \) denoting the transpose of a vector and \( x_m \) is the \( m \times 1 \) vector
\[
x_m = \left( \log 1 - \log m, \log 2 - \log m, \ldots, \log m - \log m \right)^T, \quad \log m := \frac{1}{m} \sum_{k=1}^{m} \log k,
\]
and \( \Lambda_p = \{ \lambda_p^{k,v} \}_{k,v=1}^{m} \) is a real-valued \( m \times m \) matrix with entries
\[
\lambda_p^{k,v} = \lim_{n \to \infty} n \cdot \text{Cov} \left( \hat{\gamma}_p(k/n; B^H), \hat{\gamma}_p(v/n; B^H) \right), \quad k, v = 1, 2, \ldots, m,
\]
where \( \hat{\gamma}(.; B^H) \) is given by (2.3.3) when the underlying process is a fractional Brownian motion with Hurst parameter \( H = \alpha + \frac{1}{2} \), and similarly for \( \gamma_p(.; B^H) \).

**Remark 2.3.2.** The limit in (2.3.4) exists for \( k, v = 1, \ldots, m \) by Breuer and Major (1983), Theorem 1. See also Remark 3.3. in Corcuera et al. (2013).

**Remark 2.3.3.** As mentioned in the introduction, a result similar to Theorem 2.3.1 holds for \( \alpha \geq 1/4 \) but with a different convergence rate and limiting distribution. When \( \alpha = 1/4 \), the convergence rate is \( \sqrt{n \log n} \) and \( Z_p \) is zero-mean Gaussian with an asymptotic variance different from \( \sigma^2_{m,p} \). When \( \alpha > 1/4 \) the convergence rate is \( n^{1-2\alpha} \) and the distribution of \( Z_p \) is of the Rosenblatt type, see Taqqu (1979).

**Remark 2.3.4.** The most relevant setup for us when conducting asymptotic analysis is the case \( m = 2 \). In this case we have
\[
\sigma^2_{2,p} = \frac{\lambda^{11}_p + \lambda^{22}_p - 2\lambda^{12}_p}{p^2 \log(2)^2}.
\]
Further, for \( p = 2 \) we can calculate the matrix \( \Lambda_p = \{ \lambda_p^{k,v} \}_{k,v=1}^{2} \) explicitly. The details are given in Appendix A.1.

Perhaps surprisingly, Theorem 2.3.1 shows that the asymptotic distribution of the OLS estimator does not depend on the precise structure of the DGP of \( X \), but only on the value of the fractal index \( \alpha \), through the correlation structure of the increments of an fBm with Hurst index \( H = \alpha + 1/2 \). The reason for this is that the small scale behavior of a process \( X \) fulfilling assumption (A1), will have the same small scale behavior as increments of the fBm. To see this, write
\[
r_n(j) := \text{Corr} \left( X_{j+1/n} - X_{j/n}, X_{1/n} - X_0 \right)
= \frac{\gamma_2((j+1)/n) - 2\gamma_2(j/n) + \gamma_2((j-1)/n)}{2\gamma_2(1/n)}
= -\frac{1}{2} \left( |j+1|^{2\alpha+1} - 2|j|^{2\alpha+1} + |j-1|^{2\alpha+1} \right), \quad n \to \infty,
\]
by assumption (A1) and the properties of slowly varying functions. We recognize (2.3.5) as the correlation function of the increments of an fBm with Hurst index...
$H = \alpha + 1/2$. As shown in the proof of Theorem 2.3.1, this will imply that the asymptotic variance of the estimator, $\sigma_m^2$, is the same for all Gaussian processes fulfilling assumptions (A1)–(A4), including the fBm. This fact will be the basis of our Monte Carlo algorithm: we will approximate the asymptotic distribution of $Z_p$ using a fractional Brownian motion and, by Theorem 2.3.1, this will allow us to conduct correct inference, even though the DGP of $X$ itself might be very different from the fBm.

Theorem 2.3.1, and the discussion preceding it, alludes to some of the problems with asymptotic analysis for conducting inference on $\alpha$ in the general, semiparametric, setting (2.2.1). Some advantages of a simulation-based approach, as compared to asymptotic analysis include:

1. Our methods trivially allow for any $p > 0$. In contrast, a general and feasible asymptotic theory seems to be lacking in this case (although see Coeurjolly, 2001). For instance, the calculation of $\Lambda_p$ in Theorem 2.3.1, cf. (2.3.4), appears to be infeasible for $p \neq 2$. Having an approach that allows for any $p > 0$ is desirable, as other values, e.g. $p = 1/2$ or $p = 1$, have been shown to be preferable to use in some cases. In particular, Gneiting et al. (2012) recommend using $p = 1$, as this choice is robust to outliers in the data, while Corcuera et al. (2013) recommend using several values for $p$ for checking the robustness of $\hat{\alpha}$.

2. It is well known that the asymptotic distribution of the OLS estimator (2.3.2) of $\alpha$ is Gaussian for $\alpha \in (-1/2, 1/4]$, but non-standard for $\alpha > 1/4$. In the latter case, the distribution is given by the Rosenblatt distribution, as outlined in Taqqu (1975). As explained in Remark 2.3.3, the convergence rate is also different depending on the true value of $\alpha$. This is the reason for assuming $\alpha \in (-1/2, 1/4)$ in Theorem 2.3.1. In contrast, we find that our Monte Carlo approach works well for the whole range $\alpha \in (-1/2, 1/2)$. This is because this approach directly approximates the distribution of $\hat{\alpha} - \alpha$, i.e. it automatically replicates any non-Gaussianity in the limiting distribution, induced by values of $\alpha$ larger than $1/4$, as well as differing convergence rates.

3. The methods of this paper are applicable and easily implemented for any choice of bandwidth parameter $m \geq 2$. In particular, implementation is not encumbered by choosing $m$ large. In contrast, asymptotic theory is significantly more complex when $m = 3$; this is illustrated in Theorem 2.3.1, where the matrix $\Lambda_p$ is an $m \times m$ matrix. This will allow us to investigate, through simulations, which value of $m$ is preferred to use in practice. In Section 2.5 simulations will show that the root mean squared error of the OLS estimator of $\alpha$ is minimized for $m \geq 3$; also, when the sample size is moderate to large, choosing $m = 3$ or $m = 4$ will result in hypothesis tests with better size and power properties. These findings are in contrast with the conventional wisdom for asymptotic analysis, where it is usually recommended to let $m = 2$. 


4. As mentioned in Section 3.1. of Gneiting et al. (2012), a general asymptotic theory for non-Gaussian $X$ seems to be lacking. In Section 2.4, we will introduce non-Gaussianity of $X$, through volatility modulation, and show how our bootstrap approach is valid for a large class of non-Gaussian processes.

Two different bootstrap (or, more accurately, Monte Carlo) approaches, are detailed in the following two subsections. The idea behind the procedures is to simulate an auxiliary fractional process, an fBm, and estimate the fractal index of this process using (2.3.2) and then perform this operation $B \in \mathbb{N}$ times to approximate the distribution of $\hat{\alpha} - \alpha$. By Theorem 2.3.1 this approach will work no matter the actual DGP of $X$, as long as $X$ is a Gaussian process satisfying (A1)–(A4). In other words, when $X$ is Gaussian, we can use auxiliary fBms to estimate the distribution of $Z_p$ in (2.1.2), even though the DGP of $X$ might be very different from the fBm.

It turns out, however, that when $X$ contains stochastic volatility (thus inducing non-Gaussianity of $X$) $\hat{\alpha}$ will have a limiting distribution which is different from the one which were described in Theorem 2.3.1. This is formalized in Theorem 2.4.1 below, where the volatility modulated Brownian semistationary process is introduced. For such processes, it is generally the case that $S_p \neq 1$ in (2.1.2) and the distribution of $Z_p$ thus needs to be corrected by this factor to obtain the correct distribution of $\hat{\alpha}$. For this purpose we propose the following simple estimator of $S_p$:

$$\hat{S}_p = \sqrt{\frac{m_p^{-1} \hat{\gamma}_p(1/n)}{m_p^{-1} \gamma_p(1/n)}}, \quad m_s := E[|U|^s], \quad U \sim N(0, 1), \quad s > 0, \quad (2.3.6)$$

where $\hat{\gamma}$ is defined in (2.3.3). The $\hat{S}_p$ factor provides a “heteroskedasticity correction” and will be rigorously motivated in Theorem 2.4.1 below. In Appendix A.2 the following is shown.

**Proposition 2.3.2.** Suppose $X$ is a Gaussian process satisfying (A1)–(A3). Let $p > 0$. Now,

$$\hat{S}_p \overset{p}{\to} 1, \quad n \to \infty.$$

**Remark 2.3.5.** The fBm fulfills the assumptions of Proposition 2.3.2.

Theorem 2.3.1 showed that when $X$ does not contain heteroskedasticity, no correction by $S_p$ is needed. Proposition 2.3.2 tells us that in these cases correction by $\hat{S}_p$ is harmless asymptotically. We will see in Section 2.5 that correction by $S_p$ is crucial to obtain valid inference using the bootstrap when the underlying DGP contains significant variability, e.g. because of the presence of stochastic volatility. In practice, we therefore recommend to always correct with $\hat{S}_p$. In the algorithms and simulations below, bootstraps both without the correcting factor $\hat{S}_p$ (termed "$S^*$") and with the correcting factor $\hat{S}_p$ (termed "$T^*$") will be considered; we think of the latter as a studentized version of the former.
Bootstrap procedure 1

To test the hypothesis \( H_0 : \alpha = \alpha_0 \) for some \( \alpha_0 \in \left(-\frac{1}{2}, \frac{1}{2}\right) \) consider the following bootstrap approach:

1. Estimate \( \hat{\alpha} \) using the OLS regression in (2.3.2).
2. Compute \( S = \hat{\alpha} - \alpha_0 \) and \( T = (\hat{\alpha} - \alpha_0)/\hat{S}_p \), where \( \hat{S}_p \) is given by (2.3.6).
3. For \( b = 1, 2, \ldots, B \).
   a) Simulate \( n \) observations of an fBm with Hurst index \( H^* = \alpha_0 + \frac{1}{2} \).
   b) Estimate \( \hat{\alpha}^*_b \) from the path of this fBm using the OLS regression in (2.3.2).
   c) Compute \( R^*_b = \hat{\alpha}^*_b - \alpha_0 \).
4. Compare \( S \) and \( T \) with the appropriate percentile of \( \{R^*_b\}_{b=1}^B \). P-values can be obtained by determining the relative position of \( S \) or \( T \) in the sorted version of \( \{R^*_b\}_{b=1}^B \).

Bootstrap procedure 2

Instead of the above, consider the following bootstrap procedure which is more appropriate when one wishes to compute confidence intervals for \( \hat{\alpha} \).

1. Estimate \( \hat{\alpha} \) using the OLS regression in (2.3.2).
2. Compute \( S = \hat{\alpha} - \alpha_0 \) and \( T = (\hat{\alpha} - \alpha_0)/\hat{S}_p \), where \( \hat{S}_p \) is given by (2.3.6).
3. For \( b = 1, 2, \ldots, B \).
   a) Simulate \( n \) observations of an fBm with Hurst index \( H^* = \hat{\alpha} + \frac{1}{2} \).
   b) Estimate \( \hat{\alpha}^*_b \) from the path of this fBm using the OLS regression in (2.3.2).
   c) Compute \( R^*_b = \hat{\alpha}^*_b - \hat{\alpha} \).
4. Compare \( S \) and \( T \) with the appropriate percentiles of \( \{R^*_b\}_{b=1}^B \). To obtain confidence intervals at the level \( 1 - \delta \), e.g. a 95% confidence interval for \( \delta = 0.05 \), one computes
   \[
   CI^{S}_{1-\delta/2} = \left[ \hat{\alpha} - R^*_b(1-\delta/2), \hat{\alpha} - R^*_b(\delta/2) \right],
   \]
or
   \[
   CI^{T}_{1-\delta/2} = \left[ \hat{\alpha} - \hat{S}_p \cdot R^*_b(1-\delta/2), \hat{\alpha} - \hat{S}_p \cdot R^*_b(\delta/2) \right],
   \]
where \( R^*_b \) is the \( x \)'th empirical percentile of \( \{R^*_b\}_{b=1}^B \), and where \( \hat{S}_p \) were calculated in step 2.
Remark 2.3.6. In bootstrap procedure 1, the bootstrap auxiliary variables are generated under $H_0$ by setting $\alpha = \alpha_0$. This will minimize the probability of a Type I error, i.e. rejection of $H_0$ when $H_0$ is true (Davidson and MacKinnon, 1999). We therefore expect this method to result in superior size properties of the hypothesis test as compared to procedure 2.

2.4 Extensions

This section gives two extensions to the theory discussed above. First the case of non-Gaussianity, through volatility modulation of $X$, is considered and then the case where the increments of $X$ are non-stationary.

Extension to stochastic volatility processes

A flexible way to introduce non-Gaussianity of processes for which the theory of the fractal index continues to hold is through volatility modulation. A very convenient process for this purpose is the Brownian semistationary process.

Brownian semistationary process

Let $X$ be the (volatility modulated) Brownian semistationary (BS$S$) process (Barndorff-Nielsen and Schmiegel, 2007, 2009),

$$X_t = \int_{-\infty}^{t} g(t-s)\sigma_s dW_s, \quad t \geq 0,$$

where $W$ is a Brownian motion on $\mathbb{R}$, $\sigma = \{\sigma_t\}_{t \in \mathbb{R}}$ a stationary process, and $g$ a Borel measurable function such that $\int_{-\infty}^{t} g(t-s)^2 \sigma_s^2 ds < \infty$ a.s. See Appendix A.3 for details of the BS$S$ process and for the specifications of the stochastic volatility process $\sigma$ considered in this paper.

When $\sigma$ is a stochastic process, the marginal distribution of $X$ will be non-Gaussian; indeed, by conditioning on $\sigma$ and assuming independence of $W$, it is seen that the marginal distribution of $X$ is a normal mean-variance mixture:

$$X_t | (\sigma_s, s \leq t) \sim N\left(0, \int_{0}^{\infty} g(x)^2 \sigma_{t-x}^2 dx\right), \quad t \geq 0.$$

This is a convenient way of extending the theory of fractal processes beyond the Gaussian framework, encompassing a large class of non-Gaussian models. For instance, Barndorff-Nielsen et al. (2013a) show that for a particular choice of kernel function $g$ and stochastic volatility process $\sigma$, $X$ will have a marginal distribution of the ubiquitous Normal Inverse Gaussian type.

If the theory considered in this paper is to hold for BS$S$ processes, some technical assumptions on the kernel function $g$ are required. These assumptions are as follows.
(A5) It holds that

a) \( g(x) = x^\alpha L_g(x) \), where \( L_g \) is slowly varying at zero.

b) \( g'(x) = x^{\alpha-1} L_{g'}(x) \), where \( L_{g'} \) is slowly varying at zero, and, for any \( \epsilon > 0 \), we have \( g' \in L^2((\epsilon, \infty)) \). Also, for some \( a > 0 \), \( |g'| \) is non-increasing on the interval \((a, \infty)\).

c) For any \( t > 0 \),

\[
F_t := \int_1^\infty |g'(x)|^2 \sigma_{t-x}^2 dx < \infty.
\]

The kernel function gives the \( \mathcal{BFS} \) framework great flexibility. A particularly useful kernel function which has been applied in a number of studies, e.g. Barndorff-Nielsen et al. (2013a) and Bennedsen (2015), is the so-called gamma kernel.

**Example 2.4.1** (\( \Gamma \)-\( \mathcal{BFS} \) process). Let \( g \) be the gamma kernel, i.e. \( g(x) = x^\alpha e^{-\lambda x} \) for \( \alpha \in (-1/2, 1/2) \) and \( \lambda > 0 \). The resulting process

\[
X_t = \int_{-\infty}^t (t-s)^\alpha e^{-\lambda(t-s)} \sigma_s dW_s, \quad t \geq 0,
\]

is called the (volatility modulated) \( \Gamma \)-\( \mathcal{BFS} \) process. It is not hard to show that this process fulfills assumptions (A1)–(A5), see, e.g., Example 1.2.1. in Bennedsen et al. (2015).

Another requirement is that the modulating process \( \sigma = \{\sigma_t\}_{t\in \mathbb{R}} \) not be “too rough”. As in Corcuera et al. (2013), the following assumption is therefore introduced.

(A6) For any \( q > 0 \), it holds that

\[
\mathbb{E}[|\sigma_t - \sigma_s|^q] \leq C_q |t-s|^{\xi q}, \quad t, s \in \mathbb{R},
\]

for some \( \xi > 0 \) and \( C_q > 0 \). For a power parameter \( p > 0 \), cf. equation (2.3.1), we further require that \( \xi \cdot \min\{p,1\} > 1/2 \).

**Remark 2.4.1.** In Bennedsen et al. (2015) it was shown that \( \mathcal{BFS} \) processes satisfying (A1)–(A6) will have the same fractal and continuity properties as their Gaussian counterparts. That is, for such a \( \mathcal{BFS} \) process Propositions 2.2.1 and 2.2.2 continues to hold. In other words, \( X \) will have a modification with Hölder continuous trajectories of order \( \phi \) for all \( \phi \in (0, \alpha + 1/2) \) and \( X \) will be a non-semimartingale when \( \alpha \neq 0 \).

---

\(^2\)Again following Bennedsen et al. (2016a), in the case \( \alpha = 0 \) an alternative assumption is adopted: (A5b') \( g'(x) = L_{g'}(x) \), where \( L_{g'} \) is as in (A5b).
Processes of the $BIS$ type have been extensively studied recently. In particular, using theory developed in Barndorff-Nielsen et al. (2009a, 2011), see also Barndorff-Nielsen, Corcuera, and Podolskij (2013b) and Corcuera et al. (2013), we can prove $BIS$ analogues to the Gaussianity-based results presented above. The following proposition shows the consistency of the estimator of the fractal index, when the underlying process is a $BIS$ process; it is proved in the same way as Proposition 2.3.1 and we therefore skip the details.

**Proposition 2.4.1.** Suppose $X$ is a $BIS$ process satisfying (A1)–(A6) with fractal index $\alpha \in (-1/2, 1/2)$. Fix $p > 0$, $m \in \mathbb{N}$ and let $\hat{\alpha} = \hat{\alpha}_{p,m}$ be the OLS estimator of $\alpha$ from (2.3.2). Now, as $n \to \infty$,

$$\hat{\alpha} \xrightarrow{P} \alpha,$$

where "$\xrightarrow{P}$" refers to convergence in $P$-probability.

Likewise, the following theorem provides a CLT for the estimator of the fractal index when $X$ is a $BIS$ process and is thus an analogue to Theorem 2.3.1. The proof is given in Appendix A.2.

**Theorem 2.4.1.** Suppose $X$ is a $BIS$ process satisfying (A1)–(A6) with fractal index $\alpha \in (-1/2, 1/4)$. Fix $p > 0$, $m \in \mathbb{N}$, and let $\hat{\alpha} = \hat{\alpha}_{p,m}$ be the OLS estimator of $\alpha$ from (2.3.2). Now, as $n \to \infty$,

$$\sqrt{n} (\hat{\alpha} - \alpha) \xrightarrow{st} Z_p \cdot S_p, \quad Z_p \sim N\left(0, \sigma^2_{m,p}\right),$$

where $\sigma^2_{m,p}$ is as in Theorem 2.3.1 and

$$S_p = \sqrt{\int_0^1 \sigma^2_p s \, ds \over \int_0^1 \sigma^2_p \, ds}.$$

Here "$st$" denotes stable convergence (in law), see e.g. Rényi (1963).

**Remark 2.4.2.** Theorem 2.4.1 is similar to Theorem 2.3.1 except that the limiting distribution now has a heteroskedasticity term, $S_p$. By Theorem 3.1. in Corcuera et al. (2013) it is not hard to deduce that

$$\hat{S}_p = \sqrt{\frac{m^{-1} \hat{\gamma}_{2p} (1/n)} \over m^{-1} \hat{\gamma}_{p} (1/n)} \xrightarrow{P} S_p, \quad n \to \infty,$$

where $\hat{S}_p$ is given by (2.3.6). This fact provides the theoretical justification for the choice of $\hat{S}_p$ in the bootstrap algorithms in Section 2.3. Further, the stable convergence result of Theorem 2.4.1 allows us to deduce that,

$$\sqrt{n} \frac{\hat{\alpha} - \alpha}{\hat{S}_p \sqrt{\sigma^2_{m,p}}} \xrightarrow{d} N(0,1), \quad n \to \infty,$$
CHAPTER 2. SEMIPARAMETRIC INFERENCE ON THE FRAC TAL INDEX OF GAUSSIAN AND CONDITIONALLY GAUSSIAN TIME SERIES DATA

which will be the basis of the CLT used for comparison with our Monte Carlo approach in the simulation study of Section 2.5. Recall that the factor $\sigma^2_{m,p}$ is only feasible to calculate when $p = 2$ and it gets increasingly cumbersome to calculate as $m$ increases. We will therefore consider only the case with $p = 2$ and $m = 2$ when implementing the CLT in the simulation exercise below. The details were given in Remark 2.3.4.

The $\mathcal{BI}$ process introduced above is ideal for including stochastic volatility (SV) in the process. However, it is also possible to include SV in processes based on other Gaussian processes, such as the fBm. For instance, consider processes of the form

$$X_t = X_0 + \int_0^t \sigma_s dG_s, \quad t \geq 0,$$

(2.4.1)

where $\sigma$ is a SV process and $G$ is a fractal Gaussian process with centered and stationary increments, e.g. a Gaussian $\mathcal{BI}$ process, an fBm, a Matérn process, etc. The theory of Barndorff-Nielsen et al. (2009a), and therefore also the bootstrap methods of this paper, holds also for such processes.

**Extension to processes with non-stationary increments**

When the increments of $X$ are non-stationary an approach similar to what was done in Bennedsen et al. (2015) can be adopted as follows. Define the time-dependent variogram

$$\gamma_p(h, t) := \mathbb{E}[|X_{t+h} - X_t|^p], \quad h, t \in \mathbb{R},$$

and, analogously to (2.2.1), assume that

$$\gamma_p(h, t) \sim C_{p,t}|h|^{(2\alpha+1)p/2}L(h), \quad t > 0, \quad h \downarrow 0,$$

(2.4.2)

where again $C_{p,t} > 0$, $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, and $L$ is a slowly varying function at zero. The bootstrap methods considered in this paper also apply to such processes. An example is the truncated Brownian semistationary process.

**Example 2.4.2** (Truncated $\mathcal{BI}$ process, Bennedsen et al. (2015)). Let

$$X_t = X_0 + \int_0^t g(t-s)\sigma_s dW_s, \quad t \geq 0,$$

where $X_0 \in \mathbb{R}$, $W$ is a Brownian motion, and $\sigma$ a stochastic volatility process. Bennedsen et al. (2015) call such a process a truncated $\mathcal{BI}$ process. When $X$ satisfies (A1)–(A6), Bennedsen et al. (2015) show that $\alpha$ is indeed the fractal index of $X$, in the sense of $\gamma_p(h, t)$ satisfying (2.4.2).
2.5 Simulation study

This section presents the results of a Monte Carlo simulation study where the various approaches to inference considered above are compared. Tables 2.1 and 2.2 contain results of the finite sample properties of the size and power of the hypothesis test $H_0 : \alpha = \alpha_0$ against the double-sided alternative $H_1 : \alpha \neq \alpha_0$. A number of different DGPs are considered: three Gaussian DGPs satisfying (A1)–(A4) and three BS $\mathcal{P}$ processes satisfying (A1)–(A6). The Gaussian processes are the fBm, the Matérn and the powered exponential process. Details on the powered exponential process are given in Appendix A.3, where it is also shown that it indeed satisfies assumptions (A1)–(A3) and also (A4) when $\alpha > -1/4$. The three versions of the BS $\mathcal{P}$ process are all with the gamma kernel, as given in Example 3.3.2, but encompassing three different stochastic volatility regimes: “NoSV” refers to the process without stochastic volatility, while “SV1F” and “SV2F” refer to the BS $\mathcal{P}$ process with one- and two-factor stochastic volatility, respectively. Especially SV2F induces very strong kurtosis in the stochastic volatility process. Further details of all DGPs and SV processes considered in the Monte Carlo simulations are given in Appendix A.3.

When testing $H_0$, five different approaches are considered: the asymptotics-based inference of Theorem 2.4.1 (labeled CLT), un-studentized bootstrap inference ($S_p^*$ for bootstrap method 1 and 2, respectively), and studentized bootstrap inference ($T_p^*$ for bootstrap method 1 and 2, respectively). Table 2.1 contains size calculations, that is rejection rates of $H_0$ when $H_0$ is true. Table 2.2 contains power calculations of the test; we test $H_0 : \alpha = 0$ but now simulate under the alternative, with the true value of the fractal index $\alpha$ as indicated in the table. Here, the null hypothesis $H_0 : \alpha = 0$ is chosen because it seems to be the most empirically relevant, as it can be used to test for semimartingality, cf. Section 2.2.

The tables show that, generally, all methods perform well when $p = 2$ and $\alpha < 1/4$. For small sample sizes, the CLT has slight size distortions, while the bootstrap methods retain the nominal size. This fact was also found for a different estimator and CLT in Bennedsen et al. (2016a). When SV is present, however, especially the two-factor SV process SV2F, the performance of the un-studentized bootstraps ($S_p^*$ statistics) deteriorates. In contrast, both the CLT and the studentized bootstrap approach ($T_p^*$ statistics) work very well and we conclude that studentization is crucial for accurate bootstrap inference, at least when stochastic volatility/heteroskedasticity is present in the data.

We experimented with different values of the power parameter $p$; although the CLT is infeasible for $p \neq 2$, the bootstrap methods work well for all values of $p > 0$. An illustration of this in the case of $p = 1$ is shown in the tables. As discussed in Section 2.3, the fact that the bootstrap methods provide accurate inference for all $p > 0$ is highly desirable, as different values of $p$ can be used for estimation that is outlier-robust ($p = 1$ in particular, see also Section 3.2) and for checking the general robustness of an $\alpha$ estimate by considering several values for $p$. 
The case where $\alpha > 1/4$ is also investigated. In particular, the tables show results for $\alpha = 3/8 = 0.3750$, where the limiting distribution is now non-Gaussian. In this case, the CLT is not valid. Bootstrap method 1 works very well but Bootstrap method 2 suffers from size distortions; this is likely because this latter method simulates the auxiliary bootstrap fBms with Hurst index $H^* = \hat{\alpha} + 0.5$. When, by normal variation in the OLS estimator, $\hat{\alpha} \leq 1/4$ is estimated from the observations of $X$, the asymptotic distribution of the auxiliary bootstrap estimator $\hat{\alpha}^*_b$ will therefore be Gaussian and lead to wrong inference. We conclude, that one should be wary of using bootstrap method 2, when the process under study is expected to be very smooth, i.e. when $\alpha$ is large.

**Choosing the bandwidth**

It is an open question how to choose the bandwidth, $m$, optimally. For this reason we conduct a simulation experiment, shown in Figure 2.2, to shed some light on this. In (a) we plot the root mean squared error (RMSE) of the OLS estimator of $\alpha$ coming from (2.3.2), as a function of $m$. It seems that the RMSE is minimized for $m = 4$ or $m = 5$, depending on the number of observations, although for $m \in \{3, 4, 5\}$ there is not much difference in RMSE. Note, that $m = 2$ results in a significantly higher RMSE, especially for the smaller sample sizes. In (b) we plot the standard deviation of the random variable $Z_p$, i.e. $\sigma_{m,p}$, for $p = 2$. This is computed from $B = 999$ simulations of an fBm: we simulate an fBm, estimate $\alpha$ by the OLS regression (2.3.2) and compute the (numerical) standard deviation of $\hat{\alpha}$ over the $B = 999$ instances. This process is repeated $MC = 5000$ times and the numbers for $\sigma_{m,p}$ used in the plot is the average over these $MC$ values. The same conclusions hold for the standard deviation of $Z_p$ as for RMSE. What is more, the magnitude of $\sigma_{m,p}$ is almost exactly equal to the corresponding RMSE, indicating that the bias in the estimator is negligible. In (c) and (d) the finite sample properties of the hypothesis test $H_0 : \alpha = \alpha_0$ as a function of $m$ is explored. The power is maximized for $m \in \{3, 4, 5\}$ but somewhat lower for $m = 2$. Choosing $m \geq 3$ will result in size distortions for the test, as compared to $m = 2$, when the number of observations is small ($n = 40, 80$), but when there is a moderate number of observations, $n \geq 100$ say, choosing $m \in \{3, 4, 5\}$ will result in a hypothesis test with more accurate size, as compared to the case $m = 2$.

From these investigations we conclude that choosing a low value for the bandwidth $m$ is prudent. It seems, however, that there are gains from choosing $m \geq 3$. This is definitely the case when moderately many observations are available but likely true for a small number of observations as well, cf. Figure 2.2 (a). In Figure 2.2 the DGP underlying the simulations is a $\mathcal{BIS}$ process, but the conclusions hold for other DGPs as well. In particular, we confirmed the results using the fBm, the Matérn process, and the powered exponential process, for various values of $\alpha \in (-1/2, 1/2)$. This is slightly at odds with Davies and Hall (1999) where the authors found, also via simulations, that the mean squared error of $\hat{\alpha}$ was minimized for $m = 2$ when the
### Table 2.1: Rejection rates under $H_0$.

**DGP: Gaussian processes ($\alpha = -0.1250; p = 2$)**

<table>
<thead>
<tr>
<th>$n$</th>
<th>Bm</th>
<th>Matérn ($\beta = 1$)</th>
<th>Powered exp. ($c = 0.5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
</tr>
<tr>
<td>20</td>
<td>0.087</td>
<td>0.051</td>
<td>0.078</td>
</tr>
<tr>
<td>40</td>
<td>0.071</td>
<td>0.056</td>
<td>0.068</td>
</tr>
<tr>
<td>80</td>
<td>0.061</td>
<td>0.052</td>
<td>0.058</td>
</tr>
<tr>
<td>160</td>
<td>0.055</td>
<td>0.051</td>
<td>0.052</td>
</tr>
<tr>
<td>320</td>
<td>0.050</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td>640</td>
<td>0.049</td>
<td>0.048</td>
<td>0.048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>noSV</th>
<th>SV1F</th>
<th>SV2F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
</tr>
<tr>
<td>20</td>
<td>0.099</td>
<td>0.052</td>
<td>0.076</td>
</tr>
<tr>
<td>40</td>
<td>0.067</td>
<td>0.051</td>
<td>0.065</td>
</tr>
<tr>
<td>80</td>
<td>0.067</td>
<td>0.052</td>
<td>0.057</td>
</tr>
<tr>
<td>160</td>
<td>0.054</td>
<td>0.044</td>
<td>0.051</td>
</tr>
<tr>
<td>320</td>
<td>0.054</td>
<td>0.049</td>
<td>0.053</td>
</tr>
<tr>
<td>640</td>
<td>0.050</td>
<td>0.047</td>
<td>0.050</td>
</tr>
</tbody>
</table>

**DGP: Γ - ARFIMA ($\alpha = -0.1250, \lambda = 1; p = 2$)**

<table>
<thead>
<tr>
<th>$n$</th>
<th>noSV</th>
<th>SV1F</th>
<th>SV2F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
</tr>
<tr>
<td>20</td>
<td>0.100</td>
<td>0.044</td>
<td>0.073</td>
</tr>
<tr>
<td>40</td>
<td>0.078</td>
<td>0.053</td>
<td>0.066</td>
</tr>
<tr>
<td>80</td>
<td>0.067</td>
<td>0.049</td>
<td>0.057</td>
</tr>
<tr>
<td>160</td>
<td>0.058</td>
<td>0.049</td>
<td>0.053</td>
</tr>
<tr>
<td>320</td>
<td>0.052</td>
<td>0.049</td>
<td>0.053</td>
</tr>
<tr>
<td>640</td>
<td>0.052</td>
<td>0.049</td>
<td>0.053</td>
</tr>
</tbody>
</table>

**DGP: Γ - ARFIMA ($\alpha = 0.3750, \lambda = 1; p = 2$)**

<table>
<thead>
<tr>
<th>$n$</th>
<th>noSV</th>
<th>SV1F</th>
<th>SV2F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
</tr>
<tr>
<td>20</td>
<td>0.056</td>
<td>0.056</td>
<td>0.066</td>
</tr>
<tr>
<td>40</td>
<td>0.055</td>
<td>0.056</td>
<td>0.065</td>
</tr>
<tr>
<td>80</td>
<td>0.054</td>
<td>0.055</td>
<td>0.065</td>
</tr>
<tr>
<td>160</td>
<td>0.054</td>
<td>0.056</td>
<td>0.068</td>
</tr>
<tr>
<td>320</td>
<td>0.052</td>
<td>0.057</td>
<td>0.098</td>
</tr>
<tr>
<td>640</td>
<td>0.054</td>
<td>0.056</td>
<td>0.106</td>
</tr>
</tbody>
</table>

**DGP: Γ - ARFIMA ($\alpha = -0.1250, \lambda = 1; p = 1$)**

<table>
<thead>
<tr>
<th>$n$</th>
<th>noSV</th>
<th>SV1F</th>
<th>SV2F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
<td>$S_1^*$</td>
</tr>
<tr>
<td>20</td>
<td>0.056</td>
<td>0.058</td>
<td>0.065</td>
</tr>
<tr>
<td>40</td>
<td>0.053</td>
<td>0.054</td>
<td>0.075</td>
</tr>
<tr>
<td>80</td>
<td>0.051</td>
<td>0.052</td>
<td>0.066</td>
</tr>
<tr>
<td>160</td>
<td>0.048</td>
<td>0.048</td>
<td>0.056</td>
</tr>
<tr>
<td>320</td>
<td>0.052</td>
<td>0.051</td>
<td>0.055</td>
</tr>
<tr>
<td>640</td>
<td>0.048</td>
<td>0.048</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Finite sample size properties of the test $H_0: \alpha = a_0$ against $H_1: \alpha \neq a_0$ at a nominal size of 0.05. The underlying DGP is simulated under $H_0$, i.e. the numbers in the table are the ‘size’ of the hypothesis test. The true parameter values used for the DGP, as well as the value of the power index $p$ used for estimation, is given above the respective panels. For the CLT we choose the bandwidth $m = 2$, while the bootstrap methods set $m = 3$. 5000 Monte Carlo runs, each with $B = 999$ bootstrap replications, were conducted.
### Table 2.2. Rejection rates under $H_1$.  

<table>
<thead>
<tr>
<th>$n$</th>
<th>noSV</th>
<th>SV1F</th>
<th>SV2F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLT</td>
<td>$S_1^*$</td>
<td>$T_1^*$</td>
<td>$S_2^*$</td>
</tr>
<tr>
<td>20</td>
<td>0.209</td>
<td>0.084</td>
<td>0.118</td>
</tr>
<tr>
<td>40</td>
<td>0.264</td>
<td>0.163</td>
<td>0.185</td>
</tr>
<tr>
<td>80</td>
<td>0.365</td>
<td>0.306</td>
<td>0.322</td>
</tr>
<tr>
<td>160</td>
<td>0.584</td>
<td>0.583</td>
<td>0.590</td>
</tr>
<tr>
<td>320</td>
<td>0.850</td>
<td>0.886</td>
<td>0.885</td>
</tr>
<tr>
<td>640</td>
<td>0.989</td>
<td>0.996</td>
<td>0.995</td>
</tr>
</tbody>
</table>

Finite sample size properties of the test $H_0 : \alpha = \alpha_0$ against $H_1 : \alpha \neq \alpha_0$ at a nominal size of 0.05. The underlying DGP is simulated under the alternative, i.e. the numbers in the table are the ‘power’ of the hypothesis test. The true parameter values used for the DGP, as well as the value of the power index $p$ used for estimation, is given above the respective panels. For the CLT we choose the bandwidth $m = 2$, while the bootstrap methods set $m = 3$. 5000 Monte Carlo runs, each with $B = 999$ bootstrap replications, were conducted.
DGP is the powered exponential process.$^3$

We recommend choosing $m = 3$ or $m = 4$, which seems to offer a good tradeoff between bias and variance. Choosing $m = 2$ essentially amounts to estimating $\alpha$ by drawing a straight line between only two points, $\log \hat{\gamma}_p(1/n)$ and $\log \hat{\gamma}_p(2/n)$, when running the OLS regression in (2.3.2). While this is tempting from a bias viewpoint – the scaling relationship (2.2.1) is only assumed to hold for small lag values – it seems to introduce more variance by relying on just two points.$^4$ However, because the bias is very small, it makes sense to focus on reducing variance, which is done by using more points to draw the line, i.e. choosing $m \geq 3$.

### 2.6 Empirical application: testing the semimartingale assumption on financial time series

This section studies price data from the Trades and Quote (TAQ) Database; in total, 29 different U.S. stocks, which are identified by their ticker symbol in Table 2.3, are analyzed. The data is sampled at a daily frequency: the log-price is the closing price on each day, while the volatility measure is the daily Realized Kernel (RK, Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008), constructed from intra-day, high-frequency, returns. See Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009b) for implementation of the RK. The data runs from January 2, 1997, until December 31, 2013, excluding weekends and holidays. For some assets, some days have been discarded due to limited trading during the day. All in all, we end up with an average of 4166 daily observations per asset.

For all data series we estimate $\alpha$ using the OLS regression (2.3.2), calculate 95% confidence intervals using bootstrap method 2, and calculate $p$-values for the null hypothesis $H_0 : \alpha = 0$ using bootstrap method 1. For the bootstraps we use the (studentized) $T^*$ statistic. For the log-prices our alternative hypothesis is $H_1 : \alpha \neq 0$, while it for the log-volatility seems more relevant to choose $H_1 : \alpha < 0$, to assess whether volatility is rough as has been conjectured.

In Section 2.6, we will pick out a particular asset, JPM (JPMorgan Chase), and investigate it in more depth, focusing in particular on the role of the power parameter $p$.

---

$^3$The reason for the discrepancy between the recommendations regarding the choice of bandwidth is likely related to sample sizes. Although Davies and Hall (1999) do not report the sample size $n$ used in their simulation study, it is likely quite large, making the correlation between the variogram computed at adjacent lags very high. In this case there is little gain in including extra terms in the regression. In contrast, as illustrated in this section, when the number of observations is smaller, there might be gains in choosing $m$ slightly larger than 2. In other words, the recommendation of this paper, i.e. of choosing $m = 3$ or $m = 4$ as opposed to the conventional $m = 2$, is based on finite sample considerations.

$^4$The point about bias being increasing in $m$ was also made in Constantine and Hall (1994).
Chapter 2. Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data

Figure 2.2. Investigations of the impact of the bandwidth parameter \( m \). (a): Root mean squared error (RMSE) of the OLS estimator of \( \alpha \) from (2.3.2). (b): The standard deviation of \( Z_p, \sigma_{m,p} \), obtained by simulation of 999 replications of the fBm. (c): The size of the hypothesis test \( H_0: \alpha = -0.1250 \) against the double sided \( H_1: \alpha \neq -0.1250 \). (d): The power of the hypothesis test \( H_0: \alpha = 0 \) against the double sided \( H_1: \alpha \neq 0 \). The method for the tests is bootstrap procedure 1, using the \( T^* \) statistic with nominal size 0.05. We let \( \alpha = -0.1250, p = 2, B = 999 \), and perform 5000 Monte Carlo replications. DGP: NoSV \( \Gamma - \mathcal{B}\mathcal{S}\mathcal{S} \) with \( \lambda = 1 \) (see Appendix A.3).

Application to log-prices

The application of the bootstrap methods of this paper to log-prices of financial assets can be motivated by the following stochastic volatility model. Let the stock price be denoted by \( S = \{S_t\}_{t \geq 0} \), and suppose that its logarithm satisfies

\[
\log S_t = \log S_0 + \int_0^t \sigma_s dG_s + A_t, \quad t \geq 0, \tag{2.6.1}
\]

where \( \sigma = \{\sigma_t\}_{t \geq 0} \) is a stochastic volatility process, \( G = \{G_t\}_{t \geq 0} \) is a Gaussian process, such that the integral exists, and \( A = \{A_t\}_{t \geq 0} \) is a smooth drift process, e.g. \( A_t = -\frac{1}{2} \int_0^t \sigma_s^2 \, ds \). The most well-known instance of this model is of course the (Black-Scholes-type) case where \( G = B \) is a Brownian motion, and this model has been extensively studied and has proven useful in many areas of economics and finance (see Shephard, 2005, for a book-length treatment). Although the model (2.6.1) includes a drift term \( A \), this will not influence the estimation of \( \alpha \), as long as the process \( A \) is sufficiently smooth (see e.g. Lemma 3.5. of Corcuera et al., 2013).
Recall, that Proposition 2.2.2 showed that $\alpha \neq 0$ implies that the underlying process is not a semimartingale. Standard theory of asset prices states that log-prices which are not semimartingales will imply arbitrage opportunities, theoretically at least (Delbaen and Schachermayer, 1994). In what follows, we test the semimartingale null on log-prices using the approach outlined above.

The results are presented in Table 2.3. For all assets, the estimates of $\alpha$ are fairly close to 0, as one would expect under a no-arbitrage assumption. However, when $p = 2$ (left-most columns) there is huge variation in the estimator, as evidenced by the wide confidence intervals. Further, the average p-value is found to be quite large (0.81). The reason for this is most likely that the price data is not filtered for stock splits and mergers. Therefore, there may occasionally be large jumps in the quoted prices. This is illustrated for a particular stock in Section 2.6 (Figure 2.3) below, where we take a closer look at the JPM asset, but similar outliers are found in the time series of the other assets producing wide confidence intervals (not shown).

Outliers will bias estimates of $\alpha$ downwards and possibly invalidate any inference. Therefore, we also run the same exercise but with $p = 1$ (right-most columns of Table 2.3), which has been found to be robust to outliers in the data. In contrast to the former case, the confidence intervals are now fairly tight, indicating that the deteriorating effect of the stock splits and mergers are mitigated by choosing $p = 1$. What is more, when setting $p = 1$, estimates of $\alpha$ goes from being slightly negative to be scattered around zero, as seen in Table 2.3. Likewise, the average of the bootstrapped $p$-values is now 0.36, much closer to 0.50 which is the mean of the $p$-value if $H_0$ is true. This allows for much more precise inference, even in the presence of outliers, by choosing $p = 1$.\footnote{Of course, in practice a better approach would be to clean the data for stock splits and mergers before estimating $\alpha$, but as our goal is to illustrate the advantages of using $p = 1$, we find that the experiment works best when these events are included.}

Recall, that in this case there is no feasible asymptotic theory available, so a Monte Carlo approach, such as presented in this paper, is the only viable option to the best of our knowledge.

The test $H_0 : \alpha = 0$ is rejected for 4 out of 29 assets (14%): AIG, C, CAT, and XOM. For the first three, the estimate of the fractal index is positive, while for XOM it is negative. As mentioned in the introduction, under a self-similarity hypothesis a positive fractal index will imply long memory of the process under study (since the Hurst index is then given as $H = \alpha + 1/2$); hence the findings of this section could indicate that the price series of AIG, C, and CAT exhibit a high degree of memory. However, given that the actual estimates of $\alpha$ are quite close to zero also for these assets, we believe it is very likely that these rejections are simply false-positives.

**Application to log-volatility**

In this section, application of the bootstrap methods can be motivated from mathematical models of return volatility suggested previously in the finance/econometrics
Empirical investigations concerning the roughness properties of financial log-prices. Stocks are identified by their ticker symbol, with "Avg" being the mean over all assets. $\alpha$ is estimated with the OLS regression (2.3.2) with $m = 3$. The confidence intervals (CI) are computed using bootstrap method 2 with $B = 999$. The columns with "p-val" are simulated p-values using bootstrap method 1; bold face numbers denote rejections of $H_0: \alpha = 0$ at a 5% level. The alternative is $H_1: \alpha \neq 0$.

Table 2.3. Testing $H_0: \alpha = 0$ on log-prices.

<table>
<thead>
<tr>
<th>Asset</th>
<th>$n$</th>
<th>$\hat{\alpha}$</th>
<th>95% CI</th>
<th>p-val</th>
<th>$\hat{\alpha}$</th>
<th>95% CI</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>4277</td>
<td>0.00</td>
<td>(-0.14, 0.14)</td>
<td>1.00</td>
<td>0.03</td>
<td>(-0.00, 0.06)</td>
<td>0.07</td>
</tr>
<tr>
<td>AIG</td>
<td>4277</td>
<td>0.03</td>
<td>(-0.34, 0.42)</td>
<td>0.89</td>
<td>0.07</td>
<td>(0.02, 0.13)</td>
<td>0.00</td>
</tr>
<tr>
<td>AXP</td>
<td>4277</td>
<td>-0.03</td>
<td>(-0.25, 0.20)</td>
<td>0.78</td>
<td>-0.01</td>
<td>(-0.04, 0.02)</td>
<td>0.38</td>
</tr>
<tr>
<td>BA</td>
<td>4277</td>
<td>-0.01</td>
<td>(-0.16, 0.15)</td>
<td>0.90</td>
<td>0.02</td>
<td>(-0.01, 0.04)</td>
<td>0.32</td>
</tr>
<tr>
<td>BAC</td>
<td>4277</td>
<td>0.01</td>
<td>(-0.10, 0.12)</td>
<td>0.85</td>
<td>0.02</td>
<td>(-0.01, 0.05)</td>
<td>0.17</td>
</tr>
<tr>
<td>C</td>
<td>4263</td>
<td>0.03</td>
<td>(-0.38, 0.42)</td>
<td>0.91</td>
<td>0.05</td>
<td>(0.01, 0.10)</td>
<td>0.01</td>
</tr>
<tr>
<td>CAT</td>
<td>4277</td>
<td>-0.01</td>
<td>(-0.17, 0.16)</td>
<td>0.96</td>
<td>0.05</td>
<td>(0.02, 0.08)</td>
<td>0.00</td>
</tr>
<tr>
<td>CVX</td>
<td>3078</td>
<td>-0.06</td>
<td>(-0.32, 0.22)</td>
<td>0.69</td>
<td>-0.02</td>
<td>(-0.06, 0.01)</td>
<td>0.20</td>
</tr>
<tr>
<td>DD</td>
<td>4278</td>
<td>-0.02</td>
<td>(-0.20, 0.16)</td>
<td>0.78</td>
<td>0.01</td>
<td>(-0.02, 0.04)</td>
<td>0.47</td>
</tr>
<tr>
<td>DIS</td>
<td>4278</td>
<td>-0.03</td>
<td>(-0.32, 0.28)</td>
<td>0.83</td>
<td>0.01</td>
<td>(-0.03, 0.04)</td>
<td>0.69</td>
</tr>
<tr>
<td>GE</td>
<td>4277</td>
<td>-0.00</td>
<td>(-0.30, 0.28)</td>
<td>0.98</td>
<td>0.02</td>
<td>(-0.02, 0.05)</td>
<td>0.28</td>
</tr>
<tr>
<td>GM</td>
<td>3906</td>
<td>-0.01</td>
<td>(-0.56, 0.59)</td>
<td>0.98</td>
<td>0.05</td>
<td>(-0.01, 0.11)</td>
<td>0.14</td>
</tr>
<tr>
<td>HD</td>
<td>4277</td>
<td>0.01</td>
<td>(-0.12, 0.14)</td>
<td>0.95</td>
<td>0.03</td>
<td>(-0.00, 0.06)</td>
<td>0.07</td>
</tr>
<tr>
<td>IBM</td>
<td>4277</td>
<td>-0.03</td>
<td>(-0.26, 0.17)</td>
<td>0.79</td>
<td>0.01</td>
<td>(-0.02, 0.05)</td>
<td>0.42</td>
</tr>
<tr>
<td>INTC</td>
<td>4277</td>
<td>-0.01</td>
<td>(-0.17, 0.14)</td>
<td>0.84</td>
<td>-0.00</td>
<td>(-0.03, 0.03)</td>
<td>0.99</td>
</tr>
<tr>
<td>JNJ</td>
<td>4277</td>
<td>-0.02</td>
<td>(-0.29, 0.31)</td>
<td>0.90</td>
<td>-0.00</td>
<td>(-0.03, 0.03)</td>
<td>0.95</td>
</tr>
<tr>
<td>JPM</td>
<td>4277</td>
<td>-0.05</td>
<td>(-0.33, 0.25)</td>
<td>0.72</td>
<td>-0.02</td>
<td>(-0.05, 0.02)</td>
<td>0.32</td>
</tr>
<tr>
<td>KO</td>
<td>4277</td>
<td>-0.00</td>
<td>(-0.26, 0.26)</td>
<td>0.98</td>
<td>0.01</td>
<td>(-0.02, 0.04)</td>
<td>0.40</td>
</tr>
<tr>
<td>MCD</td>
<td>4277</td>
<td>-0.02</td>
<td>(-0.28, 0.28)</td>
<td>0.91</td>
<td>0.02</td>
<td>(-0.01, 0.05)</td>
<td>0.32</td>
</tr>
<tr>
<td>MMM</td>
<td>4277</td>
<td>-0.02</td>
<td>(-0.25, 0.24)</td>
<td>0.83</td>
<td>0.01</td>
<td>(-0.02, 0.04)</td>
<td>0.64</td>
</tr>
<tr>
<td>MRK</td>
<td>4277</td>
<td>-0.02</td>
<td>(-0.21, 0.19)</td>
<td>0.88</td>
<td>0.02</td>
<td>(-0.01, 0.05)</td>
<td>0.21</td>
</tr>
<tr>
<td>MSFT</td>
<td>4277</td>
<td>-0.02</td>
<td>(-0.20, 0.17)</td>
<td>0.83</td>
<td>0.01</td>
<td>(-0.02, 0.05)</td>
<td>0.38</td>
</tr>
<tr>
<td>PG</td>
<td>4277</td>
<td>-0.02</td>
<td>(-0.27, 0.26)</td>
<td>0.88</td>
<td>-0.01</td>
<td>(-0.05, 0.03)</td>
<td>0.55</td>
</tr>
<tr>
<td>SPY</td>
<td>4278</td>
<td>-0.06</td>
<td>(-0.09, -0.02)</td>
<td>0.00</td>
<td>-0.01</td>
<td>(-0.03, 0.02)</td>
<td>0.59</td>
</tr>
<tr>
<td>T</td>
<td>4270</td>
<td>-0.01</td>
<td>(-0.17, 0.14)</td>
<td>0.94</td>
<td>0.02</td>
<td>(-0.02, 0.05)</td>
<td>0.30</td>
</tr>
<tr>
<td>UTX</td>
<td>4277</td>
<td>-0.01</td>
<td>(-0.23, 0.21)</td>
<td>0.94</td>
<td>-0.01</td>
<td>(-0.04, 0.03)</td>
<td>0.70</td>
</tr>
<tr>
<td>VZ</td>
<td>3394</td>
<td>-0.03</td>
<td>(-0.07, 0.01)</td>
<td>0.14</td>
<td>-0.02</td>
<td>(-0.05, 0.00)</td>
<td>0.10</td>
</tr>
<tr>
<td>WMT</td>
<td>4277</td>
<td>-0.01</td>
<td>(-0.20, 0.17)</td>
<td>0.94</td>
<td>-0.01</td>
<td>(-0.03, 0.02)</td>
<td>0.72</td>
</tr>
<tr>
<td>XOM</td>
<td>3542</td>
<td>-0.07</td>
<td>(-0.31, 0.19)</td>
<td>0.58</td>
<td>-0.05</td>
<td>(-0.09, -0.02)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Avg 4166 -0.02 0.81 0.01 0.36

literature. In particular, consider the following simple stochastic volatility (Black-Scholes) model for the stock price $S$:

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad t \geq 0,$$  \hspace{1cm} (2.6.2)

where $\sigma = \{\sigma_t\}_{t \geq 0}$ is a stochastic volatility process and $B = \{B_t\}_{t \geq 0}$ a Brownian motion. The volatility process can be modeled as a log-fractal process, i.e.

$$\sigma_t = \exp(X_t), \quad t \geq 0,$$

where $X = \{X_t\}_{t \geq 0}$ is a stochastic process with fractal index $\alpha \in (-1/2, 1/2)$. Such a model was proposed in Comte and Renault (1996, 1998), where it was suggested that $X$ is a fractional Brownian motion with fractal index $\alpha > 0$ (to allow for long memory). Conversely, Gatheral et al. (2014) suggested to model $X$ by an fBm with $\alpha < 0$ (to allow for roughness). Most recently, Bennedsen et al. (2016b) have proposed to model $X$ by a BS S process with $\alpha < 0$, as this process can accommodate both roughness and long memory, in contrast to the fBm, which (by the self-similarity property) has either roughness or long memory but never both. Although these models have by now been much studied in the literature, so far little inference on $\alpha$ has been conducted. The present section therefore tests the semimartingale null for log-volatility by testing whether stochastic volatility is rough, i.e. running the hypothesis test $H_0 : \alpha = 0$ against the alternative $H_1 : \alpha < 0$.

Similar to the case of the log-prices, $\alpha < 0$ would imply, through Proposition 2.2.2, that volatility is a non-semimartingale. Contrary to the case of prices, however, such a fact would not imply arbitrage opportunities, since the log-prices themselves – even in the presence of non-semimartingale volatility – could still be semimartingales. This is for instance the case in the model (2.6.2) above.

The results are given in Table 2.4. We find $\hat{\alpha} \approx -0.33$, which is the overall mean of the estimates when $p = 2$. This is in line with what was found for similar time series in Gatheral et al. (2014) and Bennedsen et al. (2016b). The test $H_0 : \alpha = 0$ is rejected in favor of the alternative $H_1 : \alpha < 0$ for all series for $p = 2$ as well as for $p = 1$. Also, the confidence intervals are tight in both cases, increasing our belief in the robustness of the results. In light of the findings for the log-prices above, it might also provide tentative evidence that there are not any (large) jumps in volatility.

In summary, we conclude that we find very strong evidence in favor of the recently advanced hypothesis that stochastic volatility is rough (Gatheral et al., 2014).

A closer look at the JPM stock

Let us now take a closer look at the JPM (JPMorgan Chase) asset, which was a case where the confidence bands were particularly wide when $\alpha$ was estimates from the log-prices with $p = 2$, cf. Table 2.3. The top plot of Figure 2.3 shows the (closing) price evolution of JPM from January 2, 1997 to December 31, 2013.
The columns with "p-val" are simulated p-values using bootstrap method 1; bold face numbers denote rejections of $H_0: \alpha = 0$ at a 5% level. The alternative is $H_1: \alpha < 0$.

<table>
<thead>
<tr>
<th>Asset</th>
<th>$n$</th>
<th>$\hat{\alpha}$</th>
<th>95% CI</th>
<th>p-val</th>
<th>$\hat{\alpha}$</th>
<th>95% CI</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>4278</td>
<td>-0.36</td>
<td>(-0.39, -0.33)</td>
<td>0.00</td>
<td>-0.37</td>
<td>(-0.39, -0.34)</td>
<td>0.00</td>
</tr>
<tr>
<td>AIG</td>
<td>4277</td>
<td>-0.30</td>
<td>(-0.33, -0.27)</td>
<td>0.00</td>
<td>-0.31</td>
<td>(-0.33, -0.28)</td>
<td>0.00</td>
</tr>
<tr>
<td>AXP</td>
<td>4277</td>
<td>-0.33</td>
<td>(-0.35, -0.30)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.35, -0.30)</td>
<td>0.00</td>
</tr>
<tr>
<td>BA</td>
<td>4277</td>
<td>-0.33</td>
<td>(-0.36, -0.31)</td>
<td>0.00</td>
<td>-0.34</td>
<td>(-0.36, -0.31)</td>
<td>0.00</td>
</tr>
<tr>
<td>BAC</td>
<td>4277</td>
<td>-0.29</td>
<td>(-0.31, -0.26)</td>
<td>0.00</td>
<td>-0.27</td>
<td>(-0.30, -0.25)</td>
<td>0.00</td>
</tr>
<tr>
<td>C</td>
<td>4263</td>
<td>-0.31</td>
<td>(-0.34, -0.29)</td>
<td>0.00</td>
<td>-0.31</td>
<td>(-0.33, -0.28)</td>
<td>0.00</td>
</tr>
<tr>
<td>CAT</td>
<td>4277</td>
<td>-0.33</td>
<td>(-0.36, -0.31)</td>
<td>0.00</td>
<td>-0.32</td>
<td>(-0.34, -0.30)</td>
<td>0.00</td>
</tr>
<tr>
<td>CVX</td>
<td>3078</td>
<td>-0.30</td>
<td>(-0.33, -0.27)</td>
<td>0.00</td>
<td>-0.30</td>
<td>(-0.33, -0.27)</td>
<td>0.00</td>
</tr>
<tr>
<td>DD</td>
<td>4278</td>
<td>-0.34</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.36, -0.31)</td>
<td>0.00</td>
</tr>
<tr>
<td>DIS</td>
<td>4278</td>
<td>-0.34</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
<td>-0.35</td>
<td>(-0.38, -0.33)</td>
<td>0.00</td>
</tr>
<tr>
<td>GE</td>
<td>4277</td>
<td>-0.33</td>
<td>(-0.35, -0.31)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.35, -0.31)</td>
<td>0.00</td>
</tr>
<tr>
<td>GM</td>
<td>3906</td>
<td>-0.33</td>
<td>(-0.36, -0.31)</td>
<td>0.00</td>
<td>-0.34</td>
<td>(-0.36, -0.32)</td>
<td>0.00</td>
</tr>
<tr>
<td>HD</td>
<td>4277</td>
<td>-0.34</td>
<td>(-0.37, -0.31)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.35, -0.31)</td>
<td>0.00</td>
</tr>
<tr>
<td>IBM</td>
<td>4277</td>
<td>-0.33</td>
<td>(-0.35, -0.30)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.35, -0.30)</td>
<td>0.00</td>
</tr>
<tr>
<td>INTC</td>
<td>4277</td>
<td>-0.29</td>
<td>(-0.31, -0.26)</td>
<td>0.00</td>
<td>-0.29</td>
<td>(-0.31, -0.27)</td>
<td>0.00</td>
</tr>
<tr>
<td>JNJ</td>
<td>4277</td>
<td>-0.36</td>
<td>(-0.38, -0.34)</td>
<td>0.00</td>
<td>-0.35</td>
<td>(-0.38, -0.33)</td>
<td>0.00</td>
</tr>
<tr>
<td>JPM</td>
<td>4277</td>
<td>-0.31</td>
<td>(-0.34, -0.29)</td>
<td>0.00</td>
<td>-0.30</td>
<td>(-0.33, -0.28)</td>
<td>0.00</td>
</tr>
<tr>
<td>KO</td>
<td>4277</td>
<td>-0.35</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
<td>-0.35</td>
<td>(-0.37, -0.33)</td>
<td>0.00</td>
</tr>
<tr>
<td>MCD</td>
<td>4277</td>
<td>-0.36</td>
<td>(-0.38, -0.33)</td>
<td>0.00</td>
<td>-0.35</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
</tr>
<tr>
<td>MMM</td>
<td>4277</td>
<td>-0.34</td>
<td>(-0.37, -0.31)</td>
<td>0.00</td>
<td>-0.32</td>
<td>(-0.35, -0.30)</td>
<td>0.00</td>
</tr>
<tr>
<td>MRK</td>
<td>4277</td>
<td>-0.34</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
<td>-0.35</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
</tr>
<tr>
<td>MSFT</td>
<td>4277</td>
<td>-0.31</td>
<td>(-0.34, -0.29)</td>
<td>0.00</td>
<td>-0.31</td>
<td>(-0.33, -0.28)</td>
<td>0.00</td>
</tr>
<tr>
<td>PG</td>
<td>4277</td>
<td>-0.33</td>
<td>(-0.37, -0.30)</td>
<td>0.00</td>
<td>-0.34</td>
<td>(-0.36, -0.32)</td>
<td>0.00</td>
</tr>
<tr>
<td>SPY</td>
<td>4278</td>
<td>-0.30</td>
<td>(-0.32, -0.27)</td>
<td>0.00</td>
<td>-0.30</td>
<td>(-0.32, -0.27)</td>
<td>0.00</td>
</tr>
<tr>
<td>T</td>
<td>4270</td>
<td>-0.34</td>
<td>(-0.36, -0.31)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.35, -0.31)</td>
<td>0.00</td>
</tr>
<tr>
<td>UTX</td>
<td>4277</td>
<td>-0.35</td>
<td>(-0.38, -0.33)</td>
<td>0.00</td>
<td>-0.35</td>
<td>(-0.37, -0.32)</td>
<td>0.00</td>
</tr>
<tr>
<td>VZ</td>
<td>3394</td>
<td>-0.34</td>
<td>(-0.37, -0.31)</td>
<td>0.00</td>
<td>-0.33</td>
<td>(-0.35, -0.30)</td>
<td>0.00</td>
</tr>
<tr>
<td>WMT</td>
<td>4276</td>
<td>-0.38</td>
<td>(-0.40, -0.35)</td>
<td>0.00</td>
<td>-0.36</td>
<td>(-0.39, -0.34)</td>
<td>0.00</td>
</tr>
<tr>
<td>XOM</td>
<td>3542</td>
<td>-0.32</td>
<td>(-0.35, -0.29)</td>
<td>0.00</td>
<td>-0.32</td>
<td>(-0.34, -0.29)</td>
<td>0.00</td>
</tr>
<tr>
<td>Avg</td>
<td>4166</td>
<td>-0.33</td>
<td></td>
<td>0.00</td>
<td>-0.33</td>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2.4. Testing $H_0: \alpha = 0$ on log-volatility.

Empirical investigations concerning the roughness properties of financial log-volatility. Stocks are identified by their ticker symbol, with "Avg" being the mean over all assets. $\alpha$ is estimated with the OLS regression (2.3.2) with $m = 3$. The confidence intervals (CI) are computed using bootstrap method 2 with $B = 999$. The columns with "p-val" are simulated p-values using bootstrap method 1; bold face numbers denote rejections of $H_0: \alpha = 0$ at a 5% level. The alternative is $H_1: \alpha < 0$. 

Figure 2.3. Top: Stock price evolution of JPMorgan Chase (JPM) from January 2, 1997, until December 31, 2013. On December 31, 2000, J. P. Morgan & Co. merged with The Chase Manhattan Group. Bottom: Estimation of $\alpha$ in time by using a rolling window of 250 days and $m = 3$. Solid lines are estimates of $\alpha$; dashed lines are 95% (point-wise) confidence bands calculated using the $T^*$ statistic of bootstrap procedure 2 with $B = 999$ bootstrap replications.

In the period under study, JPM has undergone several stock splits, which are not visible in the plot, and one large merger which is clearly visible. Indeed, on December 31st, 2000, J. P. Morgan & Co. merged with The Chase Manhattan Group causing a major change in the quoted price of the stock, as seen in the figure.

The effect of the large outlier in the JPM data is further analyzed in the bottom plot of Figure 2.3, where $\alpha$ is estimated from log-prices through time by using a rolling window with 250 days. It is evident how the confidence bands (dashed lines) get wide when the large outlier enters the data. This is alleviated somewhat by letting $p = 1$, and, to an even greater extend, when $p = 1/2$, thus providing further evidence that letting $p \leq 1$ will result in inference, which is more robust to outliers than $p = 2$.

This conclusion is strengthened in Figure 2.4, which shows estimates of $\alpha$, as a function of $p$. In plot (a) the data is log-prices: it is clear how the values of $\hat{\alpha}$ is close to zero with tight confidence bands for $p \leq 1$, while the outlier will cause the estimate of $\alpha$ to be biased downwards, and the confidence bands to be extremely wide, for $p > 1$. Conversely, in plot (b) the data is log-volatility: here the estimates of $\alpha$ are stable, as a function of $p$, and the confidence bands are very tight, as we would expect from such a large sample ($n = 4277$).
Figure 2.4. Estimates of $\alpha$ (crosses) with associated 95% confidence bands (vertical lines) calculated using the $T^*$ statistic of bootstrap procedure 2 with $m = 3$ and $B = 999$ bootstrap replications. The data is from the JPMorgan (JPM) stock as explained in the text. (a): log-prices. (b): log-volatility.

2.7 Conclusion

This paper has developed simulation-based methods to conduct inference on the fractal index of a time series. We considered a particular semiparametric estimator of the fractal index, based on OLS regression, but the same methods can be applied straightforwardly to different estimators of the fractal index, as long as the (asymptotic) distribution of the estimator does not rely on the particular characteristics on the DGP of the observations, but only on the fractal index $\alpha$ and a possible heteroskedasticity factor.

The paper also answers a call for a general approach to inference on $\alpha$ for non-Gaussian processes (Gneiting et al., 2012). Our methods allow for this, at least when the non-Gaussianity is volatility induced, which is the case in many applications of interest in (e.g.) economics, econometrics, and finance. In the empirical section, we considered two instances of such applications by looking at time series of log-prices and log-volatility. We saw that $p \leq 1$ will result in estimation and inference, which is more robust to outliers, as compared to $p > 1$. Semimartingale hypotheses were tested for the time series. Although no evidence of non-semimartingality was found in log-prices, strong evidence of non-semimartingality was found in log-volatility, thereby supporting the claim that volatility is rough (Gatheral et al., 2014).

Acknowledgements

I would like to thank Professor Asger Lunde and Dr. Mikko S. Pakkanen for insightful discussions relating to fractal processes and Solveig Sørensen for competent and precise proof reading of the manuscript. The research has been supported by CRE-
ATES (DNRF78), funded by the Danish National Research Foundation, by Aarhus University Research Foundation (project “Stochastic and Econometric Analysis of Commodity Markets”), and by Aage and Ylva Nimbs Foundation.

2.8 References


Matérn, B., 1960. Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. Meddelanden från Statens Skogsforskningsinstitut 49 (5).


Appendix

A.1 Proofs

Proof of Proposition 2.2.1. Note first, that since \(X\) is Gaussian, assumption (A1) implies that for \(n \geq 1\),
\[
E[|X_t - X_s|^{2n}] = C_2 n |t - s|^{(2\alpha + 1)n} L(t - s)^{2n}, \quad t, s \in \mathbb{R},
\]
where \(x \mapsto L(x)\) is slowly varying at zero. Let now \(K \subset (0, \infty)\) be a compact set and consider \(t, s \in K\). By the properties of slowly varying functions (Bingham et al., 1989, Theorem 1.5.6(ii)), for all \(\epsilon > 0\) we can find \(a > 0\) such that
\[
E[|X_t - X_s|^{2n}] \leq \tilde{C}_{1,n} |t - s|^{1 + (2\alpha + 1)n - 1 - 2n\epsilon}, \quad t - s \in (0, a],
\]
for a constant \(\tilde{C}_{1,n} > 0\). Conversely, since \(L\) is continuous on \((0, \infty)\), we also have that
\[
E[|X_t - X_s|^{2n}] \leq \tilde{C}_{2,n} |t - s|^{1 + (2\alpha + 1)n - 1 - 2n\epsilon}, \quad t - s > a,
\]
for a constant \(\tilde{C}_{2,n} > 0\). Putting these two observations together, we have that there exists a constant \(\tilde{C}_{3,n} > 0\) such that for all \(\epsilon > 0\) we have
\[
E[|X_t - X_s|^{2n}] \leq \tilde{C}_{3,n} |t - s|^{1 + (2\alpha + 1)n - 1 - 2n\epsilon}, \quad t, s \in K.
\]
Using this, we deduce that for \(n\) sufficiently large, the continuity criterion of Kolmogorov shows that \(X\) has a modification which is locally Hölder continuous of order \(\phi\) for all \(\phi \in \left(0, \frac{2\alpha + 1}{2} - 1 - 2n\epsilon \right) = \left(0, \alpha + 1/2 - \frac{1}{2n} - \epsilon \right)\). Letting \(n \uparrow \infty, \epsilon \downarrow 0\), yields the desired result. \(\square\)

Proof of Proposition 2.2.2. Recall the \(p\)-variation
\[
V_p(0, T) := \sup_{\Pi} \sum_{j=1}^{n} |X_{t_j} - X_{t_{j-1}}|^p,
\]
where the supremum is taken over all partitions with mesh going to zero. Consider the associated index of \(p\)-variation:
\[
I(X, [0, T]) := \inf\{p > 0 | V_p(0, T) < \infty\}.
\]
It is a well-known fact that continuous semimartingales have \(I(X, [0, T]) \in [0, 1] \cup \{2\}\) with probability one. Likewise, it is known that if a process, \(X\), has paths which are \(\phi\)-Hölder continuous, then the index of \(p\)-variation is \(\frac{1}{\phi}\) (e.g. Norvaiša, 2006). Now, by Proposition 2.2.1 we get that
\[
p^* := I(X, [0, T]) = \frac{1}{\alpha + 1/2} \in (1, \infty) \setminus \{2\}.
\]
In conclusion, \(X\) is therefore not a semimartingale. Note that \(\alpha = 0 \Rightarrow p^* = 2\), so that the present result does not hold for \(\alpha = 0\). \(\square\)
Proof of Proposition 2.3.1. Note first, that since $\alpha = a/p - 1/2$ we can write

$$\hat{\alpha} - \alpha = \frac{1}{px_m^T x_m} x_m^T (U^m + \epsilon^m),$$

where

$$U^m := (U_{1/n}, U_{2/n}, \ldots, U_{m/n})^T = \left(\log \left(\frac{\hat{\gamma}_p(1/n)}{\gamma_p(1/n)}\right), \log \left(\frac{\hat{\gamma}_p(2/n)}{\gamma_p(2/n)}\right), \ldots, \log \left(\frac{\hat{\gamma}_p(m/n)}{\gamma_p(m/n)}\right)\right)^T,$$

and

$$\epsilon^m := (\epsilon_{1/n}, \epsilon_{2/n}, \ldots, \epsilon_{m/n})^T = \left(\log L_p(1/n), \log L_p(2/n), \ldots, \log L_p(m/n)\right)^T.$$

To see that the term $x_m^T \epsilon^m$ vanishes as $n \to \infty$, note that

$$\sum_{k=1}^m x_{m,k} = \sum_{k=1}^m \left(\log k - \log m\right) = 0$$

and therefore

$$x_m^T \epsilon^m = \sum_{k=1}^m x_{m,k} \log L_p(k/n) = \sum_{k=1}^m x_{m,k} \log \left(\frac{L_p(k/n)}{L_p(1/n)}\right) \to 0, \quad n \to \infty,$$

since $\lim_{n \to \infty} \frac{L_p(k/n)}{L_p(1/n)} = 1$ by the property of slowly varying functions.

The required result now follows by noting that

$$\frac{\hat{\gamma}_p(k/n)}{\gamma_p(k/n)} = \frac{\hat{\gamma}_p(k/n)}{\gamma_p(k/n)} \to 1, \quad n \to \infty, \quad k \geq 1,$$

by Proposition 1 in Barndorff-Nielsen et al. (2009a).

Proof of Theorem 2.3.1. Using Theorem 2 in Barndorff-Nielsen et al. (2011), see also Barndorff-Nielsen et al. (2009a) Theorem 7, and the limit in equation (2.3.5), this paper, we get

$$\sqrt{n} \begin{pmatrix} \frac{\hat{\gamma}_p(1/n)}{\gamma_p(1/n)} - 1 \\ \vdots \\ \frac{\hat{\gamma}_p(m/n)}{\gamma_p(m/n)} - 1 \end{pmatrix} \to N(0, \Lambda_p), \quad n \to \infty,$$

where $\Lambda_p = \{\lambda_{p}^{k,v}\}_{k,v=1}^m$ is a $m \times m$ matrix with entries

$$\lambda_{p}^{k,v} = \lim_{n \to \infty} n \cdot Cov \left( \frac{\hat{\gamma}_p(k/n; B^H)}{\gamma_p(k/n; B^H)}, \frac{\hat{\gamma}_p(v/n; B^H)}{\gamma_p(v/n; B^H)} \right), \quad k, v = 1, 2, \ldots, m, \quad (A.3)$$
with $\gamma_p(\cdot; B^H)$ denoting the $p$'th order variogram for a fractional Brownian motion with Hurst index $H = \alpha + 1/2$, and similarly for $\hat{\gamma}_p$. Note that the limit in (A.3) exists for $k, \nu = 1, 2, \ldots, m$, by Breuer and Major (1983), Theorem 1, see also Corcuera et al. (2013), Remark 3.3.

As we will show below, assumption (A4) implies that

$$\sqrt{n}x_m^T e^m \rightarrow 0, \quad n \rightarrow \infty, \quad (A.4)$$

which means that from (A.1) we get, using (A.2) and the delta method,

$$\sqrt{n}(\hat{\alpha} - \alpha) \overset{d}{\rightarrow} N\left(0, \frac{x_m^T \Lambda_p x_m}{(x_m^T x_m)^2} \right), \quad n \rightarrow \infty,$$

which is what we wanted to show. To see that (A.4) holds, use the rule of l'Hôpital, and the properties of slowly varying functions, to conclude

$$\lim_{n \rightarrow \infty} \sqrt{n} \log \left( \frac{L_p(k/n)}{L_p(1/n)} \right) = \lim_{n \rightarrow \infty} 2n^{-1/2} \left( \frac{L'_p(k/n)}{L_p(k/n)} k - \frac{L'_p(1/n)}{L_p(1/n)} \right) = 0,$$

by assumption (A4). This concludes the proof.

**Proof of Proposition 2.3.2.** Note that we can write

$$\hat{S}_p = \sqrt{\frac{m^{-1}_2 \hat{\gamma}_p(1/n) / \gamma_2(1/n)^{2p}}{m^{-1}_p \hat{\gamma}_p(1/n) / \gamma_2(1/n)^{p/2}}}.$$

The result now follows from Proposition 1 in Barndorff-Nielsen et al. (2009a).

**Proof of Theorem 2.4.1.** Note first, that by Theorem 4 of Barndorff-Nielsen et al. (2011), see also Theorem 3.2. and Remark 3.4. of Corcuera et al. (2013), we get

$$\sqrt{n} \left( \begin{array}{c} \hat{\gamma}_p(1/n) / \gamma_p(1/n) - 1 \\ \vdots \\ \hat{\gamma}_p(m/n) / \gamma_p(m/n) - 1 \end{array} \right) \overset{st}{\rightarrow} \int_0^1 \sigma_s^p \Lambda_p dB_s,$$

where $B$ is an $m$-dimensional Brownian motion, defined on an extension of the original probability space, $\Omega, \mathcal{F}, \mathbb{P}$, independent of $\mathcal{F}$. The matrix $\Lambda_p$ is identical to the one of Theorem 2.3.1, i.e. the covariances in $\Lambda_p$ is calculated using the covariance structure of the fBm, not the $\mathcal{B}\mathcal{F}\mathcal{F}$ process.

We proceed as in the proof of Theorem 2.3.1. In particular, invoking the delta method we get

$$\sqrt{n}(\hat{\alpha} - \alpha) \overset{st}{\rightarrow} \frac{x_m^T \Lambda_p}{x_m^T x_m p} \int_0^1 \sigma_s^p dB_s,$$
We illustrate how to derive the result using the case \( k \Lambda_2 \). 

\[
\sqrt{n}(\hat{\alpha} - \alpha) \overset{s_t}{\rightarrow} Z_p \cdot S_p, \quad S_p := \frac{\int_0^1 \sigma_s^{2p} \, ds}{\int_0^1 \sigma_s^p \, ds},
\]

and where \( Z_p \) is as in Theorem 2.3.1. This concludes the proof. \( \square \)

### A.2 Deriving an expression for the \( \Lambda \) matrix of Theorem 2.3.1

Below we give expressions for the entries in the matrix \( \Lambda_p = \{\lambda_{p, k, v}\}_{k, v=1}^2 \) of Theorem 2.3.1 when \( p = 2 \). Let \( B^H \) be an fBm with Hurst index \( H \in (0, 1) \). Recall that

\[
\gamma_2(k; n; B^H) := \mathbb{E} \left[ \left| B^H_{k/n} - B^H_0 \right|^2 \right] = (k/n)^{2H}.
\]

Therefore,

\[
\lambda_{2, k, v} = \lim_{n \to \infty} n^{1+4H} (kv)^{-2H} \text{Cov} \left( \hat{\gamma}_2(k; n; B^H), \hat{\gamma}_2(v; n; B^H) \right), \quad k, v = 1, 2.
\]

We illustrate how to derive the result using the case \( k = 1, v = 2 \) as example. Since \( B^H \) is Gaussian, we can use Isserlis’ theorem to calculate (cross) moments of the increments of \( X \). Brute force calculations yield

\[
\text{Cov} \left( \hat{\gamma}_2(1; n; B^H), \hat{\gamma}_2(2; n; B^H) \right) = \frac{1}{(n-1)(n-2)} \text{Cov} \left( \sum_{i=1}^{n-1} |X_{(i+1)/n} - X_{i/n}|^2, \sum_{j=1}^{n-1} |X_{(j+2)/n} - X_{j/n}|^2 \right)
\]

\[
= \frac{2^{-1}n^{-4H}}{(n-1)(n-2)} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \left( |j-i+2|^{2H} - |j-i+1|^{2H} - |j-i|^{2H} + |j-i-1|^{2H} \right)^2,
\]

where it was used that

\[
\mathbb{E}[B_{i/n}^H B_{j/n}^H] = \frac{1}{2} n^{-2H} \left( |i|^{2H} + |j|^{2H} - |i - j|^{2H} \right),
\]

since \( B^H \) is an fBm with Hurst index \( H \). Now, deduce that

\[
\lambda_{2, 1, 2} = 2^{-2H-1} \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n-2} \sum_{j=1}^{n-2} \left( |j-i+2|^{2H} - |j-i+1|^{2H} - |j-i|^{2H} + |j-i-1|^{2H} \right)^2.
\]

Similarly, it can be shown that

\[
\lambda_{2, k, k} = 2^{-1} k^{-4H} \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n-k} \sum_{j=1}^{n-k} \left( |j-i+k|^{2H} - 2|j-i|^{2H} + |j-i-k|^{2H} \right)^2, \quad k = 1, 2.
\]

These expressions are convergent by Theorem 1 in Breuer and Major (1983), see also Remark 3.3. in Corcuera et al. (2013), but in our implementation we simply use the finite sample version by implementing the finite sums as given above.
A.3 Simulation setup

Section 2.5 requires simulating $n \in \mathbb{N}$ equidistant observations of the process $X$ on the unit interval $[0, 1]$. The following subsections give details concerning the various DGPs we consider for $X$. Simulation of the Gaussian DGPs can be done exactly using the Cholesky decomposition of the covariance matrix; since the processes are stationary (in case of the fBm the increments are stationary) one can get this covariance matrix from the correlation function $\rho$. When $X$ contains stochastic volatility, the process is non-Gaussian and the Cholesky method is not valid. In this case, we use the hybrid scheme of Bennedsen et al. (2015) which is a fast and accurate simulation scheme for certain fractal processes with or without SV.

Brownian semistationary process

The Brownian semistationary ($BIFS$) process was introduced in Barndorff-Nielsen and Schmiegel (2007, 2009) and has been widely studied (e.g. Pakkanen, 2011; Barndorff-Nielsen et al., 2011, 2013b; Corcuera et al., 2013; Bennedsen et al., 2015) and applied to a number of empirical applications (e.g. Veraart and Veraart, 2014; Bennedsen, 2015).

The driftless $BIFS$ process is defined as

$$X_t = \int_{-\infty}^{t} g(t-s)\sigma_s dW_s, \quad t \geq 0,$$

where $g$ is a square integrable kernel function, $\sigma$ is a stochastic volatility process which is possibly correlated with the driving Brownian motion $W$. The next section will present the different specifications we consider for $\sigma$.

For simulation of the volatility modulated $BIFS$ process we utilize the hybrid scheme of Bennedsen et al. (2015).

Stochastic volatility regimes

For the stochastic volatility process $\sigma = \{\sigma_t\}_{t \in \mathbb{R}}$, we consider three different specifications: (i) constant volatility, labeled NoSV; (ii) one-factor stochastic volatility, labeled SV1F; and (iii) two-factor stochastic volatility, labeled SV2F. For the NoSV model we take for $t \in \mathbb{R}$,

$$\sigma_t = 1,$$

while in the SV1F model take, following Barndorff-Nielsen et al. (2008),

$$\sigma_t = \exp(\beta_0 + \beta_1 \tau_t),$$

$$d\tau_t = \xi \tau_t dt + dB_t,$$

$$\mathbb{E}[dW_t dB_t] = \rho dt,$$
Chapter 2. Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data

where $B$ is a standard Brownian motion and $\beta_1 = 0.125$, $\xi = -0.025$, $\beta_0 = \frac{\beta_1^2}{\xi} = -0.3125$ and $\rho = -0.3$. Lastly, for the SV2F model we take, following Huang and Tauchen (2005) and Barndorff-Nielsen et al. (2008),

$$
\sigma_t = s \cdot \exp(\beta_0 + \beta_1 \tau_{1t} + \beta_2 \tau_{2t}),
$$

$$
d\tau_{1t} = \xi_1 \tau_{1t} dt + dB_{1t},
$$

$$
d\tau_{2t} = \xi_2 \tau_{2t} dt + (1 + \phi \tau_{2t}) dB_{2t},
$$

$$
E[dW_t dB_{1t}] = \rho_1 dt,
$$

$$
E[dW_t dB_{2t}] = \rho_2 dt,
$$

where $B_1$, $B_2$ are standard Brownian motions and the function $s \cdot \exp$ is given by

$$
s \cdot \exp(x) = \begin{cases} 
\exp(x), & x \leq \log(1.5), \\
\frac{3}{2} \sqrt{1 - \log(1.5)} + x^2 / \log(1.5), & x > \log(1.5), 
\end{cases}
$$

and the parameters are set to $(\beta_0, \beta_1, \beta_2)^T = (-1.20, 0.040, 1.50)^T$, $\phi = 0.250$, $\rho_1 = \rho_2 = -0.30$, and $(\xi_1, \xi_2)^T = (-0.00137, -1.386)^T$.

We note that in the NoSV case the process $X$ is Gaussian and can thus be simulated exactly using a Cholesky decomposition of its variance-covariance matrix, which is what we do in our simulations. The stochastic processes of SV1F and SV2F can be simulated exactly using methods in Glasserman (2003), see also the Simulation Appendix to Barndorff-Nielsen et al. (2008).

**Fractional Brownian motion**

The fractional Brownian motion (fBm, Mandelbrot and Van Ness, 1968) is the most well-known fractal process. It is the zero-mean Gaussian process $B^H$, starting at zero, with covariance function

$$
\text{Cov}(B^H_t, B^H_s) = \frac{1}{2} \left( |t|^{2H} + |s|^{2H} - |t - s|^{2H} \right), \quad t, s \geq 0,
$$

where $H \in (0, 1)$ is the *Hurst index*. The fBm is self-similar, in the sense that $B^H(at) \overset{d}{=} |a|^H B^H(t)$, where “$\overset{d}{=}”$ means equality in distribution. As mentioned in the introduction, this self-similarity implies a deterministic one-to-one relationship between the Hurst index $H$ and the fractal index $\alpha$ (Gneiting and Schlather, 2004). Indeed, we have $H = \alpha + \frac{1}{2}$. What this means is that the small-scale behavior (as governed by $\alpha$) is determined by the large-scale behavior (as governed by $H$) and vice versa: when $H < 1/2$ then $B^H$ has rough paths (i.e. $\alpha < 0$) and short memory, while $H > 1/2$ implies that $B^H$ has smooth paths (i.e. $\alpha > 0$) and long memory (in the sense of a non-integrable autocorrelation function). When $H = 1/2$ $B^H$ is a Brownian motion.

The increments of fBm are stationary, and the increment process $X^H_t := B^H_{t+1} - B^H_t$ is called *fractional Gaussian noise* (fGn). The correlation function of the fGn was given in (2.3.5).
**Matérn process**

The Matérn process (Matérn, 1960; Handcock and Stein, 1993) $X$ is the zero-mean, unit variance, stationary Gaussian process with correlation function

$$\rho(h) = \frac{2^{-a+1/2}}{\Gamma(a-1/2)} (\beta h)^{a+1/2} K_{a+1/2} (\beta h), \quad h \in \mathbb{R},$$

with $a \in (-1/2, 1/2)$, $\beta > 0$, and where $K_v$ is the (modified) Bessel function of the third kind with index $v$.

Using Barndorff-Nielsen (2012) and properties of Bessel functions (e.g., Gradshteyn and Ryzhik, 2007, sections 8.4 and 8.5) one can show that the Matérn process fulfills assumptions (A1)–(A3). Also, by the theorem on page 8 in Barndorff-Nielsen (2012) one can deduce that assumption (A4) is satisfied when $a \in (-1/2, 1/4)$. The calculations are somewhat tedious, so we instead illustrate the procedure on the powered exponential process below.

**Powered exponential process**

The powered exponential process $X$ is the zero-mean, unit variance, stationary Gaussian process with correlation function

$$\rho(h) = \exp\left(-|ch|^{2\alpha+1}\right), \quad h \in \mathbb{R},$$

with $\alpha \in (-1/2, 1/2)$ and $c > 0$. Let us verify that the powered exponential process satisfies assumptions (A1)–(A3). To check (A1), let $x \geq 0$ and write

$$\gamma_2(x) = 2\left(1 - \rho(x)\right) = x^{2\alpha+1} L(x),$$

where

$$L(x) = 2 \frac{1 - \rho(x)}{x^{2\alpha+1}} = 2 \frac{1 - \exp\left(-|cx|^{2\alpha+1}\right)}{x^{2\alpha+1}}. \quad (A.1)$$

Using the rule of L’Hôpital it is easy to show that

$$\lim_{x \to 0} L(x) = 2c^{2\alpha+1} > 0,$$

implying that $L$ is slowly varying at zero. To see that (A2) holds, note that we can write

$$\frac{d^2}{dx^2} \gamma_2(x) = x^{2\alpha-1} L_2(x),$$

where

$$L_2(x) = 2c^{2\alpha+1} (2\alpha + 1) \rho(x) \left[2\alpha - c^{2\alpha+1} (2\alpha + 1) x^{2\alpha+1}\right].$$
Again, it is easy to see that
\[
\lim_{x \to 0} L_2(x) = c^{2\alpha+1}4\alpha(2\alpha + 1) \in \mathbb{R},
\]
so that also \( L_2 \) is slowly varying at zero. Lastly, note that
\[
L'_2(x) := \frac{d}{dx} L_2(x) = -2c^{4\alpha+2}(2\alpha + 1)^2 \rho(x)x^{2\alpha} \left[ 4\alpha + 1 - c^{2\alpha+1}(2\alpha + 1)x^{2\alpha+1} \right],
\]
so that, around a neighborhood of \( x = 0 \) we have \( L'_2(x) < 0 \) if \( \alpha > -1/4 \) and \( L'_2(x) > 0 \) if \( \alpha \leq -1/4 \). In other words, in a neighborhood of zero, \( L_2 \) is either decreasing or increasing. In both cases we can therefore conclude that for all \( b \in (0, 1) \),
\[
\sup_{y \in [x, xb]} \left| \frac{L_2(y)}{L_2(x)} \right| \to |4\alpha(2\alpha + 1) < \infty, \quad x \to 0,
\]
which shows that (A3) is fulfilled as well. Finally, by expanding the exponential function in (A.1) it is not hard to deduce that assumption (A4) holds when \( \alpha > -1/4 \), but is violated otherwise.
Abstract

We study the empirical properties of realized volatility of the E-mini S&P 500 futures contract at various time scales, ranging from a few minutes to one day. Our main finding is that intraday volatility is remarkably rough and persistent. What is more, by further studying daily realized volatility measures of close to five thousand individual US equities, we find that both roughness and persistence appear to be universal properties of volatility. Inspired by the empirical findings, we introduce a new class of continuous-time stochastic volatility models, capable of decoupling roughness (short-term behavior) from long memory and persistence (long-term behavior) in a simple and parsimonious way, which allows us to successfully model volatility at all intraday time scales. Our prime model is based on the so-called Brownian semistationary process and we derive a number of theoretical properties of this process, relevant to volatility modeling. As an illustration of the usefulness our new
models, we conduct an extensive forecasting study; we find that the models proposed in this paper outperform a wide array of benchmarks considerably, indicating that it pays off to exploit both roughness and persistence in volatility forecasting.

**Keywords:** Stochastic volatility; high-frequency data; rough volatility; persistence; long memory; forecasting; Brownian semistationary process.

**JEL Classification:** C22, C51, C53, C58

**MSC 2010 Classification:** 60G10, 60G15, 60G17, 60G22, 62M09, 62M10

### 3.1 Introduction

Intraday modeling of asset return volatility is of importance in several applications, including derivatives pricing, high-frequency trading, and risk management (Andersen, Bollerslev, and Cai, 2000; Rossi and Fantazzini, 2015). The time scale at which volatility should be assessed depends on the intended application. On the one hand, in the context of high-frequency trading, say, the relevant time scale can be very short; a few minutes or even less. On the other hand, if the objective is to forecast daily volatility, models using moderate frequencies, such as one hour, outperform alternative models using very high frequencies (e.g., 5 minutes) or very low frequencies (e.g., one day), as shown in Andersen, Bollerslev, and Lange (1999).

The goal of this paper is to introduce a class of continuous-time models of volatility that are consistent with empirical features of realized volatility at all time scales. As we will see, this requires a model incorporating both roughness (irregular behavior at short time scales) and persistence (strong dependence at longer time scales). While the latter property is well-established in the literature on volatility (e.g., Bollerslev and Wright, 2000; Andersen, Bollerslev, Diebold, and Labys, 2003), the importance of the former has been highlighted only very recently, most notably in Gatheral et al. (2014) and Bayer et al. (2016).

Our paper has two main contributions. The first is an in-depth empirical study of the time series behavior of realized volatility at a wide range of time scales, ranging from a few minutes to one day. The conclusion of this study in a nutshell is that realized volatility is rough, persistent, and non-Gaussian.

The second contribution of the paper is to put forward a class of stochastic models that are able to capture these three key features of volatility. In particular, we advocate the use of Brownian semistationary (BSS) processes (Barndorff-Nielsen and Schmiegel, 2007, 2009) as models of logarithmic volatility. These processes are flexible in the sense that they allow for decoupling of the fine properties (roughness) from the long-term behavior (memory/persistence). As we shall see, under suitable specifications, BSS processes can accommodate both bona fide long memory (i.e., non-integrable autocorrelations) and short memory (exponentially decaying
Introduction

We derive some general theoretical results concerning the memory properties of these processes in this paper. Under rather general conditions, \( \mathcal{B}_{\alpha} \) processes are stationary, and they allow for easy inclusion of non-Gaussianity and the leverage effect. Moreover, fast and efficient simulation schemes are available (Bennedsen et al., 2015).

Recently, there has been considerable interest in rough models of volatility. This is due to both theoretical developments in implied volatility modeling (Fukasawa, 2016) as well as empirical evidence based on realized volatility (Gatheral et al., 2014; Bennedsen, 2016). An important contribution to this literature is the model suggested by Gatheral et al. (2014), which is inspired by the fractional stochastic volatility (FSV) model of Comte and Renault (1996). Both of these models are based on a fractional Ornstein–Uhlenbeck process, driven by a fractional Brownian motion (fBm). While the FSV model of Comte and Renault (1996) uses an fBm with Hurst index \( H > 1/2 \), giving rise to long memory, Gatheral et al. (2014), by contrast, switch to an fBm with \( H < 1/2 \) to allow for roughness. They therefore term their model the rough fractional stochastic volatility (RFSV) model. Bayer et al. (2016), motivated by option pricing and implied volatility smile modeling, introduce another rough volatility model, the rough Bergomi (rBergomi) model, where logarithmic volatility is modeled by a (non-stationary) Riemann–Liouville type Gaussian process, which is often referred to as a “Type II” fractional Brownian motion in the literature on time series models with long memory (Marinucci and Robinson, 1999).

The RFSV model has the limitation that it does not capture the long memory property of volatility, which is often regarded as a stylized fact (e.g., Andersen et al., 2003), while the rBergomi model, being non-stationary, would not exhibit realistic long-term behavior of volatility. When designing a realistic model of volatility that allows for both roughness and flexible long-term properties, it is important to be aware of the principle, pointed out by Gneiting and Schlather (2004), that self-similar processes, such as the fBm, have either long memory or rough sample paths. In contrast, the models presented in this work conveniently decouples the behavior of volatility at short and long time scales; in particular, they accommodate both roughness and long memory. In spite of their generality, the suggested models are simple, in terms of their mathematical structure, and parsimonious, relying on only two parameters controlling short- and long-term behavior, respectively. Moreover, the models are easy to estimate, simulate, and forecast.

In a forecasting study, we find that our proposed models outperform a wide array of benchmark models, especially at intraday time scales. This indicates that it is important to carefully model both small- and large-scale behavior of stochastic volatility in forecasting.

The rest of the paper is structured as follows. In Section 3.2 we introduce the E-mini S&P 500 data set and present our first empirical results, demonstrating that volatility is rough, very persistent, and non-Gaussian at intraday time scales. Section
3.3 introduces a class of stochastic volatility models, based on \( \mathcal{BSF} \) processes, that are able to parsimoniously capture these empirical findings. This section also reviews simulation methods for rough stochastic volatility models. Section 3.4 presents estimation results of the various models we consider, using both parametric and semiparametric estimation procedures. In Section 3.5, we additionally apply our models to volatility data on close to 5 000 individual US equities from the Trades and Quotes (TAQ) database. Section 3.6 presents a forecasting study, where we compare our new models to existing volatility forecasting models. Section 3.7 concludes. Proofs of the technical results, as well as some auxiliary mathematical derivations, are given in Appendices A.1 and A.2.

### 3.2 Empirical behavior of realized volatility

We consider a simple model for high-frequency asset returns, modeling the efficient log price \( Y = (Y_t)_{t \geq 0} \) of an asset by an Itô semimartingale

\[
dY_t = \mu_t dt + \sigma_t dB_t + dJ_t, \quad t \geq 0, \tag{3.2.1}
\]

where \( B = (B_t)_{t \geq 0} \) is a standard Brownian motion, \( \mu = (\mu_t)_{t \geq 0} \) a drift process, \( J = (J_t)_{t \geq 0} \) a jump process, and \( \sigma = (\sigma_t)_{t \geq 0} \) a volatility process, satisfying the usual assumptions of adaptedness and local boundedness. We assume that the observations of the logarithm of the asset price \( S \) is the efficient price, contaminated by market microstructure noise,

\[
\log S_t = Y_t + U_t, \quad t \geq 0,
\]

where \( U = (U_t)_{t \geq 0} \) is a microstructure noise process.

Our focus in this paper will be on the volatility process \( \sigma \), hence we do not go into much detail regarding the jump process \( J \) or the noise process \( U \). We require only mild assumptions on these processes, such that the jump- and noise-robust estimator of the volatility process, explained below, will be consistent. Sufficient conditions for this are, e.g., that the jump term is given by a compound Poisson process and the noise term is iid with zero-mean and finite fourth moment, and independent of the efficient price \( Y \) (Christensen, Oomen, and Podolskij, 2014, Proposition 1). Also, in the context of high-frequency returns, the drift process \( \mu \) is empirically negligible, and will be ignored from now on. The following sections will explain the data we use for \( S \), how we estimate the latent volatility process \( \sigma \), and the subsequent empirical findings on this process.

### Extraction of latent intraday volatility and description of the data

We seek to extract the realized spot volatility process \( (\sigma_t)_{t \in [0,T]} \), for some time horizon \( T > 0 \), from high-frequency observations of the asset price \( S \). As \( \sigma \) is not directly
observable, we need to construct a proxy for it. In particular, we are interested in assessing intraday variation of volatility. This should be contrasted with, e.g., Gatheral et al. (2014), who consider volatility proxies computed at daily frequency.

To this end, we first specify a step size $\Delta > 0$ such that $T = n\Delta$ for some large $n \in \mathbb{N}$. Then we aim to estimate the integrated variance (IV),

$$IV_t^\Delta := \int_{t-\Delta}^t \sigma_s^2 ds, \quad t = \Delta, 2\Delta, \ldots, n\Delta.$$  

Estimators of IV have been extensively studied, prominent examples being realized variance (Andersen, Bollerslev, Diebold, and Labys, 2001b; Barndorff-Nielsen and Shephard, 2002), realized kernels (Barndorff-Nielsen et al., 2008), two-scale estimators (Zhang, Mykland, and Aït-Sahalia, 2005), and pre-averaging methods (Jacod, Li, Mykland, Podolskij, and Vetter, 2009). Except for the first, these methods are robust to market microstructure effects, which is crucial when using prices sampled at higher frequencies (e.g., Hansen and Lunde, 2006). Further, as we will see below, the pre-averaging methods are straightforwardly adapted to handle jumps in the price process, which is our main reason for choosing this particular approach.

By letting $\Delta$ be sufficiently small, and assuming that volatility does not vary too much in each time interval of size $\Delta$, we can use the proxy

$$\hat{\sigma}_t^2 = \Delta^{-1} \hat{IV}_t^\Delta, \quad t = \Delta, 2\Delta, \ldots, n\Delta,$$

where $\hat{IV}_t^\Delta$ is an estimate of IV derived from one of the methods mentioned. The proxy (3.2.2) can be seen as a (finite difference) time derivative of the estimate of integrated variance. Related estimators of spot volatility have been suggested in the literature; see, e.g., Kristensen (2010), Bos, Janus, and Koopman (2012), and Zu and Boswijk (2014). In this paper we will restrict attention to (3.2.2) where we estimate IV using pre-averaging measures, developed in Jacod et al. (2009). We briefly review the implementation, following the exposition in Christensen et al. (2014) closely.

Suppose we want to estimate IV in some interval $[(i-1)\Delta, i\Delta]$ for $i \geq 1$ and we have $N+1$ (tick-by-tick) observations, $Z_0, Z_1, \ldots, Z_N$, available of the log-price process $Z = \log S$ in this interval. We define the pre-averaged log-returns,

$$r_{j,K}^* = \frac{1}{K} \left( \sum_{k=K/2}^{K-1} Z_{(j+k)} - \sum_{k=0}^{K/2-1} Z_{(j+k)} \right), \quad j = 0, 1, \ldots, N - K,$$

where $K \geq 2$ is even. For the asymptotics to work, it is required that $K = \theta \sqrt{N} + o \left( N^{-1/4} \right)$, and in our implementation we set $\theta = 1$ and $K = \lfloor \sqrt{N} \rfloor$ if $\lfloor \sqrt{N} \rfloor$ is an even number and $K = \lceil \sqrt{N} \rceil + 1$ otherwise.\(^1\) Here, $\lfloor x \rfloor$ means the largest integer smaller

\(^1\) Setting the tuning parameter $\theta$ equal to 1 was found to work well for data similar to ours in Christensen et al. (2014). We come to the same conclusion from both looking at simulated data mimicking our setup, as well as on our actual data treated later in the paper.
than, or equal to, $x \in \mathbb{R}$. Using these pre-averaged returns, we suggest the following estimators of IV, which are robust to market microstructure noise:

$$RV^\Delta_* = \frac{N}{N - K} \sum_{j=0}^{N-K+1} |r^\Delta_j|^2 - \frac{\hat{\omega}^2}{\theta^2 \psi_K},$$

$$BV^\Delta_* = \frac{N}{N - 2K} \sum_{j=0}^{N-2K+1} |r^\Delta_j||r^\Delta_{j+1}| - \frac{\hat{\omega}^2}{\theta^2 \psi_K},$$

where $\psi_K := (1 + 2K^{-2})/12$ and $t \in [(i-1)\Delta, i\Delta]$. The term $\frac{\hat{\omega}^2}{\theta^2 \psi_K}$ is a bias-correction, where $\hat{\omega}^2$ is an estimate of the variance of the microstructure noise process $U$. We use the estimator of Oomen (2006):

$$\hat{\omega}_{AC}^2 = -\frac{1}{N-1} \sum_{j=2}^{N} r_i r_{i-1},$$

where $r_i = Z_i - Z_{i-1}$ is the $i$’th log return. The objects $RV^\Delta_*$ and $BV^\Delta_*$ are the pre-averaged analogous of the realized variance (Andersen et al., 2001b) and bipower variation (Barndorff-Nielsen and Shephard, 2004) estimators of IV, respectively. While both estimators are robust to market microstructure noise, only $BV^\Delta_*$ is also robust to jumps in the price process. For this reason, we will use the $BV^\Delta_*$ estimate of IV in the following; we note, that we also conducted the analyses using $RV^\Delta_*$ and found largely similar results and came to similar conclusions as below, indicating that the effects of jumps are small in the data studied here (cf. also Christensen et al., 2014).

We analyze tick-by-tick transaction data on the front month E-mini S&P 500 futures contract, traded on the CME Globex electronic trading platform, from January 2, 2013 until December 31, 2014 excluding weekends and holidays, which results in 516 trading days. Of these days, 18 days were not full trading days; we removed these days to arrive at a total of 498 days in our sample. As we are interested in assessing volatility at very high frequencies, we rely on there being a lot of trading activity on the underlying asset. For this reason we restrict our attention to the period of the day when most trading is taking place; this is when the New York Stock Exchange (NYSE) is open, from 9.30 a.m. until 4 p.m. Eastern Standard Time (EST).

It is well-known that intraday volatility displays significant seasonality (e.g., Andersen and Bollerslev, 1997, 1998). In particular, the “U-shape” is ubiquitous, where volatility is high at the opening and at the close of the market, while being lower around midday. It is important to control for this seasonality before performing any further analyses as subsequent estimates could be affected if one does not take this into account (Rossi and Fantazzini, 2015). We use a multiplicative decomposition

$$\sigma_t = \sigma^*_t \tilde{\sigma}_t, \quad t \geq 0,$$

\footnote{We also ran the analyses using the realized kernel estimator of Barndorff-Nielsen et al. (2008) and we again found similar results, although this estimator is not robust to jumps either. We do not provide the details here, but they are available upon request.}
3.2. Empirical behavior of realized volatility

where \( \sigma^s \) is the seasonal component and \( \tilde{\sigma} \) is the deseasonalized stochastic process we are interested in. To estimate \( \sigma^s \) we use the flexible Fourier form (FFF) approach of Andersen and Bollerslev (1997, 1998); Figure 3.1 shows the output of this estimation procedure in the case \( \Delta = 15 \) minutes (see also the similar findings in Andersen, Bondarenko, Kyle, and Obizhaeva, 2016, Figure 1). The familiar U-shape of volatility is evident. We then estimate \( \tilde{\sigma}^2_t \) by

\[
\tilde{\sigma}^2_t = \Delta^{-1} \sum_{i=1}^\Delta \left( \frac{\tilde{V}_{t,i}}{\tilde{\sigma}^2_t} \right)^2 = \Delta^{-1} \sum_{i=1}^\Delta \left( \frac{BV_{t,i}}{\tilde{\sigma}^2_t} \right)^2, \quad t = \Delta, 2\Delta, \ldots, n\Delta,
\]

and will from now on be working with these de-seasonalized data. Further, we abuse notation slightly and will write \( \sigma \) even though we actually refer to the de-seasonalized process \( \tilde{\sigma} \). Table 3.1 contains some simple descriptive statistics of this process and how it behaves as \( \Delta \) increases.\(^3\) While the skewness is positive, kurtosis becomes closer to 3, indicating that the volatility estimates look more Gaussian as \( \Delta \) is increased. We will present additional evidence of this in Section 3.2 below.

**Stationarity of volatility**

Apart from intraday seasonality, volatility is widely believed to be stationary. In particular, one does not expect volatility to wander without bound but instead to revert to some “typical” level. In Table 3.1 we present results of several unit root tests applied to the data with various values of \( \Delta \). The two classical unit root tests (ADF and PP) always reject the null of a unit root. The Hansen and Lunde (2013) test (which is appropriate in the present case as our estimate of \( \sigma^2_t \) is measured with error) also rejects the presence of a unit root. This supports the hypothesis that volatility is stationary.

\(^3\)The reason for considering \( \Delta = 65 \) minutes and \( \Delta = 130 \) minutes in our analysis, instead of one and two hours, is to make sure that the 6.5-hour trading day can be divided into an integer number of time periods each of length \( \Delta \).
Table 3.1. Descriptive statistics and unit root tests

<table>
<thead>
<tr>
<th>Δ</th>
<th>n</th>
<th>Mean</th>
<th>Skew</th>
<th>Kurt</th>
<th>ADF</th>
<th>PP</th>
<th>t</th>
<th>n(t – 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 minutes</td>
<td>19422</td>
<td>0.022</td>
<td>0.039</td>
<td>4.184</td>
<td>0.000</td>
<td>0.001</td>
<td>0.974</td>
<td>–500.3</td>
</tr>
<tr>
<td>15 minutes</td>
<td>12948</td>
<td>0.031</td>
<td>0.233</td>
<td>3.808</td>
<td>0.000</td>
<td>0.001</td>
<td>0.970</td>
<td>–394.7</td>
</tr>
<tr>
<td>30 minutes</td>
<td>6474</td>
<td>0.048</td>
<td>0.363</td>
<td>3.662</td>
<td>0.000</td>
<td>0.001</td>
<td>0.966</td>
<td>–220.5</td>
</tr>
<tr>
<td>65 minutes</td>
<td>2988</td>
<td>0.054</td>
<td>0.454</td>
<td>3.633</td>
<td>0.000</td>
<td>0.001</td>
<td>0.960</td>
<td>–120.6</td>
</tr>
<tr>
<td>130 minutes</td>
<td>1494</td>
<td>0.054</td>
<td>0.486</td>
<td>3.523</td>
<td>0.000</td>
<td>0.001</td>
<td>0.950</td>
<td>–75.1</td>
</tr>
<tr>
<td>1 day</td>
<td>498</td>
<td>0.061</td>
<td>0.420</td>
<td>3.410</td>
<td>0.000</td>
<td>0.001</td>
<td>0.830</td>
<td>–84.5</td>
</tr>
</tbody>
</table>

Descriptive statistics and unit root tests of log volatility of the E-mini S&P 500 data set. ADF and PP refer to the P-values of the Augmented Dickey-Fuller test with automatic lag selection (no constant, no trend) and the Phillips-Perron test, respectively. The symbol \( \pi \) denotes the persistence parameter of Hansen and Lunde (2013) and \( n(\pi - 1) \) is the unit root test statistic from the same paper. The 1% and 5% critical values of this test are –20.7 and –14.1, respectively.

Roughness of volatility

Recent studies have provided evidence that volatility is rough; see, e.g., Gatheral et al. (2014) and Bennedsen (2016). What we mean by this is that the autocorrelation function (ACF) \( \rho \) of log volatility, assuming that it is covariance-stationary, adheres to the asymptotic relationship

\[
1 - \rho(h) := 1 - \text{Corr}(\log \sigma_t, \log \sigma_{t+h}) \sim \frac{1}{2\alpha+1} |h|, \quad |h| \to 0, \tag{3.2.3}
\]

for some \( \alpha \in \left(-\frac{1}{2}, 0\right) \). Here “\( \sim \)” indicates that the ratio between the left- and right-hand side tends to a non-zero constant. We call \( \alpha \) the roughness index of the log volatility process. In general, \( \alpha \) takes values in \( \left(-\frac{1}{2}, \infty\right) \) but, as we shall see, only negative values of \( \alpha \) will be relevant for us. For stationary Gaussian processes, the relationship (3.2.3) implies that the process has a modification with \( \phi \)-Hölder continuous trajectories for any \( \phi \in (0, \alpha + 1/2) \), where the index \( \phi \) can be seen as a measure of roughness, with small values indicating more roughness.\(^4\) It is worth recalling here that a standard Brownian motion has \( \phi \)-Hölder continuous trajectories for any \( \phi \in (0, 1/2) \). Thus, negative values of \( \alpha \) suggest trajectories rougher than those of a standard Brownian motion.

Rough models of volatility are consistent with some empirically observed features of implied volatility surfaces (Gatheral, 2006). In particular, as shown in Fukasawa (2016), such models can accurately capture the short-time behavior of the at-the-money volatility skew, which conventional local/stochastic volatility models based on Itô diffusions fail to capture. To model the roughness of volatility, earlier studies have mainly relied on the “canonical” rough process, the fractional Brownian motion (fBm) with Hurst index \( H \in (0, 1/2) \). For the fBm, the simple relationship \( H = \alpha + 1/2 \) holds, which means that \( H < 1/2 \) implies roughness.

\(^4\)See, e.g., Proposition 2.1 of Bennedsen (2016).
3.2. Empirical behavior of realized volatility

Estimating the roughness parameter of log volatility

Consider the (second-order) variogram of log volatility:

\[
\gamma_2(h) := \mathbb{E}\left[|\log \sigma_{t+h} - \log \sigma_t|^2\right], \quad t, h \in \mathbb{R}.
\]

For a covariance-stationary stochastic process, the variogram and the ACF are connected by the relationship

\[
\gamma_2(h) = 2 \text{Var}(\log \sigma_t) (1 - \rho(h)), \quad h \in \mathbb{R},
\]

showing that when the underlying process is covariance-stationary, the asymptotics (3.2.3) hold for the variogram as well. Indeed,

\[
\gamma_2(h) \sim |h|^{2\alpha + 1}, \quad |h| \to 0. \tag{3.2.4}
\]

This suggests a straightforward semiparametric estimation procedure for \( \alpha \). Namely, consider the regression

\[
\log \hat{\gamma}_2(h) = c + a \log |h| + \epsilon_h, \quad h = \Delta, 2\Delta, \ldots, m\Delta, \tag{3.2.5}
\]

for some step size \( \Delta > 0 \) and an integer bandwidth parameter \( m \geq 2 \), where \( \hat{\gamma}(h) \) is the empirical estimate of the variogram at lag \( h \in \mathbb{R} \). The relationship \( a = 2\alpha + 1 \) allows us to estimate \( \alpha \) using \( \hat{\alpha} = \frac{\hat{a} - 1}{2} \), where \( \hat{a} \) is the OLS estimate of \( a \) in (3.2.5).

The error term in the regression (3.2.5) does not satisfy the usual assumptions of OLS estimation, but inference (confidence intervals) on the parameter \( \alpha \) can be conducted using the bootstrap method of Bennedsen (2016), specifically tailored to estimation of rough processes. Choosing the bandwidth parameter, \( m \), optimally is an open question; it should preferably be small, as the asymptotics in (3.2.3) is only assumed to hold for smaller values of \( m \). However, as argued in Bennedsen (2016), choosing \( m \) slightly larger than the minimum value of 2, can improve the robustness of the OLS estimation in (3.2.5) (since one is using more points in the regression). In what follows, we will set \( m = 6 \). We performed the analyses below with several values of \( m \) and found that the results are very robust to the choice of bandwidth.

Unfortunately, this estimation approach, which is standard in the literature on rough processes, is infeasible in our case, since we do not have access to observations of the log volatility process itself, but only an estimate thereof (here through the preaveraged measure \( BV^* \)), as explained in Section 3.2. To see why this might invalidate the OLS-based procedure, recall that the measure \( BV^* \) is a noisy estimate of \( IV \); suppose that \( BV^* \) is representable as

\[
BV^*_t = \hat{I}V^*_t = IV^*_t \eta_t, \quad t = \Delta, 2\Delta, \ldots, n\Delta,
\]

where \( \{\eta_{k\Delta}\}_{k=1}^n \) is a positive iid noise sequence, independent of \( IV^\Delta \). Clearly, this entails (ignoring the deterministic seasonal factor)

\[
\log \hat{\sigma}_t^2 = \log(D^{-1}BV^*_t) = \log(D^{-1}IV^*_t \eta_t) = \log(D^{-1}IV^*_t) + \epsilon_t, \quad t = \Delta, 2\Delta, \ldots, n\Delta, \tag{3.2.6}
\]
CHAPTER 3. DECOUPLING THE SHORT- AND LONG-TERM BEHAVIOR OF
STOCHASTIC VOLATILITY

where now $\epsilon_t := \log \eta_t$ is an iid noise sequence.

From this it is evident that, even if the term $\Delta^{-1} IV^\Lambda_t$ is a good approximation of the latent volatility process $\sigma^2_t$, the estimates of the roughness parameter coming from observations of the time series $\log \hat{\sigma}^2_t$ will be impacted by the noise-term $\epsilon_t$. Indeed, one may wonder whether the findings of roughness of volatility is driven by this fact.

To see that this is not the case, at least for values of $\Delta$ which are not too small, we propose a noise-robust estimator of the roughness parameter $\alpha$. First, notice that according to (3.2.6), the variogram of the observations can be expressed as

$$\gamma^*_2(h) := E[|\log \hat{\sigma}^2_{t+h} - \log \hat{\sigma}^2_t|^2]$$

$$= E[|\log (\Delta^{-1} IV^\Lambda_{t+h}) - \log (\Delta^{-1} IV^\Lambda_t)|^2] + E[|\epsilon_{t+h} - \epsilon_t|^2]$$

$$\approx \gamma_2(h) + 2\sigma^2_\epsilon,$$  (3.2.7)

where $\sigma^2_\epsilon = \text{Var}(\epsilon_t)$ is the variance of the noise sequence $\epsilon$; the approximation in the third line comes from using $\Delta^{-1} IV^\Lambda_t$ as an approximation of $\log \sigma^2_t$. Thus, even though the variogram of log volatility behaves as in (3.2.4), the logarithm of the variogram of the observations, $\log \gamma^*_2(h)$, will not be linear in $\log h$, as is the case when no noise is present, cf. Equation (3.2.5). Because the OLS estimator relies on this linearity, estimates resulting from the OLS procedure might be biased. Furthermore, at least in the case of an iid noise sequence $\epsilon$, it is easy to show that this bias is downwards. In other words, applying the OLS estimator of $a$ to noisy data might lead the researcher to infer that the observations come from a process which is more rough than what is actually the case for the true underlying process.

To remedy this, we suggest to estimate $\alpha$ in a noise-robust way, by running the non-linear least squares (NLLS) regression, inspired by (3.2.7) and (3.2.4),

$$\hat{\gamma}^*_2(h) = a + b|h|^{2\alpha+1} + \xi_h, \quad h = \Delta, 2\Delta, \ldots, m\Delta,$$  (3.2.8)

where $\xi_h$ is a noise sequence, $m$ is the bandwidth, and $\hat{\gamma}^*_2(h)$ the empirical estimate of the variogram using the (noisy) observations. This regression has three parameters $-a, b,$ and $\alpha$ – and although we are chiefly interested in the roughness parameter $\alpha$, it is useful to note that we can estimate $2\sigma^2_\epsilon$ by the NLLS estimate of the constant $a$. A related estimator of the fractional difference parameter $d$ of a (discrete time) fractional process perturbed by noise, was studied in Phillips and Sun (2003). Although simulations show (not given here, but available upon request) that the NLLS estimator is less precise (in terms of standard deviation of the estimates) and more sensitive to

---

Actually, using the approximation $\Delta^{-1} IV^\Lambda_t = \Delta^{-1} \int_{t-1}^t \sigma^2_s ds \approx \sigma^2_t$ will result in estimates of the roughness parameter of $\sigma^2_t$ which are biased upwards, since integration is a smoothing operation, see e.g. Appendix C of Gatheral et al. (2014) for an analysis of this phenomenon. In other words, any findings of roughness in the time series of scaled integrated log volatility, i.e. of $\log (\Delta^{-1} IV^\Lambda_t)$, will hold a fortiori for the time series of the (latent) spot log volatility, $\log \sigma^2_t$. 

---
the choice of bandwidth $m$ than the OLS estimator, it is unbiased in the presence of noise in the observations as in (3.2.6), which the OLS estimator is not.

The above discussion suggests two approaches for estimating the roughness parameter $\alpha$ of log volatility; one based on OLS regression (where the noise in the estimates of IV is ignored) and one based on NLLS regression. Examples of the OLS regressions for $\Delta = 15, 65$ minutes are presented in Figure 3.2. The plots show that $\log \hat{\gamma}^2(h)$ is approximately linear in $h$ for small values of $h$, which is evidence for the relationship (3.2.3) holding for time series of log volatility (or, more accurately, for time series of pre-averaged measures of intraday integrated variance) for these values of $\Delta$. In Figure 3.3 we apply both the OLS and the NLLS estimators to the E-mini S&P 500 data to obtain a signature plot for $\alpha$, i.e., the estimated value of $\alpha$ is plotted as a function of the step size $\Delta$ (on a log-scale). The figure shows that for $\Delta < 5$ minutes, both estimators seem to be severely biased downwards. It is likely that anything less than 5 minutes is simply too small a time interval to get reasonable estimates of IV. For $\Delta \in [5, 10]$ minutes, however, the NLLS procedure results in estimates of $\alpha$
which are larger than the corresponding OLS estimates. Indeed, for these values of $\Delta$, the NLLS estimates lie outside the 95% confidence bands of the OLS estimates. This is an indication of downwards bias in the OLS estimates, resulting from the noisy measurements of IV, as discussed above. When $\Delta \geq 15$ minutes, both estimators are in good agreement. In particular, the NLLS estimates now lie within the confidence bands of the OLS estimates.

We conclude that there seems to be some downward bias in the OLS estimates of $\alpha$, induced by the noisy measures of integrated variance used as proxy of the underlying volatility. The NLLS estimator can alleviate this bias for moderate values of $\Delta$, e.g. $\Delta \geq 5$ minutes. For $\Delta \geq 15$ minutes both estimators agree, indicating that the effect of noise in the estimates of IV does not influence the OLS estimator of $\alpha$ much for these values of $\Delta$. Below we will only consider $\Delta \geq 10$ minutes; since the two estimators give roughly the same estimates in these cases, and since the OLS estimator (3.2.5) is faster, easier to implement, and more accurate in general (when no noise is present in the data), we will use this estimator going forward.

The above estimates of $\alpha$ have been computed using time series obtained by concatenating the volatility proxy data on successive days, ignoring overnight volatility. To make sure that the estimates of $\alpha$ have not been influenced by the concatenation of the data we also estimate $\alpha$ within each day, which is feasible when we have enough observations of volatility per day (when $\Delta$ is less than 15 minutes, say). The results, presented in Figure 3.4, indicate that the estimates of $\alpha$ on individual days are consistent with the results on the concatenated data, seen in Figure 3.3. In particular, the strong evidence of the roughness of volatility is carried over.

**Does roughness change over time?**

One may wonder whether the degree of roughness of volatility changes over time. This very question was also studied by Gatheral et al. (2014), who divided their volatility data into two parts and found tentative evidence that volatility was less rough during a period that overlapped with the financial crises of 2008 and 2011.
Our methodology and data allow us to investigate this question more precisely and systematically. To this end, we still use transaction data on E-mini S&P 500, but now over a longer period from January 3, 2005 until December 31, 2014. Figure 3.5 provides results of rolling-window estimation of $\alpha$, where the window length was 10 days, using the OLS regression (3.2.5) of Section 3.2 and $\Delta = 15$ minutes for the volatility proxy. We experimented also with different window sizes and different values of $\Delta$, but the results were rather consistent. Figure 3.5 additionally displays the overall median over the entire period, as well as a smoothed version of the estimates. The figure shows that $\alpha$ does indeed appear to vary in time. In particular, we observe several peaks of “smoothness” that coincide with periods of market turmoil. The plot indicates the respective dates of the onset of the subprime mortgage crisis of 2008, the Flash Crash of May 6, 2010, and the nadir of the Greek debt crisis in 2011, which in essence concur with the peaks of the estimated value of $\alpha$.

While it seems imprudent to draw any definite conclusions from a single time series, the findings presented here, together with the empirical evidence of Gatheral et al. (2014), seem to indicate that volatility exhibits less roughness during periods of market turmoil, possibly due to more sustained trading occurring in such times (see also Jaisson and Rosenbaum, 2016, for a microstructure-based theory on the connection between trading and rough volatility).

**Strong persistence of volatility**

That volatility is very persistent has long been a well-established fact (e.g., Bollerslev and Wright, 2000; Andersen et al., 2003). A large body of literature has therefore...
focused on modeling (log) volatility using long memory models, i.e., volatility models whose autocorrelations decay at a slow polynomial rate:

\[ \rho(h) = \text{Corr}(\log \sigma_t, \log \sigma_{t+h}) \sim |h|^{-\beta}, \quad |h| \to \infty, \]  \tag{3.2.9} 

for some \( \beta \in (0, \infty) \). When \( \beta \in (0, 1) \), the correlation function \( \rho \) is not integrable, that is, \( \int_0^\infty |\rho(h)| \, dh = \infty \), and we say that \( \sigma_t \) has the long memory property.

Models of log volatility that are able to reproduce the long memory property have been traditionally built using an fBm with Hurst index \( H = 1 - \beta/2 \in (1/2, 1) \) as driving noise; see, e.g., Comte and Renault (1996), Comte (1996), Comte and Renault (1998), and Comte, Coutin, and Renault (2012) for literature on continuous-time volatility models with long memory.

Paralleling the approach of Section 3.2, the relationship (3.2.9) provides a basis for semiparametric estimation of \( \beta \) via the OLS regression

\[ \log \hat{\rho}(h) = c + b \log |h| + \epsilon_h, \quad h = M \Delta, (M + 1) \Delta, \ldots, M' \Delta, \]  \tag{3.2.10} 

where \( \hat{\rho}(h) \) is the empirical autocorrelation function at lag \( h \in \mathbb{R} \), \( b = -\beta \), and \( M, M' \in \mathbb{N} \) are such that \( M' > M \) and \( M \Delta \) is large. Figure 3.6 presents the OLS regressions of the form (3.2.10) with step sizes \( \Delta = 15, 65 \) minutes. Further, Figure 3.7 shows a signature plot for \( \beta \) using the above estimator, i.e., the estimated value of \( \beta \) is plotted as a function of the step size \( \Delta \) (on a log scale). It should be noted that semiparametric estimates of the long memory parameter \( \beta \) tend to be somewhat imprecise; and we also found that the particular estimates were quite dependent on the values chosen for the thresholds \( M \) and \( M' \). For this reason, we also apply a parametric estimator, the details of which are presented in Section 3.4.

We observe that at very short time scales, volatility has a high degree of persistence and possibly long memory, with \( \hat{\beta} \approx 0.10 \). At longer time scales the semiparametric persistence estimates weaken somewhat, increasing the estimated values of \( \beta \). The parametric estimates stabilize, however, as was the case with the roughness estimates, and these indicate that volatility is very persistent for all values of \( \Delta \).

**Non-Gaussianity of log volatility**

There is some evidence that increments of log volatility follow a Gaussian distribution. This was for instance found in the seminal papers by Andersen et al. (2001b), Andersen, Bollerslev, Diebold, and Ebens (2001a), and Barndorff-Nielsen and Shephard (2002); see also Gatheral et al. (2014) for a more recent analysis. Specifically, these papers examine the empirical distribution of the increments of daily logarithmic realized variance, and find that a Gaussian distribution fits well to the empirical distribution.

We perform a similar analysis, but we consider also intraday time scales and we fit a normal-inverse Gaussian (NIG) distribution, which is a flexible distribution with
3.2. Empirical Behavior of Realized Volatility

Figure 3.6. Estimates of $\hat{\beta}$ using the OLS regression (3.2.10) with $M = \lfloor n^{1/4} \rfloor$ and $M' = \lfloor n^{1/3} \rfloor$.

Figure 3.7. A signature plot for $\beta$, where $\hat{\beta}$ is plotted as a function of $\Delta$. The semiparametric estimation was done using the OLS regression (3.2.10) with $M = \lfloor n^{1/4} \rfloor$ and $M' = \lfloor n^{1/3} \rfloor$. The parametric estimates come from the Cauchy model, as explained in Section 3.4. The leftmost point corresponds to $\Delta = 1$ minute and the rightmost point to $\Delta = 1$ day. Note that the x-axis is on a log scale.

semi-heavy tails, to volatility data (see Barndorff-Nielsen, 1997, for details on the NIG distribution and its application to stochastic volatility modelling). The results, given in Figure 3.8, suggest that increments of the $BV^{\Delta^*}$ at intraday time scales ($\Delta < 1$ day) are clearly leptokurtic, and thus non-Gaussian, while the NIG distribution, estimated by maximum likelihood, appears to fit quite well. At the standard time scale of one day, however, we find that, in accordance with the papers cited above, a Gaussian distribution fits adequately to the $BV^{\Delta^*}$ data, indicating some kind of aggregational Gaussianity in volatility. These findings are confirmed by the QQ plots in Figure 3.9.

We additionally studied the empirical distributions of the $BV^{\Delta^*}$s of the 26 US equities whose volatility we analyze below in Section 3.5. While the detailed results are available upon request, we found that a significant number of them also exhibit non-Gaussian behavior of log volatility, even at the longest time scale of one day. The E-mini S&P 500 contract is very liquid and less volatile compared to most individual equities. Therefore, it seems plausible that riskier assets, such as equities, display more non-Gaussian volatility, and that the conclusions about non-Gaussianity of volatility at intraday time scales apply, a fortiori, also to them.
Chapter 3. Decoupling the Short- and Long-Term Behavior of Stochastic Volatility

3.3 Models of Stochastic Volatility that Decouple Short- and Long-Term Behavior

Motivated by our empirical findings in Section 3.2, we seek stationary real-valued stochastic processes with arbitrary roughness index $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and long-term memory structure that is independent of the value of $\alpha$. Given a process $X$ with such properties, we then define our model for volatility as

$$\sigma_t = \xi \exp\left(X_t\right), \quad t \geq 0, \quad (3.3.1)$$

where $\xi > 0$ is a free parameter.

Models for Log Volatility

We consider two main candidates for $X$: the Cauchy process and the so-called Brownian semistationary process. Common to the two processes is that they, in the setting we consider here, both have two parameters, controlling the short- and long-term
3.3. MODELS OF STOCHASTIC VOLATILITY THAT DECOUPLE SHORT- AND LONG-TERM BEHAVIOR

Figure 3.9. QQ plots for increments of log volatility, i.e., $y_k = \log(BV^\Delta_k) - \log(BV^\Delta_{k-1})$. If the data fits a particular distribution — solid red for the Gaussian distribution, dashed green for the NIG distribution — the curve will be close to the diagonal solid black line.

behavior, respectively. We will see that the latter process, in particular, is ideally suited to volatility modeling.

The Cauchy process

A flexible Gaussian process that decouples the short- and long-term behavior can be obtained by using the Cauchy class of autocorrelation functions (Gneiting and Schlather, 2004). The resulting Cauchy process is a centered, stationary Gaussian process $G = (G_t)_{t \in \mathbb{R}}$ with autocorrelation function

$$\rho(h) = \left(1 + |h|^{2\alpha+1}\right)^{-\frac{\beta}{2\alpha+1}}, \quad h \in \mathbb{R}, \quad (3.3.2)$$

where $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and $\beta > 0$. In particular, the process $G$ satisfies (3.2.3) and (3.2.9).

However, the limitation of this process, from a modeling point of view, is its inherent Gaussianity. Yet it is possible to go beyond Gaussianity by volatility modulation,
that is, by specifying a process

\[ X_t = \int_0^t \nu_s dG_s, \quad t \geq 0, \tag{3.3.3} \]

using the Cauchy process \( G \) and some additional process \( \nu = (\nu_t)_{t \in \mathbb{R}} \) that models the volatility of volatility. Under suitable assumptions, \( X \) will inherit the roughness properties of \( G \), see, e.g., Barndorff-Nielsen et al. (2009a, Example 3). It is worth stressing that since \( G \) is typically a non-semimartingale — and with parameter values relevant to rough volatility modeling it indeed is — the stochastic integral in (3.3.3) cannot be defined as an Itô integral, but pathwise Young integration (Dudley and Norvaiša, 2011) needs to be used, which requires some additional assumptions on the roughness of \( \nu \). In the case where the Cauchy process \( G \) is rough, these assumptions become rather restrictive, unfortunately — for example, \( \nu \) cannot be a semimartingale (with non-zero quadratic variation).

**The Brownian semistationary process**

A stochastic process that is able to capture all of our desiderata, consisting of roughness, strong persistence, stationarity, and non-Gaussianity is the Brownian semistationary process (BS S), which was introduced in Barndorff-Nielsen and Schmiegel (2007, 2009). This process is defined via the moving-average representation

\[ X_t = \int_{-\infty}^t g(t-s)\nu_s dW_s, \quad t \geq 0, \tag{3.3.4} \]

where \( W = (W_t)_{t \in \mathbb{R}} \) is a standard Brownian motion defined on \( \mathbb{R} \), \( g : (0, \infty) \to \mathbb{R} \) is a square-integrable kernel function, and \( \nu = (\nu_t)_{t \in \mathbb{R}} \) is an adapted, covariance-stationary volatility (of volatility) process. Note that when \( \nu \) is deterministic, \( X \) is Gaussian, while a stochastic \( \nu \) makes \( X \) non-Gaussian. In particular, when \( \nu \) is independent of \( W \), we have

\[ X_t|_{\nu_s \leq t} \sim N \left( 0, \int_0^\infty g(x)^2 \nu_{t-x}^2 dx \right), \]

showing that the marginal distribution of \( X_t \) is a normal mean-variance mixture with conditional variance governed by \( \nu \) and \( g \). It is shown in Barndorff-Nielsen et al. (2013) that when \( g \) is given by the so-called gamma kernel (Example 3.3.2 below), we can choose the normal-inverse Gaussian (NIG) distribution as the marginal distribution of \( X \). As we saw in Section 3.2, the NIG distribution seems to fit to the empirical distribution of log volatility very well. This is an encouraging property of the BS S framework, and using this as a guide to specify a model for volatility of volatility is a promising approach, but beyond the scope of the present study. We therefore leave such extensions for future work.

Under the assumptions given above, the process \( X \) is already well-defined and covariance-stationary. For stationarity, integration from \(-\infty\) in (3.3.4) is crucial. We
3.3. MODELS OF STOCHASTIC VOLATILITY THAT DECOUPLE SHORT- AND LONG-TERM BEHAVIOR

will now introduce additional assumptions concerning the properties of the kernel function $g$, which enable us to derive some theoretical results for $X$.

(A1) For some $\alpha \in (-1/2, 1/2) \backslash \{0\},$

\[
g(x) = x^\alpha L_0(x), \quad x \in (0, 1],
\]

(3.3.5)

where the function $L_0$ is continuously differentiable, bounded away from zero, and slowly varying function at zero in the sense that $\lim_{x \downarrow 0} \frac{L_0(tx)}{L_0(x)} = 1$ for all $t > 0$.\(^6\) Furthermore, the derivative $L_0'$ of $L_0$ satisfies

\[
|L_0'(x)| \leq C(1 + x^{-1}), \quad x \in (0, 1],
\]

for some constant $C > 0$.

(A2) The function $g$ is continuously differentiable with derivative $g'$ that is ultimately monotonic and satisfies $\int_1^\infty g'(x)^2 \, dx < \infty$.

(A3) For some $\lambda \geq 0$ and $\gamma \in \mathbb{R}$, such that $\gamma > 1/2$ when $\lambda = 0,$

\[
g(x) = e^{-\lambda x} x^{-\gamma} L_1(x), \quad x \in (1, \infty),
\]

(3.3.6)

where $L_1$ is slowly varying at infinity and bounded away from zero and $\infty$ on any finite interval.

Assumptions (A1) and (A3) refine the earlier standing assumption that $g$ is square-integrable. Indeed, in the case $\lambda = 0$, a simple application of the so-called Potter bounds (Bingham et al., 1989, Theorem 1.5.6(ii)) shows that $\alpha > -1/2$ and $\gamma > 1/2$ are sufficient conditions for $g$ to be square integrable under the specifications (3.3.5) and (3.3.6). Similarly, in the case $\lambda > 0$, the conditions $\alpha > -1/2$ and $\gamma \in \mathbb{R}$, under (3.3.5) and (3.3.6), suffice for square integrability. The assumptions (A1), (A2), and (A3) are similar to those used in Bennedsen et al. (2015), the only difference being that (A3) is slightly more specific compared to the corresponding assumption in that paper.

The following proposition shows that under assumptions (A1) and (A2), the $\mathcal{B,F}_t^F$ process $X$, defined by (3.3.4), has $\alpha$ as its roughness index in the sense of Equation (3.2.3). The result is a straightforward adaptation of Proposition 2.1 in Bennedsen et al. (2015) and we refer to that paper for a proof. Below, and in what follows, we denote by $\rho_X$ the autocorrelation function of $X$.

**Proposition 3.3.1.** If the kernel function $g$ satisfies (A1) and (A2), then

\[
1 - \rho_X(h) \sim L_0(|h|)|h|^{2\alpha + 1}, \quad |h| \downarrow 0.
\]

\(^6\)We refer to Bingham et al. (1989) for an extensive treatment of slowly varying (and regularly varying) functions.
The tail behavior of the kernel function \( g \) at infinity, as specified by (A3), controls the long-term memory properties of \( X \). We consider first the case \( \lambda = 0 \), where the exponential damping factor in (3.3.6) disappears. In this case, the parameter \( \gamma \) controls the asymptotic memory properties of the \( \mathcal{B}_\mathcal{F} \) process \( X \).

**Proposition 3.3.2.** Suppose that the kernel function \( g \) satisfies (A3) with \( \lambda = 0 \).

(i) If \( \gamma \in (1, \infty) \), then

\[
\rho_X(h) \sim L_1(|h|)|h|^{-\gamma}, \quad |h| \to \infty.
\]

(ii) If \( \gamma \in (1/2, 1) \), then

\[
\rho_X(h) \sim L_1(|h|)|h|^{1-2\gamma}, \quad |h| \to \infty.
\]

**Remark 3.3.1.** In the critical case \( \gamma = 1 \), the asymptotic behavior of \( \rho_X \) is indeterminate under (A3), and would require additional assumptions on the slowly varying function \( L_1 \).

**Remark 3.3.2.** It follows from Proposition 3.3.2 that if the kernel function \( g \) satisfies (A3) with \( \lambda = 0 \), the resulting \( \mathcal{B}_\mathcal{F} \) process will — up to a slowly varying factor — have an autocorrelation function which decays polynomially as \( h \to \infty \). Indeed, letting the rate of polynomial decay be denoted by \( \beta \) as in (3.2.9), we see that \( \gamma = \beta \) when \( \gamma > 1 \), while \( \beta = 2\gamma - 1 \) when \( \gamma \in (1/2, 1) \). From this it also follows that if \( \gamma \in (1/2, 1) \), then

\[
\int_0^\infty \rho_X(h) \, dh = \infty,
\]

i.e., \( X \) has the long memory property.

In contrast to the case \( \lambda = 0 \), the assumption (A3) with \( \lambda > 0 \) allows for models where autocorrelations decay to zero exponentially fast, leading to short memory, as shown in the following result.

**Proposition 3.3.3.** If the kernel function \( g \) satisfies (A3) with \( \lambda > 0 \) and \( \gamma \in \mathbb{R} \), then

\[
\rho_X(h) \sim e^{-\lambda|h|}|h|^{-\gamma}L_1(|h|), \quad |h| \to \infty.
\]

An example of a kernel function that satisfies (A1), (A2), and (A3), which will be important for us later on, is the power law kernel.

**Example 3.3.1** (Power law kernel). Let \( g \) be the power law kernel

\[
g(x) = x^\alpha(1 + x)^{-\gamma - \alpha}, \quad x > 0, \quad \alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right), \quad \gamma \in \left(\frac{1}{2}, \infty\right). \tag{3.3.7}
\]
Bennedsen et al. (2015, Example 2.2) show that this kernel function indeed satisfies (A1), (A2), and (A3). In particular, with this kernel function, the BS process $X$ has roughness index $\alpha$ and memory properties controlled by $\gamma$, as expounded in Proposition 3.3.2. In the following, we will refer to the BS process with the power law kernel as the Power-BS process.

Later we will need the correlation structure of the Power-BS process. By covariance-stationarity of $v$, the autocovariance function of the general BS process (3.3.4) is

$$c_X(h) := \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[v_0^2] \int_0^\infty g(x)g(x+|h|)dx, \quad h \in \mathbb{R}. \quad (3.3.8)$$

From this we deduce that when $g$ is given as in (3.3.7) we have

$$c_X(0) := \text{Var}(X_t) = \mathbb{E}[v_0^2] \int_0^\infty x^{2\alpha}(1+x)^{-2\gamma-2\alpha} dx$$

$$= \mathbb{E}[v_0^2] \mathcal{B}(2\alpha + 1, 2\gamma - 1),$$

where $\mathcal{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt = \int_0^\infty t^{x-1}(1+t)^{-x-y} dt$ is the beta function (e.g., Gradshteyn and Ryzhik, 2007, formula 8.380.3). To calculate the correlation function $\rho_X(h) = c_X(h)/c_X(0)$ we resort to numerical integration of (3.3.8). Note that $\rho_X$ does not depend on $\mathbb{E}[v_0^2]$.

Another example of a kernel function that satisfies equations (A1), (A2), and (A3), which will also be important in the sequel, is the gamma kernel.

**Example 3.3.2 (Gamma kernel).** Let $g$ be the gamma kernel

$$g(x) = x^\alpha e^{-\lambda x}, \quad x > 0, \quad \alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right), \quad \lambda \in (0, \infty), \quad (3.3.9)$$

which satisfies (A1), (A2), and (A3), as shown in Bennedsen et al. (2015, Example 2.1). With this kernel function, the process $X$ has roughness index $\alpha$ and memory properties controlled by $\lambda$, as per Proposition 3.3.3. In the following, we will call the BS process with the gamma kernel the Gamma-BS process.

We will also need the correlation structure of the Gamma-BS process. We easily find

$$c_X(0) := \text{Var}(X_t) = \mathbb{E}[v_0^2] \int_0^\infty x^{2\alpha} e^{-2\lambda} dx$$

$$= \mathbb{E}[v_0^2] \Gamma(2\alpha + 1) \lambda^{-2\alpha - 1} \Gamma(2\alpha + 1),$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the gamma function. For general $h \in \mathbb{R}$, we have the autocovariance function, using Gradshteyn and Ryzhik (2007, formula 3.383.8),

$$c_X(h) := \text{Cov}(X_{t+h}, X_t) = \mathbb{E}[v_0^2] \frac{\Gamma(2\alpha + 1)}{\pi} \left(\frac{|h|}{2\lambda}\right)^{\alpha+1/2} K_{\alpha+1/2}(\lambda|h|),$$

where $K_n(x) = \int_0^\infty e^{-t} t^{-n/2} (1+t)^{-n/2} dt$ is the modified Bessel function of the second kind.
where $K_v(x)$ is the modified Bessel function of the third kind with index $v$, evaluated at $x$ (see e.g., Gradshteyn and Ryzhik, 2007, section 8.4), and the autocorrelation function can be computed using the identity $\rho_X(h) = c_X(h)/c_X(0)$. We note that $c_X$ is the Matérn covariance function (Matérn, 1960; Handcock and Stein, 1993), which is widely used in many areas of statistics, e.g., spatial statistics, geostatistics, and machine learning.

**Remark 3.3.3.** These two examples of kernel functions exemplify the theoretical distinction between long and short memory. In particular, by Proposition 3.3.3, the Gamma-BS process adheres to $\rho_X(h) \sim e^{-\lambda h} h^\alpha$, $h \to \infty$, i.e., it has short memory, while the Power-BS process will have polynomially decaying ACF and, in particular, the long memory property when $\gamma < 1$, cf. Remark 3.3.2. Although, theoretically, the Gamma-BS process has short memory, by selecting very small values of $\lambda$, it is possible to specify processes with a very high degree of persistence, mimicking long memory on finite time intervals. Empirically, these two BS models allow us to assess if there is any gain from using a model with bona fide long memory, as opposed to a highly persistent model with (technically) short memory, in particular in terms of forecasting accuracy.

**Implications for raw volatility**

The following results show that when log volatility is a BS process, the roughness and memory properties will carry over to the volatility process itself. A result related to Theorem 3.3.1(i) below was given in Comte and Renault (1998) in the case where $X$ is an fBm. Our results here are stated for a Gaussian BS process, i.e., when $v_t = v > 0$ is constant for all $t$, but we conjecture that the results hold also for more general BS processes, under suitable assumptions.

Let $\sigma = (\sigma_t)_{t \geq 0}$ be as in (3.3.1) and

$$\rho(h) = \text{Corr}(\sigma_{t+h}, \sigma_t), \quad h \in \mathbb{R}.$$ 

The first part of the following theorem shows that $\sigma$ inherits the roughness properties of the BS process $X$. The second part shows that the same is true of the long-term memory properties.

**Theorem 3.3.1.** Let $\sigma$ be given by (3.3.1) where $X$ is a BS process satisfying (A1), (A2), and (A3), with $v_t = v > 0$ for all $t$. Then,

(i) as $|h| \to 0$,

$$1 - \rho(h) \sim |h|^{2\alpha+1}L_0(|h|).$$

(ii) as $|h| \to \infty$,

$$\rho(h) \sim \rho_X(h).$$
Simulation of the stochastic volatility model

Fast and efficient simulation of a stochastic volatility model is advantageous for a number of reasons. For instance, one might wish to conduct simulation experiments to assess the properties of the model, or one might wish to price derivatives by Monte Carlo simulation. We discuss here briefly how our model can be simulated rather easily and efficiently. Effective and economical simulation methods for rough volatility models, such as the ones considered in this paper, should not be taken for granted, however. Rough volatility models are typically non-Markovian, depending on the entire history of the process, which makes conventional recursive simulation methods inapplicable. What is more, the possibility of non-Gaussianity of the \( \mathcal{BIFI} \) process poses further problems, as this rules out simulation methods based on Gaussianity, such as Cholesky factorization and circulant embedding methods (e.g., Asmussen and Glynn, 2007, Chapter XI).

According to our underlying assumptions, cf. equations (3.2.1) and (3.3.1), the model to be simulated is

\[
S_t = S_0 \exp \left( \int_0^t \sigma_s dB_s - \frac{1}{2} \int_0^t \sigma_s^2 ds \right),
\]

\[
\sigma_t = \xi \exp \{X_t\},
\]

where \( \xi > 0 \) and \( X \) is one of our candidate models for log volatility, presented in Section 3.3. We can simulate \( S \) on a grid using Riemann-sum approximations of the integrals. To this end, we need to first simulate \( B \) and \( \sigma \) on the same grid, which boils down to simulating \( B \) and \( X \). As we typically want to make these processes correlated, to capture the leverage effect, it is necessary to simulate \( B \) and \( X \) jointly.

When \( X \) is Gaussian, for instance a Cauchy process or a \( \mathcal{BIFI} \) process with constant volatility, it can be simulated exactly using, e.g., a Cholesky factorization of the covariance matrix of the observations (Asmussen and Glynn, 2007, pp. 311–314). One can additionally compute the covariance structure of the Gaussian bivariate process \( (B, X) \) and simulate \( B \) and \( X \) jointly, and in this way account for correlation between the two processes. This was the approach taken in Bayer et al. (2016). However, as the authors also note, the Cholesky factorization is computationally expensive and can become even infeasible if the number of observations to be simulated is very large. Instead, we recommend using the circulant embedding method (Asmussen and Glynn, 2007, pp. 314–316) in the Gaussian case or the hybrid scheme of Bennedsen et al. (2015) in the general case. The hybrid scheme is tailor-made for \( \mathcal{BIFI} \) processes and its advantages are that (i) simulation is fast and in most cases accurate\(^7\) (although approximate), (ii) it allows for non-Gaussianity of \( X \) through volatility (of

\[\text{footnote: The hybrid scheme requires truncating the integral representation (3.3.4) “near” infinity. If the kernel function has extremely slow decay, this truncation may lead to some loss of accuracy, with regards to the memory properties of the process.}\]
volatility) modulation, and (iii) inclusion of leverage, i.e., correlation between \(X\) and \(B\), is straightforward.

We refer to Bennedsen et al. (2015) for an exposition of the hybrid scheme for the \(BIS\) process \(X\), defined by (3.3.4), under assumptions (A1), (A2), and (A3). The authors explain in the paper (Bennedsen et al., 2015, Section 3.1) how to incorporate correlation between \(v\) and \(W\), while the same procedure can be used to introduce correlation between \(W\) and \(B\) or, indeed, between \(W\), \(B\), and \(v\).

### 3.4 Estimating the models

Estimating the new models presented in Section 3.3 involves no complications. In particular, \(\alpha\) can be estimated semiparametrically by the OLS regression (3.2.5) and \(\beta\), similarly, by (3.2.10). The memory parameters \(\gamma\) and \(\lambda\) of the Power- and Gamma-\(BIS\) process, respectively, can be estimated by a method-of-moments procedure, described next.

As mentioned earlier, semiparametric estimation of the memory parameter \(\beta\) using (3.2.9) can be somewhat unreliable in finite samples, unfortunately. Therefore, we also estimate \(\beta\) parametrically, using a method-of-moments approach, by fitting the theoretical autocorrelation function (ACF) of the model to the empirical ACF. More specifically, we use here the parametric ACFs of the Cauchy (3.3.2) and Power-\(BIS\) (3.3.8) processes, respectively.\(^8\) One could simultaneously estimate \(\alpha\) following the method-of-moments approach, but simulation results (available upon request) indicate that the OLS estimator of \(\alpha\) is very precise, suggesting that the best performance overall is achieved when \(\alpha\) is first estimated by OLS and then \(\beta\) is estimated by the method of moments. That is, to estimate the memory parameter, we plug \(\hat{\alpha}_{OLS}\) into the theoretical ACF so that it becomes a function of \(\beta\) (or \(\gamma\) or \(\lambda\)) only, and we then minimize the sum of squared distances between this function and the empirical ACF \(\hat{\beta}\). We used \(L = \lceil n^{1/3} \rceil\) lags in the ACF for this estimation procedure. While this choice is somewhat arbitrary, the results were rather robust to alternative choices. See Figure 3.10 for the fitted ACFs and Section 3.4 below for a discussion of the results.

In Table 3.2, Panel A, we report the estimates of \(\alpha\), \(\lambda\), and \(\beta\) for the different models using the data on log volatility of the E-mini S&P 500 futures contracts, extracted as explained in Section 3.2. In Sections 3.2 and 3.2, we have already seen the OLS estimates of \(\alpha\) and \(\beta\), indicating roughness and strong persistence of log volatility, which are corroborated here. Indeed, the parametric estimates \(\hat{\beta}_{Cauchy}\) using the Cauchy model agree rather well with the semiparametric estimates \(\hat{\beta}_{OLS}\) (apart from daily frequency). The \(\hat{\beta}_{BSS}\) estimates from the Power-\(BIS\) model are a somewhat larger than \(\hat{\beta}_{OLS}\) and \(\hat{\beta}_{Cauchy}\), at least for the higher frequencies, but the

---

\(^8\)In the case of the Power-\(BIS\) process, we actually estimate the parameter \(\gamma\) using the method-of-moments procedure and express the estimate in terms of \(\beta\) using the relationship described in Remark 3.3.2.
Table 3.2. Estimating the models on log and raw volatility

Panel A: Log volatility

<table>
<thead>
<tr>
<th>Δ</th>
<th>$\hat{\alpha}_{OLS}$</th>
<th>95% CI</th>
<th>$\hat{\lambda}_{BSS}$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{OLS}$</th>
<th>$\hat{\beta}_{BSS}$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{Cauchy}$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 minutes</td>
<td>-0.38 (-0.39, -0.37)</td>
<td>0.12 (0.12,0.12)</td>
<td>0.19 (0.52,0.54)</td>
<td>0.25 (0.24,0.26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 minutes</td>
<td>-0.34 (-0.35, -0.33)</td>
<td>0.19 (0.19,0.20)</td>
<td>0.24 (0.68,0.72)</td>
<td>0.33 (0.32,0.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 minutes</td>
<td>-0.30 (-0.32, -0.27)</td>
<td>0.22 (0.20,0.24)</td>
<td>0.21 (0.72,0.83)</td>
<td>0.40 (0.39,0.41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 hour</td>
<td>-0.29 (-0.33, -0.26)</td>
<td>0.16 (0.14,0.18)</td>
<td>0.23 (0.60,0.70)</td>
<td>0.38 (0.36,0.40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 hours</td>
<td>-0.33 (-0.37, -0.28)</td>
<td>0.10 (0.09,0.10)</td>
<td>0.54 (0.46,0.57)</td>
<td>0.33 (0.30,0.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>-0.30 (-0.36, -0.24)</td>
<td>0.10 (0.09,0.11)</td>
<td>1.14 (0.35,0.87)</td>
<td>0.41 (0.32,0.51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Raw volatility

<table>
<thead>
<tr>
<th>Δ</th>
<th>$\hat{\alpha}_{OLS}$</th>
<th>95% CI</th>
<th>$\hat{\lambda}_{BSS}$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{OLS}$</th>
<th>$\hat{\beta}_{BSS}$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{Cauchy}$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 minutes</td>
<td>-0.40 (-0.51, -0.29)</td>
<td>0.47 (0.43,0.50)</td>
<td>0.20 (1.13,1.24)</td>
<td>0.37 (0.36,0.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 minutes</td>
<td>-0.43 (-0.50, -0.26)</td>
<td>0.24 (0.22,0.27)</td>
<td>0.18 (0.81,0.98)</td>
<td>0.28 (0.27,0.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 minutes</td>
<td>-0.37 (-0.49, -0.24)</td>
<td>0.31 (0.27,0.36)</td>
<td>0.25 (0.95,1.11)</td>
<td>0.43 (0.41,0.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 hour</td>
<td>-0.40 (-0.54, -0.26)</td>
<td>0.13 (0.12,0.14)</td>
<td>0.37 (0.58,0.70)</td>
<td>0.32 (0.29,0.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 hours</td>
<td>-0.37 (-0.48, -0.25)</td>
<td>0.11 (0.11,0.12)</td>
<td>0.74 (0.50,0.70)</td>
<td>0.35 (0.31,0.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>-0.26 (-0.46, -0.07)</td>
<td>0.17 (0.15,0.19)</td>
<td>1.57 (0.43,1.33)</td>
<td>0.55 (0.40,0.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimates of $\alpha$ using the OLS regression (3.2.5) and of $\beta$ using the OLS regression (3.2.10), as well as method-of-moments estimates obtained by matching the empirical ACF with the theoretical ACF in our three parametric models, the Cauchy, the Power- and Gamma-$BSS$ models. We used $L = \lceil n^{1/3} \rceil$ lags to estimate the memory parameters $\lambda_{BSS}$, $\beta_{BSS}$, and $\beta_{Cauchy}$. The estimate of $\beta_{BSS}$ is calculated from the estimate of method-of-moments estimate of $\gamma$ of the Power-$BSS$ process; that is, when $\hat{\gamma} > 1$, set $\hat{\beta}_{BSS} = \hat{\gamma}$, otherwise set $\hat{\beta}_{BSS} = 2\hat{\gamma} - 1$.

estimates are still rather small and less than one in all cases. These results strengthen our confidence in the semiparametric estimates of $\beta$, and underscore the strong persistence of volatility.

Similarly, when we estimate the Gamma-$BSS$ model, which does not have the long memory property (cf. Remark 3.3.3), we get estimates of $\lambda_{BSS}$ ranging from 0.0027 to 0.13, which indicates very slow decay of the ACF.

Lastly, in Table 3.2, Panel B, we report results from a similar analysis, but now using raw (i.e., non-logarithmic) volatility as the data. As suggested by Theorem 3.3.1, raw volatility largely inherits the roughness and memory properties of log volatility, cf. Panel A.

The autocorrelation of (log) volatility

We plot the ACFs of the estimated $BSS$ models together with the empirical ACFs of log volatility data in Figure 3.10. We observe that the models fit to the empirical ACFs very well indeed. Although this might not be surprising considering the estimation procedure, it is nonetheless a desirable feature of the $BSS$ models that they have such tractable and flexible autocorrelation functions.

We also observe that there is no discernible difference between the goodness of fit of the Power-$BSS$ model and that of the Gamma-$BSS$ model. If anything, the
CHAPTER 3. DECOUPLING THE SHORT- AND LONG-TERM BEHAVIOR OF STOCHASTIC VOLATILITY

Figure 3.10. Empirical autocorrelation functions of log volatility (light blue bars and circles) and fitted ACFs of the Power-BSS (solid red line) and Gamma-BSS (dashed green line) models. We used \( L = \lceil n^{1/3} \rceil \) lags when fitting the ACFs; this corresponds to the rightmost point of each plot.

latter fits the data slightly better. This indicates that we are not able to confirm that log volatility has the long memory property, as has been sometimes suggested in the literature, by inspecting the fit of the ACF. Indeed, the explanation by Gatheral et al. (2014) that this might simply be "spurious" long memory remains plausible. The model with exponentially decaying (and therefore summable) ACF fits the data at least as well as the model with proper long memory, a property that in fact extends beyond the lags displayed in Figure 3.10.

3.5 Application to individual equities

To see whether the findings above are only valid for the E-mini S&P 500 futures contract, or whether they hold true more generally, we have also studied volatility data on a large number of US equities. The data consist of daily pre-averaged bipower variation measures (i.e. \( BV^{\Delta^*} \) with \( \Delta = 1 \) day), computed using transaction prices obtained from the Trades and Quotes (TAQ) database.

We first study data from 26 of the largest blue-chip companies in the data set. These are analyzed similarly as the E-mini S&P 500 data in the previous sections. Additionally, we analyze a much larger subset of close to five thousand assets in the data set, focusing in particular on the roughness and persistence properties of the volatility estimates.
Estimating the models on volatility data on 26 major equities

We first consider the data on 26 major US equities, identified by their ticker symbols in Table 3.3. The data cover the period from January 2, 2003 to December 31, 2013, excluding weekends and public holidays. For some equities, some days have been discarded due to limited trading during the day. All in all, we have slightly more than 3 000 observations per asset.

Table 3.3 presents the estimates of the parameters, in a similar fashion as in Section 3.4, and the results parallel the earlier findings. Indeed, we find that volatility is rough with an average roughness parameter estimate of $-0.35$. Interestingly, the estimates of $\alpha$ do not seem to vary much from asset to asset, indicating that $\alpha \approx -0.35$ is a reasonable estimate for the roughness index of daily volatility for most equities in our sample. This is also consistent with what was found in Gatheral et al. (2014). Also, like with E-mini S&P 500 data, we see a high degree of persistence in volatility. The average estimate of the parameter $\lambda$ in the Gamma-$\beta$-$\phi$ model is just 0.0038, while the estimates of the memory parameters $\beta_{\text{OLS}}$, $\beta_{\text{BSS}}$, and $\beta_{\text{Cauchy}}$ are 0.13, 0.15, and 0.11, respectively.

Universal roughness and persistence of volatility: evidence from close to five thousand equities

The data set at our disposal runs from January 4, 1993 to December 31, 2013, while the data on some assets might begin later than this start date or end earlier than the end date. In total there are 10 744 assets in the sample, classified into ten industry sectors according to the Global Industry Classification Standard.\(^9\) To make sure that the data are recent, we consider only the period from January 2, 2003 to December 31, 2013. Additionally, to ensure the reliability of the volatility estimates, we retain only the assets that are traded on at least 400 days. After this selection procedure, we are left with 4812 assets.

Since the assets are rather heterogeneous in terms of liquidity, we form a reference group of “liquid assets” satisfying the following conditions:

(a) The asset is traded on at least 400 days.

(b) The maximum number of days on which the asset is not traded (on average) every 5 minutes is 19.

The estimates of the roughness parameter $\alpha$ of volatility for the assets in the sample are summarized in the box plot in Figure 3.11. A few features of this plot are worth highlighting. Firstly, the estimates of $\alpha$ are all negative, mostly in the range $(-0.5, -0.3)$, indicating pronounced roughness. As we would expect, the estimates for the liquid assets are less spread out; they are also slightly less negative, lying mainly

\(^9\)See msci.com/gics.
### Table 3.3. Estimating the models on daily log $B^V$ measures of individual US stocks

<table>
<thead>
<tr>
<th>Asset</th>
<th>$n$</th>
<th>$\hat{a}_{OLS}$</th>
<th>95% CI</th>
<th>$\hat{\lambda}_{BSS} \times 10^2$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{OLS}$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{BSS}$</th>
<th>95% CI</th>
<th>$\hat{\beta}_{Cau}$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>3273</td>
<td>0.35 (−0.37,−0.33)</td>
<td>0.29 (0.27,0.30)</td>
<td>0.11 (−0.94,1.24)</td>
<td>0.11 (0.10,0.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIG</td>
<td>3273</td>
<td>0.31 (−0.34,−0.28)</td>
<td>0.23 (0.21,0.24)</td>
<td>0.09 (−1.18,1.51)</td>
<td>0.08 (0.07,0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXP</td>
<td>3273</td>
<td>0.34 (−0.36,−0.31)</td>
<td>0.09 (0.09,0.09)</td>
<td>0.07 (−1.44,1.75)</td>
<td>0.07 (0.06,0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>3273</td>
<td>0.37 (−0.39,−0.35)</td>
<td>0.35 (0.33,0.37)</td>
<td>0.14 (−0.75,1.04)</td>
<td>0.12 (0.12,0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>3273</td>
<td>0.30 (−0.32,−0.27)</td>
<td>0.21 (0.19,0.23)</td>
<td>0.06 (−1.22,1.55)</td>
<td>0.07 (0.07,0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>3273</td>
<td>0.29 (−0.31,−0.26)</td>
<td>0.32 (0.30,0.35)</td>
<td>0.06 (−1.08,1.41)</td>
<td>0.08 (0.08,0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT</td>
<td>3273</td>
<td>0.36 (−0.38,−0.34)</td>
<td>0.18 (0.18,0.19)</td>
<td>0.13 (−1.10,1.39)</td>
<td>0.10 (0.09,0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVX</td>
<td>3273</td>
<td>0.33 (−0.35,−0.31)</td>
<td>0.46 (0.45,0.48)</td>
<td>0.16 (−0.80,1.11)</td>
<td>0.12 (0.11,0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>3273</td>
<td>0.35 (−0.37,−0.33)</td>
<td>0.38 (0.37,0.39)</td>
<td>0.17 (−0.81,1.12)</td>
<td>0.12 (0.11,0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIS</td>
<td>3273</td>
<td>0.34 (−0.37,−0.31)</td>
<td>0.50 (0.47,0.54)</td>
<td>0.12 (−0.69,1.00)</td>
<td>0.13 (0.13,0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>3273</td>
<td>0.34 (−0.36,−0.32)</td>
<td>0.19 (0.18,0.19)</td>
<td>0.11 (−1.17,1.47)</td>
<td>0.09 (0.08,0.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD</td>
<td>3273</td>
<td>0.36 (−0.38,−0.33)</td>
<td>0.21 (0.20,0.22)</td>
<td>0.10 (−1.05,1.35)</td>
<td>0.10 (0.10,0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td>3273</td>
<td>0.35 (−0.37,−0.33)</td>
<td>0.45 (0.44,0.47)</td>
<td>0.19 (−0.72,1.02)</td>
<td>0.13 (0.12,0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTC</td>
<td>3273</td>
<td>0.33 (−0.35,−0.31)</td>
<td>0.77 (0.71,0.82)</td>
<td>0.17 (−0.48,0.80)</td>
<td>0.14 (0.14,0.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JNJ</td>
<td>3273</td>
<td>0.36 (−0.39,−0.34)</td>
<td>0.45 (0.43,0.47)</td>
<td>0.19 (−0.66,0.95)</td>
<td>0.13 (0.12,0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPM</td>
<td>3273</td>
<td>0.31 (−0.34,−0.27)</td>
<td>0.28 (0.27,0.30)</td>
<td>0.09 (−1.09,1.42)</td>
<td>0.08 (0.08,0.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KO</td>
<td>3273</td>
<td>0.36 (−0.38,−0.33)</td>
<td>0.56 (0.52,0.60)</td>
<td>0.14 (−0.55,0.85)</td>
<td>0.14 (0.13,0.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCD</td>
<td>3273</td>
<td>0.39 (−0.41,−0.37)</td>
<td>0.09 (0.09,0.10)</td>
<td>0.08 (−1.26,1.53)</td>
<td>0.09 (0.09,0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMM</td>
<td>3273</td>
<td>0.34 (−0.37,−0.32)</td>
<td>0.57 (0.54,0.60)</td>
<td>0.18 (−0.60,0.91)</td>
<td>0.13 (0.13,0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRK</td>
<td>3273</td>
<td>0.36 (−0.39,−0.33)</td>
<td>0.52 (0.50,0.55)</td>
<td>0.17 (−0.57,0.86)</td>
<td>0.14 (0.13,0.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSFT</td>
<td>3274</td>
<td>0.35 (−0.37,−0.33)</td>
<td>0.56 (0.54,0.58)</td>
<td>0.19 (−0.58,0.88)</td>
<td>0.14 (0.13,0.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>3273</td>
<td>0.36 (−0.40,−0.33)</td>
<td>0.62 (0.60,0.64)</td>
<td>0.20 (−0.46,0.76)</td>
<td>0.15 (0.14,0.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTX</td>
<td>3273</td>
<td>0.35 (−0.38,−0.33)</td>
<td>0.54 (0.51,0.56)</td>
<td>0.15 (−0.59,0.89)</td>
<td>0.14 (0.13,0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VZ</td>
<td>3273</td>
<td>0.36 (−0.38,−0.33)</td>
<td>0.38 (0.36,0.41)</td>
<td>0.12 (−0.78,1.07)</td>
<td>0.12 (0.12,0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMT</td>
<td>3273</td>
<td>0.38 (−0.40,−0.35)</td>
<td>0.25 (0.24,0.26)</td>
<td>0.14 (−0.89,1.18)</td>
<td>0.11 (0.11,0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XOM</td>
<td>3273</td>
<td>0.33 (−0.35,−0.31)</td>
<td>0.61 (0.59,0.63)</td>
<td>0.20 (−0.65,0.97)</td>
<td>0.13 (0.12,0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td></td>
<td>−0.35</td>
<td>0.38</td>
<td>0.13</td>
<td>0.15</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimates of $\alpha$ using the OLS regression (3.2.5) and of $\beta$ using the OLS regression (3.2.10) as well as method-of-moments estimates obtained by matching the empirical ACF with the theoretical ACF in our three parametric models, the Cauchy, Power- and Gamma-$\mathcal{BFP}$ models. The estimates of $\lambda$ have been multiplied by 100. The data consist of estimated daily log volatility measures on 26 major equities traded on the US market, as indicated by their ticker symbols. The last row, "Avg", provides the average parameter estimate across all assets.
in the interval \((-0.45, -0.35)\), indicating that the estimates for the less liquid assets might be more noisy than those for the liquid assets. Secondly, there does not seem to be significant differences in the roughness estimates across sectors.

Given the findings on time-varying roughness in Section 3.2, we conjecture that there is a connection between how heavily an asset is traded and how rough its volatility is. In particular, we expect the volatility of a highly liquid asset to be smoother than the volatility of a less liquid asset. To investigate this, Figure 3.12 plots the estimates of \(\alpha\) against two measures of liquidity (log daily trading frequency and log daily volume, respectively). Indeed, the fitted lines indicate a slightly increasing pattern, lending support to the conjecture. This finding is intriguing and could possibly be due to the effect of parent orders (also called metaorders), i.e. by the splitting of large trades into many smaller ones. Indeed, as shown in the recent works Jaisson and Rosenbaum (2016) and Euch et al. (2016), such parent orders of the asset can generate roughness in its volatility process. Further investigation into this possible explanation for the findings of this section, as well as for the roughness of volatility in general, would be very interesting but is beyond the scope of the present paper.

Turning now to persistence properties, Figure 3.13 contains box plots of the estimator of \(\beta\). As explained above, the semiparametric estimator of \(\beta\) can be inaccurate, and in such a large data set as studied here, we found very unstable estimates (not reported here). Instead, we therefore consider the parametric estimator \(\hat{\beta}_{\text{Cauchy}}\) based
Figure 3.12. Top: $\hat{\alpha}$ as a function of log(mean number of daily trades). Bottom: $\hat{\alpha}$ as a function of log(mean daily volume). The crosses correspond to individual assets in the data set, with blue color denoting a liquid asset, while the remaining assets are plotted in red. The lines have been fitted by OLS.

on the Cauchy model as detailed in Section 3.4. Again, our analysis confirms the findings seen earlier. Indeed, the estimates show that log volatility is very persistent, with estimates of $\beta$ mainly in the interval (0, 0.4). In particular, there is convincing evidence for very strong persistence in volatility.

### 3.6 Application to volatility forecasting

In this section we apply the $B$-$B$ process and Cauchy models to forecast intraday volatility of the E-mini S&P 500 futures contract, comparing the results with a number of benchmark models. The benchmark models can roughly be divided into three categories:

(i) standard models,

(ii) highly persistent models (possibly with long memory),

(iii) rough volatility models.

\footnote{We obtained similar results using the estimator $\hat{\beta}_{BSS}$ derived from the Power-$B$-$B$ process, but for brevity these results are not reported here.}
3.6. Application to Volatility Forecasting

The category (i) consists of the random walk (RW), autoregressive models (AR) and an ARMA(1, 1) model; (ii) of the (log) heterogeneous autoregressive (log-HAR) model of Corsi (2009) as well as two ARFIMA models; (iii) contains only the rough fractional stochastic volatility (RFSV) model of Gatheral et al. (2014). As for the models suggested in this paper, we consider the Cauchy process and the Power-$\mathcal{B}$-$\mathcal{F}$-$\mathcal{S}$ and Gamma-$\mathcal{B}$-$\mathcal{F}$-$\mathcal{S}$ models. As discussed above, these three processes all decouple long- and short-term behavioral characteristics, so with suitable parameter values, they can be seen as members of both (ii) and (iii).

The ARMA and ARFIMA models were estimated by maximum likelihood. The AR models, as well as the log-HAR model, were estimated by OLS. The log-HAR model we consider is an “intraday version” of the standard log-HAR model: the horizons used in constructing the factors of the model will represent the (intraday) sampling interval $\Delta$, one day, one week, and one month. In this way, the factors of the model will depend on the value of $\Delta$. To be precise, our log-HAR regression is

$$
\log(BV_{t+\Delta h}^\Delta) = a_0 + a_1 \log(BV_t^\Delta) + a_2 \log(BV_t^{\Delta,\text{day}}) + a_3 \log(BV_t^{\Delta,\text{week}}) + a_4 \log(BV_t^{\Delta,\text{month}}) + \epsilon_{t+\Delta h},
$$

(3.6.1)

Figure 3.13. Box plot for $\hat{\beta}_{\text{Cauchy}}$ by sector. Bandwidth $L = \lceil n^{1/3} \rceil$.

11The ARMA model was estimated and forecasted using the MFE toolbox of Kevin Sheppard, see kevinsheppard.com; the ARFIMA models were estimated and forecasted using the MATLAB package “ARFIMA(p,d,q) estimator” available from MATLAB Central.
where

\[ BV^*_{t,x} := \frac{1}{q} \sum_{k=0}^{q-1} BV^*_{t-k\Delta}, \quad x = \text{day, week, month,} \]

and \( q \) is an integer such that \( q\Delta = x \). For instance, when \( \Delta = 65 \) minutes then \( q = 6 \) for \( x = \) day (6 periods of 65 minutes on a trading day) whereas \( q = 30 \) for \( x = \) week (5 trading days, each of 6 periods, in a week), and so on.\(^\text{12}\)

The estimate of the Hurst index \( H \) in the RFSV model is set to \( \hat{H} = \hat{\alpha}_{OLS} + 0.5 \), and for the Cauchy and \( \mathcal{B} \mathcal{F} \mathcal{S} \) processes we use the parametric estimates of \( \beta, \gamma \) or \( \lambda \) along with \( \hat{\alpha}_{OLS} \). Forecasting the AR and log-HAR models is standard. To forecast the RFSV model, we use the the following approximation (cf. Gatheral et al., 2014, Equation (5.1))

\[
\mathbb{E}\left[ \log \sigma^2_{t+h\Delta} | \mathcal{F}_t \right] \\
\approx \frac{\cos(H\pi)}{\pi} (h\Delta)^{H+1/2} \int_{-\infty}^{t} \frac{\log \sigma^2_{s}}{(t-s+h\Delta)(t-s)^{H+1/2}} ds \\
= \frac{\cos(H\pi)}{\pi} (h\Delta)^{H+1/2} \sum_{j=1}^{\infty} \int_{t-j\Delta}^{t} \frac{\log \sigma^2_{s}}{(t-s+h\Delta)(t-s)^{H+1/2}} ds \\
\approx \frac{\cos(H\pi)}{\pi} (h\Delta)^{H+1/2} \sum_{j=1}^{\infty} \log \sigma^2_{t-j\Delta} \int_{t-j\Delta}^{t} \frac{1}{(t-s+h\Delta)(t-s)^{H+1/2}} ds,
\]

where \( \mathcal{F}_t \) is the information set (\( \sigma \)-algebra) generated by the fBm driving the model up to time \( t \). The integrals in this approximation can be expressed using the hypergeometric function (see Appendix A.1) and can therefore be evaluated efficiently using standard software packages. We find that calculating the integrals in the sum, as opposed to approximating them by Riemann sums, turns out to be important for small values of the index \( j \), i.e., for intervals near the singularity of the integrand, as these are the terms that contribute the most to the sum. In our implementation, we therefore calculate the integral explicitly for the first three terms, \( j = 1, 2, 3 \), while for terms \( j \geq 4 \), we simply approximate the integrand by its mid-point value. This idea of explicitly calculating the integrals that contribute the most to the sum, while approximating the remaining terms by Riemann sums, is inspired by the hybrid scheme of Bennedsen et al. (2015), where a similar approach is used to approximate \( \mathcal{B} \mathcal{F} \mathcal{S} \) processes. We note that this refines the approach of Gatheral et al. (2014), who used merely a Riemann-sum approximation of the integral. In the forecasting study below we found that this refinement substantially improved the performance of the forecasts of the RFSV model.

To forecast the Cauchy and \( \mathcal{B} \mathcal{F} \mathcal{S} \) processes we rely on the well known result that for a zero-mean Gaussian vector \( (x_{t+h}, x_t, x_{t-1}, \ldots, x_{t-m})^T \) the distribution of

\(^{12}\)Although not indicated here, in the expressions for the log-HAR model, all pre-averaging estimates of integrated volatility are de-seasonalized when we estimate and forecast the model. The seasonal factor is re-introduced after estimation and forecasting.
3.6. Application to Volatility Forecasting

\[ x_{t+h} \text{ conditionally on } (x_t, x_{t-1}, \ldots, x_{t-m})^T = a \in \mathbb{R}^{m+1} \text{ is} \]

\[ x_{t+h} | \{(x_t, x_{t-1}, \ldots, x_{t-m})^T = a\} \sim N(\mu, \xi^2), \]

where

\[ \mu = \Gamma_{12} \Gamma_{22}^{-1} a, \]

with \( \Gamma_{22} \) being the correlation matrix of the vector \((x_t, x_{t-1}, \ldots, x_{t-m})^T\), and

\[ \Gamma_{12} := \left( \text{Corr}(x_{t+h}, x_t), \text{Corr}(x_{t+h}, x_{t-1}), \ldots, \text{Corr}(x_{t+h}, x_{t-m}) \right). \]

Since the processes we consider here are stationary, the variance \( \xi^2 \) of the conditional distribution is

\[ \xi^2 = \text{Var}(x_t) \left( 1 - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right), \]

where \( \Gamma_{21} = \Gamma_{12}^T \).

To implement this procedure for the Cauchy and BS\( \mathcal{S} \) models, we assume Gaussianity of the process and use these results, where the correlation matrices and vectors above are calculated from the theoretical correlation structure of the process in question, implied by the estimated parameters. When forecasting log volatility, only the conditional mean \( \mu \) needs to be calculated. However, as we will argue in the next section, the conditional variance term \( \xi^2 \) will be important in forecasting raw volatility. These results rely on \( X \) having mean zero, so in our forecasting experiment we de-mean the data before conducting the experiment.\(^{13}\)

**Forecasting intraday integrated variance**

The previous section lays out methods of forecasting log volatility or, equivalently, the process \( X \), cf. Equation (3.3.1). However, it is in practice more relevant to forecast raw volatility. Before presenting the forecasting results for this quantity, we briefly explain our approach.

As we are now interested in \( \mathbb{E}[\exp(X_{t+\Delta})|\mathcal{F}_t] \), instead of \( \exp(\mathbb{E}[X_{t+\Delta}|\mathcal{F}_t]) \), it is worth reminding that it is a flawed strategy to simply forecast log volatility as above and then exponentiate the forecast. Indeed, by Jensen’s inequality we know this approach to be biased. However, we can often correct the exponentiated forecasts following a simple approach. For the BS\( \mathcal{S} \) and Cauchy models we follow the strategy of the preceding section. That is, if we again assume Gaussianity, we have

\[ \mathbb{E}[\exp(X_{t+\Delta})|\mathcal{F}_t] = \exp \left( \mathbb{E}[X_{t+\Delta}|\mathcal{F}_t] + \frac{1}{2} \text{Var}[X_{t+\Delta}|\mathcal{F}_t] \right). \]  \(^{13}\)

\(^{13}\)In our model for volatility, this de-meaning essentially means removing the term \( \log \xi \), cf. equation (3.3.1). We reintroduce this term after forecasting \( X \).
We will then approximate the former term in the exponential function by \( \mu \) and the latter by \( \frac{1}{2} \xi^2 \). Note that \( \xi^2 \) depends on the (stationary) variance of the process, \( \text{Var}(x_t) \); this factor we simply estimate from the (unconditional) variance of the time series being forecasted.

As for the other models, Gatheral et al. (2014) proposed a similar correction to their RFSV model (see Gatheral et al., 2014, Section 5.2), which we use in the following. As the log-HAR model is estimated by OLS using the log realized kernel data, cf. (3.6.1), we exponentiate these estimates and make a correction similar to (3.6.2), where the variance factor is estimated as the variance of the error term in the OLS regression (3.6.1). The remaining models are estimated using directly the raw (de-seasonalized) pre-averaging data \( BV_{\Delta^*} \), so no correction is needed.

We use the methodology described above to forecast integrated variance, as this is most often the object of interest in applications. Since integrated variance is not actually observable, as a feasible forecast object (FO) we use

\[
\text{FO}_t(\Delta, h) := \sum_{k=1}^h BV_{t+k\Delta} \approx \int_t^{t+h\Delta} \sigma_s^2 \, ds, \quad h = 1, 2, 5, 10, 20, \quad (3.6.3)
\]

where \( BV_{t+k\Delta} \) is the estimated value of integrated variance using the pre-averaged bipower variation estimator, cf. Section 3.2.

To forecast the FO in (3.6.3), we compute the \( h \) individual components, \( \sigma_{t+k\Delta|t} \), \( k = 1, \ldots, h \), multiply by \( \Delta \) and the seasonal component, and sum them up:

\[
\hat{\text{FO}}_t(\Delta, h) = \sum_{k=1}^h \sigma^2_{t+k\Delta|t} \left( \sigma^2_{t+k\Delta|t} \right) \Delta,
\]

where \( \sigma^2_t \) is the (deterministic) seasonal component of volatility we extracted in the preliminary step of our analysis, as explained in Section 3.2, and \( \sigma^2_{t+k\Delta|t} \) is the forecast of volatility, as detailed above.\(^{14}\)

In the forecasting experiments we consider various step sizes \( \Delta \), ranging from 15 minutes to 1 day, and various forecast horizons \( h \in \{1, 2, 5, 10, 20\} \). We start the estimation after an initial period of \( m \in \mathbb{N} \) time steps and compare the performance of the forecasts using two different loss functions:

- **Mean Squared Error:**
  \[
  \text{MSE}(\Delta, h) = \frac{1}{n-h-m+1} \sum_{t=m}^{n-h} \left| \hat{\text{FO}}_t(\Delta, h) - \text{FO}_t(\Delta, h) \right|^2,
  \]

- **QLIKE**:\(^{14}\)
  \[
  \text{QLIKE}(\Delta, h) = \frac{1}{n-h-m+1} \sum_{t=m}^{n-h} \left( \log \hat{\text{FO}}_t(\Delta, h) + \frac{\text{FO}_t(\Delta, h)}{\text{FO}_t(\Delta, h)} \right).
  \]

\(^{14}\)To keep the setup realistic, in the out-of-sample exercise below, the seasonal component is estimated using only data prior to the time point at which a forecast is constructed.
3.6. Application to Volatility Forecasting

As discussed in Section 3.2, the pre-averaging estimate $BV^{\Delta^*}$, our FO, is a noisy estimate of integrated variance, but Patton (2011) shows that the MSE and QLIKE loss functions still yield consistent rankings of the forecasting models even for integrated variance, in spite of the noisy estimates used to evaluate the loss functions. We calculate MSE, QLIKE, and also the model confidence set (MCS) of Hansen, Lunde, and Nason (2011), which is a procedure to construct a “set of best models” with a certain probability, as measured by the specific loss function in question, that avoids the problems that arise from doing multiple comparisons by pairwise tests. For instance, the best model is contained in the 90% MCS — when understood as a random set — with 90% probability. In the tables below, the models included in the MCS are denoted by grey background; the dark grey color corresponds to models in the 75% MCS, while the light grey color corresponds to models in the 90% MCS. The model that minimizes the loss is given in bold.

In-sample forecasting experiment

We first report results from an in-sample forecasting experiment. That is, we use the entire E-mini S&P 500 data set to estimate the parameters of the models and then forecast integrated variance as described above. We use the above loss functions and an initial period of $m = 200$ time steps (to allow for prior values for the HAR, ARFIMA, RFSV, BJSF, and Cauchy models, i.e., the non-Markovian models).

The results of the in-sample forecasting exercise are presented in Tables 3.5 and 3.6. It is evident that at intraday time scales (i.e., $\Delta < 1$ day), our models considerably outperform the benchmark models. Indeed, in almost all cases, one of our models has the smallest loss and one or more of them are almost always in the 75% MCS. The log-HAR model also fares fairly well, but is not as consistently in the MCS as our proposed models. For $\Delta = 1$ day, the picture is a little different. Although our models still perform best, except over the very longest forecast horizon, where the conventional long memory models (log-HAR, ARFIMA(1, $d$, 1)) perform slightly better, almost all models considered are at least in the 90% MCS – and many of them even in the 75% MCS. The reason for this is likely because of the low number of forecast periods when $\Delta = 1$ day (less than 300), making it is difficult to reach any definite conclusions in this case.

Out-of-sample forecasting experiment

We consider also a forecasting scenario that is undoubtedly more relevant in practice. At time $t$, we use only the data observed so far to estimate the models and forecast $h$ steps ahead, for $h = 1, 2, 5, 10, 20$. We then move one step forward in time, to $t + \Delta$, re-estimate the models and compute new forecasts $h$ steps ahead. We start forecasting after $m$ time steps and use $N$ observations, where $N \leq m$, at each step to estimate the models, to allow for time-varying parameter values. The values of $m$ and $N$ we
For step size $\Delta$, we have $n$ observations in total. We begin forecasting at time $t = m\Delta$ and the loss functions are calculated using $n - m - h + 1 \approx n - m$ forecasts, $h \in \{1, 2, 5, 10, 20\}$. We use a rolling window of $N$ observations for estimation at each time $t$ to forecast over period $t + h$. The final date for the forecasting exercise is December 31, 2014; the start date is $t = m\Delta$ after January 2, 2013.

The results of the out-of-sample forecasting exercise, given in Tables 3.7 and 3.8, corroborate the findings from the in-sample exercise to a very large degree. Indeed, our models again outperform the benchmark models in almost all cases; in fact, for the intraday frequencies ($\Delta < 1$ day) either the Cauchy or the Power-$\mathcal{B} \mathcal{I} \mathcal{S}$ model has the smallest loss value for all horizons and both loss metrics. All three of our candidate models are often in the MCS, where especially the Power-$\mathcal{B} \mathcal{I} \mathcal{S}$ model is impressive since it is almost always in the 75% MCS. None of the benchmarks are consistently in the MCS. The results for $\Delta = 1$ day show that the Gamma-$\mathcal{B} \mathcal{I} \mathcal{S}$ generally perform best for this value of $\Delta$, although here the RFSV model also performs quite well. The good performance of the RFSV model at the daily frequency is consistent with the results of the forecasting exercise in Gatheral et al. (2014). (As in the in-sample experiment, the MCS at both 75% and 90% confidence levels is not very selective in the case $\Delta = 1$ day, due to the relatively low number of forecast periods.)

From this exercise, we conclude that when forecasting, it is beneficial to consider a model with both roughness and persistence. When comparing our three candidate rough models in terms of the MCS, it seems that, at least for the intraday sampling periods, there is a slight improvement from considering a model with genuine long memory (the Cauchy and Power-$\mathcal{B} \mathcal{I} \mathcal{S}$ models), as opposed to technically short, but still persistent memory (Gamma-$\mathcal{B} \mathcal{I} \mathcal{S}$), although any gains in terms of the actual loss values are rather small.
3.7 Conclusions and further research

In this paper, we have presented a thorough investigation of the empirical characteristics of volatility, focusing especially on intraday time scales. Having examined intraday volatility measurements on the E-mini S&P 500 futures contract, we can conclude that volatility is *rough, highly persistent, and non-Gaussian.*

Moreover, by also looking at volatility measurements on more than five thousand individual US equities, we corroborated these findings, suggesting that both roughness and strong persistence are universal features of financial market volatility.

We have also presented mathematical models that are able to capture the key empirical features we find in the data. In particular, we advocate using a stochastic process that decouples short- and long-term behavior to model log volatility. Our results indicate that the Brownian semistationary process is an ideal model for this purpose. We illustrated one of the practical advantages of such a model in a forecasting experiment and found that the model, with just two parameters controlling the short- and long-term behavior, respectively, outperforms the benchmark models in almost all scenarios.

We believe that the models we have presented here can be utilized in a wide range of applications, beyond the ones seen in this paper. For instance, we think that they should be useful in some areas of pricing and hedging of financial derivatives, as well as in forecasting (intraday) value-at-risk. We leave such extensions for future work, but remark that related work has already begun in Bayer et al. (2016), where the authors study option pricing and smile modeling under rough volatility models, i.e., models closely related to the ones considered in the present work.
### Table 3.5. In-sample forecasting of intraday integrated variance

**Panel A: Δ = 15 minutes**

<table>
<thead>
<tr>
<th></th>
<th>$h = 1$</th>
<th></th>
<th>$h = 2$</th>
<th></th>
<th>$h = 5$</th>
<th></th>
<th>$h = 10$</th>
<th></th>
<th>$h = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
</tr>
<tr>
<td>RW</td>
<td>$\times 10^{11}$</td>
<td>-12.873</td>
<td>0.832</td>
<td>-12.144</td>
<td>0.576</td>
<td>-11.146</td>
<td>0.379</td>
<td>-10.366</td>
<td>0.156</td>
</tr>
<tr>
<td>AR5</td>
<td></td>
<td>0.168</td>
<td></td>
<td>0.487</td>
<td></td>
<td>0.256</td>
<td></td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>AR10</td>
<td></td>
<td>0.174</td>
<td></td>
<td>0.506</td>
<td></td>
<td>0.252</td>
<td></td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td></td>
<td>0.174</td>
<td></td>
<td>0.509</td>
<td></td>
<td>0.254</td>
<td></td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>log-HAR3</td>
<td></td>
<td>0.158</td>
<td>-12.930</td>
<td>0.448</td>
<td>-12.216</td>
<td>0.296</td>
<td>-11.254</td>
<td>0.064</td>
<td>-10.513</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td></td>
<td>0.168</td>
<td>-12.934</td>
<td>0.485</td>
<td>-12.218</td>
<td>0.234</td>
<td>-11.254</td>
<td>0.080</td>
<td>-10.511</td>
</tr>
<tr>
<td>ARFIMA(1,d,1)</td>
<td></td>
<td>0.169</td>
<td>-12.934</td>
<td>0.491</td>
<td>-12.219</td>
<td>0.241</td>
<td>-11.256</td>
<td>0.085</td>
<td>-10.513</td>
</tr>
<tr>
<td>RFSV</td>
<td></td>
<td>0.164</td>
<td>-12.928</td>
<td>0.489</td>
<td>-12.213</td>
<td>0.251</td>
<td>-11.250</td>
<td>0.090</td>
<td>-10.506</td>
</tr>
<tr>
<td>Cauchy</td>
<td></td>
<td>0.161</td>
<td>-12.935</td>
<td>0.461</td>
<td>-12.221</td>
<td>0.209</td>
<td>-11.259</td>
<td>0.066</td>
<td>-10.517</td>
</tr>
<tr>
<td>Power-BSS</td>
<td></td>
<td>0.159</td>
<td>-12.937</td>
<td>0.453</td>
<td>-12.222</td>
<td>0.210</td>
<td>-11.259</td>
<td>0.067</td>
<td>-10.517</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td></td>
<td>0.158</td>
<td>-12.936</td>
<td>0.454</td>
<td>-12.221</td>
<td>0.211</td>
<td>-11.256</td>
<td>0.068</td>
<td>-10.516</td>
</tr>
</tbody>
</table>

**Panel B: Δ = 30 minutes**

<table>
<thead>
<tr>
<th></th>
<th>$h = 1$</th>
<th></th>
<th>$h = 2$</th>
<th></th>
<th>$h = 5$</th>
<th></th>
<th>$h = 10$</th>
<th></th>
<th>$h = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
<td>MSE</td>
<td>QLIKE</td>
</tr>
<tr>
<td>RW</td>
<td>$\times 10^{11}$</td>
<td>-12.169</td>
<td>0.285</td>
<td>-11.428</td>
<td>0.307</td>
<td>-10.409</td>
<td>0.101</td>
<td>-9.631</td>
<td>0.418</td>
</tr>
<tr>
<td>AR5</td>
<td></td>
<td>0.546</td>
<td></td>
<td>0.179</td>
<td></td>
<td>0.093</td>
<td></td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td>AR10</td>
<td></td>
<td>0.519</td>
<td>-12.209</td>
<td>0.164</td>
<td>-11.484</td>
<td>0.081</td>
<td>-10.504</td>
<td>0.025</td>
<td>-9.758</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td></td>
<td>0.558</td>
<td>-12.207</td>
<td>0.183</td>
<td>-11.482</td>
<td>0.098</td>
<td>-10.500</td>
<td>0.030</td>
<td>-9.744</td>
</tr>
<tr>
<td>log-HAR3</td>
<td></td>
<td>0.495</td>
<td>-12.213</td>
<td>0.152</td>
<td>-11.490</td>
<td>0.074</td>
<td>-10.511</td>
<td>0.024</td>
<td>-9.764</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td></td>
<td>0.534</td>
<td>-12.211</td>
<td>0.171</td>
<td>-11.486</td>
<td>0.086</td>
<td>-10.506</td>
<td>0.027</td>
<td>-9.756</td>
</tr>
<tr>
<td>ARFIMA(1,d,1)</td>
<td></td>
<td>0.535</td>
<td>-12.210</td>
<td>0.172</td>
<td>-11.485</td>
<td>0.087</td>
<td>-10.504</td>
<td>0.027</td>
<td>-9.756</td>
</tr>
<tr>
<td>RFSV</td>
<td></td>
<td>0.553</td>
<td>-12.208</td>
<td>0.188</td>
<td>-11.482</td>
<td>0.110</td>
<td>-10.499</td>
<td>0.033</td>
<td>-9.749</td>
</tr>
<tr>
<td>Cauchy</td>
<td></td>
<td>0.502</td>
<td>-12.215</td>
<td>0.153</td>
<td>-11.491</td>
<td>0.070</td>
<td>-10.514</td>
<td>0.022</td>
<td>-9.769</td>
</tr>
<tr>
<td>Power-BSS</td>
<td></td>
<td>0.500</td>
<td>-12.215</td>
<td>0.155</td>
<td>-11.490</td>
<td>0.072</td>
<td>-10.513</td>
<td>0.022</td>
<td>-9.767</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td></td>
<td>0.504</td>
<td>-12.214</td>
<td>0.157</td>
<td>-11.489</td>
<td>0.074</td>
<td>-10.510</td>
<td>0.023</td>
<td>-9.764</td>
</tr>
</tbody>
</table>

In-sample Mean Squared Forecast Error (MSE) and QLIKE for all models considered in the paper. Bold numbers indicate the models with the smallest forecast error (column-wise). The forecast object is the sum of realized kernels (3.6.3), approximating integrated variance, as explained in the text. We vary the step size $h$ and the forecast horizon $\Delta$. Grey cells indicate models which are in the Model Confidence Set (column-wise); the dark grey denotes the 75% MCS, while the light grey denotes the 90% MCS. The MCS uses a block bootstrap method with 25,000 bootstrap replications and a block length of 6 time steps.
### Table 3.6. In-sample forecasting of intraday integrated variance

<table>
<thead>
<tr>
<th>Panel A: ( \Delta = 65 ) minutes</th>
<th>( h = 1 )</th>
<th>( h = 2 )</th>
<th>( h = 5 )</th>
<th>( h = 10 )</th>
<th>( h = 20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE ( \times 10^{10} )</td>
<td>QLIKE</td>
<td>MSE ( \times 10^{9} )</td>
<td>QLIKE</td>
<td>MSE ( \times 10^8 )</td>
</tr>
<tr>
<td>RW</td>
<td>0.422</td>
<td>−11.324</td>
<td>0.254</td>
<td>−10.558</td>
<td>0.104</td>
</tr>
<tr>
<td>AR5</td>
<td>0.256</td>
<td>−11.363</td>
<td>0.083</td>
<td>−10.623</td>
<td>0.036</td>
</tr>
<tr>
<td>AR10</td>
<td>0.253</td>
<td>−11.364</td>
<td>0.082</td>
<td>−10.624</td>
<td>0.035</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.283</td>
<td>−11.361</td>
<td>0.090</td>
<td>−10.623</td>
<td>0.037</td>
</tr>
<tr>
<td>log-HAR3</td>
<td>0.245</td>
<td>−11.373</td>
<td>0.081</td>
<td>−10.632</td>
<td>0.034</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td>0.206</td>
<td>−11.364</td>
<td>0.087</td>
<td>−10.623</td>
<td>0.037</td>
</tr>
<tr>
<td>ARFIMA(1,d,1)</td>
<td>0.209</td>
<td>−11.364</td>
<td>0.087</td>
<td>−10.624</td>
<td>0.036</td>
</tr>
<tr>
<td>RFSV</td>
<td>0.288</td>
<td>−11.367</td>
<td>0.105</td>
<td>−10.620</td>
<td>0.041</td>
</tr>
<tr>
<td>Cauchy</td>
<td>0.235</td>
<td>−11.372</td>
<td>0.070</td>
<td>−10.633</td>
<td>0.031</td>
</tr>
<tr>
<td>Power-BSS</td>
<td>0.237</td>
<td>−11.373</td>
<td>0.071</td>
<td>−10.633</td>
<td>0.031</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td>0.242</td>
<td>−11.372</td>
<td>0.074</td>
<td>−10.630</td>
<td>0.031</td>
</tr>
</tbody>
</table>

| Panel B: \( \Delta = 1 \) day |

<table>
<thead>
<tr>
<th></th>
<th>( h = 1 )</th>
<th>( h = 10 )</th>
<th>( h = 20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE ( \times 10^9 )</td>
<td>QLIKE</td>
<td>MSE ( \times 10^7 )</td>
</tr>
<tr>
<td>RW</td>
<td>0.782</td>
<td>−9.525</td>
<td>0.329</td>
</tr>
<tr>
<td>AR5</td>
<td>0.640</td>
<td>−9.582</td>
<td>0.234</td>
</tr>
<tr>
<td>AR10</td>
<td>0.629</td>
<td>−9.580</td>
<td>0.231</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.648</td>
<td>−9.584</td>
<td>0.238</td>
</tr>
<tr>
<td>log-HAR3</td>
<td>0.661</td>
<td>−9.586</td>
<td>0.228</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td>0.670</td>
<td>−9.583</td>
<td>0.246</td>
</tr>
<tr>
<td>ARFIMA(1,d,1)</td>
<td>0.638</td>
<td>−9.582</td>
<td>0.232</td>
</tr>
<tr>
<td>RFSV</td>
<td>0.657</td>
<td>−9.578</td>
<td>0.242</td>
</tr>
<tr>
<td>Cauchy</td>
<td>0.761</td>
<td>−9.580</td>
<td>0.259</td>
</tr>
<tr>
<td>Power-BSS</td>
<td>0.714</td>
<td>−9.587</td>
<td>0.245</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td>0.658</td>
<td>−9.591</td>
<td>0.225</td>
</tr>
</tbody>
</table>

In-sample Mean Squared Forecast Error (MSE) and QLIKE for all models considered in the paper. Bold numbers indicate the models with the smallest forecast error (column-wise). The forecast object is the sum of realized kernels (3.6.3), approximating integrated variance, as explained in the text. We vary the step size \( \Delta \) and the forecast horizon \( h \). Grey cells indicate models which are in the Model Confidence Set (column-wise); the dark grey denotes the 75% MCS, while the light grey denotes the 90% MCS. The MCS uses a block bootstrap method with 25 000 bootstrap replications and a block length of 6 time steps.
Table 3.7. Out-of-sample forecasting of intraday integrated variance

Panel A: $\Delta = 15$ minutes

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE $\times 10^{11}$</th>
<th>QLIKE</th>
<th>MSE $\times 10^{10}$</th>
<th>QLIKE</th>
<th>MSE $\times 10^{9}$</th>
<th>QLIKE</th>
<th>MSE $\times 10^{8}$</th>
<th>QLIKE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>0.243</td>
<td>−12.861</td>
<td>0.839</td>
<td>−12.132</td>
<td>0.600</td>
<td>−11.132</td>
<td>0.412</td>
<td>−10.354</td>
</tr>
<tr>
<td>AR5</td>
<td>0.200</td>
<td>−12.903</td>
<td>0.558</td>
<td>−12.174</td>
<td>0.290</td>
<td>−11.219</td>
<td>0.108</td>
<td>−10.473</td>
</tr>
<tr>
<td>AR10</td>
<td>0.221</td>
<td>−12.905</td>
<td>0.615</td>
<td>−12.186</td>
<td>0.311</td>
<td>−11.223</td>
<td>0.109</td>
<td>−10.478</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.185</td>
<td>−12.905</td>
<td>0.337</td>
<td>−12.186</td>
<td>0.274</td>
<td>−11.221</td>
<td>0.107</td>
<td>−10.472</td>
</tr>
<tr>
<td>log-HAR3</td>
<td>0.164</td>
<td>−12.914</td>
<td>0.481</td>
<td>−12.198</td>
<td>0.254</td>
<td>−11.232</td>
<td>0.105</td>
<td>−10.485</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td>0.165</td>
<td>−12.911</td>
<td>0.477</td>
<td>−12.195</td>
<td>0.228</td>
<td>−11.232</td>
<td>0.076</td>
<td>−10.491</td>
</tr>
<tr>
<td>RFSV</td>
<td>0.166</td>
<td>−12.916</td>
<td>0.494</td>
<td>−12.200</td>
<td>0.260</td>
<td>−11.236</td>
<td>0.097</td>
<td>−10.494</td>
</tr>
<tr>
<td>Cauchy</td>
<td>0.163</td>
<td>−12.921</td>
<td>0.466</td>
<td>−12.206</td>
<td>0.214</td>
<td>−11.245</td>
<td>0.069</td>
<td>−10.505</td>
</tr>
<tr>
<td>Power-BSS</td>
<td><strong>0.160</strong></td>
<td>−12.924</td>
<td><strong>0.459</strong></td>
<td>−12.209</td>
<td><strong>0.215</strong></td>
<td>−11.247</td>
<td><strong>0.070</strong></td>
<td>−10.506</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td><strong>0.160</strong></td>
<td>−12.924</td>
<td><strong>0.460</strong></td>
<td>−12.209</td>
<td><strong>0.217</strong></td>
<td>−11.246</td>
<td><strong>0.071</strong></td>
<td>−10.505</td>
</tr>
</tbody>
</table>

Panel B: $\Delta = 30$ minutes

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE $\times 10^{11}$</th>
<th>QLIKE</th>
<th>MSE $\times 10^{10}$</th>
<th>QLIKE</th>
<th>MSE $\times 10^{9}$</th>
<th>QLIKE</th>
<th>MSE $\times 10^{8}$</th>
<th>QLIKE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>0.749</td>
<td>−12.145</td>
<td>0.304</td>
<td>−11.407</td>
<td>0.351</td>
<td>−10.388</td>
<td>0.114</td>
<td>−9.613</td>
</tr>
<tr>
<td>AR5</td>
<td>0.647</td>
<td>−12.173</td>
<td>0.220</td>
<td>−11.448</td>
<td>0.114</td>
<td>−10.466</td>
<td>0.034</td>
<td>−9.720</td>
</tr>
<tr>
<td>AR10</td>
<td>0.653</td>
<td>−12.173</td>
<td>0.214</td>
<td>−11.450</td>
<td>0.106</td>
<td>−10.471</td>
<td>0.033</td>
<td>−9.726</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.567</td>
<td>−12.174</td>
<td>0.197</td>
<td>−11.447</td>
<td>0.114</td>
<td>−10.460</td>
<td>0.036</td>
<td>−9.705</td>
</tr>
<tr>
<td>log-HAR3</td>
<td><strong>0.531</strong></td>
<td>−12.188</td>
<td><strong>0.171</strong></td>
<td>−11.464</td>
<td><strong>0.095</strong></td>
<td>−10.481</td>
<td><strong>0.037</strong></td>
<td>−9.731</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td>0.544</td>
<td>−12.183</td>
<td>0.172</td>
<td>−11.459</td>
<td>0.085</td>
<td>−10.481</td>
<td>0.026</td>
<td>−9.738</td>
</tr>
<tr>
<td>RFSV</td>
<td>0.574</td>
<td>−12.187</td>
<td>0.200</td>
<td>−11.462</td>
<td>0.128</td>
<td>−10.479</td>
<td>0.040</td>
<td>−9.731</td>
</tr>
<tr>
<td>Cauchy</td>
<td><strong>0.525</strong></td>
<td>−12.183</td>
<td><strong>0.161</strong></td>
<td>−11.470</td>
<td><strong>0.076</strong></td>
<td>−10.494</td>
<td><strong>0.024</strong></td>
<td>−9.751</td>
</tr>
<tr>
<td>Power-BSS</td>
<td><strong>0.519</strong></td>
<td>−12.195</td>
<td><strong>0.163</strong></td>
<td>−11.471</td>
<td><strong>0.079</strong></td>
<td>−10.493</td>
<td><strong>0.024</strong></td>
<td>−9.749</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td><strong>0.522</strong></td>
<td>−12.194</td>
<td><strong>0.164</strong></td>
<td>−11.469</td>
<td><strong>0.082</strong></td>
<td>−10.490</td>
<td><strong>0.025</strong></td>
<td>−9.745</td>
</tr>
</tbody>
</table>

Out-of-sample Mean Squared Forecast Error (MSE) and QLIKE for all models considered in the paper. Bold numbers indicate the model with the smallest forecast error (column-wise). The forecast object is the sum of realized kernels (3.6.3), approximating integrated variance, as explained in the text. We vary the step size $\Delta$ and the forecast horizon $h$. Grey cells indicate models which are in the Model Confidence Set (column-wise); the dark grey denotes the 75% MCS, while the light grey denotes the 90% MCS. The MCS uses a block bootstrap method with 25 000 bootstrap replications and a block length of 6 time steps.
Table 3.8. Out-of-sample forecasting of intraday integrated variance

Panel A: $\Delta = 65$ minutes

<table>
<thead>
<tr>
<th></th>
<th>$h = 1$</th>
<th></th>
<th>$h = 2$</th>
<th></th>
<th>$h = 5$</th>
<th></th>
<th>$h = 10$</th>
<th></th>
<th>$h = 20$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE $\times 10^6$</td>
<td>QLIKE</td>
<td>MSE $\times 10^3$</td>
<td>QLIKE</td>
<td>MSE $\times 10^3$</td>
<td>QLIKE</td>
<td>MSE $\times 10^5$</td>
<td>QLIKE</td>
<td>MSE $\times 10^7$</td>
<td>QLIKE</td>
</tr>
<tr>
<td>RW</td>
<td>0.468</td>
<td>-11.318</td>
<td>0.284</td>
<td>-10.553</td>
<td>0.114</td>
<td>-9.540</td>
<td>0.489</td>
<td>-8.812</td>
<td>0.200</td>
<td>-8.072</td>
</tr>
<tr>
<td>AR5</td>
<td>0.321</td>
<td>-11.341</td>
<td>0.104</td>
<td>-10.605</td>
<td>0.045</td>
<td>-9.625</td>
<td>0.191</td>
<td>-8.914</td>
<td>0.094</td>
<td>-8.195</td>
</tr>
<tr>
<td>AR10</td>
<td>0.336</td>
<td>-11.280</td>
<td>0.106</td>
<td>-10.583</td>
<td>0.046</td>
<td>-9.621</td>
<td>0.216</td>
<td>-8.916</td>
<td>0.114</td>
<td>-8.195</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.296</td>
<td>-11.342</td>
<td>0.105</td>
<td>-10.604</td>
<td>0.044</td>
<td>-9.619</td>
<td>0.182</td>
<td>-8.899</td>
<td>0.082</td>
<td>-8.170</td>
</tr>
<tr>
<td>log-HAR3</td>
<td>0.266</td>
<td>-11.364</td>
<td>0.092</td>
<td>-10.620</td>
<td>0.044</td>
<td>-9.633</td>
<td>0.212</td>
<td>-8.911</td>
<td>0.123</td>
<td>-8.165</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td>0.259</td>
<td>-11.351</td>
<td>0.082</td>
<td>-10.615</td>
<td>0.035</td>
<td>-9.639</td>
<td>0.135</td>
<td>-8.927</td>
<td>0.054</td>
<td>-8.208</td>
</tr>
<tr>
<td>RFSV</td>
<td>0.305</td>
<td>-11.362</td>
<td>0.116</td>
<td>-10.618</td>
<td>0.046</td>
<td>-9.633</td>
<td>0.184</td>
<td>-8.928</td>
<td>0.079</td>
<td>-8.211</td>
</tr>
<tr>
<td>Cauchy</td>
<td>0.243</td>
<td>-11.367</td>
<td>0.073</td>
<td>-10.631</td>
<td>0.032</td>
<td>-9.653</td>
<td>0.120</td>
<td>-8.942</td>
<td>0.048</td>
<td>-8.220</td>
</tr>
<tr>
<td>Power-BSS</td>
<td>0.244</td>
<td>-11.369</td>
<td>0.075</td>
<td>-10.631</td>
<td>0.032</td>
<td>-9.652</td>
<td>0.120</td>
<td>-8.941</td>
<td>0.048</td>
<td>-8.218</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td>0.249</td>
<td>-11.367</td>
<td>0.077</td>
<td>-10.628</td>
<td>0.033</td>
<td>-9.648</td>
<td>0.124</td>
<td>-8.937</td>
<td>0.051</td>
<td>-8.212</td>
</tr>
</tbody>
</table>

Panel B: $\Delta = 1$ day

<table>
<thead>
<tr>
<th></th>
<th>$h = 1$</th>
<th></th>
<th>$h = 2$</th>
<th></th>
<th>$h = 5$</th>
<th></th>
<th>$h = 10$</th>
<th></th>
<th>$h = 20$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE $\times 10^9$</td>
<td>QLIKE</td>
<td>MSE $\times 10^9$</td>
<td>QLIKE</td>
<td>MSE $\times 10^7$</td>
<td>QLIKE</td>
<td>MSE $\times 10^6$</td>
<td>QLIKE</td>
<td>MSE $\times 10^5$</td>
<td>QLIKE</td>
</tr>
<tr>
<td>RW</td>
<td>0.078</td>
<td>-9.525</td>
<td>0.329</td>
<td>-8.820</td>
<td>0.222</td>
<td>-7.846</td>
<td>0.101</td>
<td>-7.059</td>
<td>0.047</td>
<td>-6.191</td>
</tr>
<tr>
<td>AR5</td>
<td>0.100</td>
<td>-9.567</td>
<td>0.483</td>
<td>-8.863</td>
<td>0.266</td>
<td>-7.911</td>
<td>0.076</td>
<td>-7.163</td>
<td>0.018</td>
<td>-6.405</td>
</tr>
<tr>
<td>AR10</td>
<td>0.102</td>
<td>-8.179</td>
<td>0.500</td>
<td>-7.408</td>
<td>0.274</td>
<td>-7.906</td>
<td>0.078</td>
<td>-7.159</td>
<td>0.019</td>
<td>-6.402</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.085</td>
<td>-9.574</td>
<td>0.384</td>
<td>-8.867</td>
<td>0.359</td>
<td>-7.913</td>
<td>0.309</td>
<td>-7.163</td>
<td>0.237</td>
<td>-6.396</td>
</tr>
<tr>
<td>log-HAR3</td>
<td>0.066</td>
<td>-9.381</td>
<td>0.236</td>
<td>-8.877</td>
<td>0.126</td>
<td>-7.923</td>
<td>0.043</td>
<td>-7.175</td>
<td>0.014</td>
<td>-6.397</td>
</tr>
<tr>
<td>ARFIMA(0,d,0)</td>
<td>0.071</td>
<td>-9.574</td>
<td>0.255</td>
<td>-8.866</td>
<td>0.139</td>
<td>-7.910</td>
<td>0.051</td>
<td>-7.160</td>
<td>0.017</td>
<td>-6.390</td>
</tr>
<tr>
<td>RFSV</td>
<td>0.064</td>
<td>-9.581</td>
<td>0.232</td>
<td>-8.881</td>
<td>0.126</td>
<td>-7.932</td>
<td>0.048</td>
<td>-7.185</td>
<td>0.019</td>
<td>-6.403</td>
</tr>
<tr>
<td>Cauchy</td>
<td>0.077</td>
<td>-9.575</td>
<td>0.262</td>
<td>-8.867</td>
<td>0.131</td>
<td>-7.915</td>
<td>0.044</td>
<td>-7.169</td>
<td>0.014</td>
<td>-6.397</td>
</tr>
<tr>
<td>Power-BSS</td>
<td>0.072</td>
<td>-9.583</td>
<td>0.248</td>
<td>-8.875</td>
<td>0.126</td>
<td>-7.921</td>
<td>0.043</td>
<td>-7.174</td>
<td>0.014</td>
<td>-6.404</td>
</tr>
<tr>
<td>Gamma-BSS</td>
<td>0.067</td>
<td>-9.587</td>
<td>0.232</td>
<td>-8.882</td>
<td>0.119</td>
<td>-7.928</td>
<td>0.041</td>
<td>-7.182</td>
<td>0.013</td>
<td>-6.414</td>
</tr>
</tbody>
</table>

Out-of-sample Mean Squared Forecast Error (MSE) and QLIKE for all models considered in the paper. Bold numbers indicate the model with the smallest forecast error (column-wise). The forecast object is the sum of realized kernels (3.6.3), approximating integrated variance, as explained in the text. We vary the step size $\Delta$ and the forecast horizon $h$. Grey cells indicate models which are in the Model Confidence Set (column-wise); the dark grey denotes the 75% MCS, while the light grey denotes the 90% MCS. The MCS uses a block bootstrap method with 25 000 bootstrap replications and a block length of 6 time steps.
Acknowledgements

The results of this paper have been presented at the 9th Annual SoFiE Conference in Hong Kong in June 2016, at the ICMS workshop At the Frontiers of Quantitative Finance in Edinburgh in June 2016, and at the 9th Bachelier World Congress in New York in July 2016; and we thank the audiences for stimulating comments and questions. Our research has been partially supported by CREATES (DNRF78), funded by the Danish National Research Foundation, by Aarhus University Research Foundation (project “Stochastic and Econometric Analysis of Commodity Markets”) and by the Academy of Finland (project 258042).

3.8 References


CHAPTER 3. DECODING THE SHORT- AND LONG-TERM BEHAVIOR OF
STOCHASTIC VOLATILITY


dependencies: The role of high-frequency data. Journal of Econometrics 98 (1),
81–106.

application to high-frequency jump-testing. Journal of Financial Econometrics
10 (2), 356–389.

Christensen, K., Oomen, R. C. A., Podolskij, M., 2014. Fact or friction: Jumps at ultra


Comte, F., Renault, E., 1996. Long memory continuous time models. Journal of Econo-
metrics 73 (1), 101–149.

Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Jour-


Euch, O. E., Fukasawa, M., Rosenbaum, M., 2016. The microstructural foundations
org/abs/1609.05177.


Gatheral, J., Jaisson, T., Rosenbaum, M., 2014. Volatility is rough, working paper,

Gneiting, T., Schlather, M., 2004. Stochastic models that separate fractal dimension

Gradshteyn, I. S., Ryzhik, I. M., 2007. Table of integrals, series, and products, seventh
3.8. REFERENCES


A.1 Evaluation of the RFSV integral

For any $H \in (0, 1/2)$, we need to evaluate numerically integrals of the form

$$
\int_{u}^{v} \frac{1}{(y + h\Delta) y^{H + 1/2}} dy,
$$

(A.1)

where $\Delta > 0$, $h \in \mathbb{N}$, and $u \geq v$. We may set $u = 0$, as the value with $u \neq 0$ can be computed as difference of two integrals of the form (A.1) with $u = 0$. According to Gradshteyn and Ryzhik (2007, formula 3.194.5), we get

$$
\int_{0}^{v} \frac{1}{(y + h\Delta) y^{H + 1/2}} dy = (h\Delta)^{-1} \frac{\nu^{1/2-H}}{1/2 - H} 2F_1 \left( 1, 1/2 - H; 3/2 - H; -\nu/(h\Delta) \right),
$$

where $2F_1$ is the (Gauss) hypergeometric function.

A.2 Proofs

Proofs of Propositions 3.3.2 and 3.3.3

**Proof of Proposition 3.3.2.** Recall that we can write

$$
\rho_X(h) = \frac{\int_{0}^{\infty} g(x)g(x + |h|)dx}{\int_{0}^{\infty} g(x)^2 dx}, \quad h \in \mathbb{R}.
$$

(A.1)

Clearly we can take $h > 1$ and now

$$
\int_{0}^{\infty} g(x)g(x + h)dx = \int_{0}^{1} x^{\alpha} L_0(x)(x + h)^{-\beta} L_1(x + h)dx + \int_{1}^{\infty} x^{-\beta}(x + h)^{-\beta} L_1(x) L_1(x + h)dx
$$

=: I_{1,h} + I_{2,h},

where

$$
I_{1,h} = \int_{0}^{1} x^{\alpha} L_0(x)(x + h)^{-\beta} L_1(x + h)dx,
$$

$$
I_{2,h} = \int_{1}^{\infty} x^{-\beta}(x + h)^{-\beta} L_1(x) L_1(x + h)dx.
$$

(i) Our strategy is to show that $I_{1,h} \sim h^{-\beta} L_1(h)$ and $I_{2,h} = o \left( h^{-\beta} L_1(h) \right)$ as $h \to \infty$, from which we, together with (A.1), get the desired convergence rate, $\rho(h) \sim h^{-\beta} L_1(h)$. 
Take first $I_{1,h}$. Since $\beta > 0$, 
\[
I_{1,h} \leq h^{-\beta} \int_0^1 x^\alpha L_0(x) L_1(x + h) \, dx \\
= h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \frac{L_1((x/h + 1)h)}{L_1(h)} \, dx \\
\sim h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \, dx, \quad h \to \infty,
\]
by the properties of slowly varying functions and where we applied the dominated convergence theorem, which is valid since for all $\epsilon > 0$ and large enough $h$,
\[
x^\alpha L_0(x) \frac{L_1((x/h + 1)h)}{L_1(h)} < x^\alpha L_0(x)(1 + \epsilon), \quad x \in (0, 1),
\]
which is integrable over $(0, 1)$ since $\alpha > -1/2$. Similarly, we can make the opposite evaluation:
\[
I_{1,h} \geq (1 + h)^{-\beta} \int_0^1 x^\alpha L_0(x) L_1(x + h) \, dx \\
= (1 + h)^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \frac{L_1((x/h + 1)h)}{L_1(h)} \, dx \\
\sim h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \, dx, \quad h \to \infty.
\]
Putting the two inequalities together we get that $I_{1,h} \sim h^{-\beta} L_1(h)$ as $h \to \infty$. For the second integral we get 
\[
I_{2,h} \leq (1 + h)^{-\beta} \int_1^{\infty} x^{-\beta}(x + h)^{-\beta} L_1(x) L_1(x + h) \, dx \\
= (1 + h)^{-\beta} L_1(h) \int_1^{\infty} x^{-\beta} L_1(x) \frac{L_1((x/h + 1)h)}{L_1(h)} \, dx \\
\sim h^{-\beta} L_1(h) \int_1^{\infty} x^{-\beta} L_1(x) \, dx, \quad h \to \infty,
\]
where $\int_1^{\infty} x^{-\beta} L_1(x) \, dx < \infty$ since $\beta > 1$.

(ii) We may write 
\[
I_{2,h} = h^{-\beta} \int_1^{\infty} x^{-\beta} \left(1 + \frac{x}{h}\right)^{-\beta} L_1(x) L_1(x + h) \, dx \\
= h^{-2\beta + 1} L_1(h)^2 \int_{1/h}^{\infty} y^{-\beta} (1 + y)^{-\beta} \frac{L_1(h y) L_1(h(1 + y))}{L_1(h) L_1(h)} \, dy, \\
=: \delta_h(y)
\]
15In both asymptotic estimates, the constant corresponding to the relation “$\sim$” equals one.
where the second equality follows by substituting \( y = x/h \). Fix now \( \delta \in (0, \beta - 1/2) \subset (0, 1/2) \). By the Potter bounds (Bingham et al., 1989, Theorem 1.5.6(ii)), under (A3) there exists a constant \( C_\delta > 0 \) such that

\[
\frac{L_1(hy)}{L_1(h)} \leq C_\delta \max\{y^{-\delta}, y^\delta\}, \quad y > 1/h, \quad h > 1.
\]

Accordingly, we find a dominant for \( k_h \), given by

\[
k_h(y) \leq \bar{K}(y) := \begin{cases} 
C_\delta y^{-\delta} (1 + y)^{-\beta + \delta}, & y \in (0, 1], \\
C_\delta y^{2(\beta - \delta)}, & y \in (1, \infty),
\end{cases}
\]

for any \( h > 1 \). Note that the dominant \( \bar{K} \) is integrable on \((0, 1]\), since \(-\beta - \delta > -1 - 1/2 = -1/2\), as well as on \((1, \infty)\), since \(-2(\beta - \delta) < -1\). Applying the dominated convergence theorem, we get

\[
I_{2,h} \sim h^{-2\beta + 1} L_1(h)^2 \int_0^\infty y^\beta (1 + y)^\beta \, dy, \quad h \to \infty.
\]

Moreover, the asymptotic estimate \( I_{1,h} \sim h^{-\beta} L_1(h) \), as \( h \to \infty \), remains valid also under the present assumptions. Observing that \(-\beta < -2\beta + 1\) when \( \beta < 1 \), we deduce that \( I_{1,h} = o(h^{-2\beta + 1} L_1(h)^2) \) as \( h \to \infty \). Therefore \( \rho_X(h) \sim I_{2,h} \) as \( h \to \infty \), and the assertion follows. \( \square \)

**Proof of Proposition 3.3.3.** As in the proof of Proposition 3.3.2, we take \( h > 1 \) and write

\[
\int_0^\infty g(x) g(x + h) \, dx = I_{1,h} + I_{2,h},
\]

where

\[
I_{1,h} = \int_0^1 x^\alpha L_0(x) (x + h)^{-\beta} e^{-\lambda(x+h)} L_1(x + h) \, dx,
\]

\[
I_{2,h} = \int_1^\infty x^{-\beta} e^{-\lambda x} L_1(x)(x + h)^{-\beta} e^{-\lambda(x+h)} L_1(x + h) \, dx.
\]

Our strategy is to show that \( I_{2,h} \sim e^{-\lambda h} h^{-\beta} L_1(h) \) and \( I_{1,h} = o\left(e^{-\lambda h} h^{-\beta} L_1(h)\right) \) as \( h \to \infty \), from which we, together with (A.1) get the desired convergence rate, \( \rho(h) \sim e^{-\lambda h} h^{-\beta} L_1(h) \). Consider first \( I_{1,h} \):

\[
I_{1,h} \leq e^{-\lambda h} h^{-\beta} \int_0^1 x^\alpha L_0(x) L_1(x + h) \, dx
\]

\[
= e^{-\lambda h} h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \frac{L_1(x + h)}{L(h)} \, dx
\]

\[
\sim e^{-\lambda h} h^{-\beta} L_1(h),
\]
as $h \to \infty$, where we use the dominated convergence theorem as in Proposition 3.3.2. For $I_{2,h}$, we get

$$I_{2,h} = e^{-\lambda h} \int_{1}^{\infty} x^{-\beta}(x + h)^{-\beta} e^{-2\lambda x} L_1(x) L_1(x + h) dx$$

$$= e^{-\lambda h} h^{-\beta} L_1(h) \int_{1}^{\infty} x^{-\beta}(x/h + 1)^{-\beta} e^{-2\lambda x} L_1(x) \frac{L_1(x) L_1(x + h)}{L_1(h)} dx$$

$$\sim e^{-\lambda h} h^{-\beta} L_1(h),$$

as $h \to \infty$.

Proof of Theorem 3.3.1

We first state and prove an elementary lemma, that we will need below.

Lemma A.2.1. As $x \downarrow 0$,

$$\sum_{k=2}^{\infty} \frac{x^k}{k!} = o(x).$$

Proof of Lemma A.2.1. By Taylor expansion of the exponential we have

$$\sum_{k=2}^{\infty} \frac{x^k}{k!} = e^x - 1 - x,$$

and, using l'Hôpital's rule,

$$\frac{e^x - 1 - x}{x} \to 0,$$

as $x \downarrow 0$. This concludes the proof.

Proof of Theorem 3.3.1. (i) Suppose w.l.o.g. that $\xi = 1$. Since $\sigma$ is covariance-stationary, it suffices to study

$$\rho(h) = Corr(\sigma_t, \sigma_{t+h}) = \frac{Cov(\sigma_t, \sigma_{t+h})}{\text{Var}(\sigma_0)}, \quad h \geq 0.$$ 

We get using the definition (3.3.1) of $\sigma$

$$Cov(\sigma_t, \sigma_{t+h}) = E[\sigma_t \sigma_{t+h}] - E[\sigma_0]^2$$

$$= E[\exp(X_t + X_{t+h})] - E[\exp(X_0)]^2.$$ 

Now, by the fact that $X$ is a zero-mean Gaussian process we get, using the moment generating function of the Gaussian distribution,

$$E[\exp(X_t + X_{t+h})] = \exp \left( \frac{1}{2} \text{Var}(X_t + X_{t+h}) \right)$$

$$= \exp \left( \text{Var}(X_0) + \text{cov}(X_0, X_{t+h}) \right)$$

$$= \exp \left( \gamma_X(0) + \gamma_X(h) \right).$$
and

\[ \mathbb{E}[\exp(X_0)] = \exp\left(\frac{1}{2} \gamma_X(0) \right), \]

where we write \( \gamma_X(h) \) for \( \text{Cov}(X_h, X_0) = \mathbb{E}[X_h X_0] \). Putting this together, we arrive at

\[ \rho(h) = \frac{\exp(\gamma_X(h)) - 1}{\exp(\gamma(0)) - 1}. \]

Now, using generic positive constants \( C, C_1, C_2 \) that may vary from line to line,

\[
1 - \rho(h) = \frac{\exp(\gamma_X(0)) - 1 - (\exp(\gamma_X(h)) - 1)}{\exp(\gamma(0)) - 1} \\
= \frac{\exp(\gamma_X(0)) - \exp(\gamma_X(h))}{\exp(\gamma(0)) - 1} \\
= C \left[ 1 - \exp(\gamma_X(h) - \gamma_X(0)) \right] \\
= C \left[ 1 - \exp(-\gamma_X(0)(1 - \rho_X(h))) \right] \\
= C_1(1 - \rho_X(h)) + C_2 \sum_{k=2}^{\infty} \frac{(\gamma_X(0)(1 - \rho_X(h)))^k}{k!} \\
\sim |h|^{2\alpha + 1} L_0(h),
\]

by the assumption on \( 1 - \rho_X(h) \) and Lemma A.2.1. (On the penultimate line, we Taylor expanded the exponential.) This concludes the proof of (i).

(ii) Suppose w.l.o.g. that \( \xi = 1 \). Using the same approach as in part (i), we get by Taylor expansion:

\[
\rho(h) = \frac{\exp(\gamma_X(h)) - 1}{\exp(\gamma(0)) - 1} \\
= C_1 \gamma_X(h) + C_2 \sum_{k=2}^{\infty} \frac{\gamma_X(h)^k}{k!} \\
= C_1 \gamma_X(0) \rho_X(h) + C_2 \sum_{k=2}^{\infty} \frac{\gamma_X(h)^k}{k!} \\
\sim \rho_X(h), \quad h \to \infty,
\]

by Lemma A.2.1. \( \square \)
PhD dissertations since 1 July 2011

2011-4 Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics
2011-5 Christian Bach: The Game of Risk
2011-6 Stefan Holst Bache: Quantile Regression: Three Econometric Studies
2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management
2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization
2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders
2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade
2012-1 Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting
2012-2 Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping
2012-3 Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing
2012-4 Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing
2012-5 Lasse Frisgaard Gunnersen: Income Redistribution Policies
2012-6 Miriam Wüst: Essays on early investments in child health
2012-7 Yukai Yang: Modelling Nonlinear Vector Economic Time Series
2012-9 Henrik Nørholm: Structured Retail Products and Return Predictability
2012-10 Signe Frederiksen: Empirical Essays on Placements in Outside Home Care
2012-11 Mateusz P. Dziubinski: Essays on Financial Econometrics and Derivatives Pricing
<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-12</td>
<td>Jens Riis Andersen</td>
<td>Option Games under Incomplete Information</td>
</tr>
<tr>
<td>2012-13</td>
<td>Margit Malmose</td>
<td>The Role of Management Accounting in New Public Management Reforms: Implications in a Socio-Political Health Care Context</td>
</tr>
<tr>
<td>2012-14</td>
<td>Laurent Callot</td>
<td>Large Panels and High-dimensional VAR</td>
</tr>
<tr>
<td>2012-15</td>
<td>Christian Rix-Nielsen</td>
<td>Strategic Investment</td>
</tr>
<tr>
<td>2013-1</td>
<td>Kenneth Lykke Sørensen</td>
<td>Essays on Wage Determination</td>
</tr>
<tr>
<td>2013-2</td>
<td>Tue Rauff Lind Christensen</td>
<td>Network Design Problems with Piecewise Linear Cost Functions</td>
</tr>
<tr>
<td>2013-3</td>
<td>Dominyka Sakalauskaite</td>
<td>A Challenge for Experts: Auditors, Forensic Specialists and the Detection of Fraud</td>
</tr>
<tr>
<td>2013-4</td>
<td>Rune Bysted</td>
<td>Essays on Innovative Work Behavior</td>
</tr>
<tr>
<td>2013-5</td>
<td>Mikkel Nørlem Hermansen</td>
<td>Longer Human Lifespan and the Retirement Decision</td>
</tr>
<tr>
<td>2013-6</td>
<td>Jannie H.G. Kristoffersen</td>
<td>Empirical Essays on Economics of Education</td>
</tr>
<tr>
<td>2013-7</td>
<td>Mark Strøm Kristoffersen</td>
<td>Essays on Economic Policies over the Business Cycle</td>
</tr>
<tr>
<td>2013-8</td>
<td>Philipp Meinen</td>
<td>Essays on Firms in International Trade</td>
</tr>
<tr>
<td>2013-9</td>
<td>Cédric Gorinas</td>
<td>Essays on Marginalization and Integration of Immigrants and Young Criminals – A Labour Economics Perspective</td>
</tr>
<tr>
<td>2013-10</td>
<td>Ina Charlotte Jäkel</td>
<td>Product Quality, Trade Policy, and Voter Preferences: Essays on International Trade</td>
</tr>
<tr>
<td>2013-11</td>
<td>Anna Gerstrøm</td>
<td>World Disruption - How Bankers Reconstruct the Financial Crisis: Essays on Interpretation</td>
</tr>
<tr>
<td>2013-12</td>
<td>Paola Andrea Barrientos Quiroga</td>
<td>Essays on Development Economics</td>
</tr>
<tr>
<td>2013-13</td>
<td>Peter Bodnar</td>
<td>Essays on Warehouse Operations</td>
</tr>
<tr>
<td>2013-14</td>
<td>Rune Vammen Lesner</td>
<td>Essays on Determinants of Inequality</td>
</tr>
<tr>
<td>2013-15</td>
<td>Peter Arendorf Bache</td>
<td>Firms and International Trade</td>
</tr>
<tr>
<td>2013-16</td>
<td>Anders Laugesen</td>
<td>On Complementarities, Heterogeneous Firms, and International Trade</td>
</tr>
</tbody>
</table>
2013-17  Anders Bruun Jonassen: Regression Discontinuity Analyses of the Disincentive Effects of Increasing Social Assistance

2014-1  David Sloth Pedersen: A Journey into the Dark Arts of Quantitative Finance


2014-3  Lukas Bach: Routing and Scheduling Problems - Optimization using Exact and Heuristic Methods

2014-4  Tanja Groth: Regulatory impacts in relation to a renewable fuel CHP technology: A financial and socioeconomic analysis

2014-5  Niels Strange Hansen: Forecasting Based on Unobserved Variables

2014-6  Ritwik Banerjee: Economics of Misbehavior

2014-7  Christina Annette Gravert: Giving and Taking – Essays in Experimental Economics

2014-8  Astrid Hanghøj: Papers in purchasing and supply management: A capability-based perspective

2014-9  Nima Nonejad: Essays in Applied Bayesian Particle and Markov Chain Monte Carlo Techniques in Time Series Econometrics

2014-10  Tine L. Mundbjerg Eriksen: Essays on Bullying: an Economist’s Perspective

2014-11  Sashka Dimova: Essays on Job Search Assistance

2014-12  Rasmus Tangsgaard Varneskov: Econometric Analysis of Volatility in Financial Additive Noise Models

2015-1  Anne Floor Brix: Estimation of Continuous Time Models Driven by Lévy Processes

2015-2  Kasper Vinther Olesen: Realizing Conditional Distributions and Coherence Across Financial Asset Classes

2015-3  Manuel Sebastian Lukas: Estimation and Model Specification for Econometric Forecasting

2015-4  Sofie Theilade Nyland Brodersen: Essays on Job Search Assistance and Labor Market Outcomes

2015-6  Sanni Nørgaard Breining: The Sibling Relationship Dynamics and Spillovers
2015-7  Marie Herly: Empirical Studies of Earnings Quality
2015-8  Stine Ludvig Bech: The Relationship between Caseworkers and Unemployed Workers
2015-9  Kaleb Girma Abreha: Empirical Essays on Heterogeneous Firms and International Trade
2015-10 Jeanne Andersen: Modelling and Optimisation of Renewable Energy Systems
2015-11 Rasmus Landersø: Essays in the Economics of Crime
2015-12 Juan Carlos Parra-Alvarez: Solution Methods and Inference in Continuous-Time Dynamic Equilibrium Economies (with Applications in Asset Pricing and Income Fluctuation Models)
2015-13 Sakshi Girdhar: The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis
2015-14 Wenjing Wang: Corporate Innovation, R&D Personnel and External Knowledge Utilization
2015-16 Jonas Maibom: Structural and Empirical Analysis of the Labour Market
2015-17 Sylvanus Kwaku Afesorgbor: Essays on International Economics and Development
2015-18 Orimar Sauri: Lévy Semistationary Models with Applications in Energy Markets
2015-20 Jonas Nygaard Eriksen: Business Cycles and Expected Returns
2015-21 Simon Juul Hviid: Dynamic Models of the Housing Market
2016-1  Silvia Migali: Essays on International Migration: Institutions, Skill Recognition, and the Welfare State
2016-2  Lorenzo Boldrini: Essays on Forecasting with Linear State-Space Systems
2016-4  Camilla Pisani: Volatility and Correlation in Financial Markets: Theoretical Developments and Numerical Analysis
2016-5 Anders Kronborg: Methods and Applications to DSGE Models
2016-6 Morten Visby Krægpøth: Empirical Studies in Economics of Education
2016-7 Anne Odile Peschel: Essays on Implicit Information Processing at the Point of Sale: Evidence from Experiments and Scanner Data Analysis
2016-8 Girum Dagnachew Abate: Essays in Spatial Econometrics
2016-9 Kai Rehwald: Essays in Public Policy Evaluation
2016-10 Reza Pourmoayed: Optimization Methods in a Stochastic Production Environment
2016-11 Sune Lauth Gadegaard: Discrete Location Problems – Theory, Algorithms, and Extensions to Multiple Objectives
2016-12 Lisbeth Palmhøj Nielsen: Empirical Essays on Child Achievement, Maternal Employment, Parental Leave, and Geographic Mobility
2016-13 Louise Voldby Beuchert-Pedersen: School Resources and Student Achievement: Evidence From Social and Natural Experiments
2016-14 Mette Trier Damgaard: Essays in Applied Behavioral Economics
2016-15 Andrea Barletta: Consistent Modeling and Efficient Pricing of Volatility Derivatives
2016-16 Thorvardur Tjörvi Ólafsson: Macrofinancial Linkages and Crises in Small Open Economies
2016-17 Carlos Vladimir Rodríguez Caballero: On Factor Analysis with Long-Range Dependence
2016-18 J. Eduardo Vera-Valdés: Essays in Long Memory
2016-19 Magnus Sander: Returns, Dividends, and Optimal Portfolios
2016-20 Ioana Daniela Neamtu: Wind Power Effects and Price Elasticity of Demand for the Nordic Electricity Markets
2016-21 Anne Brink Nandrup: Determinants of Student Achievement and Education Choice
2016-22 Jakob Guldbæk Mikkelsen: Time-Varying Loadings in Factor Models: Theory and Applications
2016-23 Dan Nguyen: Formidability and Human Behavior: An Interdisciplinary Approach

2016-25  Samira Mirzaei: Optimization Algorithms for Multi-Commodity Routing and Inventory Routing Problems

2017-1   Viktoryia Buhayenko: Determining Dynamic Discounts for Supply Chain Coordination

2017-2   Mikkel Bennedsen: Rough Continuous-Time Processes: Theory and Applications