Decoupling the short- and long-term behavior of stochastic volatility

Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen

CREATES Research Paper 2017-26

Decoupling the short- and long-term behavior of stochastic volatility[*]

Mikkel Bennedsen[†]   Asger Lunde[‡]   Mikko S. Pakkanen[§]

First version: October 2, 2016
This version: July 28, 2017

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 two 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.

[†]Department of Economics and Business Economics and CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. E-mail: mbennedsen@econ.au.dk

[‡]Department of Economics and Business Economics and CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. E-mail: alunde@econ.au.dk

[§]Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK and CREATES, Aarhus University, Denmark. E-mail: m.pakkanen@imperial.ac.uk

1 Introduction

Intraday modeling of asset return volatility is of importance in several applications, including derivatives pricing, high-frequency trading, and risk management (Andersen et al., 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 et al. (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 et al., 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* ($\mathcal{BSS}$) 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). We show that $\mathcal{BSS}$ processes, under suitable specifications, can accommodate both bona fide long memory (i.e., non-integrable autocorrelations) and short memory (exponentially decaying autocorrelations). We derive some general theoretical results concerning the memory properties of these processes in this paper. Under rather general conditions, $\mathcal{BSS}$ 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., 2017).

Recently, there has been considerable interest in rough models of volatility. This is due to both theoretical developments in implied volatility modeling (Alòs et al., 2007; Fukasawa, 2017) 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 logaritmic 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 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 extends this empirical analysis to daily volatility data on 1 944 individual US equities from the Trades and Quotes (TAQ) database, documenting that roughness and persistence are universal features of the volatility of financial equities. Section 4 introduces a class of stochastic volatility models, based on $\mathcal{BSS}$ processes, that are able to parsimoniously capture these empirical findings. This section also reviews simulation methods for rough stochastic volatility models. Section 5 presents estimation results of the various models we consider, using both parametric and semiparametric estimation procedures. Section 6 presents a forecasting study, where we compare our new models to existing volatility forecasting

models. Section 7 concludes. Proofs of the technical results are given in Appendix A.

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

$$dY_t = \mu_t dt + \sigma_t dB_t + dJ_t, \quad t \geq 0, \tag{2.1}$$

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. Since we sample prices at high frequency, we follow the standard practice of high-frequency financial econometrics by treating the data as noisy observations of the efficient log price Y, 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 σ, hence we do not go into much detail regarding the jump process J or the noise process U. We would 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, but not necessary, 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 et al., 2014, Proposition 1). Also, in the context of high-frequency returns, the drift process μ 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 σ, and the subsequent empirical findings on this process.

2.1 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 σ 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 et al., 2001; Barndorff-Nielsen and Shephard, 2002), realized kernels (Barndorff-Nielsen

4

et al., 2008), two-scale estimators (Zhang et al., 2005), and pre-averaging methods (Jacod et al., 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 Δ be sufficiently small, and assuming that volatility does not vary too much in each time interval of size Δ, we can use the proxy

$$\hat{\sigma}_t^2 = \Delta^{-1}\widehat{IV}_t^{\Delta}, \quad t = \Delta, 2\Delta, \dots n\Delta, \tag{2.2}$$

where $\widehat{IV}_t^{\Delta}$ is an estimate of IV derived from one of the methods mentioned. A priori, we do not want to fix any specific value of the step size Δ to sample volatility, as the choice of this parameter would vary from application to another, and as there is no canonical choice when Δ is less than a day. For this reason, we perform the subsequent analyses for various values of Δ and, as will be seen, our empirical findings hold for essentially all intraday, as well as daily, time scales.

The proxy (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 et al. (2012), and Zu and Boswijk (2014). In this paper we will restrict attention to (2.2) where we estimate IV using *pre-averaged measures*, developed in Jacod et al. (2009). We briefly review the implementation, following the methodology 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, $Z_0, Z_1, \dots, Z_N$, where $Z_i = \log S_{t_i}$, of the log price process $Z = \log S$ in this interval. We define the *pre-averaged log returns*,

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

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 $K = \lfloor\sqrt{N}\rfloor + 1$ otherwise.[1] Here, $\lfloor x \rfloor$ means the largest integer smaller 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:

$$RV_t^{\Delta*} = \frac{N}{N-K+2}\frac{1}{K\psi_K}\sum_{j=0}^{N-K+1}|r_{j,K}^*|^2 - \frac{\hat{\omega}^2}{\theta^2\psi_K},$$

$$BV_t^{\Delta*} = \frac{N}{N-2K+2}\frac{1}{K\psi_K}\frac{\pi}{2}\sum_{j=0}^{N-2K+1}|r_{j,K}^*||r_{j+K,K}^*| - \frac{\hat{\omega}^2}{\theta^2\psi_K}, \tag{2.3}$$

---

[1]Setting the tuning parameter θ equal to 1 was found to work well for data similar to ours in Christensen et al. (2014). We drew the same conclusion from both simulated data mimicking our setup, as well from the actual data analyzed later in the paper.

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 statistics $RV^{\Delta*}$ and $BV^{\Delta*}$ are the pre-averaged analogs of the realized variance (Andersen et al., 2001) 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 our study. We note, however, that we also conducted analogous analyses using $RV^{\Delta*}$ and obtained largely similar results and drew essentially identical conclusions as below, suggesting that the impact of jumps is negligible in the data studied here (cf. also Christensen et al., 2014).[2]

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

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

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

$$\widehat{\tilde{\sigma}_t^2} = \Delta^{-1} \widehat{IV}_t^{\Delta} / \widehat{(\sigma_t^s)^2} = \Delta^{-1} BV_t^{\Delta*} / \widehat{(\sigma_t^s)^2}, \quad t = \Delta, 2\Delta, \ldots, n\Delta,$$

---

[2]We also ran the analyses using the realized kernel estimator of Barndorff-Nielsen et al. (2008), obtaining similar results, although this estimator is not robust to jumps either. The details can be found in the first version of this paper, available at: https://arxiv.org/1610.00332v1.
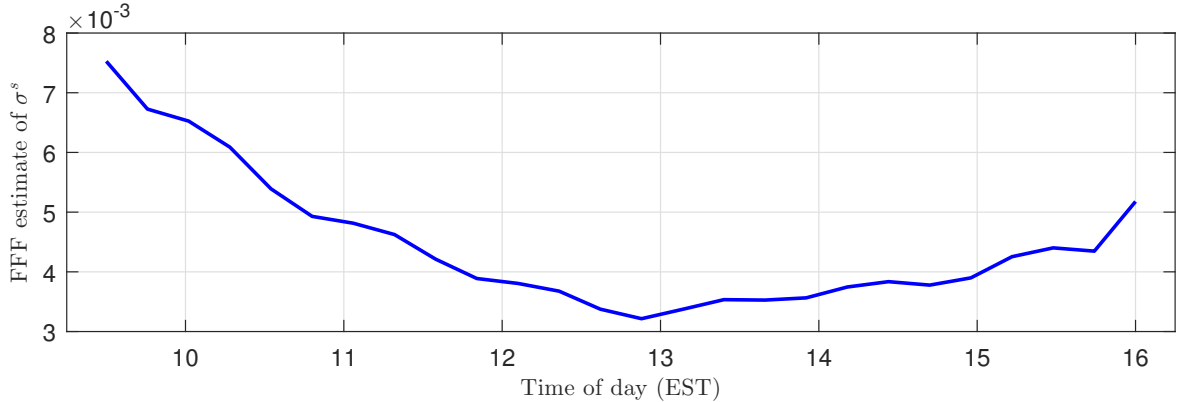
Figure 1: *Seasonality of volatility with $\Delta = 15$ minutes.*

and will from now on be working with these de-seasonalized data. Further, we abuse notation slightly and will write σ even though we actually refer to the de-seasonalized process $\tilde{\sigma}$. Table 1 contains some simple descriptive statistics of this process and how it behaves as Δ increases.[3] While the skewness is positive, kurtosis becomes closer to 3, indicating that the volatility estimates look more Gaussian as Δ is increased. We will present additional evidence of this in Section 2.5 below.

2.2 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 1 we present results of several unit root tests applied to the data with various values of Δ. 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 σ_t^2 is measured with error) also rejects the presence of a unit root. This supports the hypothesis that volatility is stationary.

2.3 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) ρ of log volatility, assuming that it is covariance-stationary, adheres to the asymptotic relationship

$$1 - \rho(h) := 1 - Corr(\log \sigma_t, \log \sigma_{t+h}) \sim c|h|^{2\alpha+1}, \quad |h| \to 0, \tag{2.4}$$

for a constant $c > 0$ and some $\alpha \in \left(-\frac{1}{2}, 0\right)$. Here and below "$\sim$" indicates that the ratio between the left- and right-hand side tends to one — when the constant in this relationship is immaterial,

---

[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 Δ.

Table 1: *Descriptive statistics and unit root tests*

Δ	n	Mean	Skew	Kurt	ADF	PP	$\hat{\pi}$	$n(\hat{\pi}-1)$
10 minutes	19422	0.022	0.039	4.184	0.000	0.001	0.974	-500.3
15 minutes	12948	0.031	0.233	3.808	0.000	0.001	0.970	-394.7
30 minutes	6474	0.048	0.363	3.662	0.000	0.001	0.966	-220.5
65 minutes	2988	0.054	0.454	3.633	0.000	0.001	0.960	-120.6
130 minutes	1494	0.054	0.486	3.523	0.000	0.001	0.950	-75.1
1 day	498	0.061	0.420	3.410	0.000	0.001	0.830	-84.5

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 π 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% critial values of this test are -20.7 and -14.1, respectively.

we will denote it by the generic c, which may vary from from formula to another. We call α the *roughness index* of the log volatility process. In general, α takes values in $\left(-\frac{1}{2}, \infty\right)$ but, as we shall see, only negative values of α will be relevant for us. For stationary Gaussian processes, the relationship (2.4) implies that the process has a modification with locally ϕ-Hölder continuous trajectories for any $\phi \in (0, \alpha + 1/2)$, where the index ϕ 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 locally ϕ-Hölder continuous trajectories for any $\phi \in (0, 1/2)$. Thus, negative values of α suggest trajectories rougher that 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 Alòs et al. (2007) and Fukasawa (2017), 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.

2.3.1 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

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

---

[4]See, e.g., Proposition 2.1 of Bennedsen (2016).

showing that when the underlying process is covariance-stationary, the asymptotics (2.4) hold for the variogram as well. Indeed,

$$\gamma_2(h) \sim c|h|^{2\alpha+1}, \quad |h| \to 0. \tag{2.5}$$

This suggests a straightforward semiparametric estimation procedure for α. Namely, consider the regression

$$\log \hat{\gamma}_2(h) = b + a \log |h| + \epsilon_h, \quad h = \Delta, 2\Delta, \ldots, m\Delta, \tag{2.6}$$

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 α using $\hat{\alpha} = \frac{\hat{a}-1}{2}$, where $\hat{a}$ is the OLS estimate of a in (2.6). The error term in the regression (2.6) does not satisfy the usual assumptions of OLS estimation, but inference (confidence intervals) on the parameter α can be conducted using the bootstrap method of Bennedsen (2016), specifically tailored to estimation of rough processes. Choosing the bandwidth parameter, m, optimally requires some care. It should preferably be small, as the asymptotics in (2.4) may only hold for relatively small values of m. However, as argued in Bennedsen (2016), choosing m slightly larger than the minimum value of 2 makes the OLS estimation in (2.6) more robust, as it leaves more data points for the regression. In our study, we use $m = 6$, while we also experimented with several other values of m and found that the results are very robust to the choice of bandwidth.

Unfortunately, this OLS 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 actual log volatility process, but only an estimate thereof (here through the pre-averaged measure BV^*), as explained in Section 2.1. To see why this might invalidate the OLS-based procedure, recall that BV^* is a noisy estimate of IV. Suppose, for the sake of argument, that BV^* is representable as

$$BV_t^{\Delta*} = \widehat{IV}_t^\Delta = IV_t^\Delta \eta_t, \quad t = \Delta, 2\Delta, \ldots, n\Delta,$$

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

$$\log \widehat{\sigma_t^2} = \log \left(\Delta^{-1} BV_t^{\Delta*}\right) = \log \left(\Delta^{-1} IV_t^\Delta \eta_t\right) = \log \left(\Delta^{-1} IV_t^\Delta\right) + \epsilon_t, \quad t = \Delta, 2\Delta, \ldots, n\Delta, \tag{2.7}$$

where now $\epsilon_t := \log \eta_t$ defines another iid noise sequence.

From this it is evident that, even if the term $\Delta^{-1} IV_t^\Delta$ is a good approximation of the latent volatility process σ_t^2, the estimates of the roughness parameter coming from observations of the time series $\log \widehat{\sigma_t^2}$ will be influenced by the noise-term ϵ_t. Indeed, one might even wonder whether the findings of roughness of volatility are explained by this effect.

To demonstrate that this is not the case, at least for values of Δ that are not too small, we propose below a *noise-robust* estimator of the roughness parameter α. To motivate the estimator,

9

first notice that according to (2.7), the variogram of the observations can be expressed as

$$\gamma_2^*(h) := \mathbb{E}[|\log \widehat{\sigma^2}_{t+h} - \log \widehat{\sigma^2}_t|^2]$$
$$= \mathbb{E}[|\log\left(\Delta^{-1}IV_{t+h}^\Delta\right) - \log\left(\Delta^{-1}IV_t^\Delta\right)|^2] + \mathbb{E}[|\epsilon_{t+h} - \epsilon_t|^2]$$
$$\approx \gamma_2(h) + 2\sigma_\epsilon^2, \tag{2.8}$$

where $\sigma_\epsilon^2 = Var(\epsilon_t)$ is the variance of the noise sequence ϵ and the approximation on the third line stems from using $\Delta^{-1}IV_t^\Delta$ as an approximation of σ_t^2.[5] Thus, even though the variogram of log volatility behaves as in (2.5), 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. the relationship (2.6). Because the OLS estimator hinges on linearity, estimates using the OLS procedure may be biased. Furthermore, at least in the case of an iid noise sequence ϵ, it is easy to show that this bias is *downwards*. In other words, applying the OLS estimator of α to noisy data might lead us to overestimate the roughness of the underlying process.

To mitigate the bias, we endeavor to estimate α in a noise-robust way, by running a non-linear least squares (NLLS) regression, inspired by (2.8) and (2.5),

$$\hat{\gamma}_2^*(h) = a + b|h|^{2\alpha+1} + \xi_h, \quad h = \Delta, 2\Delta, \ldots, m\Delta, \tag{2.9}$$

where ξ 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 α, and although we are chiefly interested in the roughness parameter α, it is useful to note that we can estimate $2\sigma_\epsilon^2$ by the NLLS estimate of the constant a. The following theorem establishes the consistency of the NLLS estimator under stylized high-level conditions. Its proof is carried out in Appendix A.

Theorem 2.1. *Let X be a continuous process with stationary increments and variogram*

$$\gamma_2(h) = \mathbb{E}[|X_{t+h} - X_t|^2] = c_0 h^{2\alpha_0+1}, \quad h > 0,$$

for a constant $c_0 > 0$ and a roughness parameter $\alpha_0 \in (-1/2, 1/2)$. Fix $\Delta > 0$ and let

$$Z_{i\Delta} = X_{i\Delta} + u_i, \quad i = 1, \ldots, n,$$

be equidistant observations; here $u = \{u_i\}_{i=1}^n$ is a zero-mean iid sequence, independent of X, with $Var(u_1) = \sigma_u^2 > 0$. Denote by $\gamma_2^*(h)$ the variogram of Z, let Θ denote a compact subset of $\mathbb{R}_+ \times \mathbb{R}_+ \times (-1/2, 1/2)$ such that $(2\sigma_u^2, c_0, \alpha_0)$ is in the interior of Θ, and define the NLLS estimate of $\theta := (a, c, \alpha)$ as

$$\hat{\theta} = (\hat{a}, \hat{c}, \hat{\alpha}) := \arg\inf_{\theta \in \Theta} \sum_{k=1}^m \left(\hat{\gamma}_2^*(k, \Delta) - a - c(k\Delta)^{2\alpha+1}\right)^2,$$

---

[5] Actually, using the approximation $\Delta^{-1}IV_t^\Delta = \Delta^{-1}\int_{t-\Delta}^t \sigma_s^2 ds \approx \sigma_t^2$ will result in estimates of the roughness parameter of σ_t^2 that 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 $\log\left(\Delta^{-1}IV_t^\Delta\right)$, will apply *a fortiori* for the time series of the (latent) spot log volatility $\log \sigma_t^2$.

with a fixed bandwidth $m \geq 3$ and

$$\hat{\gamma}_2^*(k, \Delta) := \frac{1}{n-k} \sum_{j=1}^{n-k} |Z_{(j+k)\Delta} - Z_{j\Delta}|^2, \quad k = 1, 2, \ldots, m.$$

Then

$$(\hat{a}, \hat{c}, \hat{\alpha}) \to (2\sigma_u^2, c_0, \alpha_0) \quad \text{in probability,} \quad \text{as} \quad n \to \infty.$$

Remark 2.1. The assumption that

$$\gamma_2(h) = c_0 h^{2\alpha_0 + 1}, \quad h > 0,$$

in Theorem 2.1 is not completely innocuous and is violated by many models of interest. For instance, below we will propose models conforming to

$$\gamma_2(h) = h^{2\alpha_0 + 1} L(h), \quad h > 0,$$

where the function L is *slowly varying at zero*, in the sense that $\lim_{x \to 0} \frac{L(tx)}{L(x)} = 1$ for all $t > 0$. It is possible that the presence of the function L could bias the NLLS estimates. However, we conjecture that such bias should typically be small, since, by the properties of slowly varying functions (e.g., Bingham et al., 1989), it is possible to find positive constants C_1 and C_2 such that for all $\epsilon > 0$ and h sufficiently small,

$$C_1 h^{2\alpha_0 + 1 + \epsilon} \leq \gamma_2(h) \leq C_2 h^{2\alpha_0 + 1 - \epsilon},$$

indicating that any potential bias in the estimate of α_0 can be made arbitrarily small by setting Δ and m small enough. Developing an asymptotic result for the NLLS estimators under these more general assumptions is beyond the scope of the present paper. However, we remark that in a series of simulation studies (not reported here, but available upon request), we found strong evidence for the conjecture that introducing the function L has only negligible impact on the NLLS estimates of α_0 in settings similar to the one considered in this paper.

The above discussion puts forward two methods for estimation of the roughness parameter α of log volatility — one based on OLS regression (where the noise in the estimates of IV is ignored) and another based on NLLS regression. Examples of the OLS regressions for $\Delta = 15, 65$ minutes are presented in Figure 2. The plots show that $\log \hat{\gamma}_2^*(h)$ is approximately linear in h for small values of h, which is evidence of the relationship (2.4) 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 Δ. In Figure 3 we apply both the OLS and the NLLS estimators to the E-mini S&P 500 data to obtain a signature plot for α, i.e., the estimated value of α is plotted as a function of the step size Δ (on a log scale). The figure shows that for $\Delta < 5$ minutes, both estimators produce estimates that are only very slightly above the lower bound $-\frac{1}{2}$, corresponding to maximal roughness. It is

plausible that these values are fraught with bias stemming from measurement noise, as explained above. Indeed, it appears that anything less than 5 minutes is simply too short a time interval to get reasonably low-noise estimates of IV using the current methodology.

For Δ between 5 and 10 minutes, however, the NLLS procedure results in estimates of α that are higher than the corresponding OLS estimates, lying outside the 95% confidence bands of the OLS estimates. While the bias of the OLS estimates persists, the NLLS procedure is able to mitigate it, resulting in more credible estimates that are compatible with those obtained with larger values of Δ. Finally, 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. This indicates that the impact of noise in the estimates of IV is no longer an issue then. Below we will only consider $\Delta \geq 10$ minutes, as the two estimators then give roughly the same estimates. Since the OLS estimator (2.6) is faster, easier to implement, and less noisy in general, we will use this estimator going forward.

It is worth pointing out that our estimates of α 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 α have not been influenced by the concatenation of the data we also estimate α within each day, which is feasible when we have enough observations of volatility per day (when Δ is less than 15 minutes, say). The results, presented in Figure 4, indicate that the estimates of α on individual days are consistent with the results on the concatenated data, seen in Figure 3. In particular, the strong evidence of the roughness of volatility is carried over.

2.3.2   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 5 provides results of rolling-window estimation of α, where the window length was 10 days, using the OLS regression (2.6) of Section 2.3 and $\Delta = 15$ minutes for the volatility proxy. We experimented also with different window sizes and different values of Δ, but the results were rather consistent. Figure 5 additionally displays the overall median over the entire period, as well as a smoothed version of the estimates. The figure shows that α 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 α.
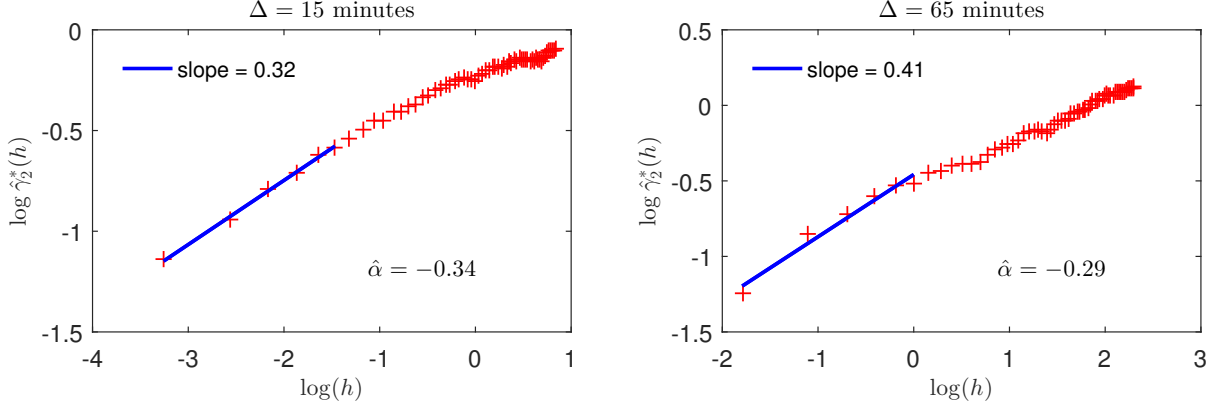
Figure 2: *Estimates of α using the OLS regression* (2.6) *with m = 6 lags.*

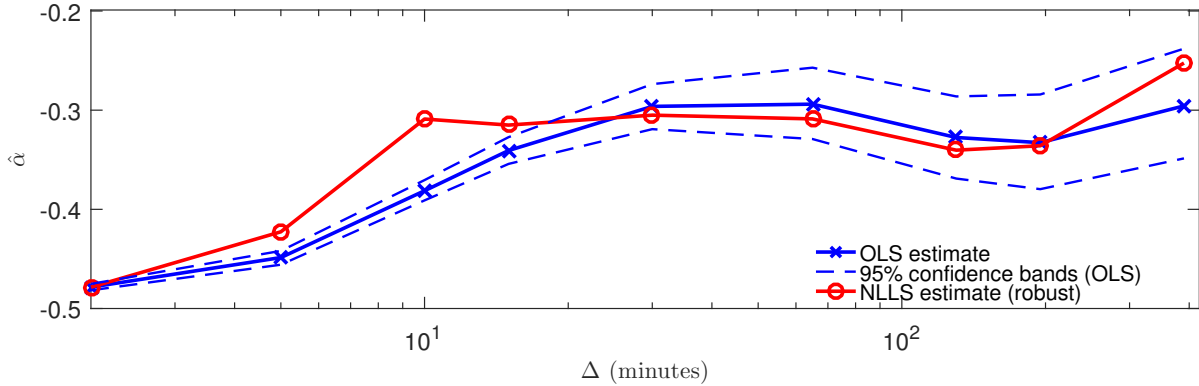Figure 3: *A signature plot for α, where the estimate α̂ is plotted as a function of Δ. Estimation was done using the OLS regression* (2.6) *with m = 6 lags (solid blue line with crosses). Confidence bands (dashed blue lines) are calculated using the bootstrap method of* Bennedsen (2016) *with B = 999 bootstrap replications. Also shown are estimates using the robust NLLS estimation procedure (solid red line with circles) as given in* (2.9). *The leftmost point corresponds to Δ = 1 minute and the rightmost point to Δ = 1 day. Note that the x-axis uses log scale.*

While it seems imprudent to 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).

2.4 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
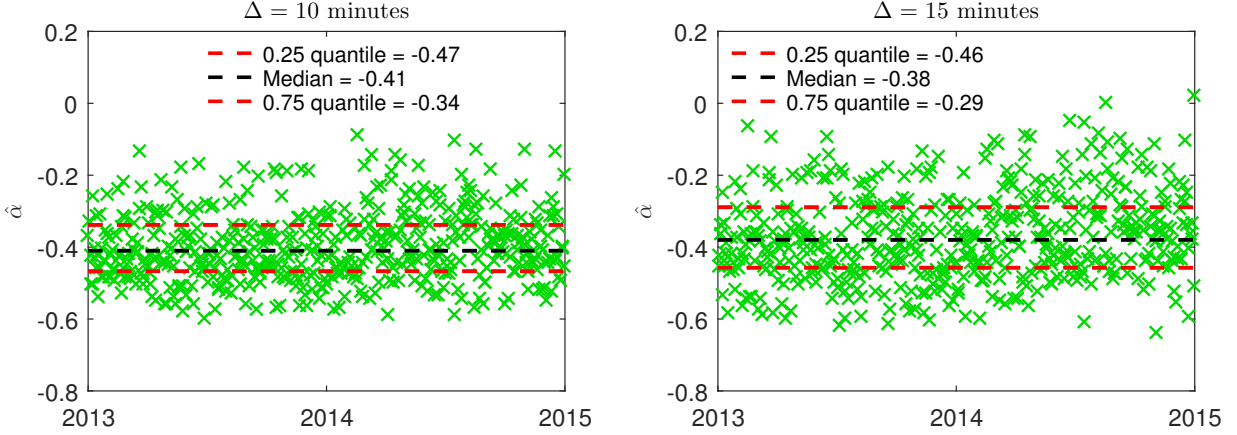
Figure 4: *Daily estimates of α using the OLS regression* (2.6) *with $m = 6$ lags.*
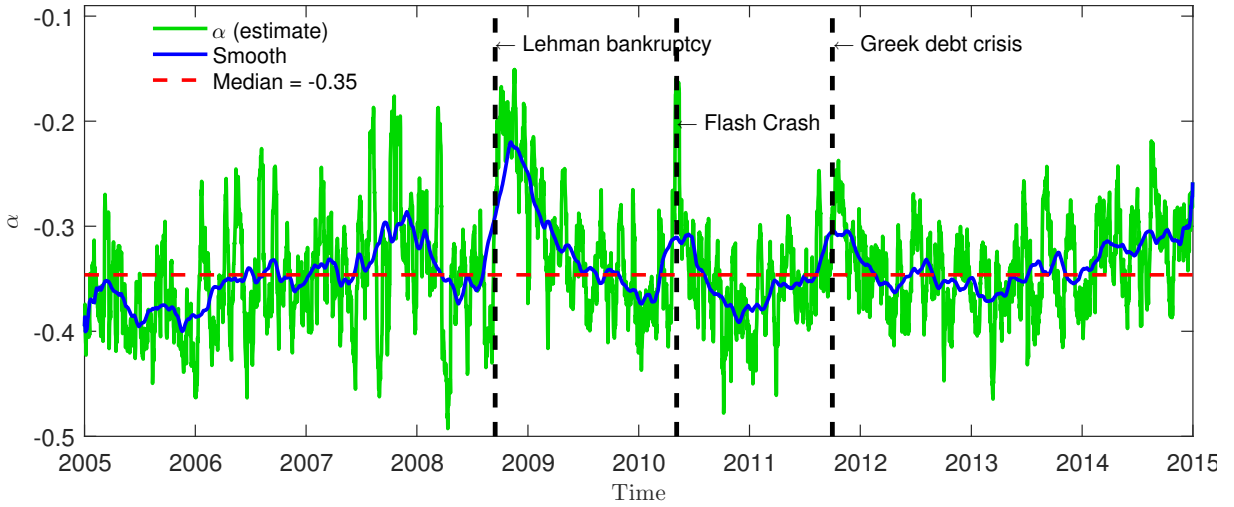
Figure 5: *Rolling-window estimates of α, as explained in Section 2.3.2. The volatility proxy is computed using $\Delta = 15$ minutes and the smoothed version is a simple moving average filter using 75 observations on both sides of the particular estimate. The vertical dashed black lines indicate three periods of market turmoil: September 15, 2008 (Lehman Brothers filing for bankruptcy at the onset of the subprime mortgage crisis), May 6, 2010 (the Flash Crash), and October 1, 2011 (in the midst of the Greek debt crisis), respectively.*

polynomial rate:

$$\rho(h) = Corr(\log \sigma_t, \log \sigma_{t+h}) \sim c|h|^{-\beta}, \quad |h| \to \infty, \tag{2.10}$$

for a constant $c > 0$ and some $\beta \in (0, \infty)$. When $\beta \in (0, 1)$, the correlation function ρ is not integrable, that is, $\int_0^\infty |\rho(h)|dh = \infty$, and we say that $\sigma = (\sigma_t)_{t \geq 0}$ 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 et al. (2012) for literature

14

on continuous-time volatility models with long memory.

Paralleling the approach of Section 2.3, the relationship (2.10) provides a basis for semiparametric estimation of β via the OLS regression

$$\log \hat{\rho}(h) = a + b \log |h| + \epsilon_h, \quad h = M\Delta, (M+1)\Delta, \ldots, M'\Delta, \tag{2.11}$$

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 6 presents the OLS regressions of the form (2.11) with step sizes $\Delta = 15, 65$ minutes. Further, Figure 7 shows a signature plot for β using the above estimator, i.e., the estimated value of β is plotted as a function of the step size Δ (with log scale). It should be noted that semiparametric estimates of the memory parameter β tend to be somewhat imprecise. We also found that these estimates were quite dependent on the values chosen for the thresholds M and M'. For this reason, we also apply a parametric estimator, obtained by fitting an autocorrelation function from the *Cauchy class*,

$$\rho_{Cau}(h) = \left(1 + |h|^{2\alpha+1}\right)^{-\frac{\beta}{2\alpha+1}}, \quad \alpha \in (-1/2, 1/2), \quad \beta > 0, \quad h \in \mathbb{R},$$

to the empirical autocorrelations. As argued in Gneiting and Schlather (2004), processes with this autocorrelation function obey both (2.5) and (2.10). Further details of the parametric estimation procedure are presented in Section 5.

As when estimating the roughness parameter α, we endeavor to estimate β in a way that is robust to the noise introduced by using the pre-averaged measure BV^* in place of the actual integrated variance IV. A similar estimation problem was considered in Hansen and Lunde (2013), albeit in a different setup. Working again under the stylized assumption (2.7), the autocorrelation function of the observations is

$$\rho^*(h) := Corr\left(\log \widehat{\sigma^2}_{t+h}, \log \widehat{\sigma^2}_t\right) = \frac{Var\left(\log\left(\Delta^{-1}IV^\Delta_{t+h}\right), \log\left(\Delta^{-1}IV^\Delta_t\right)\right)}{Var\left(\log\left(\Delta^{-1}IV^\Delta_t\right)\right) + Var\left(\epsilon_t\right)} \approx \frac{\rho(h)}{1+\delta^2}, \tag{2.12}$$

where

$$\delta^2 = \frac{Var\left(\epsilon_t\right)}{Var\left(\log\left(\Delta^{-1}IV^\Delta_t\right)\right)}$$

is the noise-to-signal ratio and the approximation in (2.12) again stems from using $\Delta^{-1}IV^\Delta_t$ as a proxy for σ^2_t. Equation (2.12) shows that the OLS estimator (2.11) of β is actually robust to noise, as the log transformation of ρ^* entails that the logarithm of the deflating factor $(1+\delta^2)^{-1}$ enters the regression (2.11) as an additive constant, so that it does not affect the estimation of β. When employing the parametric estimator, one simply needs to include an intercept in the regression, so that the function $\rho^*_{Cau}(h) := c \cdot \rho_{Cau}(h)$, where $c \in (0, 1]$ is a constant, instead of $\rho_{Cau}(h)$ is fitted to the empirical autocorrelations.[6]

---

[6]We found it to be helpful to work with the logarithm of both the empirical autocorrelations and $\rho^*_{Cau}(h)$ when implementing the estimation procedure numerically.
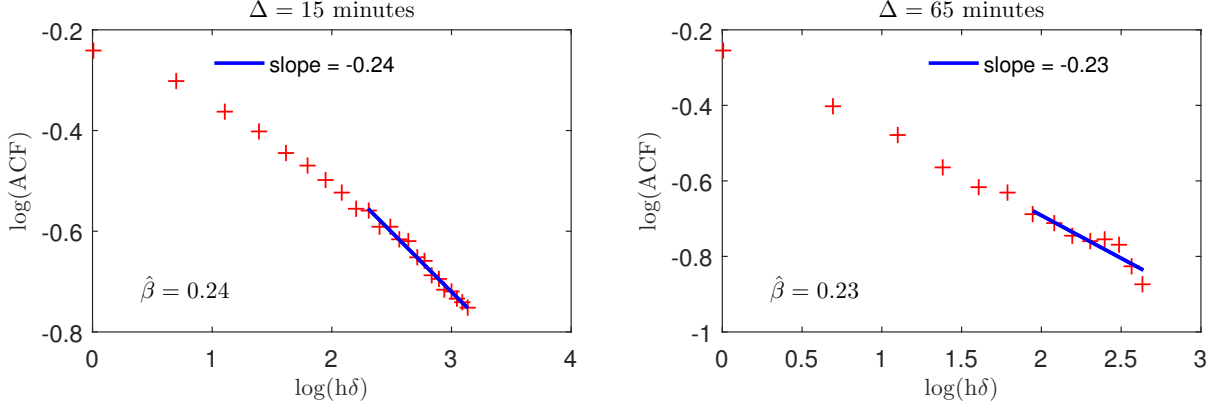
Figure 6: *Estimates of β using the OLS regression* (2.11) *with $M = \lfloor n^{1/4} \rfloor$ and $M' = \lfloor n^{1/3} \rfloor$.*
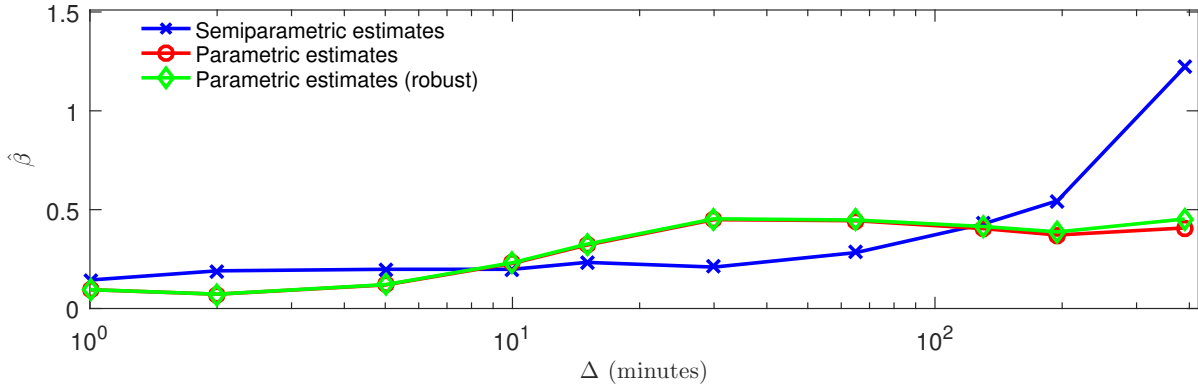
Figure 7: *A signature plot for β, where $\hat{\beta}$ is plotted as a function of Δ. The semiparametric estimation was done using the OLS regression* (2.11) *with $M = \lfloor n^{1/4} \rfloor$ and $M' = \lfloor n^{1/3} \rfloor$. The parametric estimates come from the Cauchy model, as explained in Section 5. The green curve is the noise-robust parametric estimator, cf. equation* (2.12). *The leftmost point corresponds to $\Delta = 1$ minute and the rightmost point to $\Delta = 1$ day.*

In Figure 7, 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 β. 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 Δ. Lastly, the difference between the parametric estimates using ρ_{Cau} and the noise-robust estimates using ρ_{Cau}^* is practically non-existent, indicating that the impact of the noisy measurements of IV on the parametric estimates of β, is negligible.

2.5 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. (2001), Andersen et al. (2001), 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 semi-heavy tails, to log volatility data (see Barndorff-Nielsen, 1997, for details on the NIG distribution and its application to stochastic volatility modelling). The results, given in Figure 8, suggest that increments of the log $BV^{\Delta*}$ statistics 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 log $BV^{\Delta*}$ data, indicating some kind of aggregational Gaussianity in log volatility. These findings are confirmed by the QQ plots in Figure 9.

We additionally studied the empirical distributions of the log $BV^{\Delta*}$ statistics of the 26 blue-chip US equities whose volatility we analyze below in Section 3. 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 log volatility at intraday time scales apply, *a fortiori*, also to them.

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

Our goal here is to study whether the findings above are only valid for the E-mini S&P 500 futures contract, or whether they hold true more generally. We therefore study 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.

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.[7] 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 most liquid assets. Here, a *liquid asset* is characterized by the following criteria.

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

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

---

[7]See: https://msci.com/gics

Figure 8: *Kernel density estimates for increments of log volatility, i.e., $y_k = \log(BV_{k\Delta}^{\Delta*}) - \log(BV_{\Delta(k-1)}^{\Delta*})$, (solid blue lines), fitted Gaussian distributions (dashed red lines), and fitted NIG distributions (patterned green lines).*

19.

(c) The estimate BV^* of IV is strictly positive on every day.[8]

After discarding the assets that do not fulfil these criteria, we are left with 1 944 liquid assets for our analysis.[9]

The estimates of the roughness parameter α of volatility for the assets in the sample are summarized in a box plot in Figure 10. A few features of this plot are worth highlighting. Firstly, the estimates of α are all negative, mostly in the range $(-0.45, -0.3)$, indicating pronounced roughness. Secondly, there does not seem to be significant differences in the roughness estimates across sectors.

Turning now to persistence properties, Figure 11 contains box plots of the estimates of β. As explained above, the semiparametric estimator of β can be inaccurate, and in such a large

---

[8]Although IV is by definition never negative, BV^* can become negative due to the bias correction term appearing in its definition (2.3).

[9]In an earlier version of this paper, available at: https://arxiv.org/1610.00332v1, we employed the realized kernel estimator of Barndorff-Nielsen et al. (2008), which is guaranteed to be non-negative, enabling us to analyze 5 071 assets. The results therein are very similar to the results of the present paper.
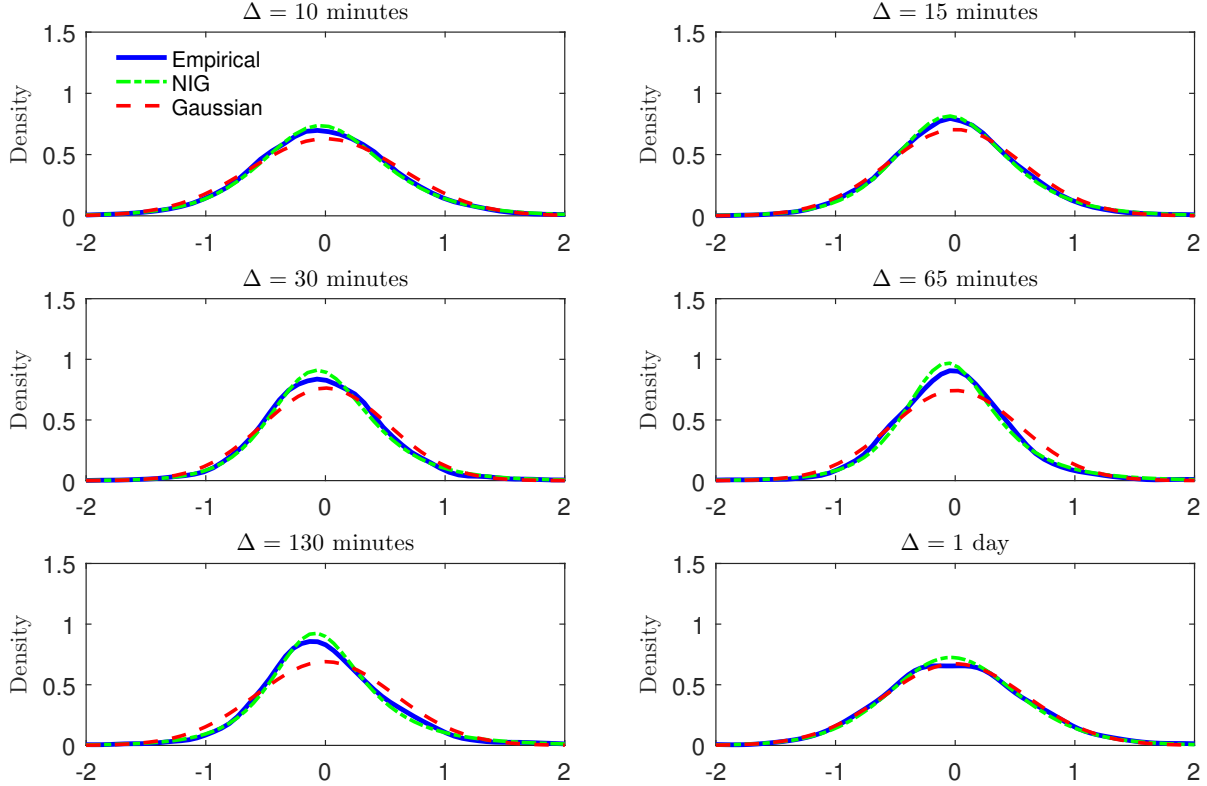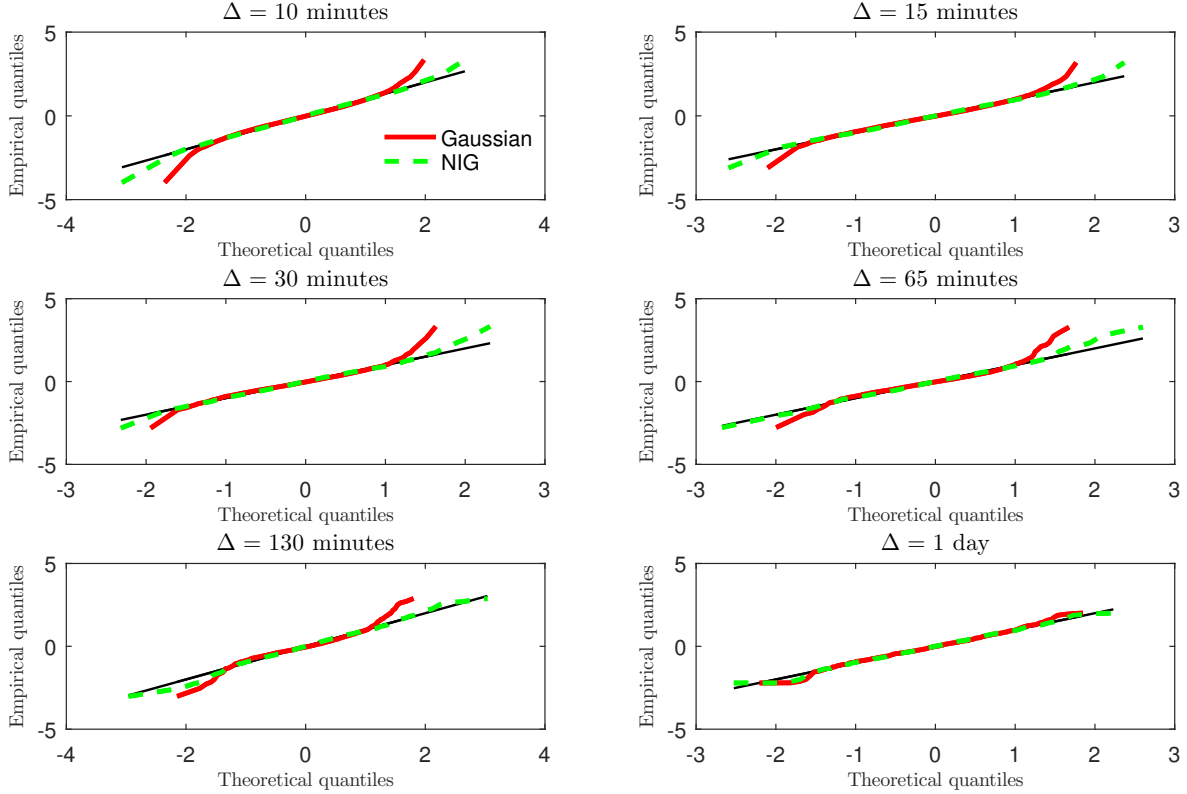
Figure 9: *QQ plots for increments of log volatility, i.e., $y_k = \log(BV_{k\Delta}^{\Delta*}) - \log(BV_{\Delta(k-1)}^{\Delta*})$. If the data fits a particular distribution — solid red for the Gaussian distribution, dashed green for the NIG distribution — the curve will be close to the diagonal solid black line.*
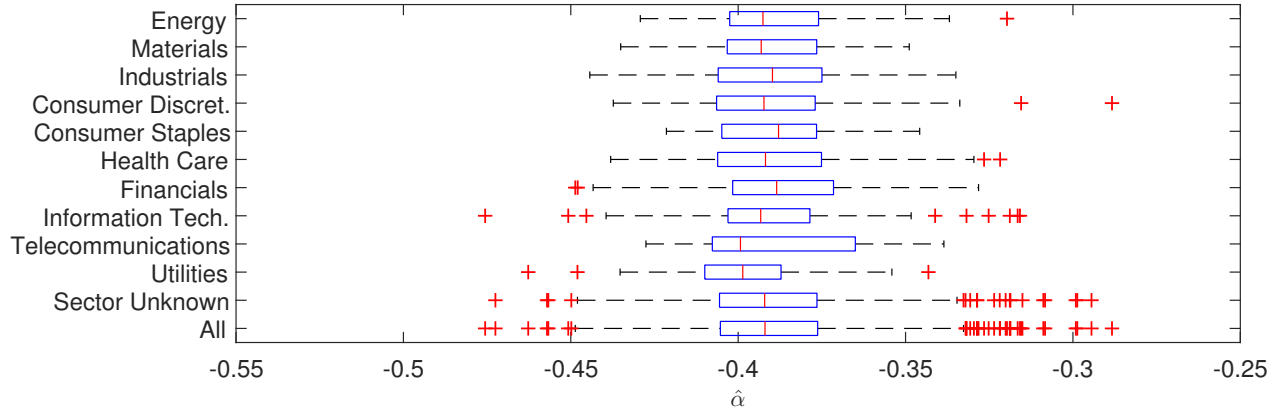
Figure 10: *Box plot for $\hat{\alpha}$ by sector. Bandwidth $m = 6$.*

data set as studied here, we obtained highly dispersed estimates (not reported here). Therefore, we instead consider the parametric estimator $\hat{\beta}_{Cauchy}$ based on the Cauchy model as detailed in Section 5.[10] Again, our analysis confirms the findings seen earlier. Indeed, the estimates show that

---

[10]We obtained similar results using the estimator $\hat{\beta}_{BSS}$ derived from the Power-$\mathcal{BSS}$ process, also discussed in Section 5, but for brevity these results are omitted.
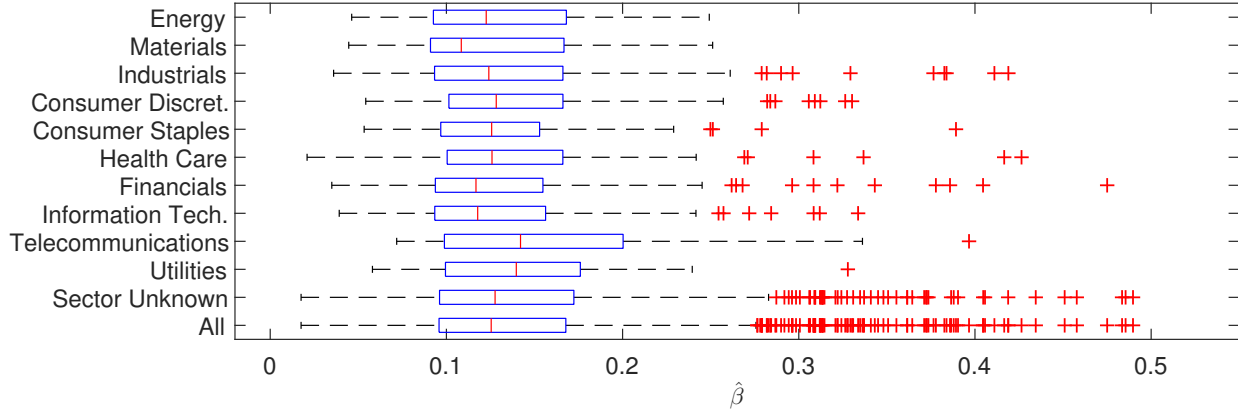
Figure 11: *Box plot for $\hat{\beta}_{Cauchy}$ by sector. Bandwidth $L = \lceil n^{1/3} \rceil$.*

log volatility is very persistent, with estimates of β mainly in the interval $(0, 0.4)$. In particular, there is compelling evidence of very strong persistence in volatility.

In summary, we find that roughness and persistence of log volatility extends to the volatility of individual equities as well. Indeed, these properties appear to be *universal* properties of the (log) volatility of equities.

4 Models of stochastic volatility that decouple short- and long-term behavior

Motivated by our empirical findings above, 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 α. Given a process X with such properties, we then define our model for volatility as

$$\sigma_t = \xi \exp\left(X_t\right), \quad t \geq 0, \tag{4.1}$$

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

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

4.1.1 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

20

function

$$\rho(h) = \left(1 + |h|^{2\alpha+1}\right)^{-\frac{\beta}{2\alpha+1}}, \quad h \in \mathbb{R}, \tag{4.2}$$

where $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and $\beta > 0$. In particular, the process G satisfies (2.4) and (2.10).

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

$$X_t = \int_0^t v_s dG_s, \quad t \geq 0, \tag{4.3}$$

using the Cauchy process G and some additional process $v = (v_t)_{t \in \mathbb{R}}$ that models the volatility of volatility. Under suitable assumptions, X will inherit the roughness properties of G, see, e.g., Barndorff-Nielsen et al. (2009, 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 (4.3) 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 v. In the case where the Cauchy process G is rough, these assumptions become rather restrictive, unfortunately — for example, v cannot be a semimartingale (with non-zero quadratic variation).

In sum, the Cauchy class provides a convenient model that decouples roughness and memory properties. Since the Cauchy class is characterized by a closed-form ACF, it is very easy to work with. Moreover, as we will see in the forecasting experiment below, volatility forecasts derived using a Cauchy class model perform rather well. However, this class of models is not easy to extend beyond Gaussianity, which limits its applicability as a model of log volatility, in view of the results in Section 2.5. Next, we propose a different modeling framework, where non-Gaussianity is easy to attain.

4.1.2 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 ($\mathcal{BSS}$), which was introduced in Barndorff-Nielsen and Schmiegel (2007, 2009). This process is defined via the moving-average representation

$$X_t = \int_{-\infty}^t g(t-s)v_s dW_s, \quad t \geq 0, \tag{4.4}$$

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

X non-Gaussian. In particular, when v is independent of W, we have

$$X_t|(v_s)_{s \leq t} \sim N\left(0, \int_0^\infty g(x)^2 v_{t-x}^2 dx\right),$$

showing that the marginal distribution of X_t is a normal mean-variance mixture with conditional variance governed by v and g. It is shown in Barndorff-Nielsen et al. (2013) that when g is given by the so-called gamma kernel (Example 4.2 below), we can choose the normal-inverse Gaussian (NIG) distribution as the marginal distribution of X. As we saw in Section 2.5, the NIG distribution seems to fit to the empirical distribution of log volatility very well. This is an encouraging property of the $\mathcal{BSS}$ 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 (4.4) is crucial. We 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 $\alpha \in (-1/2, 1/2)\backslash\{0\}$,

$$g(x) = x^\alpha L_0(x), \quad x \in (0, 1], \tag{4.5}$$

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 $t > 0$.[11] 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 $\lambda = 0$,

$$g(x) = e^{-\lambda x} x^{-\gamma} L_1(x), \quad x \in (1, \infty), \tag{4.6}$$

where L_1 is slowly varying at infinity and bounded away from zero and ∞ 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

---

[11]We refer to Bingham et al. (1989) for an extensive treatment of slowly varying (and regularly varying) functions.

integrable under the specifications (4.5) and (4.6). Similarly, in the case $\lambda > 0$, the conditions $\alpha > -1/2$ and $\gamma \in \mathbb{R}$, under (4.5) and (4.6), suffice for square integrability. The assumptions (A1), (A2), and (A3) are similar to those used in Bennedsen et al. (2017), 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 $\mathcal{BSS}$ process X, defined by (4.4), has α as its roughness index in the sense of Equation (2.4). The result is a straightforward adaptation of Proposition 2.1 in Bennedsen et al. (2017) and we refer to that paper for a proof. Below, and in what follows, we denote by ρ_X the autocorrelation function of X.

Proposition 4.1. *If the kernel function g satisfies* (A1) *and* (A2), *then*

$$1 - \rho_X(h) \sim \left(\frac{\frac{1}{2\alpha+1} + \int_0^\infty \left((x+1)^\alpha - x^\alpha\right)^2 dx}{\int_0^\infty g(x)^2 dx} \right) L_0(|h|)^2 |h|^{2\alpha+1}, \quad |h| \downarrow 0.$$

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 (4.6) disappears. In this case, the parameter γ controls the asymptotic memory properties of the $\mathcal{BSS}$ process X.

Proposition 4.2. *Suppose that the kernel function g satisfies* (A3), *with $\lambda = 0$, and $\int_0^1 g(x)dx < \infty$.*

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

$$\rho_X(h) \sim \left(\frac{\int_0^\infty g(x)dx}{\int_0^\infty g(x)^2 dx} \right) L_1(|h|)|h|^{-\gamma}, \quad |h| \to \infty.$$

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

$$\rho_X(h) \sim \left(\frac{\int_0^\infty x^{-\gamma}(1+x)^{-\gamma} dx}{\int_0^\infty g(x)^2 dx} \right) L_1(|h|)^2 |h|^{1-2\gamma}, \quad |h| \to \infty.$$

Remark 4.1. In the critical case $\gamma = 1$, the asymptotic behavior of ρ_X is indeterminate under (A3), and would require additional assumptions on the slowly varying function L_1.

Remark 4.2. It follows from Proposition 4.2 that if the kernel function g satisfies (A3) with $\lambda = 0$, the resulting $\mathcal{BSS}$ 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 β as in (2.10), 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 4.3. *Suppose that the kernel function g satisfies* (A3) *with $\lambda > 0$ and $\gamma \in \mathbb{R}$, such that $\int_0^1 g(x)dx < \infty$. Then*

$$\rho_X(h) \sim \left(\frac{\int_0^\infty g(x)e^{-\lambda x}dx}{\int_0^\infty g(x)^2 dx} \right) e^{-\lambda|h|}|h|^{-\gamma}L_1(|h|), \quad |h| \to \infty.$$

Remark 4.3. Assumption (A1) is a sufficient condition for the requirement that $\int_0^1 g(x)dx < \infty$ in Propositions 4.2 and 4.3.

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 4.1 (Power law kernel)**.** Let g be the power law kernel

$$g(x) = x^\alpha(1+x)^{-\gamma-\alpha}, \qquad x > 0, \qquad \alpha \in \left(-\frac{1}{2}, \frac{1}{2} \right), \quad \gamma \in \left(\frac{1}{2}, \infty \right). \tag{4.7}$$

Bennedsen et al. (2017, Example 2.2) show that this kernel function indeed satisfies (A1), (A2), and (A3). In particular, with this kernel function, the $\mathcal{BSS}$ process X has roughness index α and memory properties controlled by γ, as expounded in Proposition 4.2. In the following, we will refer to the $\mathcal{BSS}$ process with the power law kernel as the *Power-$\mathcal{BSS}$* process.

Later we will need the correlation structure of the Power-$\mathcal{BSS}$ process. By covariance-stationarity of v, the autocovariance function of the general $\mathcal{BSS}$ process (4.4) is

$$c_X(h) := Cov(X_t, X_{t+h}) = \mathbb{E}[v_0^2] \int_0^\infty g(x)g(x+|h|)dx, \quad h \in \mathbb{R}. \tag{4.8}$$

From this we deduce that when g is given as in (4.7) we have

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

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 (4.8). Note that ρ_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 4.2 (Gamma kernel)**.** Let g be the gamma kernel

$$g(x) = x^\alpha e^{-\lambda x}, \qquad x > 0, \qquad \alpha \in \left(-\frac{1}{2}, \frac{1}{2} \right), \quad \lambda \in (0, \infty), \tag{4.9}$$

24

which satisfies (A1), (A2), and (A3), as shown in Bennedsen et al. (2017, Example 2.1). With this kernel function, the process X has roughness index α and memory properties controlled by λ, as per Proposition 4.3. In the following, we will call the $\mathcal{BSS}$ process with the gamma kernel the *Gamma-$\mathcal{BSS}$* process.

We will also need the correlation structure of the Gamma-$\mathcal{BSS}$ process. We easily find

$$
\begin{aligned}
c_X(0) := Var(X_t) = \mathbb{E}[v_0^2] \int_0^\infty x^{2\alpha} e^{-2\lambda} dx \\
= \mathbb{E}[v_0^2](2\lambda)^{-2\alpha-1}\Gamma(2\alpha+1),
\end{aligned}
$$

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

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

where $K_\nu(x)$ is the modified Bessel function of the third kind with index ν, 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 4.4. These two examples of kernel functions exemplify the theoretical distinction between long and short memory. In particular, by Proposition 4.3, the Gamma-$\mathcal{BSS}$ process adheres to

$$
\rho_X(h) \sim c \cdot e^{-\lambda h} h^\alpha, \quad h \to \infty,
$$

i.e., it has short memory, while the Power-$\mathcal{BSS}$ process will have polynomially decaying ACF and, in particular, the long memory property when $\gamma < 1$, cf. Remark 4.2. Although, theoretically, the Gamma-$\mathcal{BSS}$ process has short memory, by selecting very small values of λ, it is possible to specify processes with a very high degree of persistence, mimicking long memory on finite time intervals. Empirically, these two $\mathcal{BSS}$ 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.

4.2 Implications for raw volatility

The following results show that when log volatility is a $\mathcal{BSS}$ process, the roughness and memory properties will carry over to the volatility process itself. A result related to Theorem 4.1(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 $\mathcal{BSS}$ process, i.e., when $v_t = v > 0$ is constant for all t, but we conjecture that the results hold also for more general $\mathcal{BSS}$ processes, under suitable assumptions.

Let $\sigma = (\sigma_t)_{t \geq 0}$ be as in (4.1) and

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

The first part of the following theorem shows that σ inherits the roughness properties of the $\mathcal{BSS}$ process X. The second part shows that the same is true of the long-term memory properties.

Theorem 4.1. *Let σ be given by (4.1) where X is a $\mathcal{BSS}$ process satisfying (A1), (A2), and (A3), with $v_t = v > 0$ for all t. Then,*

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

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

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

$$\rho(h) \sim c \cdot \rho_X(h).$$

4.3   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 efficient 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 $\mathcal{BSS}$ 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 (2.1) and (4.1), 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 4. We can simulate S on a grid using Riemann-sum approximations of the integrals. To this end, we need to first simulate B and σ 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 $\mathcal{BSS}$ 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. (2017) in the general case. The hybrid scheme is tailor-made for $\mathcal{BSS}$ processes and its advantages are that (i) simulation is fast and in most cases accurate[12] (although approximate), (ii) it allows for non-Gaussianity of X through volatility (of volatility) modulation, and (iii) inclusion of leverage, i.e., correlation between X and B, is straightforward.

We refer to Bennedsen et al. (2017) for an exposition of the hybrid scheme for the $\mathcal{BSS}$ process X, defined by (4.4), under assumptions (A1), (A2), and (A3). The authors explain in the paper (Bennedsen et al., 2017, 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.

5   Estimating the models

Estimating the new models presented in Section 4 involves no complications. In particular, α can be estimated semiparametrically by the OLS regression (2.6) and β, similarly, by (2.11). The memory parameters γ and λ of the Power- and Gamma-$\mathcal{BSS}$ process, respectively, can be estimated by a method-of-moments procedure, described next.

As mentioned earlier, semiparametric estimation of the memory parameter β using (2.10) can be somewhat unreliable in finite samples, unfortunately. Therefore, we also estimate β 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 (4.2) and Power-$\mathcal{BSS}$ (4.8) processes, respectively.[13] One could simultaneously estimate α following the method-of-moments approach, but simulation results (available upon request) indicate that the OLS estimator of α is very precise, suggesting that the best performance overall is achieved when α is first estimated by OLS and then β 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 β (or γ or λ) only, and we then minimize the sum of squared distances between this function and the empirical ACF $\hat{\rho}$. We used $L = \lceil n^{1/3} \rceil$ lags in the ACF for this estimation procedure. While

---

[12]The hybrid scheme requires truncating the integral representation (4.4) "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.

[13]In the case of the Power-$\mathcal{BSS}$ process, we actually estimate the parameter γ using the method-of-moments procedure and express the estimate in terms of β using the relationship described in Remark 4.2.

Table 2: *Estimating the models on log and raw volatility*

Panel A: Log volatility

Δ	$\hat{\alpha}_{OLS}$	95% CI	$\hat{\lambda}_{BSS}$	95% CI	$\hat{\beta}_{OLS}$	$\hat{\beta}_{BSS}$	95% CI	$\hat{\beta}_{Cauchy}$	95% CI
10 minutes	-0.38	$(-0.39, -0.37)$	0.12	$(0.12, 0.12)$	0.19	0.53	$(0.52, 0.54)$	0.25	$(0.24, 0.26)$
15 minutes	-0.34	$(-0.35, -0.33)$	0.19	$(0.19, 0.20)$	0.24	0.70	$(0.68, 0.72)$	0.33	$(0.32, 0.34)$
30 minutes	-0.30	$(-0.32, -0.27)$	0.22	$(0.20, 0.24)$	0.21	0.78	$(0.72, 0.83)$	0.40	$(0.39, 0.41)$
65 minutes	-0.29	$(-0.33, -0.26)$	0.16	$(0.14, 0.18)$	0.23	0.65	$(0.60, 0.70)$	0.38	$(0.36, 0.40)$
130 minutes	-0.33	$(-0.37, -0.28)$	0.10	$(0.09, 0.10)$	0.54	0.52	$(0.46, 0.57)$	0.33	$(0.30, 0.36)$
1 day	-0.30	$(-0.36, -0.24)$	0.10	$(0.09, 0.11)$	1.14	0.61	$(0.35, 0.87)$	0.41	$(0.32, 0.51)$

Panel B: Raw volatility

Δ	$\hat{\alpha}_{OLS}$	95% CI	$\hat{\lambda}_{BSS}$	95% CI	$\hat{\beta}_{OLS}$	$\hat{\beta}_{BSS}$	95% CI	$\hat{\beta}_{Cauchy}$	95% CI
10 minutes	-0.40	$(-0.51, -0.29)$	0.47	$(0.43, 0.50)$	0.20	1.18	$(1.13, 1.24)$	0.37	$(0.36, 0.39)$
15 minutes	-0.43	$(-0.59, -0.26)$	0.24	$(0.22, 0.27)$	0.18	0.89	$(0.81, 0.98)$	0.28	$(0.27, 0.29)$
30 minutes	-0.37	$(-0.49, -0.24)$	0.31	$(0.27, 0.36)$	0.25	1.03	$(0.95, 1.11)$	0.43	$(0.41, 0.44)$
65 minutes	-0.40	$(-0.54, -0.26)$	0.13	$(0.12, 0.14)$	0.37	0.64	$(0.58, 0.70)$	0.32	$(0.29, 0.34)$
130 minutes	-0.37	$(-0.48, -0.25)$	0.11	$(0.11, 0.12)$	0.74	0.60	$(0.50, 0.70)$	0.35	$(0.31, 0.39)$
1 day	-0.26	$(-0.46, -0.07)$	0.17	$(0.15, 0.19)$	1.57	0.88	$(0.43, 1.33)$	0.55	$(0.40, 0.70)$

Estimates of α using the OLS regression (2.6) and of β using the OLS regression (2.11), 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-$\mathcal{BSS}$ models. We used $L = \lceil n^{1/3} \rceil$ lags to estimate the memory parameters λ_{BSS}, β_{BSS}, and β_{Cauchy}. The estimate of β_{BSS} is calculated from the estimate of method-of-moments estimate of γ of the Power-$\mathcal{BSS}$ process; that is, when $\hat{\gamma} > 1$, set $\hat{\beta}_{BSS} = \hat{\gamma}$, otherwise set $\hat{\beta}_{BSS} = 2\hat{\gamma} - 1$.

this choice is somewhat arbitrary, the results were rather robust to alternative choices. See Figure 12 for the fitted ACFs and Section 5.1 below for a discussion of the results.

In Table 2, Panel A, we report the estimates of α, λ, and β for the different models using the data on log volatility of the E-mini S&P 500 futures contracts, extracted as explained in Section 2.1. In Sections 2.3 and 2.4, we have already seen the OLS estimates of α and β, 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-$\mathcal{BSS}$ model are a somewhat larger than $\hat{\beta}_{OLS}$ and $\hat{\beta}_{Cauchy}$, at least for the higher frequencies, but the estimates are still rather small and less than one in all cases. These results strengthen our confidence in the semiparametric estimates of β, and underscore the strong persistence of volatility.

Similarly, when we estimate the Gamma-$\mathcal{BSS}$ model, which does not have the long memory property (cf. Remark 4.4), we get estimates of λ_{BSS} ranging from 0.0027 to 0.13, which indicates very slow decay of the ACF.

Lastly, in Table 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 4.1, raw volatility largely inherits the roughness and memory properties of log volatility, cf. Panel A.
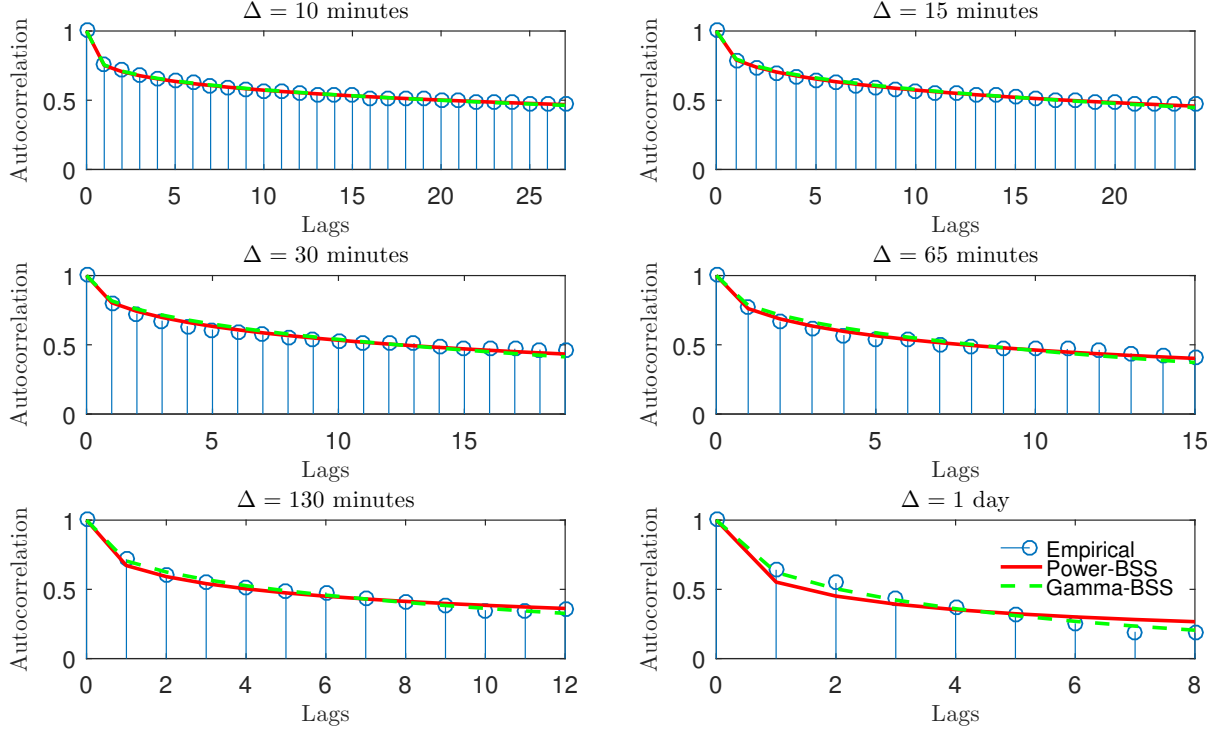
Figure 12: *Empirical autocorrelation functions of log volatility (light blue bars and circles) and fitted ACFs of the Power-$\mathcal{BSS}$ (solid red line) and Gamma-$\mathcal{BSS}$ (dashed green line) models. We used $L = \lceil n^{1/3} \rceil$ lags when fitting the ACFs; this corresponds to the rightmost point of each plot.*

5.1 The autocorrelation of (log) volatility

We plot the ACFs of the estimated $\mathcal{BSS}$ models together with the empirical ACFs of log volatility data in Figure 12. 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 $\mathcal{BSS}$ 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-$\mathcal{BSS}$ model and that of the Gamma-$\mathcal{BSS}$ model. If anything, the 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 12.

5.2 Estimating the models on volatility data on 26 major equities

To further assess the different models introduced above and their estimates, we also apply the models to a number of individual equities. More precisely, we analyze a subset of 26 assets extracted

from the pool of 1 944 US equities studied in Section 3. We select 26 blue-chip equities identified by their ticker symbols in Table 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 presents also the estimates of the parameters, in a similar fashion as in Section 5, 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 α 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 λ in the Gamma-$\mathcal{BSS}$ model is just 0.0038, while the average estimates of the memory parameters β_{OLS}, β_{BSS}, and β_{Cauchy} are 0.13, 0.15, and 0.11, respectively.

6 Application to volatility forecasting

In this section we apply the $\mathcal{BSS}$ 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.

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-$\mathcal{BSS}$ and Gamma-$\mathcal{BSS}$ 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).[14]

The ARMA and ARFIMA models were estimated by maximum likelihood.[15] The AR models, as well as the log-HAR model, were estimated by OLS. We apply an intraday adaptation of the

---

[14]While the $\mathcal{BSS}$ models are initially defined in *continuous time* for the log volatility process $X_t = \log(\sigma_t/\xi)$, $t \geq 0$, cf. equation (4.1), we can apply them, or rather their correlation structure, here in *discrete time* to forecast IV via its estimate $BV^{\Delta*}$, as motivated by the approximation of σ_t by $\hat{\sigma}_t^2 = \Delta^{-1}\widehat{IV}_t^{\Delta}$, cf. equation (2.2).

[15]The ARMA model was estimated and forecasted using the MFE toolbox of Kevin Sheppard, see: https://www.kevinsheppard.com. The ARFIMA models were estimated and forecasted using the MATLAB package "ARFIMA(p,d,q) estimator" available from MATLAB Central.

Table 3: *Estimating the models on daily log BV* measures of individual US stocks*

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

Estimates of α using the OLS regression (2.6) and of β using the OLS regression (2.11) 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-BSS models. The estimates of λ 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.

standard log-HAR model. The time periods used in constructing the HAR regressors are still one day, one week, and one month, respectively, but the regressors need to be adapted to the step size Δ, which may now be less than a day. More precisely, our log-HAR regression is

$$\log\!\big(BV_{t+\Delta h}^{\Delta *}\big) = a_0 + a_1 \log\big(BV_t^{\Delta *}\big) + a_2 \log\Big(BV_t^{\Delta *,\text{day}}\Big) + a_3 \log\Big(BV_t^{\Delta *,\text{week}}\Big) \tag{6.1}$$
$$+ a_4 \log\Big(BV_t^{\Delta *,\text{month}}\Big) + \epsilon_{t+\Delta h},$$

where

$$BV_t^{\Delta *,x} := \frac{1}{q}\sum_{k=0}^{q-1} BV_{t-k\Delta}^{\Delta *}, \quad x = \text{day, week, month,}$$

and q is an integer such that $q\Delta = x$. For instance, when $\Delta = 65$ minutes then $q = 6$ for $x = \text{day}$ (6 periods of 65 minutes during a trading day) whereas $q = 30$ for $x = \text{week}$ (5 trading days, each consisting of 6 periods, in a week).[16]

The estimate of the Hurst index H in the RFSV model is derived via $\hat{H} = \hat{\alpha}_{OLS} + 0.5$, and for the Cauchy and $\mathcal{BSS}$ processes we use the parametric estimates of β, γ, and λ 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))

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

where $\mathcal{F}_t$ is the information set (σ-algebra) generated by the fBm driving the model up to time t. The integrals are approximated by Riemann sums.

To forecast the Cauchy and $\mathcal{BSS}$ processes we rely on the elementary result that for a zero-mean Gaussian random vector $(x_{t+h}, x_t, x_{t-1}, \ldots, x_{t-m})^T$ the distribution of x_{t+h} conditionally on $(x_t, x_{t-1}, \ldots, x_{t-m})^T = a \in \mathbb{R}^{m+1}$ is

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

where

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

---

[16]Although not indicated here, in the variables entering the log-HAR model, all pre-averaged estimates of integrated variance are de-seasonalized when we estimate and forecast the model for $\Delta < 1$ day. The seasonal factor is re-introduced after estimation and forecasting.

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

$$\Gamma_{12} := (Corr(x_{t+h}, x_t), Corr(x_{t+h}, x_{t-1}), \ldots, Corr(x_{t+h}, x_{t-m})).$$

Since the processes we consider here are stationary, the variance ξ^2 of the conditional distribution is

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

where $\Gamma_{21} = \Gamma_{12}^T$.

To implement this procedure for the Cauchy and $\mathcal{BSS}$ 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 μ needs to be calculated. However, as we will argue in the next section, the conditional variance term ξ^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.[17]

6.1 Forecasting intraday integrated variance

The previous section lays out methods of forecasting log volatility or, equivalently, the process X, cf. equation (4.1). 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 $\mathcal{BSS}$ and Cauchy models we follow the strategy of the preceding section. That is, if we again assume Gaussianity, we have

$$\mathbb{E}[\exp(X_{t+\Delta})|\mathcal{F}_t] = \exp\left(\mathbb{E}[X_{t+\Delta}|\mathcal{F}_t] + \frac{1}{2}Var[X_{t+\Delta}|\mathcal{F}_t]\right). \tag{6.2}$$

We will then approximate the former term in the exponential function by μ and the latter by $\frac{1}{2}\xi^2$. Note that ξ^2 depends on the (stationary) variance of the process, $Var(x_t)$; 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 $BV^{\Delta*}$ data, cf. equation (6.1), we exponentiate these estimates and make a correction similar to (6.2), where the variance factor is estimated as the variance of the error term in the OLS regression (6.1). The remaining models are estimated using directly the raw (de-seasonalized) $BV^{\Delta*}$ data, so no correction is needed.

---

[17]In our model for volatility, this de-meaning essentially means removing the term $\log \xi$, cf. equation (4.1). We reintroduce this term after forecasting X.

6.1.1 Forecast setup

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

$$\mathrm{FO}_t(\Delta, h) := \sum_{k=1}^{h} BV_{t+k\Delta}^{\Delta*} \approx \int_t^{t+h\Delta} \sigma_s^2 ds, \quad h = 1, 2, 5, 10, 20, \tag{6.3}$$

where $BV_t^{\Delta*}$ is the estimated value of integrated variance using the pre-averaged bipower variation estimator, cf. Section 2.1.

To forecast the FO in (6.3), we compute the h individual components, $\widehat{\sigma^2}_{t+k\Delta|t}$, $k = 1, \ldots, h$, multiply by Δ and the seasonal component, and sum them up:

$$\widehat{\mathrm{FO}}_t(\Delta, h) = \sum_{k=1}^{h} \widehat{\sigma^2}_{t+k\Delta|t} \left(\sigma_{t+k\Delta}^s\right)^2 \Delta,$$

where σ_t^s is the (deterministic) seasonal component of volatility we extracted in the preliminary step of our analysis, as explained in Section 2.1, and $\widehat{\sigma^2}_{t+k\Delta|t}$ is the forecast of volatility, as detailed above.[18]

In the forecasting experiments we consider various step sizes Δ, 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 (MSE):

$$\mathrm{MSE}(\Delta, h) = \frac{1}{n-h-m+1} \sum_{t=m}^{n-h} \left| \widehat{\mathrm{FO}}_t(\Delta, h) - \mathrm{FO}_t(\Delta, h) \right|^2,$$

- "Quasi-likelihood" (QLIKE):

$$\mathrm{QLIKE}(\Delta, h) = \frac{1}{n-h-m+1} \sum_{t=m}^{n-h} \left(\log \widehat{\mathrm{FO}}_t(\Delta, h) + \frac{\mathrm{FO}_t(\Delta,h)}{\widehat{\mathrm{FO}}_t(\Delta,h)} \right).$$

As discussed in Section 2.1, the the pre-averaged 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 et al. (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,

---

[18]To keep the procedure realistic, in the out-of-sample experiment discussed below, the seasonal component is estimated in a non-anticipative fashion, using only the observations that would be available at the time when the forecast is produced.

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.

We set up our forecasting experiment so that it is realistic and mirrors the situation a practitioner would face if they were to forecast intraday volatility. 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, to allow for time variation in the parameters, and compute new forecasts h steps ahead. In re-estimation, we used a rolling window of 200 observations. While this window length is somewhat arbitrary, brief experimentation with other window lengths suggested that our results are not particularly dependent on this choice.

We assess the forecasting performance of the different models over two time periods using the E-mini S&P 500 data set. First we forecast over a long out-of-sample period from from January 3, 2005 to December 31, 2014, using the the entire data set at our disposal (cf. Section 2.3.2, where we presented evidence of the roughness parameter varying in time). This period contains a couple of abrupt market crashes (e.g., the Lehman Brothers bankruptcy and Flash Crash), precipitated by circumstances that would have been hard to predict by virtually any model based on historical data. We observed that failing to forecast the volatility bursts during such episodes leads to disproportionately large losses with both MSE and QLIKE, and especially with MSE, these individual losses can represent a significant proportion of the total loss. This phenomenon also affects the MCS procedure, which would not be very selective over this long out-of-sample period. For this reason, we also experiment with another, shorter out-of-sample period from from January 2, 2013 to December 31, 2014. Volatility over this period is less pronounced, exemplifying "non-stressed" market conditions. (Recall that we initially studied this period in Section 2.)

6.1.2 Results of out-of-sample forecasting experiment

Figure 13 plots the cumulative QLIKE losses for select models in the case of forecasting intraday ($\Delta = 15$ minutes) integrated variance over the long out-of-sample period from 2005 to 2014. We employ the QLIKE loss function since it is less prone to exaggerating the impact of large forecast errors than MSE (Patton, 2011). We present our results on the cumulative losses relative to those arising when the Gamma-$\mathcal{BSS}$ model is used. More precisely, the graphs present, as a function of time t,

$$\text{Cumulative QLIKE (vs. Gamma-BSS)}_t^x := \sum_{k=1}^{\lfloor t/\Delta \rfloor} \left(\text{QLIKE}(x)_k - \text{QLIKE}(\text{Gamma-}\mathcal{BSS})_k \right),$$

where $x \in \{\text{HAR}, \text{RFSV}, \text{Cauchy}, \text{Power-BSS}\}$ denotes the model being compared to the Gamma-$\mathcal{BSS}$ model, and $\text{QLIKE}(x)_k$ stands for the k-th loss using model x. With this definition, positive numbers indicate that the model performs worse than the Gamma-$\mathcal{BSS}$ model and vice versa for
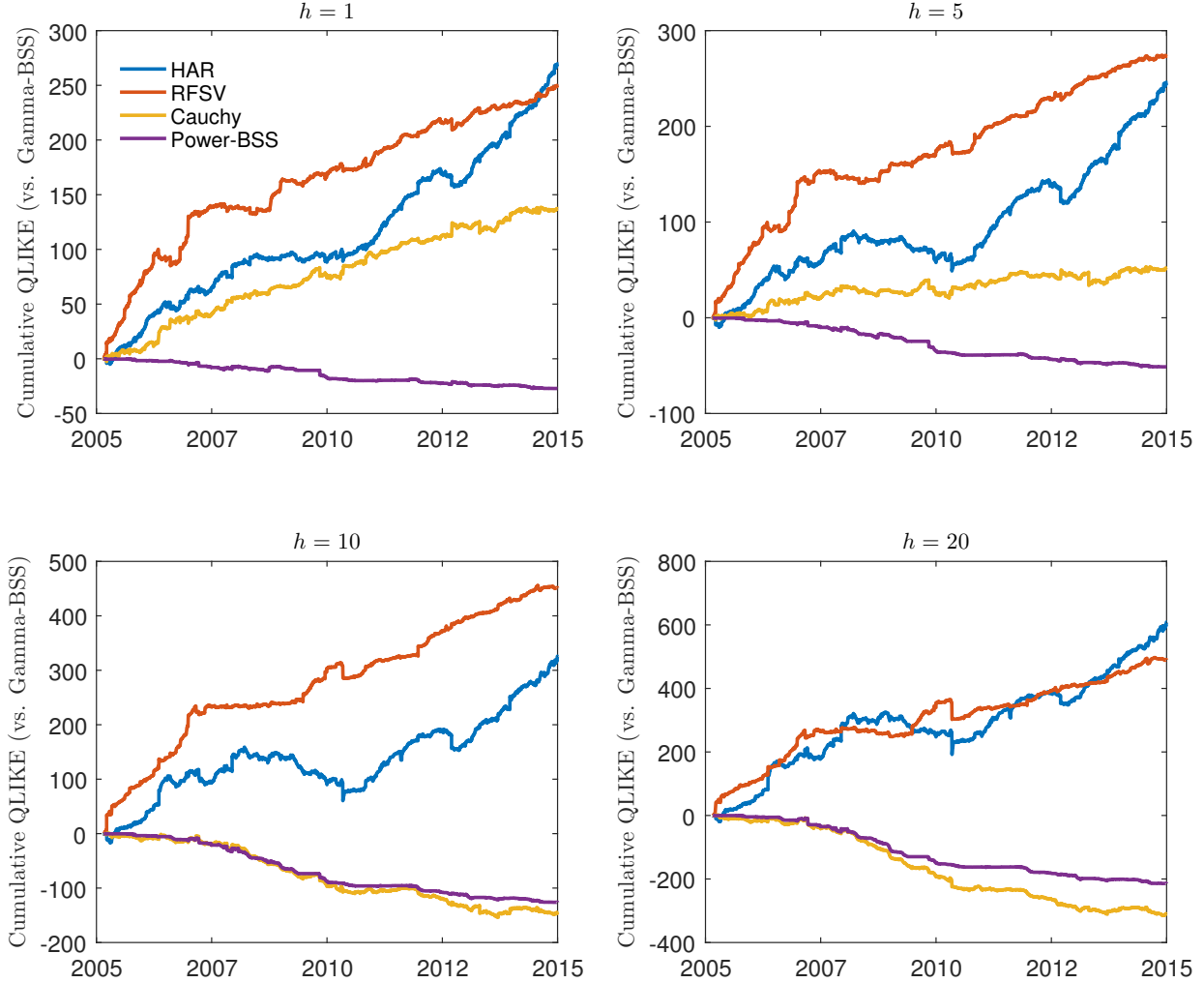
Figure 13: *Cumulative QLIKE forecast errors through time. Here, h is the forecast horizon (i.e., volatility is forecasted over a time interval of length $h\Delta$) and $\Delta = 15$ minutes.*

negative numbers.

We observe that both the Gamma-$\mathcal{BSS}$ and especially the Power-$\mathcal{BSS}$ models perform well over the entire period. Interestingly, as the forecasting horizon h is increased, the Cauchy model performs better relative to the Gamma-$\mathcal{BSS}$ model, suggesting that the long memory property of the Cauchy model, which is absent from the Gamma-$\mathcal{BSS}$ model, becomes relevant.

Turning now to the shorter out-of-sample period from January 2, 2013 to December 31, 2014 that represents calm market conditions, we present more detailed forecasting results for $\Delta = 15, 30, 65$ minutes and $\Delta = 1$ day in Tables 4 and 5. There, a boldface number identifies the model with the smallest loss, while light grey background highlights the models in the 90% MCS and dark grey background those in the 75% MCS. The models proposed in this paper generally outperform the benchmarks over this period. Indeed, in most cases either the Cauchy or the Power-$\mathcal{BSS}$ model has the lowest MSE and QLIKE losses. Further, these two models are very often in the

36

75% MCS. These findings are most convincing for the intraday values of Δ, while, interestingly, RFSV turns out to be the best performing model when $\Delta = 1$ day. (Unfortunately, the MCS is not very selective in this case, at least when using MSE as the loss function.) This result agrees with Gatheral et al. (2014), where the authors demonstrated that the RFSV model can outperform a range of benchmark models when forecasting daily integrated variance one step ahead. All in all, these results suggest that it is in general advantageous to exploit the roughness of volatility in forecasting, and at intraday time scales, careful modeling of persistence can further improve the quality of forecasts.

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 often has slightly heavier tails than the lognormal distribution.

Moreover, by also looking at volatility measurements on almost two 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 stochastic 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, particularly at intraday time scales.

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 and optimal execution. 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 4: *Out-of-sample forecasting of intraday integrated variance*

Panel A: Δ = 15 minutes

	h = 1		h = 2		h = 5		h = 10		h = 20	
	MSE ×10^11	QLIKE	MSE ×10^11	QLIKE	MSE ×10^10	QLIKE	MSE ×10^9	QLIKE	MSE ×10^8	QLIKE
RW	0.229	−12.883	0.733	−12.153	0.471	−11.147	0.245	−10.355	0.102	−9.553
AR5	0.178	−12.928	0.472	−12.214	0.220	−11.252	0.079	−10.509	0.027	−9.754
AR10	0.187	−12.931	0.494	−12.217	0.226	−11.257	0.077	−10.514	0.027	−9.760
ARMA(1,1)	0.159	−12.932	0.433	−12.217	0.203	−11.255	0.074	−10.509	0.027	−9.748
log-HAR3	0.149	−12.941	0.408	−12.225	0.198	−11.260	0.079	−10.512	0.038	−9.745
ARFIMA(0, d, 0)	0.145	−12.937	0.393	−12.222	0.176	−11.262	0.059	−10.522	0.021	−9.772
RFSV	0.145	−12.947	0.384	−12.232	0.170	−11.268	0.057	−10.525	0.020	−9.774
Cauchy	0.143	−12.948	0.382	−12.234	0.164	−11.274	0.053	−10.533	**0.019**	**−9.783**
Power-BSS	**0.142**	**−12.950**	**0.377**	**−12.235**	**0.163**	**−11.274**	**0.053**	**−10.533**	0.019	−9.783
Gamma-BSS	0.142	−12.949	0.377	−12.234	0.164	−11.273	0.054	−10.531	0.019	−9.780

Panel B: Δ = 30 minutes

	h = 1		h = 2		h = 5		h = 10		h = 20	
	MSE ×10^11	QLIKE	MSE ×10^10	QLIKE	MSE ×10^9	QLIKE	MSE ×10^9	QLIKE	MSE ×10^8	QLIKE
RW	0.596	−12.181	0.220	−11.435	0.170	−10.404	0.612	−9.617	0.264	−8.874
AR5	0.500	−12.213	0.154	−11.490	0.074	−10.512	0.259	−9.758	0.113	−9.037
AR10	0.501	−12.214	0.151	−11.491	0.073	−10.513	0.262	−9.761	0.117	−9.043
ARMA(1,1)	0.459	−12.215	0.142	−11.492	0.073	−10.511	0.261	−9.753	0.119	−9.023
log-HAR3	0.433	−12.227	0.131	−11.501	0.070	−10.517	0.300	−9.762	0.160	−9.032
ARFIMA(0, d, 0)	0.432	−12.222	0.129	−11.498	0.062	−10.522	0.218	−9.772	0.088	−9.055
RFSV	0.438	−12.230	0.128	−11.504	0.060	−10.523	0.206	−9.772	0.089	−9.059
Cauchy	0.418	−12.229	0.122	−11.506	**0.057**	**−10.530**	0.199	**−9.781**	**0.080**	**−9.067**
Power-BSS	**0.416**	**−12.231**	**0.122**	**−11.506**	0.057	−10.529	0.199	−9.780	0.081	−9.065
Gamma-BSS	0.418	−12.230	0.124	−11.504	0.059	−10.526	0.205	−9.776	0.085	−9.059

Out-of-sample Mean Squared Forecast Error (MSE) and QLIKE for all models considered in the paper. Forecasting period: January 2, 2013, to December 31, 2014. Bold numbers indicate the model with the smallest forecast error (column-wise). The forecast object is the sum of realized kernels (6.3), approximating integrated variance, as explained in the text. We vary the step size Δ 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 5: *Out-of-sample forecasting of intraday integrated variance*

Panel A: Δ = 65 minutes

	h = 1		h = 2		h = 5		h = 10		h = 20	
	MSE $\times 10^{10}$	QLIKE	MSE $\times 10^{9}$	QLIKE	MSE $\times 10^{9}$	QLIKE	MSE $\times 10^{8}$	QLIKE	MSE $\times 10^{7}$	QLIKE
RW	0.314	−11.355	0.147	−10.579	0.675	−9.557	0.297	−8.822	0.127	−8.079
AR5	0.237	−11.389	0.074	−10.648	0.367	−9.659	0.158	−8.947	0.078	−8.226
AR10	0.247	−11.337	0.076	−10.615	0.378	−9.661	0.178	−8.949	0.093	−8.226
ARMA(1,1)	0.221	−11.390	0.071	−10.649	0.347	−9.656	0.154	−8.935	0.078	−8.206
log-HAR3	0.207	−11.404	0.067	−10.657	0.370	−9.662	0.178	−8.938	0.104	−8.189
ARFIMA(0, d, 0)	0.205	−11.397	0.063	−10.658	0.302	−9.673	0.116	−8.960	0.047	−8.240
RFSV	0.201	**−11.409**	0.062	−10.664	0.281	−9.676	0.109	−8.966	0.047	**−8.250**
Cauchy	0.199	−11.406	0.059	−10.668	0.285	**−9.683**	0.108	**−8.970**	**0.043**	−8.248
Power-BSS	**0.195**	−11.409	**0.058**	**−10.668**	**0.279**	−9.682	**0.107**	−8.970	0.044	−8.247
Gamma-BSS	0.197	−11.407	0.059	−10.665	0.282	−9.679	0.109	−8.967	0.046	−8.242

Panel B: Δ = 1 day

	h = 1		h = 2		h = 5		h = 10		h = 20	
	MSE $\times 10^{9}$	QLIKE	MSE $\times 10^{8}$	QLIKE	MSE $\times 10^{7}$	QLIKE	MSE $\times 10^{6}$	QLIKE	MSE $\times 10^{5}$	QLIKE
RW	0.614	−9.491	0.253	−8.783	0.170	−7.818	0.076	−7.045	0.035	−6.218
AR5	0.717	−9.531	0.326	−8.826	0.179	−7.882	0.053	−7.146	0.013	−6.405
AR10	0.729	−8.717	0.338	−7.976	0.185	−7.876	0.055	−7.142	0.014	−6.403
ARMA(1,1)	0.621	−9.537	0.266	−8.831	0.232	−7.883	0.189	−7.146	0.496	−6.400
log-HAR3	0.506	−9.543	0.179	−8.838	0.096	−7.890	0.034	−7.154	0.011	−6.403
ARFIMA(0, d, 0)	0.542	−9.537	0.191	−8.830	0.104	−7.881	0.038	−7.144	0.013	−6.395
RFSV	**0.493**	−9.546	**0.171**	**−8.843**	0.091	**−7.898**	0.034	**−7.160**	0.014	−6.402
Cauchy	0.579	−9.537	0.195	−8.831	0.098	−7.884	0.034	−7.151	0.011	−6.401
Power-BSS	0.545	−9.543	0.186	−8.837	0.095	−7.889	0.033	−7.154	0.011	−6.405
Gamma-BSS	0.513	**−9.547**	0.176	−8.842	**0.091**	−7.895	**0.032**	−7.159	**0.010**	**−6.412**

Out-of-sample Mean Squared Forecast Error (MSE) and QLIKE for all models considered in the paper. Forecasting period: January 2, 2013, to December 31, 2014. Bold numbers indicate the model with the smallest forecast error (column-wise). The forecast object is the sum of realized kernels (6.3), approximating integrated variance, as explained in the text. We vary the step size Δ 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.

References

Alòs, E., J. A. León, and J. Vives (2007). On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics 11*(4), 571–589.

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance 4*(2–3), 115–158.

Andersen, T. G. and T. Bollerslev (1998). DM-dollar volatility: Intraday activity patterns, macroeconomic announcements and longer run dependencies. *Journal of Finance 53*(1), 219–265.

Andersen, T. G., T. Bollerslev, and J. Cai (2000). Intraday and interday volatility in the Japanese stock market. *Journal of International Financial Markets, Institutions and Money 10*, 107–130.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*(453), 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Andersen, T. G., T. Bollerslev, and S. Lange (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance 6*(5), 457–477.

Andersen, T. G., O. Bondarenko, A. S. Kyle, and A. A. Obizhaeva (2016). Intraday trading invariance in the E-mini S&P 500 futures market. Working paper, available at: http://ssrn.com/abstract=2693810.

Asmussen, S. and P. W. Glynn (2007). *Stochastic simulation: algorithms and analysis*. New York: Springer.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*(1), 1–13.

Barndorff-Nielsen, O. E., F. E. Benth, and A. E. D. Veraart (2013). Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. *Bernoulli 19*(3), 803–845.

Barndorff-Nielsen, O. E., J. M. Corcuera, and M. Podolskij (2009). Power variation for Gaussian processes with stationary increments. *Stochastic Processes and their Applications 119*(6), 1845–1865.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(8), 1481–1536.

Barndorff-Nielsen, O. E. and J. Schmiegel (2007). Ambit processes: with applications to turbulence and tumour growth. In *Stochastic analysis and applications*, Volume 2 of *Abel Symp.*, pp. 93–124. Berlin: Springer.

Barndorff-Nielsen, O. E. and J. Schmiegel (2009). Brownian semistationary processes and volatility/intermittency. In *Advanced financial modelling*, Volume 8 of *Radon Series on Computational and Applied Mathematics*, pp. 1–25. Berlin: Walter de Gruyter.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*(2), 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 1*(2), 1–48.

Bayer, C., K. Friz, and J. Gatheral (2016). Pricing under rough volatility. *Quantitative Finance 16*(6), 887–904.

Bennedsen, M. (2016). Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data. Working paper, available at: `http://arxiv.org/abs/1608.01895`.

Bennedsen, M., A. Lunde, and M. S. Pakkanen (2017). Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, to appear: `http://dx.doi.org/10.1007/s00780-017-0335-5`.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular variation*. Cambridge: Cambridge University Press.

Bollerslev, T. and J. H. Wright (2000). Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics 98*(1), 81–106.

Bos, C. S., P. Janus, and S. J. Koopman (2012). Spot variance path estimation and its application to high-frequency jump-testing. *Journal of Financial Econometrics 10*(2), 356–389.

Christensen, K., R. C. A. Oomen, and M. Podolskij (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics 114*(3), 576–599.

Comte, F. (1996). Simulation and estimation of long memory continuous time models. *Journal of Time Series Analysis 17*(1), 19–36.

Comte, F., L. Coutin, and E. Renault (2012). Affine fractional stochastic volatility models. *Annals of Finance 8*(3), 337–378.

Comte, F. and E. Renault (1996). Long memory continuous time models. *Journal of Econometrics 73*(1), 101–149.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*(4), 291–323.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Dudley, R. M. and R. Norvaiša (2011). *Concrete functional calculus*. New York: Springer.

Fukasawa, M. (2017). Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance 17*(2), 189–198.

Gatheral, J. (2006). *The volatility surface*. Hoboken: Wiley.

Gatheral, J., T. Jaisson, and M. Rosenbaum (2014). Volatility is rough. Working paper, available at: `http://arxiv.org/abs/1410.3394`.

Gneiting, T. and M. Schlather (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review 46*(2), 269–282.

Gradshteyn, I. S. and I. M. Ryzhik (2007). *Table of integrals, series, and products* (Seventh ed.). Amsterdam: Academic Press.

Handcock, M. S. and M. L. Stein (1993). A Bayesian analysis of kriging. *Technometrics 35*(4), 403–410.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics 24*(2), 127–161.

Hansen, P. R. and A. Lunde (2013). Estimating the persistence and the autocorrelation function of a time series that is measured with noise. *Econometric Theory 30*(1), 1–34.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications 119*(7), 2249–2276.

Jaisson, T. and M. Rosenbaum (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability 26*(5), 2860–2882.

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory 23*(1), 60–93.

Marinucci, D. and P. M. Robinson (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference 80*(1-2), 111–122.

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

Oomen, R. C. A. (2006). Comment on 2005 JBES invited address "Realized variance and market microstructure noise" by Peter R. Hansen and Asger Lunde. *Journal of Business and Economic Statistics 2*(24), 195–202.

Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics 160*(1), 246–256.

Rossi, E. and D. Fantazzini (2015). Long memory and periodicity in intraday volatility. *Journal of Financial Econometrics 13*(4), 922–961.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.

Zu, Y. and H. P. Boswijk (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics 181*(2), 117–135.

A Proofs

Proof of Theorem 2.1. Let

$$Q_n(\theta) = Q_n(a, b, \alpha) := \sum_{k=1}^{m} \left(\hat{\gamma}_2^*(k\Delta) - a - b(k\Delta)^{2\alpha+1} \right)^2,$$

and

$$Q(\theta) = Q(a, b, \alpha) := \sum_{k=1}^{m} \left(a_0 + b_0(k\Delta)^{2\alpha_0+1} - a - b(k\Delta)^{2\alpha+1} \right)^2,$$

where $a_0 := 2\sigma_u^2$. Since, by the assumptions on X and Z, $\hat{\gamma}_2^*(k\Delta)$ is a consistent estimator of $a_0 + b_0(k\Delta)^{2\alpha_0+1}$, it is not hard to show that

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \to 0, \quad n \to \infty. \tag{A.1}$$

Because $m \geq 3$, $\theta_0 := (a_0, b_0, \alpha_0)$ is the unique minimizer of $Q(\theta)$ (in fact $Q(\theta_0) = 0$), the rest of the proof therefore follows standard arguments, which we give here for completeness.

Let $\epsilon > 0$. For n sufficiently large, we have, because of the uniform convergence (A.1), that

$$|Q(\hat{\theta}) - Q_n(\hat{\theta})| < \epsilon.$$

Using first that $Q(\theta_0) = 0$, then that $\hat{\theta}$ minimizes $Q_n(\theta)$, and finally the uniform convergence (A.1) again, it holds that

$$|Q_n(\hat{\theta}) - Q(\theta_0)| = Q_n(\hat{\theta}) \leq Q_n(\theta_0) < \epsilon,$$

for n sufficiently large. Let B be an open neighborhood of θ_0. Now,

$$\begin{aligned}
\mathbb{P}(\hat{\theta} \in B^c \cap \Theta) &\leq \mathbb{P}(|Q(\hat{\theta}) - Q(\theta_0)| > 0) \\
&\leq \mathbb{P}(|Q(\hat{\theta}) - Q_n(\hat{\theta})| + |Q_n(\hat{\theta}) - Q(\theta_0)| > 0) \\
&\to 0, \quad n \to \infty,
\end{aligned}$$

by the results above, since $\epsilon > 0$ was arbitrary. $\qquad\square$

A.1 Proofs of Propositions 4.2 and 4.3

Recall first that we can write

$$\rho_X(h) = \frac{\int_0^\infty g(x)g(x + |h|)dx}{\int_0^\infty g(x)^2 dx}, \quad h \in \mathbb{R}. \tag{A.2}$$

Proof of Proposition 4.2. (i) We may assume that $h > 1$, since we let $h \to \infty$. By the assumption (A3), we may write

$$\begin{aligned}
\int_0^\infty g(x)g(x + h)dx &= \int_0^\infty g(x)(x + h)^{-\gamma}L_1(x + h)dx \\
&= h^{-\gamma}L_1(h) \int_0^\infty g(x)(x/h + 1)^{-\gamma}\frac{L_1(x + h)}{L(h)}dx \\
&= h^{-\gamma}L_1(h) \int_0^\infty g(x)(x/h + 1)^{-\gamma}\frac{L_1(h(x/h + 1))}{L(h)}dx
\end{aligned}$$

$$\sim h^{-\gamma} L_1(h) \int_0^\infty g(x) dx, \quad h \to \infty,$$

using 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,

$$g(x) \frac{L_1\left((x/h + 1)h\right)}{L_1(h)} < g(x)(1 + \epsilon), \quad x \in (0, \infty),$$

which is integrable over $(0, 1)$ by assumption and over $[1, \infty)$ since $\gamma > 1$.

(ii) Let first

$$\int_0^\infty g(x) g(x + h) dx = \int_0^1 g(x)(x + h)^{-\gamma} L_1(x + h) dx$$
$$+ \int_1^\infty x^{-\gamma}(x + h)^{-\gamma} L_1(x) L_1(x + h) dx$$
$$=: I_{1,h} + I_{2,h},$$

where

$$I_{1,h} = \int_0^1 g(x)(x + h)^{-\gamma} L_1(x + h) dx,$$
$$I_{2,h} = \int_1^\infty x^{-\gamma}(x + h)^{-\gamma} L_1(x) L_1(x + h) dx.$$

We may write

$$I_{2,h} = h^{-\gamma} \int_1^\infty x^{-\gamma}\left(1 + \frac{x}{h}\right)^{-\gamma} L_1(x) L_1(x + h) dx$$
$$= h^{-2\gamma+1} L_1(h)^2 \int_{1/h}^\infty \underbrace{y^{-\gamma}(1 + y)^{-\gamma} \frac{L_1(hy)}{L_1(h)} \frac{L_1(h(1 + y))}{L_1(h)}}_{=:k_h(y)} dy,$$

where the second equality follows by substituting $y = x/h$. Fix now $\delta \in (0, 1 - \gamma) \subset (0, 1/2)$. By the Potter bounds (Bingham et al., 1989, Theorem 1.5.6(ii)), under (A3) there exists a constant $C_\delta > 0$ such that

$$\frac{L_1(hy)}{L_1(h)} \le C_\delta \max\{y^{-\delta}, y^\delta\}, \quad y > 1/h, \quad h > 1.$$

Accordingly, we find a dominant for k_h, given by

$$k_h(y) \le \overline{k}(y) := \begin{cases} C_\delta y^{-\gamma-\delta}(1 + y)^{-\gamma+\delta}, & y \in (0, 1], \\ C_\delta y^{-2(\gamma-\delta)}, & y \in (1, \infty), \end{cases}$$

for any $h > 1$. Note that the dominant $\overline{k}$ is integrable on $(0, 1]$, since $-\gamma - \delta > -\gamma - (1 - \gamma) = -1$, as well as on $(1, \infty)$, since $-2(\gamma - \delta) < -1$. Applying the dominated convergence theorem, we get

$$I_{2,h} \sim h^{-2\gamma+1} L_1(h)^2 \int_0^\infty y^{-\gamma}(1 + y)^{-\gamma} dy, \quad h \to \infty.$$

Moreover, the asymptotic estimate $I_{1,h} \sim ch^{-\gamma}L_1(h)$, as $h \to \infty$, can be deduced using the same strategy as in (i). Observing that $-\gamma < -2\gamma+1$ when $\gamma < 1$, we deduce that $I_{1,h} = o\big(h^{-2\gamma+1}L_1(h)^2\big)$ as $h \to \infty$, so the assertion follows. □

Proof of Proposition 4.3. As in the proof of Proposition 4.2, we take $h > 1$ and write

$$\int_0^\infty g(x)g(x+h)dx = \int_0^\infty g(x)(x+h)^{-\gamma}e^{-\lambda(x+h)}L_1(x+h)dx$$

$$= e^{-\lambda h}h^{-\gamma}L_1(h)\int_0^\infty g(x)(x/h+1)^{-\gamma}e^{-\lambda x}\frac{L_1(x+h)}{L(h)}dx$$

$$\sim e^{-\lambda h}h^{-\gamma}L_1(h)\int_0^\infty g(x)e^{-\lambda x}dx,$$

where we used the dominated convergence theorem, which can be justified in the same manner as in the proof of Proposition 4.2. □

A.2 Proof of Theorem 4.1

We first recall an elementary result that we will need below. Namely, as $x \downarrow 0$,

$$\sum_{k=2}^\infty \frac{x^k}{k!} = o(x). \tag{A.3}$$

This result is easy to prove by examining the Taylor expansion of the exponential function and applying l'Hôpital's rule.

Proof of Theorem 4.1. (i) Suppose w.l.o.g. that $\xi = 1$. Since σ is covariance-stationary, it suffices to study

$$\rho(h) = Corr(\sigma_t, \sigma_{t+h}) = \frac{Cov(\sigma_t, \sigma_{t+h})}{Var(\sigma_0)}, \quad h \geq 0.$$

We get using the definition (4.1) of σ

$$Cov(\sigma_t, \sigma_{t+h}) = \mathbb{E}[\sigma_t\sigma_{t+h}] - \mathbb{E}[\sigma_0]^2$$

$$= \mathbb{E}[\exp(X_t + X_{t+h})] - \mathbb{E}[\exp(X_0)]^2.$$

Now, by the fact that X is a zero-mean Gaussian process we get, using the moment generating function of the Gaussian distribution,

$$\mathbb{E}[\exp(X_t + X_{t+h})] = \exp\left(\frac{1}{2}Var(X_t + X_{t+h})\right)$$

$$= \exp\left(Var(X_0) + cov(X_0, X_{t+h})\right)$$

$$= \exp\left(\gamma_X(0) + \gamma_X(h)\right),$$

and

$$\mathbb{E}[\exp(X_0)] = \exp\left(\frac{1}{2}\gamma_X(0)\right),$$

where we write $\gamma_X(h)$ for $Cov(X_h, X_0) = \mathbb{E}[X_h X_0]$. 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,

$$
\begin{aligned}
1 - \rho(h) &= \frac{\exp(\gamma_X(0)) - 1 - (\exp(\gamma_X(h)) - 1)}{\exp(\gamma(0)) - 1} \\
&= \frac{\exp(\gamma_X(0)) - \exp(\gamma_X(h))}{\exp(\gamma(0)) - 1} \\
&= C\left[1 - \exp(\gamma_X(h) - \gamma_X(0))\right] \\
&= C\left[1 - \exp(-\gamma_X(0)(1 - \rho_X(h)))\right] \\
&= C_1(1 - \rho_X(h)) + C_2 \sum_{k=2}^{\infty} \frac{(-\gamma_X(0)(1 - \rho_X(h)))^k}{k!} \\
&\sim C|h|^{2\alpha+1} L_0(h),
\end{aligned}
$$

by the assumption on $1 - \rho_X(h)$ and the asymptotic result (A.3). (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:

$$
\begin{aligned}
\rho(h) &= \frac{\exp(\gamma_X(h)) - 1}{\exp(\gamma(0)) - 1} \\
&= C_1 \gamma_X(h) + C_2 \sum_{k=2}^{\infty} \frac{\gamma_X(h)^k}{k!} \\
&= C_1 \gamma_X(0) \rho_X(h) + C_2 \sum_{k=2}^{\infty} \frac{\gamma_X(h)^k}{k!} \\
&\sim C\rho_X(h), \quad h \to \infty,
\end{aligned}
$$

where we have finally applied (A.3). $\qquad\square$

Research Papers
2017

CREATES
Center for Research in Econometric
Analysis of Time Series

2017-09:	Thomas Quistgaard Pedersen and Erik Christian Montes Schütte: Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations
2017-10:	Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors' Sentiment and Return Predictability
2017-11:	Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models
2017-12:	Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term
2017-13:	Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers: Picking Funds with Confidence
2017-14:	Martin M. Andreasen and Anders Kronborg: The Extended Perturbation Method: New Insights on the New Keynesian Model
2017-15:	Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: A Non-Structural Investigation of VIX Risk Neutral Density
2017-16:	Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Does the ARFIMA really shift?
2017-17:	Massimo Franchi and Søren Johansen: Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles
2017-18:	Matias D. Cattaneo, Michael Jansson and Kenichi Nagasawa: Bootstrap-Based Inference for Cube Root Consistent Estimators
2017-19:	Daniel Borup and Martin Thyrsgaard: Statistical tests for equal predictive ability across multiple forecasting methods
2017-20:	Tommaso Proietti and Alessandro Giovannelli: A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices
2017-21:	Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Variance swap payoffs, risk premia and extreme market conditions
2017-22:	Jakob Guldbæk Mikkelsen: Testing for time-varying loadings in dynamic factor models
2017-23:	Roman Frydman, Søren Johansen, Anders Rahbek and Morten Nyboe Tabor: The Qualitative Expectations Hypothesis: Model Ambiguity, Concistent Representations of Market Forecasts, and Sentiment
2017-24:	Giorgio Mirone: Inference from the futures: ranking the noise cancelling accuracy of realized measures
2017-25:	Massimiliano Caporin, Gisle J. Natvik, Francesco Ravazzolo and Paolo Santucci de Magistris: The Bank-Sovereign Nexus: Evidence from a non-Bailout Episode
2017-26:	Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Decoupling the short- and long-term behavior of stochastic volatility