Volatility modelling: decoupling the short- and long-term behavior of stochastic volatility

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

January 10, 2016

Abstract

We study estimates of return volatility sampled at different time scales ranging from one minute to one day. Our main finding is that as sampling frequency increases, volatility becomes rougher and displays stronger memory. Traditional continuous-time volatility models invoked to capture these features are based on the fractional Brownian motion, and thus, necessarily, display *either* roughness *or* long memory. These models are therefore inadequate to explain the data. We propose new continuous-time models of log-volatility capable of decoupling roughness properties (local behavior) from memory properties (global behavior) in a simple and parsimonious way, which allows us to adequately model volatility at all intraday time scales. Our main candidate model will be based on the Brownian semistationary process, and we derive some theoretical properties of this model related to our setting. We compare these new models with existing alternatives in a forecasting exercise, and find that for the higher intraday sampling frequencies our models outperform the competition.

Keywords: Stochastic volatility; roughness; long memory; forecasting; Brownian semistationary process.

1 Introduction

Intraday modelling of volatility is of importance in several applications, including derivatives pricing, algorithmic trading, and risk management (Andersen et al., 2000; Rossi and Fantazzini, 2015). The frequency at which one samples the volatility process might vary depending on the goal at hand: for instance, a high-frequency (algorithmic) trader might be concerned with stock returns over very short periods, such as a minute or even less, and would therefore like to assess volatility over the same time scale. On the other hand, for different purposes it might make sense to sample

[*]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. E-mail: m.pakkanen@imperial.ac.uk

the process at a lower frequency; for instance, Andersen et al. (1999) show that if the goal is to forecast *daily* volatility then sampling at moderate frequencies, such as one hour, outperforms the alternative choices of either very high frequencies (e.g. one minute) or very low frequencies (e.g. one day).

The present work has two main contributions. The first is an in-depth examination of the time series behavior of (log) volatility at a wide range of scales: from a sampling frequency less than a minute all the way to daily sampling. Three main conclusions emerge from this exercise: volatility is *rough*, it has significant *memory* and it is *non-Gaussian*.[1] What is more, these facts become more pronounced as sample frequency increases; indeed, when sampling at very high frequencies volatility is extremely rough and has very strong long memory. The second contribution is to suggest mathematical models which can accomodate these findings. In particular, we propose to use either the *Cauchy class* (Gneiting and Schlather, 2004) or the *Brownian semistationary* ($\mathcal{BSS}$) *process* (Barndorff-Nielsen and Schmiegel, 2007, 2009). These processes are flexible in the sense that they allow decoupling of the short-scale behavior (roughness) from the global behavior (memory). Both candidate processes are stationary, but only the $\mathcal{BSS}$ process accomodates easy inclusion of non-Gaussianity, of leverage effects, and of fast and efficient simulation. Consequently, the $\mathcal{BSS}$ process will be our main candidate for a model of log-volatility, and we derive some general theoretical results related to the memory properties of this process.

Recently, there has been considerable interest in rough models of volatility. This is due to both theoretical developments (Fukasawa, 2015) as well as empirical evidence (Gatheral et al., 2014, and Section 2.3, this paper). A prominent contribution to this literature is the seminal Gatheral et al. (2014) which modifies the fractional stochastic volatility (FSV) model of Comte and Renault (1996). Both of these models rely on the fractional Brownian motion (fBm) as the driving stochastic process, but while Comte and Renault (1996) model volatility using an fBm with Hurst index $H > 1/2$ to arrive at a model with the long memory property, Gatheral et al. (2014), by contrast, model with an fBm with $H < 1/2$ to allow for roughness. They thus term their model the rough fractional stochastic volatility model (RFSV). In Bayer et al. (2015) the RFSV model is extended to the rough Bergomi (rBergomi) model to allow for pricing of financial contracts in a rough volatility setup. The downside of the RFSV and rBergomi models is that they do not allow for the long memory property, which is a well-established stylized fact of volatility (e.g. Andersen et al., 2003). As we shall see below this is actually a deficiency of any model based on a self-similar process such as the fBm. In fact, such models have *either* long memory *or* rough paths (Gneiting and Schlather, 2004). To overcome the dichotomy long memory/roughness, Corlay et al. (2014) model volatility using a multifractional Brownian motion (mBm), which allows the Hurst parameter to change through time. The effect of this is decoupling of the local regularity properties from the long-range dependence properties – i.e. exactly what we aim to do in this paper – but

[1] 'Roughness' refers, more or less, to the Hölder regularity of the paths of volatility. Section 2.3 will give the exact mathematical definition of what we mean by 'roughness'.

the price to pay is intractability and complicated estimation procedures relying on quantization-based curbature methods. In contrast, the models presented in this work are parsimonious, easy to estimate, easy to simulate, and easy to forecast.

The rest of the paper is structured as follows. In Section 2 we present our data and perform empirical investigations which show that volatility is rough, has strong memory, and is non-Gaussian across all relevant intradaily scales. Section 3 go on to present mathematical models of stochastic processes which are able to capture these findings. This section also discusses simulation of the resulting stochastic volatility model. Section 4 presents estimation results of the various models we consider, using both parametric and semiparametric estimation procedures. Section 5 presents a forecasting exercise where we compare our new models to existing alternatives. Finally, section 6 concludes. Proofs of the technical results are relegated to the Appendix.

2 Empirical behavior of log-volatility

We consider a simple model for high-frequency asset returns,

$$\frac{dS_t}{S_t} = \sigma_t dB_t, \quad t \geq 0, \tag{2.1}$$

where $B = (B_t)_{t\geq 0}$ is a Brownian motion and $\sigma = (\sigma_t)_{t\geq 0}$ is a volatility process. The following sections will explain the data we use for S, how we estimate the latent σ process, and the subsequent empirical findings for this process.

2.1 Data description and extraction of latent volatility

We consider E-mini futures data 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. We consider only the time of day when the NYSE is open: from 9.30 a.m. until 4 p.m. Eastern Standard Time (EST). Below we will calculate estimates of integrated variance (IV) based on the realized kernel (RK) over intervals as short as one minute. In some of these intervals the RK will be zero due to no trading taking place in that particular interval; this will cause problems as we will work on the logarithm of the RK. When such a zero occurs we will replace it by the lowest non-zero value of the RK obtained on the same day. The effect of this data cleaning should be minimal, as there are not many zeros observed when the NYSE is open. For instance, when the sampling frequency is one minute this results in $498 \times 6.5 \times 60 = 194,220$ observations of which only around 1900 – or less that 1% – were zeros. When sampling at a 2 minute frequency we have 210 zeros out of $97,110$ observations, or roughly 0.2% of the observations.

We seek to approximate the latent volatility process σ_t; we can not observe this process directly, but instead have access to high frequency prices observed every second. For a time step $\Delta > 0$ we

therefore seek to estimate the *integrated variance* (IV),

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

Estimators of IV are much studied in the literature, prominent examples being realized variance (Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2002), realized kernels (Barndorff-Nielsen et al., 2008), two-scale estimators (Zhang et al., 2005), and pre-averaging methods (Jacod et al., 2009). Except for the former, these methods are robust to market microstructure effects, which is crucial when sampling at higher frequencies. By choosing Δ sufficiently small and assuming that volatility does not vary too much in each time interval of size Δ we approximate

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

where $\widehat{IV}_t^{\Delta}$ is an estimate of IV coming from one of the methods mentioned. In this paper we will restrict attention to realized kernel measures, i.e. $\widehat{IV}_t^{\Delta} = RK_t^{\Delta}$, where RK_t^{Δ} is the realized kernel estimator of IV over the time interval $[(t-1)\Delta, t\Delta]$. For implementation of the RK we refer to Barndorff-Nielsen et al. (2008) and Barndorff-Nielsen et al. (2009).

We will consider $T = 1$ to be one day and will take $\Delta \in (0, 1]$, i.e. we estimate *intraday* volatility; this should be contrasted with e.g. Gatheral et al. (2014) who consider data sampled at daily frequency. 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 low around midday.

It is important to account for this seasonality before performing any further analyses as subsequent estimates could be severly biased if one does not take this into account (e.g. Rossi and Fantazzini, 2015). We assume a multiplicative seasonality:

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

where σ^s is the seasonal component and $\tilde{\sigma}$ is the stochastic process we are interested in. To estimate σ^s we use the Flexible Fourier Form (FFF) approach of Andersen and Bollerslev (1997, 1998). We then estimate $\tilde{\sigma}_t^2$ by $\widehat{IV}_t / \widehat{(\sigma_t^s)^2}$ and will from now on be working on this data. Further, we will abuse notation and continue to write σ even though we are now actually considering the de-seasonalized process $\tilde{\sigma}$. Table 1 contains some simple descriptive statistics of this process and how it behaves as Δ increases; overall, skewness lessens and kurtosis becomes closer to 3, indicating that the data looks more Gaussian as we sample less frequently. We will see further evidence of this in Section 2.5 below.

2.2 Stationarity of volatility

Volatility is widely believed to be stationary. In particular, one does not expect the volatility to wander without bound but instead to revert to some 'reasonable' level. In Table 1 we perform unit

Table 1: *Desriptive statistics and unit root tests*

Δ	n	Mean	Skew	Kurt	ADF	PP	$\hat{\pi}$	$n(\hat{\pi}-1)$
1 minute	194220	−0.093	−1.902	11.125	0.000	0.001	0.989	−2129.3
2 minutes	97110	−0.036	−1.256	11.322	0.000	0.001	0.990	−932.9
5 minutes	38844	0.008	0.130	5.432	0.000	0.001	0.989	−443.2
15 minutes	12948	0.033	0.524	4.264	0.000	0.001	0.976	−313.0
30 minutes	6474	0.047	0.547	4.066	0.000	0.001	0.969	−203.6
1 hour	3486	0.064	0.484	3.713	0.000	0.001	0.942	−201.3
2 hours	1992	0.059	0.451	3.521	0.000	0.001	0.937	−125.4
1 day	498	0.037	0.307	3.296	0.000	0.001	0.837	−81.2

Descriptive statistics and unit root tests of log-volatility of the E-mini data set. ADF and PP refers to the P-values of the Augmented Dickey-Fuller test with automatic lag selection and the Phillips-Perron test, respectively. π is 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 are −20.7 and −14.1, respectively.

root tests for the different sampling frequencies, 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 in volatility

Recent studies have provided evidence that volatility is *rough*, see e.g. Gatheral et al. (2014) and Bayer et al. (2015). What we mean by this is that the correlation function of log-volatility, ρ, adheres to the asymptotic relationship

$$1-\rho(h) := 1-Corr(\log\sigma_t, \log\sigma_{t+h}) \sim |h|^{2\alpha+1}, \quad |h|\to 0, \tag{2.2}$$

for some $\alpha \in \left(-\frac{1}{2}, 0\right)$. We call α the *fractal index* of the log-volatility process; usually one allows α to take values in $\left(-\frac{1}{2}, \frac{1}{2}\right)$ but, as we shall see, only negative values of α will be relevant for us. This implies that volatility is *rough*, meaning informally that it has a low degree of Hölder continuity.[2]

Rough models of volatility are consistent with some empirical aspects of the implied volatility surface (Gatheral, 2006); in particular, as shown in Fukasawa (2015), such models can accurately capture the short-time at-the-money volatility skew, which traditional local/stochastc volatility models based on Itō diffusions fail to do, see also Bayer et al. (2015). To model this roughness earlier studies have mainly relied on the canonical fractional process: the fractional Brownian motion (fBm) with Hurst index $H \in (0,1)$. For the fBm the simple relationship $H = \alpha + 1/2$ holds which means that $H < 1/2$ implies roughness.

The relationship (2.2) suggests a straight forward semiparametric estimation procedure of α. Namely, consider the regression

$$\log\left(1-\hat{\rho}(h)\right) = c + a\log|h| + \epsilon_h, \quad h = \Delta, 2\Delta, \ldots, L\Delta, \tag{2.3}$$

[2]See e.g. Proposition 2.1. of Bennedsen (2015b) for a statement on the precise connection between the fractal index and the degree of Hölder continuity.

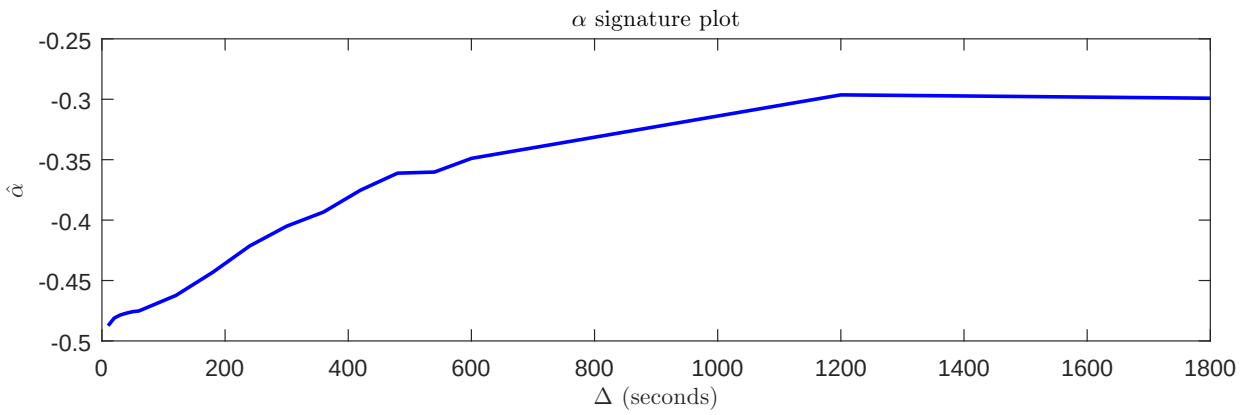

Figure 1: *$\hat{\alpha}$ as a function of Δ. Estimation was done using the OLS regression (2.3) with $L = \max\{3, \lceil n^{1/3} \rceil\}$ lags.*

for some small step size $\Delta > 0$ and a number of lags $L \in \mathbb{N}$ such that $L\Delta$ is still small, and where $\hat{\rho}(h)$ is the empirical estimate of the correlation function at lag $h \in \mathbb{R}$. The relationship $a = 2\alpha + 1$ lets us estimate $\hat{\alpha} = \frac{\hat{a}-1}{2}$ from simple OLS regression of (2.3). Confidence intervals and hypothesis testing of α can be bootstrapped using the theory in Bennedsen (2015b), who also recommend using $L = \max\{3, \lceil n^{1/3} \rceil\}$, with n being the number of observations.

Figure 1 shows an "α signature plot" using the above estimator, i.e. the estimated value of α as a function of sampling frequency, Δ. For very low sampling frequencies we find $\hat{\alpha} \approx -1/2$ and then the estimate increases up to $\hat{\alpha} \approx -0.3$ at around $\Delta = 1200$ seconds (20 minutes), from which point it flats out.[3] Indeed, although it is not shown in the figure, we will see in Section 4 (Table 2), that for all sampling frequencies up to $\Delta = 1$ day we get $\hat{\alpha} \approx -0.3$.

To summarize: volatility is very rough, but the roughness becomes less pronounced as we sample less often.

2.4 Long memory in 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 modelling (log) volatility using models with strong memory, i.e. volatility models which has autocorrelations decaying at a slow polynomial rate:

$$\rho(h) := Corr(\log \sigma_t, \log \sigma_{t+h}) \sim |h|^{-\beta}, \quad |h| \to \infty, \tag{2.4}$$

for some $\beta \in (0, \infty)$. When $\beta \in (0, 1)$ the correlation function in (2.4) is not integrable and we say that $\sigma = (\sigma_t)_{t \in \mathbb{R}}$ has the long memory propterty with Hurst index $H = 1 - \beta/2 \in \left(\frac{1}{2}, 1\right)$. A

[3] The upward slope of the α signature plot is to be expected as we are proxying σ_t^2 by an estimate of *integrated* volatility, i.e. by $\Delta^{-1} \int_{t-\Delta}^{t} \sigma_s^2 ds$. This integration is a smoothing operation and we thus expect the time series to be smoother as we integrate over a larger interval, i.e. as Δ grows, see also Appendix C of Gatheral et al. (2014).

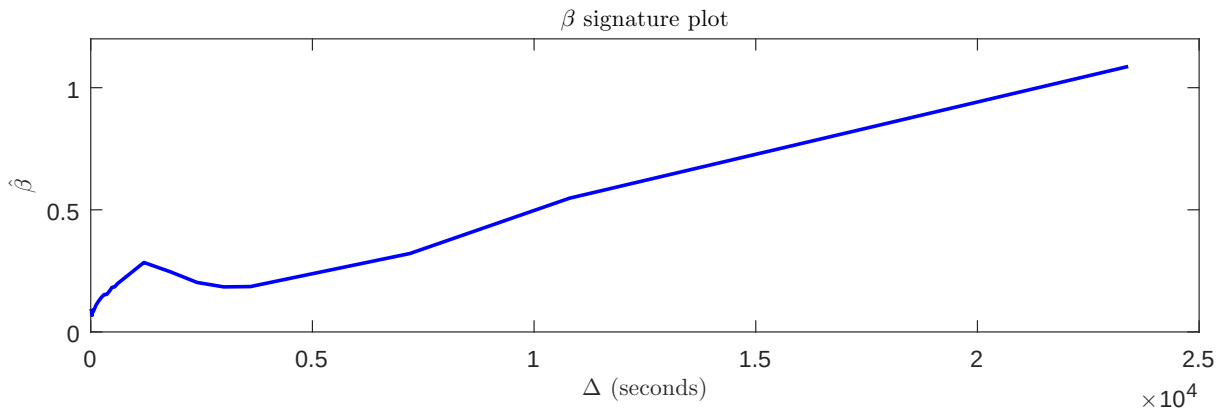

Figure 2: $\hat{\beta}$ *as a function of* Δ. *Estimation was done using the OLS regression* (2.5) *with* $M = \lfloor n^{1/4} \rfloor$ *and* $M' = \lfloor n^{1/3} \rfloor$.

somewhat tractable stochastic process which has precisely this behavior is the fractional Brownian motion and in a series of papers, e.g. Comte and Renault (1996), Comte (1996), Comte and Renault (1998), Comte et al. (2012) (and many more), log-volatility was modelled using an fBm.

As in the previous section, (2.4) suggests estimating $\beta = -b$ semiparametrically using the OLS regression

$$\log \hat{\rho}(h) = c + b \log |h| + \epsilon_h, \quad h = M\Delta, (M+1)\Delta, \dots, M'\Delta, \tag{2.5}$$

for some numbers $M, M' \in \mathbb{N}$ with $M' > M$ such that $M\Delta$ is large, and where $\hat{\rho}(h)$ is the empirical estimate of the correlation function at lag $h \in \mathbb{R}$.

Analogous to the previous section, Figure 2 shows a "β signature plot" using the above estimator, i.e. it shows the estimated value of β as a function of sampling frequency, Δ. In this case we see that for very high frequency sampling we have $\hat{\beta} \approx 0.10$ which is very strong long memory indeed. As sampling frequency decreases the estimated value of β increases, that is, the memory becomes less strong. Recall that $\beta < 1$ corresponds to long memory; the cross-over point between long memory and non-long memory seems to be just before $\Delta = 1$ day (although not technically "long memory", the memory is still rather strong in this case). It should be noted that estimates of the long memory parameter are notoriously imprecise and we also found that the particular estimates were very much dependent on what values one chooses for M and M'. Regardless of these tuning parameters, however, the overall picture of Figure 2 does remain: volatility has significant memory, and the memory is decreasing as Δ increases. In Section 4 (Table 2) we will also estimate the β parameter in a parametric setting, which corroborates this conclusion.

2.5 (Non-)Gaussianity of log-volatility

There is some evidence that increments of log-volatility is Gaussian; this was for instance found in the now classical papers Andersen et al. (2001), Andersen et al. (2001), and Barndorff-Nielsen and

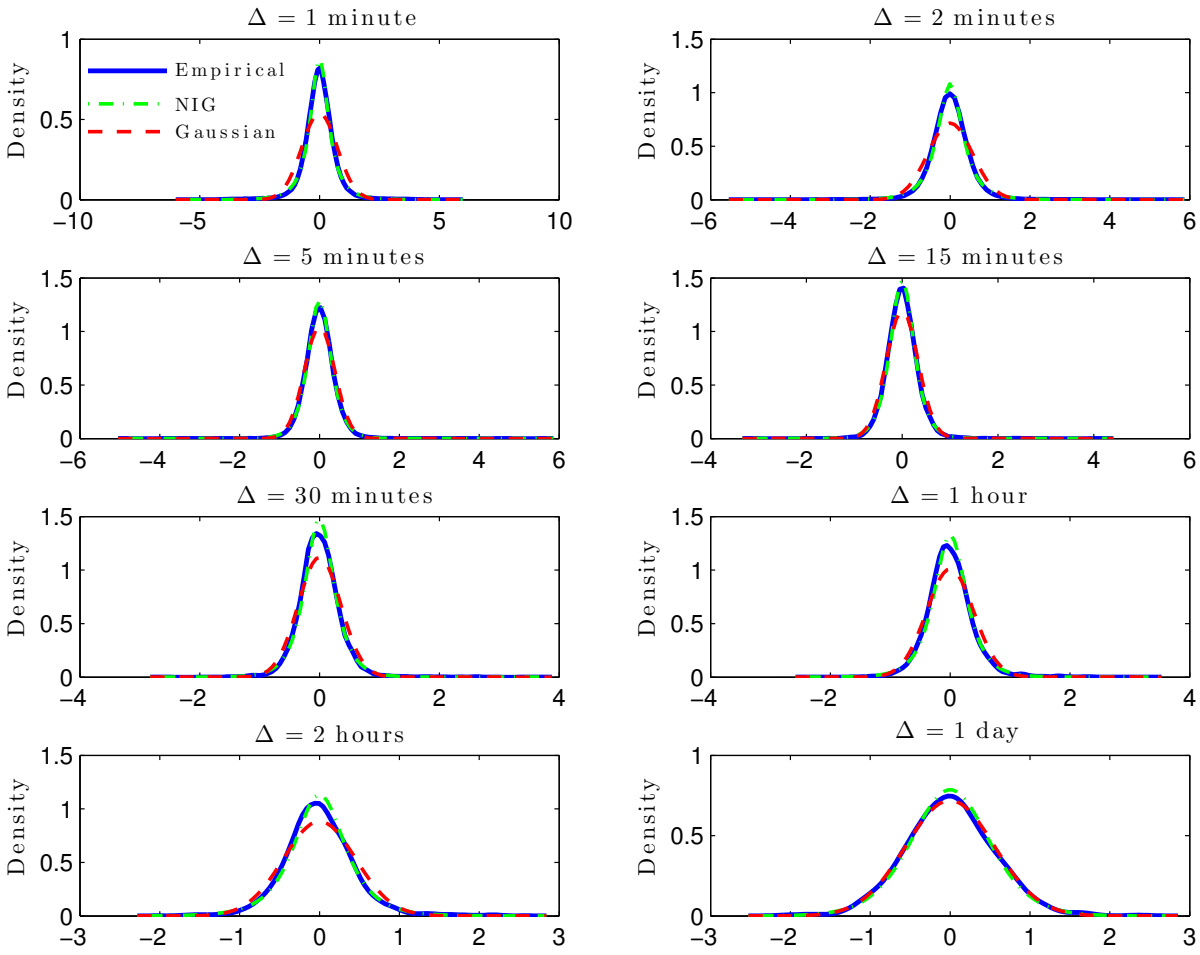

Figure 3: *Empirical distribution of the increments of log volatility, i.e.* $y_t = \log(RK_t) - \log(RK_{t-1})$.

Shephard (2002), see also Gatheral et al. (2014). Specifically, Andersen et al. (2001), Andersen et al. (2001), and Barndorff-Nielsen and Shephard (2002) looked at the empirical distribution of the time series of the increments of daily log-realized variance, i.e of time series similar to the ones studied in this paper, and found that the Gaussian distribution fitted this empirical distribution well. These conclusions were largely drawn from looking at empirical histograms of daily increments of log-realized variance and comparing with the Gaussian distribution. In Figure 3 we perform the same analysis but now looking at several sampling frequencies. We also compare with the fitted Normal Inverse Gaussian distribution (NIG) distribution. We see that when sampling RK at intraday frequencies the resulting distribution is highly non-Gaussian. In contrast, the NIG distribution seems to fit quite well. However, when sampling at a daily frequency the Gaussian distribution provides a great fit, indicating some kind of "aggregation to Gaussianity"; this last observation is similar to the findings in the above cited papers.

2.6 Empirical conclusions

As discussed above the fBm is the standard choice when modelling long memory in a continuous-time framework. Similarly, fBm-based models are also standard choices when modelling roughness. However, the conclusion we can draw from the previous sections is that an fBm-based model is a poor choice when modelling intraday volatility. This is so for two reasons. Firstly, since the fBm is *self-similar* the small-scale behavior (roughness, as controlled by α) is fully determined by the large-scale behavior (memory, as controlled by β) and vice versa (Gneiting and Schlather, 2004). In fact, it is necessarily the case that $H = \alpha + \frac{1}{2}$, where $H = 1 - \beta/2$ is the Hurst index of the fBm, meaning that $\beta = 1 - 2\alpha$ for these processes. Clearly, this restrictive relationship is violated for the data we consider here. Indeed, for an fBm, roughness ($\alpha < 0$) implies short memory ($\beta > 1$). Similarly, long memory ($\beta < 1$) implies a smooth model ($\alpha > 0$).[4]

Secondly, the fBm is a Gaussian process and although the findings of previous studies, discussed in Section 2.5, have justified the use of Gaussian processes as models of log-volatility, we saw in the same section that this assumption is violated for intraday sampling frequencies. Therefore, a model that can accomodate non-Gaussianity is preferable.

In conclusion, traditional models such as the ones in Comte and Renault (1996), Comte (1996), Comte and Renault (1998), Comte et al. (2012), Gatheral et al. (2014) and Bayer et al. (2015) are inadequate due to their failure to decouple the behavior at short and long time scales; additionally the implicit Gaussianity assumption of these models make them inadequate for intraday volatility modelling. The next section will present stationary stochastic processes that are more adequate in the present context.

3 Models of stochastic volatility that decouples short- and long-term behavior

We now present our specification for the stochastic volatility process σ. Let X be a zero-mean process that decouples fractal index from the memory parameter, i.e. let X satisfy

$$\rho_X(h) := Corr(X_{t+h}, X_t) \sim \begin{cases} 1 - |h|^{2\alpha+1} L_0(h), & h \to 0, \\ |h|^{-\beta} L_1(h), & h \to \infty, \end{cases} \tag{3.1}$$

where $\alpha \in (-1/2, 1/2)$, $\beta > 0$, and L_0, L_1 are slowly varying functions at zero and infinity, respectively.[5] Intuitively, and informally, speaking this means that as $|h| \to \infty$, $\rho_X(h)$ decays slower than $|h|^{-\beta-\epsilon}$ but faster than $|h|^{-\beta+\epsilon}$ for all $\epsilon > 0$. A similar conclusion holds for the case $|h| \to 0$.

[4] Gatheral et al. (2014) show how their rough fBm-based model can generate *spurious* long memory. While this is a possible way to account for roughness and slowly decaying autocorrelations, we prefer to model roughness and *bona fide* long memory.

[5] Recall that a function, L, is slowly varying at $c \in \mathbb{R} \cup \{\infty\}$ if $\lim_{x \to c} \frac{L(tx)}{L(x)} = 1$ for all $t > 0$. For many additional details of slowly varying functions see Bingham et al. (1989).

Define now

$$\sigma_t = \xi_t \exp\left(X_t - \frac{1}{2}\mathbb{E}[X_t^2]\right), \quad t \in \mathbb{R}, \tag{3.2}$$

where $\xi = \{\xi_t\}_{t\in\mathbb{R}}$ is the so-called forward variance swap curve (e.g. Bayer et al., 2015). As our focus is on the process X we set $\xi_t = \xi > 0$ for all t; going into detail with the ξ_t process is beyond the scope of this article. Let $c(h)$ be the covariance function of the volatility process,

$$c(h) = Cov(\sigma_{t+h}, \sigma_t), \quad h \geq 0,$$

and let the autocorrelation function be $\rho(h) := c(h)/c(0)$. It has been found that, empirically, both the log-variance as well as the raw variance have similar roughness and memory properties, and we confirmed this in our data (not reported here). Our mathematical model (3.2) accomodates this, as we now show.

The following theorem shows that σ inherets the roughness properties of the process X.

Theorem 3.1. *Let σ be given by (3.2) where X is a Gaussian process satisfying (3.1) with $\alpha \in (-1/2, 1/2)$. Now,*

$$\rho(h) \sim 1 - |h|^{2\alpha+1} L_0(h), \quad |h| \to 0.$$

The following theorem shows that σ also inherits the memory properties of X.

Theorem 3.2. *Let σ be given by (3.2) where X is a Gaussian process satisfying (3.1) with $\beta \in (0, \infty)$. Now,*

$$\rho(h) \sim |h|^{-\beta} L_1(h), \quad |h| \to \infty.$$

3.1 Models for X

Motivated by our empirical findings we seek stationary stochastic processes, X, with arbitrary fractal index $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and memory parameter $\beta \in (0, \infty)$. Many possibilities exist, but we focus here on two: the Cauchy process (also called the Cauchy class) and the Brownian semistationary process.

3.1.1 The Cauchy process

A flexible process that decouples small-scale behavior and global behavior is processes of the *Cauchy class* (Gneiting and Schlather, 2004). Such a process is a stationary Gaussian process, $G = (G_t)_{t\in\mathbb{R}}$, with correlation function

$$\rho(h) = \left(1 + |h|^{2\alpha+1}\right)^{-\frac{\beta}{2\alpha+1}}, \quad h \in \mathbb{R}, \tag{3.3}$$

where $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and $\beta > 0$. In particular, this process obeys (2.2) and (2.4). The downside of modelling with such a process is the inherent Gaussianity; indeed, we saw above the intraday

data we consider is *not* Gaussian. A way to circumvent this is through volatility modulation. One could e.g. specify a process

$$X_t = \int_0^t v_s dG_s, \quad t \geq 0, \tag{3.4}$$

where G is of the Cauchy class and $v = (v_t)_{t\in\mathbb{R}}$ is a stochastic process, the volatility of volatility. See e.g. Barndorff-Nielsen et al. (2009) Example 3 for how to formalize such an approach. The downside of a model such as (3.4) is that it is non-stationary and thus not ideally suited for modelling of log-volatility as discussed above.

3.1.2 The Brownian semistationary process

A stochastic process which is able to capture all desiderata of roughness, long memory, 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 as

$$X_t = \int_{-\infty}^t g(t-s)v_s dW_s, \quad t \geq 0, \tag{3.5}$$

where W is a standard Brownian motion defined on $\mathbb{R}$, g is a square integrable kernel function and $v = \{v_t\}_{t\in\mathbb{R}}$ is a (stationary) volatility process. Note that when v is deterministic X is Gaussian, while a stochastic v will induce non-Gaussianity of X. To ensure existence of the process (3.5) we assume g and v to be given such that for all $t \geq 0$,

$$\int_0^\infty g^2(x)v_{t-x}^2 dx < \infty, \quad \text{a.s.}$$

The kernel function g is supposed to fulfil the following assumptions,

$$g(x) = x^\alpha L_0(x), \quad x \in (0,1], \quad \alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right), \tag{3.6}$$

$$g(x) = x^{-\beta} L_1(x), \quad x \in (1,\infty), \quad \beta \in \left(\frac{1}{2}, \infty\right), \tag{3.7}$$

where L_0, L_1 are functions, which are slowly varying at zero and infinity, respectively. The restriction on β, i.e. $\beta > 1/2$ as opposed to $\beta > 0$, is to ensure square integrability of g. Indeed, a simple application of Karamata's theorem (Bingham et al., 1989, Proposition 1.5.10.) shows that $\alpha > -1/2$ and $\beta > 1/2$ are necessary and sufficient conditions for g to be square integrable under the specifications (3.6)-(3.7). The integration from $-\infty$ in (3.5) is to ensure stationarity of X (which holds when v is stationary as well). Proposition 2.1 in Bennedsen et al. (2015) shows that a $\mathcal{BSS}$ process satisfying the above assumptions will have α as its fractal index in the sense of Equation (2.2). The following propositions show how β controls the long-term memory properties of X.

Proposition 3.1. *Let X be a $\mathcal{BSS}$ process satisfying (3.6)-(3.7) with $\beta \in (1, \infty)$ and let ρ_X be the autocorrelation function of X. Now,*

$$\rho_X(h) \sim L_1(h)|h|^{-\beta}, \quad |h| \to \infty.$$

Even though Proposition 3.1 does unfortunately not provide the exact asymptotic behavior when $\beta \in (1/2, 1)$, it can be shown that for these values the $\mathcal{BSS}$ process does indeed have the long memory property, as we would expect.

Proposition 3.2. *Let X be a $\mathcal{BSS}$ process satisfying (3.6)-(3.7) with $\beta \in (1/2, 1)$ and let ρ_X be the autocorrelation function of X. Now,*

$$\int_0^\infty \rho_X(h)dh = \infty,$$

i.e. X has the long memory property.

An example of a kernel function that satisfies equations (3.6) and (3.7) and which will be important for us later is the *power law kernel*,

Example 3.1 (Power law kernel)**.** Let g be given by the *power law kernel*:

$$g(x) = x^\alpha (1+x)^{-\beta-\alpha}, \qquad \alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right), \quad \beta \in \left(\frac{1}{2}, \infty\right). \tag{3.8}$$

The power law kernel satisfies the assumptions above. In particular, it has roughness index α and memory properties controlled by β, as expounded in Propositions 3.1 and 3.2.

Later we will need the correlation structure of the $\mathcal{BSS}$ process with power law kernel (3.8). By stationarity of v, the autocovariance function of the general $\mathcal{BSS}$ process (3.5) is

$$c_X(h) := Cov(X_t, X_{t+h}) = \mathbb{E}[v_0^2] \int_0^\infty g(x)g(x+h)dx, \quad h \in \mathbb{R}. \tag{3.9}$$

From this we deduce that when g is given as in (3.8) we have

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

where $\mathcal{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \int_0^\infty t^{x-1}(1+t)^{-x-y}dt$ is the Beta function (e.g. Gradshteyn and Ryzhik, 2007, formula 8.380.3.). To calculate the correlation function $\rho_X(h) = c_X(h)/c_X(0)$ we resort to numerical integration of (3.9). Note that ρ_X does not depend on $\mathbb{E}[v_0^2]$.

3.2 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 design simulation experiments to gauge the properties of the model, or one might wish to price financial claims by Monte Carlo simulation. We here explain how our model can be simulated rather easily. It should be noted that easy simulation methods for models such as the ones considered in this paper are not a given; indeed, rough models generally rely on the entire history of the process and therefore ordinary recursive simulation methods are inadequate. What is more, the possibility of non-Gaussianity of the $\mathcal{BSS}$ process poses further problems, as this excludes simulation methods based on Gaussianity such as Cholesky factorization and circulant embedding methods (e.g. Asmussen and Glynn, 2007).

According to our underlying assumptions, cf. equations (2.1) and (3.2), the model to be simulated is

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

where X is one of our candidate models for log-volatility presented in the sections above. As our models are stationary we have that $\mathbb{E}[X_t^2] = \mathbb{E}[X_0^2]$ for all t. We will simulate S on a grid by Riemann-sum approximation of the integrals. Consequently we need to simulate B and σ on this grid as well, which boils down to simulating B and X. As these processes might be correlated (due to the leverage effect), it is necessary to simulate B and X jointly.

When X is Gaussian, for instance a Cauchy process or a non-volatility modulated $\mathcal{BSS}$ process, it can be simulated exactly using e.g. a Cholesky factorization of its covariance matrix. One can even compute the joint covariance structure of the Gaussian bi-variate process (B, X) and jointly simulate B and X, and in this way account for correlation between the two processes. This was the approach taken in Bayer et al. (2015). However, as the authors also note, this approach is slow and unwieldy: if one is interested in simulating many observations the Cholesky decomposition will be slow or infeasible to implement. Instead, we recommend the use of the *Hybrid Scheme* of Bennedsen et al. (2015) which is a simulation scheme designed for $\mathcal{BSS}$ processes. The advantages of using this approach are that (i) simulation is fast and accurate (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 straight forward.

We refer to Bennedsen et al. (2015) for the implementation of the simulation scheme for the general $\mathcal{BSS}$ process, X, in Equation (3.5). Note, that the authors explain how to include correlation between v and W. However, exactly the same method can be used to implement correlation between W and B or, indeed, between W, B, and v.

Table 2: *Estimating the models*

Δ	$\hat{\alpha}_{OLS}$	95% CI	$\hat{\beta}_{OLS}$	$\hat{\beta}_{BSS}$	95% CI	$\hat{\beta}_{Cauchy}$	95% CI
1 minute	-0.48	$(-0.49, -0.46)$	0.08	0.62	$(0.61, 0.63)$	0.09	$(0.08, 0.09)$
2 minutes	-0.46	$(-0.47, -0.46)$	0.11	0.58	$(0.58, 0.58)$	0.09	$(0.09, 0.09)$
5 minutes	-0.41	$(-0.42, -0.39)$	0.15	0.64	$(0.64, 0.64)$	0.15	$(0.15, 0.16)$
15 minutes	-0.32	$(-0.34, -0.30)$	0.28	0.79	$(0.78, 0.79)$	0.30	$(0.29, 0.32)$
30 minutes	-0.30	$(-0.37, -0.25)$	0.25	0.81	$(0.80, 0.81)$	0.34	$(0.33, 0.36)$
1 hour	-0.31	$(-0.46, -0.20)$	0.19	0.76	$(0.75, 0.77)$	0.34	$(0.32, 0.36)$
2 hours	-0.32	$(-0.45, -0.22)$	0.32	0.73	$(0.71, 0.74)$	0.31	$(0.28, 0.34)$
1 day	-0.29	$(-0.35, -0.25)$	1.09	0.75	$(0.70, 0.80)$	0.34	$(0.27, 0.43)$

Semiparametric estimates of α using the OLS regression (2.3) and of β using the OLS regression (2.5). Also parametric method of moments estimators coming from matching the empirical ACF with the theoretical ACF in our two parametric models, the $\mathcal{BSS}$ model and the Cauchy model. The confidence intervals for $\hat{\alpha}_{OLS}$ are calculated using the bootstrap method of Bennedsen (2015b) with $B = 999$ bootstrap replications.

4 Estimating the models

Estimation of the new models presented here is straight forward. In particular, α can be estimated semiparametrically by the OLS regression (2.3) and similarly β from (2.5). These parameters are the only ones we need for the models we consider, cf. the Cauchy class (3.3) and the $\mathcal{BSS}$ process with the power law kernel, (3.8). This simple estimation procedure should be contrasted with competing approaches such as the rBergomi model of Bayer et al. (2015) and the mBm model of Corlay et al. (2014). Indeed, in Bayer et al. (2015) the authors had to *guess* at the parameter values of the model; in Corlay et al. (2014) the authors employed complicated quantization-based curbature methods.

Unfortunately, accurate estimation of the long memory parameter β can be very difficult in finite samples. Therefore, we also estimate β parametrically, using a method of moments approach, by fitting the theoretical autocorrelation function (ACF) of the model to the empirically observed autocorrelations. One could simultaneously estimate α from this method of moments procedure, but simulation results (not presented here) show that the OLS estimator of α is very precise which cause the best overall performance to be when α is first estimated by OLS and then β is estimated by method of moments. That is, we plug $\hat{\alpha}_{OLS}$ into the corresponding ACF so that it is only a function of β, and then we minimize the sum of squared differences between the theoretical ACF, as a function of β, and the empirically observed ACF. The number of lags used in this estimation was chosen so that the theoretical ACFs fitted the empirical ACF in a, subjectively, satisfying manner, see Figure 4. In Table 2 we estimate α and β for the different models using the different estimation procedures. In Sections 2.3 and 2.4 we have already seen the OLS estimates of α and β, indicating roughness and long memory of log-volatility which is corroborated here.

Using the estimates from the Cauchy and $\mathcal{BSS}$ models we plot the empirical autocorrelation of the data together with the theoretical ones from the models. The result is seen in Figure 4. We see that the models fit the empirical ACFs very well indeed; especially the $\mathcal{BSS}$ process is able to

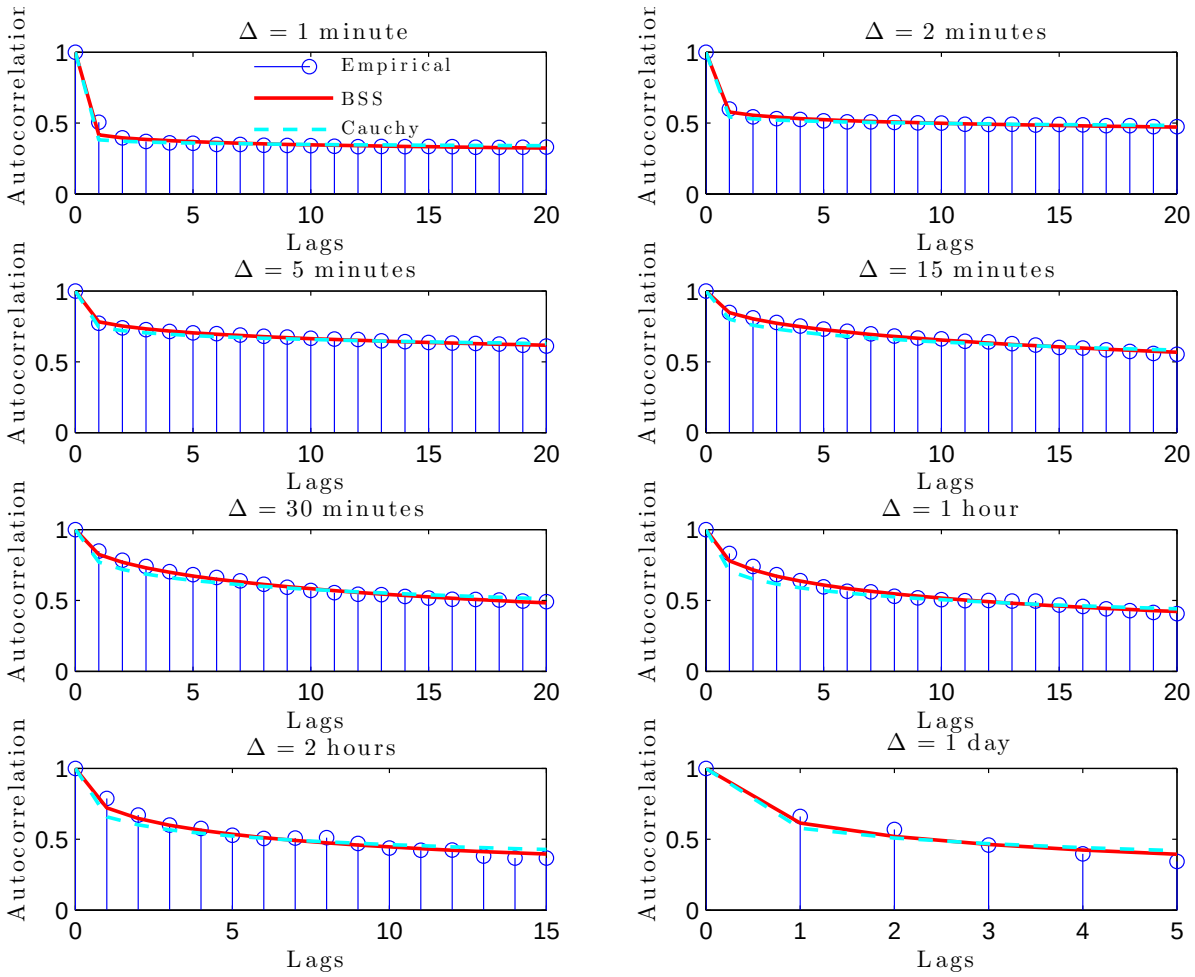

Figure 4: *Empirical autocorrelation from log-volatility (blue bars) and fitted ACFs from the Cauchy (cyan, broken line) and $\mathcal{BSS}$ (red, full line) models. The number of lags used in the estimation of the β parameter is the right-most lag in the plots; e.g. for $\Delta = 1$ minute we used 25 lags, while for $\Delta = 1$ hour we used 20 lags.*

fit the ACF at all time scales. Although this might not be surprising, considering the estimation procedure, it is nonetheless a desirable feature of the Cauchy and $\mathcal{BSS}$ models that they have such tractable and flexible autocorrelation functions. This should be contrasted with e.g. the OU-fBm model which does not have this luxury, but where only (asymptotic) approximations are available.

5 Forecasting volatility

In this section we will conduct a forecasting exercise using the models proposed in this paper and compare with existing ones. In total we consider the models: Random Walk (RW), AR(1), AR(5), AR(10), HAR(3), ARFIMA$(0, d, 0)$, RFSV, Cauchy, and $\mathcal{BSS}$ with the power law kernel (3.8). We also include a $\mathcal{BSS}$ process with the so-called *gamma kernel* (labeled Γ-$\mathcal{BSS}$), i.e. where

$g(x) = x^\alpha e^{-\lambda x}$ for $\lambda > 0$ (see Bennedsen, 2015a, for an application of this model as well as the estimation procedure). The Γ-$\mathcal{BSS}$ process also has roughness properties dictated by α, but has exponentially decaying autocorrelations i.e. very short memory; by comparing this with the $\mathcal{BSS}$ process with power law kernel we can assess the effect of including strong memory in these models. See Corsi (2009) for details of the HAR(3) model and Gatheral et al. (2014) for details of the RFSV model. Our performance metrics will be mean squared error (MSE) as well as mean absolute error (MAE).

In the following we will forecast both log-volatility, or basically the X process, and actual volatility. The former will let us assess how well the various models fit the log-volatility data, as measured by out-of-sample forecast performance. The latter might be more relevant in real-world applications, some of which were discussed in the introduction.

What we actually forecast in the following is $\log(RK_t)$ and RK_t, respectively, i.e. our *estimates* of integrated volatility.[6] Since we use the raw estimates coming from the realized kernel, it is important to remember to correct for the estimated seasonality before comparing our forecasts of X_t with $\log(RK_t)$ or $\exp(X_t)$ with RK_t.

We estimate the AR models as well as the HAR model by OLS. The estimate of H in the RFSV model is set to $\hat{H} = \hat{\alpha}_{OLS} + 0.5$, and for the Cauchy and $\mathcal{BSS}$ processes we use the parametric estimates of β along with $\hat{\alpha}_{OLS}$. Forecasting the AR and HAR models is straight forward. To forecast the RFSV model we use the Riemann-sum version of the following approximation (cf. Gatheral et al., 2014, Equation (5.1))

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

where $\mathcal{F}_t$ is the filtration generated by the driving fBm.

To predict the Cauchy and $\mathcal{BSS}$ processes we rely on the *best linear predictor*, which is a linear combination of past observations of the process, where the weights of the previous values are derived from the ACF of the process. The best linear predictor of $X_{t+h\Delta}$ given $X_0, X_\Delta, \ldots, X_t$ is (e.g. Grimmett and Stirzaker, 2001, Section 9.2.)

$$\hat{X}_{t+h\Delta} = \sum_{i=0}^{t} a_i X_{t-i\Delta}, \quad h \geq 1,$$

where $a_0, a_1, \ldots, a_t \in \mathbb{R}$ solves

$$\sum_{i=0}^{t} a_i \rho_X(|i-j|\Delta) = \rho(h\Delta + j\Delta), \quad 0 \leq j \leq t.$$

Note, that this result relies on X being zero-mean, so in our forecasting exercise we de-mean the data before running the exercise.

[6] Recall that our estimate of integrated volatility only differs from our estimate of spot volatility by a factor of Δ^{-1}, so there is no loss of generality in forecasting integrated variance as opposed to spot variance.

5.1 Predicting log-volatility [To be completed]

Tables 3 and 4 shows the forecasting results using the MSE metric. The corresponding results for the MAE metric are in Tables 5 and 6. A clear picture emerges: for shorter time horizons ($\Delta \leq 30$ minutes) the $\mathcal{BSS}$ model outperforms the competition and is best in practically all situations. For $\Delta = 1$ hour the $\mathcal{BSS}$ process is still best for shorter forecast horizons, i.e. $h \leq 5$. For $\Delta \geq 2$ hours the HAR(3) model is generally best; this is not surprising as this model has proven very capable indeed (e.g. Corsi, 2009). The same conclusion holds for the MAE metric.

5.2 Predicting volatility [To be completed]

Tables 7 and 8 shows the forecasting results using the MSE metric. The corresponding results for the MAE metric are in Tables 9 and 10. We see roughly the same picture as in the previous section on forecasting log-volatility; namely, for MSE the $\mathcal{BSS}$ process is mainly best when Δ is small[7], while the HAR(3), and now also the ARFIMA model, is a bit better for large sample horizons, Δ.

For the MAE metric things change a little bit; in particular the $\mathcal{BSS}$ process seems to fare even better as compared to the competition.

6 Conclusion

In this paper we looked at the empirical characteristics of log-volatility and how they related to the intraday sampling frequency. We found evidence of roughness *and* strong memory. This calls for new models, as traditional ones based on the fBm are inadequate; we proposed two candidates, the Cauchy process and the $\mathcal{BSS}$ process. Especially the latter is very promising as it allows for stationarity, non-Gaussianity, leverage, roughness, and long memory while being easy to both estimate and simulate. Further, we saw that this model outperformed the competition in terms of out-of-sample forecasting, at least when roughness and memory were both very strong (i.e. for high sampling frequencies). This paper considered mainly the $\mathcal{BSS}$ process with a power law kernel, but other kernels could have been considered. We believe that there is potential for even better results from using other kernels than the simple power law (3.8); research into this is currently ongoing.

References

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance 4*, 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*, 219–265.

[7]Here the $\mathcal{BSS}$ model is not always best when $\Delta = 5, 15$ minutes, although the difference between the best model and the $\mathcal{BSS}$ model is at the 4'th decimal place (i.e. not visible in the table).

Table 3: *Predicting* $\log(RK)$*, MSE*

$\Delta = 1$ minute

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.548	0.412	0.413	0.411	0.373	**0.359**	0.388	0.491	0.368	0.361
$h = 2$	0.669	0.478	0.419	0.404	0.388	0.388	0.398	0.518	0.381	**0.380**
$h = 5$	0.710	0.541	0.430	0.411	0.397	0.392	0.398	0.542	0.389	**0.388**
$h = 10$	0.730	0.553	0.438	0.418	0.404	0.398	0.400	0.551	0.395	**0.394**
$h = 20$	0.742	0.553	0.446	0.427	0.413	0.407	0.404	0.553	0.404	**0.402**

$\Delta = 2$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.310	0.248	0.218	0.215	0.199	0.198	0.232	0.252	0.198	**0.194**
$h = 2$	0.353	0.286	0.224	0.216	0.207	0.208	0.227	0.277	0.206	**0.204**
$h = 5$	0.374	0.358	0.233	0.223	0.215	0.216	0.223	0.317	0.215	**0.213**
$h = 10$	0.387	0.384	0.242	0.230	0.223	0.226	0.227	0.352	0.223	**0.221**
$h = 20$	0.406	0.387	0.255	0.242	0.236	0.240	0.236	0.377	0.236	**0.233**

$\Delta = 5$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.145	0.128	0.112	0.125	0.102	0.101	0.161	0.107	0.099	**0.098**
$h = 2$	0.166	0.151	0.116	0.125	0.110	0.109	0.147	0.123	0.109	**0.108**
$h = 5$	0.188	0.219	0.130	0.135	0.124	0.124	0.143	0.154	0.125	**0.122**
$h = 10$	0.212	0.288	0.148	0.148	0.141	0.142	0.151	0.195	0.142	**0.140**
$h = 20$	0.247	0.316	0.173	0.171	0.165	0.167	0.169	0.247	0.166	**0.163**

$\Delta = 15$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.112	0.103	0.132	0.174	0.102	0.089	0.207	0.090	0.089	**0.088**
$h = 2$	0.139	0.129	0.139	0.175	0.119	0.110	0.187	0.114	0.111	**0.108**
$h = 5$	0.198	0.204	0.166	0.181	0.155	0.152	0.186	0.162	0.151	**0.148**
$h = 10$	0.249	0.290	0.196	0.213	0.189	0.190	0.202	0.213	0.187	**0.184**
$h = 20$	0.330	0.357	0.247	0.246	0.237	0.238	0.235	0.280	0.233	**0.231**

Mean squared errors for the predicting $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 4: *Predicting* $\log(RK)$*, MSE*

$\Delta = 30$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.127	0.117	0.188	0.273	0.135	0.109	0.253	0.111	0.112	**0.108**
$h = 2$	0.181	0.164	0.191	0.263	0.162	0.149	0.237	0.153	0.150	**0.146**
$h = 5$	0.267	0.251	0.232	0.263	0.211	0.208	0.243	0.224	0.206	**0.203**
$h = 10$	0.362	0.348	0.273	0.274	0.257	0.263	0.270	0.296	0.259	**0.257**
$h = 20$	0.431	0.412	0.303	0.300	**0.294**	0.311	0.300	0.362	0.302	0.299

$\Delta = 1$ hour

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.157	0.144	0.286	0.351	0.186	0.142	0.295	0.148	0.151	**0.141**
$h = 2$	0.245	0.215	0.283	0.319	0.225	0.202	0.286	0.212	0.207	**0.200**
$h = 5$	0.382	0.325	0.302	0.294	0.284	0.284	0.307	0.310	0.283	**0.280**
$h = 10$	0.469	0.413	0.325	0.339	**0.321**	0.337	0.335	0.382	0.334	0.330
$h = 20$	0.565	0.469	0.393	0.386	**0.383**	0.413	0.389	0.448	0.403	0.396

$\Delta = 2$ hours

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.209	0.187	0.335	0.407	0.230	**0.181**	0.321	0.198	0.195	0.184
$h = 2$	0.327	0.275	0.316	0.370	0.272	0.253	0.321	0.278	0.258	**0.253**
$h = 5$	0.464	0.383	0.334	0.380	**0.323**	0.331	0.347	0.380	0.331	0.328
$h = 10$	0.560	0.467	0.398	0.414	**0.388**	0.399	0.393	0.456	0.398	0.393
$h = 20$	0.663	0.501	0.443	0.473	**0.442**	0.477	0.455	0.496	0.468	0.461

$\Delta = 1$ day

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.322	0.269	0.395	0.534	0.294	**0.256**	0.366	0.285	0.269	0.259
$h = 2$	0.400	0.331	0.418	0.517	0.339	**0.319**	0.373	0.367	0.332	0.323
$h = 5$	0.567	0.447	0.447	0.492	**0.413**	0.422	0.419	0.463	0.426	0.420
$h = 10$	0.781	0.501	0.491	0.520	**0.481**	0.535	0.506	0.503	0.522	0.514
$h = 20$	1.040	0.539	0.543	0.550	**0.536**	0.635	0.610	0.539	0.604	0.592

Mean squared errors for the predicting $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 5: *Predicting* $\log(RK)$*, MAE*

$\Delta = 1$ minute

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.490	0.430	0.419	0.416	0.395	0.392	0.405	0.469	0.392	**0.389**
$h = 2$	0.530	0.460	0.420	0.410	0.401	0.401	0.406	0.483	0.398	**0.397**
$h = 5$	0.547	0.497	0.426	0.414	0.406	0.402	0.405	0.498	0.403	**0.402**
$h = 10$	0.555	0.504	0.431	0.419	0.411	**0.405**	0.406	0.503	0.407	0.405
$h = 20$	0.563	0.504	0.436	0.425	0.417	0.411	**0.409**	0.504	0.413	0.410

$\Delta = 2$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.375	0.341	0.311	0.309	0.295	0.296	0.325	0.343	0.295	**0.292**
$h = 2$	0.396	0.370	0.314	0.309	0.301	0.301	0.317	0.361	0.300	**0.299**
$h = 5$	0.408	0.424	0.322	0.314	0.308	0.307	0.312	0.393	0.308	**0.305**
$h = 10$	0.419	0.442	0.330	0.321	0.315	0.315	0.316	0.419	0.316	**0.313**
$h = 20$	0.431	0.444	0.341	0.331	0.326	0.328	0.324	0.437	0.328	**0.324**

$\Delta = 5$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.280	0.263	0.243	0.255	0.230	0.230	0.296	0.236	0.227	**0.226**
$h = 2$	0.297	0.287	0.247	0.256	0.238	0.239	0.280	0.253	0.238	**0.236**
$h = 5$	0.318	0.350	0.260	0.265	0.253	0.254	0.274	0.287	0.254	**0.251**
$h = 10$	0.339	0.405	0.279	0.278	0.271	0.274	0.282	0.326	0.274	**0.270**
$h = 20$	0.365	0.425	0.303	0.302	**0.296**	0.301	0.302	0.372	0.300	0.296

$\Delta = 15$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.245	0.236	0.264	0.306	0.230	0.215	0.345	0.217	0.215	**0.213**
$h = 2$	0.273	0.266	0.271	0.308	0.248	0.240	0.326	0.245	0.240	**0.237**
$h = 5$	0.325	0.340	0.297	0.315	0.287	0.287	0.323	0.298	0.285	**0.281**
$h = 10$	0.370	0.411	0.330	0.347	0.325	0.329	0.341	0.349	0.326	**0.322**
$h = 20$	0.434	0.458	0.379	0.379	0.373	0.377	0.372	0.405	0.371	**0.368**

Mean absolute errors for the predicting $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 6: *Predicting* $\log(RK)$*, MAE*

$\Delta = 30$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.254	0.246	0.317	0.391	0.264	0.236	0.383	0.239	0.239	**0.233**
$h = 2$	0.306	0.296	0.321	0.387	0.293	0.281	0.368	0.286	0.281	**0.276**
$h = 5$	0.380	0.380	0.360	0.389	0.345	0.345	0.374	0.356	0.344	**0.338**
$h = 10$	0.451	0.453	0.399	0.405	0.391	0.397	0.399	0.417	0.393	**0.390**
$h = 20$	0.510	0.494	0.428	0.426	**0.422**	0.438	0.425	0.463	0.430	0.426

$\Delta = 1$ hour

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.284	0.275	0.401	0.461	0.320	0.275	0.417	0.282	0.287	**0.274**
$h = 2$	0.361	0.346	0.402	0.439	0.358	0.338	0.410	0.345	0.344	**0.334**
$h = 5$	0.463	0.438	0.426	0.423	0.414	0.414	0.427	0.427	0.413	**0.409**
$h = 10$	0.534	0.498	0.447	0.457	**0.445**	0.458	0.451	0.478	0.455	0.451
$h = 20$	0.588	0.532	0.491	0.486	**0.485**	0.510	0.489	0.519	0.501	0.496

$\Delta = 2$ hours

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.333	0.319	0.448	0.501	0.366	**0.317**	0.439	0.335	0.334	0.320
$h = 2$	0.427	0.395	0.439	0.481	0.406	0.386	0.439	0.404	0.393	**0.386**
$h = 5$	0.527	0.484	0.456	0.492	**0.449**	0.455	0.463	0.482	0.455	0.453
$h = 10$	0.583	0.534	0.498	0.505	**0.492**	0.502	0.493	0.528	0.500	0.496
$h = 20$	0.643	0.554	**0.519**	0.537	0.520	0.545	0.530	0.551	0.539	0.534

$\Delta = 1$ day

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.439	0.405	0.503	0.595	0.440	**0.403**	0.479	0.423	0.417	0.407
$h = 2$	0.508	**0.456**	0.509	0.588	0.463	0.459	0.486	0.480	0.468	0.461
$h = 5$	0.601	0.526	0.524	0.547	**0.506**	0.517	0.507	0.536	0.513	0.509
$h = 10$	0.722	0.559	0.548	0.569	**0.542**	0.585	0.564	0.560	0.573	0.568
$h = 20$	0.804	0.581	0.582	0.585	**0.580**	0.628	0.618	0.581	0.613	0.608

Mean absolute errors for the predicting $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 7: *Predicting RK, MSE* $\times 1000$

$\Delta = 1$ minute

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.129	0.127	0.116	0.115	0.108	**0.102**	0.128	0.161	0.110	0.106
$h = 2$	0.153	0.153	0.121	0.116	0.112	0.110	0.124	0.167	0.112	**0.109**
$h = 5$	0.187	0.172	0.127	0.120	0.116	0.117	0.122	0.173	0.115	**0.114**
$h = 10$	0.200	0.175	0.131	0.124	0.120	0.122	0.122	0.175	0.118	**0.117**
$h = 20$	0.208	0.175	0.135	0.128	0.124	0.127	0.124	0.175	0.122	**0.121**

$\Delta = 2$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.095	0.093	0.085	0.087	0.082	**0.078**	0.104	0.111	0.083	0.080
$h = 2$	0.127	0.117	0.089	0.088	0.085	0.085	0.100	0.119	0.086	**0.084**
$h = 5$	0.149	0.141	0.094	0.091	0.090	0.092	0.098	0.131	0.090	**0.089**
$h = 10$	0.158	0.147	0.098	0.095	0.094	0.097	0.098	0.140	0.094	**0.093**
$h = 20$	0.169	0.147	0.104	0.101	0.100	0.103	0.102	0.145	0.099	**0.098**

$\Delta = 5$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.069	0.059	0.063	0.070	0.059	**0.054**	0.086	0.063	0.058	0.056
$h = 2$	0.098	0.078	0.065	0.070	0.063	**0.062**	0.082	0.072	0.063	0.062
$h = 5$	0.112	0.101	0.071	0.074	0.069	0.068	0.079	0.084	0.069	**0.068**
$h = 10$	0.127	0.115	0.078	0.078	0.076	0.074	0.080	0.095	0.074	**0.074**
$h = 20$	0.151	0.120	0.088	0.085	0.082	0.079	0.084	0.107	0.079	**0.079**

$\Delta = 15$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.077	0.059	0.068	0.080	0.057	**0.051**	0.086	0.053	0.052	0.051
$h = 2$	0.091	0.066	0.070	0.075	0.062	**0.058**	0.082	0.061	0.059	0.058
$h = 5$	0.122	0.084	0.076	0.071	0.070	**0.066**	0.079	0.073	0.066	0.066
$h = 10$	0.176	0.098	0.074	0.075	0.072	**0.071**	0.080	0.083	0.071	0.071
$h = 20$	0.131	0.107	0.083	0.083	0.081	0.079	0.083	0.096	0.079	**0.079**

Mean squared errors ($\times 1000$) *for the predicting* $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 8: *Predicting RK, MSE* $\times 1000$

$\Delta = 30$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.064	0.050	0.069	0.067	0.051	**0.044**	0.077	0.046	0.045	0.044
$h = 2$	0.090	0.059	0.061	0.065	0.054	**0.051**	0.073	0.055	0.051	0.052
$h = 5$	0.158	0.075	0.062	0.065	0.059	**0.058**	0.071	0.066	0.059	0.058
$h = 10$	0.103	0.088	0.069	0.070	0.067	0.066	0.073	0.079	0.067	**0.066**
$h = 20$	0.159	0.095	0.076	0.076	**0.075**	0.076	0.078	0.089	0.076	0.076

$\Delta = 1$ hour

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.047	0.037	0.052	0.080	0.039	**0.033**	0.066	0.037	0.036	0.034
$h = 2$	0.089	0.050	0.050	0.061	0.044	**0.043**	0.063	0.047	0.044	0.043
$h = 5$	0.080	0.065	0.057	0.055	0.054	0.052	0.063	0.063	0.054	**0.052**
$h = 10$	0.168	0.078	0.062	0.064	**0.061**	0.063	0.066	0.074	0.064	0.063
$h = 20$	0.108	0.084	0.072	0.072	**0.071**	0.073	0.072	0.082	0.073	0.073

$\Delta = 2$ hours

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.056	0.039	0.057	0.074	0.039	**0.033**	0.064	0.040	0.037	0.034
$h = 2$	0.100	0.052	0.054	0.060	0.046	**0.043**	0.062	0.052	0.046	0.044
$h = 5$	0.103	0.069	0.057	0.064	**0.056**	0.057	0.064	0.068	0.059	0.058
$h = 10$	0.151	0.079	0.067	0.070	**0.066**	0.067	0.069	0.077	0.068	0.067
$h = 20$	0.119	0.082	0.075	0.078	**0.075**	0.077	0.076	0.081	0.077	0.076

$\Delta = 1$ day

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	0.038	0.037	0.051	0.068	0.041	**0.035**	0.056	0.045	0.042	0.039
$h = 2$	0.058	0.050	0.057	0.066	0.048	**0.046**	0.056	0.055	0.050	0.049
$h = 5$	0.083	0.063	0.062	0.065	0.060	**0.059**	0.060	0.064	0.061	0.060
$h = 10$	0.109	0.069	0.068	0.070	**0.066**	0.070	0.069	0.069	0.070	0.070
$h = 20$	0.147	0.076	0.075	0.075	**0.074**	0.083	0.081	0.076	0.081	0.080

Mean squared errors ($\times 1000$) *for the predicting* $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 9: *Predicting RK, MAE* $\times 1000$

$\Delta = 1$ minute

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	6.655	6.075	5.718	5.658	5.398	5.339	5.820	6.833	5.415	**5.326**
$h = 2$	7.046	6.649	5.800	5.640	5.505	5.474	5.742	7.026	5.494	**5.435**
$h = 5$	7.291	7.214	5.926	5.728	5.604	5.579	5.676	7.225	5.580	**5.532**
$h = 10$	7.455	7.312	6.021	5.823	5.696	5.682	5.689	7.300	5.663	**5.616**
$h = 20$	7.621	7.316	6.133	5.942	5.818	5.823	5.762	7.316	5.780	**5.734**

$\Delta = 2$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	5.021	4.724	4.233	4.240	4.050	4.027	4.687	4.961	4.079	**4.005**
$h = 2$	5.317	5.281	4.305	4.252	4.144	4.147	4.531	5.230	4.173	**4.111**
$h = 5$	5.570	6.067	4.456	4.357	4.280	4.306	4.441	5.660	4.307	**4.250**
$h = 10$	5.784	6.306	4.609	4.490	4.428	4.475	4.488	6.002	4.449	**4.395**
$h = 20$	6.034	6.327	4.815	4.701	4.640	4.688	4.629	6.241	4.644	**4.595**

$\Delta = 5$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	3.579	3.358	3.189	3.434	3.022	2.952	3.960	3.129	2.968	**2.929**
$h = 2$	3.908	3.767	3.255	3.442	3.149	3.112	3.763	3.391	3.130	**3.095**
$h = 5$	4.291	4.589	3.485	3.573	3.387	3.367	3.674	3.846	3.380	**3.344**
$h = 10$	4.648	5.195	3.754	3.760	3.650	3.631	3.771	4.321	3.635	**3.598**
$h = 20$	5.064	5.416	4.090	4.061	3.959	3.954	3.991	4.833	3.948	**3.910**

$\Delta = 15$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	2.982	2.789	3.219	3.696	2.778	2.527	3.910	2.574	2.545	**2.516**
$h = 2$	3.406	3.145	3.285	3.648	2.994	2.844	3.727	2.918	2.850	**2.827**
$h = 5$	4.118	3.864	3.553	3.606	3.388	3.321	3.680	3.455	3.300	**3.283**
$h = 10$	4.662	4.504	3.737	3.850	3.660	3.686	3.806	3.919	3.648	**3.620**
$h = 20$	4.995	4.914	4.138	4.133	4.061	4.131	4.068	4.420	4.059	**4.023**

Mean absolute errors ($\times 1000$) *for the predicting* $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

Table 10: *Predicting RK, MAE* $\times 1000$

$\Delta = 30$ minutes

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	2.997	2.775	3.569	4.013	2.955	2.597	3.933	2.646	2.641	**2.589**
$h = 2$	3.674	3.262	3.477	3.907	3.186	3.032	3.799	3.093	3.039	**3.020**
$h = 5$	4.555	3.921	3.668	3.893	3.526	3.547	3.805	3.648	3.526	**3.502**
$h = 10$	4.778	4.474	3.968	4.063	3.896	3.965	3.988	4.126	3.921	**3.883**
$h = 20$	5.543	4.814	4.255	4.254	**4.216**	4.387	4.232	4.529	4.291	4.252

$\Delta = 1$ hour

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	3.143	2.882	3.852	4.666	3.164	**2.791**	3.985	2.859	2.894	2.794
$h = 2$	4.113	3.483	3.759	4.242	3.417	3.312	3.891	3.350	3.342	**3.292**
$h = 5$	4.622	4.096	4.016	3.977	3.878	3.869	4.000	3.978	3.872	**3.812**
$h = 10$	5.836	4.594	4.189	4.310	**4.177**	4.319	4.215	4.410	4.265	4.232
$h = 20$	5.780	4.857	4.569	**4.507**	4.510	4.740	4.535	4.745	4.636	4.590

$\Delta = 2$ hours

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	3.640	3.205	4.290	4.902	3.399	**3.045**	4.063	3.170	3.192	3.060
$h = 2$	4.636	3.738	4.127	4.478	3.753	3.585	4.033	3.705	3.657	**3.574**
$h = 5$	5.412	4.417	4.220	4.545	**4.151**	4.231	4.239	4.357	4.211	4.185
$h = 10$	6.196	4.775	4.544	4.616	**4.491**	4.605	4.491	4.702	4.555	4.528
$h = 20$	6.249	4.910	**4.664**	4.816	4.675	4.921	4.767	4.881	4.828	4.790

$\Delta = 1$ day

	RW	AR(1)	AR(5)	AR(10)	HAR(3)	ARFIMA	RFSV	$\Gamma - \mathcal{BSS}$	Cauchy	$\mathcal{BSS}$
$h = 1$	3.452	3.223	3.966	4.737	3.455	**3.208**	3.755	3.336	3.328	3.245
$h = 2$	4.152	**3.589**	4.039	4.591	3.636	3.639	3.802	3.738	3.689	3.636
$h = 5$	4.962	4.062	4.060	4.261	**3.930**	4.060	3.938	4.114	3.991	3.958
$h = 10$	6.080	4.370	4.311	4.452	**4.263**	4.563	4.413	4.374	4.456	4.421
$h = 20$	6.885	4.631	4.641	4.667	**4.629**	4.972	4.905	4.631	4.837	4.800

Mean absolute errors ($\times 1000$) *for the predicting* $\log(RK)$ *h-period ahead. The first* 200 *periods were used for initial estimation.*

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*, 579–625.

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

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

Barndorff-Nielsen, O. E., J. M. Corcuera, and M. Podolskij (2009). Power variation for Gaussian processes with stationary increments. *Stochastic Process. Appl. 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., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal 12*, C1–C32.

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 Ser. Comput. Appl. Math.*, 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 64*, 253–280.

Bayer, C., K. Friz, and J. Gatheral (2015). Pricing under rough volatility. *Working paper available at SSRN: http://ssrn.com/abstract=2554754*.

Bennedsen, M. (2015a). Rough electricity: a new fractal multi-factor model of electricity spot prices. *Working Paper available at SSRN: 2636829*.

Bennedsen, M. (2015b). Semiparametric bootstrap approach to estimation of the fractal dimension and the Hurst coefficient with application to testing for self-similarity and semimartingality. *Working Paper*.

Bennedsen, M., A. Lunde, and M. S. Pakkanen (2015). Hybrid scheme for Brownian semistationary processes. *Working paper, arXiv:1507.03004*.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular Variation*. 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*, 81–106.

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*, 101–149.

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

Corlay, S., J. Lebovits, and J. L. Véhel (2014). Multifractional stochastic volatility models. *Mathematical Finance 24*(2), 364–402.

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

Fukasawa, M. (2015). Short-time at-the-money skew and rough fractional volatility. *Working paper available at arXiv: 1501.06980*.

Gatheral, J. (2006). *The Volatility Surface*. Wiley Finance.

Gatheral, J., T. Jaisson, and M. Rosenbaum (2014). Volatility is rough. *Working paper*.

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.

Grimmett, G. and D. Stirzaker (2001). *Probability and Random Processes* (3rd ed.). Oxford University Press.

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*, 1–34.

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*, 2249–2276.

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.

A Proofs

Proof of Theorem 3.1. Since σ is stationary are interested in

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

Suppose for simplicity that $\xi_t = 1$ for all t. We get using the definition (3.2) of σ

$$\begin{aligned} Cov(\sigma_t, \sigma_{t+h}) &= \mathbb{E}[\sigma_t \sigma_{t+h}] - \mathbb{E}[\sigma_0]^2 \\ &= \mathbb{E}[\exp(X_t + X_{t+h})] \exp(-\mathbb{E}[X_0^2]) - 1. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}[\exp(X_t + X_{t+h})] &= \exp\left(\frac{1}{2} Var(X_t + X_{t+h})\right) \\ &= \exp\left(Var(X_0) + cov(X_0, X_{t+h})\right) \\ &= \exp\left(\gamma_X(0) + \gamma_X(h)\right), \end{aligned}$$

where we write $\gamma_X(h)$ for $Cov(X_h, X_0) = \mathbb{E}[X_h X_0]$. Putting it all together, and using that $Var(\sigma_0) = \mathbb{E}[\sigma_0^2] - 1 = \exp(\gamma_X(0)) - 1$ we arrive at

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

where we in the second-to-last line Taylor expanded the exponential and and in the last line we used the assumption (3.1). □

Proof of Theorem 3.2. Using the same approach as in the proof of Theorem 3.1, we get by Taylor expansion:

$$\begin{aligned} \rho(h) &= \frac{\exp(\gamma_X(h)) - 1}{\exp(\gamma(0)) - 1} \\ &= \frac{1}{\exp(\gamma(0)) - 1} \sum_{k=1}^{\infty} \frac{\gamma_X(h)^k}{k!} \end{aligned}$$

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

by the assumption (3.1). □

Proof of Proposition 3.1. Recall that we can write

$$
\rho_X(h) = \frac{\int_0^\infty g(x)g(x+h)dx}{\int_0^\infty g(x)^2 dx}, \quad h \geq 0. \tag{A.1}
$$

Clearly we can take $h > 1$ and now

$$
\begin{aligned}
\int_0^\infty g(x)g(x+h)dx &= \int_0^1 x^\alpha L_0(x)(x+h)^{-\beta} L_1(x+h)dx + \int_1^\infty x^{-\beta}(x+h)^{-\beta} L_1(x)L_1(x+h)dx \\
&:= I_{1,h} + I_{2,h},
\end{aligned}
$$

where

$$
\begin{aligned}
I_{1,h} &= \int_0^1 x^\alpha L_0(x)(x+h)^{-\beta} L_1(x+h)dx, \\
I_{2,h} &= \int_1^\infty x^{-\beta}(x+h)^{-\beta} L_1(x)L_1(x+h)dx.
\end{aligned}
$$

Our strategy is to show that $I_{1,h} \sim h^{-\beta} L_1(h)$ and $I_{2,h} = o\left(h^{-\beta} L_1(h)\right)$ as $h \to \infty$, from which we, together with (A.1) get the desired convergence rate, $\rho(h) \sim h^{-\beta} L_1(h)$. Take first $I_{1,h}$. Since $\beta > 0$,

$$
\begin{aligned}
I_{1,h} &\leq h^{-\beta} \int_0^1 x^\alpha L_0(x) L_1(x+h)dx \\
&= h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \frac{L_1\left((x/h+1)h\right)}{L_1(h)} dx \\
&\sim h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x)dx, \quad h \to \infty,
\end{aligned}
$$

by the properties of slowly varying functions and where we applied the dominated convergence theorem. Similarly, we can make the opposite evaluation:

$$
\begin{aligned}
I_{1,h} &\geq (1+h)^{-\beta} \int_0^1 x^\alpha L_0(x) L_1(x+h)dx \\
&= (1+h)^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x) \frac{L_1\left((x/h+1)h\right)}{L_1(h)} dx \\
&\sim h^{-\beta} L_1(h) \int_0^1 x^\alpha L_0(x)dx, \quad h \to \infty.
\end{aligned}
$$

Putting the two inequalities together we get that $I_{1,h} \sim h^{-\beta} L_1(h)$ as $h \to \infty$. For the second integral we get

$$\begin{aligned} I_{2,h} &\leq (1+h)^{-\beta} \int_1^\infty x^{-\beta}(x+h)^{-\beta} L_1(x) L_1(x+h) dx \\ &= (1+h)^{-\beta} L_1(h) \int_1^\infty x^{-\beta} L_1(x) \frac{L_1\left((x/h+1)h\right)}{L_1(h)} dx \\ &\sim h^{-\beta} L_1(h) \int_1^\infty x^{-\beta} L_1(x) dx, \quad h \to \infty, \end{aligned}$$

where $\int_1^\infty x^{-\beta} L_1(x) dx < \infty$ since $\beta > 1$. □

Proof of Proposition 3.2. As we intend h to diverge to infinity we may clearly take $h > 1$. Recall that

$$\rho_X(h) = \frac{\int_0^\infty g(x) g(x+h)}{\int_0^\infty g(x)^2 dx}.$$

Now, by Tonelli's Theorem, we may interchange the order of integration in the following,

$$\int_0^\infty \int_0^\infty g(x) g(x+h) dx dh = \int_0^\infty g(x) \int_0^\infty g(x+h) dh dx.$$

Since $h > 1$, we get, using Karamata's theorem (Bingham et al., 1989, Proposition 1.5.10.),

$$\int_0^y g(x+h) dh = \int_0^y (x+h)^{-\beta} L_1(x+h) dh \sim (x+y)^{1-\beta} L_1(x+y), \quad y \to \infty.$$

Since $1 - \beta > 0$ by assumption, we have $\int_0^\infty g(x+h) dh = \infty$. Now, since for large enough x we have $g(x) = x^\beta L_1(x)$ we easily conclude that there is a set $A \subset \mathbb{R}_+$ with $Leb(A) > 0$ such that $g(x) > \epsilon > 0$ for $x \in A$. Therefore,

$$\begin{aligned} \int_0^\infty \int_0^\infty g(x) g(x+h) dx dh &= \int_A g(x) \int_0^\infty g(x+h) dh dx + \int_{A^c} g(x) \int_0^\infty g(x+h) dh dx \\ &\geq \epsilon \int_A \int_0^\infty g(x+h) dh dx + \int_{A^c} g(x) \int_0^\infty g(x+h) dh dx \\ &= \infty, \end{aligned}$$

since $Leb(A) > 0$. □