

# MODELLING AND FORECASTING FRACTIONAL VOLATILITY

By Anine Eg Bolko

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

August 2020





## PREFACE

This dissertation was written as a part of my PhD programme at the Department of Economics and Business Economics at Aarhus University during the period September 2017 to September 2020. I am thankful to the Department of Economics and Business Economics and to the Center for Research in Econometric Analysis of Time Series (CREATES) for a pleasant research environment and financial support.

Through my enrolment I have been surrounded by inspiring people, whom I owe a thank. First and foremost, my main supervisor Kim Christensen. I am grateful for the massive support, valuable insights and plenty of helpful conversation on anything and everything. I would also like to thank my co-supervisor, Bezirgen Veliyev, for useful discussions, nice teamwork and friendly support throughout the years, and my co-supervisor, Mark Podolskij, for always having an open door. I have learned a lot from all of you.

From January through August 2019, I had the pleasure to visit Mikko Pakkanen at Department of Mathematics, Imperial College London. I would like to thank Mikko Pakkanen for being a welcoming host, the many good conversations on the research while I was there, and the continuing mentoring and teamwork afterwards.

Also a huge thanks to all colleagues at the department for creating an inspiring research and social environment. In that regard, a special thank to Solveig Nygaard Sørensen for everything from organising practical matters to proofreading. And of course also a huge thanks to my fellow PhD students, without you all to share lunches, coffee breaks, board games and even yoga classes with, I would not have enjoyed the time here to the same extend. Special thanks to my former office mates Kristoffer Bertelsen and Nikolaj Udengaard and my current office mates, Frederik Okslund, Mathias Siggaard and Sebastian Jensen, for many fun memories and countless of more or less useful discussions.

Finally, my family and friends outside university have a part in this, thank you for being supportive and patient, when I have disconnected from reality. An extra thank to my mum, Lene Eg, for long supportive conversations, my dad, Jakob Bolko, for curious phone calls, and of course my sister, Josefine, for sharing busy evenings with me. Last, but not least I have reserved a heartfelt thank to my boyfriend, Martin Lundgaard, who has always listened and been consistently supportive. Without you I

have not managed to finish this.

*Anine Eg Bolko*

*Aarhus, September 2020*

# CONTENTS

<b>Summary</b>	<b>v</b>
<b>Danish summary</b>	<b>ix</b>
<b>1 Roughness in spot variance? A GMM approach for estimation of log-normal stochastic volatility models using realized measures</b>	<b>1</b>
1.1 Introduction . . . . .	2
1.2 The setting . . . . .	4
1.3 GMM Estimation . . . . .	8
1.4 Simulation study . . . . .	18
1.5 Empirical application . . . . .	23
1.6 Conclusion . . . . .	26
1.A Proofs . . . . .	26
1.2 References . . . . .	44
<b>2 Option pricing using rough realized measures</b>	<b>49</b>
2.1 Introduction . . . . .	50
2.2 The Theoretical Framework . . . . .	51
2.3 Derivative pricing . . . . .	55
2.4 Estimation: Methodology . . . . .	57
2.5 Conclusion . . . . .	62
2.A Proofs . . . . .	62
2.2 References . . . . .	64
<b>3 Forecasting realized variance using a simple ARIMA model</b>	<b>67</b>
3.1 Introduction . . . . .	68
3.2 Settings . . . . .	69
3.3 Forecasting procedure . . . . .	75
3.4 Empirical analysis . . . . .	76
3.5 Conclusion . . . . .	82
3.A Appendix . . . . .	84
3.2 References . . . . .	86



## SUMMARY

This dissertation contains three independent chapters on modeling and forecasting stochastic asset return variance. The common theme of all three chapters is the focus on fractal index and memory properties. Whether it is for derivative pricing or forecasting purposes, both these features are of great importance and are often closely related, even though they describe very different properties. Fractional Brownian motions (fBm) are obvious examples of self-similar processes which can generate long memory and rough behaviour, though not simultaneously. Therefore they are convenient as building blocks in the continuous time stochastic volatility (SV) framework. Empirical investigations of the variance process conclude that it has a strong memory, while recent findings suggest that the variance process is rough as well. Since this combination is not compatible with the use of fBm, three cases arise: The first option is that the fBm is a bad building block for modeling stochastic volatility, and a model which can decompose the memory and the regularity is necessary. Perhaps the long memory property is spurious or perhaps the observed rough behaviour is falsely detected for example caused by noise. This dissertation illuminates these kinds of questions and investigates whether such models are useful in practice. Even though all three chapters have this topic in common, the methods and applications are very different across the chapters.

The first chapter is joint work with Professor Kim Christensen, Associate Professor Mikko Pakkanen and Associate Professor Bezirgen Veliyev. It deals with the question whether spot volatility is rough by developing a parametric estimation procedure for log-normal SV models, which works simultaneously with fractal indexes in both ends of the spectrum. It is tailored to the financial asset market, where the spot volatility is latent, but information on it can be deduced from high-frequency trade data and realized measures. To circumvent problems with ultra high-frequency data and microstructure noise, our generalized method of moments (GMM) estimation approach is based on daily integrated variance rather than spot variance. This also ensures accessibility to a Central Limit Theorem. In contrast to conventional estimators in this kind of continuous time models, the estimation is done in one step for all involved parameters instead of the usual pre-estimation of the fractal index by periodogram- or variogram-type estimators. This choice ensures that if other memory parts are present, they will be taken into account, and the estimate of the fractal index stays

unbiased. Therefore, it is possible to decide if the slowly decaying autocorrelation function observed in the market is best fitted by the outcome of a genuine long memory fBm or if it is a spurious feature implied by a slowly decaying but short-memory process and a rough fBm. In practice, the GMM estimator is impacted by a small sample bias, but we cancel out the effect of this by introducing a correction term depending on the sampling frequency of the realized measure. In a simulation study using 5-minute realized variance it is clear that the correction is not only useful, but strictly necessary to obtain accurate results. We apply the estimation procedure to 29 indexes from all around the world, and the answer is clear. Spot variance is very rough, with a Hurst index around 0.05, and the slowly decaying autocorrelation function might be a result of a very slow mean-reversion mechanism.

Based on the findings of roughness in the spot variance, it is natural to consider an equivalent model for derivative pricing. For this purpose it is necessary to know the dynamics of the asset price under the risk neutral pricing measure. Even after assuming a parametric fractional setup, the calibration of the model parameters is not straightforward. In the fractional log-normal SV models neither a closed form density function nor a characteristic function is available, and poor computational efficiency of simulation schemes makes calibration by means of Monte Carlo hard if not impossible. The second chapter deals with an alternative way to determine the values of the model parameters to be used for option pricing. The procedure relies on estimation of the parameters under the physical probability measure and defines the measure change between this and the risk neutral pricing measure. A one-to-one mapping between the physical and risk neutral model parameters is established in a variance model driven by a fractional Ornstein-Uhlenbeck process, which only depends on the model parameters and the market price of log-volatility risk. The latter is restricted such that the model preserves its structure from the real world to the risk neutral pricing model, and estimation of the market price of log-volatility risk leads to knowledge of the pricing parameters, and thereby makes option pricing possible.

The third and final chapter of this dissertation returns focus to the slowly decaying autocorrelation observed in the index return variance. Standard unit-root tests and semi-parametric estimation of the long memory parameter reveal that log-realized variance is close to, if not quite, non-stationary. Instead of considering close to non-stationary SV models by slow mean-reversion as for example in log-normal rough SV models, it is likely that forecasting based on the surely stationary log-variance increment series performs well. A visual inspection of the autocorrelation function backed by estimation of the fractal index reveals that this series is anti-persistent and very similar to a fBm or an ARFIMA(0,  $d$ , 0) with fractal index,  $d < 0$ . Instead of dealing with a complicated and possibly biased semi-parametric estimator, this chapter suggests an approximation of this structure by a simple ARIMA(0, 1,  $q$ ) model for log-realized variance and the performance for out-of-sample forecasting is investigated



to see if these models perform equally well as popular benchmark models. The answer is yes, by selecting quite a small number of moving average lags, these simple models perform equally well to both the fractional model and other relevant benchmark models.



## DANISH SUMMARY

Denne afhandling består af tre uafhængige kapitler, der hver især omhandler modellering og forudsigelse af stokastisk varians for afkast på finansielle aktiver. De tre kapitler har det tilfælles, at de alle fokuserer på fraktalindeks og hukommelsesegenskaber for variansprocessen. Uanset om det er med øje for prisfastsættelse af derivater eller i forbindelse med forudsigelse af fremtidig varians, er disse egenskaber vigtige og ofte tæt forbundet, selvom de beskriver vidt forskellig træk. Fraktionelle Brownske bevægelser (fBm) er selv-similære processor, der kan generere både lang hukommelse og ru udfaldsstier, dog ikke samtidigt. Det er derfor hensigtsmæssige byggesten i kontinuerte modeller for stokastisk volatilitet. Empiriske undersøgelser viser at variansprocessen har en stærk hukommelse, og senest er der også fundet evidens for at processen har ru udfaldsstier. Denne kombination er ikke mulig at konstruere udelukkende ved brug af fBm, og det medfører tre mulige scenarier. Den første mulighed er, at fBm en dårlig byggesten til at modellere stokastisk varians, og at det er nødvendigt at benytte processer der kan adskille hukommelse og regularitet. Måske er enten den lange hukommelse ikke opfyldt fra et matematisk synspunkt, eller også er den observerede ruhed bare et resultat af støj. Afhandlingen her belyser spørgsmål af denne slags, og undersøger om fraktionelle modeller er givende i praksis. På trods af at alle tre kapitler har dette fælles emne, er metoderne og anvendelserne i de forskellige kapitler meget forskellige.

Det første kapitel er udarbejdet i samarbejde med professor Kim Christensen, lektor Mikko Pakkanen og lektor Bezirgen Veliyev. Det behandler spørgsmålet om hvorvidt spotvariansprocessen er ru ved at udvikle en parametrisk estimeringsprocedure for log-normale SV modeller, der kan håndtere fraktalindeks i begge ender af skalaen på samme tid. De er skræddersyet til det finansielle aktiemarked, hvor spotvarians ikke er observerbar, men hvor information omkring variansen kan udledes ved brug af højfrekvensdata og realiserede mål for variansen. For at undgå problemer med ultra-højfrekvensdata og støj forbundet med mikrostrukturer i markedet tager vores generaliserede moment (GMM) baserede estimeringsmetode udgangspunkt i integreret varians i stedet for spotvarians. Det sikrer yderligere adgang til en Central Grænseværdisætning. Modsat konventionelle estimators for denne type kontinuerttids modeller, så bygger metoden på en et-steps procedure for alle involverede parametre i stedet for at præ-estimere fraktalindekset med periodogram- eller vario-

grambaserede estimatorer. Denne fremgangsmåde sikrer at selv ved tilstedeværelse af andre typer hukommelse vil estimeringen tage højde for dette og undgå skævhed i estimatet for fraktalindekset. Derfor er det muligt at tage højde for den langsomt faldende autokorrelationsfunktion observeret i markedet og bestemme hvorvidt ægte lang hukommelse i fBm afspejler variansen bedst eller om effekten er uægte og i stedet et resultat af en process med kort hukommelse og en ru fBm. I praktisk er GMM-estimatoren påvirket af en skævhed i endeligt data, men vi eliminerer denne effekt ved at introducere en korrektion, der afhænger af datafrekvensen benyttet til at konstruere det realiserede mål for varians. I et simulationsstudie bruger vi 5-minutters realiseret varians, og det er herfra klart at korrektionen ikke blot er hjælp-som, men strengt nødvendig for at opnå akkurate resultater. Vi anvender estimatoren på data fra 29 indekser fra hele verden, og svaret er klart. Spotvariens er meget ru med et Hurst indeks omkring 0.05, og den langsomt aftagende autokorrelationsfunktion kan være et resultat af en meget langsom mekanisme, der trækker variansen mod dens ubetingede middelværdi.

Baseret på undersøgelser der viser at spotvariens er ru, er det naturligt at undersøge en tilsvarende model til at fastsætte priser på derivater. For at kunne det, er det nødvendigt at kende dynamikker for prisprocessen under det risikoneutrale prisfastsættelsesmål. Selv med en parametriske form af den fraktionelle opsætning er kalibrering af modellens parametre kompliceret. I den fraktionelle log-normale SV model er der hverken adgang til en lukket løsning for tæthedsfunktionen eller den karakteristiske funktion, og mangel på hurtige beregningsmetoder af simuleringsskemaer gør kalibrering ved hjælp af Monte Carlo metoder tæt på umulig. Det andet kapitel fokuserer på en alternativ metode til at bestemme værdier for modelparametre, der kan benyttes til prisfastsættelse. Proceduren udnytter at parametrene kan estimeres under det fysiske sandsynlighedsmål og definerer et målskifte mellem dette og det risikoneutrale prisfastsættelsesmål. En én-til-én afbildning mellem de to modeller bliver etableret i en model hvor log-variens er drevet af en fraktionel Ornstein-Uhlenbeck proces, og denne afhænger udelukkende af modelparametrene og af markedets pris på log-volatilitetsrisiko. Formen på sidstnævnte er begrænset på en sådan måde at modelstrukturen er bevaret ved målskiftet mellem den virkelige verden og risikoneutrale verden, og estimering af markedets pris på log-volatilitetsrisiko medfører viden om parametrene og gør derved prisfastsættelse muligt.

Det tredje og sidste kapitel i denne afhandling vender fokus tilbage til den langsomt faldende autokorrelationsfunktion observeret for variens i afkast på indeks. Standard unit-root test og semi-parametriske estimering af hukommelsesparameteren afslører at log-realiseret variens er tæt på, hvis ikke faktisk, ikke-stationær. I stedet for at betragte tæt på ikke-stationære SV modeller, der vender langsomt tilbage til deres ubetingede middelværdi, som for eksempel log-normale ru SV modeller, er det sandsynligt at gode forudsigelser for variansen kan opnås ved at betragte den stationære serie af tilvækster i log-variansen. Et kig på autokorrelationsfunktionen bakket op

af estimering af fraktalindekset afslører at denne serie er meget antipersistent og har egenskaber, der ligner en fBm eller en ARFIMA( $0, d, 0$ ) proces med fraktalindeks,  $d < 0$ . I stedet for at behandle komplicerede og muligvis skæve semiparametriske estimatorer, foreslår dette kapitel en approksimation af strukturen ved hjælp af simple ARIMA( $0, 1, q$ ) modeller log-realiseret varians og præstationen i forbindelse med forudsigelser "out-of-sample" bliver undersøgt for at se om disse modeller performer på samme niveau som andre populære modeller. Svaret er ja. Ved at vælge et ret lille antal moving average forsinkelser kan disse modeller hamle op med både fraktionelle modeller og andre modeller der er relevante at sammenligne med.



**ROUGHNESS IN SPOT VARIANCE?  
A GMM APPROACH FOR ESTIMATION OF  
LOG-NORMAL STOCHASTIC VOLATILITY MODELS  
USING REALIZED MEASURES**

**Anine Eg Bolko**

*Aarhus University and CREATES*

**Kim Christensen**

*Aarhus University and CREATES*

**Mikko Pakkanen**

*Imperial College London and CREATES*

**Bezirgen Veliev**

*Aarhus University and CREATES*

**Abstract**

In this paper, we develop a generalized method of moments approach for joint estimation of the parameters of a fractional log-normal stochastic volatility model. We show that with an arbitrary Hurst exponent an estimator based on integrated variance is consistent. Moreover, under stronger conditions we also derive a central limit theorem. These results stand even when integrated variance is replaced with a realized measure of volatility calculated from discrete high-frequency data. However, in practice a realized estimator contains sampling error, the effect of which is to skew the fractal coefficient toward “roughness”. We construct an analytical approach to

control this error. In a simulation study, we demonstrate convincing small sample properties of our approach based both on integrated and realized variance over the entire memory spectrum. We show that the bias correction attenuates any systematic deviance in the estimated parameters. Our procedure is applied to empirical high-frequency data from numerous leading equity indexes. With our robust approach the Hurst index is estimated around 0.05, confirming roughness in integrated variance.

**JEL Classification:** C10; C50.

**Keywords:** GMM estimation; realized variance; rough volatility; stochastic volatility.

## 1.1 Introduction

Stochastic volatility (SV) models are pervasive in finance. Over the years, a variety of different models—each with its own volatility dynamics—were developed such as the log-normal model (e.g. Taylor, 1986), the square-root diffusion (e.g. Heston, 1993), or more complicated processes where volatility is driven by a non-Gaussian pure-jump component, e.g. Barndorff-Nielsen and Shephard (2001); Todorov and Tauchen (2011).

In this paper, we investigate the log-normal SV model, which has been extensively studied in previous work, e.g. Alizadeh, Brandt and Diebold (2002). This class is a promising starting point, because the unconditional distribution of realized variance is close to log-normal (see e.g. Andersen, Bollerslev, Diebold and Ebens, 2001; Andersen, Bollerslev, Diebold and Labys, 2003). However, while there is general agreement that log-normal volatility offers a decent description of return variation in financial asset prices, there is no consensus on the properties of the background driving Gaussian process. In a standard setting, it is assumed to be a Brownian motion. The mean-reversion and volatility-of-volatility parameters of the model then control both the local properties of volatility and also determine its longer-run persistence. In this instance, there are multiple papers dealing with estimation of the parameters of the log-normal SV model, for example using the method of moments- or likelihood-based approaches (e.g. Taylor, 1986; Melino and Turnbull, 1990; Duffie and Singleton, 1993; Harvey, Ruiz and Shephard, 1994; Gallant, Hsieh and Tauchen, 1997; Fridman and Harris, 1998). In the context of generalized method of moments (GMM) estimation, Andersen and Sørensen (1996) offer further advice on how to select moment criteria and the weighting matrix in order to get good results in small samples.

When the driving process is a fractional Brownian motion, which does not have independent increments, less is known. In this non-Markovian setting, part of the memory in volatility is reallocated to the background driving process via an additional parameter, the fractal index or Hurst exponent (after Hurst, 1951). Comte and Renault (1998) propose such a version of the log-normal SV model, where the Hurst exponent



is larger than one-half—as implied by a standard Brownian motion—thus inducing positive serial correlation in the increments of the process. Bennedsen (2016); Euch and Rosenbaum (2018); Gatheral, Jaisson and Rosenbaum (2018), among others, study roughness in volatility captured by a fractal index smaller than one-half, rendering volatility anti-persistent and highly erratic at short time scales. The typical estimation approach in the fractional setting is a semi-parametric two-stage procedure, where the Hurst index is pre-estimated, before the other parameters are recovered. While this procedure may yield consistent parameter estimates, it is generally inefficient and may be severely biased in finite samples.

In this paper, we extend the GMM procedure for joint estimation of the parameters of the log-normal SV model with a general fractal index. We show that our proposed estimator is consistent and asymptotically normal. An attractive feature of our procedure is that moment expressions are derived in near closed-form facilitating the implementation without recourse to simulation-based approaches. As in many papers before this one, we appeal to the time series properties of integrated variance to construct our estimator, an idea pioneered by Bollerslev and Zhou (2002); Corradi and Distaso (2006); Todorov (2009).

In practice, the integrated variance is unobserved. Realized variance, which is computed from high-frequency data, is a consistent estimator of integrated variance and can replace it in the calculations. In previous work, this is handled by showing convergence in probability of the parameter estimator in a doubly-asymptotic in-fill and long-span setting, such that the volatility discretization error is small enough to be ignored. In the subsequent applications, the volatility proxy then enters directly in place of integrated variance.

Substitution of the latent volatility with a proxy, however, entails a measurement error. This obfuscates the underlying integrated variance dynamics, which can be detrimental to the estimation procedure if unaccounted for (e.g. Meddahi, 2002; Hansen and Lunde, 2014). In particular, the imposed moment conditions for integrated variance are not valid for the proxy. Barndorff-Nielsen and Shephard (2002) employ a state-space system and the Kalman filter to smooth out realized variance prior to maximum quasi-likelihood estimation of their SV model, see also Meddahi (2003). In this paper, we construct an analytic bias correction that controls for the measurement error. Following Patton (2011) we introduce a high-level assumption employing a generic realized measure to proxy for integrated variance. We show in the GMM setting that the additional sampling variation is captured and corrected by adding the measurement error variance to the second moment of integrated variance. Our main asymptotic theory is therefore long-span with time going to infinity but high-frequency data sampled at a fixed frequency. As an aside, we complement the analysis by deriving the double-asymptotic result, where the correction is immaterial.

We investigate our estimator in a simulation study, where various configurations of a fractional log-normal stochastic volatility model with different Hurst parameters

covering the rough and long-memory setting are inspected. We note that our procedure is both unbiased and relatively accurate for the unknown parameters, once the above bias correction is adopted. In an empirical application, we study an extensive selection of major equity indexes and confirm roughness in the volatility process, even after smoothing out the effect of noise in the volatility proxy. In those data we consistently locate a roughness parameter around 0.05, in line with the findings of recent work by Fukasawa, Takabatake and Westphal (2019).

The rest of this paper is organized as follows. Section 1.2 presents the general log-normal SV model and studies the properties of integrated variance within this framework. The GMM procedure is introduced in Section 1.3, along with theoretical results of the estimation. Section 1.4 examines the performance of our estimator in a Monte Carlo study. In Section 1.5, we apply the estimation procedure to real data and compare our findings with the previous literature. We conclude in Section 1.6 and leave some theoretical derivations to the Appendix.

## 1.2 The setting

We model the log-price of a financial asset,  $X = (X_t)_{t \geq 0}$ , as an adapted continuous-time stochastic process defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ . We suppose a standard arbitrage-free market, in which asset prices are of semimartingale form (e.g., Back, 1991; Delbaen and Schachermayer, 1994). We assume  $X$  can be described by an Itô process:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \quad t \geq 0, \quad (1.1)$$

where  $X_0$  is  $\mathcal{F}_0$ -measurable,  $\mu = (\mu_t)_{t \geq 0}$  is a predictable drift process,  $\sigma = (\sigma_t)_{t \geq 0}$  is a càdlàg volatility process and  $W = (W_t)_{t \geq 0}$  is a standard Brownian motion.

The spot variance  $\sigma^2 = (\sigma_t^2)_{t \geq 0}$  is given by:

$$\sigma_t^2 = \xi \exp\left(Y_t - \frac{1}{2}\kappa(0)\right), \quad t \geq 0, \quad (1.2)$$

where  $\xi \in \Xi \subset (0, \infty)$  is a scale parameter, representing the unconditional mean of the stochastic variance, while  $Y = (Y_t)_{t \geq 0}$  is a zero-mean stationary Gaussian process with covariance function  $\kappa(u) = \text{cov}(Y_0, Y_u) = \kappa_\phi(u)$ ,  $u \geq 0$ , parameterized by  $\phi \in \Phi \subset \mathbb{R}^p$ . We assume  $\Xi$  and  $\Phi$  are compact, so that  $\Theta = \Xi \times \Phi \subset \mathbb{R}^{p+1}$  is compact, and write  $\theta = (\xi, \phi) \in \Theta$ .

Note that we do not restrict the model to a Markovian volatility process nor a semimartingale setting.<sup>1</sup> This is not a problem for absence of arbitrage and existence

<sup>1</sup>The log-normal distribution is invariant to (non-zero) power transformations. This implies that “volatility,” which in financial economics is more often associated with the standard deviation—or the square-root of the variance—is also log-normal if the variance is (and vice versa). Hence, volatility is applied loosely here to mean either variance or standard deviation. The meaning should be apparent from the context and not cause much confusion in this particular setting.

of an equivalent risk-neutral probability measure (although it is not unique in our setup), since the volatility itself is not the price of a tradable asset.<sup>2</sup>

To maintain a streamlined exposition, we exclude a jump component in  $X$ . The theory should at least be robust to the addition of finite-activity jumps, but then one needs to pay attention to the practical implementation.<sup>3</sup>

The integrated variance on day  $t$  is defined as:

$$IV_t = \int_{t-1}^t \sigma_s^2 ds, \quad t \in \mathbb{N}, \quad (1.3)$$

and holds information on the parameters of the model. Our estimation procedure exploits this by measuring integrated variance on daily basis and  $t$  indicates the end of a day. We later substitute integrated variance with a realized measure of volatility computed from intraday high-frequency data of  $X$ .

Note that we exploit the dynamics of integrated variance in this paper. This follows previous work of Fukasawa, Takabatake and Westphal (2019) on rough volatility, but is in contrast to the application of spot variance in, e.g., Bennedsen, Lunde and Pakkanen (2017a); Gatheral, Jaisson and Rosenbaum (2018). While spot variance is more ideal, it is associated with numerous pitfalls in practice. First, spot variance estimation requires ultra high-frequency data, which may not readily be available. Even if they are, sampling at the highest frequency may induce an accumulation of microstructure noise that can distort the analysis (e.g., Hansen and Lunde, 2006). The calculation of microstructure noise-robust estimators is complicated and they suffer from poor rates of convergence (e.g., Barndorff-Nielsen, Hansen, Lunde and Shephard, 2008; Jacod, Li, Mykland, Podolskij and Vetter, 2009; Zhang, Mykland and Aït-Sahalia, 2005). Secondly, intraday spot variance is driven by a pronounced deterministic diurnal pattern, which needs to be controlled for if the properties of the underlying stochastic process are to be uncovered (Andersen and Bollerslev, 1997, 1998b). Working at the daily frequency sidesteps this problem. Thirdly, spot variance estimators converge at a slow rate—relative to estimators of integrated variance—and, in the context of our model, often lack associated CLTs. The smoothing entailed by integrating spot variance overcomes this issue.

### 1.2.1 Properties of integrated variance

In this section, we derive some basic properties of integrated variance in the framework of the general log-normal SV model (1.1) – (1.2). This serves as the foundation for our GMM approach to estimate the parameters.

<sup>2</sup>In 2004, CBOE launched derivatives on the VIX index, which is a weighted average of implied volatility from a basket of S&P 500 options, rendering volatility at least partially tradable (see, e.g., the white paper available at <https://www.cboe.com/micro/vix/vixwhite.pdf> for an explanation of VIX products).

<sup>3</sup>In particular, the realized variance introduced below is not a consistent estimator for integrated variance in the presence of jumps. In such instances, the bipower variation of Barndorff-Nielsen and Shephard (2004) can be exploited. Below, we derive a bias correction term for bipower variation, which is practical for this purpose.

Henceforth, we denote asymptotic equivalence with  $f(\ell) \sim g(\ell)$  meaning that  $f(\ell)/g(\ell) \rightarrow 1$  as  $\ell \rightarrow \infty$ .

**Theorem 1.2.1.** *Suppose that (1.1) – (1.2) hold. Then, the integrated variance process  $(IV_t)_{t \in \mathbb{N}}$  is stationary with the following first and second-order moment structure:*

$$\mathbb{E}[IV_t] = \xi,$$

$$\mathbb{E}[IV_t IV_{t+\ell}] = \xi^2 \int_0^1 (1-y) \left[ \exp(\kappa(\ell+y)) + \exp(\kappa(|\ell-y|)) \right] dy,$$

for  $\ell \in \mathbb{N} \cup \{0\}$ . In addition, suppose the following conditions hold:

- (a)  $\lim_{\ell \rightarrow \infty} \kappa(\ell) = 0$ ,
- (b) there exists an integrable function  $\phi : [-1, 1] \rightarrow \mathbb{R}$  such that  $\frac{\kappa(\ell+y)}{\kappa(\ell)} \rightarrow \phi(y)$  as  $\ell \rightarrow \infty$  for any  $y \in [-1, 1]$ ,
- (c)  $\limsup_{\ell \rightarrow \infty} \sup_{y \in [-1, 1]} \left| \frac{\kappa(\ell+y)}{\kappa(\ell)} \right| < \infty$ .

Then, as  $\ell \rightarrow \infty$ :

$$\mathbb{E}[(IV_t - \xi)(IV_{t+\ell} - \xi)] \sim \xi^2 \kappa(\ell) \int_{-1}^1 (1-|y|) \phi(y) dy. \quad (1.4)$$

The integral describing the second-order moments of integrated variance depends on  $Y$ . In many realistic models, the integral does not possess an analytic solution and has to be either approximated or solved numerically. Moreover, while moments of higher order can be expressed in this way as well, increasing the order of integration by one makes the resulting expressions unwieldy to work with in practice. Hence, our estimation procedure relies on low-order moments.

### 1.2.1.1 Examples

**Example 1.2.2** (Fractional SV (fSV) model). *In a fSV model, the volatility process is driven by a fractional Ornstein-Uhlenbeck (fOU) process:*

$$Y_t = v \int_{-\infty}^t e^{-\lambda(t-s)} dB_s^H, \quad t \geq 0, \quad (1.5)$$

where  $v, \lambda > 0$ , and  $B^H = (B_t^H)_{t \geq 0}$  is a fractional Brownian motion (fBm) with Hurst index,  $H \in (0, 1)$ .<sup>4</sup> This model reduces to a standard log-normal SV model for  $H = 1/2$ .

---

<sup>4</sup>A fBm started at the origin ( $B_0^H = 0$ ) with Hurst exponent  $H \in (0, 1)$  can be constructed as a weighted infinite moving average of past increments to a standard Brownian motion following the representation in Mandelbrot and Van Ness (1968, Definition 2.1):  $B_t^H = \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^0 \left[ (t-s)^{H-1/2} - (-s)^{H-1/2} \right] dB_s + \int_0^t (t-s)^{H-1/2} dB_s \right\}$ , where  $\Gamma(\cdot)$  is the Gamma function. This is also termed a fractional (or Weyl) integral of  $B$ . As readily seen, the fBm reduces to a standard Brownian motion for  $H = 1/2$ .

The fractional version was introduced by Comte and Renault (1998) in a long-memory setting ( $H > 1/2$ ) and more recently in a rough setting ( $H < 1/2$ ) by Gatheral, Jaisson and Rosenbaum (2018).

Below, we describe the covariance structure of the fSV model.

**Lemma 1.2.3.** *If  $Y$  follows the fSV model, we deduce that for  $\ell \geq 0$ :*

$$\begin{aligned}\kappa(0) &= \frac{v^2}{2\lambda^{2H}} \Gamma(1 + 2H), \\ \kappa(\ell) &= \kappa(0) \cosh(\lambda\ell) - \frac{v^2 \ell^{2H}}{2} {}_1F_2\left(1; H + \frac{1}{2}, H + 1; \frac{\lambda^2 \ell^2}{4}\right), \quad \ell \geq 0,\end{aligned}\tag{1.6}$$

where  ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$  is the generalized hypergeometric function with  $p$  parameters of type 1 and  $q$  parameters of type 2.

With  $H = 1/2$ , the above covariance function reduces to

$$\kappa(\ell) = \frac{v^2}{2\lambda} e^{-\lambda\ell},\tag{1.7}$$

which is the standard formula the log-normal SV model with an exponentially decaying autocovariance function.

**Remark 1.2.4.** *In the fSV model, the second-order moment structure of integrated variance can be approximated by the following expression for  $\ell \geq 1$ :*

$$\begin{aligned}\mathbb{E}[IV_t^2] &\approx \xi^2 \exp(\kappa(0)) \left(1 - \kappa(0) + \frac{2\kappa(0)}{\lambda^2} (\cosh(\lambda) - 1) - c {}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2}{4}\right)\right), \\ \mathbb{E}[IV_t IV_{t+\ell}] &\approx \xi^2 \exp(\kappa(\ell)) \left(1 - \kappa(\ell) + \frac{2\kappa(0)}{\lambda^2} \cosh(\lambda\ell) (\cosh(\lambda) - 1) \right. \\ &\quad - \xi^2 \exp(\kappa(\ell)) \frac{c}{2} (\ell + 1)^{2H+2} {}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2 (\ell + 1)^2}{4}\right) \\ &\quad - \xi^2 \exp(\kappa(\ell)) \frac{c}{2} (\ell - 1)^{2H+2} {}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2 (\ell - 1)^2}{4}\right) \\ &\quad \left. + \xi^2 \exp(\kappa(\ell)) c \ell^{2H+2} {}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2 \ell^2}{4}\right)\right),\end{aligned}\tag{1.8}$$

where  $c = \frac{v^2}{(2H+1)(2H+2)}$ .

Note that the approximation in Remark 1.2.4 is based on the Taylor approximation,  $\exp(\kappa(\ell + y) - \kappa(\ell)) \approx 1 + \kappa(\ell + y) - \kappa(\ell)$  for  $y \in (0, 1)$ , and performs well when  $\kappa(\ell + y) - \kappa(\ell)$  is close to zero. The performance of the approximation is therefore best in areas where  $\kappa(\ell)$  is slowly decreasing. When  $H$  is very small, the correlation

function decreases fast for small  $\ell$ , and the performance of the approximation is poor. Therefore, it is important to be extremely careful with the approximation. However, it is fast to calculate compared to the integral representation in Theorem 1.2.1, and is therefore relevant for practical purposes in deliberate regions of the parameter space.

**Example 1.2.5** (Brownian semistationary (BSS) SV model). *In this model,  $Y$  is a BSS process, i.e. a Gaussian process constructed with a serial correlation that is locally equivalent to a fBm, whereas the global dependence structure can differ a lot:*

$$Y_t = v \int_{-\infty}^t h(t-s) dB_s, \quad t \geq 0, \quad (1.9)$$

where  $v > 0$  and  $h : (0, \infty) \rightarrow \mathbb{R}$  is a kernel function (subject to suitable regularity conditions). A popular choice is the gamma kernel  $h(x) = x^\alpha e^{-\lambda x}$  with  $\alpha > -1/2$  and  $\lambda > 0$  (Gamma-BSS). This model has local properties that are quite similar to the fSV model, and while not formally long-memory it does allow for substantial persistence in the process.

The precise covariance structure of the Gamma-BSS model was derived in Bensen, Lunde and Pakkanen (2017a) and given by the following result.

**Lemma 1.2.6.** *If  $Y$  follows the Gamma-BSS process of Example 1.2.5,  $\kappa(\ell)$  has the form:*

$$\begin{aligned} \kappa(0) &= \frac{v^2}{(2\lambda)^{2\alpha+1}} \Gamma(2\alpha+1), \\ \kappa(\ell) &= \frac{v^2 \Gamma(\alpha+1)}{\sqrt{\pi}} \left( \frac{\ell}{2\lambda} \right)^{\alpha+\frac{1}{2}} K_{\alpha+1/2}(\lambda\ell), \quad \ell > 0, \end{aligned} \quad (1.10)$$

where  $K_\alpha(x)$  is the Bessel function of the third kind. In addition, as  $\ell \rightarrow \infty$ , it follows that

$$\mathbb{E}[(IV_t - \xi)(IV_{t+\ell} - \xi)] \sim \frac{v^2 \xi^2 \Gamma(\alpha+1) (\exp(\lambda) - 1)^2}{2^{\alpha+1} \lambda^{\alpha+2}} \ell^\alpha \exp(-\lambda(\ell+1)). \quad (1.11)$$

### 1.3 GMM Estimation

In this section, for technical convenience we define all processes also for negative time indices.

#### 1.3.1 Assumptions and Examples

As described above, the spot variance  $\sigma^2 = (\sigma_t^2)_{t \in \mathbb{R}}$  depends on the parameter vector  $\theta = (\xi, \phi) \in \Theta$ . The true value is denoted by  $\theta_0 \in \Theta$  and is fixed. We write  $\mathbb{P}_\theta$  for the probability measure induced by  $\theta$  and  $\mathbb{E}_\theta$  is the corresponding expectation operator. Additionally, we denote by  $\mathcal{F}^\sigma$  the  $\sigma$ -algebra generated by  $\sigma^2$  or, equivalently,  $Y$ .

We now introduce our main assumption about  $Y$ .

**Assumption 1.** *The Gaussian process  $Y$  and its covariance function  $\kappa$  satisfy the following conditions:*

- (i)  *$Y$  has continuous sample paths for any  $\phi \in \Phi$ ,*
- (ii)  *$(u, \phi) \mapsto \kappa_\phi(u)$  is a continuous function.*

Condition (i) is natural for stationary Gaussian processes, since if  $Y$  was discontinuous, its sample paths would in fact be *unbounded* almost surely by a classical result of Belyaev (1961). Condition (ii) is crucial below in ensuring that the moments of the model are continuous with respect to  $\theta$ . It is worth pointing out that neither these conditions nor the stationarity of  $Y$  say much about the long-term behavior of volatility under this model. We return to this in Assumption 2 below.

**Example 1.3.1.** *In the context of the fOU process used in the fSV model, condition (i) has been shown, e.g., in Proposition 3.4 of Kaarakka and Salminen (2011), while condition (ii) follows from the continuity of the hyperbolic cosine and the hypergeometric function  ${}_1F_2$  that appears in the covariance function of the process.*

*A general result establishing condition (i) for BSS processes is derived in Bennedsen, Lunde and Pakkanen (2017b, Proposition 2.2). This also covers the Gamma-BSS process. Condition (ii) follows for the Gamma-BSS process by noting that the modified Bessel function of the second kind,  $K_\alpha$ , that appear in its covariance function is continuous.*

As the main object of interest, integrated variance, is not observable in practice, it needs to be estimated. We strive for a general framework applicable to realized measures *at large*, while still remaining analytically tractable. We postulate that we observe a noisy proxy of  $IV_t$  given by

$$\widehat{IV}_t = IV_t + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (1.12)$$

where  $\varepsilon_t$  is a random variable capturing the measurement error, which needs to adhere to a set of stylized technical conditions given in Assumption 2. We remark that such a high-level approach to describing measurement error between a realized measure and the corresponding integrated variance is similar in spirit to what Patton (2011) uses for the analysis of noisy volatility proxies in the context of forecast evaluation.

To formalize our assumptions about the process  $(\varepsilon_t)_{t \in \mathbb{Z}}$ , we require a filtration  $\mathcal{F}_t^{\sigma, \varepsilon} = \mathcal{F}_t^\sigma \vee \mathcal{F}_t^\varepsilon$ , where  $\mathcal{F}_t^\varepsilon = \sigma(\{\varepsilon_t, \varepsilon_{t-1}, \dots\})$ ,  $t \in \mathbb{Z}$ , is the  $\sigma$ -algebra generated by the errors up to time  $t$ . We also introduce a key assumption about the joint long-term behavior of  $(IV_t)_{t \in \mathbb{Z}}$  and  $(\varepsilon_t)_{t \in \mathbb{Z}}$ .

**Assumption 2.** *The processes  $(IV_t)_{t \in \mathbb{Z}}$  and  $(\varepsilon_t)_{t \in \mathbb{Z}}$  satisfy the following conditions:*

- (i)  *$(IV_t, \varepsilon_t)_{t \in \mathbb{Z}}$  is a stationary and ergodic process under  $\mathbb{P}_\theta$  for any  $\theta \in \Theta$ ,*

(ii)  $\theta \mapsto c(\theta) \equiv \mathbb{E}_\theta[\varepsilon_1^2]$  is a finite-valued, continuous function on  $\Theta$ ,

(iii)  $\mathbb{E}_\theta[\varepsilon_t \mid \mathcal{F}_{t-1}^{\sigma, \varepsilon}] = 0$  for any  $t \in \mathbb{Z}$  and any  $\theta \in \Theta$ .

Regarding condition (i), we remark that the Gaussian process  $Y$  is ergodic by the classical result of Maruyama (1949) provided

$$\kappa_\phi(u) \rightarrow 0, \quad (1.13)$$

as  $u \rightarrow \infty$ , which is evidently true for all models considered in this paper. The processes  $\sigma^2$  and  $(IV_t)_{t \in \mathbb{Z}}$  readily inherit the stationarity and ergodicity of  $Y$ . The joint ergodicity of  $(IV_t, \varepsilon_t)_{t \in \mathbb{Z}}$  is a more delicate matter, since even if  $(IV_t)_{t \in \mathbb{Z}}$  and  $(\varepsilon_t)_{t \in \mathbb{Z}}$  are ergodic on their own right and mutually independent, it does not follow that  $(IV_t, \varepsilon_t)_{t \in \mathbb{Z}}$  is ergodic (see Lindgren, 2006, Exercise 5.13). But if additionally  $(IV_t)_{t \in \mathbb{N}}$  or  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is weakly mixing, then their joint ergodicity holds (Lindgren, 2006, see Exercise 5.14). That said, in practical applications the mutual independence of  $(IV_t)_{t \in \mathbb{Z}}$  and  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is too strong an assumption, since the level of measurement error typically depends on the underlying level of volatility. Condition (iii) is a martingale-difference property for  $(\varepsilon_t)_{t \in \mathbb{Z}}$ , which implies  $\mathbb{E}_\theta[\varepsilon_t] = 0$ , i.e. the proxy  $\widehat{IV}_t$  is unbiased. This is obviously a somewhat stylized assumption, which is not exactly satisfied by many realized measures. However, we can expect it to hold approximately and, in any case, it is crucial for the analytical tractability of the setup.<sup>5</sup>

We now demonstrate that a particular structural form of the error term,  $\varepsilon_t$ , conveniently accommodates concrete realized measures as proxies in an approximate sense while satisfying Assumption 2. More specifically, let

$$\varepsilon_t = h(Z_t, (\sigma_{s+t-1}^2)_{s \in [0,1]}), \quad t \in \mathbb{Z}, \quad (1.14)$$

where  $Z_t$ ,  $t \in \mathbb{Z}$ , are i.i.d.  $d$ -dimensional random vectors, for some  $d \in \mathbb{N}$ , that are independent of  $\mathcal{F}^\sigma$  and  $h: \mathbb{R}^d \times C([0,1]) \rightarrow \mathbb{R}$  is a continuous functional such that

$$\mathbb{E}_\theta[h(Z_1, f)] = 0, \quad (1.15)$$

for any  $\theta \in \Theta$  and  $f \in C([0,1])$ . Then condition (i) in Assumption 2 can be proved using standard ergodic theory arguments, see, e.g., Lindgren (2006, Section 5.4), while condition (iii) is readily implied by (1.15). We can check condition (ii) on a case-by-case basis below by computing  $c(\theta) = \mathbb{E}_\theta[\varepsilon_1^2]$  explicitly and invoking condition (ii) of Assumption 1 to establish its continuity in  $\theta$ .

In the following examples, we construct  $\varepsilon_t$  and  $Z_t$  only for  $t \in \mathbb{N}$ , but we can extend them to negative indices by stationarity.

---

<sup>5</sup>Meddahi (2002, Section 4) studies the properties of the measurement error of realized variance under a class of log-normal volatility models including drift. He finds that the mean of the measurement error is negligible at 5-minute sampling frequency.



**Example 1.3.2** (Realized variance, CLT approximation). *Suppose that we estimate the integrated variance  $IV_t$  with the realized variance (see, e.g., Andersen and Bollerslev, 1998a; Barndorff-Nielsen and Shephard, 2002):*

$$RV_t^n = \sum_{i=1}^n \left( X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}} \right)^2, \quad (1.16)$$

for any  $t \in \mathbb{N}$ . Under standard technical conditions, the central limit theorem (CLT)

$$\sqrt{n}(RV_t^n - IV_t) \xrightarrow[n \rightarrow \infty]{d_{st}} \sqrt{2} \int_{t-1}^t \sigma_s^2 dB_s^\perp, \quad (1.17)$$

holds jointly for all  $t \in \mathbb{N}$ , where  $\xrightarrow{d_{st}}$  denotes stable convergence in distribution and  $(B_s^\perp)_{s \geq 0}$  is a Brownian motion independent of  $X$  and  $\sigma$ . Note that the limiting random variables  $\sqrt{2} \int_{t-1}^t \sigma_s^2 dB_s^\perp$ ,  $t \in \mathbb{N}$ , are conditionally independent given  $\mathcal{F}^\sigma$  with

$$\sqrt{2} \int_{t-1}^t \sigma_s^2 dB_s^\perp \Big| \mathcal{F}^\sigma \sim N(0, 2IQ_t), \quad t \in \mathbb{N}, \quad (1.18)$$

where

$$IQ_t = \int_{t-1}^t \sigma_s^4 ds \quad (1.19)$$

is the integrated quarticity. Thus the random variables

$$Z_t = \frac{\int_{t-1}^t \sigma_s^2 dB_s^\perp}{IQ_t^{1/2}} \sim N(0, 1), \quad t \in \mathbb{N}, \quad (1.20)$$

are both mutually independent and independent of  $\mathcal{F}^\sigma$ .

Informally, the CLT (1.17) says that, for any  $t \in \mathbb{N}$ ,

$$RV_t^n \stackrel{d}{\approx} IV_t + \left( \frac{2}{n} IQ_t \right)^{1/2} Z_t \quad (1.21)$$

for large  $n$ , where “ $\stackrel{d}{\approx}$ ” denotes approximate equality in distribution, as used, e.g., in Zhang et al. (2005, Section 1.2). Thus, for any  $t \in \mathbb{N}$ , the proxy  $\widehat{IV}_t = IV_t + \varepsilon_t$  with  $\varepsilon_t = \left( \frac{2}{n} IQ_t \right)^{1/2} Z_t$  approximates  $RV_t^n$  for large  $n$ . Such a proxy is analogous to what Fukasawa et al. (2019) employ in their estimation framework. We can represent the error term as  $\varepsilon_t$  in the form (1.14) using the continuous functional

$$h(z, f) = \left( \frac{2}{n} \int_0^1 f(s)^2 ds \right)^{1/2} z, \quad z \in \mathbb{R}, \quad f \in C([0, 1]). \quad (1.22)$$

Then (1.15) holds given that  $Z_1 \sim N(0, 1)$ . We can compute  $c(\theta)$  explicitly using Tonelli's theorem. The expression is reported in Table 1.1 and  $\theta \mapsto c(\theta)$  is evidently continuous under Assumption 1.

**Example 1.3.3** (Realized variance, no drift or leverage effect). *In general, the measurement error  $RV_t^n - IV_t$  is analytically hard to analyze unless we resort to asymptotic approximation with  $n \rightarrow \infty$  as in Example 1.3.2 (see also Remark 1.3.4 below). However, in a simple specific case, we can actually work with the exact error  $\varepsilon_t = RV_t^n - IV_t$ , that is  $\widehat{IV}_t = RV_t^n$ , without losing analytical tractability.*

*Namely, suppose that the log-price  $X = (X_t)_{t \geq 0}$  of the asset follows a drift-free Itô process*

$$X_t = X_0 + \int_0^t \sigma_s dW_s, \quad t \geq 0, \quad (1.23)$$

*where  $W = (W_t)_{t \geq 0}$  is a standard Brownian motion independent of  $\mathcal{F}^\sigma$ , i.e. ruling out any dependence between  $W$  and the spot variance process  $\sigma^2$ , stemming from the leverage effect for instance (e.g., Christie, 1982). Then, for any  $t \in \mathbb{N}$ ,*

$$\begin{aligned} RV_t^n - IV_t &= \sum_{i=1}^n \left( \left( \int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s dW_s \right)^2 - \int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s^2 ds \right) \\ &= \sum_{i=1}^n (Z_{t,i}^2 - 1) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma_{s+t-1}^2 ds, \end{aligned} \quad (1.24)$$

*where*

$$Z_{t,i} = \frac{\int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s dW_s}{\left( \int_{\frac{i-1}{n}+t-1}^{\frac{i}{n}+t-1} \sigma_s^2 ds \right)^{1/2}}, \quad t \in \mathbb{N}, \quad i = 1, \dots, n. \quad (1.25)$$

*Since  $W$  is independent of  $\mathcal{F}^\sigma$ , conditional on  $\mathcal{F}^\sigma$  the random variables  $Z_{t,i}$ ,  $t \in \mathbb{N}$ ,  $i = 1, \dots, n$ , are mutually independent and follow a standard normal distribution. Consequently, they are i.i.d. standard normal also unconditionally and independent of  $\mathcal{F}^\sigma$ .*

*Thanks to (1.24), we can represent the measurement error  $\varepsilon_t = RV_t^n - IV_t$  in the form (1.14) via the functional*

$$h((z_1, \dots, z_n), f) = \sum_{i=1}^n (z_i^2 - 1) \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(s) ds, \quad (z_1, \dots, z_n) \in \mathbb{R}^n, \quad f \in C([0, 1]), \quad (1.26)$$

*and i.i.d. random vectors*

$$Z_t = (Z_{t,1}, \dots, Z_{t,n}), \quad t \in \mathbb{N}, \quad (1.27)$$

*with components given by (1.25), so that  $d = n$ . The property (1.15) then holds, while an integral functional representation of  $c(\theta)$  is given in Table 1.1 and its continuity in  $\theta$  follows from the dominated convergence theorem under Assumption 1.*

**Remark 1.3.4.** *The moments of the measurement error  $\varepsilon_t = RV_t^n - IV_t$  can be analyzed under the leverage effect using Malliavin calculus and chaos expansions, see, e.g., Peccati and Taqqu (2011). However, the resulting formulae are not very convenient for numerical use, which is why we do not pursue this approach here.*

**Table 1.1:** Formulae for  $c(\theta) = \mathbb{E}_\theta[\varepsilon_1^2]$ .

Proxy	Setting	$c(\theta) = c(\xi, \phi)$
Realized variance	CLT approximation (Example 1.3.2)	$\frac{2\xi^2}{n} \exp(\kappa_\phi(0))$
	No drift or leverage (Example 1.3.3)	$\frac{4\xi^2}{n} \int_0^1 (1-y) \exp(\kappa_\phi(\frac{y}{n})) dy$
Bipower variation	CLT approximation (Example 1.3.5)	$\frac{(\frac{\pi^2}{4} + \pi - 3)\xi^2}{n} \exp(\kappa_\phi(0))$

*Note.* In the case of Example 1.3.3, we use Theorem 1.2.1 to derive the expression.

**Example 1.3.5** (Bipower variation, CLT approximation). *In the context of Example 1.3.2, the realized variance can be substituted with the bipower variation estimator of Barndorff-Nielsen and Shephard (2004), which is defined as:*

$$BV_t^n = \frac{\pi}{2} \sum_{i=2}^n |X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}}| |X_{t-1+\frac{i-1}{n}} - X_{t-1+\frac{i-2}{n}}|, \quad n \in \mathbb{N}, \quad (1.28)$$

for any  $t \in \mathbb{N}$ . Under standard technical conditions

$$\sqrt{n}(BV_t^n - IV_t) \xrightarrow[n \rightarrow \infty]{d_{st}} \sqrt{\frac{\pi^2}{4} + \pi - 3} \int_{t-1}^t \sigma_s^2 dB_s^\perp, \quad (1.29)$$

jointly for all  $t \in \mathbb{N}$ , where the structure of the limit is identical to the one in (1.17).  $BV_t^n$  is then approximated for large  $n$  by the proxy  $\widehat{IV}_t = IV_t + \varepsilon_t$  with error term  $\varepsilon_t = \left(\frac{\pi^2 + \pi - 3}{4n} IQ_t\right)^{1/2} Z_t$ , where  $Z_t$ ,  $t \in \mathbb{N}$ , are as in Example 1.3.2. Retracing the arguments in Example 1.3.2, we can then show that  $\varepsilon_t$  can be cast in the form (1.25) so Assumption 2 holds.

### 1.3.2 Consistency

Turning to the consistency result for our GMM estimator, we introduce the moment structure of the  $IV_t$  process, which is defined by:

$$g_0^{(1)}(\theta) = \mathbb{E}[IV_t(\theta)], \quad g_0^{(2)}(\theta) = \mathbb{E}[IV_t^2(\theta)], \quad g_\ell(\theta) = \mathbb{E}[IV_t(\theta)IV_{t-\ell}(\theta)], \quad \ell \in \mathbb{Z}, \quad \theta \in \Theta, \quad (1.30)$$

for a fixed  $k \in \mathbb{N}$ , which we collect in the vector

$$G(\theta) = (g_0^{(1)}(\theta), g_0^{(2)}(\theta), g_1(\theta), \dots, g_k(\theta)), \quad \theta \in \Theta. \quad (1.31)$$

We also define

$$\begin{aligned} \mathbb{V}_t &= (IV_t, IV_t^2, IV_t IV_{t-1}, \dots, IV_t IV_{t-k}), \quad t \in \mathbb{Z}, \\ \widehat{\mathbb{V}}_t &= (\widehat{IV}_t, \widehat{IV}_t^2, \widehat{IV}_t \widehat{IV}_{t-1}, \dots, \widehat{IV}_t \widehat{IV}_{t-k}), \quad t \in \mathbb{Z}, \end{aligned} \quad (1.32)$$

which by condition (i) of Assumption 2 are stationary and ergodic processes.

By conditions (ii)-(iii) of Assumption 2, we find that for any  $\theta \in \Theta$ ,  $t \in \mathbb{Z}$  and  $\ell \in \mathbb{Z}$ :

$$\begin{aligned} \mathbb{E}_\theta [\widehat{IV}_t] &= g_0^{(1)}(\theta), \\ \mathbb{E}_\theta [\widehat{IV}_t \widehat{IV}_{t-\ell}] &= \begin{cases} g_0^{(2)}(\theta) + c(\theta), & \ell = 0, \\ g_\ell(\theta), & \ell \neq 0. \end{cases} \end{aligned} \quad (1.33)$$

The expressions in (1.33) show that application of a noisy proxy  $\widehat{IV}_t$  leads to biased estimation of a single moment: the variance of integrated variance,  $g_0^{(2)}(\theta)$ . The other moments are unbiased, because the errors are mean zero and serially uncorrelated. In principle, we can thus avoid the negative impact of measurement errors by excluding  $g_0^{(2)}(\theta)$  from the selected second-order moments. More generally, however, it is often preferable to add the variance or absolute value to the moment conditions, because low-order moments are highly informative about the parameters of SV models (Andersen and Sørensen, 1996). To avoid any systematic deviance in the estimated values of the parameters, it is then necessary to correct the appropriate entries in the moment vector as detailed above (dealing with the measurement error is of course much more complicated for the absolute value than for the square).

We therefore propose to compare the sample moments of  $\widehat{IV}_t$  to a corrected moment function

$$G_c(\theta) = G(\theta) + (0, c(\theta), 0, \dots, 0), \quad \theta \in \Theta, \quad (1.34)$$

to ensure an unbiased and consistent GMM estimator. We define a random function:

$$\widehat{m}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \widehat{IV}_t - G_c(\theta), \quad \theta \in \Theta, \quad (1.35)$$

which, in view of (1.33), has

$$\mathbb{E}_{\theta_0} [\widehat{m}_T(\theta)] = G_c(\theta_0) - G_c(\theta) \quad (1.36)$$

so that

$$\mathbb{E}_{\theta_0} [\widehat{m}_T(\theta_0)] = 0. \quad (1.37)$$

Our GMM estimator is then defined as:

$$\widehat{\theta}_T = \operatorname{argmin}_{\theta \in \Theta} \widehat{m}_T(\theta)' \mathbb{W}_T \widehat{m}_T(\theta), \quad (1.38)$$

where  $\mathbb{W}_T$  is a random  $(k+2) \times (k+2)$  weight matrix.

We need additional conditions for the consistency of  $\widehat{\theta}_T$ . Firstly, we introduce a standard assumption about the limiting behavior of the weight matrix.

**Assumption 3.**  $\mathbb{W}_T = A_T' A_T$  for a random  $(k+2) \times (k+2)$  matrix  $A_T$ , which under  $\mathbb{P}_{\theta_0}$  converges almost surely to a non-random matrix  $A$  as  $T \rightarrow \infty$ .

Note that any weight matrix  $\mathbb{W}_T$ , which is a continuous function of the sample statistic

$$\frac{1}{T} \sum_{t=1}^T w(\widehat{\mathbb{V}}_t), \quad (1.39)$$

where  $w : \mathbb{R}^{k+2} \rightarrow \mathbb{R}^{d'}$  is a measurable function such that  $\mathbb{E}_{\theta_0} \left[ \|w(\widehat{\mathbb{V}}_{k+1})\|_{\mathbb{R}^{d'}} \right] < \infty$ , for some  $d' \in \mathbb{N}$ , fulfills the convergence criterion in Assumption 3 by the stationarity and ergodicity of  $(\widehat{\mathbb{V}}_t)_{t \in \mathbb{Z}}$ .

Secondly, we assume that the parameter  $\theta$  is identifiable.

**Assumption 4.**  $A(G_c(\theta) - G_c(\theta_0)) = 0$  if and only if  $\theta = \theta_0$ , where  $A$  is the limiting matrix in Assumption 3.

Assumption 4 is an identifying condition, which is equivalent to  $A\mathbb{E}_{\theta_0}[\widehat{m}_T(\theta)] = 0$  if and only if  $\theta = \theta_0$ . It is difficult to check this in practice, because there are no closed-form expressions for the moments of our model, i.e. for the components of  $G_c(\theta)$ .

**Theorem 1.3.6.** *Suppose Assumptions 1 – 4 hold. As  $T \rightarrow \infty$*

$$\widehat{\theta}_T \xrightarrow{a.s.} \theta_0. \quad (1.40)$$

In the above, our analysis assumed that the number of observations per day,  $n$ , is fixed and then relies on the noisy proxy idea. Now, following, e.g., Bollerslev and Zhou (2002); Corradi and Distaso (2006); Todorov (2009), we also cover the theory of the GMM estimator in a double asymptotic setting with  $T \rightarrow \infty$  and  $n \rightarrow \infty$ .

To this end, we denote with  $V_t^n$  some consistent realized measure of integrated variance (e.g., realized variance, bipower variation, or the pre-averaging estimator). For fixed  $k \in \mathbb{N}$ , we denote

$$\mathbb{V}_t^n = (V_t^n, (V_t^n)^2, V_t^n V_{t-1}^n, \dots, V_t^n V_{t-k}^n), \quad t \in \mathbb{Z}, \quad (1.41)$$

with associated sample moments

$$\widetilde{m}_{n,T}(\theta) = \frac{1}{T} \sum_{t=1}^T \mathbb{V}_t^n - G(\theta), \quad (1.42)$$

where we employ the moments of  $G(\theta)$  instead of the corrected version  $G_c(\theta)$ , which is of no consequence for the following result since  $n \rightarrow \infty$ .

Then,

$$\widetilde{\theta}_{n,T} = \arg \min_{\theta \in \Theta} \widetilde{m}_{n,T}(\theta)' \mathbb{W}_T \widetilde{m}_{n,T}(\theta), \quad (1.43)$$

is our GMM estimator.

In this setting, we replace Assumption 2 with the following requirement.

**Assumption 5.** *The processes  $(IV_t)_{t \in \mathbb{Z}}$  and  $(V_t^n)_{t \in \mathbb{Z}, n \in \mathbb{N}}$  admit the following:*

- (i)  $(IV_t)_{t \in \mathbb{Z}}$  is a stationary and ergodic process under  $\mathbb{P}_\theta$  for any  $\theta \in \Theta$ ,
- (ii)  $\sup_{t \in \mathbb{Z}} \mathbb{E}[(V_t^n - IV_t)^2] \rightarrow 0$  as  $n \rightarrow \infty$ .

**Theorem 1.3.7.** *Suppose Assumptions 1 and 3–5 hold. As  $T \rightarrow \infty$  and  $n \rightarrow \infty$*

$$\tilde{\theta}_{n,T} \xrightarrow{\mathbb{P}} \theta_0. \quad (1.44)$$

This result is related to Theorem 1 (and Corollary 1) in Todorov (2009) and Theorem 1 in Corradi and Distaso (2006).

**Remark 1.3.8.** *In Appendix 1.A.8, we show that under mild assumptions*

$$\sup_{t \in \mathbb{Z}} \mathbb{E}[(RV_t^n - IV_t)^2] \leq Cn^{-1}, \quad (1.45)$$

for some  $C > 0$ , hence Assumption 5 holds for  $RV_t^n$ .

### 1.3.3 Asymptotic normality

To establish asymptotic normality of our GMM estimator, for technical reasons we assume that under  $\mathbb{P}_{\theta_0}$  the Gaussian process  $Y$  admits a causal moving average representation

$$Y_t = \int_{-\infty}^t K(t-u) dB_u, \quad t \in \mathbb{R}, \quad (1.46)$$

for a two-sided standard Brownian motion  $B = (B_t)_{t \in \mathbb{R}}$  and measurable kernel function  $K : (0, \infty) \rightarrow \mathbb{R}$  such that  $\int_0^\infty K(u)^2 du < \infty$ . We can extend  $K$  to the entire real line by setting  $K(u) = 0$  for  $u \leq 0$  when necessary. (1.46) is not very restrictive, since a stationary Gaussian process admits such a representation under weak conditions. In particular, the moving average structure exists if and only if  $Y$  satisfies a mild, albeit somewhat technical, condition known as *pure non-determinism*, see Karhunen (1950, Satz 5) and Dym and McKean (1976, Section 4.5). The SV models incorporated in this paper adhere to this form. The  $\mathcal{BSS}$  model is already expressed in this way, while the fractional Ornstein-Uhlenbeck process also has such a representation (e.g., Barndorff-Nielsen and Basse-O'Connor, 2011).

In the above, it is the asymptotic behavior of  $K(u)$  as  $u \rightarrow \infty$  that governs the long-term memory of  $Y$ . To derive the asymptotic normality of our GMM estimator, we need to constrain that memory, which we do by the following:

**Assumption 6.**  $K(u) = O(u^{-\gamma})$  as  $u \rightarrow \infty$  for some  $\gamma > 1$ .

The Gamma- $\mathcal{BSS}$  model achieves Assumption 6 in the entire parameter space. Moreover, Garnier and Sølna (2018) showed that the kernel  $K(u)$  in the moving average representation of the fOU process is asymptotically, as  $u \rightarrow \infty$ , proportional to  $u^{H-3/2}$ . Thereby, the fSV model requires the restriction  $H < 1/2$  to be covered by

Assumption 6, allowing for rough volatility to be included but ruling out the long-memory version.

We believe the constraint in Assumption 6 is nearly optimal in the sense that if  $K(u)$  is asymptotically proportional to  $u^{-\gamma}$  for  $\gamma \in (0, 1)$ , e.g. with the fSV model for  $H > 1/2$ , then asymptotic normality ceases to hold. In this case, we can show that the expression for the asymptotic covariance matrix in our central limit theorem (Proposition 1.3.9) does not converge. It is possible that a non-central limit theorem with non-standard scaling holds, a common phenomenon in the realm of long-memory processes, see, e.g., Taqqu (1975). Proving such an extension is rather non-trivial, however, and beyond the scope of the present paper.

Additionally, we introduce stronger assumptions about the error process  $(\varepsilon_t)_{t \in \mathbb{Z}}$ . In what follows, we write  $\|X\|_{L^2(\mathbb{P}_\theta)} = \mathbb{E}_{\mathbb{P}_\theta} [X^2]^{1/2}$  for any square integrable random variable  $X$  and work with the filtrations  $\mathcal{F}_t^{\widehat{\mathbb{V}}} = \sigma\{\widehat{\mathbb{V}}_t, \widehat{\mathbb{V}}_{t-1}, \dots\}$ ,  $t \in \mathbb{Z}$ , and  $\mathcal{F}_t^{B, \varepsilon} = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \dots\} \vee \sigma\{B_u : u \leq t\}$ ,  $t \in \mathbb{Z}$ .

**Assumption 7.** *The processes  $B$  and  $(\varepsilon_t)_{t \in \mathbb{Z}}$  satisfy the following conditions:*

- (i)  $\mathbb{E}[\varepsilon_1^4] < \infty$ ,
- (ii)  $\left\| \mathbb{E}_{\theta_0} \left[ \varepsilon_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [\varepsilon_1^2] \right\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2})$  as  $r \rightarrow \infty$ ,
- (iii)  $B$  has independent increments with respect to  $(\mathcal{F}_t^{B, \varepsilon})_{t \in \mathbb{Z}}$  (i.e., for any  $t \in \mathbb{Z}$  the process  $(B_u - B_t)_{u \geq t}$  is independent of  $\mathcal{F}_t^{B, \varepsilon}$ ).

Condition (ii) constrains the memory in the squared measurement error. In the high-frequency setting, the measurement error usually depends on volatility (as exemplified in Example 1.3.2, 1.3.3, and 1.3.5 above). So here Assumption 6 implies condition (ii), see Proposition 1.A.6 in Appendix 1.A.9. Condition (iii) ensures that the measurement error does not anticipate future increments of the driving Brownian motion  $W$ , which is not very restrictive anyway. It is evidently true in the above examples.

The next result presents the central limit theorem for the sample mean of our statistic.

**Proposition 1.3.9.** *Suppose that Assumptions 1, 2, 6, and 7 hold. Then, as  $T \rightarrow \infty$ , under  $\mathbb{P}_{\theta_0}$ ,*

$$T^{1/2} \widehat{m}_T(\theta_0) \xrightarrow{d} N(0, \Sigma_{\widehat{\mathbb{V}}}), \quad (1.47)$$

where  $\Sigma_{\widehat{\mathbb{V}}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\widehat{\mathbb{V}}}(\ell)$  with  $\Gamma_{\widehat{\mathbb{V}}}(\ell) = \mathbb{E}_{\theta_0} [(\widehat{\mathbb{V}}_1 - G_c(\theta_0))(\widehat{\mathbb{V}}_{1+\ell} - G_c(\theta_0))']$ ,  $\ell \in \mathbb{Z}$ .

A final assumption for the CLT of our GMM estimator is presented next. Here, we introduce the function  $\mathbf{g} : \mathbb{R}^{k+2} \times \Theta \rightarrow \mathbb{R}$  via  $\mathbf{g}(x, \theta) = x - G_c(\theta)$ .

**Assumption 8.** *We have*

- (i)  $\theta_0$  is an interior point of  $\Theta$ .
- (ii)  $G'WG$  is non-singular, where  $G = \mathbb{E} \left[ \nabla_{\theta} \mathbf{g}(\widehat{\mathbb{V}}_1, \theta_0) \right]$  and  $W = A' A$ .
- (iii) The function  $\theta \mapsto \mathbf{g}(x, \theta)$  is continuously differentiable. In addition,  $\mathbb{E}[\|\mathbf{g}(\widehat{\mathbb{V}}_1, \theta_0)\|^2] < \infty$  and  $\mathbb{E}[\sup_{\theta \in \Theta} \|\nabla_{\theta} \mathbf{g}(\widehat{\mathbb{V}}_1, \theta)\|] < \infty$ .

Now, we are ready to present the asymptotic distribution of  $\widehat{\theta}_T$ .

**Theorem 1.3.10.** *Suppose Assumptions 1 – 8 hold. As  $T \rightarrow \infty$ ,*

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1} G'W\Sigma_{\widehat{\mathbb{V}}}WG(G'WG)^{-1}). \quad (1.48)$$

To finish this section, we also study the CLT of our GMM estimator in the double-asymptotic setting, where  $T \rightarrow \infty$  and  $n \rightarrow \infty$ , such that the discretization error is negligible.

**Assumption 9.** *The processes  $(IV_t)_{t \in \mathbb{Z}}$  and  $(V_t^n)_{t \in \mathbb{Z}, n \in \mathbb{N}}$  satisfy the following conditions:*

- (i)  $(IV_t)_{t \in \mathbb{Z}}$  is a stationary and ergodic process under  $\mathbb{P}_{\theta}$  for any  $\theta \in \Theta$ ,
- (ii)  $T \sup_{t \in \mathbb{Z}} \mathbb{E}[(V_t^n - IV_t)^2] \rightarrow 0$  as  $n \rightarrow \infty, T \rightarrow \infty$ .

**Theorem 1.3.11.** *Suppose Assumptions 1, 3 – 9 hold. As  $T \rightarrow \infty$  and  $n \rightarrow \infty$ ,*

$$\sqrt{T}(\tilde{\theta}_{n,T} - \theta_0) \xrightarrow{d} N(0, (\tilde{G}'W\tilde{G})^{-1} \tilde{G}'W\Sigma_{\widehat{\mathbb{V}}}W\tilde{G}(\tilde{G}'W\tilde{G})^{-1}), \quad (1.49)$$

where  $\Sigma_{\widehat{\mathbb{V}}} = \sum_{\ell=-\infty}^{\infty} \Gamma_{\widehat{\mathbb{V}}}(\ell)$  with  $\Gamma_{\widehat{\mathbb{V}}}(\ell) = \mathbb{E}_{\theta_0}[(\widehat{\mathbb{V}}_1 - G_c(\theta_0))(\widehat{\mathbb{V}}_{1+\ell} - G_c(\theta_0))']$ ,  $\ell \in \mathbb{Z}$ , and  $\tilde{G} = \mathbb{E}[\nabla_{\theta} \mathbf{g}(\widehat{\mathbb{V}}_1, \theta_0)]$ .

## 1.4 Simulation study

In the above, we developed a full-blown large sample GMM framework for estimation of the log-normal fSV model with a general Hurst index. We now review the finite sample properties of our approach. The aim is to disclose the accuracy of the procedure in a realistic setup. We inspect both the infeasible setting where estimation is based on integrated variance and a feasible implementation relying on realized variance. For the latter, we gauge the performance both with and without the quarticity correction introduced in (1.34).

We assume the log-price,  $X_t$ , evolves as a driftless diffusion:

$$dX_t = \sigma_t dW_t, \quad t \geq 0, \quad (1.50)$$

with initial condition  $X_0 \equiv 0$ . Here,  $\sigma_t$  is the spot volatility and  $W_t$  is a standard Brownian motion.



The log-variance,  $Y_t = \ln(\sigma_t^2)$ , is a fOU process:

$$dY_t = -\lambda(Y_t - \eta)dt + \nu dB_t^H, \quad (1.51)$$

where  $B_t^H$  is a fbM. We assume  $W \perp\!\!\!\perp B^H$ , so there is no leverage effect.

We draw 500 independent replications of this model with a path length of  $T = 4,000$  days as a default. In each simulation, the log-variance process is started at random from its stationary distribution,  $Y_0 \sim N(\eta, \text{var}(Y_t))$ , where  $\text{var}(Y_t) = \frac{\nu^2}{2\lambda^{2H}}\Gamma(1 + 2H)$ . To get an almost continuous-time realization of the processes and minimize the discretization bias, we partition  $[t - 1, t]$ , for  $t = 1, \dots, T$ , into  $N = 23,400$  discrete points of length  $\Delta = 1/N$ . In the US equity market, this roughly amounts to a 16 year sample of the stock price recorded every second in a 6.5 hour trading day. We then discretize  $X$  via an Euler scheme.

The SDE in (1.51) is solved to get a more convenient expression for  $Y$ :

$$Y_t = \eta + (Y_{t-\Delta} - \eta)e^{-\lambda\Delta} + \nu \int_{t-\Delta}^t e^{-\lambda(t-s)} dB_s^H. \quad (1.52)$$

The stochastic integral is approximated as  $\int_{t-\Delta}^t e^{-\lambda(t-s)} dB_s^H \simeq e^{-\lambda\Delta/2} \int_{t-\Delta}^t dB_s^H$  meaning that increments to a discretely sampled fbM are required. These can be produced in many ways to get an exact discretization, e.g. Cholesky factorization or circulant embedding (see Asmussen and Glynn, 2007). While the former has complexity  $O([TN]^3)$ , the latter entails a markedly lower budget of  $O(TN \log(TN))$  and is our preferred algorithm.

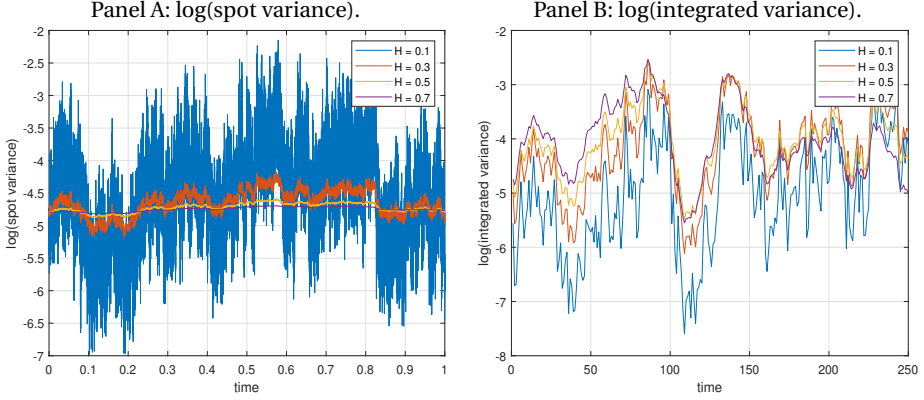
Our procedure is inspected on several distinct sets of parameters to gauge its robustness. Throughout, we set  $\eta = \ln(\xi) - 0.5\text{var}(Y_t)$ , where  $\xi = E(\sigma_t^2) = 0.0225$ . This ensures that the unconditional mean of the variance process is identical across settings and implies that the annualized standard deviation  $\sigma_t$  is about 15% on average, close to the aggregate level of volatility in the data analyzed in Section 1.5. As we are particularly attentive to estimation of the Hurst index, we choose  $H = [0.1, 0.3, 0.5, 0.7]$ , thus covering both the rough, standard and long-memory case. We calibrate  $\lambda$  and  $\nu$  by equating  $\text{std}(IV_t) = 0.05$ , which is on par with the corresponding attribute of the bipower variation of the .SPX index in the empirical part. This effectively locks in  $\nu/\lambda^H$  for each  $H$ , and as an identifying restriction we match the lag 100 autocovariance of integrated variance to the sample autocovariance of .SPX bipower variation.

The parameters are presented in Table 1.2. A realization of the spot and integrated variance processes from each model are plotted in Figure 1.1. While the pathwise properties of volatility are notably different at a microscopic scale, they are much harder to discriminate after we integrate them up to the daily horizon.

---

<sup>6</sup>The ploy is as always to use Itô's Lemma with the integrating factor  $e^{\lambda t} Y_t$ . The math is a bit more involved here though, since we are dealing with a fractional Brownian motion, where a stochastic calculus may not exist. Nevertheless, it goes through in this particular instance, see, e.g., Cheridito, Kawaguchi and Maejima (2003).

**Figure 1.1:** Sample path of spot and integrated variance.



*Note.* In Panel A, we simulate a sample path of the log-spot variance for a single day as a function of  $H$ . In Panel B, we show the associated integrated variance dynamics over 250 trading days.

In addition to integrated variance we also collect realized variance with  $n = 78$ , i.e. with 5-minute data. The advantage of this choice is that there is no concern about microstructure noise at this sampling frequency in practice. The input to the optimizer is therefore either  $(IV_t)_{t=1}^T$  or  $(RV_t^n)_{t=1}^T$ . We restrict the description of the implementation details below to the feasible setting with realized variance.

The unknown parameter vector is  $\theta_0 = (\xi, \lambda, \nu, H)$ , which we estimate with the non-gradient-based Nelder-Mead algorithm available via `fminsearch` in MatLab. That function does not accept boundary conditions, so we reparameterize the model by log-transforming  $\nu$  and  $\lambda$ , while  $H$  is bounded by the logistic function. We launch the engine at initial values determined as follows:  $\xi$  is started at the unconditional mean of realized variance, i.e.  $\bar{RV} = T^{-1} \sum_{t=1}^T RV_t^n$ . To set  $H$  and  $\nu$  we exploit the auxiliary two-stage procedure proposed in Gatheral, Jaisson and Rosenbaum (2018), which relies on the scaling law:

$$\gamma_h \equiv \mathbb{E}[|Y_{t+h} - Y_t|^q] \rightarrow K_q \nu^q |h|^{qH}, \quad (1.53)$$

as  $h \rightarrow 0$ , where  $K_q = 2^{q/2} \frac{\Gamma(\frac{q+1}{2})}{\sqrt{\pi}}$  is the  $q$ 'th moment of the absolute value of a standard normal random variable. This entails a log-linear relationship between  $\gamma_h$  and  $|h|$ :  $\ln(\gamma_h) = \ln(K_q \nu^q) + qH \ln(|h|)$ . We employ  $RV_t^n$  as a proxy for the instantaneous variance and substitute the left-hand side of (1.53) by the sample mean:

$$\hat{\gamma}_h = \frac{1}{T-m} \sum_{t=1}^{T-m} |\ln(RV_{t+h}^n) - \ln(RV_t^n)|^q, \quad (1.54)$$

for  $h = 1, \dots, m$ .  $H$  and  $\nu$  are then estimated by OLS with  $q = 2$  and  $m = 6$ . The results are rather robust against this configuration. At last,  $\lambda$  is pre-estimated such that the

theoretical variance of  $Y_t$  equals the sample variance of  $\ln(RV_t^n)$ .

As shown in Table 1.2, the initial values display very low variation between replications, but they are often highly biased. For instance, using  $IV_t$  the starting point of  $H$  increases with the true value, but as expected it is too high on average, whereas for  $RV_t^n$  it is largely unaffected by the actual roughness of the model.

As such, there is still a lot of work left for the GMM procedure. We match the  $\ell$ 'th sample autocovariance of  $RV_t^n$  with the second-order moment structure of  $IV_t$  with  $\ell = [0, 1, 2, 5, 10, 20, 50]$ .<sup>7</sup> In the region where  $H > 0.3$ , the approximate second-order moment structure —available in closed-form from Remark 1.2.4— is performing indistinguishable to the numerical integration of the exact moments from Theorem 1.2.1. Therefore, when the algorithm searches for the optimal set of parameters in this region, it compares sample moments to its analytical approximate counterpart to speed up the procedure. For  $H \leq 0.3$ , we rely on numerical integration of the exact moments presented in 1.2.1.<sup>8</sup> Likewise, the quarticity correction is implemented using its parametric form given in Table 1.1. Finally, the weight matrix,  $W_T$ , is a data-driven Newey and West (1987) HAC-type estimator computed with a Bartlett kernel and a lag length equal to  $[T^{1/3}]$ .

The results are presented in Table 1.2. We report the mean parameter estimate (standard error in parenthesis) both for the initial and final value, where all calculations are done across simulations. The left-hand side shows the outcome based on integrated variance, whereas the right-hand side is for realized variance with and without the correction in (1.34). Several interesting findings emerge. Firstly, the setting with integrated variance leads to parameter estimates that are generally close to their population counterparts across models, thus verifying the robustness and accuracy of our procedure. Secondly, we record a significant deterioration in the estimation of  $H$  for realized variance without the quarticity adjustment. As explained, realized variance is a noisy proxy for integrated variance and this translates to additional roughness in the  $(RV_t^n)_{t=1}^T$  process. In that case, we further note  $v$  increases, while  $\lambda$  decreases in order to “compensate” for the spurious loss of memory, even as  $H$  is decreasing. Including the bias correction leads to a huge improvement in the estimates that are surprisingly accurate also for small  $H$ . While a marginal downward bias is retained, the typical estimate is nevertheless within about a standard error of the true value and very much in line with the infeasible results recovered using the benchmark integrated variance.

---

<sup>7</sup>In a robustness check, and to better capture the persistence of log-variance with  $H = 0.7$ , we also attempted to include lag 100 and 200. However, the results did not move much. This is consistent with Andersen and Sørensen (1996), who note that estimation of SV models does not always improve by adding more information.

<sup>8</sup>The threshold,  $H = 0.3$ , causes no sign of estimates clustered in the threshold supporting that the difference between the approximation and numerical integration is negligible here.

**Table 1.2:** Parameter estimation of the fSV model in simulated data.

parameter	value	integrated variance		realized variance		
		initial value	sim. average	initial value	sim. average no correction	sim. average correction
Panel A:						
$\xi$	0.0225	0.0268 (0.0246)	0.0251 (0.0222)	0.0268 (0.0244)	0.0249 (0.0213)	0.0251 (0.0213)
$\lambda$	0.0020	0.0189 (0.0066)	0.0029 (0.0029)	0.0155 (0.0059)	0.0025 (0.0104)	0.0034 (0.0036)
$\nu$	1.2080	0.5963 (0.0076)	1.3686 (0.3147)	0.6499 (0.0084)	2.1987 (1.1309)	1.1734 (0.1588)
$H$	0.1000	0.3005 (0.0102)	0.0817 (0.0440)	0.2680 (0.0098)	0.0492 (0.0394)	0.0969 (0.0330)
Panel B:						
$\xi$	0.0225	0.0222 (0.0067)	0.0212 (0.0061)	0.0223 (0.0067)	0.0206 (0.0056)	0.0211 (0.0061)
$\lambda$	0.0062	0.0187 (0.0041)	0.0074 (0.0043)	0.0107 (0.0029)	0.0035 (0.0028)	0.0075 (0.0048)
$\nu$	0.4395	0.3183 (0.0040)	0.4541 (0.0485)	0.3907 (0.0048)	0.6729 (0.1183)	0.4505 (0.0640)
$H$	0.3000	0.4420 (0.0108)	0.2706 (0.0474)	0.3493 (0.0110)	0.1708 (0.0461)	0.2726 (0.0540)
Panel C:						
$\xi$	0.0225	0.0229 (0.0081)	0.0220 (0.0075)	0.0229 (0.0081)	0.0201 (0.0069)	0.0218 (0.0074)
$\lambda$	0.0126	0.0205 (0.0038)	0.0159 (0.0085)	0.0051 (0.0015)	0.0032 (0.0034)	0.0168 (0.0124)
$\nu$	0.2116	0.1761 (0.0022)	0.2168 (0.0294)	0.2814 (0.0033)	0.4665 (0.0948)	0.2178 (0.0369)
$H$	0.5000	0.5867 (0.0103)	0.4800 (0.0709)	0.3594 (0.0114)	0.2166 (0.0754)	0.4813 (0.0895)
Panel D:						
$\xi$	0.0225	0.0223 (0.0101)	0.0213 (0.0097)	0.0223 (0.0101)	0.0189 (0.0085)	0.0212 (0.0096)
$\lambda$	0.0300	0.0300 (0.0041)	0.0381 (0.0203)	0.0046 (0.0013)	0.0041 (0.0060)	0.0355 (0.0205)
$\nu$	0.1455	0.1268 (0.0017)	0.1614 (0.0258)	0.2503 (0.0029)	0.4244 (0.1470)	0.1694 (0.0339)
$H$	0.7000	0.7123 (0.0096)	0.6577 (0.0912)	0.3683 (0.0120)	0.2356 (0.1074)	0.6321 (0.0977)

*Note.* We simulate 500 replications of a fractional Ornstein-Uhlenbeck process  $dt_f = -\lambda(Y_t - \eta)dt + \nu d\tilde{B}_t^H$  on  $[0, T]$  with  $T = 4,000$  and a discretization step of  $\Delta = 1/23,400$ . The true model parameters  $\theta_0 = (\xi, \lambda, \nu, H)$  appear in Panel A – D, where  $\xi = e^{(T+0.5)\text{var}(Y)}$ . We estimate  $\theta_0$  with the GMM procedure developed in the main text, where the theoretical mean and autocovariance (at lag 0, 1, 2, 5, 10, 20, and 50) of integrated variance is matched with the sample. The optimizer is launched with initial values from the two-stage procedure in Gatheral, Jaisson and Rosenbaum (2018). We report the average of the initial values and the associated parameters estimates, based on integrated variance (left) and realized variance (right). The latter is computed both without and with the correction in (1.34). Standard deviation across simulations appear in parenthesis.

## 1.5 Empirical application

The log-normal fSV model is estimated from empirical high-frequency data covering a comprehensive selection of asset return series. We downloaded version 0.3 of the Oxford-Man Institute's "realized library" via: <https://realized.oxford-man.ox.ac.uk/>. The website tracks thirty-one leading stock indexes covering major financial markets. At the end of each trading day, the library is refreshed with information from Thomson Reuters DataScope Tick History and several non-parametric volatility estimators are calculated and appended to the database. We here employ the daily bipower variation defined in (1.28). This is a consistent and jump-robust measure of integrated variance with  $\text{avar}[\sqrt{n}(BV_t^n - IV_t)] = 2.6\mathbb{E}[\int_{t-1}^t \sigma_s^4 ds]$ . In line with our comments above, a 5-minute sampling frequency corresponding to  $n = 78$  has been set to suppress microstructure noise.

An overview of the data is presented in Table 1.3. It reports the starting date of each index and the sample size. We include information up to 31 July 2019 and exclude KSE og STI from our investigation, as there are sizable gaps in their data series.

The GMM estimation of the model parameters follows the setup from the simulation section. The right-hand side of Table 1.3 shows the outcome for individual stock indexes, where the bottom row presents the cross-sectional average of each parameter. The  $\bar{\xi} = 0.0214$  estimate corresponds to about 14.49% annualized volatility in the aggregate stock market.

The reported Hurst exponents suggest a very rough volatility process with an average level of  $\bar{H} = 0.035$ . This is on par with Bayer, Friz and Gatheral (2016) and Fukasawa, Takabatake and Westphal (2019) but slightly smaller compared to Bennesen, Lunde and Pakkanen (2017a) and Gatheral, Jaisson and Rosenbaum (2018). A possible reason for this discrepancy is that the latter employ realized variance as a proxy for spot variance. However, the former is a consistent estimator of the integrated variance, which is much smoother than instantaneous variance (see Figure 1.1). This ought to bias their  $H$  estimates to the upside. We do not suffer from that problem here, as we directly compare bipower variation to the dynamics of integrated variance in the fSV model, so the averaging "cancels out."

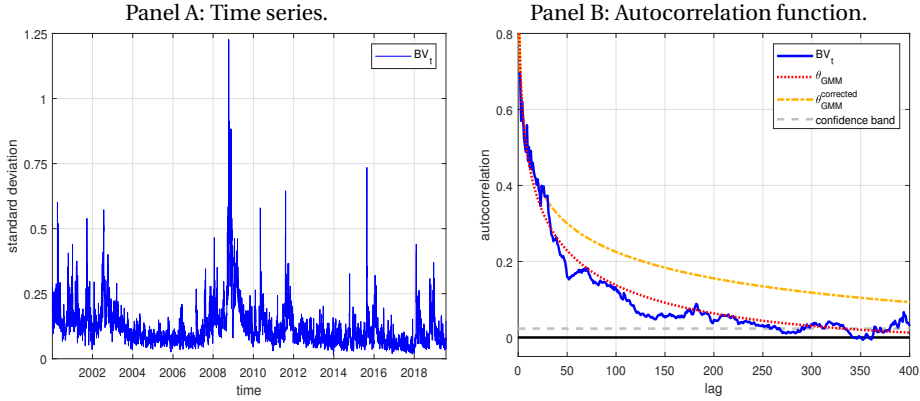
Looking at the table, the results are remarkably stable across assets. We do observe a minor deviation for the Finnish OMXHPI index. On manual inspection of the data, we found that its sample autocorrelation function (acf) differs notably from the other stock indexes. This can possibly be explained by the fact that its evolution has for a large portion of our sample been dominated by a single stock, following the rise and fall of Nokia.

To gauge the statistical fit of the model, we calculate a  $\mathcal{J}$ -test for overidentifying restrictions. The test statistic has an asymptotic  $\chi^2$ -distribution with four degrees of freedom under  $\mathcal{H}_0$ . The results appear in the last column of Table 1.3. Overall, the  $P$ -values are relatively high, so the fSV process does a good job in describing the data.

**Table 1.3:** Parameter estimation of the log-normal ISV model from stock index return data.

code	index	location	starting date	sample size	parameter estimate				
					$\xi$	$\lambda \times 100$	$v$	$H$	$\beta$ -test
AEX	AEX	Netherlands	2000-01	4,992	0.025	0.046	1.545	0.046	0.929
AORD	All Ordinaries	Australia	2000-01	4,942	0.010	0.031	2.064	0.025	0.984
BEX	BEL 20	Belgium	2000-01	4,989	0.019	0.119	1.708	0.039	0.934
BSESN	BSE Sensex	India	2000-01	4,858	0.031	0.020	1.954	0.026	0.699
BVLG	PSI All-Share	Portugal	2012-10	1,731	0.010	0.019	1.898	0.015	0.944
BVSP	Bovespa	Brazil	2000-01	4,822	0.031	0.118	1.789	0.030	0.907
DJI	Dow Jones Industrial Average	USA	2000-01	4,909	0.018	0.043	1.986	0.032	0.745
FCHI	CAC 40	France	2000-01	4,991	0.030	0.075	1.796	0.037	0.989
FTMIB	FTSE MIB	Italy	2009-06	2,582	0.025	0.019	1.969	0.016	0.895
FTSE	FTSE 100	United Kingdom	2000-01	4,937	0.020	0.033	2.482	0.021	0.885
GDAXI	DAX	Germany	2000-01	4,966	0.034	0.052	1.679	0.041	0.906
.GSPTSE	TSX Composite	Canada	2002-05	4,316	0.011	0.019	1.828	0.042	0.821
.HSI	Hang Seng	Hong Kong	2000-01	4,794	0.021	0.041	1.875	0.030	0.968
IBEX	IBEX 35	Spain	2000-01	4,959	0.031	0.060	1.904	0.026	0.941
IXIC	Nasdaq 100	USA	2000-01	4,907	0.028	0.027	1.901	0.033	0.973
KS11	KOSPI	South Korea	2000-01	4,818	0.027	0.064	1.442	0.053	0.676
MXX	IPC Mexico	Mexico	2000-01	4,912	0.015	0.029	2.215	0.017	0.871
N225	Nikkei 225	Japan	2000-02	4,765	0.021	0.100	1.910	0.031	0.820
NSEI	Nifty 50	India	2000-01	4,855	0.024	0.004	2.078	0.020	0.757
OMXC20	OMXC20	Denmark	2005-10	3,438	0.021	0.090	1.223	0.065	0.881
OMXHPI	OMX Helsinki	Finland	2005-10	3,470	0.021	0.147	0.511	0.186	0.872
OMXSPI	OMX Stockholm	Sweden	2005-10	3,470	0.017	0.018	1.836	0.030	0.789
OSEAX	Oslo Exchange	Norway	2001-09	4,465	0.019	0.025	1.826	0.032	0.867
RUT	Russel 2000	USA	2000-01	4,907	0.014	0.072	1.980	0.035	0.748
SMSI	Madrid General	Spain	2005-07	3,589	0.023	0.140	2.005	0.029	0.962
SPX	S&P 500	USA	2000-01	4,913	0.019	0.055	1.579	0.048	0.779
SSEC	Shanghai Composite	China	2000-01	4,726	0.036	0.009	1.833	0.025	0.777
SSMI	Swiss Market Index	Switzerland	2000-01	4,906	0.018	0.046	2.146	0.027	0.898
STOXX50E	EURO STOXX 50	Europe	2000-01	4,991	0.031	0.060	1.792	0.036	0.954
Average					0.021	0.051	1.702	0.035	

Note: "code" is based on the Oxford-Man Institutes naming convention.  $\xi$  is the average level of the spot variance process,  $\lambda$  is the speed of mean reversion (multiplied by 100),  $v$  is the volatility-of-volatility, while  $H$  is the Hurst exponent. The cross-sectional average of the estimated parameter vector is reported in the bottom row.  $\beta$ -test is the  $p$ -value of the Sargan-Hansen test of overidentifying restrictions, which is asymptotically  $\chi^2$ -distributed with four degrees of freedom.

**Figure 1.2:** Properties of .SPX bipower variation.

*Note.* In Panel A, we plot the bipower variation of .SPX converted to standard deviation per annum. In Panel B, we show the sample acf of bipower variation with a 95% white noise confidence band. We compare this to the theoretical acf of the log-normal fSV model implied by the estimated parameter vector  $\hat{\theta}_{\text{GMM}}$ , where the latter is reported with and without bias correction from (1.34).

As an illustration of our findings we zoom in at SPX, which represents the S&P 500 index and is therefore related to developments in the US stock market. In Panel A of Figure 1.2, we show the bipower variation of .SPX (the raw estimator has been converted to standard deviation per annum for convenience). It displays the customary volatility clustering present in most financial asset return series. In Panel B, we plot the first 400 lags of the associated acf of  $BV_t^n$  together with a Bartlett one-sided 95% white noise confidence band. The slow decay is consistent with significant memory in integrated variance. As a comparison, we superimpose the model-implied acf recovered from the GMM parameter estimation, where the latter is shown both with and without the bias correction in (1.34). The acf of the uncorrected estimator tracks the sample counterpart based on bipower variation closely both at the short and long end. Meanwhile, the effect of the bias correction is to lift the acf higher, indicating a larger amount of memory in integrated variance. Note that this was to be expected, since the impact of measurement error in a time series is to attenuate the acf (e.g., Hansen and Lunde, 2014). Meanwhile, it retains the impression of an exponential decline, which suggests that short-memory components are driving the serial correlation, in contrast to a hyperbolic decay symptomatic of true long-memory, where the latter entails a Hurst exponent above 0.5.<sup>9</sup>

In sum, our empirical results point toward a very erratic volatility process in line

<sup>9</sup>We also estimated the fSV model with a driving standard Brownian motion, i.e. pre-imposing  $H = 0.5$ . The remaining parameters were  $(\hat{\xi}, \hat{\lambda}, \hat{\nu}) = (0.016, 0.612, 1.048)$ , which broadly aligns with previous studies, e.g., Tegnér and Poulsen (2018). Intuitively, to fit the sample acf of bipower variation the GMM procedure has to select a larger mean-reversion parameter  $\lambda$  to compensate for the extra memory induced by forcing  $H$  to one-half.

with—or even exceeding—previous research. As these findings are not induced by microstructure noise nor discretization error, we are bound to conclude there is roughness in integrated variance.

## 1.6 Conclusion

We propose a GMM framework for estimation of the log-normal stochastic volatility model governed by a general fractional Brownian motion. Our procedure is built from the dynamic properties of integrated variance in this model, but it employs a generic realized measure of volatility computed from high-frequency as a noisy proxy. We explicitly account for the inherent measurement error in the selected estimator by adjusting an appropriate moment condition. We prove consistency and asymptotic normality of our estimator in a classical long-span setting. A Monte Carlo study shows our proposed routine is capable of recovering the parameters of the model across the entire memory spectrum. We implement the approach on a vast array of high-frequency data from leading equity market indexes and confirm the presence of substantial roughness in the stochastic variance process, as consistent with recent findings in the literature.

In future work, we envision our theoretical results can be extended to other classes of fSV models. The Heston (1993) model, for example, has been studied in both rough and long-memory form (e.g., Comte, Coutin and Renault, 2012; Guennoun, Jacquier, Roome and Shi, 2018). With its affine structure and appeal to option pricing (Duffie, Pan and Singleton, 2000; Euch and Rosenbaum, 2018), it should be interesting to adapt the GMM estimation routine outlined in this paper to the generalized fractional version of that model. We leave that for another endeavour.

## 1.A Proofs

### 1.A.1 Auxiliary Result

To prove Theorem 1.2.1, we need the following auxiliary result that enables to express certain two-dimensional integrals in a one-dimensional form.

**Lemma 1.A.1.** *Assume  $f : [0, \infty) \rightarrow \mathbb{R}$  is a continuous function and let  $k \in \mathbb{N}$ . Then,*

$$\int_{k-1}^k \int_0^1 f(|s-t|) ds dt = \int_0^1 (1-y) (f(|k-1-y|) + f(k-1+y)) dy.$$

*Proof.* Write

$$\int_{k-1}^k \int_0^1 f(|s-t|) ds dt = \iint_{[k-1, k] \times [0, 1]} f(|s-t|) ds dt$$



and introduce the linear (bijective) change of variables:

$$\begin{bmatrix} s \\ t \end{bmatrix} = \frac{1}{2} \begin{bmatrix} (u+v) \\ (-u+v) \end{bmatrix} = \underbrace{\frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\equiv A} \begin{bmatrix} u \\ v \end{bmatrix} \equiv \begin{bmatrix} \varphi_1(u, v) \\ \varphi_2(u, v) \end{bmatrix} \equiv \varphi(u, v).$$

Applying this to the inequalities  $k-1 \leq s \leq k$  and  $0 \leq t \leq 1$ , we find they are equivalent to

$$v \leq 2k - u, \quad v \geq 2(k-1) - u, \quad v \geq u, \quad \text{and} \quad v \leq u + 2.$$

Therefore, the set

$$B = \{(u, v) \in \mathbb{R}^2 : v \leq 2k - u, \quad v \geq 2(k-1) - u, \quad v \geq u, \quad \text{and} \quad v \leq u + 2\}$$

is mapped by  $\varphi$  to  $[k-1, k] \times [0, 1]$ . Note also that  $B = B_1 \cup B_2$ , where

$$B_1 \equiv \{(u, v) \in \mathbb{R}^2 : k-2 \leq u < k-1 \quad \text{and} \quad 2(k-1) - u \leq v \leq u + 2\}$$

$$B_2 \equiv \{(u, v) \in \mathbb{R}^2 : k-1 \leq u \leq k \quad \text{and} \quad u \leq v \leq 2k - u\}$$

are disjoint. Now, the Jacobian  $(D\varphi)(u, v)$  of  $\varphi$  equals  $A$  for any  $(u, v) \in \mathbb{R}^2$ , whereby

$$|\det(D\varphi)(u, v)| = |\det(A)| = \frac{1}{2},$$

and since  $s - t = \varphi_1(u, v) - \varphi_2(u, v) = \frac{1}{2}(u + v) - \frac{1}{2}(-u + v) = u$ , we get by multivariate integration by substitution:

$$\begin{aligned} \iint_{[k-1, k] \times [0, 1]} f(|s - t|) \, ds \, dt &= \iint_{\varphi(B)} f(|s - t|) \, ds \, dt \\ &= \iint_B f(|\varphi_1(u, v) - \varphi_2(u, v)|) |\det(D\varphi)(u, v)| \, du \, dv \\ &= \frac{1}{2} \left( \iint_{B_1} f(|u|) \, du \, dv + \iint_{B_2} f(|u|) \, du \, dv \right). \end{aligned}$$

Firstly,

$$\begin{aligned} \iint_{B_1} f(|u|) \, du \, dv &= \int_{k-2}^{k-1} \left( \int_{2(k-1)-u}^{u+2} f(|u|) \, dv \right) du = 2 \int_{k-2}^{k-1} (u - (k-2)) f(|u|) \, du \\ &= 2 \int_0^1 (1 - y) f(|k-1-y|) \, dy, \end{aligned}$$

via the substitution  $y = k-1-u$ . Secondly,

$$\begin{aligned} \iint_{B_2} f(|u|) \, du \, dv &= \int_{k-1}^k \left( \int_u^{2k-u} f(|u|) \, dv \right) du = 2 \int_{k-1}^k (k-u) f(u) \, du \\ &= 2 \int_0^1 (1-y) f(k-1+y) \, dy, \end{aligned}$$

by substituting  $y = u - (k - 1)$  and noting  $u \geq k - 1 \geq 0$ . Thus, the asserted formula follows.  $\square$   $\square$

### 1.A.2 Proof of Theorem 1.2.1

To prove the first part of the theorem, we note that since the variance process  $(\sigma_t^2)_{t \geq 0}$  is stationary, Fubini's Theorem yields that

$$\mathbb{E}[IV_t] = \int_{t-1}^t \mathbb{E}[\sigma_s^2] ds = \mathbb{E}[\sigma_0^2] = \xi.$$

We proceed with the second-order moments of  $IV_t$  by noting that

$$\begin{aligned} \mathbb{E}[\sigma_t^2 \sigma_s^2] &= \xi^2 \mathbb{E}[\exp(Y_t + Y_s - \kappa(0))] \\ &= \xi^2 \exp(\kappa(|t - s|)), \end{aligned}$$

where the last equation follows from  $Y_t + Y_s \sim N(0, 2\kappa(|t - s|) + 2\kappa(0))$ . We deduce that

$$\begin{aligned} \mathbb{E}[IV_1 IV_{1+\ell}] &= \int_{\ell}^{\ell+1} \int_0^1 \mathbb{E}[\sigma_s^2 \sigma_t^2] ds dt \\ &= \xi^2 \int_{\ell}^{\ell+1} \int_0^1 \exp(\kappa(|t - s|)) ds dt \\ &= \xi^2 \int_0^1 (1 - y) [\exp(\kappa(|\ell - y|)) + \exp(\kappa(\ell + y))] dy, \end{aligned}$$

as a consequence of Lemma 1.A.1.

As for the second part of Theorem 1.2.1, note that from condition (a) there exists  $\ell_0 > 0$  such that  $|\kappa(u)| \leq 1$  for any  $u \geq \ell_0 - 1$ . Denoting  $\gamma_{\ell+1,1} = \mathbb{E}[IV_t IV_{t+\ell}] - \xi^2$ , we thus find that

$$\gamma_{\ell+1,1} = \xi^2 \int_0^1 (1 - y) (\exp(\kappa(|\ell - y|)) - 1 + \exp(\kappa(\ell + y)) - 1) dy.$$

Introducing  $r(x) \equiv \exp(x) - 1 - x$ ,  $x \in \mathbb{R}$  allows to further write

$$\frac{\gamma_{\ell+1,1}}{\xi^2 \kappa(\ell)} = \underbrace{\int_0^1 (1 - y) \left( \frac{\kappa(\ell - y)}{\kappa(\ell)} + \frac{\kappa(\ell + y)}{\kappa(\ell)} \right) dy}_{\equiv I_1} + \underbrace{\int_0^1 (1 - y) \left( \frac{r(\kappa(\ell - y))}{\kappa(\ell)} + \frac{r(\kappa(\ell + y))}{\kappa(\ell)} \right) dy}_{\equiv I_2},$$

for any  $\ell \geq \ell_0$ .

As  $|r(x)| \leq 3x^2$ ,  $x \in [0, 1]$ , it follows that

$$|I_2| \leq 3 \sup_{y \in [-1, 1]} \left| \frac{\kappa(\ell + y)}{\kappa(\ell)} \right| \int_0^1 \underbrace{(1 - y) (|\kappa(|\ell - y|)| + |\kappa(\ell + y)|)}_{\equiv \nu_{\ell}(y)} dy,$$

where for any  $y \in [0, 1] : 0 \leq v_\ell(y) \leq 1$ , while  $v_\ell(y) \rightarrow 0$ , as  $\ell \rightarrow \infty$  by (1.2.1). Applying the dominated convergence theorem and condition (c) implies that:

$$\limsup_{\ell \rightarrow \infty} |I_2| \leq 3 \limsup_{\ell \rightarrow \infty} \sup_{\bar{y} \in [-1, 1]} \left| \frac{\kappa(\ell + \bar{y})}{\kappa(\ell)} \right| \lim_{\ell \rightarrow \infty} \int_0^1 v_\ell(y) dy = 0.$$

Finally, for  $y \in [0, 1]$  the integrand  $u_\ell(y) \equiv (1 - y) \left( \frac{\kappa(\ell - y)}{\kappa(\ell)} + \frac{\kappa(\ell + y)}{\kappa(\ell)} \right)$  in  $I_1$  is bounded uniformly in  $\ell$  by some constant from condition (c), while

$$\lim_{\ell \rightarrow \infty} u_\ell(y) = (1 - y)(\phi(-y) + \phi(y)), \quad y \in [0, 1],$$

by condition (b). Thus, by dominated convergence

$$\lim_{\ell \rightarrow \infty} I_1 = \int_0^1 (1 - y)(\phi(-y) + \phi(y)) dy = \int_{-1}^1 (1 - |y|)\phi(y) dy,$$

which concludes the proof.  $\square$

### 1.A.3 Proof of Lemma 1.2.3

In view of Remark 2.4 in Cheridito, Kawaguchi and Maejima (2003), for any  $\ell \geq 0$

$$\kappa(\ell) = \frac{v^2 \Gamma(2H + 1) \sin(\pi H)}{2\pi} \int_{-\infty}^{\infty} e^{i\ell z} \frac{|z|^{1-2H}}{\lambda^2 + z^2} dz.$$

Appealing to Euler's formula  $e^{ix} = \cos(x) + i \sin(x)$  and the cosine (sine) function being even (odd), we further conclude that:

$$\begin{aligned} \kappa(\ell) &= \frac{v^2 \Gamma(2H + 1) \sin(\pi H)}{\pi} \int_0^{\infty} \cos(\ell z) \frac{z^{1-2H}}{\lambda^2 + z^2} dz \\ &= \frac{v^2 \Gamma(2H + 1) \sin(\pi H)}{\pi} \int_0^{\infty} \cos(\ell \lambda x) \frac{\lambda^{1-2H} x^{1-2H}}{\lambda^2 (1 + x^2)} \lambda dx \\ &= \frac{v^2 \Gamma(2H + 1) \sin(\pi H)}{\lambda^{2H} \pi} \int_0^{\infty} \cos(\ell \lambda x) \frac{x^{1-2H}}{1 + x^2} dx. \end{aligned}$$

To compute the variance term ( $\ell = 0$ ), this expression is evaluated using Gradshteyn and Ryzhik (2007), formula 8.380 (3), 8.384 (1) and 8.334 (3):

$$\begin{aligned} \kappa(0) &= \frac{v^2 \Gamma(2H + 1) \sin(\pi H)}{\lambda^{2H} \pi} \int_0^{\infty} \frac{x^{1-2H}}{1 + x^2} dx \\ &= \frac{v^2 \Gamma(2H + 1) \sin(\pi H)}{2\lambda^{2H} \pi} B(1 - H, H) \\ &= \frac{v^2}{2\lambda^{2H}} \Gamma(2H + 1), \end{aligned}$$

where  $B(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1} dt$  is the Beta function.

Now, we proceed to compute the covariance, i.e.  $\ell > 0$ . In view of Gradshteyn and Ryzhik (2007), formula 3.773 (4) [with  $\mu = 1$  and  $\nu = 1/2 - H$ ], this reduces to:

$$\begin{aligned} \int_0^\infty \cos(\ell \lambda x) \frac{x^{1-2H}}{1+x^2} dx &= \frac{1}{2} B(1-H, H) {}_1F_2\left(1-H; 1-H, \frac{1}{2}; \frac{\ell^2 \lambda^2}{4}\right) \\ &\quad + \frac{\sqrt{\pi}(\ell \lambda)^{2H}}{4^{H+1/2}} \frac{\Gamma(-H)}{\Gamma(H+\frac{1}{2})} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\ell^2 \lambda^2}{4}\right) \\ &= \frac{\pi \cosh(\ell \lambda)}{2 \sin(\pi H)} - \frac{\sqrt{\pi}(\ell \lambda)^{2H}}{2^{2H+1} H} \frac{\Gamma(1-H)}{\Gamma(H+\frac{1}{2})} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\ell^2 \lambda^2}{4}\right) \\ &= \frac{\pi \cosh(\ell \lambda)}{2 \sin(\pi H)} - \frac{(\ell \lambda)^{2H}}{2 \Gamma(2H+1)} \frac{\pi}{\sin(\pi H)} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\ell^2 \lambda^2}{4}\right), \end{aligned}$$

where we used the identity  $\Gamma(z)\Gamma(1-z) = \pi / \sin(\pi z)$  and the duplication formula  $\Gamma(z)\Gamma(z+1/2) = 2^{1-2z} \sqrt{\pi} \Gamma(2z)$ . This yields

$$\begin{aligned} \kappa(\ell) &= \frac{v^2 \Gamma(2H+1) \sin(\pi H)}{\lambda^{2H} \pi} \int_0^\infty \cos(\ell \lambda x) \frac{x^{1-2H}}{1+x^2} dx \\ &= \frac{v^2 \Gamma(2H+1) \cosh(\ell \lambda)}{2 \lambda^{2H}} - \frac{v^2 \ell^{2H}}{2} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\ell^2 \lambda^2}{4}\right), \end{aligned}$$

and we are done.  $\square$

#### 1.A.4 Derivation of Remark 1.2.4

We start with the variance. By a standard linear Taylor approximation  $\exp(x) \approx 1 + x$  and setting  $x = \kappa(y) - \kappa(0)$ , it follows that

$$\begin{aligned} \mathbb{E}[IV_t^2] &= 2\xi^2 \int_0^1 (1-y) \exp(\kappa(y)) dy \\ &\approx 2\xi^2 \exp(\kappa(0)) \int_0^1 (1-y)(1+\kappa(y)-\kappa(0)) dy \\ &= \xi^2 \exp(\kappa(0)) \left(1 - \kappa(0) + 2\kappa(0) \int_0^1 (1-y) \cosh(\lambda y) dy\right) \\ &\quad - \xi^2 \exp(\kappa(0)) v^2 \int_0^1 (1-y) y^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 y^2}{4}\right) dy. \end{aligned}$$

The first integral is calculated as

$$\int_0^1 (1-y) \cosh(\lambda y) dy = \frac{\cosh(\lambda) - 1}{\lambda^2}.$$

To handle the second integral, the substitution  $x = y^2$  is done, and hereafter we apply Gradshteyn and Ryzhik (2007), formula 7.512 (11) twice [with  $\mu = 1, \nu = H+1$  and

then  $\mu = 1, \nu = H + \frac{1}{2}$ ,

$$\begin{aligned}
& \int_0^1 (1-y)y^{2H} {}_1F_2\left(1; H + \frac{1}{2}, H + 1; \frac{\lambda^2 y^2}{4}\right) dy \\
&= \frac{1}{2} \int_0^1 (x^{H-1/2} - x^H) {}_1F_2\left(1; H + \frac{1}{2}, H + 1; \frac{\lambda^2 x}{4}\right) dx \\
&= \frac{\Gamma(H + \frac{1}{2})}{2\Gamma(H + \frac{3}{2})} {}_1F_2\left(1; H + \frac{3}{2}, H + 1; \frac{\lambda^2}{4}\right) \\
&\quad - \frac{\Gamma(H + 1)}{2\Gamma(H + 2)} {}_1F_2\left(1; H + \frac{1}{2}, H + 2; \frac{\lambda^2 x}{4}\right) \\
&= \frac{1}{2} \frac{(H + 1) {}_1F_2\left(1; H + \frac{3}{2}, H + 1; \frac{\lambda^2 x}{4}\right) - (H + \frac{1}{2}) {}_1F_2\left(1; H + \frac{1}{2}, H + 2; \frac{\lambda^2}{4}\right)}{(H + \frac{1}{2})(H + 1)} \\
&= \frac{{}_1F_2\left(1; H + \frac{3}{2}, H + 2; \frac{\lambda^2 x}{4}\right)}{(2H + 1)(2H + 2)},
\end{aligned}$$

where the property  $\Gamma(z + 1) = z\Gamma(z)$  is applied repeatedly along with the definition of the generalized hypergeometric function. At last, setting  $c = \frac{v^2}{(2H + 1)(2H + 2)}$  completes the derivation of the variance term.

For each  $\ell \geq 1$ , we approximate the covariance as:

$$\begin{aligned}
\mathbb{E}[IV_t IV_{t+\ell}] &\approx \xi^2 \exp(\kappa(\ell)) \int_0^1 (1-y) (2 + \kappa(\ell + y) - \kappa(\ell) + \kappa(\ell - y) - \kappa(\ell)) dy \\
&= \xi^2 \exp(\kappa(\ell)) \left( 1 - \kappa(\ell) + \int_0^1 (1-y) (\kappa(\ell + y) + \kappa(\ell - y)) dy \right) \\
&= \xi^2 \exp(\kappa(\ell)) \left( 1 - \kappa(\ell) + \kappa(0) \int_0^1 (1-y) (\cosh(\ell + y) + \cosh(\ell - y)) dy \right) \\
&\quad - \xi^2 \exp(\kappa(\ell)) v^2 / 2 \int_0^1 (1-y) (\ell + y)^{2H} {}_1F_2\left(1; H + \frac{1}{2}, H + 1; \frac{\lambda^2 (\ell + y)^2}{4}\right) dy \\
&\quad - \xi^2 \exp(\kappa(\ell)) v^2 / 2 \int_0^1 (1-y) (\ell - y)^{2H} {}_1F_2\left(1; H + \frac{1}{2}, H + 1; \frac{\lambda^2 (\ell - y)^2}{4}\right) dy.
\end{aligned}$$

We calculate each of the above integrals in turn. First,

$$\begin{aligned}
I_1 &= \int_0^1 (1-y) \cosh(\lambda(\ell + y)) dy + \int_0^1 (1-y) \cosh(\lambda(\ell - y)) dy \\
&= \frac{\cosh(\lambda(\ell + 1)) - \cosh(\lambda\ell) - \lambda \sinh(\lambda\ell)}{\lambda^2} + \frac{\cosh(\lambda(\ell - 1)) - \cosh(\lambda\ell) + \lambda \sinh(\lambda\ell)}{\lambda^2} \\
&= \frac{\cosh(\lambda(\ell + 1)) - 2 \cosh(\lambda\ell) + \cosh(\lambda(\ell - 1))}{\lambda^2}
\end{aligned}$$

$$= \frac{2}{\lambda^2} \cosh(\lambda l) (\cosh(\lambda) - 1).$$

Secondly,

$$\begin{aligned} I_2 &= \int_0^1 (1-y)(\ell+y)^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+y)^2}{4}\right) dy \\ &= \int_\ell^{\ell+1} (1+\ell-x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &= \int_0^{\ell+1} (1+\ell-x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &\quad - \int_0^\ell (1+\ell-x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx. \end{aligned}$$

Thirdly, we express the last integral as

$$\begin{aligned} I_3 &= \int_0^1 (1-y)(\ell-y)^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell-y)^2}{4}\right) dy \\ &= \int_{\ell-1}^\ell (1-\ell+x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &= \int_0^\ell (1-\ell+x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &\quad - \int_0^{\ell-1} (1-\ell+x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx. \end{aligned}$$

Hence,

$$\begin{aligned} I_2 + I_3 &= \int_0^{\ell+1} (1+\ell-x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &\quad - 2 \int_0^\ell (\ell-x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &\quad + \int_0^{\ell-1} (\ell-1-x)x^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 x^2}{4}\right) dx \\ &= (\ell+1)^{2H+2} \int_0^1 (1-z)z^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+1)^2 z^2}{4}\right) dz \\ &\quad - 2\ell^{2H+2} \int_0^1 (1-z)z^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2 \ell^2 z^2}{4}\right) dz \\ &\quad + (\ell-1)^{2H+2} \int_0^1 (1-z)z^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell-1)^2 z^2}{4}\right) dz. \end{aligned}$$

At this stage, we proceed as for the variance term and apply Gradshteyn and Ryzhik (2007), formula 7.512 (11). This leads to the evaluation

$$I_2 + I_3 = \frac{(\ell+1)^{2H+2}}{(2H+1)(2H+2)} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{\lambda^2(\ell+1)^2}{4}\right)$$

$$\begin{aligned}
& - \frac{2\ell^{2H+2}}{(2H+1)(2H+2)} {}_1F_2\left(1; H + \frac{1}{2}, H+1; \frac{\lambda^2 \ell^2}{4}\right) \\
& + \frac{(\ell-1)^{2H+2}}{(2H+1)(2H+2)} {}_1F_2\left(1; H + \frac{1}{2}, H+1; \frac{\lambda^2 (\ell-1)^2}{4}\right).
\end{aligned}$$

Plugging the terms into the original expression delivers the acclaimed approximation.  $\square$

### 1.A.5 Proof of Lemma 1.2.6

Recall that  $\kappa(\ell) = \text{cov}(Y_t, Y_{t+\ell})$ , where for the Gamma- $\mathcal{BSS}$  process:

$$\kappa(\ell) = v^2 \int_0^\infty h(x)h(x+\ell)dx.$$

Inserting the Gamma kernel,  $h(x) = x^\alpha e^{-\lambda x}$ , we deduce the identity:

$$\begin{aligned}
\kappa(0) &= v^2 \int_0^\infty x^{2\alpha} \exp(-2\lambda x) dx \\
&= v^2 (2\lambda)^{-2\alpha-1} \int_0^\infty z^{2\alpha} \exp(-2z) dz \\
&= v^2 (2\lambda)^{-2\alpha-1} \Gamma(2\alpha+1).
\end{aligned}$$

Now, for each  $\ell > 0$ ,

$$\begin{aligned}
\kappa(\ell) &= v^2 \exp(-\lambda\ell) \int_0^\infty x^\alpha (x+\ell)^\alpha \exp(-2\lambda x) dx \\
&= \frac{v^2 \Gamma(\alpha+1)}{\sqrt{\pi}} \left(\frac{\ell}{2\lambda}\right)^{\alpha+\frac{1}{2}} K_{\alpha+1/2}(\lambda\ell),
\end{aligned}$$

where the last equality follows from Gradshteyn and Ryzhik (2007), formula 3.383 (8).

Concerning the final part of the statement, we show that conditions (a)-(c) in Theorem 1.2.1 are fulfilled. Since<sup>10</sup>

$$K_a(x) \sim \sqrt{\frac{\pi}{2x}} \exp(-x), \quad x \rightarrow \infty, \quad (1.55)$$

for all  $a \in \mathbb{R}$ , it follows that

$$\kappa(\ell) \sim \frac{v^2 \Gamma(\alpha+1)}{(2\lambda)^{\alpha+1}} \ell^\alpha \exp(-\lambda\ell), \quad \ell \rightarrow \infty,$$

and condition (a) is immediate. Subsequently, for  $y \in [-1, 1]$ ,

$$\frac{\kappa(\ell+y)}{\kappa(\ell)} \sim \frac{(\ell+y)^\alpha \exp(-\lambda(\ell+y))}{\ell^\alpha \exp(-\lambda\ell)} = \left(1 + \frac{y}{\ell}\right)^\alpha \exp(-\lambda y) \sim \exp(-\lambda y), \quad \ell \rightarrow \infty,$$

<sup>10</sup>See the Digital Library of Mathematical Functions (DLMF) §10.25(iii) at <https://dlmf.nist.gov/10.25>.

whereby condition (b) holds with  $\phi(y) = \exp(-\lambda y)$ . In addition, for any  $y \in [-1, 1]$ ,

$$\left| \frac{\kappa(\ell + y)}{\kappa(\ell)} \right| = \left| \frac{\ell + y}{\ell} \right|^{\alpha + \frac{1}{2}} \left| \frac{K_{\alpha + \frac{1}{2}}(\lambda(\ell + y))}{K_{\alpha + \frac{1}{2}}(\lambda\ell)} \right| \leq 2^{\alpha + \frac{1}{2}} \left| \frac{K_{\alpha + \frac{1}{2}}(\lambda(\ell - 1))}{K_{\alpha + \frac{1}{2}}(\lambda\ell)} \right|,$$

invoking the monotonicity of  $x \mapsto K_a(x)$  for  $a \geq 0$ .<sup>11</sup> Thus, thanks to (1.55), we find that

$$\lim_{\ell \rightarrow \infty} \left| \frac{K_{\alpha + \frac{1}{2}}(\lambda(\ell - 1))}{K_{\alpha + \frac{1}{2}}(\lambda\ell)} \right| = \lim_{\ell \rightarrow \infty} \left| \frac{\ell \exp(\lambda)}{\ell + 1} \right| = \exp(\lambda),$$

establishing condition (c).

Hence, Theorem 1.2.1 applies with

$$\int_{-1}^1 (1 - |y|) \phi(y) dy = \int_{-1}^1 (1 - |y|) \exp(-\lambda y) dy = \frac{\exp(-\lambda)(\exp(\lambda) - 1)^2}{\lambda^2},$$

so it follows that

$$\gamma_{\ell+1,1} \sim F(\ell; \alpha, \lambda, \nu, \xi), \quad \ell \rightarrow \infty,$$

where

$$F(\ell; \alpha, \lambda, \nu, \xi) \equiv \underbrace{\frac{\nu^2 \xi^2 \Gamma(\alpha + 1) (\exp(\lambda) - 1)^2}{2^{\alpha+1} \lambda^{\alpha+2}}}_{>0} \ell^\alpha \exp(-\lambda(\ell + 1))$$

for  $\ell > 0$ ,  $\alpha > -\frac{1}{2}$ ,  $\lambda > 0$ ,  $\nu > 0$  and  $\xi > 0$ . □

### 1.A.6 Proof of Theorem 1.3.6

We apply Theorem 2.1 of Hansen (1982), the sufficient conditions of which are implied by our Assumptions 1 – 4. It remains to verify  $G_c(\theta)$  is continuous in  $\theta$ , which also renders the random function  $\hat{m}_T(\theta)$  continuous in  $\theta$ . Next, note that the moduli of continuity of  $\hat{m}_T(\theta)$  and  $G_c(\theta)$  coincide, so  $G_c(\theta)$  being continuous readily implies the so-called first moment continuity of  $\hat{m}_T(\theta)$ , see Hansen (1982, Definition 2.2).

To establish continuity of  $G_c(\theta)$  in  $\theta = (\xi, \phi)$ , note that  $c(\theta)$  is continuous by Assumption 2, whereby it suffices to prove the continuity of  $G(\theta)$ . The first component of  $G(\theta)$  is  $g_0^{(1)}(\theta) = \xi$ , which is evidently continuous, while the remaining components are given in integral form in Theorem 1.2.1. Their continuity is then a consequence of the dominated convergence theorem, given condition (ii) of Assumption 1. □

### 1.A.7 Proof of Theorem 1.3.7

We introduce the notation:

$$\tilde{Q}_{n,T}(\theta) = \tilde{m}_{n,T}(\theta)' \mathbb{W}_T \tilde{m}_{n,T}(\theta),$$

---

<sup>11</sup>See e.g. Digital Library of mathematical functions (DLMF) §10.37 at <https://dlmf.nist.gov/10.37>.



$$Q(\theta) = m(\theta)' \mathbb{W} m(\theta).$$

where  $m(\theta) = G(\theta_0) - G(\theta)$  and  $\mathbb{W} = A' A$ . The claim then follows under the conditions of Theorem 2.1 of Newey and McFadden (1994):

- (i)  $Q(\theta)$  is uniquely minimized at  $\theta_0$ ,
- (ii)  $\Theta$  is compact,
- (iii)  $\theta \rightarrow Q(\theta)$  is continuous, and
- (iv)  $\sup_{\theta \in \Theta} |\tilde{Q}_{n,T}(\theta) - Q(\theta)| \xrightarrow{\mathbb{P}} 0$ .

We note that condition (i) is implied by Assumption 4, since for  $\theta \neq \theta_0$ :

$$Q(\theta) = (Am(\theta))' Am(\theta) > 0 = Q(\theta_0).$$

Condition (ii) is immediate. We already showed condition (iii) in the proof of Theorem 1.3.6. Now, we pass to the last condition (iv). In view of the Cauchy-Schwarz inequality,

$$\begin{aligned} |\tilde{Q}_{n,T}(\theta) - Q(\theta)| &\leq |(\tilde{m}_{n,T}(\theta) - m(\theta))' \mathbb{W}_T (\tilde{m}_{n,T}(\theta) - m(\theta))| + |m(\theta)' (\mathbb{W}_T + \mathbb{W}_T') (\tilde{m}_{n,T}(\theta) - m(\theta))| \\ &\quad + |m(\theta)' (\mathbb{W}_T - \mathbb{W}) m(\theta)| \\ &\leq \|\tilde{m}_{n,T}(\theta) - m(\theta)\|^2 \|\mathbb{W}_T\| + 2 \|m(\theta)\| \|\mathbb{W}_T\| \|\tilde{m}_{n,T}(\theta) - m(\theta)\| \\ &\quad + \|m(\theta)\|^2 \|\mathbb{W}_T - \mathbb{W}\|. \end{aligned}$$

Then, in view of Assumption 3, it suffices to prove that

$$\sup_{\theta \in \Theta} \|\tilde{m}_{n,T} - m(\theta)\| \xrightarrow{\mathbb{P}} 0, \quad \text{as } T \rightarrow \infty \text{ and } n \rightarrow \infty.$$

Let  $m_T(\theta) = T^{-1} \sum_{t=1}^T [\mathbb{V}_t - G(\theta)]$ . Since the convergence  $\sup_{\theta \in \Theta} \|m_T(\theta) - m(\theta)\| \xrightarrow{\mathbb{P}} 0$  was already covered by the proof of Theorem 1.3.6 (setting  $\varepsilon_t = c(\theta) = 0$ ), it remains to show

$$\sup_{\theta \in \Theta} \|\tilde{m}_{n,T}(\theta) - m_T(\theta)\| \xrightarrow{\mathbb{P}} 0, \quad \text{as } T \rightarrow \infty \text{ and } n \rightarrow \infty.$$

To this end, we observe that

$$\begin{aligned} \|\tilde{m}_{n,T}(\theta) - m_T(\theta)\| &\leq \frac{1}{T} \sum_{t=1}^T \|\mathbb{V}_t^n - \mathbb{V}_t\| \\ &\leq \frac{1}{T} \sum_{t=1}^T \left[ |V_t^n - IV_t| + \sum_{j=0}^k |V_t^n V_{t-j}^n - IV_t IV_{t-j}| \right] \\ &\leq \frac{1}{T} \sum_{t=1}^T |V_t^n - IV_t| (1 + |V_t^n| + IV_t) \end{aligned}$$

$$+ \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^k |V_t^n - IV_t| |V_{t-j}^n| + IV_t |V_{t-j}^n - IV_{t-j}|.$$

From Assumption 5 and the Cauchy-Schwarz inequality, we deduce that:

$$\mathbb{E} \left[ \sup_{\theta \in \Theta} \|\tilde{m}_{n,T}(\theta) - m_T(\theta)\| \right] \rightarrow 0, \quad \text{as } T \rightarrow \infty \text{ and } n \rightarrow \infty.$$

which was to be shown.  $\square$

### 1.A.8 Derivation of Remark 1.3.8

Suppose that  $\sup_{s \in \mathbb{R}} \mathbb{E}[\mu_s^4] + \sup_{s \in \mathbb{R}} \mathbb{E}[\sigma_s^4] < \infty$ . Then, there exists a constant  $C$  such that

$$\sup_{t \in \mathbb{Z}} \mathbb{E}[(RV_t^n - IV_t)^2] \leq Cn^{-1}.$$

To see this, we apply Itô's Lemma to get

$$(X_{t-1+\frac{i}{n}} - X_{t-1+\frac{i-1}{n}})^2 = 2 \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} (X_s - X_{t-1+\frac{i-1}{n}}) dX_s + \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} \sigma_s^2 ds.$$

Consequently,

$$\begin{aligned} RV_t^n - IV_t &= 2 \sum_{i=1}^n \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} (X_s - X_{t-1+\frac{i-1}{n}}) dX_s \\ &= 2 \sum_{i=1}^n \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} (X_s - X_{t-1+\frac{i-1}{n}}) \mu_s ds + 2 \sum_{i=1}^n \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} (X_s - X_{t-1+\frac{i-1}{n}}) \sigma_s dW_s. \end{aligned}$$

In turn, this combined with Cauchy-Schwarz and Jensen's inequality leads to

$$\begin{aligned} \mathbb{E}[(RV_t - IV_t)^2] &\leq 4 \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \mathbb{E}[(X_s - X_{\frac{i-1}{n}})^2 \mu_s^2] ds + 4 \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \mathbb{E}[(X_s - X_{\frac{i-1}{n}})^2 \sigma_s^2] ds \\ &\leq Cn^{-1}, \end{aligned}$$

where in the last inequality we exploited  $\sup_{s \geq 0} \mathbb{E}[\mu_s^4] + \sup_{s \geq 0} \mathbb{E}[\sigma_s^4] < \infty$  along with Burkholder's inequality:  $\sup_{s \in [\frac{i-1}{n}, \frac{i}{n}]} \mathbb{E}[(X_s - X_{\frac{i-1}{n}})^4] \leq Cn^{-2}$ .  $\square$

### 1.A.9 Proof of Proposition 1.3.9

The proof relies on a martingale approximation central limit theorem of Peligrad and Utev (2006) and requires some preparation. First, we state and prove a couple of generic, elementary lemmas.

**Lemma 1.A.2.** *Suppose  $X$  is a random variable such that  $\mathbb{E}[X^2] < \infty$  and let  $\mathcal{F}$  and  $\mathcal{G}$  be  $\sigma$ -algebras such that  $\mathcal{F} \subset \mathcal{G}$ . Then,*

$$\|\mathbb{E}[X \mid \mathcal{F}]\|_{L^2(\mathbb{P})} \leq \|\mathbb{E}[X \mid \mathcal{G}]\|_{L^2(\mathbb{P})}.$$

*Proof.* Since  $\mathcal{F} \subset \mathcal{G}$ , we get by the tower property of conditional expectations,

$$\|\mathbb{E}[X | \mathcal{F}]\|_{L^2(\mathbb{P})}^2 = \mathbb{E}[\mathbb{E}[X | \mathcal{F}]^2] = \mathbb{E}[\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{F}]^2].$$

Applying Jensen's inequality for conditional expectations,

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{F}]^2 \leq \mathbb{E}[\mathbb{E}[X | \mathcal{G}]^2 | \mathcal{F}].$$

Hence,

$$\mathbb{E}[\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{F}]^2] \leq \mathbb{E}[\mathbb{E}[\mathbb{E}[X | \mathcal{G}]^2 | \mathcal{F}]] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}]^2] = \|\mathbb{E}[X | \mathcal{G}]\|_{L^2(\mathbb{P})}^2.$$

□

□

**Lemma 1.A.3.** *Suppose that  $X \sim N(\mu, \lambda^2)$  for some  $\mu \in \mathbb{R}$  and  $\lambda > 0$ . Then,*

$$\mathbb{E}[(e^X - 1)^2] \leq (e^{\mu + \lambda^2} + 1)^2 (8|\mu| + 6\lambda^2).$$

*Proof.* Note that

$$\mathbb{E}[(e^X - 1)^2] = e^{2(\mu + \lambda^2)} - 2e^{\mu + \frac{1}{2}\lambda^2} + 1 \leq e^{2(\mu + \lambda^2)} + 2e^{\mu + \lambda^2} + 1 = (e^{\mu + \lambda^2} + 1)^2,$$

while

$$e^{2(\mu + \lambda^2)} - 2e^{\mu + \frac{1}{2}\lambda^2} + 1 = e^{2(\mu + \lambda^2)} - 1 + 2(1 - e^{\mu + \frac{1}{2}\lambda^2}) \leq 8|\mu| + 6\lambda^2 \leq \underbrace{(e^{\mu + \lambda^2} + 1)^2}_{\geq 1} (8|\mu| + 6\lambda^2),$$

for  $|\mu| + \lambda^2 < \frac{1}{2}$  due to the elementary inequality  $|e^x - 1| \leq 2|x|$ , for  $|x| \leq 1$ . However, if  $|\mu| + \lambda^2 \geq \frac{1}{2}$ , then  $8|\mu| + 6\lambda^2 \geq 1$ , so the inequality holds also unconditionally. □ □

Secondly, we exploit these results to prove the following two technical lemmas that estimate the memory of integrated variance and its noisy proxy.

**Lemma 1.A.4.** *Suppose that Assumptions 1 and 6 hold. Moreover, suppose that  $(\mathcal{F}_t)_{t \in \mathbb{R}}$  is a filtration such that  $W$  is adapted and has independent increments with respect to it (cf. condition (iii) in Assumption 7). Then, for any  $p > 0$ ,  $0 \leq s \leq t$  and  $0 \leq s' \leq t'$ ,*

$$(i) \quad \left\| \mathbb{E}_{\theta_0} \left[ \int_{s+r}^{t+r} \sigma_u^2 du \mid \mathcal{F}_0 \right] - \mathbb{E}_{\theta_0} \left[ \int_s^t \sigma_u^2 du \right] \right\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2}),$$

$$(ii) \quad \left\| \mathbb{E}_{\theta_0} \left[ \int_{s+r}^{t+r} \sigma_u^2 du \int_{s'+r}^{t'+r} \sigma_{u'}^2 du' \mid \mathcal{F}_0 \right] - \mathbb{E}_{\theta_0} \left[ \int_s^t \sigma_u^2 du \int_{s'}^{t'} \sigma_{u'}^2 du' \right] \right\|_{L^2(\mathbb{P}_{\theta_0})} = O(r^{-\gamma+1/2}),$$

as  $r \rightarrow \infty$ .

*Proof.* We only prove (ii) as the proof of (i) is analogous. In explicit terms, Assumption 6 says that there exist constants  $u_0 \geq 0$  and  $c > 0$  such that

$$|K(u)| \leq cu^{-\gamma}, \quad u \geq u_0. \quad (1.56)$$

Without loss of generality, assume  $r \geq u_0$  from now on. By Tonelli's theorem,

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[ \int_{s+r}^{t+r} \sigma_u^2 du \int_{s'+r}^{t'+r} \sigma_{u'}^2 du' \mid \mathcal{F}_0 \right] - \mathbb{E}_{\theta_0} \left[ \int_s^t \sigma_u^2 du \int_{s'}^{t'} \sigma_{u'}^2 du' \right] \\ = \int_s^t \int_{s'}^{t'} (\mathbb{E}_{\theta_0} [\sigma_{u+r}^2 \sigma_{u'+r}^2 \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0} [\sigma_u^2 \sigma_{u'}^2]) du du', \end{aligned} \quad (1.57)$$

where

$$\sigma_v^2 \sigma_{v'}^2 = \xi^2 e^{-\kappa(0)} \exp \left( \int_{-\infty}^{\infty} K^+(v, v', \tau) dB_\tau \right)$$

with  $K^+(v, v', \tau) = K(v - \tau) + K(v' - \tau)$  for any  $v, v' \geq 0$  (recall we set  $K(v) = 0$  for any  $v \leq 0$ ). Subsequently,

$$\mathbb{E}_{\theta_0} [\sigma_u^2 \sigma_{u'}^2] = \xi^2 e^{-\kappa(0)} \exp \left( \frac{1}{2} \int_{-\infty}^{\infty} K^+(u, u', \tau)^2 d\tau \right),$$

while, by the assumed properties of the Brownian motion  $B$  and the filtration  $(\mathcal{F}_t)_{t \in \mathbb{R}}$ ,

$$\mathbb{E}_{\theta_0} [\sigma_{u+r}^2 \sigma_{u'+r}^2 \mid \mathcal{F}_0] = \xi^2 e^{-\kappa(0)} \exp \left( \int_{-\infty}^0 K^+(u, u', \tau - r) dB_\tau + \frac{1}{2} \int_0^{\infty} K^+(u, u', \tau - r)^2 d\tau \right),$$

using the property  $K^+(u + r, u' + r, \tau) = K^+(u, u', \tau - r)$ .

Therefore,

$$\mathbb{E}_{\theta_0} [\sigma_{u+r}^2 \sigma_{u'+r}^2 \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0} [\sigma_u^2 \sigma_{u'}^2] = \xi^2 e^{-\kappa(0)} \exp \left( \frac{1}{2} \int_{-\infty}^{\infty} K^+(u, u', \tau)^2 d\tau \right) (\exp(Y_r^{u, u'}) - 1), \quad (1.58)$$

with

$$Y_r^{u, u'} = \int_{-\infty}^0 K^+(u, u', \tau - r) dB_\tau - \frac{1}{2} \overline{K}^{u, u'}(r) \sim N \left( -\frac{1}{2} \overline{K}^{u, u'}(r), \overline{K}^{u, u'}(r) \right) \quad (1.59)$$

and  $\overline{K}^{u, u'}(r) = \int_{-\infty}^{-r} K^+(u, u', \tau)^2 d\tau = \int_{-\infty}^0 K^+(u, u', \tau - r)^2 d\tau$ . Applying Tonelli's theorem and Jensen's inequality to (1.57), we conclude that:

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[ \left( \mathbb{E}_{\theta_0} \left[ \int_{s+r}^{t+r} \sigma_u^2 du \int_{s'+r}^{t'+r} \sigma_{u'}^2 du' \mid \mathcal{F}_0 \right] - \mathbb{E}_{\theta_0} \left[ \int_s^t \sigma_u^2 du \int_{s'}^{t'} \sigma_{u'}^2 du' \right] \right)^2 \right] \\ \leq (t - s)(t' - s') \int_s^t \int_{s'}^{t'} \mathbb{E}_{\theta_0} [(\mathbb{E}_{\theta_0} [\sigma_{u+r}^2 \sigma_{u'+r}^2 \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0} [\sigma_u^2 \sigma_{u'}^2])^2] du du', \end{aligned}$$

where, by (1.58) – (1.59) and Lemma 1.A.3,

$$\begin{aligned} \mathbb{E}_{\theta_0} [(\mathbb{E}_{\theta_0} [\sigma_{u+r}^2 \sigma_{u'+r}^2 \mid \mathcal{F}_0] - \mathbb{E}_{\theta_0} [\sigma_u^2 \sigma_{u'}^2])^2] \\ = e^{-2\kappa(0)} \exp \left( \int_{-\infty}^{\infty} K^+(u, u', \tau)^2 d\tau \right) \mathbb{E}_{\theta_0} [(\exp(Y_r^{u, u'}) - 1)^2] \\ \leq 14 e^{-2\kappa(0)} \exp \left( \int_{-\infty}^{\infty} K^+(u, u', \tau)^2 d\tau \right) (e^{\frac{3}{2} \overline{K}^{u, u'}(r)} + 1)^2 \overline{K}^{u, u'}(r) \end{aligned}$$

$$\leq 14(e^{3\kappa(0)} + 1)^2 \bar{K}^{u,u'}(r),$$

after observing that

$$\bar{K}^{u,u'}(r) \leq \int_{-\infty}^{\infty} K^+(u, u', \tau)^2 d\tau \leq 2 \int_0^{\infty} K(\tau)^2 d\tau = 2\kappa(0).$$

Moreover, if  $\tau \leq -r$  then  $-\tau \geq r \geq u_0$ , whereby  $u - \tau \geq u_0$  and  $u' - \tau \geq u_0$  since  $u \geq s \geq 0$  and  $u' \geq s' \geq 0$ . Thus, by (1.56),

$$\begin{aligned} \bar{K}^{u,u'}(r) &= \int_{-\infty}^{-r} (K(u - \tau) - K(u' - \tau))^2 d\tau \\ &\leq 2c^2 \int_{-\infty}^{-r} ((u - \tau)^{-2\gamma} + (u' - \tau)^{-2\gamma}) d\tau \\ &\leq 2c^2 \int_r^{\infty} \tau^{-2\gamma} d\tau = \frac{2c^2}{1 - 2\gamma} r^{-2\gamma+1}. \end{aligned}$$

Consequently,

$$\begin{aligned} &\left\| \mathbb{E}_{\theta_0} \left[ \int_{s+r}^{t+r} \sigma_u^2 du \int_{s'+r}^{t'+r} \sigma_{u'}^2 du' \mid \mathcal{F}_0 \right] - \mathbb{E}_{\theta_0} \left[ \int_s^t \sigma_u^2 du \int_{s'}^{t'} \sigma_{u'}^2 du' \right] \right\|_{L^2(\mathbb{P}_{\theta_0})} \\ &\leq (t - s)(t' - s') \left( \frac{28}{1 - 2\gamma} \right)^{1/2} (e^{3\kappa(0)} + 1) c r^{-\gamma+1/2} = O(r^{-\gamma+1/2}), \end{aligned}$$

as  $r \rightarrow \infty$ , which concludes the proof of (ii).  $\square$   $\square$

**Lemma 1.A.5.** *If Assumptions 1 – 2 and 6 – 7 hold, then*

$$\sum_{r=1}^{\infty} r^{-1/2} \left\| \mathbb{E}_{\theta_0} \left[ \widehat{\mathbb{V}}_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - G_c(\theta_0) \right\|_{L^2(\mathbb{P}_{\theta_0})} < \infty.$$

*Proof.* Let  $r \geq 1$ . First, we consider:

$$\mathbb{E}_{\theta_0} \left[ \widehat{IV}_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - g_0^{(1)}(\theta_0) = \mathbb{E}_{\theta_0} \left[ IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [IV_1] + \mathbb{E}_{\theta_0} \left[ \varepsilon_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right],$$

where

$$\mathbb{E}_{\theta_0} \left[ \varepsilon_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = \mathbb{E}_{\theta_0} \left[ \mathbb{E}_{\theta_0} [\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma, \varepsilon}] \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = 0$$

by Assumption 2 and the tower property of conditional expectations, which is applicable since  $\mathcal{F}_0^{\widehat{\mathbb{V}}} \subset \mathcal{F}_{r-1}^{\sigma, \varepsilon}$ . Therefore,

$$\begin{aligned} \left\| \mathbb{E}_{\theta_0} \left[ \widehat{IV}_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - g_0^{(1)}(\theta_0) \right\|_{L^2(\mathbb{P}_{\theta_0})} &= \left\| \mathbb{E}_{\theta_0} \left[ IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [IV_1] \right\|_{L^2(\mathbb{P}_{\theta_0})} \\ &\leq \left\| \mathbb{E}_{\theta_0} \left[ IV_r \mid \mathcal{F}_0^{W, \varepsilon} \right] - \mathbb{E}_{\theta_0} [IV_1] \right\|_{L^2(\mathbb{P}_{\theta_0})} \quad (1.60) \\ &= O(r^{-\gamma+1/2}), \quad r \rightarrow \infty, \end{aligned}$$

which follows by Lemma 1.A.2, again since  $\mathcal{F}_0^{\widehat{\mathbb{V}}} \subset \mathcal{F}_0^{W,\varepsilon}$ , and Lemma 1.A.4(i).

Secondly,

$$\begin{aligned} & \mathbb{E}_{\theta_0} \left[ \widehat{IV}_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - g_0^{(2)}(\theta_0) - c(\theta_0) \\ &= \mathbb{E}_{\theta_0} \left[ IV_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [IV_1^2] + 2\mathbb{E}_{\theta_0} \left[ \varepsilon_r IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] + \mathbb{E}_{\theta_0} \left[ \varepsilon_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [\varepsilon_1^2], \end{aligned}$$

where

$$\mathbb{E}_{\theta_0} \left[ \varepsilon_r IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = \mathbb{E}_{\theta_0} \left[ \mathbb{E}_{\theta_0} [\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}] IV_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = 0,$$

by the tower property, since  $\mathcal{F}_0^{\widehat{\mathbb{V}}} \subset \mathcal{F}_{r-1}^{\sigma,\varepsilon}$ , and Assumption 2. By virtue of condition (7) in Assumption 7 and Minkowski's inequality:

$$\begin{aligned} \left\| \mathbb{E}_{\theta_0} \left[ \widehat{IV}_r^2 \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - g_0^{(2)}(\theta_0) - c(\theta_0) \right\|_{L^2(\mathbb{P}_{\theta_0})} &= \left\| \mathbb{E}_{\theta_0} \left[ IV_1^2 \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [IV_1^2] \right\|_{L^2(\mathbb{P}_{\theta_0})} + O(r^{-\gamma+1/2}) \\ &\leq \left\| \mathbb{E}_{\theta_0} \left[ IV_r^2 \mid \mathcal{F}_0^{W,\varepsilon} \right] - \mathbb{E}_{\theta_0} [IV_1^2] \right\|_{L^2(\mathbb{P}_{\theta_0})} + O(r^{-\gamma+1/2}) \\ &= O(r^{-\gamma+1/2}), \end{aligned} \tag{1.61}$$

as  $r \rightarrow \infty$ , by Lemma 1.A.2 and 1.A.4(ii).

Lastly, for  $\ell = 1, \dots, k$  (assuming  $r > k$  without loss of generality):

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[ \widehat{IV}_r \widehat{IV}_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - g_\ell(\theta_0) &= \mathbb{E}_{\theta_0} \left[ IV_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [IV_1 IV_{1-\ell}] \\ &\quad + \mathbb{E}_{\theta_0} \left[ \varepsilon_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] + \mathbb{E}_{\theta_0} \left[ IV_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] + \mathbb{E}_{\theta_0} \left[ \varepsilon_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right], \end{aligned}$$

where

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[ \varepsilon_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] &= \mathbb{E}_{\theta_0} \left[ \mathbb{E}_{\theta_0} [\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}] IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = 0, \\ \mathbb{E}_{\theta_0} \left[ IV_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] &= \mathbb{E}_{\theta_0} \left[ IV_r \mathbb{E}_{\theta_0} [\varepsilon_{r-\ell} \mid \mathcal{F}_{r-\ell-1}^{\sigma,\varepsilon}] \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = 0, \\ \mathbb{E}_{\theta_0} \left[ \varepsilon_r \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] &= \mathbb{E}_{\theta_0} \left[ \mathbb{E}_{\theta_0} [\varepsilon_r \mid \mathcal{F}_{r-1}^{\sigma,\varepsilon}] \varepsilon_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] = 0, \end{aligned}$$

from tower property, because  $\mathcal{F}_0^{\widehat{\mathbb{V}}} \subset \mathcal{F}_{r-\ell-1}^{\sigma,\varepsilon} \subset \mathcal{F}_r^{\sigma,\varepsilon}$ , and Assumption 2. Thus, applying yet again Lemma 1.A.2, we get

$$\begin{aligned} \left\| \mathbb{E}_{\theta_0} \left[ \widehat{IV}_r \widehat{IV}_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - g_\ell(\theta_0) \right\|_{L^2(\mathbb{P}_{\theta_0})} &= \left\| \mathbb{E}_{\theta_0} \left[ IV_r IV_{r-\ell} \mid \mathcal{F}_0^{\widehat{\mathbb{V}}} \right] - \mathbb{E}_{\theta_0} [IV_1 IV_{1-\ell}] \right\|_{L^2(\mathbb{P}_{\theta_0})} \\ &\leq \left\| \mathbb{E}_{\theta_0} \left[ IV_r IV_{r-\ell} \mid \mathcal{F}_0^{W,\varepsilon} \right] - \mathbb{E}_{\theta_0} [IV_1 IV_{1-\ell}] \right\|_{L^2(\mathbb{P}_{\theta_0})} \\ &= O(r^{-\gamma+1/2}), \end{aligned} \tag{1.62}$$

as  $r \rightarrow \infty$ , due to Lemma 1.A.4(ii).

Combining (1.60) – (1.62), we deduce that

$$r^{-1/2} \left\| \mathbb{E}_{\theta_0} \left[ \widehat{\mathbb{V}}_r \mid \mathcal{F}_1^{\widehat{\mathbb{V}}} \right] - G_c(\theta_0) \right\|_{L^2(\mathbb{P}_{\theta_0})} = r^{-1/2} O(r^{-\gamma+1/2}) = O(r^{-\gamma}),$$

as  $r \rightarrow \infty$ , whereby the result follows since  $\gamma > 1$ .  $\square$

*Proof of Proposition 1.3.9.* By a Cramér–Wold device, we can reduce the multivariate convergence in (1.47) to a univariate problem. To this end, fix arbitrary  $a \in \mathbb{R}^{k+2}$  and set  $S_T = \sum_{t=1}^T \xi_t$ , where  $\xi_t = a'(\widehat{\mathbb{V}}_t - G_c(\theta_0))$ ,  $t \in \mathbb{Z}$ , so that

$$T^{-1/2} S_T = T^{1/2} a' \widehat{m}_T(\theta_0).$$

Note that  $(\xi_t)_{t \in \mathbb{Z}}$  inherits the stationarity and ergodicity of  $(\widehat{\mathbb{V}}_t)_{t \in \mathbb{Z}}$ . Moreover,  $\mathbb{E}_{\theta_0}[\xi_1] = 0$  and  $\mathbb{E}_{\theta_0}[\xi_1^2] < \infty$  under the present assumptions. As the natural filtration  $\mathcal{F}_t^\xi = \sigma\{\xi_t, \xi_{t-1}, \dots\}$ ,  $t \in \mathbb{Z}$ , of  $(\xi_t)_{t \in \mathbb{Z}}$  has  $\mathcal{F}_t^\xi \subset \mathcal{F}_t^{\widehat{\mathbb{V}}}$  for any  $t \in \mathbb{Z}$ , then by Lemmas 1.A.2 and 1.A.5,

$$\sum_{r=1}^{\infty} r^{-1/2} \|\mathbb{E}[\xi_r \mid \mathcal{F}_0^\xi]\|_{L^2(\mathbb{P}_{\theta_0})} \leq \|a\|_{\mathbb{R}^{k+2}} \sum_{r=1}^{\infty} r^{-1/2} \left\| \mathbb{E}[\widehat{\mathbb{V}}_r \mid \mathcal{F}_0^{\widehat{\mathbb{V}}}] - G_c(\theta_0) \right\|_{L^2(\mathbb{P}_{\theta_0})} < \infty.$$

Appealing to Theorem 1 and Corollary 2 of Peligrad and Utev (2006),

$$T^{-1/2} S_T \xrightarrow[T \rightarrow \infty]{\mathcal{L}} N\left(0, \sum_{\ell=-\infty}^{\infty} \mathbb{E}_{\theta_0}[\xi_1 \xi_{1+\ell}]\right),$$

with long-run variance

$$\sum_{\ell=-\infty}^{\infty} \mathbb{E}_{\theta_0}[\xi_1 \xi_{1+\ell}] = \sum_{\ell=-\infty}^{\infty} \mathbb{E}_{\theta_0}[a'(\widehat{\mathbb{V}}_1 - G_c(\theta_0))(\widehat{\mathbb{V}}_{1+\ell} - G_c(\theta_0))' a] = a' \Sigma_{\widehat{\mathbb{V}}} a,$$

and the proposition is verified.  $\square$

**Proposition 1.A.6.** *Suppose that Assumptions 1 and 6 hold. Then condition (ii) of Assumption 7 applies in Examples 1.3.2, 1.3.3, and 1.3.5.*

*Proof.* The error terms in Examples 1.3.2 and 1.3.5 differ only by scaling factor, so it suffices to look at the former. Then,

$$\varepsilon_t = \left( \frac{2}{n} \int_{t-1}^t \sigma_u^2 du \right)^{1/2} Z_t, \quad t \in \mathbb{Z},$$

so that, using the filtration  $\mathcal{F}_t^{B,Z} = \sigma\{Z_t, Z_{t-1}, \dots\} \vee \sigma\{B_u : u \leq t\}$ ,  $t \in \mathbb{Z}$ :

$$\mathbb{E}_{\theta_0}[\varepsilon_r^2 \mid \mathcal{F}_0^{B,Z}] - \mathbb{E}_{\theta_0}[\varepsilon_1^2] = \frac{2}{n} \left( \mathbb{E}_{\theta_0} \left[ \int_{r-1}^r \sigma_u^2 du \mid \mathcal{F}_0^{B,Z} \right] - \mathbb{E}_{\theta_0} \left[ \int_0^1 \sigma_u^2 du \right] \right), \quad r \geq 1,$$

since  $\int_{r-1}^r \sigma_u^2 du$  and  $Z_r$  are conditionally independent on  $\mathcal{F}_0^{B,Z}$ . We can then apply Lemma 1.A.4(i) and 1.A.2 to show the conjecture of this part.

In Example 1.3.3,

$$\varepsilon_t = \sum_{i=1}^n (Z_{t,i}^2 - 1) \int_{t-1+\frac{i-1}{n}}^{t-1+\frac{i}{n}} \sigma_u^2 du, \quad t \in \mathbb{Z},$$

whereby for any  $r \geq 1$ ,

$$\mathbb{E}_{\theta_0} \left[ \varepsilon_r^2 \mid \mathcal{F}_0^{B,Z} \right] - \mathbb{E}_{\theta_0} [\varepsilon_1^2] = 2 \sum_{i=1}^n \left( \mathbb{E}_{\theta_0} \left[ \left( \int_{r-1+\frac{i-1}{n}}^{r-1+\frac{i}{n}} \sigma_u^2 du \right)^2 \mid \mathcal{F}_0^{B,Z} \right] - \mathbb{E}_{\theta_0} \left[ \left( \int_{i-1}^i \sigma_u^2 du \right)^2 \right] \right).$$

Applying Minkowski's inequality and Lemmas 1.A.4(i) and 1.A.2 concludes the proof.  $\square$   $\square$

### 1.A.10 Proof of Theorem 1.3.10

We set  $Q_T = \hat{m}_T(\theta)' \mathbb{W}_T \hat{m}_T(\theta)$  and note that  $Q_T$  attains its minimum value at  $\hat{\theta}_T$ . Combining this with the mean value theorem yields that

$$0 = \nabla_{\theta} Q_T(\hat{\theta}_T) = \nabla_{\theta} Q_T(\theta_0) + \nabla_{\theta\theta}^2 Q_T(\bar{\theta}_T)(\hat{\theta}_T - \theta_0),$$

where  $\bar{\theta}_T$  lies between  $\hat{\theta}_T$  and  $\theta_0$ . Now,

$$\begin{aligned} \nabla_{\theta} Q_T(\theta) &= 2 \nabla_{\theta} \hat{m}_T(\theta)' \mathbb{W}_T \hat{m}_T(\theta), \\ \nabla_{\theta\theta}^2 Q_T(\theta) &= 2 \nabla_{\theta\theta} \hat{m}_T(\theta)' \mathbb{W}_T \hat{m}_T(\theta) + 2 \nabla_{\theta\theta} \hat{m}_T(\theta)' \mathbb{W}_T \nabla_{\theta} \hat{m}_T(\theta) \end{aligned}$$

This leads to:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = (\nabla_{\theta\theta}^2 Q_T(\bar{\theta}_T))^{-1} 2 \nabla_{\theta} \hat{m}_T(\theta_0)' \mathbb{W}_T \sqrt{T} \hat{m}_T(\theta_0). \quad (1.63)$$

Invoking the assumptions of the theorem, it follows that  $\bar{\theta}_T \xrightarrow{\mathbb{P}} \theta$  as  $T \rightarrow \infty$ . In addition, and recalling Proposition 1.3.9, we deduce that as  $T \rightarrow \infty$ :

$$\begin{aligned} \sqrt{T} \hat{m}_T(\theta_0) &\xrightarrow{d} N(0, \Sigma_{\mathbb{W}}), \\ \nabla_{\theta} \hat{m}_T(\theta_0) &\xrightarrow{\mathbb{P}} G, \\ \nabla_{\theta\theta}^2 Q_T(\bar{\theta}_T) &\xrightarrow{\mathbb{P}} G' W G, \end{aligned}$$

where the last part uses that  $\hat{m}_T(\theta_0) \xrightarrow{\mathbb{P}} 0$ . Then, Slutsky's theorem finishes the proof.  $\square$

### 1.A.11 Proof of Theorem 1.3.11

Proceeding as above, and denoting  $\tilde{Q}_{n,T} = \tilde{m}_{n,T}(\theta)' \mathbb{W}_T \tilde{m}_{n,T}(\theta)$ , we find that

$$\sqrt{T}(\tilde{\theta}_{n,T} - \theta_0) = (\nabla_{\theta\theta}^2 \tilde{Q}_{n,T}(\check{\theta}_{n,T}))^{-1} 2 \nabla_{\theta} \tilde{m}_{n,T}(\theta_0)' \mathbb{W}_T \sqrt{T} \tilde{m}_{n,T}(\theta_0).$$



Then, as  $T \rightarrow \infty$  and  $n \rightarrow \infty$ ,

$$\begin{aligned}\sqrt{T} \tilde{m}_{n,T}(\theta_0) &\xrightarrow{d} N(0, \Sigma_{\mathbb{V}}), \\ \nabla_{\theta} \tilde{m}_{n,T}(\theta_0) &\xrightarrow{\mathbb{P}} \tilde{G}, \\ \nabla_{\theta\theta}^2 \tilde{Q}_{n,T}(\check{\theta}_{n,T}) &\xrightarrow{\mathbb{P}} \tilde{G}' W \tilde{G},\end{aligned}$$

To wrap up, we again exploit Slutsky's theorem. □

## 1.2 References

- Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047–1092.
- Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115–158.
- Andersen, T.G., Bollerslev, T., 1998a. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T.G., Bollerslev, T., 1998b. Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53, 219–265.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andersen, T.G., Sørensen, B.E., 1996. GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 14, 328–352.
- Asmussen, S., Glynn, P.W., 2007. *Stochastic Simulation: Algorithms and Analysis*. 1st ed., Springer-Verlag.
- Back, K., 1991. Asset prices for general processes. *Journal of Mathematical Economics* 20, 371–395.
- Barndorff-Nielsen, O.E., Basse-O'Connor, A., 2011. Quasi ornstein–uhlenbeck processes. *Bernoulli* 17, 916–941.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian Orstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B* 63, 167–241.
- Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B* 64, 253–280.
- Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–48.

- Bayer, C., Friz, P., Gatheral, J., 2016. Pricing under rough volatility. *Quantitative Finance* 16, 887–904.
- Belyaev, Y.K., 1961. Continuity and Hölder's conditions for sample functions of stationary Gaussian processes, in: *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Volume II*. University of California Press, Berkeley, pp. 23–33.
- Bennedsen, M., 2016. Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data. Working paper. Aarhus University.
- Bennedsen, M., Lunde, A., Pakkanen, M.S., 2017a. Decoupling the short- and long-term behavior of stochastic volatility. Working paper. Aarhus University.
- Bennedsen, M., Lunde, A., Pakkanen, M.S., 2017b. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics* 21, 931–965.
- Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33–65.
- Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability* 8, 1–14.
- Christie, A.A., 1982. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics* 10, 407–432.
- Comte, F., Coutin, L., Renault, E., 2012. Affine fractional stochastic volatility models. *Annals of Finance* 8, 337–378.
- Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8, 291–323.
- Corradi, V., Distaso, W., 2006. Semi-parametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies* 73, 635–667.
- Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300, 463–520.
- Duffie, D., Pan, J., Singleton, K.J., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.
- Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of markov models of asset prices. *Econometrica* 61, 929–952.
- Dym, H., McKean, H.P., 1976. Gaussian processes, function theory, and the inverse spectral problem. Academic Press, New York.

- Euch, O.E., Rosenbaum, M., 2018. The characteristic function of rough Heston models. *Mathematical Finance* 29, 3–38.
- Fridman, M., Harris, L., 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economic Statistics* 16, 284–291.
- Fukasawa, M., Takabatake, T., Westphal, R., 2019. Is volatility rough? Working paper. Osaka University.
- Gallant, A.R., Hsieh, D.A., Tauchen, G.E., 1997. Estimation of stochastic volatility with diagnostics. *Journal of Econometrics* 81, 159–192.
- Garnier, J., Sølna, K., 2018. Option pricing under fast-varying and rough stochastic volatility. *Annals of Finance* 14, 489–516.
- Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. *Quantitative Finance* 18, 933–949.
- Gradshteyn, I.S., Ryzhik, I.M., 2007. Table of integrals, series, and products. 7th ed., Academic Press.
- Guenoun, H., Jacquier, A., Roome, P., Shi, F., 2018. Asymptotic behavior of the fractional Heston model. *SIAM Journal of Financial Mathematics* 9, 1017–1045.
- Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* 24, 127–161.
- Hansen, P.R., Lunde, A., 2014. Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory* 30, 60–93.
- Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. *Review of Financial Studies* 6, 247–264.
- Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–343.
- Hurst, H.E., 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770–799.
- Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications* 119, 2249–2276.

- Kaarakka, T., Salminen, P., 2011. On fractional ornstein-uhlenbeck processes. *Communications on Stochastic Analysis* 5, 121–133.
- Karhunen, K., 1950. über die struktur stationärer zufälliger funktionen. *Arkiv för Matematik* 1, 141–160.
- Lindgren, G., 2006. Lectures on stationary stochastic processes. Lectures notes, Lund University, available at: <http://www.maths.lth.se/matstat/staff/georg/Publications/lecture2006.pdf>.
- Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional brownian motions, fractional noises and applications. *SIAM Review* 10, 422–437.
- Maruyama, G., 1949. The harmonic analysis of stationary stochastic processes. *Memoirs of the Faculty of Science, Kyūsyū University, Series A* 4, 45–106.
- Meddahi, N., 2002. A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479–508.
- Meddahi, N., 2003. ARMA representation of integrated and realized variances. *Econometrics Journal* 6, 335–356.
- Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.
- Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis. *Handbook of Econometrics*, IV, Edited by R. Engle and D. McFadden , 2112–2245.
- Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
- Peccati, G., Taqqu, M.S., 2011. *Wiener chaos: moments, cumulants and diagrams*. Springer, Milan.
- Peligrad, M., Utev, S., 2006. Central limit theorem for stationary linear processes. *Annals of Probability* 34, 1608–1622.
- Taqqu, M.S., 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31, 287–302.
- Taylor, S.J., 1986. *Modelling Financial Time Series*. 1 ed., John Wiley & Sons.

Tegnér, M., Poulsen, R., 2018. Volatility is log-normal – but not for the reason you think. *Risks* 6, –.

Todorov, V., 2009. Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics* 148, 131–148.

Todorov, V., Tauchen, G., 2011. Volatility jumps. *Journal of Business and Economic Statistics* 29, 356–371.

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

# 2

CHAPTER

## OPTION PRICING USING ROUGH REALIZED MEASURES

**Anine Eg Bolko**

*Aarhus University and CREATES*

### **Abstract**

This paper presents an option pricing method in a rough log-normal stochastic volatility model based on parameters estimated from historical realized variance. We establish a model-preserving measure change from the physical probability space into a corresponding risk-neutral pricing one, and define the mapping of the parameters into this space in the case where the fractional Ornstein-Uhlenbeck process is considered. The market is incomplete, and information on the market price of volatility risk is extracted using VIX data and parametric forecasting of realized variance.

**Keywords:** Rough volatility; high-frequency data; realized variance; stochastic volatility; option pricing

## 2.1 Introduction

Stochastic volatility models are important in the field of mathematical finance, because they capture a realistic framework for pricing financial assets and related derivatives. In this category, plenty of different designs have been suggested including the popular ones by Heston (1993) and Hull and White (1987). Prudent modeling of the asset prices involve the ability to generate crucial properties equivalent to those observed in the market. Thus, in the context of derivative pricing using stochastic volatility models, performance depends on the ability to replicate the Black-Scholes option implied volatility surface. For options with very short time to maturity, the implied volatility surface reveals an explosive steep volatility smile, which classical continuous stochastic volatility models cannot replicate (e.g. Carr and Wu (2003)). However, this behavior of the implied volatility surface can be generated by continuous volatility processes, which in the Hölder sense, have rough trajectories, (see e.g. Bayer, Friz and Gatheral (2016); Gatheral, Jaisson and Rosenbaum (2018); Fukasawa (2017)). They demonstrate that these models are even flexible enough to construct realistic implied volatility surfaces, not only for the part with short time to maturity, but can simultaneously also produce the flatter smile for the long time to maturity options. Therefore, the class of rough stochastic volatility models is promising for derivative pricing. Yet, the non-Markovian structure, which is essential to create roughness, entails deficiencies with the mathematical tractability. Only in the rough Heston model, the affine structure provides closed form solutions of the characteristic function (Euch and Rosenbaum, 2019), while the more appealing log-normal framework suffers from lack of closed form density function and characteristic function. This implies that calibration procedures are complicated, and makes analytical methods for option pricing infeasible in most cases, driving us to Monte Carlo based pricing. However, even calibration by means of Monte Carlo is cumbersome, caused by the computationally heavy simulation schemes. Instead, parameters used for option pricing in the rough log-normal stochastic volatility models have been chosen by qualified guesses and machine learning techniques (e.g. Bayer et al. (2016); Bayer, Horvath, Muguruza, Stemper and Tomas (2019)).

The volatility is the main ingredient associated with option pricing. However, knowledge of this quantity is not only available from the derivative market, but also through the historical asset price itself. Realized variance is an obvious choice to extract that information, and rough models have appropriate features for modeling the spot volatility under the physical probability measure as well (see e.g. Bennedsen, Lunde and Pakkanen (2017); Gatheral, Jaisson and Rosenbaum (2018)). Using realized variance, different time series estimation procedures are available in the rough framework, for example by a two step-approach with a Whittle-type estimator in Fukasawa, Takabatake and Westphal (2019) or as proposed here, the generalized method of moments procedure in Bolko, Christensen, Pakkanen and Veliyev (2020).

This paper suggests circumventing the hard calibration task involving option



data and instead taking advantage of the information on the spot volatility from historical return data. A similar approach using real world data for option pricing is considered in Stentoft (2008) for a stochastic volatility setup and for Heterogeneous Autoregressive (HAR) type of model by Corsi, Fusari and Vecchia (2013). To establish the necessary measure change, it is crucial to consider under which risk neutral pricing measure the market participants operate. We assume that the model structure is preserved under the measure change, which is supported by for example Livieri, Mouti, Pallavicini and Rosenbaum (2018). The risk neutral pricing measure is thus defined by the market price of risk and the market price of volatility risk, where we ensure the model-preserving structure by a couple of mild restrictions on these terms. This is common practice, see for example Carr and Wu (2009) and Bollerslev, Tauchen and Zhou (2009).

The results of this paper are two-fold. First, we show that a model preserving risk neutral measure is available in the fractional log-normal setup under mild conditions. Specifically, we consider the fractional Ornstein-Uhlenbeck process, and show that the model structure is preserved when the market price of volatility risk is linear in log-volatility. A result which in more general sense applies to log-normal volatility models driven by stationary Gaussian Volterra processes. Secondly, an overview of necessary estimation methodology for this approach to be implemented is gathered.

The paper is organized as follows. Section 2.2 presents the theoretical framework of the model, and establishes the theory behind the measure change within this framework. Section 2.3 goes over the general theory on option pricing in this framework, while Section 2.4 explains the estimation procedure applicable to the physical dynamics of the stock price. Finally, Section 2.5 concludes.

## 2.2 The Theoretical Framework

Consider a financial market defined on the physical probability space,  $(\Omega, \mathcal{F}, \mathbb{P})$ , filtered by the right-continuous and complete filtration,  $(\mathcal{F}_t)_{t \in [0, T]}$ , containing market information. By assumption, there is a risk-free asset in the market whose price,  $B = (B_t)_{t \in [0, T]}$ , depends on a deterministic interest rate,  $r = (r_t)_{t \in [0, T]}$ , in the following way,

$$dB_t = r_t B_t dt, \quad \text{for } t \in [0, T]. \quad (2.1)$$

Additionally, the market contains a risky asset with the price process  $S = (S_t)_{t \in [0, T]}$ . The dynamics of this process are described by a stochastic volatility model, such that the price process itself is an Itô semi-martingale,

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_{1,t}^{\mathbb{P}}, \quad (2.2)$$

$$d \log(\sigma_t) = \lambda (\theta - \log(\sigma_t)) dt + \nu dW_{2,t}^{H, \mathbb{P}}, \quad \text{for } t \in [0, T]. \quad (2.3)$$

The driving process  $W_1^{\mathbb{P}} = (W_{1,t}^{\mathbb{P}})_{t \in [0, T]}$  is a standard Brownian motion, and the drift,  $\mu = (\mu_t)_{t \in [0, T]}$ , is locally bounded and predictable. The volatility process,  $\sigma =$

$(\sigma_t)_{t \in [0, T]}$ , is the exponential of a fractional Ornstein-Uhlenbeck process with speed of mean-reversion,  $\lambda > 0$ , volatility of volatility parameter,  $\nu > 0$ , and long-run mean,  $\theta \in \mathbb{R}$ .<sup>1</sup> The driving process,  $W_2^{H, \mathbb{P}} = (W_{2,t}^{H, \mathbb{P}})_{t \in [0, T]}$ , is a fractional Brownian motion with Hurst index,  $H \in (0, 1)$ . The fractional Brownian motion is a centered Gaussian process and can be interpreted as a Volterra type of process by the representation of Norros, Valkeila and Virtamo (1999):

$$W_t^H = \int_0^t k_H(t, s) dW_s, \quad (2.4)$$

where

$$k_H(t, s) = c_H \left( \left( \frac{t}{s} \right)^{H-\frac{1}{2}} (t-s)^{H-\frac{1}{2}} - \left( H - \frac{1}{2} \right) s^{H-\frac{1}{2}} \int_s^t u^{H-\frac{3}{2}} (u-s)^{H-\frac{1}{2}} du \right), \quad (2.5)$$

and

$$c_H = \sqrt{\frac{2H\Gamma(\frac{3}{2}-H)}{\Gamma(H+\frac{1}{2})\Gamma(2-2H)}}. \quad (2.6)$$

For  $H = \frac{1}{2}$ , the kernel reduces to the indicator function from zero to  $t$ ,  $k_{1/2}(t, s) = \mathbb{1}_{(0, t)}(s)$ , and the process becomes a standard Brownian motion. In that case, log-volatility is both Markovian and a semi-martingale, and standard Itô calculus applies. For  $H \neq \frac{1}{2}$  this is not the case, and instead other tools from white noise theory or Malliavin calculus are available. For an overview of the theory, see Nualart (2006). However, note that for all  $H$  the solution to the system in (2.3) exists in the Riemann-Stieltje sense (Cheridito, Kawaguchi and Maejima, 2003), and is given by

$$\log(\sigma_t) = e^{-\lambda t} \log(\sigma_0) + \theta(1 - e^{-\lambda t}) + \nu \int_0^t e^{-\lambda(t-s)} dW_{2,s}^{H, \mathbb{P}}. \quad (2.7)$$

The integral with respect to the fractional Brownian motion is well-defined in this case by the regularity of the exponential function. It is Lipschitz continuous or equivalently Hölder-continuous of order one, while the fractional Brownian motion is Hölder-continuous of order up to  $H$ .<sup>2</sup> Notice also that moments of the log-volatility exist and are finite, ensuring that the stock price dynamic in (2.2) is well-defined. One should be aware if implementing a stochastic volatility of volatility process instead of a deterministic term,  $\nu$ , since this might disrupt the regularity, and it is necessary to define the integral by the use of Malliavin Calculus. This case is however outside the scope of this paper.

The stock return and the volatility process are generally negatively correlated. This effect is referred to as the leverage effect, and is of great importance for option

<sup>1</sup> Due to the logarithmic structure, also the variance process is a fractional Ornstein-Uhlenbeck process, but with long-run mean parameter  $2\theta$  and volatility of volatility parameter  $2\nu$ .

<sup>2</sup> The integral  $\int_a^b f(x) dg(x)$  exists in the Riemann-Stieltje sense if  $f$  is Hölder continuous of order  $\alpha$  and  $g$  is Hölder continuous of order  $\beta$  such that  $\alpha + \beta > 1$ , see e.g. Young (1936).

pricing purposes, since it controls the symmetry of the implied volatility surface. We capture this property by implementing correlation between the driving Brownian motion in (2.2) and (2.4). This is done by the common decomposition,

$$W_{1,t}^{\mathbb{P}} = \rho W_{2,t}^{\mathbb{P}} + \sqrt{1 - \rho^2} W_{2\perp,t}^{\mathbb{P}} \quad \text{for } t \in [0, T], \quad (2.8)$$

where  $W_{2\perp,t}^{\mathbb{P}}$  and  $W_{2,t}^{\mathbb{P}}$  are independent Brownian motions.

The model introduced here was originally proposed by Comte and Renault (1998) in a long memory version with  $H \in (1/2, 1)$  and later by Gatheral, Jaisson and Rosenbaum (2018) in the rough settings with  $H \in (0, 1/2)$ . In both cases, it belongs to the class of log-normal stochastic volatility models, which motivated by data has potential to model the spot volatility well (e.g. Andersen, Bollerslev, Diebold and Ebens (2001); Tegnér and Poulsen (2018)). Moreover, this structure implies that we can fully define the volatility process by the first and second order moments of the driving fractional Ornstein-Uhlenbeck process. The volatility inherits the stationary nature, if the driving process is stationary, and does generally depends on its full past trajectories, when the driving process is fractional.

**Remark 2.2.1.** *The theory on the measure change introduced here, can be applied to a much wider class of log-normal stochastic volatility models. The only condition is that the system in (2.2) and (2.3) is well-defined and that the driving process for the volatility is a stationary Gaussian Volterra process.*

### 2.2.1 The measure change

Prices of financial derivatives based on the underlying asset introduced in (2.2) are obtainable by means of martingale pricing. To apply this technique, it is necessary to consider the dynamics of the price process under an equivalent martingale measure. An equivalent measure can be defined by the Radon-Nikodym derivative,  $L = (L_t)_{t \in [0, T]}$ ,

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_t = L_t. \quad (2.9)$$

To ensure that the new measure,  $\mathbb{Q}$ , is a probability measure,  $L$  must be a martingale and its expected value has to be equal to one on the compact interval,  $[0, T]$ , and thus,  $L$  must satisfy that  $L_0 = 1$  and  $\mathbb{E}^{\mathbb{P}}[L_t] = 1$  for all  $t$  in the interval. This is achieved by defining the process,  $L$ , as the Dolean-Dade exponential of a martingale,  $Z_t$ ,

$$\mathcal{E}(Z_t) := \exp(Z_t - \frac{1}{2} \langle Z \rangle_t) = L_t \quad \text{for } t \in [0, T] \quad (2.10)$$

with

$$Z_t = \int_0^t \varphi_{2,s} dW_{2,s}^{\mathbb{P}} + \int_0^t \varphi_{2\perp,s} dW_{2\perp,s}^{\mathbb{P}}. \quad (2.11)$$

By defining  $Z$  in this way, Girsanov Theorem implies that  $\mathbb{Q}$  is a well-defined probability measure, and that under this measure, each of the Brownian motions under  $\mathbb{P}$

become the Brownian motion and a drift term under  $\mathbb{Q}$  such that,

$$\begin{aligned} W_{2,t}^{\mathbb{Q}} &= W_{2,t}^{\mathbb{P}} - \int_0^t \varphi_{2,s} ds, \\ W_{2\perp,t}^{\mathbb{Q}} &= W_{2\perp,t}^{\mathbb{P}} - \int_0^t \varphi_{2\perp,s} ds, \end{aligned} \quad (2.12)$$

are  $\mathbb{Q}$ -Brownian motions. And it follows that,

$$W_{1,t}^{\mathbb{Q}} = W_{1,t}^{\mathbb{P}} - \int_0^t \varphi_{1,s} ds \quad (2.13)$$

with  $\varphi_{1,s} = \rho\varphi_{2,s} + \sqrt{1-\rho^2}\varphi_{2\perp,s}$  is a  $\mathbb{Q}$ -Brownian motion as well, which drives the price process under the new measure. As a consequence of Girsanov Theorem and the representation of fBm in (2.4), the volatility driving fractional Brownian motion is likewise a  $\mathbb{Q}$ -fractional Brownian motion and a drift:

$$\begin{aligned} W_{2,t}^{H,\mathbb{Q}} &= \int_0^t k_H(t,s) (dW_{2,s}^{\mathbb{P}} - \varphi_{2,s} ds) \\ &= W_{2,t}^{H,\mathbb{P}} - \int_0^t k_H(t,s) \varphi_{2,s} ds \\ &:= W_{2,t}^{H,\mathbb{P}} - \int_0^t \tilde{\varphi}_{2,s} ds, \end{aligned} \quad (2.14)$$

Therefore, constructing the measure change as explained above, ensures that the driving process preserves its structure down to an additional drift term. For a formal proof of the fractional measure change, see Norros et al. (1999). Finally, to arrive at a risk neutral pricing measure, it is necessary to introduce some conditions on the market price of risk,  $\varphi_1$ , and on the market price of log-volatility risk,  $\varphi_2$ .

**Assumption 10.** *The measure change is restricted such that,*

1.  $\varphi_1$  and  $\varphi_2$  satisfy the necessary regularity conditions for  $L$  to be a martingale.
2. The market price of risk is completely determined by the current state of the market,

$$\varphi_{1,t} = \frac{\mu_t - r_t}{\sigma_t}.$$

3. The market price of log-volatility risk is restricted to be a affine function of log-volatility such that for  $a, b \in \mathbb{R}$ ,

$$\tilde{\varphi}_{2,t} = a + b \log(\sigma_t).$$

Assumption 10.1 is not at all restrictive, and the necessary regularity can be imposed by satisfying Novikov's condition.<sup>3</sup> Assumption 10.2 ensures that the market

---

<sup>3</sup>The Dolean-Dade exponential,  $\mathcal{E}\left(\int_0^T \phi_t dW_t\right)$ , is a martingale if  $\mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^T \phi_t^2 dt\right)\right] < \infty$ . By independence between  $W_2$  and  $W_{2\perp}$ , each of the integrals in (2.11) satisfy this implying sufficient regularity.

is free of arbitrage (no free lunch with vanishing risk). It corresponds to letting the drift of the asset equal the risk-free rate under the new measure, and hence that the discounted asset price is a martingale. It is therefore not necessary to estimate this quantity to do option pricing as long as interest rates (and dividend yields) are observable. Note that if the asset pays out a deterministic dividend yield,  $q_t$ , it can be handled by replacing  $r_t$  with  $r_t - q_t$  in here and in the following theory. However, to ease the notion,  $q_t$  is omitted. On the other hand, the market price of log-volatility risk, does not need to be restricted in order to do arbitrage-free pricing. Instead, Assumption 10.3 is the minimum restriction needed to ensure that the pricing measure preserves the model structure after the measure change. This kind of restriction is common in the literature (see e.g. Carr and Wu (2009)), and is often restricted even more, by letting it equal a constant or simply assuming it to be zero.

**Theorem 2.2.2.** *Under Assumption 10 the model structure in (2.2) and (2.3) is preserved, and there exists a one-to-one mapping of the parameters between the two probability spaces such that for  $t \in [0, T]$*

$$\begin{aligned} dS_t &= r_t S_t dt + \sigma_t S_t dW_{1,t}^{\mathbb{Q}} \\ d \log \sigma_t &= \tilde{\lambda}(\tilde{\theta} - \log \sigma_t) dt + \nu dW_{2,t}^{H,\mathbb{Q}}, \end{aligned} \quad (2.15)$$

where  $W_2^{H,\mathbb{Q}}$  is a  $\mathbb{Q}$ -fBm and

$$\tilde{\rho} = \rho, \quad \tilde{\lambda} = \lambda + b\nu, \quad \tilde{\theta} = \frac{\lambda\theta - a\nu}{\lambda + b\nu}. \quad (2.16)$$

The proof is found in Appendix 2.A. Notice that if we restrict the market price of volatility risk to be a constant, the measure change only impacts the level of the volatility process. In general however, the affine structure affects both the level parameter and the speed of mean reversion, while correlation, volatility of volatility and the Hurst parameter are preserved under the pricing measure.

**Proposition 2.2.3.** *The market described in (2.2) and (2.3) is free of arbitrage and incomplete.*

While the market is free of arbitrage due to Assumption 10.2, it is incomplete since  $a$  and  $b$  are not uniquely determined, and each set of these parameters provides an equivalent risk neutral pricing measure. A discussion of which to apply is saved for Section 2.4.2.

## 2.3 Derivative pricing

Let a derivative instrument of a financial asset be described by its pay-off function,  $g: \mathbb{R}_+ \rightarrow \mathbb{R}$ . Our main focus is on European call options, but instruments with other pay-off structures can easily be considered by the same approach. For the European

call option, the pay-off function is  $g(S_\tau; K) = (S_\tau - K)_+$ , where  $S_\tau$  is the price of the underlying financial asset at maturity,  $\tau$  and  $K$  is the deterministic strike price of the contract. From the martingale pricing theory, the price of any such contract should be considered under a risk neutral probability measure,  $\mathbb{Q}$ , and equal the present value of the expected pay-off. The price of a European call option at time  $t$  with expiry at time  $\tau$  and strike price,  $K$  is therefore given by:

$$\begin{aligned} C_t(\tau, K) &= \exp\left(-\int_t^{t+\tau} r_s ds\right) \mathbb{E}_t^{\mathbb{Q}}[(S_{t+\tau} - K)_+] \\ &= \exp\left(-\int_t^{t+\tau} r_s ds\right) \mathbb{E}_t^{\mathbb{Q}}[(e^{X_{t+\tau}} - K)_+] \quad \text{for } 0 \leq t < t + \tau \leq T. \end{aligned} \quad (2.17)$$

Here,  $\mathbb{E}_t^{\mathbb{Q}}[\cdot]$  denotes the  $\mathbb{Q}$ -expectation conditional on the information set available on time  $t$ ,  $\mathcal{F}_t$ . We will keep this notation in the following. The first term is the discount factor, while the second term is the  $\mathbb{Q}$ -expected pay-off of the call contract. Even though the dynamics of  $S$  are well-known under  $\mathbb{Q}$ , neither the analytical form of the density function nor the Fourier transform are known due to the complicated structure of the volatility process. Instead, option prices can be determined by means of Monte Carlo.

### Dealing with the non-Markovian structure of the volatility process

The non-Markovian structure of the volatility process implies that the time  $t$  expectation not only depends on the current state of volatility, but on the full past trajectory of the driving process. This is of course not convenient since it is hard to handle long past trajectories in real data and requires at least a truncation. Instead, we can benefit from the adapted expectations in the observable forward variance following the approach in Bayer, Friz and Gatheral (2016).

**Theorem 2.3.1.** *Consider volatility in the log-normal setup driven by a stationary Gaussian process, with*

$$\sigma_t = \exp\left(\theta + \int_{-\infty}^t \psi(t-u) dW_u\right), \quad (2.18)$$

where  $\psi(\cdot)$  is continuous and well-defined. Then the volatility process can be expressed by its forward variance curve, and a martingale process. For  $s < t$ , it follows that,

$$\sigma_t = \mathbb{E}^{\mathbb{Q}}[\sigma_t | \mathcal{F}_s] \mathcal{E}\left(\int_s^t \psi(t-u) dW_u\right), \quad (2.19)$$

where  $\mathcal{E}(\cdot)$  is the Dolean-Dade exponential and  $\mathbb{E}_s^{\mathbb{Q}}[\sigma_t]$  is the time  $t$  forward volatility given the information at time  $s$ .

It is well-known that most stationary Gaussian processes have such a moving average type of representation. From Karhunen (1950) it is only necessary to add

some mild conditions on the spectral density. Barndorff-Nielsen and Basse-O'Connor (2011) shows that the fractional Ornstein-Uhlenbeck process satisfy this, and its moving average representation is given by

$$\psi(t) = \nu c_H \left( t^{H-1/2} - \lambda \int_0^t e^{-\lambda(t-u)} u^{H-1/2} du \right) \quad \text{with} \quad c_H = \frac{\sqrt{2H \sin(\pi H) \Gamma(2H)}}{\Gamma(H+1/2)}. \quad (2.20)$$

The volatility process is of course still non-Markovian, but all dependence on the past is hidden in the forward volatility curve, and given that we observe this object, simulation of the semi-martingale part can be done independently of the past. Note that the log-normal structure allows us to translate information from the forward variance curve into estimation of the forward volatility curve, since

$$\mathbb{E}_s^{\mathbb{Q}}[\sigma_t] = \sqrt{\mathbb{E}_s^{\mathbb{Q}}[\sigma_t]} \exp\left(\mathbb{E}_s^{\mathbb{Q}}\left[\int_s^t \psi(t-u) dW_u^2\right]\right) \quad (2.21)$$

When  $H = 1/2$ , it is easy to see that the expression in (2.20) reduces such that  $\psi(t) = \nu e^{-\lambda t}$ , and the volatility process takes the form of the exponentially transformed standard Ornstein-Uhlenbeck process. In that case, the volatility process becomes Markovian, and this is reflected in the forward volatility curve, which only depends on the time  $s$  state of the volatility process. Lastly, it is worth mentioning that also the class of Brownian Semi-Stationary processes are possible to deal with in this way.

## 2.4 Estimation: Methodology

The purpose of this section is to introduce an applicable estimation procedure of parameters occurring under the risk neutral model settings, and hence which either directly or indirectly are used for option pricing in empirical data. That is, estimation of model parameters in the volatility model under the physical probability measure, estimation of the market price of log-volatility risk, and finally a discussion on how to extract the forward variance curve from the option market. The methods are applied and discussed regarding SPY data from January 2001 to December 2019 extracted from NYSE trade and quote (TAQ) database.

### 2.4.1 Dynamics of spot volatility under the physical measure

The theoretical framework considered here assumes that the price process of the underlying asset follows the geometric Brownian motion in (2.2). This implies that the asset return process,  $X = (X_t)_{t \geq 0}$ , is given by

$$X_t = X_0 + \int_0^t \mu_s - \frac{1}{2} \sigma_s^2 ds + \int_0^t \sigma_s dW_{1,s}^{\mathbb{P}}, \quad t \in [0, T] \quad (2.22)$$

**Table 2.1:** Parameter estimation of log-normal fractional stochastic volatility models under  $\mathbb{P}$ .

model	parameter			
	$\theta$	$\lambda \times 10$	$\nu$	$H$
log-normal SV	-2.1120	6.3470	0.4555	0.5000
log-normal fSV	-3.2540	0.0002	0.8540	0.0297

*Note.* We report the the result of the simultaneous GMM estimation of  $(\theta, \lambda, \nu, H)$  using 5 min RV on SPY from 2001-2019.

and hence, realized variance is a consistent estimator for the quadratic variation of  $X$  (Barndorff-Nielsen and Shephard, 2002):

$$RV_t^{(n)} = \sum_{i=1}^n \left| X_{t-1+i/n} - X_{t-1+(i-1)/n} \right|^2 \rightarrow \int_{t-1}^t \sigma_s^2 ds, \quad t \in [0, T], \quad (2.23)$$

when  $n$  tends to infinity, and where  $X_{t-1+i/n}$  is the  $i$ 'th return observation at day  $t$ , with  $t$  indicating the end of that trading day. Therefore, daily realized variance holds information on the spot volatility dynamics, but is not a convenient proxy for this quantity due to its distributional properties, which differ from those attached to the quadratic variation of  $X$  under the relevant model assumptions. We consider the parametric generalized method of moments (GMM) procedure introduced for fractional log-normal stochastic volatility models in Bolko, Christensen, Pakkanen and Veliyev (2020), and which takes care of this obstacle by being based on moments of the quadratic variation. To circumvent impact on the parameter-estimates implied by either microstructure noise or by discretization errors, a relatively low 5-minute sampling frequency is chosen and additionally, the bias correction term suggested for the procedure is applied. The parameters  $\lambda$ ,  $\nu$ ,  $\theta$  and  $H$ , which are obtained using realized variance directly, are those describing the spot variance process. Recall that, due to the invariant log-normal structure, only  $\lambda$  and  $\theta$  must be adjusted to obtain parameters for the volatility process described in (2.3). The results for spot-volatility attached to SPY are collected in Table 2.1. The estimates are in line with the literature on both models (e.g. Tegnér and Poulsen (2018); Gatheral, Jaisson and Rosenbaum (2018)), and suggest that the long-run mean of the volatility in the log-normal SV model is 0.1425, while in the log-normal fSV it is almost equivalently estimated to 0.1356. In the next section, we will investigate the change in this parameter when switching to the risk-neutral pricing dynamics.

#### 2.4.2 Market price of log-volatility risk

To translate the parameters from the stochastic volatility model under  $\mathbb{P}$  into the associated model under a risk-neutral pricing measure,  $\mathbb{Q}$ , Theorem 2.2.2 tells us that it is necessary to determine the market price of log-volatility risk. Since the market is incomplete, the pricing measure is not unique, and which measure to use for pricing depends on the preferences of the market participants. Intuitively, the market price



of log-volatility risk is the premium investors demand to carry an additional risk on fluctuations in log-volatility. As seen in the theoretical framework, the premium is closely related to the difference between the  $\mathbb{P}$  and the  $\mathbb{Q}$  dynamics of the volatility process. Invoking Assumption 10.3 ensures that the model structure is preserved when switching between the two probability spaces. Here we further impose the restriction that  $b = 0$ , implying the market price of log-volatility risk,  $\tilde{\varphi}_{2,t} = a$ , equals a constant. The consequence of this is that the market price of log-volatility risk is equivalent in all types of volatility regimes, and therefore only the long-run mean parameter differs between the two spaces. We dub the difference of the unconditional long-run mean of the log-variance process under  $\mathbb{P}$  and  $\mathbb{Q}$  respectively, LMPVR, and refer to this quantity as the market price of log-variance risk from now on. It is convenient to consider LMPVR, due to its dependence to the model parameters:

$$\text{LMPVR} := \mathbb{E}^{\mathbb{P}} [\log(\sigma_t^2)] - \mathbb{E}^{\mathbb{Q}} [\log(\sigma_t^2)] = 2\theta - 2\tilde{\theta} = \frac{2\nu\tilde{\varphi}_2}{\lambda}. \quad (2.24)$$

Estimation of LMPVR is therefore possible by comparing predicted future level of log-variance under  $\mathbb{P}$  and  $\mathbb{Q}$  respectively, see e.g. Carr and Wu (2009); Bollerslev, Tauchen and Zhou (2009). Under the risk neutral pricing measure, information on the long-run mean can be extracted from the derivative market. Both variance swaps and the standardized Volatility Index (VIX) are commonly used to extract such information. Here, we choose to use the VIX index, to approximate the long-run mean parameter for the covariance-stationary variance process, using that:

$$\mathbb{E}^{\mathbb{Q}} [\sigma_s^2] = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \mathbb{E}_t^{\mathbb{Q}} \left[ \int_t^{t+\tau} \sigma_s^2 ds \right] \quad (2.25)$$

and we estimate the  $\mathbb{Q}$ -expected mean of the variance process by:

$$\widehat{\mathbb{E}^{\mathbb{Q}} [\sigma_s^2]} = \frac{1}{\tau} \mathbb{E}_t^{\mathbb{Q}} \left[ \int_t^{t+\tau} \sigma_s^2 ds \right] = \left( \frac{\text{VIX}_t}{100} \right)^2 \quad (2.26)$$

for  $\tau$  equal to 30 days. Therefore, the expression in (2.26) can be used as approximation of the long run-mean parameter even though it is only exact for  $\tau \rightarrow \infty$ . Likewise, the long-run mean parameter of the variance process under  $\mathbb{P}$ , is approximated by the  $\mathbb{P}$  forecast, such that:

$$\widehat{\mathbb{E}^{\mathbb{P}} [\sigma_s^2]} = \frac{1}{\tau} \mathbb{E}_t^{\mathbb{P}} \left[ \int_t^{t+\tau} \sigma_s^2 ds \right] \quad (2.27)$$

When  $H < 0.5$  and  $\lambda$  is small, this forecast is calculated using the approximating forecasting procedure suggested for similar models in Gatheral, Jaisson and Rosenbaum (2018) using a window of 250 days. In the special case where  $H = 1/2$ , the forecast is simple due to the Markovian structure, and it is straightforward to calculate the conditional mean. Thus, in any case, the forecasting procedure is chosen consistently with the model under consideration. Note that these quantities concern the dynamics

on variance rather than log-variance, and it is necessary to transform this knowledge. Remember that,

$$\mathbb{E}^{\mathbb{P}} [\log(\sigma_t^2)] = \log(\mathbb{E}^{\mathbb{P}} [\sigma_t^2]) - \frac{1}{2} \text{Var}^{\mathbb{P}}(\log(\sigma_t^2)). \quad (2.28)$$

An equivalent expression is available under  $\mathbb{Q}$ , but since the change of measure only affects the mean and not the variance of the process,  $\text{Var}^{\mathbb{P}}(\log(\sigma_t^2)) = \text{Var}^{\mathbb{Q}}(\log(\sigma_t^2))$  and evens out in (2.24). Thus, the market price of log-variance is obtained by,

$$LMPVR = \log(\mathbb{E}^{\mathbb{P}} [\sigma_t^2]) - \log(\mathbb{E}^{\mathbb{Q}} [\sigma_t^2]), \quad (2.29)$$

and is approximated with

$$\widehat{LMPVR} = \log(\widehat{\mathbb{E}^{\mathbb{P}} [\sigma_t^2]}) - \log(\widehat{\mathbb{E}^{\mathbb{Q}} [\sigma_t^2]}). \quad (2.30)$$

In practice, the sample mean of LMPVR over all days in the data set is considered.

In Figure 2.1, the forecasts of the variance are depicted for the log-normal SV model together with that obtained from VIX data. In general, the latter is on a higher level, yielding negative market price of variance. This is a well-known fact, and is connected to the negative correlation between returns and volatility. In the right figure, the market price of log-volatility risk is obtained and we see that the series is much more volatile using the standard log-normal volatility model compared to the rough settings. This is caused by a much stronger memory generated in the rough framework by the slow mean-reversion parameter, implying that these forecasts are not blown up to the same extent during and after extreme days. Otherwise, the series takes the same form, and we check that there is a positive correlation between market price of log-variance and the level of the variance process. This suggests that the assumption on  $b$  might be too restrictive, and it is possible that it is beneficial to relax this assumption. While we leave the dependence on log-variance outside the scope of this paper, we do relax the assumption on  $a$  slightly by allowing for a level shift. Visual inspection of the market price of log-variance risk seems to have a level shift arriving in 2010 just after the financial crisis. Here, it becomes more negative on average, possible caused by the increasing focus on variance products in the market at this point in time.

Table 2.2 summarizes the market price of variance risk extracted by VIX and reports the long-run mean parameter,  $\hat{\theta}$ , suggested by this analysis. Summing up, all model parameters for the  $\mathbb{Q}$ -dynamics are obtained, and ready to be considered on the derivative market having SPY or similarly S&P 500 index as the underlying asset.<sup>4</sup> Taking a look at the volatility formula in Theorem 2.3.1 reveals that the long-run mean parameter is only indirectly necessary for volatility pricing through the forward variance curve. While Markovian models such as the log-normal SV do not need anything but the present value of variance and the long-run mean parameter

---

<sup>4</sup>SPY is an ETF based on S&P 500 index, and therefore reflects the dynamics of SPX as well.

**Table 2.2:** Parameter estimation of market price of log-variance risk.

model	2001 – 2019		2001 – 2009		2010 – 2019	
	LMPVR	$\theta$	LMPVR	$\theta$	LMPVR	$\theta$
log-normal SV	-0.5569	-1.8335	-0.3707	-1.9266	-0.7303	-1.7468
log-normal fSV	-0.7137	-2.8971	-0.5151	-2.9964	-0.8801	-2.8139

*Note.* We LMPV estimated by comparing VIX index with 30-days parametric forecast of log-SV and log-fSV model.

to obtain closed form interpretation of the forward variance curve, the fractional analogue depends on all past information. This makes it preferable to obtain such information directly from the market.

### 2.4.3 The forward variance curve

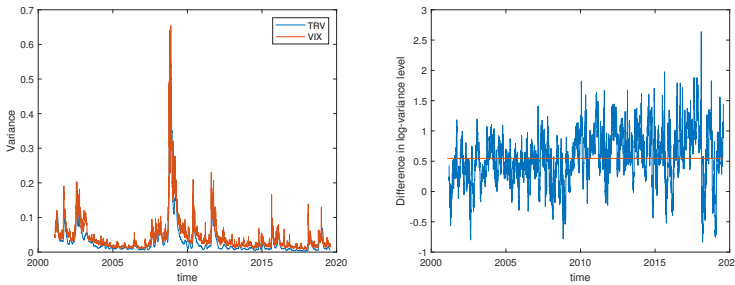
Information on the forward variance curve under the pricing measure  $\mathbb{Q}$  can either be calculated parametrically or approximated by price data in the market. In non-Markovian SV models, the first option demands information on the full past history of the volatility process, which is of course infeasible in practice. From a practical viewpoint, the forward variance curve is indirectly available in market data, through the price of log-return contracts, since

$$\mathbb{E}_t^{\mathbb{Q}} \left[ \int_t^{t+\tau} \sigma_s^2 ds \right] = -2\mathbb{E}_t^{\mathbb{Q}} \left[ \log \left( \frac{S_{t+\tau}}{F} \right) \right], \quad (2.31)$$

where  $F = e^{\int_t^{t+\tau} r_s ds} S_t$  is the forward price. All contracts, including this one can be replicated by European put and call options in a non-parametric manner, and for this type of contract, it is given by

$$\mathbb{E}_t^{\mathbb{Q}} \left[ \int_t^{t+\tau} \sigma_s^2 ds \right] = 2e^{-\int_t^{t+\tau} r_s ds} \left( \int_0^F \frac{P_t(\tau, K; S_t, t)}{K^2} dK + \int_F^\infty \frac{C_t(\tau, K)}{K^2} dK \right). \quad (2.32)$$

Thus, the forward variance curve is obtained by differentiation of this object. In practice even very liquid markets do not contain options with all different maturities

**Figure 2.1:** Market price of log-variance risk.

*Note.* Realized volatility refers to 30-days ahead average forecasts of 5-minute realized variance on SPY and VIX refers to the squared VIX-index in percentage.

and strikes, and hence interpolation is required, and can for example be solved by using the SVI surface (see Gatheral and Jacquier (2014)). The procedure is run on a day by day basis and ensures that option pricing is adjusted to the past history every day.

## 2.5 Conclusion

This paper establishes a measure change from the real world measure,  $\mathbb{P}$ , to a risk-neutral pricing measure,  $\mathbb{Q}$ , and extracts a one-to-one mapping between parameters in the log-normal fSV model between these two spaces. Due to incompleteness of the market, the mapping is not unique. However, by imposing linearity in log-volatility on the market price of log-volatility risk, the measure change is model-preserving. Such a change will only affect the long-run mean and the mean-reversion parameter. If, in addition the dependence of log-volatility on the market price of log-volatility risk is dropped, the measure change solely affects the long run mean parameter. Considering the difference of mean in the level of volatility under  $\mathbb{P}$  and  $\mathbb{Q}$ , it is clearly too restrictive to consider market price of log-volatility risk to be a constant and we suggest at least to allow for level shifts.

The one-to-one mapping is useful regarding option pricing, and we suggest to use a time series of realized variance to estimate parameters for this purpose. Based on these, the market price of log-volatility risk and the forward variance curve, all log-normal SV models driven by stationary Volterra processes are available for pricing in the derivatives market. It is left for further research to evaluate the performance of this approach in practice.

## 2.A Proofs

In this section, proofs of all results in the paper is available including some auxiliary results.

### Proof of Theorem 2.2.2

*Proof.* (Theorem 2.2.2) The stock price dynamics follows straight forward by inserting the  $\mathbb{Q}$ -Brownian motion in (2.2),

$$\begin{aligned} dS_t &= \mu_t S_t dt + \sigma_t S_t \left( dW_{1,t}^{\mathbb{Q}} - \varphi_{1,t} dt \right) \\ &= (\mu_t - \sigma_t \varphi_{1,t}) S_t dt + \sigma_t S_t dW_{1,t}^{\mathbb{Q}} \\ &= r_t S_t dt + \sigma_t S_t dW_{1,t}^{\mathbb{Q}}, \end{aligned} \tag{2.33}$$

where the last equality is established by Assumption 102. In similar manner, the dynamics of the log-volatility is

$$\begin{aligned}
 d \log \sigma_t &= \lambda(\theta - \log \sigma_t) dt + \nu(dW_{2,t}^{H,\mathbb{Q}} - \tilde{\varphi}_{2,t} dt) \\
 &= \lambda(\theta - \log \sigma_t - \nu \tilde{\varphi}_{2,t}) dt + \nu dW_{2,t}^{H,\mathbb{Q}} \\
 &= \lambda \left( \theta - \frac{a\nu}{\lambda} - (1 + \frac{b\nu}{\lambda}) \log \sigma_t \right) dt + \nu dW_{2,t}^{H,\mathbb{Q}} \\
 &= (\lambda + b\nu) \left( \frac{\lambda\theta - a\nu}{\lambda + b\nu} - \log \sigma_t \right) dt + \nu dW_{2,t}^{H,\mathbb{Q}} \\
 &= \tilde{\lambda}(\tilde{\theta} - \log \sigma_t) dt + \nu dW_{2,t}^{H,\mathbb{Q}},
 \end{aligned} \tag{2.34}$$

which ends the proof.  $\square$   $\square$

### Proof of Theorem 2.3.1

*Proof.* Given the moving average representation of the log-volatility structure, the forward volatility curve is given by,

$$\begin{aligned}
 \mathbb{E}_s^{\mathbb{Q}}[\sigma_t | \mathcal{F}_s] &= \mathbb{E}_s^{\mathbb{Q}} \left[ \exp \left( \theta + \int_{-\infty}^t \psi(t-u) dW_u \right) \right] \\
 &= \exp \left( \theta + \int_{-\infty}^s \psi(t-u) dW_u \right) \mathbb{E}_s^{\mathbb{Q}} \left[ \exp \left( \int_s^t \psi(t-u) dW_u \right) \right] \\
 &= \exp \left( \theta + \int_{-\infty}^s \psi(t-u) dW_u + \frac{1}{2} \mathbb{E}_s^{\mathbb{Q}} \left[ \left( \int_s^t \psi(t-u) dW_u \right)^2 \right] \right).
 \end{aligned} \tag{2.35}$$

Note that the expected value of the last term does not depend on the filtration up to time  $s$ , and the condition on the filtration is irrelevant. Thus inserting this in the formula of the volatility processes the desired result is obtained,

$$\begin{aligned}
 \sigma_t &= \exp \left( \theta + \int_{-\infty}^s \psi(t-u) dW_u + \int_s^t \psi(t-u) dW_u \right) \\
 &= \mathbb{E}_s^{\mathbb{Q}}[\sigma_t | \mathcal{F}_s] \exp \left( \int_s^t \psi(t-u) dW_u - \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[ \left( \int_s^t \psi(t-u) dW_u \right)^2 \right] \right) \\
 &= \mathbb{E}_s^{\mathbb{Q}}[\sigma_t | \mathcal{F}_s] \mathcal{E} \left( \int_s^t \psi(t-u) dW_u \right),
 \end{aligned} \tag{2.36}$$

where  $\mathcal{E}(\cdot)$  is the Dolean-Dade exponential, and last term is hence a martingale.  $\square$

$\square$

## 2.2 References

- Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Barndorff-Nielsen, O.E., Basse-O'Connor, A., 2011. Quasi ornstein–uhlenbeck processes. *Bernoulli* 17, 916–941.
- Barndorff-Nielsen, O.E., Shephard, N., 2002. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17, 457–477.
- Bayer, C., Friz, P., Gatheral, J., 2016. Pricing under rough volatility. *Quantitative Finance* 16, 887–904.
- Bayer, C., Horvath, B., Muguruza, A., Stemper, B., Tomas, M., 2019. On deep calibration of (rough) stochastic volatility models. *arXiv:1908.08806*.
- Bennedsen, M., Lunde, A., Pakkanen, M., 2017. Decoupling the short-and long-term behavior of stochastic volatility. *arXiv:1610.00332*.
- Bolko, A., Christensen, K., Pakkanen, M., Veliyev, B., 2020. Roughness in spot variance? a gmm approach for estimation of log-normal stochastic volatility models using realized measures. *Working Paper*.
- Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. *The Review of Financial Studies* 22, 4463–4492.
- Carr, P., Wu, L., 2003. What type of process underlies options? a simple robust test. *The Journal of Finance* 58, 2581–2610.
- Carr, P., Wu, L., 2009. Variance risk premiums. *The Review of Financial Studies* 22, 1311–1341.
- Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional ornstein-uhlenbeck processes. *Electronic Journal of probability* 8.
- Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. *Mathematical finance* 8, 291–323.
- Corsi, F., Fusari, N., Vecchia, D.L., 2013. Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics* 107, 284–304.
- Euch, O.E., Rosenbaum, M., 2019. The characteristic function of rough heston models. *Mathematical Finance* 29, 3–38.
- Fukasawa, M., 2017. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance* 17, 189–198.

- Fukasawa, M., Takabatake, T., Westphal, R., 2019. Is volatility rough? arXiv:1905.04852 .
- Gatheral, J., Jacquier, A., 2014. Arbitrage-free svi volatility surfaces. *Quantitative Finance* 14, 59–71.
- Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. *Quantitative Finance* 18, 933–949.
- Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6, 327–343.
- Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. *The Journal of Finance* 42, 281–300.
- Karhunen, K., 1950. Über die struktur stationärer zufälliger funktionen. *Arkiv för Matematik* 1, 141–160.
- Livieri, G., Mouti, S., Pallavicini, A., Rosenbaum, M., 2018. Rough volatility: evidence from option prices. *IIE transactions* 50, 767–776.
- Norros, I., Valkeila, E., Virtamo, J., 1999. An elementary approach to a girsanov formula and other analytical results on fractional brownian motions. *Bernoulli* 5, 571–587.
- Nualart, D., 2006. The Malliavin calculus and related topics. volume 1995. Springer.
- Stentoft, L., 2008. Option pricing using realized volatility. CREATES Research Paper .
- Tegnér, M., Poulsen, R., 2018. Volatility is log-normal—but not for the reason you think. *Risks* 6, 46.
- Young, L.C., 1936. An inequality of the hölder type, connected with stieltjes integration. *Acta Mathematica* 67, 251.





## FORECASTING REALIZED VARIANCE USING A SIMPLE ARIMA MODEL

**Anine Eg Bolko**

*Aarhus University and CREATES*

### Abstract

We propose a forecasting procedure for realized variance using ARIMA(0,1,q) models for its log-transformed. The setup is compatible with the recently popular rough volatility theory, but keeps a simple discrete-time interpretation, and allows for a natural rendition even when the fractal index is close to the boundary of the definition area. An in-sample analysis based on S&P 500 index data reveals that only a small number of moving-average lags are necessary to obtain a good fit. In the out-of-sample experiment, the forecasting performance of these models is evaluated. One-day-ahead forecasts are statistically equally as good or better compared to relevant benchmark models.

**JEL Classification:** C10; C50.

**Keywords:** Fractional integrated processes; realized variance; high-frequency data; stochastic volatility; forecasting

### 3.1 Introduction

The price of a financial asset fluctuates randomly over time, and this exposes market participants for uncertainty associated with investments. For financial decision making, for example within risk management, portfolio allocation, and derivative pricing, it is crucial to understand this risk exposure. Stochastic variance is a key concept in that respect, and prudent modeling of this component makes it possible to construct reliable forecasts of the expected magnitude of variability a number of periods ahead. Therefore, the literature on this subject is not surprisingly massive. An early, but thorough overview of the forecasting methods are given in Poon and Granger (2003), and most of the subsequent literature originates from the ideas herein.

A major hurdle to overcome is how to collect information on the behavior of asset price variance since it is not directly observed in the market. Instead, characteristics can be extracted through the return series of the asset. Initially, this led to conditional volatility models such as the (G)ARCH framework by Engle (1982) and Bollerslev (1986) using low frequency asset returns. Later, the accessibility of high frequency data facilitates the construction of accurate realized measures for the variance process such as realized variance (e.g. Andersen and Bollerslev (1998); Barndorff-Nielsen and Shephard (2002)). This allowed for a new desirable line of research in the forecasting literature, where realized measures are considered as proxies for asset variance, and thus variance forecasting is done directly on this historical proxy (e.g. Andersen, Bollerslev, Diebold and Labys (2003); Corsi (2009)). The approach has shown to be beneficial (e.g. Oomen (2001); Poon and Granger (2005); Koopman, Jungbacker and Hol (2005)), and this paper continues the work in that direction.

To a large extent, the variance process is predictable, due to a number of stylized facts, such as its apparent log-normality, clustering of the variance, and a slowly decaying autocorrelation function, (see e.g. Engle and Patton (2001); Andersen, Bollerslev, Diebold and Ebens (2001); Cont and Tankov (2004)). The first property makes log-transformed realized variance a convenient choice of modeling object, and is the origin for many popular models including those considered here. By contrast, the memory property has been widely discussed, and even though the autocorrelation function decays very slowly, unit-root based tests give mixed results implying different modeling strategies. For example Andersen, Bollerslev, Diebold and Labys (2003) propose a genuine long-memory Autoregressive Fractionally Integrated Moving Average (ARFIMA) framework with fractal index,  $d \in (0, 0.5)$ , while Corsi (2009) demonstrates that the Heterogeneous Autoregressive (HAR) model can construct a very similar autocorrelation function by aggregation of short memory components. Both models show auspicious performance for out-of-sample forecasting, and both have its own pros and cons (e.g. Corsi, Mitnik, Pigorsch and Pigorsch (2008); Izzeldin, Hassan, Pappas and Tsionas (2019)). The HAR model is simple to estimate by OLS regression and has an intuitive interpretation motivated by the Heterogeneous Market Hypothesis, but at the same time the long memory impact is lost. On the other

hand, the ARFIMA framework can generate long memory, but lacks the intuitive interpretation of the memory parameter, and the fractal index is estimated close to the non-stationary region. Based on this, it is very likely that the realized variance is not even stationary, and a potential profitable alternative to these models is to consider the stationary first differenced series. By estimation of  $d \in (-0.5, 0)$  this process is highly anti-persistent, and even if it is overdifferenced, this approach can contribute to better forecasts. Moreover, this setup is in line with the recent literature on rough volatility (see e.g. Gatheral, Jaisson and Rosenbaum (2018); Bennedsen, Lunde and Pakkanen (2017)), and also in the regime switching ARFIMA literature (Granger and Hyung, 2004). We suggest canceling out the non-parametric estimation problems regarding the memory index in the close to non-stationary area by considering a moving average representation of log-realized variance, which exists due to the stationarity of its increments. Note that both the ARFIMA framework and the fractional Gaussian noise have similar memory properties, and do both have a moving average representation. Moreover, since the increment series is highly anti-persistent, its autocorrelation function can be truncated after a few lags, and autocorrelation hereafter is close to insignificant. This leaves us with an ARIMA framework where simple applicable estimation procedures are available, and which can approximate popular long-memory frameworks. The ARIMA models are already widely used, not only for monthly macro data, but also for monthly stock market volatility (Schwert, 1987). The good in-sample-fit is also confirmed for realized variance based on high-frequency data.

In the empirical part of this paper an investigation of the out-of-sample forecasting performance for the simple ARIMA models are implemented and compared with popular benchmark models – the HAR and the ARFIMA model – for indexes from all around the world on different forecasting horizons. We find similar behavior of all indexes, and our conclusions applies for example to realized variance for the S&P 500 index and FTSE 100 index. While the simple ARIMA model is not consequently superior to neither the HAR model nor the ARFIMA model, it forecasts at least equally well in most cases.

The paper is organized as follows. Section 3.2 introduces the theoretical framework, the model setup and the motivation for forecasting using an ARIMA model. Hereafter, Section 3.3 goes into details with the forecasting procedure. Section 3.4 is the empirical part, where the ARIMA(0, 1,  $q$ ) models are applied to forecast four major indexes, discuss model selection and compare the results with the benchmark models. Finally, Section 3.5 concludes and makes suggestions for further research.

## 3.2 Settings

The log-price of a financial asset,  $X = (X_t)_{t \leq 0}$ , is adapted to the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ , where  $\mathcal{F}_t$  is the information contained in the market at time  $t$ .

In accordance with the no-arbitrage theory,  $X$  is described by an Itô semi-martingale (Delbaen and Schachermayer, 1994),

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \sum_{i=1}^{N_t^J} J_i \quad \text{for } t \geq 0. \quad (3.1)$$

Here,  $X_0$  is the initial value,  $\mu = (\mu_t)_{t \geq 0} \in \mathbb{R}$  is a bounded and predictable drift process,  $\sigma = (\sigma_t)_{t \geq 0}$  is an adapted càdlàg volatility process and  $W = (W_t)_{t \geq 0}$  is a standard Brownian motion. The last term allows for finite activity jumps and  $N^J = (N_t^J)_{t \geq 0}$  is a Poisson process counting the number of jumps up to time  $t$ , while  $J = (J_i)_{i=1, \dots, N_t^J}$  is a random process controlling the jump sizes.

The quadratic variation of  $X$  explains the variability of the log-price process, and is thus the component of interest. It is described by the volatility process,  $\sigma$ , and the jump component,

$$[X]_t := [X]_t^C + [X]_t^J = \int_0^t \sigma_s^2 ds + \sum_{i=1}^{N_t^J} J_i^2 \quad \text{for } t \geq 0. \quad (3.2)$$

In the case of no jumps, the quadratic variation reduces to the part attached to the continuous driving process, namely the integrated variance. There is no clear answer whether the price process includes jumps or not, but in any case, the focus is on forecasting the total quadratic variation.<sup>1</sup> Given the latent nature of the volatility process and the jump intensity, the total quadratic variation is not observed, and realized measures for this quantity are considered. A consistent estimator is the realized variance (RV), which is undoubtedly also widely used in the literature. It takes the form,

$$RV_t^{(n)} = \sum_{i=1}^n \left| X_{t+\frac{i}{n}} - X_{t+\frac{i-1}{n}} \right|^2, \quad \text{for } t \geq 0, \quad (3.3)$$

where  $X$  is the log-price process observed in  $n+1$  points each day. Here, observations are considered in an equidistant grid over the trading day on a five minute grid. The relatively low frequency of observations used to construct realized variance implies that the contribution of microstructure noise is by rule-of-thumb negligible and is thus ignored.

### 3.2.1 Memory properties of realized variance

Realized variance exhibits several stylized facts, and these are important to construct a discrete-time model, which produces good forecasts. Here, focus is on the memory property, which is, of course, important regarding forecasting, since it explains how the state of the variance process today depends on the level of the variance process

<sup>1</sup>If jumps are present, and the main interest is integrated variance, the approach described here can be applied using jump robust measures such as bipower variation (Barndorff-Nielsen and Shephard, 2004) or truncated realized variance (Mancini, 2009).

in the past. There is no doubt that the autocorrelation function decays very slowly, both for realized variance and log-realized variance, and in fact it decays slower than an exponential function. This motivates the use of long memory models or models with similar strong memory behavior.

In the mathematical sense, the long memory property is directly connected to the spectral density, or equivalently to the covariance function of the process. A covariance stationary Gaussian process has the spectral density defined by its covariance function,  $\gamma(\cdot)$  such that,

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma(k) \exp(-ik\lambda) dk \quad \text{for } \lambda \in (-\pi, \pi) \quad (3.4)$$

and  $i = \sqrt{-1}$ . The process exhibits long memory, if  $d \in (0, 1/2)$  and

$$f(\lambda) = L_f(\lambda) |\lambda|^{-2d}, \quad (3.5)$$

where  $L_f$  is a positive symmetric slowly varying function at zero.<sup>2</sup> This implies that the spectral density function decreases like a power-law function near zero. When  $d \in (-1/2, 0)$ , the process has short memory, but is still stationary. Note that the memory properties are closely connected to the more intuitive covariance function as well, which then also decays similarly to a power-law function for large lags:

$$\gamma(k) \sim C_d |k|^{2d-1} \quad \text{for } k \rightarrow \infty, \quad (3.6)$$

where " $\sim$ " denotes asymptotic equivalence and  $C_d$  is a constant. Hence, the sum of  $\gamma(k)$  over  $k$  diverges in the case of a long memory process. For a meticulous review on long memory processes and their mathematical properties, we refer to Beran, Feng, Ghosh and Kulik (2013).

It is clear from the definition above that long memory properties are closely connected to stationary processes. Even though the definition is absurd in the non-stationary case, such processes can exhibit strong memory in the sense that they do not forget the past. The most exhaustively studied case is the unit-root processes or random walks, corresponding to  $d = 1$ . However, also non-stationary Gaussian processes with memory parameter,  $d \in (1/2, 1)$  exist. It is hard to differentiate the genuine (in the mathematical sense) long memory processes from these, however the non-stationarity impacts basic properties and estimation procedures.

In the literature, it is generally agreed that the memory parameter for log-realized variance is close to the non-stationary limit, but far from the unit-root case. The memory parameter is commonly estimated in the range,  $d \in (0.4, 0.6)$ , and this implies that stationary fractionally integrated processes capable of constructing such memory properties are of interest, whether it is for the log-realized variance itself or

---

<sup>2</sup>A function is said to be slowly varying at zero if  $\lim_{x \downarrow 0} \frac{L_f(ux)}{L_f(x)} = 1$  for all  $u > 0$ , see e.g. Beran, Feng, Ghosh and Kulik (2013).

its stationary differenced series. In the next section, we present a popular example of such a fractionally integrated model, which in the context of forecasting realized variance is useful and the building block to our approach.

### 3.2.2 Fractional modeling of realized variance

While keeping a discrete-time approach for the modeling in mind, the obvious and widely applied choice to create strong memory similar to that observed for log-realized variance series, is the ARFIMA framework. It is a fractionally integrated discrete-time Gaussian process introduced by Granger and Joyeux (1980) and Hosking (1981). The process,  $(X_t)_{t=0,1,\dots,T}$  follows an ARFIMA( $p, d, q$ ) model, if:

$$(1 - L)^d \Phi(L) X_t = \Theta(L) \varepsilon_t, \quad (3.7)$$

where  $(\varepsilon_t)_{t=1,\dots,T}$  is i.i.d. Gaussian noise with variance  $\sigma_\varepsilon^2$ , and the fractional lag-operator is given by

$$(1 - L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j, \quad (3.8)$$

$\Phi(L) = 1 + \sum_{j=1}^p \phi_j L^j$  and  $\Theta(L) = 1 + \sum_{j=1}^q \theta_j L^j$  is the autoregressive and the moving average lag operator, respectively. When  $d \in (-1/2, 1/2)$  and the ARMA part has no common roots, all solutions are outside the unit circle. Thus the process is stationary with spectral density of the form:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{\Phi(e^{i\lambda})}{\Theta(e^{i\lambda})} \right| |1 - e^{i\lambda}|^{-2d} \quad \text{for } \lambda \in (-\pi, \pi]. \quad (3.9)$$

For  $\lambda \rightarrow 0$ , the process is of the form in (3.4) by Taylor expansion and the fact that the first part is attached to the aggregated exponentially decaying correlation function produced by the ARMA( $p, q$ ) part of the process, and hence satisfies the requirement of being slowly varying at zero. Note that  $(1 - L)^{d+1} = (1 - L)(1 - L)^d$ , and therefore, for  $d \in (1/2, 3/2)$  the increment series is described by a similar ARFIMA process, but with  $\tilde{d} = d - 1$ . This makes a natural translation between the stationary and non-stationary limit, and the non-stationary series can be dealt with by considering the first differenced series.

The special case with  $p = q = 0$  is of great importance, since it is closely connected to the continuous fractional Gaussian noise (fGn) process. In fact, it converges in distribution when time steps tend to zero. This makes these models compatible with the rough stochastic volatility suggested in recent literature (Gatheral, Jaisson and Rosenbaum, 2018), and a scaling similar to fractional Gaussian noise with  $H = d + \frac{1}{2} \in (0, 1/2)$  can also be handled by the corresponding ARFIMA( $0, d, 0$ ) process. At the same time, it makes a natural suggestion to move to the stationary long memory setup with  $d \in (0, 1/2)$  for realized variance, in cases where the Hurst index is estimated below 0.

In the following section we will discuss how to extract information on the memory parameter, and thus controlling whether it is desirable to consider an anti-persistent first differenced series or whether a long-memory setup is preferable.

### 3.2.3 Estimation of the memory parameter

For estimation of the memory parameter,  $d$ , described above, there are two paths to walk. The first is a parametric estimation using for example maximum likelihood (e.g. Sowell (1992)), however this restricts us to fixed model specifications, and has the undesirable consequence of being inconsistent for other model structures. The alternative is a semi-parametric estimation based on the form of either the covariance function or the spectral density presented above. While the first decays like a power-law function near infinity, the latter is easier to work with in practice, since the power-law structure is available near zero. Therefore, most of the semi-parametric approaches take advantage of the empirical analogue to the spectral density, namely the periodogram, and its relation to  $d$ . The periodogram is defined by:

$$I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t e^{-it\lambda_j} \right|^2, \quad (3.10)$$

and is evaluated in the Fourier frequencies,  $\lambda_j = 2\pi j/T$ . There exists different types of estimators on this form. Here, the focus is on the GPH estimator developed by Geweke and Porter-Hudak (1983) and Robinson (1995).<sup>3</sup> It uses the long memory condition to construct a linear relation between log-periodogram and the log-frequencies. Accounting for the use of the periodogram rather than the spectral density results in the following relation (see e.g. Geweke and Porter-Hudak (1983)):

$$\log(I(\lambda_j)) = c - 2d \log(2 \sin(\lambda_j/2)) + e_j \quad \text{for } \lambda_j \rightarrow 0,$$

where  $e_j$  is an error term. In practice, ordinary least squares (OLS) regression is applied using observations for  $m$  frequencies.  $m$  is called the bandwidth of the estimator, and should depend on  $T$ , however tending to infinity at a slower rate. It is important to select the bandwidth cleverly, due to trade-off between the bias and variance of the estimator. Also, since level-shifts or non-stationarity in general implies a spike in the periodogram at zero, avoiding the use of the spectral density evaluated here will circumvent issues regarding consistency in that case. Finally, the estimator yields consistent estimates even for  $d \in (1/2, 1)$ , and thus the covariance stationarity is not a strictly necessary assumption for the procedure to work, but must at least be satisfied for the differenced series.

Note that there are a lot of challenges using the semi-parametric approach. The presence of different short memory parts can disrupt the estimation, and level shifts

---

<sup>3</sup>Alternatively to this, the local Whittle estimator is applicable (Kunsch, 1987). In the empirical analysis similar results are found using this estimator.

can be mistaken for memory, or vice-versa. Moreover, a lot of data is necessary to obtain accurate results, making it impossible to re-estimate the parameter in small rolling windows, often considered ideal for forecasting purposes. However, the parameter does still tell a lot about the memory of the process, and hence is useful for indications regarding the stationarity in models with strong memories. Due to the tricky practice of the estimation procedure, we suggest applying an approximation of this model for forecasting purposes.

### 3.2.4 An approximation of infinite moving average structures

For forecasting purposes, models of realized variance should be able to reproduce the memory properties similar to processes with  $d \in (0, 1)$ . We suggest to take advantage of the fact that stationary Gaussian processes under mild conditions on the spectral density have a moving average representation, which makes it possible to write them as moving average processes with infinite order of lags,  $MA(\infty)$  (see e.g. Karhunen (1950)). The memory parameter of log-realized variance is usually estimated very close to the non-stationary limit of  $d = 1/2$ . Whenever  $d > 1/2$ , the first differenced series is stationary and has a moving average representation. In this case we suggest an  $ARIMA(0, 1, \infty)$  model for log-realized variance. While it is of course not possible to estimate an infinite number of moving average parameters in finite data, a truncation is needed, and the framework suggested here is an  $ARIMA(0, 1, q)$  model:

$$(1 - L)\log(RV_t) = \Theta(L)\varepsilon_t \quad \text{for } t = 1, \dots, T, \quad (3.11)$$

where  $(\varepsilon_t)_{t=1, \dots, T}$  is i.i.d. Gaussian noise with variance  $\sigma_\varepsilon^2$ , and the lag-operator for the moving average part is given by

$$\Theta(L) = 1 + \sum_{i=1}^q \theta_{t-i} L^i. \quad (3.12)$$

At first sight, the truncation seems rather restrictive, but if log-realized variance has memory parameter  $d \in (1/2, 1)$ , the increment series becomes anti-persistent and exhibits short memory. Therefore, only a couple of lags are necessary to provide a good approximation of its autocorrelation function. Finally, the ARIMA model is easy to estimate by linear regression or by maximum-likelihood techniques.

From an intuitive viewpoint, non-stationary log-realized variance seems rather strange. However, remember that in the setting explained here, it is caused by infinite strong memory of the past, which is maybe not so strange after all. In any case, this setup has favorable properties when it is used for discrete-time forecasting of realized variance.



### 3.3 Forecasting procedure

The goal of our procedure is to produce out-of-sample forecasting for data series on daily realized variance of index returns. Forecasts are produced by the proposed ARIMA(0, 1,  $q$ ) model for log-realized variance as well as some benchmark models. For forecast horizons longer than one day, the quantity of interest becomes the average realized variance over the forecast horizon of  $h$  days. That is, we want to predict:

$$\widehat{RV_{t+h}} = \frac{1}{h} \sum_{j=1}^h RV_{t+h}, \quad (3.13)$$

and we define the forecast of this quantity as the expected average value,

$$\widehat{\widehat{RV_{t+h}}} = \mathbb{E}_t \left[ \frac{1}{h} \sum_{j=1}^h RV_{t+h} \right] = \frac{1}{h} \sum_{j=1}^h \widehat{RV_{t+h}}, \quad (3.14)$$

where  $\mathbb{E}_t[\cdot]$  denotes the expected value conditional on time  $t$ . The ARIMA(0, 1,  $q$ ) model and the benchmark models have the common feature of modeling log-realized variance by a Gaussian process, and therefore, it is necessary to transform the predictions into forecasts of realized variance. For normally distributed stochastic variables, Jensen's Inequality imposes a convexity adjustment, such that

$$\widehat{RV_{t+h}} := \mathbb{E}_t [RV_{t+h}] = \exp \left( \mathbb{E}_t [\log(RV_{t+h})] + \frac{1}{2} \text{Var}_t [\log(RV_{t+h})] \right). \quad (3.15)$$

All components can be calculated in explicit form when parameters of the model are well-known, see Appendix 3.A.1 for further details. In practice, however, these are unknown, and are substituted for parameter estimates. The estimation procedure is applied to a rolling window containing data from the past  $w$  trading days.

#### 3.3.1 Forecast evaluation

The forecast performances are evaluated by the use of loss functions. Therefore, the results are sensitive to the shape of the loss function and hence, this choice is important. Here, we consider Mean Squared Error (MSE) and Quasi-Likelihood (QLIKE) loss functions,

$$\text{MSE:} \quad L(\hat{\sigma}^2, h) = (\hat{\sigma}^2 - h)^2, \quad (3.16)$$

$$\text{Q-LIKE:} \quad L(\hat{\sigma}^2, h) = \frac{\hat{\sigma}^2}{h} - \log \left( \frac{\hat{\sigma}^2}{h} \right) - 1. \quad (3.17)$$

Both of these provide a ranking among forecasts that is also true for the spot variance process, when considering a conditionally unbiased proxy of this term, such as realized variance (see Patton (2011)). The two loss functions are, however, of quite different of nature. MSE only depends on the forecast error, and is symmetric around

the volatility proxy. Opposite, the QLIKE loss function relies on the standardized forecast residuals. The latter makes Q-LIKE less sensitive to forecast errors in high volatility regimes. Moreover, it is asymmetric and penalizes too low forecasts the most. A feature that is appropriate when it is more serious to underestimate than overestimate, in this case the variance process. Evaluations for both loss functions are reported, so that the results are easily comparable with earlier literature.

Even though the forecasting exercise does not explain which model fits the data best in general, it tells us which models work the best for forecasting in this sample. For conclusions of this kind, we consider the model confidence set (MCS) by Hansen, Lunde and Nason (2011), and due to embedded models in our model set, it is implemented by considering the bootstrap approach using standard settings as explained in the original paper.

### 3.3.2 Benchmark models

We consider two different and popular forecasting models as benchmark to evaluate the pros and cons by forecasting using the ARIMA model. First, and the most obvious benchmark model based on our discussion in the theoretical part is the ARFIMA(0,  $d$ , 0) model described in (3.7). Since log-realized variance is also normally distributed in this framework, the estimation procedure is similar to that described above. The fractal index,  $d$ , is estimated using the GPH estimator.

Secondly, the popular short-memory log-HAR model is used as benchmark. Due to its high-order (but restricted) autoregressive structure, it is easy to estimate and generates a slowly decaying autocorrelation function by aggregation of short memory terms. Its structure is given by,

$$\log RV_{t+h} = c + \beta^{(1)} \log RV_t + \beta^{(5)} \log RV_t^{(5)} + \beta^{(22)} \log RV_t^{(22)} + \varepsilon_{t+1|t+h}, \quad (3.18)$$

where  $RV_t^{(J)} = \sum_{j=1}^J RV_{t+1-j}$ , corresponding to the average realized variance one day, one week and one month prior to day  $t$ . The error,  $(\varepsilon_{t+1|t+h})$  is i.i.d. Gaussian noise with variance  $\sigma_\varepsilon^2$ . The parameters  $(c, \beta^{(1)}, \beta^{(5)}, \beta^{(22)})$  are estimated by a simple OLS regression. On the evidences in earlier literature, we conduct direct forecasting for this model, re-estimating the parameters  $h$  times to do the  $h$  time ahead forecast of  $\widetilde{RV_{t+h}}$ .

## 3.4 Empirical analysis

In this section we present results of the proposed ARIMA model for forecasting in empirical data. We consider 5-minute realized variance, which are downloaded from <https://realized.oxford-man.ox.ac.uk/>, where different realized measures are available for 31 different indexes. We focus on four major of these from all around the world, namely Nasdaq 100, S&P 500, FTSE 100, and Nikkei 225. The sampling period runs in opening hours at the exchange from January 2000 until February 14, 2020,

**Table 3.1:** Descriptive statistics.

code	mean	std. dev.	skewness	kurtosis	ADF	KPSS	$\hat{d}_{GPH}$
<i>Panel A: Realized variance.</i>							
FTSE 100	0.029	0.068	17.933	546.218	-10.669 (0.000)	0.651 (< 0.01)	0.474
Nasdaq 100	0.030	0.055	7.455	94.302	-8.123 (0.000)	1.080 (< 0.01)	0.438
Nikkei 225	0.025	0.041	8.550	112.099	-9.106 (0.000)	0.330 (< 0.01)	0.483
S&P 500	0.026	0.060	12.061	264.152	-6.592 (0.000)	0.720 (< 0.01)	0.5836
<i>Panel B: Log realized variance.</i>							
FTSE 100	-4.161	0.994	0.552	3.569	-7.284 (0.000)	1.416 (< 0.01)	0.534
Nasdaq 100	-4.182	1.093	0.432	3.006	-6.433 (0.000)	1.243 (< 0.01)	0.548
Nikkei 225	-4.161	0.937	0.246	3.355	-7.891 (0.000)	0.611 (< 0.01)	0.514
S&P 500	-4.383	1.136	0.339	3.299	-6.825 (0.000)	1.273 (< 0.01)	0.574
<i>Panel C: Increments of log-realized variance.</i>							
FTSE 100	-0.000	0.690	0.068	4.423	-34.346 (0.000)	0.003 (> 0.10)	-0.472
Nasdaq 100	-0.001	0.593	0.158	3.945	-27.181 (0.000)	0.003 (> 0.10)	-0.455
Nikkei 225	-0.000	0.615	0.381	5.209	-30.200 (0.000)	0.002 (> 0.10)	-0.495
S&P 500	-0.000	0.684	0.077	3.818	-32.506 (0.000)	0.003 (> 0.10)	-0.431

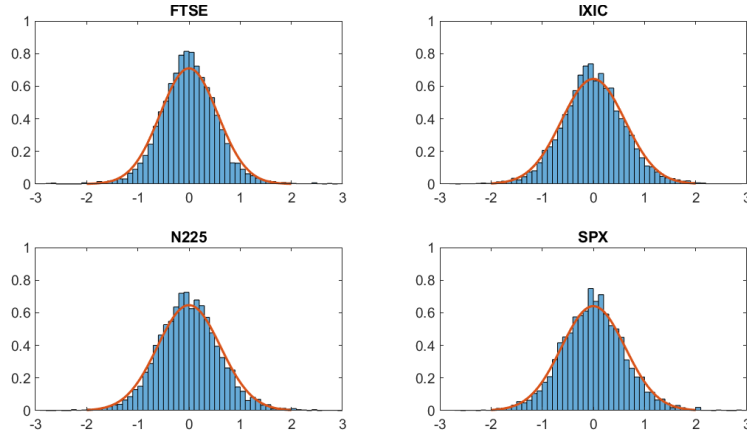
Note. Descriptive statistics for the four indexes and unit-root tests. P-values for ADF and KPSS test are reported in parenthesis.  $\hat{d}_{GPH}$  reports the GPH estimate of the fractal-index.

and the number of observations differs slightly between the indexes due to different number of bank holidays around the world, but is in all cases close to 5000.

Descriptive statistics for the realized variance, log-realized variance and increments of log-realized variance for each index are reported in Table 3.1. As expected,  $RV$  is extremely right skewed and leptokurtic, and thus very unlikely to be normally distributed for all indexes. On the other hand,  $\log(RV)$  and  $\Delta \log(RV)$  are statistically much more like the outcome of a normal distribution, but have small positive sample excess kurtosis and are only close to being symmetric. This is equivalent to earlier findings, for example by Andersen, Bollerslev, Diebold and Labys (2003). Due to the large amount of data, standard tests such as the Jarque-Bera test or the Kolmogorov-Smirnov test reject normality, however, a look at the histograms in Figure 3.1 confirms that the increment series of log-realized variance are not far from being normal, and based on this we find it acceptable to use normal log- $RV$  models.

Regarding the stationarity of the log-realized variance, the results are mixed. We report test statistics for Augmented Dickey-Fuller (ADF) test and KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992). For  $RV$  and  $\log(RV)$  series we allow for a constant and choose the lag length by the Bayesian Information Criterion (BIC), considering lag-length less than  $p_{max} = 12(T/100)^{1/4}$ .<sup>4</sup> It is a trade-off between too often rejection of a unit-root, if the lag-length is chosen too small, while the test loses power for a large number of lags. While the KPSS test rejects that  $RV$  and log- $RV$  is stationary on all relevant significance levels, the ADF test rejects the presence of a unit root. A rational explanation for this is that the process is neither stationary nor  $I(1)$ . That is for example the case for the fractional integration order,  $I(d)$ , with  $d \in (0.5, 1)$ . However, it could also be caused by the limitations of the unit-root tests. It is well-known that the ADF test rejects the null hypothesis of a unit root too often in the presence of a large negative moving average parameter, see for example Schwert (2002). Table 3.1 also

<sup>4</sup>This is the rule of thumb presented in Kwiatkowski et al. (1992).

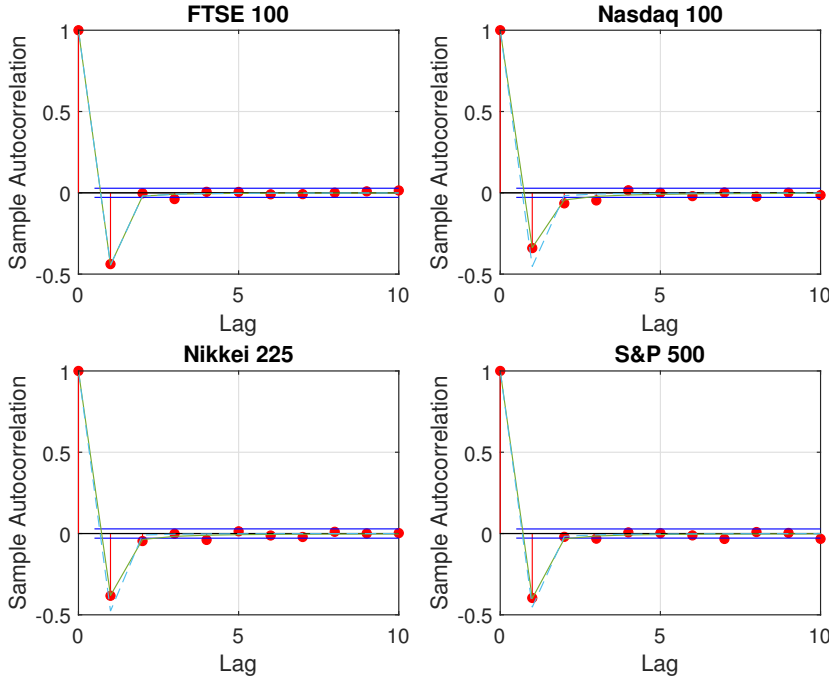


**Figure 3.1:** Histograms of increments of log-realized variance (a) FTSE 100, (b) Nasdaq 100, (c) Nikkei 225 and (d) S&P 500.

reports the GPH estimate of fractal order of integration (Geweke and Porter-Hudak, 1983) with bandwidth  $T^{0.8}$ .<sup>5</sup> For  $\log(RV)$ ,  $\hat{d}_{GPH} > 0.5$ , which is in the non-stationary region, and the long-memory property is meaningless. Instead it makes sense to investigate the first differenced series,  $\Delta \log(RV)$ , here  $\hat{d}_{GPH} \in (-0.5, -0.4)$ , suggesting that it behaves like an anti-persistent Gaussian process. Note also that the estimate of the first differenced series is not exactly equal to  $1 - d_{GPH}$  of the log-realized variance, this is due to small sample properties of the estimator, see Hurvich and Ray (1995) for a simulation study on this. So far we conclude that the log-realized series are likely non-stationary, therefore we consider the first order differenced series and apply the ARIMA(0, 1,  $q$ ) framework presented above.

The autocorrelation functions for  $\Delta \log(RV)$  is depicted in Figure 3.2, and for all indexes reveal a significant negative effect on the first lag which dissipates quite fast over the next couple of lags. Comparing the autocorrelation function to that of an ARFIMA(0,  $\hat{d}_{GPH}$ , 0) process shows that this pattern is almost equivalent to fractional models of this kind. The final take-away from this is the clear and early cut-off after the first couple of lags, suggesting that a moving average model could potentially fit well for small  $q$ . This leads to another subject, which must be handled. How many moving average lags are actually necessary to forecast realized variance well? Based on a visual inspection of the figures, only the first lag seems to be crucially significant, however from the theory above, the ARFIMA framework suggests that infinitely many lags are needed. We investigate this by an in-sample approach.

<sup>5</sup>In our case, the estimator is relatively robust to choice of bandwidth, however slightly decreasing in the exponent. We follow the standard in the forecasting literature, e.g. Andersen et al. (2003).



**Figure 3.2:** Autocorrelation function for increments of log-realized variance, and fitted autocorrelation function for  $\text{ARFIMA}(0, \hat{d}_{\text{GPH}}, 0)$  (dotted line).

### 3.4.1 Model selection

It is not clear which number of moving average lags,  $q$ , should be convenient for forecasting using the  $\text{ARIMA}(0, 1, q)$ , not even after watching the empirical autocorrelation functions. Even though this gave us a clear intuition about whether only a few lags can do the job, a clarifying in-sample study is done. In-sample estimation of the parameters are done, and hereafter the log-likelihood function is calculated together with Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The results are summarized in Table 3.2 and further details are available in Appendix 3.A.2. Since BIC penalizes a higher number of parameters more than AIC, it is not surprising that BIC chooses a model with a lower number of lags,  $q$ . The number chosen by BIC is quite close to what we might have guessed by looking at the autocorrelation function, however is larger than 1, and suggests that also significant information are available in a couple of additional lags. The suggestion of  $q$  by AIC points out that there is additional information to gather even in lags one month away, which is also in line with the ideas of the HAR model.

**Table 3.2:** In-sample model selection.

	FTSE 100	Nasdaq 100	Nikkei 225	S&P 500
$q_{BIC}$	3	6	4	7
$q_{AIC}$	18	26	24	25

*Note.* The table reports optimal number of moving average lags chosen by BIC and AIC respectively.

**Table 3.3:** One-day-ahead out-of-sample forecast evaluation.

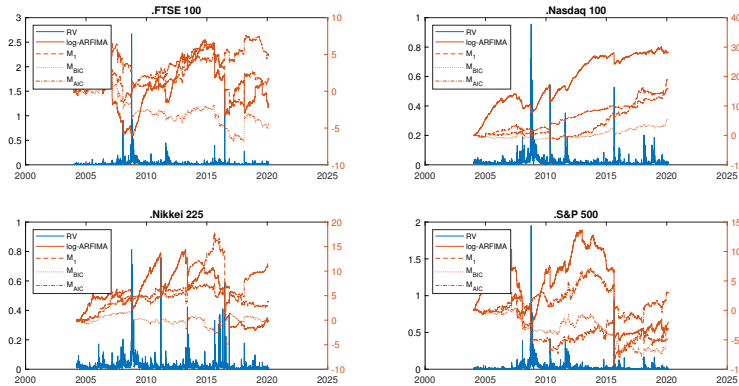
Model	FTSE 100		Nasdaq 100		Nikkei 225		S&P 500	
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
log-HAR	1.000	1.000	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	1.000
log-ARFIMA	1.011	0.998	1.043	1.045	<b>0.976</b>	1.000	1.049	0.996
$M_1$	<b>0.989</b>	1.002	1.009	1.030	1.010	1.005	1.012	0.998
$M_{BIC}$	0.993	<b>0.995</b>	<b>0.980</b>	1.009	0.976	1.001	<b>0.990</b>	<b>0.994</b>
$M_{AIC}$	1.003	1.006	1.036	1.025	0.985	1.014	1.048	1.004

*Note.* We report standardized average MSE and QLIKE. Boxes marked with light grey is included in the 75 percent MCS, while dark grey is also included in 95 percent MCS.

### 3.4.2 Forecasting results

In the out-of-sample forecasting exercise, we consider three different settings of the ARIMA(0, 1,  $q$ ) model with  $q \in \{1, q_{BIC}, q_{AIC}\}$ , respectively. The case, where  $q = 1$ , is considered on the basis of visual inspection of the autocorrelation function, while  $q_{BIC}$  and  $q_{AIC}$  are interesting, since the in-sample model inspection suggests that these models perform the best evaluated by BIC and AIC, respectively. They are compared to two benchmark models, namely the log-HAR model and the log-ARFIMA model. The forecast procedure is done using a rolling window of size  $w = 1000$  trading days for the parameter estimation, corresponding to approximately four years of data.<sup>6</sup> The results for the one-day-ahead forecasts are reported in Table 3.3. The first thing to notice is that there are no huge differences in forecast performance between any of the models, however, based on the MCS procedure, some of the models are doing better than others. The ARIMA(0, 1, 1) model generally performs worse compared to the other models, suggesting that even though the autocorrelation function seems only to be significant for the first lag, there is additional information to benefit from by using more lags. In contrast, the ARIMA(0, 1,  $q_{AIC}$ ) seems to include too many lags to perform well on the one-day-ahead forecast. The explanation could be that the estimates of higher order lag parameters are relatively small, and even though the extra lags are useful for in-sample fitting, they do not hold any useful information about the nearest future. On the other hand, in most cases the ARFIMA(0, 1,  $q_{BIC}$ ) model outperforms the other models in terms of both MSE and QLIKE. It is almost consistently in the 95% MCS, and often performs better than the

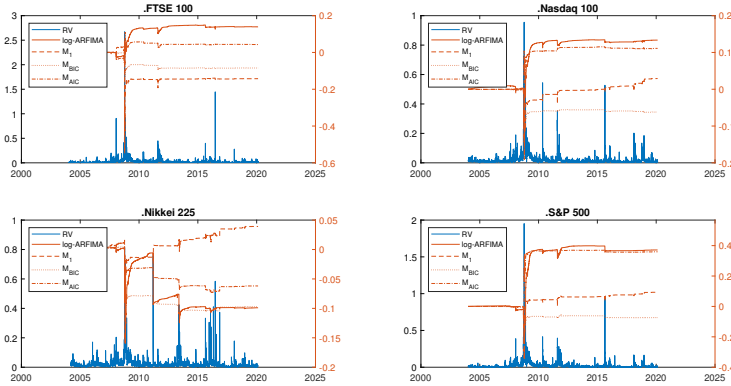
<sup>6</sup>We tried different window sizes ranging from 250 days up to 1500 days with similar ranking between the models.



**Figure 3.3:** RV (blue line) and cumulative QLIKE losses for each model minus cumulative QLIKE-losses for log-HAR model.

spurious long-memory log-ARFIMA model, while only including a small number of lags. Before drawing any definitive conclusions, we make a visual inspection of the cumulative loss function to get an impression on which days make the differences. For the cumulative QLIKE loss function depicted in Figure 3.3 a couple of days impact the performance a lot. Specifically, there is a huge negative effect for the HAR model compared to the others on specific days, for example the day of the 2015 flash-crash for the S&P 500 index. The commonalities among these days are huge increases in the level of volatility and immediately hereafter the variance returns. Due to the structure of the HAR model, the additional loss is caused by the large positive contribution to both daily, weekly and monthly average lags these days, all positive effects which drag the forecast up. In the other models, the antipersistent structure of log-realized variance increments counterbalance the effect of the fast increase, and the forecast is impacted less, however still quite a lot. For the MSE the story is a bit different. As explained in the theory above, MSE is largely affected by days with large absolute deviation from the true realized variance process and therefore large differences in forecasting performance are almost solely accumulated in periods with high variance. This also explains the shape of the cumulative MSE losses depicted in Figure 3.4. There are long periods with almost constant loss between the models, and therefore MSE is a good measure for the ability to forecast in specifically these periods. We conclude that the one-day-ahead forecasts using the ARIMA models, despite of the simple structure, predict at least as well as the popular benchmark models and often better, no matter the loss function.

The ability to forecast realized variance multi-periods ahead is also investigated, and we consider forecast horizons of one week, two weeks and one month. The results are collected in Table 3.4. The performance here is more mixed. Comparing



**Figure 3.4:** RV (blue line) and cumulative MSE losses for each model minus cumulative MSE-losses for log-HAR model.

the ARIMA models, it is clear that for long horizon ( $h \in \{10, 22\}$ ), the forecast based on  $q_{AIC}$  is often doing best. A logical explanation is that extra lags are needed to fit the memory far out, and hence are more important for distant forecasts. In many cases the ARIMA model with  $q_{BIC}$  or  $q_{AIC}$  is better than the corresponding ARFIMA model, suggesting that this long memory approximation is a good candidate for forecasting in such data series. One reason could be that the pure ARFIMA(0,  $\hat{d}_{GPH}$ , 0) needs for example additional moving average lags to fit the data well. This change however will complicate the structure further, and the ARIMA framework is preferable. Lastly, the strong memory ARIMA framework and the HAR are on average performing equally well.

### 3.5 Conclusion

We have proposed to model log-realized variance using an ARIMA(0, 1,  $q$ ) model with a view on forecasting realized variance. The motivation was to find an easy-to-implement way to do forecasting in data that exhibits strong memory properties in the non-stationary region.

In the empirical part, we consider index data for four major indexes, and use the ARIMA model to do forecasting. The performance is benchmarked against the fractional log-ARFIMA process and the log-HAR model. We show that choosing  $q$  by the Bayesian Information Criterion results in a good trade-off between lack of important information and too much useless information, and using this setting, the ARIMA model produces at least as good one-day ahead forecasts compared to the popular benchmark models.

Increasing the forecast horizon yields mixed results. For all horizons,  $q_{AIC}$  moving



**Table 3.4:** Multiple-day-ahead out-of-sample forecast evaluation.

Model	FTSE 100		Nasdaq 100		Nikkei 225		S&P 500	
	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
<i>Panel A: h = 5.</i>								
log-HAR	<b>1.000</b>	1.000	<b>1.000</b>	<b>1.000</b>	1.000	1.000	<b>1.000</b>	1.000
log-ARFIMA	1.032	1.010	1.108	1.092	0.983	<b>0.982</b>	1.137	<b>0.999</b>
$M_1$	1.062	<b>0.986</b>	1.160	1.064	1.133	1.024	1.169	1.009
$M_{BIC}$	1.012	0.988	1.008	1.022	<b>0.970</b>	1.014	1.006	1.015
$M_{AIC}$	1.021	0.987	1.072	1.020	0.973	0.987	1.106	1.012
<i>Panel B: h = 10.</i>								
log-HAR	<b>1.000</b>	1.000	<b>1.000</b>	<b>1.000</b>	1.000	1.000	<b>1.000</b>	1.000
log-ARFIMA	1.053	1.012	1.155	1.087	0.987	<b>0.980</b>	1.194	<b>0.974</b>
$M_1$	1.223	0.992	1.428	1.088	1.269	1.047	1.407	0.999
$M_{BIC}$	1.067	0.988	1.049	1.024	0.994	1.038	1.060	1.014
$M_{AIC}$	1.025	<b>0.973</b>	1.076	1.020	<b>0.961</b>	0.983	1.164	1.006
<i>Panel C: h = 22</i>								
log-HAR	<b>1.000</b>	1.000	<b>1.000</b>	<b>1.000</b>	1.000	1.000	<b>1.000</b>	1.000
log-ARFIMA	1.013	0.981	1.183	1.079	0.995	<b>0.966</b>	1.168	<b>0.953</b>
$M_1$	1.750	1.029	2.395	1.210	2.039	1.119	2.210	1.089
$M_{BIC}$	1.242	0.992	1.138	1.048	1.222	1.083	1.171	1.019
$M_{AIC}$	1.006	<b>0.943</b>	1.095	1.037	<b>0.994</b>	1.016	1.146	1.040

*Note.* We report average MSE and QLIKE. Boxes marked with light grey is included in the 75 percent MCS, while dark grey is also included in 95 percent MCS.

average lags are often preferable over  $q_{BIC}$ . In general, this model often beats the ARFIMA model, and even though it might be caused by missing moving average lags, the ARIMA framework is preferred by its simple estimation procedures.

For future research it might also be of great interest to investigate the performance of these simple ARIMA models on single stock level. Here, the log-normality assumption might be less pronounced, and it will thus be interesting to relax this assumption.

### 3.A Appendix

#### 3.A.1 Convexity-adjustment in ARIMA

**Proposition 3.A.1.** *If log-realized variance follows an ARIMA(0, 1, q), the conditional mean and variance is at the form:*

$$\mathbb{E}[\log(RV_{t+h})|\mathcal{F}_t] = \log(RV_t) + \sum_{j=1}^h \sum_{i=1 \vee j-1}^q \theta_i \varepsilon_{t+j-1-i} \quad (3.19)$$

and the conditional variance is

$$\mathbb{V}ar[\log(RV_{t+h})|\mathcal{F}_t] = \sum_{j=1}^h \left( 1 + \sum_{i=1}^{q \wedge (j-1)} \theta_i^2 \right) \sigma_\varepsilon^2 + 2 \sum_{j=1}^{h-1} \sum_{k=j+1}^h \left( \sum_{i=1}^{q-(k-j)} \theta_i \theta_{i+k-j} + \theta_{|k-j|} \right) \sigma_\varepsilon^2. \quad (3.20)$$

*Proof.* Let  $x_t$  be a random variable following an ARIMA(0,1,q) model,

$$x_{t+1} = x_t + \sum_{i=1}^q \theta_i \varepsilon_{t+1-i} + \varepsilon_{t+1}. \quad (3.21)$$

The conditional mean follows directly by linearity of the expectation operator in this structure. To calculate the variance term, we are interested in,

$$\begin{aligned} \mathbb{V}ar(x_{t+h}|\mathcal{F}_t) &= \mathbb{V}ar\left(\sum_{j=1}^h \Delta x_{t+j}|\mathcal{F}_t\right) \\ &= \mathbb{V}ar\left(\sum_{j=1}^h \Delta x_{t+j}|\mathcal{F}_t\right) + 2 \sum_{1 \leq j < k \leq h} \mathbb{C}ov(\Delta x_{t+j}, \Delta x_{t+k}|\mathcal{F}_t). \end{aligned} \quad (3.22)$$

Notice that  $\Delta x_t$  is an moving average process, and therefore

$$\mathbb{V}ar(\Delta x_{t+j}|\mathcal{F}_t) = \mathbb{V}ar\left(\sum_{i=1}^q \theta_i \varepsilon_{t+j-i} + \varepsilon_{t+j}|\mathcal{F}_t\right) = \left(1 + \sum_{i=1}^{q \wedge (j-1)} \theta_i^2\right) \sigma_\varepsilon^2 \quad (3.23)$$

and for  $k > j$ ,

$$\begin{aligned} \mathbb{C}ov(\Delta x_{t+j}, \Delta x_{t+k}|\mathcal{F}_t) &= \mathbb{C}ov\left(\sum_{i=1}^q \theta_i \varepsilon_{t+j-i} + \varepsilon_{t+j}, \sum_{l=1}^q \theta_l \varepsilon_{t+k-l} + \varepsilon_{t+k}|\mathcal{F}_t\right) \\ &= \sum_{i=1}^q \sum_{l=1}^q \theta_i \theta_l \mathbb{C}ov(\varepsilon_{t+j-i}, \varepsilon_{t+k-l}|\mathcal{F}_t) + \sum_{l=1}^q \theta_l \mathbb{C}ov(\varepsilon_{t+j}, \varepsilon_{t+k-l}|\mathcal{F}_t) \\ &= \begin{cases} \left(\sum_{i=1}^{q-(k-j)} \theta_i \theta_{i+k-j} + \theta_{|k-j|}\right) \sigma_\varepsilon^2 & \text{for } k-j < q \\ 0 & \text{else.} \end{cases} \end{aligned} \quad (3.24)$$

Inserting this in (3.22) finishes the proof.  $\square$   $\square$

### 3.A.2 Model selection

In Table 3.5, AIC and BIC for a bunch of different ARIMA models with varying number of moving average lags, are shown.

**Table 3.5:** In sample model selection.

q	FTSE 100		Nasdaq 100		Nikkei 225		S&P 500	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
0	2.0962	2.0975	1.7920	1.7933	1.8669	1.8682	2.0786	2.0799
1	1.7433	1.7459	1.5944	1.5970	1.6135	1.6162	1.8231	1.8256
2	1.7370	1.7408	1.5721	1.5760	1.5962	1.6002	1.8144	1.8183
3	1.7332	<b>1.7383</b>	1.5645	1.5696	1.5906	1.5959	1.8082	1.8134
4	1.7334	1.7398	1.5635	1.5700	1.5859	<b>1.5925</b>	1.8066	1.8130
5	1.7330	1.7407	1.5605	1.5683	1.5857	1.5937	1.8035	1.8113
6	1.7323	1.7413	1.5570	<b>1.5661</b>	1.5845	1.5937	1.7996	1.8086
7	1.7324	1.7427	1.5561	1.5664	1.5843	1.5949	1.7978	<b>1.8081</b>
8	1.7327	1.7442	1.5549	1.5665	1.5847	1.5966	1.7982	1.8098
9	1.7327	1.7456	1.5550	1.5680	1.5849	1.5981	1.7982	1.8111
10	1.7317	1.7458	1.5552	1.5695	1.5850	1.5996	1.7982	1.8124
11	1.7295	1.7450	1.5556	1.5711	1.5849	1.6008	1.7984	1.8139
12	1.7296	1.7463	1.5536	1.5705	1.5853	1.6025	1.7969	1.8137
13	1.7296	1.7476	1.5540	1.5721	1.5854	1.6040	1.7967	1.8148
14	1.7300	1.7493	1.5541	1.5736	1.5852	1.6051	1.7970	1.8164
15	1.7300	1.7506	1.5544	1.5751	1.5853	1.6066	1.7974	1.8181
16	1.7298	1.7517	1.5539	1.5759	1.5851	1.6077	1.7970	1.8190
17	1.7296	1.7528	1.5524	1.5757	1.5854	1.6093	1.7969	1.8202
18	<b>1.7294</b>	1.7539	1.5526	1.5772	1.5853	1.6105	1.7970	1.8216
19	1.7297	1.7554	1.5526	1.5784	1.5848	1.6114	1.7973	1.8232
20	1.7301	1.7571	1.5522	1.5794	1.5849	1.6127	1.7974	1.8246
21	1.7300	1.7584	1.5526	1.5810	1.5829	1.6121	1.7973	1.8258
22	1.7294	1.7590	1.5527	1.5824	1.5822	1.6127	1.7976	1.8274
23	1.7295	1.7604	1.5526	1.5836	1.5822	1.6141	1.7964	1.8275
24	1.7299	1.7621	1.5530	1.5853	<b>1.5813</b>	1.6145	1.7961	1.8284
25	1.7303	1.7638	1.5528	1.5864	1.5816	1.6161	<b>1.7957</b>	1.8293
26	1.7306	1.7653	<b>1.5520</b>	1.5869	1.5814	1.6172	1.7960	1.8309
27	1.7297	1.7658	1.5524	1.5886	1.5817	1.6189	1.7963	1.8325
28	1.7301	1.7674	1.5525	1.5900	1.5820	1.6205	1.7967	1.8341
29	1.7305	1.7691	1.5524	1.5912	1.5824	1.6223	1.7970	1.8358
30	1.7308	1.7707	1.5527	1.5928	1.5829	1.6240	1.7973	1.8373
31	1.7311	1.7723	1.5530	1.5944	1.5830	1.6254	1.7972	1.8386

*Note.* The table reports Akaike and Bayesian Information Criterion per observation based on the full sample from January 2001 to February 2020.

### 3.2 References

- Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* , 885–905.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 253–280.
- Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–37.
- Bennedsen, M., Lunde, A., Pakkanen, M., 2017. Decoupling the short-and long-term behavior of stochastic volatility. *arXiv:1610.00332* .
- Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-memory processes probabilistic properties and statistical methods .
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Cont, R., Tankov, P., 2004. Financial modelling with jump processes. Chapman & .
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. *Econometric Reviews* 27, 46–78.
- Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300, 463–520.
- Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society* , 987–1007.
- Engle, R.F., Patton, A.J., 2001. What good is a volatility model? *Quantitative Finance* 1, 237–245.
- Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. *Quantitative Finance* 18, 933–949.

- Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.
- Granger, C.W.J., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. *Journal of Empirical Finance* 11, 399–421.
- Granger, W.W.J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.
- Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. *Econometrica* 79, 453–497.
- Hosking, J.R.M., 1981. Fractional differencing. *Biometrika* 68, 165–165.
- Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* 16, 17–41.
- Izzeldin, M., Hassan, M.K., Pappas, V., Tsionas, M., 2019. Forecasting realised volatility using arfima and har models. *Quantitative Finance* , 1–12.
- Karhunen, K., 1950. Über die struktur stationärer zufälliger funktionen. *Arkiv för Matematik* 1, 141–160.
- Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.
- Kunsch, H.R., 1987. Statistical aspects of self-similar processes, in: *Proceedings of the First World Congress of the Bernoulli Society*, 1987, pp. 67–74.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159–178.
- Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36, 270–296.
- Oomen, R.C.A., 2001. Using high frequency stock market index data to calculate, model & forecast realized return variance. *European Univ., Economics Discussion Paper* .
- Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
- Poon, S., Granger, C., 2005. Practical issues in forecasting volatility. *Financial Analysts Journal* 61, 45–56.

- Poon, S., Granger, C.W.J., 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41, 478–539.
- Robinson, P.M., 1995. Log-periodogram regression of time series with long range dependence. *The annals of Statistics* , 1048–1072.
- Schwert, G.W., 1987. Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics* 20, 73–103.
- Schwert, G.W., 2002. Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics* 20, 5–17.
- Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53, 165–188.

## Declaration of co-authorship

Full name of the PhD student: Anine Eg Bolko

This declaration concerns the following article/manuscript:

Title:	Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures
Authors:	Anine Eg Bolko, Kim Christensen, Mikko Pakkanen & Bezirgen Veliyev

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☐ In preparation ☒

If published, state full reference:

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

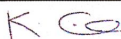
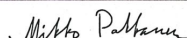
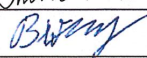
No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

- A. Has essentially done all the work
- B. Major contribution
- C. Equal contribution
- D. Minor contribution
- E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	C
3. Involvement in the experimental work/clinical studies/data collection	C
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	C
6. Finalization of the manuscript and submission	C

### Signatures of the co-authors

Date	Name	Signature
29-09-2020	Kim Christensen	
29-Sep 2020	Mikko Pakkanen	
29-09-2020	Bezirgen Veliyev	

In case of further co-authors please attach appendix

Date: 30-09-2020

  
Signature of the PhD student