Non-parametric estimation of stochastic volatility models: spot volatility, leverage and vol-of-vol.

Four essays on asymptotic error distributions, finite-sample properties and empirical applications.

SCUOLA NORMALE SUPERIORE

Thesis submitted for the degree of Ph.D. in Financial Mathematics

A.Y. 2020-2021

Candidate:
Giacomo Toscano

Advisor:
Prof. Maria Elvira Mancino

Co-Advisors:
Prof. Maria Cristina Recchioni
Prof. Simone Scotti

ABSTRACT

This thesis contains four essays on non-parametric estimators of the spot volatility, the leverage and the volatility-of-volatility. In particular, the focus of this thesis is on the study of the asymptotic properties of the estimators, the optimization of their finite-sample performance and the use of the resulting estimates in empirical applications.

Specifically, in Chapter 2 we prove a central limit theorem for the estimator of the integrated leverage based on the Fourier method of Malliavin and Mancino (2009), showing that it reaches the optimal rate of convergence and a smaller variance with respect to different estimators based on a pre-estimation of the instantaneous volatility. Then, we exploit the availability of efficient Fourier-based estimates of the integrated leverage to show, using S&P500 prices over the period 2006-2018, that adding an extra term which accounts for the leverage effect to the Heterogeneous Auto-Regressive (HAR) volatility model by Corsi (2009) increases the explanatory power of the latter.

In Chapter 3 we study the sensitivity of the leverage process to changes of the price and the volatility. In particular, under the Constant Elasticity of Variance (CEV) model by Beckers (1980), which is explicitly designed to capture leverage effects, we find that the derivatives of the leverage with respect to the log-price and the volatility can be expressed as the ratio of quantities that can be consistently estimated from sample prices, that is, as the ratio of the price-leverage covariation and, respectively, the volatility and the leverage. From the financial standpoint, this suggests that the price-leverage covariation may be interpreted as a gauge of the responsiveness of the leverage to the arrival of new information that causes changes in the price or the volatility. Additionally, we also find that the price-leverage covariation is equal to twice the vol-of-vol under the CEV model, thereby suggesting that the responsiveness of the leverage (i.e., the price-leverage covariation) is proportional to the amount of uncertainty about risk (i.e., the vol-of-vol). After reconstructing the trajectories of the volatility, the leverage, the vol-of-vol and the price-leverage covariation through the Fourier methodology by Malliavin and Mancino (2009), we provide empirical evidence supporting this financial interpretation of the price-leverage covariation in a model-free setting, using 1-second S&P500 prices over the period March, 2018-April, 2018.

In Chapter 4, we perform an analytical study to identify the sources of the finite-sample bias that typically plagues the simplest and most natural vol-of-vol estimator, the Pre-estimated Spot-variance based Realized Variance (PSRV) by Barndorff-Nielsen and Veraart (2009). Based on the full knowledge of its analytical expression, we show that the finite-sample bias of the PSRV may be substantially reduced by allowing for the overlap of consecutive local windows to pre-estimate the spot variance. In particular, we provide a feasible analytical rule for the bias-optimal selection of the length of local windows when the volatility is a process in the Chan, Karolyi, Longstaff and Sanders (CKLS) class (see Chan et al. (1992)) and show that selections based on this analytical rule match some selections prescribed in the literature, based on simulations.

In Chapter 5, we exploit efficient Fourier estimates of the path of the volatility to empirically investigate the functional link between the latter and the variance swap rate. Specifically, using S&P500 data over the period 2006-2018, we find overwhelming empirical evidence supporting the affine link analytically found by Kallsen et al. (2011) in the context of exponentially affine stochastic volatility models. Additionally, based on tests performed on yearly subsamples, we find that exponentially mean-reverting variance models provide a good fit during periods of extreme volatility, while polynomial models, introduced in Cuchiero (2011), are suited for years characterized by more frequent price jumps. These empirical results are confirmed when replacing Fourier estimates of the spot volatility with realized local estimates.

Chapter 6 concludes, summarizing the main findings of the thesis.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank professor M.E. Mancino for her invaluable guidance, I will be forever in her debt. Also, I am grateful to professor M.C. Recchioni and professor S. Scotti for their constant availability and insightful advice, I also owe them a lot.

Further, I would like to thank the coordinator of the Ph.D. in Financial Mathematics at Scuola Normale, professor S. Marmi, for funding my research. I am also thankful to professors G. Bormetti, F. Corsi, F. Lillo, M.E. Mancino, S. Marmi and D. Tantari for their lectures during the first year of the Ph.D. program.

I also thank my colleagues A. Brini, F. Cordoni and I. Raffaelli for their friendship and moral support. Lastly, I am grateful to my parents for having always encouraged me during my studies.

CONTENTS

LIST OF PUBLICATIONS

Published articles:

- Mancino, M.E., Scotti, S. and Toscano, G. (2020). Is the variance swap rate affine in the spot variance? Evidence from S&P500 data. *Applied Mathematical Finance*, 27(4): 288-313.

Articles under review:

- Recchioni, M.C. and Toscano, G. (2020). Bias-optimal vol-of-vol estimation: the role of window overlapping. *Revise & Resubmit, Decisions in Economics and Finance*.

- Mancino, M.E. and Toscano, G. (2020). Rate-efficient asymptotic normality for the Fourier estimator of the leverage process. *Submitted, Statistics and Its Interface*.

Articles to be submitted:

- Toscano, G. (2021). The price-leverage covariation as a measure of the response of the leverage effect to price and volatility changes: empirical evidence. Available at: https://ssrn.com/abstract=3776771.

1. OVERVIEW OF THE THESIS

Introduction: from volatility to second-order (and third-order) quantities

Based on the seminal work by Jacod (see, e.g., Jacod (1994)[1] and Jacod (1997)), the literature on high-frequency financial econometrics has flourished in the last three decades, as high-frequency financial data have become progressively more accessible to scholars and practitioners. In a nutshell, the objective of high-frequency financial econometrics is to obtain non-parametric estimates of financially relevant quantities, primarily the volatility[2], from high-frequency asset prices. These estimates are in turn used for various applications, e.g., forecasting, model calibration, derivatives pricing, hedging and risk management.

More specifically, the literature on high-frequency financial econometrics aims at the development of statistically efficient estimators, with a focus not only on their asymptotic properties, that is, their consistency and error distribution as the number of available price observations on a fixed interval goes to infinity (the so-called *infill asymptotics*), but also on their finite-sample sample performance, that is, their efficient implementation in practical applications. These estimators are labeled "non-parametric" because, for their asymptotic properties to hold, they only require the data-generating process to be a stochastic volatility model where the price and the variance processes are Itô semimartingles, without any parametric assumption on their drift and diffusion components. Such a data-generating process is fairly general under the no-arbitrage condition (see Delbaen and Schachermayer (1994)).

Historically, the first goal of high-frequency financial econometrics has been obtaining efficient estimates of asset volatilities. However, in the last decade the discipline has broadened its scope to include the topic of the estimation of the so-called second order quantities, that is, quantities that require the knowledge of the volatility process to be computed. These are the leverage, i.e., the covariation

[1]This manuscript has actually been published only recently, see Jacod, J. (2018). Limit of random measures associated with the increments of a Brownian semimartingale. Journal of Financial Econometrics, 16(4):526–569.

[2]Note that in this thesis we will use the terms variance and volatility interchangeably, as customary in the literature on high-frequency financial econometrics.

between the logarithmic price process and the volatility process[3], and the volatility of the volatility itself (hereinafter vol-of-vol), i.e., the quadratic variation of the variance process.

In particular, the first object of interest of the literature on high-frequency financial econometrics has been the *integrated* volatility, that is, the integral of the volatility process over a given time horizon. Specifically, based on the quadratic variation formula, the first proposed estimator of the integrated volatility of a financial asset was the realized variance, which is simply the sum of the squared increments of the logarithmic price of the asset of interest. Assuming that the price process is a continuous semimartingale, the realized volatility is the rate-optimal estimator of the integrated variance. For a detailed description of the finite-sample and asymptotic properties of the realized variance see the seminal papers of the late 1990s and early 2000s by Andersen and Bollerslev (1998), Andersen et al. (2001a), Andersen et al. (2001b), Barndorff-Nielsen and Shephard (2002a), Andersen et al. (2003).

However, empirical evidence shows that observed asset price dynamics deviate from the semimartingale hypothesis at high sampling frequencies, where their dynamics are well-modeled by the sum of a semimartingale component (termed the efficient price) and a noise component, whose presence is ascribed to market microstructure phenomena such as bid-ask bounces, discretization effects and price rounding (see Hasbrouck (2007) for an extensive statistical treatment of market microstructure phenomena). In this regard, several realized estimators of the integrated volatility have been developed that are consistent in the presence of noise, such as the two-scale realized variance by Zhang et al. (2005), the realized kernel estimator by Barndorff-Nielsen et al. (2008a) or the realized estimator based on price pre-averaging by Jacod et al. (2009), just to name a few. At the same time, Malliavin and Mancino (2002, 2009) have also introduced a volatility estimator, based on the Fourier methodology, which is robust to the presence of noise in sample observations. Additionally, note that the Fourier methodology is particularly well-suited for the estimation of multivariate volatilities from asynchronous price observations. Previously mentioned realized methods can indeed be extended, in principle, to the multivariate case by replacing squared returns with cross-returns, that is, by computing realized covariances. However, this may represent a challenging task, as it requires bivariate price series to be aligned on a synchronized grid (see Aït-Sahalia et al. (2010) for a description of the statistical challenges related to data synchronization).

Furthermore, in the presence of jumps in the price trajectory, the realized volatility is a consistent estimator of the sum of the integrated variance and squared

[3] See Chapter 2 for the motivation behind the use of the term *leverage* to indicate such a process.

jumps. As a consequence, both realized and Fourier-based estimators have been proposed to allow the consistent estimation of the integrated variance in the presence of jumps. Jump-robust realized estimators have been introduced by Barndorff-Nielsen and Shephard (2004) and Mancini (2009), based, respectively, on the use of the bipower variation and the truncation of jumps, starting a very large literature (see Chapter 4 and Section IV in Aït-Sahalia and Jacod (2014) for an extensive treatment of this topic). Instead, Cuchiero and Teichmann (2015) have proposed a jump-robust version of the Fourier volatility estimator. Estimators which are consistent in the challenging situation where both noise and jumps are present have also been proposed, see, e.g., Podolskij and Vetter (2009).

A second development of the literature on high-frequency econometrics pertains to the reconstruction of the trajectory of the spot volatility. In this regard, both local and global approaches have been proposed: the latter estimate the entire path of the volatility on a discrete grid, while the former yield point-wise volatility estimates. In particular, local estimators are basically obtained as the numerical derivative of a consistent estimator of the integrated volatility over a local window (see, e.g., Fan and Wang (2008), Mykland and Zhang (2009), Alvarez et al. (2011), Zu and Boswijk (2014) and Bandi and Renò (2018); additionally, see Chapter 8 in Aït-Sahalia and Jacod (2014) for a detailed discussion of the asymptotic properties of localized estimators). For what entails global methods, instead, estimators based on the Fourier transform (Malliavin and Mancino (2002, 2009); Mancino and Recchioni (2015); Cuchiero and Teichmann (2015)) and the Laplace transform (Curato et al. (2018)) have been proposed in the literature. We refer the reader to Chapter 5 for a more detailed illustration of local and global estimators of the spot volatility.

The availability of efficient estimates of the latent spot volatility allows treating the latter as an observable quantity and this, in turn, is crucial for many financial applications, such as, among many others, model calibration (see, e.g., Kanaya and Kristensen (2016)) or risk management (see, e.g., Zu and Boswijk (2014)). In this regard, in Chapter 5 we exploit the availability of efficient spot volatility estimates to obtain empirical evidence supporting the use of affine models in financial applications.

As mentioned, in the last decade the scope of the literature on high-frequency financial econometrics has further expanded to include the estimation of the so-called second-order quantities, i.e., the leverage and the vol-of-vol. We refer the reader to, respectively, Chapters 2 and 4 for a discussion of the estimators of second-order quantities available in the literature, the statistical challenges related to their computation and the relevance of obtaining efficient estimates of second-order quantities for financial applications.

Finally, very recently Sanfelici and Mancino (2020) have proposed a consistent estimator of a third-order quantity, the price-leverage covariation, based on the Fourier methodology. This quantity already appears in Barucci et al. (2003), where the authors propose a model-free indicator of financial instability whose formula depends, other than on the volatility and the leverage, on the price-leverage covariation. In Chapter 3 of this thesis we provide empirical evidence supporting the financial interpretation of the price-leverage covariation as a gauge of the responsiveness of the leverage effect to price and volatility changes.

Outline of the thesis

This thesis contains four essays on the asymptotic and finite-sample properties of non-parametric estimators of the spot volatility, the leverage and the volatility-of-volatility. Empirical applications of the resulting estimates are also explored.

Specifically, the thesis is organized as follows. Chapter 2 is dedicated to the study of the asymptotic normality of the estimation error of the Fourier estimator of the integrated leverage. A central limit theorem for this estimator has already been derived by Curato (2019), but the rate of convergence obtained by the author is sub-optimal. In Chapter 2, instead, we prove a central limit theorem for the Fourier estimator of the integrated leverage with optimal rate of convergence. Additionally, we obtain a smaller asymptotic variance with respect to different realized estimators based on the pre-estimation of the spot volatility. In this chapter we also provide simulation results that confirm the asymptotic result obtained.

Finally, based on the availability of efficient Fourier estimates of the integrated leverage, we perform an empirical study, where we show, using S&P500 prices over the period 2006-2018, that adding an extra term which accounts for the leverage effect to the Heterogeneous Auto-Regressive volatility (HAR) model by Corsi (2009) increases the explanatory power of the latter in a statistically-significant manner. This empirical result extends and robustifies the empirical findings by Mykland and Wang (2014), which have been obtained using a different volatility model, namely an auto-regressive model of order 2, and a different data sample, that is, Microsoft prices over the period 2008-2011.

Chapter 3 is devoted to an empirical study on the relationship between the derivatives of the leverage process with respect to the price and the volatility and the price-leverage covariation. This study is motivated by analytical results that hold under the Constant Elasticity of Variance (CEV) model by Beckers (1980), which is explicitly designed to capture leverage effects. In fact, under the CEV model, the derivatives of the leverage with respect to the log-price and the volatility are equal to ratios of quantities that can be consistently estimated from sample prices over a fixed time horizon. Specifically, such derivatives are equal to the price-leverage covariation scaled, respectively, by the volatility and the leverage itself. This analytical result suggests that the price-leverage covariation may be interpreted, from the financial standpoint, as a gauge of the responsiveness of the leverage to the arrival of new information that causes changes in the price or the volatility. Additionally, we also find that under the CEV model the price-leverage covariation is equal to twice the vol-of-vol, a finding which suggests that the responsiveness of the leverage (i.e., the price-leverage covariation) is proportional to the amount of uncertainty about risk (i.e., the vol-of-vol).

After reconstructing the trajectories of the volatility, the leverage, the vol-of-vol and the price-leverage covariation through the Fourier methodology, we provide empirical evidence supporting such financial interpretation of the price-leverage covariation in a model-free setting, using 1-second S&P500 prices over the period March, 2018-April, 2018. Specifically, we show that the theoretical, model-dependent predictions of the CEV model for the derivatives of the leverage (and for the link between the price-leverage covariation and the vol-of-vol) are quite accurate in approximating their empirical, model-free counterparts.

In Chapter 4, we perform an analytical study to identify the sources of the finite-sample bias that typically plagues the simplest and most natural vol-of-vol estimator, the Pre-estimated Spot-variance based Realized Variance (PSRV) by Barndorff-Nielsen and Veraart (2009). Inspired by the analytical study in Aït-Sahalia et al. (2013), we follow a similar approach to reconstruct the full parametric bias expression under the assumption that the volatility follows a model in the Chan, Karolyi, Longstaff and Sanders (CKLS) class (see Chan et al. (1992)). Then, based on the full knowledge of this expression, we show that the finite-sample bias of the PSRV may be substantially reduced by allowing for the overlap of consecutive local windows to pre-estimate the spot variance. In particular, we provide a feasible analytical rule for the bias-optimal selection of the length of local windows and verify its efficiency both numerically and empirically. Furthermore, we show that selections based on this analytical rule match the selections prescribed in the numerical study by Sanfelici et al. (2015).

In Chapter 5, we empirically investigate, using S&P500 sample data over the period 2006-2018, the functional link between the variance swap rate and the spot volatility, after reconstructing the trajectory of the latter via the Fourier methodology. As a proxy for the variance swap rate we use, instead, the VIX index squared (see Carr and Wu (2006)). As a result, we find overwhelming empirical evidence supporting the affine link analytically found by Kallsen et al. (2011) in the context of exponentially affine stochastic volatility models. Additionally, based on tests performed on yearly subsamples, we find that exponentially mean-reverting variance models provide a good fit during periods of extreme volatility, while polynomial models, introduced in Cuchiero (2011), are suited for years characterized by more frequent price jumps. These empirical results are confirmed when replacing

Fourier estimates of the spot volatility with realized local estimates.

Finally, Chapter 6 concludes, summarizing the main findings of the thesis. Note that, for the ease of the reader, each chapter is meant to be self-contained. As a consequence, some repetitions of the same concepts may appear throughout the thesis.

2. RATE-EFFICIENT ASYMPTOTIC NORMALITY FOR THE FOURIER ESTIMATOR OF THE LEVERAGE

2.1 Introduction

The leverage effect, introduced in the seminal paper by Black (1976), refers to the relationship between asset price returns and volatility changes, which tend to be negatively correlated on equity markets. From a financial standpoint, this phenomenon has been explained as follows (see Black (1976)): if the asset price decreases (increases), the market value of the equity of the corresponding company automatically decreases (increases) as well, making the leverage of the company, i.e., its debt-to-equity ratio, larger (smaller) and thus making the asset more risky and thus more volatile. This explanation, despite being supported by empirical evidence for some assets, has been criticized: on the one hand, the existence of a leverage effect has been shown also for some exchange rates and commodities, where no debt-to-equity ratio (no company) is directly involved; on the other hand, this mechanism implies that the price decline (raise) causes the increase (decrease) in the volatility. Other studies (see, e.g., French et al. (1987)) have in fact suggested that the causal relationship is opposite, being the volatility change the cause of the price change, based on a mechanism termed "volatility feedback effect". However, more recent empirical evidence collected in the literature shows that the leverage effect, generally interpreted as the covariation between the log-price and variance processes, is actually time-dependent and random. Accordingly, models that feature a stochastic leverage effect have been proposed. See, e.g., in this regard, Bandi and Renò (2012), Veraart and Veraart (2012), Mykland and Wang (2014), Kalnina and Xiu (2017), Curato (2019).

The estimation of the leverage effect is challenging, because it requires the knowledge of the latent volatility process. This issue can be dealt with directly by building estimators of the leverage based on the pre-estimation of the path of the spot volatility, as done, e.g., by Mykland and Wang (2014), Aït-Sahalia and Jacod (2014), Chapter 8.4, and Aït-Sahalia et al. (2017). An alternative approach is offered by the Fourier covariance estimation method by Malliavin and Mancino (2009), which requires only the pre-estimation of the Fourier coefficient of the volatility. In this chapter we prove the central limit theorem for the estimator of the integrated leverage obtained with the latter approach.

The Fourier estimation method is particularly suited to build estimators of second-order quantities. In fact, as a first step the Fourier method is applied to obtain estimates of the Fourier coefficients of the volatility. Then, the knowledge of the Fourier coefficients of the latent volatility process allows iterating the procedure in order to compute the Fourier coefficients of the covariation process between the log-price and the volatility. In particular, the integrated leverage requires to compute only the 0-th Fourier coefficient of the covariation process. As mentioned, this procedure does not require the preliminary estimation of the spot volatility path, which is typically obtained via a differentiation procedure (see Chapter 8 in Aït-Sahalia and Jacod (2014)), but only the estimation of the Fourier coefficients of the volatility, which, being integrated quantities, are less subject to numerical instabilities.

An early attempt to use the Fourier method to identify the parameters of stochastic volatility models is present in Malliavin and Mancino (2002), Barucci and Mancino (2010). Curato (2019) proves the asymptotic error normality for the Fourier estimator of the leverage with a rate lower than $1/6$. The low rate found in Curato (2019) is a consequence of the assumption that the number of frequencies M employed for the second step, namely the convolution product between the Fourier coefficients of the log-price and those of the volatility, satisfies $M^3/n \to 0$, where n is the number of price observations. As the asymptotic rate of the Fourier leverage estimator obtained by the author is $M^{1/2}$, the result found in Curato (2019) is clear. At the same time, due to the low rate of convergence, the asymptotic variance does not depend on the number of the Fourier coefficients N of the log-price used to obtain the Fourier coefficients of the volatility. Further, Curato and Sanfelici (2019) study the finite sample properties of the Fourier estimator of the integrated leverage effect in the presence of microstructure noise contamination, showing its asymptotic unbiasedness under this condition.

In this chapter we prove that, in the continuous stochastic volatility model considered in Curato (2019), a careful choice of the two cutting frequencies which define the Fourier leverage estimator allows reaching the optimal rate $1/4$. The resulting asymptotic variance depends on both the frequencies M and N, except in the case where N is chosen to be the Nyquist frequency $n/2$, which is the natural choice in the absence of microstructure noise, indeed. Furthermore, we consider two different convolution products to obtain the Fourier coefficients of the leverage, based, respectively, on the Dirichlet and Fejér kernel. As it is well known, this choice does not affect the rate of convergence, but the asymptotic variance. Both the Fourier leverage estimators reach a smaller asymptotic variance with respect to the leverage estimator in Mykland and Wang (2014), while only the Fourier leverage estimator with the convolution obtained using the Fejér kernel has a smaller asymptotic variance with respect to the leverage estimator in Aït-Sahalia and Jacod (2014). The leverage estimator in Aït-Sahalia et al. (2017) reaches an even smaller asymptotic error variance.

The analytical results derived in this chapter are corroborated by a simulation study, where we show that, as the sample size increases, the empirical distribution of the estimation error approaches the asymptotic distribution with accuracy. Furthermore, the simulation study confirms that the Fourier leverage estimator obtained by means of the Fejér kernel leads to a superior finite-sample efficiency in terms of mean squared error.

Finally, we exploit the availability of efficient leverage estimates to investigate the contribution of the leverage effect to the prediction of the future integrated volatility. In Mykland and Wang (2014), the authors suggest that adding an extra term, namely the asset return scaled by the leverage effect, to any auto-regressive model aimed at predicting next-period's volatility may increase the model's explanatory power in a statistically significant manner, based on empirical evidence obtained from Microsoft high-frequency prices over the period 2008-2011. Accordingly, in this chapter we add the extra term represented by the asset return scaled by the Fourier estimate of the leverage effect to the popular Heterogeneous Auto-Regressive (HAR) volatility model by Corsi (2009) and show, using S&P500 prices over the period 2006-2018, that the contribution of this extra term is statistically significant, thus confirming the empirical result by Mykland and Wang (2014) on a different model and data set.

The chapter is organized as follow. Section 2.2 introduces the main definitions and assumptions. In Section 2.3 we state the main theorems. Section 2.4 contains a numerical study supporting the results of Section 2.3. In Section 2.5 we empirically demonstrate the additional explanatory power of the asset return scaled by the leverage effect when added to the HAR volatility model. Section 2.6 summarizes the main findings of this chapter. Finally, the Appendix (Section 2.7) contains the proofs and some useful results on the Dirichlet and Fejér kernels.

2.2 The Fourier leverage estimator: definition and assumptions

In this section we introduce the general non-parametric stochastic volatility model which will be considered through the chapter as the data-generating process and define the Fourier estimator of the leverage process. Specifically, the assumption on the data-generating process is as follows.

Assumption 2.1. ***Data-generating process***

Let x and σ be, respectively, the logarithmic price process and the volatility process over a fixed time period $[0,T]$. We assume that x and σ are Brownian *semimartingales satisfying, respectively, the stochastic differential equations*

$$dx(t) = \sigma(t)\,dW(t) + a(t)\,dt,$$

$$d\sigma(t) = \widetilde{\gamma}(t)\,dZ(t) + \widetilde{b}(t)\,dt,$$

where W and Z are correlated Brownian motions on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. The processes σ, a, $\widetilde{\gamma}$ and $\widetilde{b}$ are continuous adapted stochastic processes, bounded in absolute value. Moreover, we assume that the process σ is bounded away from zero.

By applying the Itô formula, it is easily seen that the process $v := \sigma^2$ is a Brownian semimartingale under Assumption 2.1, satisfying the stochastic differential equation

$$dv(t) = \gamma(t)\,dZ(t) + b(t)\,dt,$$

where γ and b are expressed in terms of the drift $\widetilde{b}$ and diffusion $\widetilde{\gamma}$, and inherit their properties.

The leverage process η is defined as the covariation between the log-price process x and variance process v, i.e.,

$$\eta(t) := \frac{d\langle x, v\rangle_t}{dt}. \tag{2.1}$$

Under the no-arbitrage condition (see Delbaen and Schachermayer (1994)), the class considered in Assumption 2.1 is fairly general. In fact, it includes most continuous stochastic volatility models commonly used in financial applications and, also, is assumed in Aït-Sahalia and Jacod (2014), Chapter 8.4, and in Mykland and Wang (2014). Further, it coincides with the continuous component of the model considered in Aït-Sahalia et al. (2017), which reads

$$dx(t) = \sigma(t)\,dW(t) + a(t)\,dt,$$

$$d\sigma(t) = f(t)\,dW(t) + g(t)dB(t) + b(t)\,dt,$$

where W and B are independent Brownian motions which can be reconciled with our model by projecting Z on W and an orthogonal Brownian motion $W^{\perp}$.

As it is the case for Mykland and Wang (2014), Aït-Sahalia and Jacod (2014), Chapter 8.4, and Aït-Sahalia et al. (2017), in this chapter we are interested in the estimation of the integral of the stochastic leverage defined in (2.1) as the covariation between the processes x and v. Taking a different point of view, Veraart and Veraart (2012) and Kalnina and Xiu (2017) focus instead on the estimation of the stochastic correlation between W and Z in a more general setting which includes price and volatility jumps.

We now recall the definition of the Fourier estimator of the integrated leverage studied by Curato (2019). By re-scaling the unit of time we can always reduce ourselves to the case where the time window $[0,T]$ becomes $[0,2\pi]$. Suppose that the asset log-price x, satisfying Assumption 2.1, is observed on the discrete grid $t_{j,n} = j2\pi/n$, $j = 0,1\ldots,n$ and denote $\rho(n) := 2\pi/n$.

The Fourier transform of the discretized price is defined as

$$c_k(dx_n) := \frac{1}{2\pi}\sum_{j=0}^{n-1} e^{-\mathrm{i}kt_{j,n}}\delta_j^n(x), \tag{2.2}$$

where $\delta_j^n(x) := x(t_{j+1,n}) - x(t_{j,n})$, while i denotes the imaginary unit $\sqrt{-1}$.

Further, for any $|k| \leq N$, the estimator of the k-th Fourier coefficient of the volatility is defined as

$$c_k(v_{n,N}) := \frac{2\pi}{2N+1}\sum_{|s|\leq N} c_s(dx_n)c_{k-s}(dx_n). \tag{2.3}$$

Based on the knowledge of the Fourier coefficients of the volatility process, $c_k(v_{n,N})$, it is then possible to iteratively apply the convolution formula between the volatility process and the log-price process. The convolution formula as in (2.3), applied to the volatility and log-price processes, produces the Fourier coefficients of their covariation, that is, the leverage process. Note that the Fourier coefficients of the variance increments, $c_k(dv)$, are obtained from $c_k(v)$, via the integration by parts formula. It is now evident that only integrated quantities are necessary to define the Fourier estimator of leverage, and the estimation of spot volatility is not involved.

Using the fact that the integrated leverage is equal to 2π times the 0-th Fourier coefficient of the same function, the Fourier estimator of $\frac{1}{2\pi}\int_0^{2\pi}\eta(t)dt$ is defined as

$$\widehat{\eta}_{n,N,M} := \frac{2\pi}{2M+1}\sum_{|k|\leq M} \mathrm{i}k\, c_k(v_{n,N})\, c_{-k}(dx_n). \tag{2.4}$$

It is also possible to weight the convolution with a different kernel, like the Fejér kernel, as

$$\widetilde{\eta}_{n,N,M} := \frac{2\pi}{M+1}\sum_{|k|\leq M}\left(1-\frac{|k|}{M+1}\right)\mathrm{i}k\, c_k(v_{n,N})\, c_{-k}(dx_n). \tag{2.5}$$

In the present chapter we study the rate efficient asymptotic theorem for the estimators (2.4) and (2.5).

Note that the estimators (2.4) and (2.5) differ from the realized kernel estimators by Barndorff-Nielsen et al. (2008b). In fact, the convolution product leading to the Fourier estimators weights the auto-covariances of any order, the weight being dependent on the number of frequencies N and M, in addition to the lag between observations, see Mancino et al. (2017) for a detailed comparison.

Remark 2.1. *The definition (2.4) (as well as (2.5)) clarifies the reason why asynchronicity issues do not appear in the estimation of the leverage, despite the fact it is tout à fait a covariance. In fact, the Fourier coefficients of the volatility $c_k(v_{n,N})$ defined in (2.3) use only the log-prices observations, but all of them, without any sub-sampling or manipulations. Further, the quantities $c_k(v_{n,N})$ and $c_{-k}(dx_n)$ entering into the definition are two separately integrated quantities, thus no issue of asynchronicity matters.*

Remark 2.2. *In Lemma 2.2 of Malliavin and Mancino (2009), it is proved that the drift component of the semimartingale model gives no contribution to the convolution formula. Therefore, the drift in Assumption 2.1 can be ignored. Moreover, as observed in Malliavin and Mancino (2009), we can assume that $x(0) = x(2\pi)$ and $v(0) = v(2\pi)$. In fact, if $x(0) \neq x(2\pi)$, or $v(0) \neq v(2\pi)$, we introduce, respectively,*

$$\widetilde{x}(t) = x(t) - \frac{t}{2\pi}[x(2\pi) - x(0)]$$

and

$$\widetilde{v}(t) = v(t) - \frac{t}{2\pi}[v(2\pi) - v(0)].$$

Then $\widetilde{x}(0) = \widetilde{x}(2\pi)$, $\widetilde{v}(0) = \widetilde{v}(2\pi)$ and, moreover, volatility and co-volatilities estimations are not affected by a modification of the drift as above. From the point of view of the modeling, we may then consider

$$d\widetilde{x}(t) = \sqrt{v(t)}dW(t),$$

$$d\widetilde{v}(t) = \gamma(t)\,dZ(t).$$

In fact, for any $k \neq 0$, it holds $c_k(dv) = c_k(d\widetilde{v})$.

2.3 The Central Limit Theorems

In this Section we study the asymptotic normality of the Fourier estimators of the integrated leverage (2.4) and (2.5), showing that they reach the optimal rate of convergence $1/4$. Further, we compare the asymptotic variance of the estimators with some competitor estimators having the same rate of convergence.

Theorem 2.1. *Suppose Assumption 2.1 holds. Further, assume that $x(0) = x(2\pi)$ and $v(0) = v(2\pi)$[1]. Let $M\rho(n)^{1/2} \to c_M$ and $N\rho(n) \to c_N$, where c_M and c_N are*

[1] See comments in Remark 2.2.

positive constants. Then, for the leverage estimator (2.4), the following stable convergence in law holds:

$$\rho(n)^{-1/4}\left(\widehat{\eta}_{n,N,M} - \frac{1}{2\pi}\int_0^{2\pi}\eta(t)dt\right)$$

$$\downarrow$$

$$\mathcal{N}\left(0, \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{c_M}\left(\gamma^2(s)v(s)+\eta^2(s)\right)+\frac{1}{6}c_M\left(1+2\theta(c_N/\pi)\right)v^3(s)\,ds\right),$$

where

$$\theta(a) := \frac{1}{2a^2}r(a)(1-r(a)), \tag{2.6}$$

and $r(a) = a - [a]$, being $[a]$ the integer part of a.

The estimator (2.5) has the same asymptotic rate $1/4$, while its asymptotic variance is smaller, due to the presence of the Fejér kernel. The following result is easily obtained from Theorem 2.1.

Theorem 2.2. *Suppose Assumption 2.1 holds. Further, assume that $x(0) = x(2\pi)$ and $v(0) = v(2\pi)$. Let $M\rho(n)^{1/2} \to c_M$ and $N\rho(n) \to c_N$, where c_M and c_N are positive constants. Then, for the estimator (2.5), the following stable convergence in law holds:*

$$\rho(n)^{-1/4}\left(\widetilde{\eta}_{n,N,M} - \frac{1}{2\pi}\int_0^{2\pi}\eta(t)dt\right)$$

$$\downarrow$$

$$\mathcal{N}\left(0, \frac{1}{2\pi}\int_0^{2\pi}\frac{2}{3}\frac{1}{c_M}\left(\gamma^2(s)v(s)+\eta^2(s)\right)+\frac{2}{15}c_M\left(1+2\theta(c_N/\pi)\right)v^3(s)\,ds\right),$$

where $\theta(a)$ is defined in (2.6).

The asymptotic rate $1/4$ is the same rate reached by the integrated leverage estimators proposed in Mykland and Wang (2014), Aït-Sahalia and Jacod (2014) and Aït-Sahalia et al. (2017). We discuss now the asymptotic variances obtained. Note that if c_N is chosen to be equal to π or, equivalently, $N = n/2$ (i.e., the cutting frequency N used for the estimation of the volatility coefficients is equal to the Nyquist frequency), then $\theta(c_N/\pi) = 0$ and the asymptotic variance in Theorem 2.1 becomes

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{1}{c_M}\left(\gamma^2(s)v(s)+\eta^2(s)\right)+\frac{1}{6}c_M v^3(s)\right)ds, \tag{2.7}$$

while that in Theorem 2.2 it becomes

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{2}{3}\frac{1}{c_M}\left(\gamma^2(s)v(s)+\eta^2(s)\right)+\frac{2}{15}c_M v^3(s)\right)ds. \tag{2.8}$$

To compare (2.7) and (2.8) with the asymptotic variance of the leverage estimator in Mykland and Wang (2014), Theorem 1, note that, under the same notation and setting of Theorem 2.1 of this chapter, the latter corresponds to

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{1}{c_M}(\frac{8}{3}\gamma^2(s)v(s)+\eta^2(s))+\frac{4}{\pi^2}c_M v^3(s)\right)ds, \tag{2.9}$$

which is larger with respect to both (2.7) and (2.8). On the other side, considering the variance of the (continuous part) of the leverage estimator by Aït-Sahalia et al. (2017), Theorem 3, which reads

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{2}{3}\frac{1}{c_M}\gamma^2(s)v(s)+\frac{1}{\pi^2}c_M v^3(s)\right)ds, \tag{2.10}$$

we note that (2.10) is smaller with respect to (2.8) (which in turn is smaller than (2.7)). Finally, compared with the asymptotic variance of the leverage estimator in Aït-Sahalia and Jacod (2014), Theorem 8.14, we see that (2.8) is smaller, while (2.7) is larger.

Remark 2.3. *The leverage estimator of Curato (2019) is precisely the estimator (2.4). In Curato (2019) the central limit theorem is obtained under the conditions*[2] $M^3/N \to 0$ *and* $N\rho(n) \to c$*, where c is constant, and reads as follows*

$$\sqrt{M}\left(\widehat{\eta}_{n,N,M}-\int_0^{2\pi}\eta(t)dt\right)$$

$$\downarrow$$

$$\mathcal{N}\left(0,\frac{\pi}{2}\int_0^{2\pi}2(v(s)\gamma^2(s)+\eta^2(s))+v(0)\gamma^2(s)+(v(2\pi)-v(0))^2v(2\pi)ds\right).$$

Therefore the rate of convergence is less than $1/6$*. As for the variance, it is equal to the asymptotic variance of the covariance estimator in Malliavin and Mancino (2009) under synchronous observations, plus a terms depending on the boundary values. Due to the slow rate of convergence, or equivalently, the slower M frequency considered, the asymptotic variance does not depend on the constant ratio between n and N (i.e., c), as remarked also in Mancino et al. (2017) in the general case of Fourier covariance estimator.*

[2] Note that in Curato (2019) M and N are interchanged with respect to the present work.

2.4 Simulation study

In this section we perform a simulation study to assess the finite-sample efficiency of the estimators (2.4) and (2.5). More precisely, we investigate the asymptotic normality as a function of the sample size n and the accuracy of the leverage estimates obtained with the two estimators.

The model used for the simulation is the widely-used stochastic volatility model by Heston (1993):

$$\begin{aligned} dx(t) &= \sigma(t)\,dW(t) + (\mu - \frac{1}{2}v(t))\,dt, \\ dv(t) &= \gamma\sigma(t)\,dZ(t) + \theta(\alpha - v(t))\,dt, \end{aligned} \tag{2.11}$$

where $\mu = 0.01$, $\gamma = 0.5$, $\theta = 2$, $\alpha = 0.2$ and W, Z are correlated Brownian motions with correlation parameter $\rho = -0.8$. This choice of ρ corresponds to a strong leverage effect. However, our numerical results are robust to the selection of different values of ρ and different values of the parameters in the variance process.

The standardized estimation errors for the estimators (2.4) and (2.5) are defined as follows:

$$\widehat{\varepsilon}_{n,N,M} := \rho(n)^{-1/4}\frac{\widehat{\eta}_{n,N,M} - \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\eta(t)dt}{\sqrt{\widehat{V}}}, \tag{2.12}$$

$$\widetilde{\varepsilon}_{n,N,M} := \rho(n)^{-1/4}\frac{\widetilde{\eta}_{n,N,M} - \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\eta(t)dt}{\sqrt{\widetilde{V}}}, \tag{2.13}$$

where $\widehat{V}$ and $\widetilde{V}$ denote the asymptotic error variances derived, respectively, in the Theorem 2.1 and Theorem 2.2. Specifically, we consider the asymptotic error variances in the form given by (2.7) and (2.8), because the simulated setting does not account for the presence of a microstructure noise component and thus the optimal choice for the cutting frequency N is the Nyquist frequency $n/2$.

Regarding the choice of the second cutting frequency, M, its order with respect to n is dictated by the asymptotic theory (Theorems 2.1 and 2.2), while guidance for the choice of the constant c_M (recall that $M \cong c_M(n/2\pi)^{1/2}$) can be obtained as follows. First, we generate 10^4 trajectories of the model (2.11). Each trajectory corresponds to a trading day of length equal to 6.5 hours, i.e., to $n = 23400$ 1-second observations. Then we compute the mean squared error (MSE) of estimators (2.4) and (2.5) in correspondence of different values of c_M. Figure 2.1 compares the MSE of the two estimators as a function of c_M, showing that the MSE-optimal choice of c_M is given by $c_M = 0.4\sqrt{2\pi}$ for the estimator (2.4) and

$c_M = \sqrt{2\pi}$ for the estimator (2.5)[3]. Note that Figure 2.1 confirms that the presence of the Fejer kernel in the definition of the estimator (2.5) has the effect of reducing the error variance of the estimator.

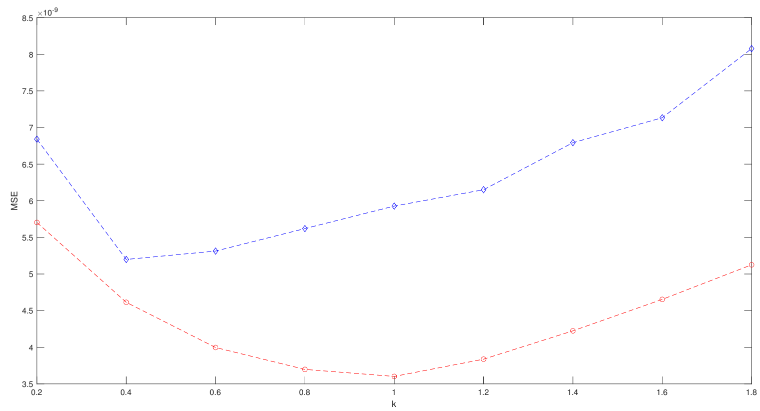

Fig. 2.1: Comparison of the MSE of estimators (2.4) (blue) and (2.5) (red) as a function of $k = \frac{c_M}{\sqrt{2\pi}}$ for n fixed and corresponding to the 1-second sampling frequency.

With this choice of the cutting frequencies N and M, we investigate the asymptotic normality as a function of the sample size n. Figure 2.2 and Table 2.1 illustrate a comparison between the empirical density of the standardized estimation error (2.12), computed from the simulated price trajectories, and the standard normal density, for different values of the sample size n. Figure 2.2 and Table 2.1 show that, as n increases, the estimation accuracy improves, thus confirming, numerically, the finding of Theorem 2.1. Very similar numerical results to those described in Figure 2.2 and Table 2.1 are obtained for the estimation error (2.13) and thus are omitted for the sake of brevity.

2.5 Empirical study

In this section we perform an empirical study where the integrated leverage estimator (2.5) is computed for the S&P500 index in order to investigate the predictive power of the leverage effect on future volatility, in the spirit of Mykland and Wang

[3] An analogous study has been conducted for the 5-minute sampling frequency which will be employed in the empirical study of Section 2.5. The optimal choice of c_M turns out to be pretty much the same as in Figure 2.1.

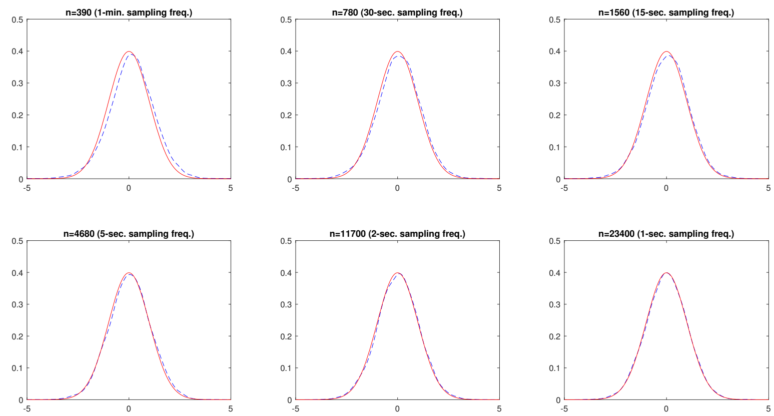

Fig. 2.2: Comparison between the density of the estimation error (2.12) (blue dashed line) and the standard normal density (red line), for different values of the sample size n and $T = 1/252$ (1 day). The price sampling frequencies corresponding to the different values of n are reported between brackets. The plots are obtained from 10^4 simulated trajectories of model (2.11), using the empirical density estimator with normal kernel. The frequencies are chosen as $N = n/2$ and $M = 0.4\sqrt{n}$.

n (sampl. freq.)	variance	mean	median	1st quartile	3rd quartile
390 (1 min.)	1.187	0.135	0.146	-0.542	0.841
780 (30 sec.)	1.110	0.056	0.075	-0.618	0.740
1560 (15 sec.)	1.078	0.040	0.063	-0.634	0.730
4680 (5 sec.)	1.054	0.033	0.046	-0.629	0.719
11700 (2 sec.)	1.031	0.029	0.037	-0.644	0.707
23400 (1 sec.)	1.011	0.003	0.009	-0.672	0.676
$n \to \infty$	1	0	0	-0.674	0.674

Tab. 2.1: Comparison of the sample statistics of the empirical distributions in Figure 2.2 with the corresponding statistics of the standard normal distribution, i.e., the limiting distribution as $n \to \infty$.

(2014). We conduct our empirical investigation on S&P500 prices sampled over the period 2006-2018. This relatively long time frame allows to assess the contribution of the leverage effect in predicting the future variance not only in times of market stability but also in times of market stress, as it encompasses a number of turmoil periods, such as the global financial crisis of 2008, the flash-crash of May

2010, the debt crisis in the Euro area of 2011 and the instability phase caused by the so-called Brexit in 2016.

Figure 2.3 shows the evolution of the estimated daily integrated leverage over the period 2006-2018, together with that of the estimated daily integrated volatility. Note that the integrated leverage changes abruptly in correspondence to volatility spikes: this happens, e.g., in correspondence to the aforementioned financial turmoil periods of 2008, 2010, 2011 and 2016 and is consistent with the findings by Bandi and Renò (2012) and Kalnina and Xiu (2017), who highlight that the leverage effect is not constant and stronger during financial crises.

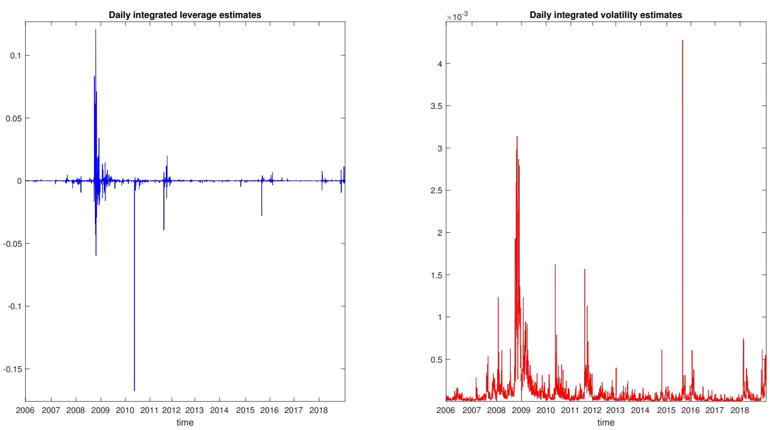

Fig. 2.3: Estimates of the S&P500 daily integrated leverage (blue, left) and daily integrated volatility (red, right) over the period 2006-2018. Leverage estimates are obtained through the estimator (2.5), while volatility estimates are obtained through the estimator of Malliavin and Mancino (2009).

In view of studying the predictive power of the leverage on future volatility, Mykland and Wang (2014) observe that the dynamics of the spot variance depend on the log-return scaled by the ratio between the leverage and the variance itself. More precisely, under the Assumption 2.1, the stochastic differential equation for the variance process can be rewritten as

$$dv(t) = \frac{\eta(t)}{v(t)}dx(t) + 2\sqrt{1-\rho^2}\sigma(t)\widetilde{\gamma}(t)dW^{\perp}(t) + (2\sigma(t)\tilde{b}(t) - 2\rho\widetilde{\gamma}(t)a(t) + \widetilde{\gamma}^2(t))dt,$$

where $W^{\perp}$ is a Brownian motion independent of W. Building on this remark, the authors empirically investigate the dependence between the log-return scaled for the leverage effect and the integrated variance. In particular, using Microsoft high-frequency prices over the period 2008-2011, they show that the explanatory power of a simple AR(2) model for the integrated variance is substantially improved by adding an extra term, represented by the log-return scaled by the time-varying leverage effect. As a conclusion to their empirical study, the authors claim that their findings strongly suggest the inclusion of the return scaled by the leverage effect into any auto-regressive volatility model.

Based on this claim, we empirically investigate whether adding the log-return scaled by the leverage-variance ratio to the HAR model by Corsi (2009) improves the explanatory power of the latter. Formally, the model reads

$$IV_{t+1} = \beta_0 + \beta_1 IV_t + \beta_2 IV_t^{(w)} + \beta_3 IV_t^{(m)} + \beta_4 \frac{IL_t}{IV_t} \Delta x_t + \varepsilon_t, \tag{2.14}$$

where IV_t is the integrated variance on day t, $IV_t^{(w)}$ is the average of the daily integrated variance on the one-week period ranging from day $t-4$ to day t, $IV_t^{(m)}$ is the average of the daily integrated variance on the one-month period ranging from day $t-20$ to day t, Δx_t is the log-return from day $t-1$ to day t, IL_t is the integrated leverage on day t and ε_t is a zero-mean random variable with a distribution that guarantees the positivity of the regressand IV_{t+1}.

Estimates of the integrated variance and leverage are obtained as follows. The sample size n is selected in correspondence to the 5-minute price sampling frequency, based on the result of the Hausman test by Aït-Sahalia and Xiu (2019) for the presence of noise in the price process, which tells that the impact of noise at the 5-minute frequency is negligible in our sample, confirming a well-known stylized fact (see Andersen et al. (2001a)). We also perform the jump-detection test by Corsi et al. (2010) on 5-minute prices and remove the days in which jumps are detected, which amount to about 10% of the 13-year sample[4]. Given this sample, integrated leverage estimates are obtained by means of the estimator (2.5), setting $N = n/2$ and $M = \sqrt{n}$ based on the numerical findings of Section 2.4, while the integrated variance is estimated with the Fourier estimator of Malliavin and Mancino (2009), setting $N = n/2$.

In the spirit of Mykland and Wang (2014), it is worth stressing that we do not aim at discovering the optimal model for volatility forecasting, but rather our goal is to confirm, using a different data set and a different volatility model, that the extra term represented by the return scaled by the leverage effect significantly improves the explanatory power of auto-regressive volatility models. For another

[4]The literature on non-parametric jump tests provides large and robust empirical evidence, mainly based on US markets, that price jumps are accompanied by volatility jumps, consistent with the presence of a leverage effect (see, e.g., Jacod and Todorov (2010); Bandi and Renò (2016); Bibinger and Winkelmann (2018)). Thus removing days with price jumps basically also takes care of jumps in the volatility.

modification of the HAR model that takes into account leverage effects, see Corsi and Renò (2012).

As in Corsi (2009), we estimate model (2.14) using standard Ordinary Least Squares coupled with Newey-West standard errors, to account for the presence of heteroskedasticity and auto-correlations in the residuals. Table 2.2 compares the results of the estimation of the unrestricted model (2.14) and the restricted simple HAR model, i.e., model (2.14) with $\beta_4 = 0$. Note that the estimation yields statistically significant coefficients at 95% level for both models considered.

	model (2.14)	simple HAR model
$\hat{\beta}_0$	$9.202 \cdot 10^{-6}$ (0.037)	$8.163 \cdot 10^{-6}$ (0.038)
$\hat{\beta}_1$	0.353 (10^{-4})	0.369 (10^{-4})
$\hat{\beta}_2$	0.311 (0.009)	0.350 (0.007)
$\hat{\beta}_3$	0.245 (0.018)	0.209 (0.041)
$\hat{\beta}_4$	$1.374 \cdot 10^{-4}$ (0.015)	-
$Adjusted\ R^2$	0.625	0.617
AIC	$-4.286 \cdot 10^4$	$-4.279 \cdot 10^4$
BIC	$-4.283 \cdot 10^4$	$-4.277 \cdot 10^4$

Tab. 2.2: Estimation results for the unrestricted model (2.14) and the simple HAR model by Corsi (2009) on 2006-2018 S&P500 data. P-values are reported between brackets.

Based of the comparison of Adjusted R^2, AIC and BIC values, model (2.14) should be preferred to the simple HAR in terms of goodness of fit. To statistically test if the additional term, namely the log-return scaled by the leverage-variance ratio, significantly improves the explanatory power of the simple HAR model, we perform the likelihood ratio test. The test returns a p-value equal to $1.776 \cdot 10^{-15}$, overwhelmingly rejecting the null hypothesis that the two models carry the same explanatory power. Thus, we have statistical evidence that including the information related to the leverage effect improves the explanatory power of auto-regressive volatility models, confirming the empirical result by Mykland and Wang (2014).

2.6 Conclusions

The main result of the chapter is proving a central limit theorem for the Fourier estimator of the leverage with the optimal rate of convergence $1/4$ and a smaller variance with respect to different estimators that are based on a pre-estimation of the spot volatility. This analytical result is supported with numerical evidence that shows the accuracy of the empirical distribution of the error of the Fourier leverage estimator in approximating the asymptotic distribution for large values of the sample size. Finally, we exploit the availability of finite-sample efficient Fourier estimates of the leverage to empirically show, using S&P500 prices over the period 2006-2018, that adding an extra term that accounts for the leverage effect to the HAR volatility model by Corsi (2009) increases the explanatory power of the latter. This result confirms and extends the empirical findings by Mykland and Wang (2014).

2.7 Appendix

2.7.1 Proofs

Proof of Theorem 2.1

We use the notation in continuous time by letting $\varphi_n(t) := \sup\{t_j : t_j \leq t\}$ and, for brevity, we denote[5]

$$D_{*,n}(s-u) := D_*(\varphi_n(s) - \varphi_n(u)), \text{ for } * = N, M, \tag{2.15}$$

being $D_*(\cdot)$ the Dirichlet kernel defined in (2.107). Then we rewrite (2.2) as

$$c_k(dx_n) := \frac{1}{2\pi} \int_0^{2\pi} e^{-\mathrm{i}k\varphi_n(t)} dx(t)$$

and, by applying Itô formula, we can rewrite (2.3) as follows

$$c_k(v_{n,N}) := A_{k,n} + B_{k,n,N} + C_{k,n,N}, \tag{2.16}$$

where

$$A_{k,n} := \frac{1}{2\pi} \int_0^{2\pi} e^{-\mathrm{i}k\varphi_n(s)} v(s) ds \tag{2.17}$$

$$B_{k,n,N} := \frac{1}{2\pi} \int_0^{2\pi} e^{-\mathrm{i}k\varphi_n(s)} Y_{n,N}(s,s)\, \sigma(s) dW(s) \tag{2.18}$$

$$C_{k,n,N} := \frac{1}{2\pi} \int_0^{2\pi} Y_{k,n,N}(s,s)\, \sigma(s) dW(s), \tag{2.19}$$

where we used the notation

$$Y_{n,N}(t,s) := \int_0^t D_{N,n}(s-u)\, \sigma(u) dW(u) \tag{2.20}$$

[5]We stress that the notation for discretized Dirichlet kernel $D_{N,n}(t-u)$ is used to denote $D_N(\varphi_n(t) - \varphi_n(u))$, not $D_N(\varphi_n(t-u))$. The notation is chosen to highlight the role of the convolution product, one of the key tools of the Fourier methodology, see Malliavin and Mancino (2009).

and

$$Y_{k,n,N}(t,s) := \int_0^t e^{-\mathrm{i}k\varphi_n(u)} D_{N,n}(s-u)\,\sigma(u)dW(u). \tag{2.21}$$

Therefore, the estimator (2.4) will be studied as the sum of three terms:

$$\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k A_{k,n}\, c_{-k}(dx_n), \tag{2.22}$$

$$\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k B_{k,n,N}\, c_{-k}(dx_n), \tag{2.23}$$

$$\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k C_{k,n,N}\, c_{-k}(dx_n). \tag{2.24}$$

Clearly, the terms (2.23) and (2.24) give analogous contribution.
The proof is divided in three steps, the first one shows that the discretization error converges to zero. The other two steps follow Jacod (1997) in order to identify the asymptotic variance and prove the stable convergence in law.

Step I: Discretization
We begin with (2.22) and consider

$$\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k A_{k,n} c_{-k}(dx_n) - \frac{1}{2\pi}\int_0^{2\pi} \eta(t)dt,$$

which is equal to the sum of the following two terms (respectively, discretization error and continuous limit)

$$\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k \left(A_{k,n}c_{-k}(dx_n) - \left(\frac{1}{2\pi}\int_0^{2\pi} e^{-\mathrm{i}ks}\nu(s)ds\right)\left(\frac{1}{2\pi}\int_0^{2\pi} e^{\mathrm{i}ks}\sigma(s)dW(s)\right)\right) \tag{2.25}$$

$$+\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k \left(\frac{1}{2\pi}\int_0^{2\pi} e^{-\mathrm{i}ks}\nu(s)ds\right)\left(\frac{1}{2\pi}\int_0^{2\pi} e^{\mathrm{i}ks}\sigma(s)dW(s)\right) - \frac{1}{2\pi}\int_0^{2\pi}\eta(t)dt. \tag{2.26}$$

In order to study (2.25), by applying the Itô formula and the stochastic Fubini theorem, we show that

$$E[\,|\rho(n)^{-1/4}\frac{1}{2\pi}\int_0^{2\pi}\int_0^{2\pi} \left(D'_{M,n}(t-u) - D'_M(t-u)\right)\nu(u)du\,\sigma(t)dW(t)|^2\,]$$

converges, for $n, M \to \infty$, to zero[6]. Applying again the Itô isometry and using the assumption of the boundedness of volatility, it is enough to prove the convergence

[6] For $z \in \mathbf{C}$, then $|z|$ denotes the modulus of z.

to zero of

$$\rho(n)^{-1/2}\int_0^{2\pi} E[|\frac{1}{2\pi}\int_0^{2\pi}\left(D'_{M,n}(t-u)-D'_M(t-u)\right)v(u)du|^2]\,dt. \tag{2.27}$$

It holds

$$\frac{1}{2\pi}\int_0^{2\pi}\left(D'_{M,n}(t-u)-D'_M(t-u)\right)v(u)du$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2M+1}\sum_{|k|\le M}\mathrm{i}k\left(e^{\mathrm{i}k(\varphi_n(t)-\varphi_n(u))}-e^{\mathrm{i}k(\varphi_n(t)-u)}\right)v(u)du \tag{2.28}$$

$$+\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2M+1}\sum_{|k|\le M}\mathrm{i}k\left(e^{\mathrm{i}k(\varphi_n(t)-u)}-e^{\mathrm{i}k(t-u)}\right)v(u)du. \tag{2.29}$$

Consider (2.28), which is equal to

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2M+1}\sum_{|k|\le M}e^{\mathrm{i}k\varphi_n(t)}\,(e^{-\mathrm{i}k\varepsilon_n(u)}-1)\,\mathrm{i}k\,e^{-\mathrm{i}ku}v(u)du \tag{2.30}$$

where $\varepsilon_n(u):=\varphi_n(u)-u$. By the Taylor expansion, for any $u\in[0,2\pi]$, $e^{-\mathrm{i}k\varepsilon_n(u)}-1=-ik\rho(n)+O(k^2\rho(n)^2)$, therefore, we need to study the order of

$$\rho(n)\frac{1}{2M+1}\sum_{|k|\le M}k\,\frac{1}{2\pi}\int_0^{2\pi}\mathrm{i}k\,e^{-\mathrm{i}ku}v(u)du \tag{2.31}$$

or, equivalently, of

$$\rho(n)\,\frac{1}{2M+1}\sum_{|k|\le M}k\,c_k(dv), \tag{2.32}$$

where we have used the integration by parts formula[7]. Finally, we obtain

$$\rho(n)^2E[|\frac{1}{2M+1}\sum_{|k|\le M}k\,c_k(dv)|^2]\le CM\rho(n)^2\sum_{|k|\le M}E[|c_k(dv)|^2]\le CM^2\rho(n)^2E[\int_0^{2\pi}\gamma^2(t)dt]. \tag{2.33}$$

Then, the term (2.33) multiplied by $\rho(n)^{-1/2}$ is $O(\rho(n)^{1/2})$.

Consider now (2.29). It holds

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2M+1}\sum_{|k|\le M}\mathrm{i}k\,e^{-\mathrm{i}ku}\left(e^{\mathrm{i}k\varphi_n(t)}-e^{\mathrm{i}kt}\right)v(u)du$$

[7] For any $k\neq 0$, the integration by parts formula gives: $c_k(v)=\frac{\mathrm{i}}{k}\left(\frac{v(2\pi)-v(0)}{2\pi}-c_k(dv)\right)=-\frac{\mathrm{i}}{k}\,c_k(dv)$.

$$= \frac{1}{2M+1} \sum_{|k| \le M} e^{\mathrm{i}kt} (e^{\mathrm{i}k(\varphi_n(t)-t)} - 1) \frac{1}{2\pi} \int_0^{2\pi} \mathrm{i}k\, e^{-\mathrm{i}ku} v(u) du.$$

Then, with analogous arguments, we have

$$E[| \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2M+1} \sum_{|k| \le M} \mathrm{i}k\, e^{-\mathrm{i}ku} \left(e^{\mathrm{i}k\varphi_n(t)} - e^{\mathrm{i}kt} \right) v(u) du|^2] \le M^2 \rho(n)^2 \frac{1}{2M+1} \sum_{|k| \le M} E[| \frac{1}{2\pi} \int_0^{2\pi} \mathrm{i}k\, e^{-\mathrm{i}ku} v(u) du|^2].$$

Finally, as before, the term (2.29) multiplied by $\rho(n)^{-1/2}$ converges to zero.

Step II: Asymptotic Variance

We identify now the asymptotic variance. First consider the term (2.26), namely:

$$\frac{2\pi}{2M+1} \sum_{|k| \le M} \mathrm{i}k\, c_k(v)\, c_{-k}(dx) - \frac{1}{2\pi} \int_0^{2\pi} \eta(t) dt. \tag{2.34}$$

Using the integration by parts formula, for any $k \neq 0$, then $\mathrm{i}kc_k(v) = c_k(dv)$, therefore (2.34) is equal to

$$\frac{2\pi}{2M+1} \sum_{|k| \le M} c_k(dv) c_{-k}(dx) - \frac{1}{2\pi} \int_0^{2\pi} \eta(t) dt \tag{2.35}$$

$$= \frac{1}{2M+1} \sum_{|k| \le M} \frac{1}{2\pi} \left(\int_0^{2\pi} \int_0^t e^{-\mathrm{i}k(t-s)} dx(s)\, dv(t) + \int_0^{2\pi} \int_0^t e^{-\mathrm{i}k(t-s)} dv(s)\, dx(t) \right),$$

where we have applied the Itô formula for the last line. Therefore, it holds

$$\frac{2\pi}{2M+1} \sum_{|k| \le M} c_k(dv) c_{-k}(dx) - \frac{1}{2\pi} \int_0^{2\pi} \eta(t) dt = A_M^{(i)}(2\pi) + A_M^{(ii)}(2\pi),$$

where we denote

$$A_M^{(i)}(u) := \frac{1}{2\pi} \int_0^u \int_0^t D_M(t-s) dv(s)\, dx(t), \tag{2.36}$$

$$A_M^{(ii)}(u) := \frac{1}{2\pi} \int_0^u \int_0^t D_M(t-s) dx(s)\, dv(t). \tag{2.37}$$

The first addend of the asymptotic variance is determined with the limit in probability of[8]

$$\langle M^{1/2}(A_M^{(i)} + A_M^{(ii)}), M^{1/2}(A_M^{(i)} + A_M^{(ii)}) \rangle_{2\pi}. \tag{2.38}$$

[8]For ease of notation, we will denote $A_M^{(i)}$ instead of $A_M^{(i)}(\cdot)$ when no confusion may appear.

Consider first $\langle M^{1/2}A_M^{(i)}, M^{1/2}A_M^{(i)}\rangle_{2\pi}$. It holds:

$$\langle M^{1/2}A_M^{(i)}, M^{1/2}A_M^{(i)}\rangle_{2\pi} = M\int_0^{2\pi}\left(\frac{1}{2\pi}\int_0^t D_M(t-s)dv(s)\right)^2 v(t)dt$$

$$= M\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^t D_M^2(t-s)\gamma^2(s)ds\, v(t)dt \tag{2.39}$$

$$+M\frac{1}{(2\pi)^2}\int_0^{2\pi} 2\int_0^t\int_0^s D_M(t-u)dv(u)\, D_M(t-s)dv(s)\, v(t)dt. \tag{2.40}$$

Using Lemma 2.2 , then (2.39) converges in probability to

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\gamma^2(t)v(t)dt. \tag{2.41}$$

As for (2.40), it gives:

$$E[(M\frac{1}{(2\pi)^2}\int_0^{2\pi} 2\int_0^t\int_0^s D_M(t-u)dv(u)\, D_M(t-s)dv(s)\, v(t)dt)^2]$$

$$\leq CM^2\int_0^{2\pi}\int_0^t\int_0^s D_M^2(t-u)du\, D_M^2(t-s)ds\, dt, \tag{2.42}$$

where we have applied Itô isometry and used the boundedness of the processes v and γ. Finally, it is enough to use Lemma 2.2 i) and iii) to see that (2.42) is $o(1)$. With the same method, it is shown that also $\langle M^{1/2}A_M^{(ii)}, M^{1/2}A_M^{(ii)}\rangle_{2\pi}$ converges to (2.41).

Consider now the bracket $\langle M^{1/2}A_M^{(i)}, M^{1/2}A_M^{(ii)}\rangle_{2\pi}$ (analogous results hold for $\langle M^{1/2}A_M^{(ii)}, M^{1/2}A_M^{(i)}\rangle_{2\pi}$). The bracket is equal to:

$$M\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^t D_M(t-s)dx(s)\int_0^t D_M(t-s)dv(s)\,\eta(t)dt$$

$$= M\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^t D_M^2(t-s)\eta(s)ds\,\eta(t)dt \tag{2.43}$$

$$+M\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^t\int_0^s D_M(t-u)dv(u)\, D_M(t-s)dx(s)\,\eta(t)dt \tag{2.44}$$

$$+M\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^t\int_0^s D_M(t-u)dx(u)\, D_M(t-s)dv(s)\,\eta(t)dt. \tag{2.45}$$

By Lemma 2.2, the term (2.43) converges to

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\eta^2(t)dt.$$

As for (2.44) and (2.45), they are $o_p(1)$, as for (2.42). Finally, noting that $\rho(n)^{-1/4} \sim (c_M)^{-1/2}M^{1/2}$, we conclude that

$$\langle \rho(n)^{-1/4}(A_M^{(i)}+A_M^{(ii)}), \rho(n)^{-1/4}(A_M^{(i)}+A_M^{(ii)})\rangle_{2\pi} \to \frac{1}{c_M}\frac{1}{2\pi}\int_0^{2\pi}(\gamma^2(s)\nu(s)+\eta^2(s))ds \tag{2.46}$$

in probability.

We consider now the other addends of the asymptotic variance. First, we study (2.23). Applying Itô formula, it holds:

$$\frac{2\pi}{2M+1}\sum_{|k|\le M} \mathrm{i}k B_{k,n,N} c_{-k}(dx_n)$$

$$= \frac{1}{2M+1}\sum_{|k|\le M} \mathrm{i}k \int_0^{2\pi}\frac{1}{2\pi}\int_0^s D_{N,n}(s-u)\sigma(u)dW(u)\,\nu(s)ds$$

$$+\frac{1}{2M+1}\sum_{|k|\le M} \mathrm{i}k\frac{1}{2\pi}\int_0^{2\pi} e^{-\mathrm{i}k\varphi_n(s)}\int_0^s e^{\mathrm{i}k\varphi_n(t)}\sigma(t)dW(t)\int_0^s D_{N,n}(s-u)\sigma(u)dW(u)\,\sigma(s)dW(s)$$

$$+\frac{1}{2M+1}\sum_{|k|\le M} \mathrm{i}k\frac{1}{2\pi}\int_0^{2\pi}\int_0^s e^{-\mathrm{i}k\varphi_n(t)}\int_0^t D_{N,n}(t-u)\sigma(u)dW(u)\,\sigma(t)dW(t)\,e^{\mathrm{i}k\varphi_n(s)}\sigma(s)dW(s).$$

Thus (2.23) is the sum of three terms:

$$B_{n,N,M}^{(i)} := D_M'(0)\frac{1}{2\pi}\int_0^{2\pi}\int_0^s D_{N,n}(s-u)\sigma(u)dW(u)\,\nu(s)ds, \tag{2.47}$$

$$B_{n,N,M}^{(ii)} := \int_0^{2\pi}\frac{1}{2\pi}\left(\int_0^s D_{M,n}'(s-t)\sigma(t)dW(t)\right)\left(\int_0^s D_{N,n}(s-u)\sigma(u)dW(u)\right)\sigma(s)dW(s), \tag{2.48}$$

$$B_{n,N,M}^{(iii)} := \int_0^{2\pi}\frac{1}{2\pi}\int_0^s\left(\int_0^t D_{N,n}(t-u)\sigma(u)dW(u)\right)D_{M,n}'(s-t)\sigma(t)dW(t)\,\sigma(s)dW(s). \tag{2.49}$$

Note that $D_M'(0) = \frac{1}{2M+1}\sum_{|k|\le M}\mathrm{i}k = 0$, thus $B_{n,N,M}^{(i)}$ is identically zero. It remains to study

$$\langle \rho(n)^{-1/4}(B_{n,N,M}^{(ii)}+B_{n,N,M}^{(iii)}), \rho(n)^{-1/4}\overline{(B_{n,N,M}^{(ii)}+B_{n,N,M}^{(iii)})}\rangle_{2\pi}. \tag{2.50}$$

First compute

$$\rho(n)^{-1/2}\langle B_{n,N,M}^{(ii)}, \overline{B_{n,N,M}^{(ii)}}\rangle_{2\pi}. \tag{2.51}$$

Using Itô formula, $\langle B^{(ii)}_{n,N,M}, \overline{B^{(ii)}_{n,N,M}}\rangle_{2\pi}$ is equal to

$$\int_0^{2\pi} \frac{1}{(2\pi)^2} \int_0^s |D'_{M,n}(s-t)|^2 v(t)dt \int_0^s D^2_{N,n}(s-u)v(u)du\, v(s)ds \tag{2.52}$$

$$+\int_0^{2\pi} \frac{2}{(2\pi)^2} \int_0^s \int_0^u D_{N,n}(s-u')\sigma(u')dW_{u'}\, D_{N,n}(s-u)\sigma(u)dW(u) \int_0^s |D'_{M,n}(s-u)|^2 v(u)du\, v(s)ds \tag{2.53}$$

$$+\int_0^{2\pi} \frac{2}{(2\pi)^2} \int_0^s \int_0^u D'_{M,n}(s-u')\sigma(u')dW_{u'}\, \overline{D'_{M,n}(s-u)}\sigma(u)dW(u) \int_0^s D^2_{N,n}(s-u)v(u)du\, v(s)ds \tag{2.54}$$

$$+\int_0^{2\pi} \frac{4}{(2\pi)^2} \int_0^s \int_0^u \overline{D'_{M,n}(s-u')}\sigma(u')dW_{u'}\, D'_{M,n}(s-u)\sigma(u)dW(u)\times \tag{2.55}$$

$$\times \int_0^s \int_0^u D_{N,n}(s-u')\sigma(u')dW_{u'}\, D_{N,n}(s-u)\sigma(u)dW(u)\, v(s)ds.$$

Consider (2.52). By Lemmas 2.1 and 2.3 it holds in probability

$$n\int_0^s D^2_{N,n}(s-u)v(u)du \to \pi(1+2\theta(c_N/\pi))v(s)$$

and

$$\frac{1}{M}\int_0^s |D'_{M,n}(s-t)|^2 v(t)dt \to \frac{\pi}{6} v(s).$$

Therefore, noting that $\rho(n)^{-1/2} \sim (c_M)^{-1}M$, the term (2.52) multiplied by $\rho(n)^{-1/2}$ gives:

$$\frac{1}{24}c_M(1+2\theta(c_N/\pi))\frac{1}{2\pi}\int_0^{2\pi}\sigma^6(s)ds. \tag{2.56}$$

Now we consider (2.53) and show that it converges to zero in probability when multiplied by $\rho(n)^{-1/2}$. The terms (2.54) and (2.55) are analogous. Using the boundedness of the volatility and the Cauchy-Schwartz inequality

$$E[|\int_0^{2\pi}\int_0^s \int_0^u D_{N,n}(s-u')\sigma(u')dW_{u'}\, D_{N,n}(s-u)\sigma(u)dW(u) \int_0^s |D'_{M,n}(s-u)|^2 v(u)du\, v(s)ds|] \tag{2.57}$$

$$\leq \int_0^{2\pi} E[(\int_0^s \int_0^u D_{N,n}(s-u')\sigma(u')dW_{u'}\, D_{N,n}(s-u)\sigma(u)dW(u))^2]^{1/2} E[(\int_0^s |D'_{M,n}(s-u)|^2 v(u)du)^2]^{1/2} ds.$$

Now, by using twice the Itô isometry and the boundedness of the volatility

$$E[(\int_0^s \int_0^u D_{N,n}(s-u')\sigma(u')dW_{u'}\, D_{N,n}(s-u)\sigma(u)dW(u))^2] \leq C\int_0^s\int_0^u D^2_{N,n}(s-u')du'\, D^2_{N,n}(s-u)du \leq \frac{C}{N^3},$$

where we have used Lemma 2.1 and the fact that, for any $\varepsilon > 0$ and $u' < s - \varepsilon$, then $D^2_{N,n}(s-u') \leq CN^{-2}$ for n large enough. Finally, using Lemma 2.3, the term (2.57) is $O(N^{-1})$. Thus, when multiplied by $\rho(n)^{-1/2}$, the term converges to zero.

Consider now

$$\rho(n)^{-1/2} \langle B^{(iii)}_{n,N,M}, \overline{B^{(iii)}_{n,N,M}} \rangle_{2\pi}. \tag{2.58}$$

Observe that (2.58) is equal to

$$\rho(n)^{-1/2} \int_0^{2\pi} \frac{1}{(2\pi)^2} |B_{N,M,n}(s)|^2 v(s)ds, \tag{2.59}$$

where

$$B_{N,M,n}(s) := \int_0^s Y_{N,n}(t,t) D'_{M,n}(s-t)\sigma(t)dW(t).$$

By applying the Itô formula, it holds that $|B_{N,M,n}(s)|^2$ is equal to

$$\int_0^s Y^2_{N,n}(t,t)|D'_{M,n}(s-t)|^2 v(t)dt \tag{2.60}$$

$$+ \int_0^s \int_0^t Y_{N,n}(t_1,t)D'_{M,n}(s-t_1)\sigma(t_1)dW_{t_1}\, Y_{N,n}(t,t)\overline{D'_{M,n}(s-t)}\sigma(t)dW(t) \tag{2.61}$$

$$+ \int_0^s \int_0^t Y_{N,n}(t_1,t)\overline{D'_{M,n}(s-t_1)}\sigma(t_1)dW_{t_1}\, Y_{N,n}(t,t)D'_{M,n}(s-t)\sigma(t)dW(t). \tag{2.62}$$

Consider (2.60). By an iterated application of the Itô formula, the first term to be studied is

$$\rho(n)^{-1/2} \int_0^{2\pi} \frac{1}{(2\pi)^2} \int_0^s \int_0^t D^2_{N,n}(t-u)v(u)du\, |D'_{M,n}(s-t)|^2 v(t)dt\, v(s)ds \tag{2.63}$$

$$+\rho(n)^{-1/2} \int_0^{2\pi} \frac{1}{(2\pi)^2} \int_0^s 2\int_0^t Y_{N,n}(u,t)D_{N,n}(t-u)\sigma(u)dW_u\, |D'_{M,n}(s-t)|^2 v(t)dt\, v(s)ds. \tag{2.64}$$

The term (2.63) is studied as (2.52). By Lemma 2.2 and Lemma 2.3 it converges in probability to the same process as (2.56), that is,

$$\frac{1}{24} c_M(1+2\theta(c_N/\pi)) \frac{1}{2\pi} \int_0^{2\pi} \sigma^6(s)ds.$$

We then study (2.64) and show that it converges to zero. Using the boundedness of the volatility (and neglecting irrelevant constants), we obtain:

$$E[|\rho(n)^{-1/2} \int_0^{2\pi} \int_0^s \int_0^t Y_{N,n}(u,t)\, D_{N,n}(t-u)\sigma(u)dW_u\, |D'_{M,n}(s-t)|^2 v(t)dt\, v(s)ds|]$$

$$\leq \rho(n)^{-1/2} \int_0^{2\pi} \int_0^s E[| \int_0^t Y_{N,n}(u,t)\, D_{N,n}(t-u)\sigma(u)dW_u|]\, |D'_{M,n}(s-t)|^2 dt\, ds.$$

Then, it is enough to note that

$$E[(\int_0^t Y_{N,n}(u,t)\, D_{N,n}(t-u)\sigma(u)dW_u)^2] \leq C \int_0^t \int_0^u D^2_{N,n}(t-u')du'\, D^2_{N,n}(t-u)du,$$

where we used the Itô isometry and the boundedness of the volatility process. Moreover, using Lemma 2.1 and the fact that, for any $\varepsilon > 0$ and $u' < t-\varepsilon$, then $D^2_{N,n}(t-u') \leq CN^{-2}$, for n large enough, this term is smaller than $Cn^{-3/2}$. Finally, the term (2.64) is $O_p(\rho(n))$.

Now we consider (2.61) and prove that

$$\rho(n)^{-1/2} \int_0^{2\pi} \int_0^s \int_0^t Y_{N,n}(t_1,t) D'_{M,n}(s-t_1)\sigma(t_1)dW_{t_1}\, Y_{N,n}(t,t)\overline{D'_{M,n}(s-t)}\sigma(t)dW(t)\, v(s)ds$$

converges to 0. By the boundedness of the volatility process, it is enough to study

$$E[|\rho(n)^{-1/2} \int_0^s \int_0^t Y_{N,n}(t_1,t) D'_{M,n}(s-t_1)\sigma(t_1)dW_{t_1}\, Y_{N,n}(t,t)\overline{D'_{M,n}(s-t)}\sigma(t)dW(t)|^2], \tag{2.65}$$

which, by using the Itô isometry, is smaller than

$$C\rho(n)^{-1} \int_0^s E[| \int_0^t Y_{N,n}(t_1,t) D'_{M,n}(s-t_1)\sigma(t_1)dW_{t_1}\, Y_{N,n}(t,t)|^2]\, |D'_{M,n}(s-t)|^2 dt$$

$$\leq C\rho(n)^{-1} \int_0^s E[| \int_0^t Y_{N,n}(t_1,t) D'_{M,n}(s-t_1)\sigma(t_1)dW_{t_1}|^4]^{1/2} E[Y^4_{N,n}(t,t)]^{1/2}\, |D'_{M,n}(s-t)|^2 dt,$$

where we have applied the Cauchy-Schwartz inequality in the last line. Observe now that by the Burkholder-Davis-Gundy inequality, Lemma 2.1 and the boundedness of the volatility, it holds

$$E[Y^4_{N,n}(t,t)]^{1/2} \leq C \int_0^t D^2_{N,n}(t-u)du \leq C\rho(n), \tag{2.66}$$

and

$$E[| \int_0^t Y_{N,n}(t_1,t) D'_{M,n}(s-t_1)\sigma(t_1)dW_{t_1}|^4]^{1/2} \leq C(\int_0^t E[Y^4_{N,n}(t_1,t)]|D'_{M,n}(s-t_1)|^4 dt_1)^{1/2} \leq CN^{-2}M^{3/2}, \tag{2.67}$$

where we have used the fact that, for any $\varepsilon > 0$ and $u < t-\varepsilon$, then $D^2_{N,n}(t-u) \leq CN^{-2}$, and Lemma 2.3. Finally, the term (2.65) is $O(M^{-3/2})$, thus goes to 0.

With the same methodology, it is proved that the contribution of the other two brackets is the same. Therefore, (2.50) converges to

$$\frac{1}{6} c_M \left(1+2\theta(c_N/\pi)\right) \frac{1}{2\pi} \int_0^{2\pi} \sigma^6(s)ds. \tag{2.68}$$

Finally, we show that

$$\langle \rho(n)^{-1/4}(A_M^{(i)} + A_M^{(ii)}), \rho(n)^{-1/4}\overline{(B_{n,N,M}^{(ii)} + B_{n,N,M}^{(iii)})}\rangle_{2\pi} \to 0 \tag{2.69}$$

in probability. We study in details the term

$$\rho(n)^{-1/2}\langle A_M^{(i)}, \overline{B_{n,N,M}^{(ii)}}\rangle_{2\pi}$$

$$= \rho(n)^{-1/2}\int_0^{2\pi}\frac{1}{(2\pi)^2}\int_0^t D_M(t-s)\gamma(s)dZ(s)\int_0^t \overline{D'_{M,n}(t-s)}\sigma(s)dW(s)\int_0^t D_{N,n}(t-s)\sigma(s)dW(s)v(t)dt. \tag{2.70}$$

The other three terms are analogous. An iterated application of Itô formula gives that (2.70) is equal to:

$$\rho(n)^{-1/2}\int_0^{2\pi}\frac{1}{(2\pi)^2}\int_0^t\int_0^s D_M(t-u)D_{N,n}(t-u)\eta(u)du\,\overline{D'_{M,n}(t-s)}\sigma(s)dW(s)\,v(t)dt \tag{2.71}$$

$$+\rho(n)^{-1/2}\int_0^{2\pi}\frac{1}{(2\pi)^2}\int_0^t\int_0^s \overline{D'_{M,n}(t-u)}\sigma(u)dW(u)D_M(t-s)D_{N,n}(t-s)\eta(s)ds\,v(t)dt \tag{2.72}$$

$$+\rho(n)^{-1/2}\int_0^{2\pi}\frac{1}{(2\pi)^2}\int_0^t\int_0^s D_M(t-u)\gamma(u)dZ(u)D_{N,n}(t-s)\sigma(s)dW(s)\int_0^t \overline{D'_{M,n}(t-s)}\sigma(s)dW(s)\,v(t)dt \tag{2.73}$$

$$+\rho(n)^{-1/2}\int_0^{2\pi}\frac{1}{(2\pi)^2}\int_0^t\int_0^s D_{N,n}(t-u)\sigma(u)dW(u)D_M(t-s)\gamma(s)dZ(s)\int_0^t \overline{D'_{M,n}(t-s)}\sigma(s)dW(s)\,v(t)dt. \tag{2.74}$$

Consider (2.71). The term (2.72) is analogous. By the boundedness of the volatility, it is enough to study

$$\rho(n)^{-1}E[|\int_0^t\int_0^s D_M(t-u)D_{N,n}(t-u)\eta(u)du\,\overline{D'_{M,n}(t-s)}\sigma(s)dW(s)|^2]$$

$$\leq C\rho(n)^{-1}\int_0^t E[(\int_0^s D_M(t-u)D_{N,n}(t-u)v(u)du)^2]|D'_{M,n}(t-s)|^2ds, \tag{2.75}$$

where, in the last inequality, we have used the Itô isometry and the boundedness of the volatility. Finally, by the Cauchy-Schwartz inequality and the fact that for any $\varepsilon > 0$, both $D_M^2(t-u) \leq CM^{-2}$ and $D_{N,n}^2(t-u) \leq CN^{-2}$ for $u < t-\varepsilon$, for n large enough, we have that (2.75) has order $\rho(n)^{-1}M^{-2}N^{-2}M$, which is $o(1)$.

Consider now (2.73). The term (2.74) is analogous. By the boundedness of the volatility and the Cauchy-Schwartz inequality:

$$E[|\int_0^{2\pi}\int_0^t\int_0^s D_M(t-u)\gamma(u)dZ(u)D_{N,n}(t-s)\sigma(s)dW(s)\int_0^t \overline{D'_{M,n}(t-s)}\sigma(s)dW(s)\,v(t)dt|]$$

$$\leq C\int_0^{2\pi} E[|\int_0^t \int_0^s D_M(t-u)\gamma(u)dZ(u)D_{N,n}(t-s)\sigma(s)dW(s)|^2]^{1/2} E[|\int_0^t \overline{D'_{M,n}(t-s)}\sigma(s)dW(s)|^2]^{1/2}\, dt.$$

Therefore, it is enough to prove the convergence to zero of

$$\rho(n)^{-1}E[|\int_0^t \int_0^s D_M(t-u)\gamma(u)dZ(u)D_{N,n}(t-s)\sigma(s)dW(s)|^2]\, E[|\int_0^t \overline{D'_{M,n}(t-s)}\sigma(s)dW(s)|^2]. \tag{2.76}$$

By the Itô isometry and the boundedness of the volatility and the volatility of volatility, the term (2.76) is smaller than

$$C\rho(n)^{-1}\int_0^t \int_0^s D_M^2(t-u)du D_{N,n}^2(t-s)ds \int_0^t |D'_{M,n}(t-s)|^2 ds \tag{2.77}$$

$$\cong \rho(n)^{-1}M^{-1}N^{-1}M\, o(1) = o(1),$$

where we have used Lemmas 2.1 $i)$ and $ii)$ and 2.3.

Step III: Orthogonality

It remains to prove that

$$\langle \rho(n)^{-1/4}(A_M^{(i)} + A_M^{(ii)} + B_{n,N,M}^{(ii)} + B_{n,N,M}^{(iii)}), W\rangle \to 0 \tag{2.78}$$

in probability. We show in details the convergence to zero of the bracket $\langle \rho(n)^{-1/4}B_{n,N,M}^{(ii)}, W\rangle$. The other terms are analogous.

It holds:

$$\langle \rho(n)^{-1/4}B_{n,N,M}^{(ii)}, W\rangle = \rho(n)^{-1/4}\int_0^{2\pi} \frac{1}{2\pi}\int_0^s Y_{N,n}(t,t)D'_{M,n}(s-t)\sigma(t)dW(t)\sigma(s)ds. \tag{2.79}$$

Thus

$$E[|\rho(n)^{-1/4}\int_0^{2\pi} \frac{1}{2\pi}\int_0^s Y_{N,n}(t,t)D'_{M,n}(s-t)\sigma(t)dW(t)\sigma(s)ds|^2]$$

$$= \rho(n)^{-1/2}E[\frac{1}{(2\pi)^2}\int_0^{2\pi}\sigma(s)\int_0^{2\pi}\sigma(s')\, Z_{N,M,n}(s,s)\, \overline{Z_{N,M,n}(s',s')}\, ds\, ds'], \tag{2.80}$$

where

$$Z_{N,M,n}(s,u) := \int_0^s Y_{N,n}(t,t)D'_{M,n}(u-t)\sigma(t)dW(t). \tag{2.81}$$

By symmetry, suppose $s' \leq s$. An application of the Itô formula gives that

$$Z_{N,M,n}(s,s)\, \overline{Z_{N,M,n}(s',s')}$$

is equal to

$$\int_0^{s'} Y_{N,n}^2(t,t) D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}v(t)dt \tag{2.82}$$

$$+\int_0^{s'} Z_{N,M,n}(t,s)\, Y_{N,n}(t,t)\overline{D'_{M,n}(s'-t)}\sigma(t)dW(t) \tag{2.83}$$

$$+\int_0^{s'} \overline{Z_{N,M,n}(t,s)}\, Y_{N,n}(t,t) D'_{M,n}(s'-t)\sigma(t)dW(t). \tag{2.84}$$

We study (2.82), which is equal to:

$$\int_0^{s'}\int_0^t D_{N,n}^2(t-u)v(u)du\, D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}v(t)dt \tag{2.85}$$

$$+\int_0^{s'} 2\int_0^t Y_{N,n}(u,t)\, D_{N,n}(t-u)\sigma(u)dW(u)\, D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}v(t)dt. \tag{2.86}$$

Consider (2.85). Remember that $s' \leq s$. Let, for any $\varepsilon > 0$, $s' < s-\varepsilon$ and consider

$$E[\int_0^{2\pi} ds\,\sigma(s)\int_0^{2\pi} ds' 1_{(s'<s-\varepsilon)}\sigma(s')\int_0^{s'}\int_0^t D_{N,n}^2(t-u)v(u)du\, D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}v(t)dt]. \tag{2.87}$$

Thus, noting that $\rho(n)^{-1}\int_0^t D_{N,n}^2(t-u)v(u)du = O_p(1)$ by Lemma 2.1, we obtain that (2.85) multiplied by $\rho(n)^{-1/2}$ reduces to the study of

$$\rho(n)^{1/2}E[\int_0^{2\pi} ds\,\sigma(s)\int_0^{2\pi} ds' 1_{(s'<s-\varepsilon)}\,\sigma(s')\int_0^{s'} D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}\,v^2(t)dt]$$

$$\leq C\rho(n)^{1/2}\int_0^{2\pi} ds\int_0^{2\pi} ds' 1_{(s'<s-\varepsilon)}E[|\int_0^{s'} D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}\,v^2(t)dt|]$$

$$\leq C\rho(n)^{1/2}\int_0^{2\pi} ds\int_0^{2\pi} ds' 1_{(s'<s-\varepsilon)}E[\int_0^{s'} |D'_{M,n}(s-t)|^2 v(t)dt]^{1/2}E[\int_0^{s'} |D'_{M,n}(s'-t)|^2 v(t)dt]^{1/2}$$

$$\cong \rho(n)^{1/2}M^{1/2}o(1)M^{1/2} = o(1),$$

where we have used Lemma 2.3 and the fact that $s' < s-\varepsilon$, for $\varepsilon > 0$.

Now consider the case for any $\varepsilon > 0$, $s-\varepsilon \leq s' \leq s$. We have

$$\rho(n)^{1/2}\int_0^{2\pi} ds\int_0^{2\pi} 1_{(s-\varepsilon\leq s'\leq s)}ds' E[|\int_0^{s'} D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}v^2(t)dt|]$$

$$\leq \rho(n)^{1/2}\int_0^{2\pi} ds\int_0^{2\pi} 1_{(s-\varepsilon\leq s'\leq s)}ds' E[\int_0^{s'} |D'_{M,n}(s-t)|^2 v(t)dt]^{1/2}E[\int_0^{s'} |D'_{M,n}(s'-t)|^2 v(t)dt]^{1/2}$$

$$\cong \rho(n)^{1/2}M^{1/2}M^{1/2}\,2\pi\varepsilon \cong \varepsilon.$$

Consider now the term (2.86). It holds

$$E[\int_0^{2\pi} ds\,\sigma(s)\int_0^{2\pi} ds'\sigma(s')\int_0^{s'} 2\int_0^t Y_{N,n}(u,t)\,D_{N,n}(t-u)\sigma(u)dW(u)D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}v(t)dt]$$

$$\leq C\int_0^{2\pi} ds\int_0^{2\pi} ds'\int_0^{s'} E[|\int_0^t Y_{N,n}(u,t)\,D_{N,n}(t-u)\sigma(u)dW(u)|]|D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}|\,dt$$

$$\leq C\int_0^{2\pi} ds\int_0^{2\pi} ds'\int_0^{s'}\left(\int_0^t\int_0^u D^2_{N,n}(t-u')du'D^2_{N,n}(t-u)du\right)^{1/2}|D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}|\,dt. \tag{2.88}$$

Now using that for any $\varepsilon>0$, $D^2_{N,n}(t-u')<C/N^2$ for $u'<t-\varepsilon$, if n is large enough, then

$$\left(\int_0^t\int_0^u D^2_{N,n}(t-u')du'D^2_{N,n}(t-u)du\right)^{1/2}=O(\rho(n)^{3/2}).$$

Moreover, it holds

$$\int_0^{2\pi} ds\int_0^{2\pi} ds'\int_0^{s'}|D'_{M,n}(s-t)\overline{D'_{M,n}(s'-t)}|\,dt \tag{2.89}$$

$$\leq\int_0^{2\pi} ds\int_0^{2\pi} ds'\left(\int_0^{s'}|D'_{M,n}(s-t)|^2dt\right)^{1/2}\left(\int_0^{s'}|D'_{M,n}(s'-t)|^2\,dt\right)^{1/2}. \tag{2.90}$$

Finally, we have obtained that (2.88) multiplied by $\rho(n)^{-1/2}$ is of order

$$\rho(n)^{-1/2}\rho(n)^{3/2}M=O(\rho(n)^{1/2}).$$

Consider now (2.83). The term (2.84) is analogous. It is enough to show that

$$\rho(n)^{-1}E[|\int_0^{s'} Z_{N,M,n}(t,s)\,Y_{N,n}(t,t)\overline{D'_{M,n}(s'-t)}\sigma(t)dW(t)|^2] \tag{2.91}$$

converges to 0. We study (2.91) using the Itô isometry and the Cauchy-Schwartz inequality. It holds

$$\rho(n)^{-1}E[\int_0^{s'}|Z_{N,M,n}(t,s)\,Y_{N,n}(t,t)\overline{D'_{M,n}(s'-t)}|^2v(t)dt]$$

$$\leq C\rho(n)^{-1}\int_0^{s'}E[|Z_{N,M,n}(t,s)|^4]^{1/2}\,E[Y^4_{N,n}(t,t)]^{1/2}\,|D'_{M,n}(s'-t)|^2\,dt.$$

Observe now that

$$E[Y^4_{N,n}(t,t)]^{1/2}\leq C\rho(n) \tag{2.92}$$

and

$$E[|Z_{N,M,n}(t,s)|^4] \leq CE[(\int_0^t |Y_{N,n}(u,u)D'_{M,n}(s-u)|^2 v(u)du)^2]$$

$$\leq CE[\int_0^t Y_{N,n}^4(u,u)v(u)du] \int_0^t |D'_{M,n}(s-u)|^4 du \leq C\rho(n)^2 M^3,$$

where we have used Burkoholder-Davis-Gundy inequality, the boundedness of the volatility and Lemma 2.3. Therefore

$$E[|Z_{N,M,n}(t,s)|^4]^{1/2} \leq C\rho(n)^{1/4}.$$

Finally, it holds that the term (2.91) is $O(\rho(n)^{1/4})$. $\square$

Proof of Theorem 2.2

The proof is obtained along the lines of Theorem 2.1 by substituting the Dirichlet kernel with the Fejér kernel (2.108). In particular, with the same notations, the estimator (2.5) is studied as the sum of three terms:

$$\frac{2\pi}{M+1}\sum_{|k|\leq M}\left(1-\frac{|k|}{M+1}\right) \mathrm{i}kA_{k,n}\, c_{-k}(dx_n) \tag{2.93}$$

$$\frac{2\pi}{M+1}\sum_{|k|\leq M}\left(1-\frac{|k|}{M+1}\right) \mathrm{i}kB_{k,n,N}\, c_{-k}(dx_n) \tag{2.94}$$

$$\frac{2\pi}{M+1}\sum_{|k|\leq M}\left(1-\frac{|k|}{M+1}\right) \mathrm{i}kC_{k,n,N}\, c_{-k}(dx_n). \tag{2.95}$$

We detail the main differences in the proof, which appear in the computation of the asymptotic variance. First of all, we have to consider the terms corresponding to (2.36) and (2.37), which read

$$\widetilde{A}_M^{(i)}(u) := \frac{1}{2\pi}\frac{1}{M+1}\int_0^u \int_0^t F_M(t-s)dv(s)\,dx(t) \tag{2.96}$$

$$\widetilde{A}_M^{(ii)}(u) := \frac{1}{2\pi}\frac{1}{M+1}\int_0^u \int_0^t F_M(t-s)dx(s)\,dv(t), \tag{2.97}$$

such that

$$\frac{2\pi}{M+1}\sum_{|k|\leq M}\left(1-\frac{|k|}{M+1}\right) c_k(dv)c_{-k}(dx) - \frac{1}{2\pi}\int_0^{2\pi}\eta(t)dt = \widetilde{A}_M^{(i)}(2\pi)+\widetilde{A}_M^{(ii)}(2\pi).$$

In order to illustrate the result, consider $\langle M^{1/2}\widetilde{A}_M^{(i)}, M^{1/2}\widetilde{A}_M^{(i)}\rangle_{2\pi}$. It holds:

$$\langle M^{1/2}\widetilde{A}_M^{(i)}, M^{1/2}\widetilde{A}_M^{(i)}\rangle_{2\pi} = M\frac{1}{(2\pi)^2}\int_0^{2\pi}\left(\int_0^t \frac{1}{M+1}F_M(t-s)dv(s)\right)^2 v(t)dt$$

$$= M\frac{1}{(2\pi)^2}\int_0^{2\pi}\frac{1}{(M+1)^2}\int_0^t F_M^2(t-s)\gamma^2(s)ds\, v(t)dt \tag{2.98}$$

$$+M\frac{1}{(2\pi)^2}\int_0^{2\pi}2\frac{1}{(M+1)^2}\int_0^t\int_0^s F_M(t-u)dv(u)\,F_M(t-s)dv(s)\,v(t)dt. \tag{2.99}$$

Using Lemma 2.4 , then (2.98) converges in probability to

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{3}\gamma^2(t)v(t)dt, \tag{2.100}$$

while (2.99) is $o_p(1)$. Considering the bracket $\langle M^{1/2}\widetilde{A}_M^{(i)}, M^{1/2}\widetilde{A}_M^{(ii)}\rangle_{2\pi}$, we obtain:

$$M\frac{1}{(2\pi)^2}\frac{1}{(M+1)^2}\int_0^{2\pi}\int_0^t F_M(t-s)dx(s)\int_0^t F_M(t-s)dv(s)\,\eta(t)dt$$

$$= M\frac{1}{(2\pi)^2}\frac{1}{(M+1)^2}\int_0^{2\pi}\int_0^t F_M^2(t-s)\eta(s)ds\,\eta(t)dt \tag{2.101}$$

$$+M\frac{1}{(2\pi)^2}\frac{1}{(M+1)^2}\int_0^{2\pi}\int_0^t\int_0^s F_M(t-u)dv(u)\,F_M(t-s)dx(s)\,\eta(t)dt \tag{2.102}$$

$$+M\frac{1}{(2\pi)^2}\frac{1}{(M+1)^2}\int_0^{2\pi}\int_0^t\int_0^s F_M(t-u)dx(u)\,F_M(t-s)dv(s)\,\eta(t)dt. \tag{2.103}$$

By Lemma 2.4, the term (2.101) converges to

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{3}\eta^2(t)dt,$$

while (2.102) and (2.103) are $o_p(1)$.

Finally, noting that $\rho(n)^{-1/4}\sim(c_M)^{-1/2}M^{1/2}$, we conclude that

$$\langle\rho(n)^{-1/4}(\widetilde{A}_M^{(i)}+\widetilde{A}_M^{(ii)}),\rho(n)^{-1/4}(\widetilde{A}_M^{(i)}+\widetilde{A}_M^{(ii)})\rangle_{2\pi}\to\frac{2}{3}\frac{1}{c_M}\frac{1}{2\pi}\int_0^{2\pi}(\gamma^2(s)v(s)+\eta^2(s))ds \tag{2.104}$$

in probability.

The second contribution to the asymptotic variance comes from the analogous of terms (2.23) and (2.24), where the Dirichlet kernel is substituted with the Fejér kernel. It is enough to check the convergence of

$$\rho(n)^{-1/2}\langle\widetilde{B}_{n,N,M}^{(ii)},\overline{\widetilde{B}_{n,N,M}^{(ii)}}\rangle_{2\pi}, \tag{2.105}$$

where

$$\widetilde{B}_{n,N,M}^{(ii)}:=\int_0^{2\pi}\frac{1}{2\pi}\left(\frac{1}{M+1}\int_0^s F'_{M,n}(s-t)\sigma(t)dW(t)\right)\left(\int_0^s D_{N,n}(s-u)\sigma(u)dW(u)\right)\sigma(s)dW(s).$$

The other terms are analogous. Using the Itô formula, it is shown that (2.105) is equal to

$$\rho(n)^{-1/2}\int_0^{2\pi}\frac{1}{(2\pi)^2}\int_0^s\frac{1}{(M+1)^2}\,|F'_{M,n}(s-t)|^2 v(t)dt\int_0^s D^2_{N,n}(s-u)v(u)du\,v(s)ds \tag{2.106}$$

plus some terms which are $o_p(1)$.

By Lemmas 2.1 and 2.4 it holds in probability

$$n\int_0^s D^2_{N,n}(s-u)v(u)du \to \pi(1+2\theta(c_N/\pi))v(s)$$

and

$$\frac{1}{M}\int_0^s\frac{1}{(M+1)^2}\,|F'_{M,n}(s-t)|^2 v(t)dt \to \frac{\pi}{15}\,v(s).$$

Therefore, noting that $\rho(n)^{-1/2}\sim(c_M)^{-1}M$, the term (2.106) gives:

$$\frac{1}{30}c_M(1+2\theta(c_N/\pi))\,\frac{1}{2\pi}\int_0^{2\pi}\sigma^6(s)ds.$$

Combining the analogous four terms, gives the result

$$\frac{2}{15}c_M(1+2\theta(c_N/\pi))\,\frac{1}{2\pi}\int_0^{2\pi}\sigma^6(s)ds.$$

The remaining parts of the proof follow along the same lines of Theorem 2.1. □

2.7.2 Properties of the Dirichlet and Fejér kernels

This subsection of the Appendix contains some results on the Dirichlet kernel, defined as

$$D_N(x) := \frac{1}{2N+1}\sum_{|k|\le N} e^{\mathrm{i}kx} = \frac{1}{2N+1}\frac{\sin((2N+1)x/2)}{\sin(x/2)} \tag{2.107}$$

and the Fejér kernel, defined as

$$F_N(x) := \sum_{|k|\le N}\left(1-\frac{|k|}{N+1}\right)e^{\mathrm{i}kx} = \frac{1}{N+1}\sum_{|k|\le N}\left(1-\frac{|k|}{N+1}\right)e^{\mathrm{i}kx}. \tag{2.108}$$

Dirichlet kernel

Lemma 2.1. *Under the condition $N/n \to a > 0$:*
i) for any $p > 1$ there exists a constant C_p such that

$$\lim_{n,N} n \sup_{x\in[0,2\pi]} \int_0^{2\pi} |D_{N,n}(x-y)|^p dy \leq C_p,$$

ii) it holds:

$$\lim_{n,N} n \int_0^x D^2_{N,n}(x-y)dy = \pi(1+2\theta(2a))$$

and, for any α-Hölder continuous function f, with $\alpha \in (0,1]$:

$$\lim_{N,n} n \int_0^x D^2_{N,n}(x-y) f(y)dy = \pi(1+2\theta(2a)) f(x),$$

where

$$\theta(a) := \frac{1}{2a^2} r(a)(1-r(a)), \tag{2.109}$$

being $r(a) = a-[a]$, with $[a]$ the integer part of a,
iii) for any $\varepsilon > 0$:

$$\lim_{N,n} n \int_0^{x-\varepsilon} D^2_{N,n}(x-y)dy = 0.$$

Proof. See Clement and Gloter (2011) Lemma 1 and Lemma 4.

Lemma 2.2. *i) Under the condition $M^2/n \to a > 0$, then*

$$\lim_{M,n} M \int_0^{2\pi} D^2_{M,n}(y)dy = \lim_{M} M \int_0^{2\pi} D^2_{M}(y)dy = 2\pi.$$

ii) For any α-Hölder continuous function f, with $\alpha \in (0,1]$ we have

$$\lim_{M,n} M \int_0^x D^2_{M,n}(x-y) f(y)dy = \lim_{M} M \int_0^x D^2_{M}(x-y) f(y)dy = \pi f(x).$$

iii) for any $\varepsilon > 0$:

$$\lim_{M,n} M \int_0^{x-\varepsilon} D^2_{M,n}(x-y)dy = \lim_{M} M \int_0^{x-\varepsilon} D^2_{M}(t-y)dy = 0.$$

Proof. The proof follows from Lemma 5.1 in Cuchiero and Teichmann (2015).

Lemma 2.3. *Let $D'_M(x)$ denote the derivative of the Dirichlet kernel. Under the condition $\lim_{n,M\to\infty} M^2/n = a > 0$, it holds:*
i)
$$\lim_{n,M} \frac{1}{M} \int_0^{2\pi} |D'_{M,n}(x)|^2 dx = \lim_{M} \frac{1}{M} \int_0^{2\pi} |D'_M(x)|^2 dx = \frac{\pi}{3},$$
ii)
$$\lim_{n,M} \frac{1}{M^3} \int_0^{2\pi} |D'_{M,n}(x)|^4 dx = \lim_{M} \frac{1}{M^3} \int_0^{2\pi} |D'_M(x)|^4 dx = \frac{4\pi}{105}.$$

Proof. As for i) it holds:
$$\frac{1}{M}\int_0^{2\pi} D'_M(x)\overline{D'_M(x)}dx = \frac{1}{M}\frac{2\pi}{(2M+1)^2}\sum_{|k|\le M} k^2 = \frac{1}{M}\frac{2\pi}{(2M+1)^2}\frac{M(M+1)(2M+1)}{3} = \frac{2}{3}\pi\frac{M+1}{2M+1} \to \frac{1}{3}\pi.$$

In order to prove ii), observe that
$$\int_0^{2\pi} |D'_{M,n}(x)|^4 dx = \frac{1}{(2M+1)^4}\sum_{|k|\le M}\sum_{|h|\le M}\sum_{|l|\le M}\sum_{|m|\le M}\int_0^{2\pi} k\,h\,l\,m\,e^{\mathrm{i}(k+h-l-m)x}dx,$$

as $\overline{\mathrm{i}ke^{\mathrm{i}kx}} = -\mathrm{i}ke^{-\mathrm{i}kx}$, for any k. Then, as the integral is null except for $m = k+h-l$, it holds:
$$\frac{1}{M^3}\int_0^{2\pi} |D'_{M,n}(x)|^4 dx = 2\pi\frac{1}{M^3}\frac{1}{(2M+1)^4}\sum_{|k|\le M}\sum_{|h|\le M}\sum_{\substack{-M\vee(-M+k+h)\le\\ \le l\le M\wedge(M+k-h)}} k\,h\,l\,(k+h-l)$$
$$=^{9} \frac{(6+30M+101M^2+113M^3+68M^4+32M^5)\pi}{105M^2(1+2M)^3} \to \frac{4\pi}{105}.$$

Fejér kernel

Lemma 2.4. *Under the condition $M^2/n \to a > 0$, it holds:*
i)
$$\lim_{M,n\to\infty}\int_0^{2\pi} F_{M,n}(x)dx = \lim_{M\to\infty}\int_0^{2\pi} F_M(x)dx = 2\pi,$$
ii)
$$\lim_{M,n\to\infty}\int_0^{2\pi}\frac{1}{M}F^2_{M,n}(x)dx = \lim_{M\to\infty}\int_0^{2\pi}\frac{1}{M}F^2_M(x)dx = \frac{4\pi}{3}, \tag{2.110}$$
iii) for any α-Hölder continuous function f, with $\alpha \in (0,1]$, we have
$$\lim_{M,n}\frac{1}{M}\int_0^{2\pi} F^2_{M,n}(x-y)f(y)dy = \lim_{M}\frac{1}{M}\int_0^{2\pi} F^2_M(x-y)f(y)dy = \frac{4\pi}{3}f(x).$$

[9] This identity can be verified using, e.g., the software Mathematica.

iv)

$$\lim_{n,M} \frac{1}{M^3} \int_0^{2\pi} |F'_{M,n}(x)|^2 dx = \lim_{M} \frac{1}{M^3} \int_0^{2\pi} |F'_M(x)|^2 dx = \frac{1}{3}\pi$$

Proof. For i), ii) and iii) see, e.g., Lemma 5.1 in Cuchiero and Teichmann (2015). As for iv), it is enough to observe that

$$\frac{1}{M^3} \int_0^{2\pi} |F'_M(x)|^2 dx = \frac{1}{M^3} 2\pi \sum_{|k|\le M} (1 - \frac{|k|}{M+1})^2 k^2 = \frac{1}{M^3} \frac{2\pi}{(M+1)^2} 2 \frac{M(M^3+4M^2+6M+4)}{30(M+1)} \to \frac{2}{15}\pi.$$

3. THE PRICE-LEVERAGE COVARIATION AS A MEASURE OF THE RESPONSE OF THE LEVERAGE EFFECT TO PRICE AND VOLATILITY CHANGES: EMPIRICAL EVIDENCE

3.1 Introduction

Empirical evidence collected in the literature suggests that the leverage effect is time-varying. For instance, Kalnina and Xiu (2017) point out that the the intensity of the leverage effect gets stronger in turbulent periods, that is, in correspondence of volatility spikes or large returns, while Bandi and Renò (2012) model the leverage process as a function of the stochastic volatility of the asset, based on empirical evidence. Thus, in order to get insight into the time-varying dynamics of the leverage process, it may be interesting to study its sensitivity to increments of the volatility or the price.

This can be done analytically in the case of the Constant Elasticity of Variance (CEV) model by Beckers (1980). The CEV model is possibly the most popular example in the class of level-dependent models, that is, models that treat the volatility process as a deterministic function of the price process. Level-dependent models represent a parsimonious and analytically-tractable tool to reproduce some stylized facts of financial markets, e.g., the implied volatility smile (see, e.g., Derman and Kani (1994); Dupire (1994); Hobson and Rogers (1998)). More recently, a level-dependent model driven by a Fractional Brownian motion has also been introduced, with the aim of reproducing the empirically-observed long-memory property of the volatility (see Araneda (2020)). Specifically, the CEV model is explicitly designed to capture leverage effects. Moreover, under the CEV model, the leverage process can be viewed as a deterministic differentiable function of either the volatility or the log-price, thereby allowing the computation of its analytical derivative with respect to any of these two processes.

In this regard, simple calculations show that both these analytical derivatives depend on the same quantity: the price-leverage covariation. In particular, it emerges that the derivative of the leverage with respect to the price (respectively, the volatility) is equal to the ratio of the price-leverage covariation and the volatility (respectively, the leverage). This is crucial from the point of view of estimation, in that such derivatives can be rewritten in terms of quantities which can be consistently estimated from asset prices sampled over a fixed time horizon. Additionally, it also emerges that the price-leverage covariation is equal to twice the vol-of-vol process. However, the result related to the analytical derivative of the leverage with respect to the volatility holds more generally. In fact, for this derivative to be equal to the ratio of the price-leverage covariation and the leverage, it is sufficient to assume that the data-generating process is any stochastic volatility model with continuous paths where the vol-of-vol process is a multiple of a power of the variance process Popular, widely-used examples of stochastic volatility models with this feature, beyond the CEV model, are the model by Heston (1993), the 3/2 model by Platen (1997) and the continuous-time GARCH model by Nelson (1990). Also, in this more general semi-parametric framework, the price-leverage covariation is still a linear function of the vol-of-vol.

The price-leverage covariation has first been studied in Barucci et al. (2003), where the authors derive a model-free indicator of financial instability whose analytical expression depends, other than on the volatility and the leverage, on the price-leverage covariation. However, only recently Sanfelici and Mancino (2020) have provided a consistent non-parametric estimator of the price-leverage covariation, based on the Fourier method by Malliavin and Mancino (2002).

The existence of a theoretical, model-dependent link between the price-leverage covariation and the sensitivity (i.e., the derivative) of the leverage process with respect to the price and the volatility motivates an empirical, model-free investigation of this link. Accordingly, in this chapter we conduct this investigation on the sample of S&P500 1-second prices over the period March, 2018 - April, 2018. As a result, we uncover the existence of a statistically-significant linear relationship between the theoretical derivative of the leverage with respect to the price (respectively, the volatility) in the CEV framework and the corresponding numerical derivative of the leverage, computed via finite differences. Remarkably, estimated regressions coefficients are close to 1, suggesting that theoretical predictions provide an accurate proxy of the true derivatives of the leverage for the sample object of study. Note that, to be able to perform this empirical investigation, we have reconstructed the unobservable paths of the volatility, the leverage and the price-leverage covariation from high-frequency prices in a non-parametric fashion through the Fourier methodology (see, respectively, Malliavin and Mancino (2002, 2009); Barucci and Mancino (2010); Sanfelici and Mancino (2020)).

Based on these empirical findings, the price-leverage covariation can be interpreted, from a financial standpoint, as a model-free measure of the responsiveness of the leverage effect to the arrival of new information on the market that causes changes in the price or in the amount of risk perceived by market participants (i.e., in the volatility). Further, additional empirical results suggest that the price-leverage covariation is approximately equal to twice the vol-of-vol for the sample of object of study. Again, this results is line with the prediction of the CEV model, which implies - as already mentioned - that the price-leverage covariation is exactly equal to twice the vol-of-vol. Interpreting the vol-of-vol as the uncertainty about the actual level of risk perceived on the market, this finding suggests that the response of the leverage effect to changes in the price or the volatility is proportional to the intensity of this uncertainty: the larger the latter, the stronger the response of the leverage (and viceversa). Finally, note that the path of the vol-of-vol has also been reconstructed non-parametrically using the Fourier methodology (see Sanfelici et al. (2015)) for this empirical analysis.

This chapter is organized as follows. In Section 3.2 we derive the analytical expressions of the derivatives of the leverage with respect to the price and the volatility under the CEV model. In Section 3.3 we give a brief description of the non-parametric Fourier estimators of the spot volatility, leverage, vol-of-vol and price-variance covariation and recall their asymptotic properties. Sections 3.4 and 3.5 contain, respectively, numerical and empirical results. Finally, Section 3.6 concludes.

3.2 Analytical derivatives of the leverage in the CEV framework

Let $X(t)$ denote the price process and assume that its dynamics follow the CEV model, that is,

$$dX(t) = \sigma X(t)^{\delta} dW(t) + \mu X(t) d(t), \tag{3.1}$$

where W is a Brownian motion on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$, satisfying the usual conditions, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\delta > 0$. Note that the role of the parameter δ is crucial, as it captures leverage effects. Specifically, if $\delta < 1$, the price and the volatility are negatively correlated, as it commonly happens on equity markets. Instead, if $\delta < 1$, the price and the volatility move in the same direction, according to the so-called inverse leverage effect, a phenomenon usually observed on commodity markets.

Define $x(t) := ln(X(t))$. Under model (3.1), the following expressions are obtained for the volatility $\nu(t)dt := \langle x,x\rangle_t$, the leverage $\eta(t)dt := \langle x,\nu\rangle_t$, the vol-of-vol $\xi(t)dt := \langle \nu,\nu\rangle_t$ and the price-leverage covariation $\chi(t)dt := \langle x,\eta\rangle_t$:

$$\nu(t) = \sigma^2 e^{2(\delta-1)x(t)}, \tag{3.2}$$

$$\eta(t) = 2(\delta-1)\nu(t)^2, \tag{3.3}$$

$$\chi(t) = 8(\delta-1)^2\nu(t)^3, \tag{3.4}$$

$$\xi(t) = 4(\delta - 1)^2 \nu(t)^3. \tag{3.5}$$

Therefore, the derivatives of the leverage $\eta(t)$ with respect to the log-price $x(t)$ and the volatility $\nu(t)$ read:

$$\frac{\partial \eta(t)}{\partial x(t)} = \frac{\chi(t)}{\nu(t)}, \tag{3.6}$$

$$\frac{\partial \eta(t)}{\partial \nu(t)} = \frac{\chi(t)}{\eta(t)}. \tag{3.7}$$

Based on equations (3.6) and (3.7), in the CEV framework $\chi(t)$ could be interpreted, from a financial point of view, as the process that captures the response of the leverage to the arrival of new information that causes changes in the volatility and/or the price.

Further, note that the derivative of the leverage with respect to the price in equation (3.6) is strictly positive, since it is equal to the ratio of two strictly positive processes, $\nu(t)$ (see (3.2)) and $\chi(t)$ (see (3.4)). This implies that on equity markets the leverage effect increases (i.e., the leverage process becomes more negative) in correspondence of a negative return, and viceversa. Instead, the sign of the derivative of the leverage with respect to the volatility in equation (3.7) depends on the sign of $\eta(t)$. Therefore, if at some point in time $\eta(t)$ is negative, it becomes more (respectively, less) negative in correspondence of an increment (respectively, reduction) of the volatility. Overall, model-dependent predictions of the sensitivity of the leverage effect to the price and the volatility in equations (3.6) and (3.7) are consistent with the empirical findings by Kalnina and Xiu (2017) and Bandi and Renò (2012) related to time-varying leverage effects.

Additionally, based on equations (3.4) and (3.5), the process $\chi(t)$ is simply equal to twice the vol-of-vol $\xi(t)$ in the CEV framework. From a financial standpoint, this linear link could be interpreted as follows. Taking the volatility as a measure of market risk and the vol-of-vol as a proxy of the uncertainty about the actual level of market risk perceived by market operators, the larger is the latter, the more intense is the response of the leverage to price and market risk changes, as captured by $\chi(t)$.

Remark 3.1. *For the result in equation (3.7) to hold and for the price-leverage covariation to be a linear function of the vol-of-vol, it is sufficient to assume that the log-price and the volatility are continuous semimartingales driven by two Brownian motions with constant non-zero correlation parameter and that the diffusion component of the volatility process is a multiple of a power of the volatility process itself. Formally, assume that*

$$\begin{aligned} dx(t) &= \sqrt{\nu(t)}dW(t) + a(t)dt \\ d\nu(t) &= \gamma\nu(t)^{\beta}dZ(t) + b(t)dt \\ d\langle W, Z\rangle_t &= \rho dt \end{aligned} \tag{3.8}$$

where: W and Z are Brownian motions on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$*, which satisfies the usual conditions; a, b and* ν *are continuous adapted processes*[1]*;* $\rho \in [-1,1] - \{0\}$*,* $\gamma \in \mathbb{R}$ *and* $\beta \in \mathbb{R}$*. Then it follows that:*

$$\xi(t) = \gamma^2 \nu(t)^{2\beta}, \tag{3.9}$$

$$\eta(t) = \rho\sqrt{\nu(t)}\sqrt{\xi(t)}, \tag{3.10}$$

$$\chi(t) = \Big(\beta + \frac{1}{2}\Big)\rho^2\xi(t). \tag{3.11}$$

Therefore:

$$\frac{\partial\eta(t)}{\partial\nu(t)} = \frac{\chi(t)}{\eta(t)}. \tag{3.12}$$

The semi-parametric specification (3.8) contains the class of stochastic volatility models where the volatility is a CKLS process (Chan et al. (1992)), such as the Heston model, the continuous-time GARCH model and the 3/2 model. Moreover, it contains also the CEV model, as in (3.1), which indeed can be rewritten as:

$$\begin{aligned} dx(t) &= \sigma(t)dW(t) + \Big(\mu - \frac{1}{2}\nu(t)\Big)dt \\ d\nu(t) &= \gamma\nu(t)^{\beta}dW(t) + \gamma\nu(t)\Big(\mu - \frac{1}{2}(\gamma+1)\nu(t)\Big)dt \end{aligned} \tag{3.13}$$

where $\gamma = 2(\delta - 1)$ *and* $\beta = 3/2$*.*

Assuming the CEV model as the data-generating process, consistent estimators of the derivatives (3.6) and (3.7) can be built as the ratio of non-parametric estimators of $\chi(t)$ and, respectively, $\nu(t)$ or $\eta(t)$. We address this aspect in the next section, using the Fourier methodology.

[1] The condition $\nu(t) > 0$ a.s., which is clearly desirable from a financial standpoint, may impose some additional constraints on the parametric form of the drift b (see, e.g., Feller (1951) for the case when b has a mean-reverting structure). However, any additional constraints on the parametric form of b do not interfere with the computations of $\xi(t)$, $\eta(t)$ and $\chi(t)$, as they depend only on the diffusion components of the price and the volatility.

3.3 Fourier-based estimation of the analytical derivatives of the leverage in the CEV framework

The Fourier method, introduced by Malliavin and Mancino (2002), is particularly well-suited to build non-parametric estimators of second-order and third-order quantities. As a first step, one obtains estimates of the Fourier coefficients of the latent volatility $\nu(t)$. Then, the knowledge of these coefficients allows iterating the procedure to compute the Fourier coefficients of the second-order quantities $\xi(t)$ and $\eta(t)$. Finally, a third iteration yields estimates of the coefficients of the third-order quantity $\chi(t)$. In this regard, it is worth noting that these progressive iterations do not involve any differentiation procedure for the pre-estimation of the spot volatility (in order to estimate second-order quantities) or the spot leverage (in order to estimate the third-order quantity $\chi(t)$). Instead, they only require the pre-estimation of integrated quantities, namely the Fourier coefficients. Given the numerical instabilities which are typically linked to differentiation procedures, this feature represents a strength of the Fourier methodology, compared to the realized approach for the estimation of spot processes (see Chapter 8 in Aït-Sahalia and Jacod (2014) for a detailed description of realized spot estimators and their asymptotic properties).

The Fourier estimators of $\nu(t), \eta(t), \xi(t)$ and $\chi(t)$, which we illustrate in the following, are termed non-parametric in that, for their asymptotic properties to hold, they only require that the processes $x(t)$, $\nu(t)$ and $\eta(t)$ are continuous semimartingales. Formally, we assume that:

$$
\begin{aligned}
dx(t) &= \sqrt{\nu(t)}dW(t) + a(t)dt \\
d\nu(t) &= \gamma(t)dZ(t) + b(t)dt \\
d\eta(t) &= \lambda(t)dY(t) + c(t)dt
\end{aligned}
$$

where W, Z and Y are correlated Brownian motions on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$, which satisfies the usual conditions, while the processes a, b, c, ν, γ and λ are continuous, adapted and bounded in absolute value.

In the following subsections, after briefly illustrating the Fourier estimators of $\nu(t)$, $\eta(t)$, and $\chi(t)$ and recalling their asymptotic properties, we derive consistent estimators of the derivatives (3.6) and (3.7) as the ratio of the Fourier estimators of $\chi(t)$ and, respectively, $\nu(t)$ or $\eta(t)$. The Fourier estimator of the vol-of-vol $\xi(t)$ is also illustrated, as it will be used in the empirical study of Section 3.5.

3.3.1 Fourier estimator of the volatility

Assume that the log-price process $x(t)$ is observable on the grid of mesh size $\rho(n) := 2\pi/n$ over the interval $[0,2\pi]$[2]. Then, for $|k| \leq N$, the k-th (discrete) Fourier coefficient of the volatility is defined as

$$c_k(\nu_{n,N}) := \frac{2\pi}{2N+1} \sum_{|s| \leq N} c_s(dx_n)\, c_{k-s}(dx_n), \tag{3.14}$$

where for any integer $k, |k| \leq 2N, c_k(dx_n)$ is the k-th (discrete) Fourier coefficient of the log-return process, namely

$$c_k(dx_n) := \frac{1}{2\pi} \sum_{j=0}^{n-1} e^{-\mathrm{i}kt_{j,n}} \delta_j^n(x), \tag{3.15}$$

where $\delta_j^n(x) := x_{t_{j+1},n} - x_{t_j,n}$, $t_{j,n} = j\frac{2\pi}{n}$, $j = 0,1,\ldots,n$, while the symbol i denotes the imaginary unit, that is, $\mathrm{i} = \sqrt{-1}$.

Once the Fourier coefficients of the volatility (3.14) have been computed, the application of the Fourier-Fejér inversion formula allows reconstructing the volatility path. The definition of the Fourier spot volatility estimator is as follows.

Definition 3.1. ***Fourier estimator of the spot volatility***

The Fourier estimator of the spot volatility process is defined as the random function of time

$$\widehat{\nu}_{n,N,S_\nu}(t) := \sum_{|k|<S_\nu} \left(1 - \frac{|k|}{S_\nu}\right) c_k(\nu_{n,N})\, e^{\mathrm{i}kt}, \tag{3.16}$$

where S_ν is a positive integer smaller than N, while $c_k(\nu_{n,N})$ is defined in (3.14).

The following theorem demonstrates the uniform consistency of the Fourier estimator of the spot volatility (3.16).

Theorem 3.1. *For any integer $|k| \leq N$, if $N/n \to 1/2$, the following convergence in probability holds*

$$\lim_{n,N\to\infty} c_k(\nu_{n,N}) = c_k(\nu),$$

[2]In applications, we can always assume that the price process $x(t)$ is observed on $[0,2\pi]$ by rescaling the actual time interval. Moreover, note that in this Chapter, to simplify the exposition, we assume that the price process is observable on an equally-spaced grid, but, in general, the Fourier method works also with unequally-spaced samples (see Mancino et al. (2017)).

where $c_k(\nu)$ is the k-th Fourier coefficient of the volatility process $\nu(t)$. Moreover, if $N/n \to 1/2$ and $S_\nu/n \to 0$, it holds in probability that

$$\lim_{n,N,S_\nu \to \infty} \sup_{t \in (0,2\pi)} |\widehat{\nu}_{n,N,S_\nu}(t) - \nu(t)| = 0.$$

Proof. See Malliavin and Mancino (2009). □

3.3.2 Fourier estimator of the leverage

As mentioned, the knowledge of the Fourier coefficients of the latent instantaneous volatility $\nu(t)$ allows treating the latter as an observable process and iterate the procedure for computing the Fourier coefficients in order to reconstruct the leverage process $\eta(t)$. In particular, to estimate the instantaneous leverage $\eta(t)$ we exploit the multivariate version of Fourier method introduced in Malliavin and Mancino (2009). Accordingly, an estimator of the Fourier coefficients of the leverage is given by

$$c_k(\eta_{n,N,M}) := \frac{2\pi}{2M+1} \sum_{|j| \le M} c_j(dx_n)\, c_{k-j}(d\nu_{n,N}), \tag{3.17}$$

where M is a positive integer smaller than N, $c_j(dx_n)$ is given in (3.15) and we use the approximation $c_j(d\nu_{n,N}) \cong \mathrm{i} j c_j(\nu_{n,N})$[3]. Then the following theorem holds.

Theorem 3.2. *If $N/n \to 1/2$ and $M^2/n \to 0$ for $n, N, M \to \infty$, then the following convergence in probability holds*

$$\lim_{n,N,M \to \infty} c_k(\eta_{n,N,M}) = c_k(\eta),$$

where $c_k(\eta)$ is the k-th Fourier coefficient of the leverage process $\eta(t)$.

Proof. See Barucci and Mancino (2010). □

Finally, a consistent estimator of the the instantaneous leverage $\eta(t)$ is obtained as

$$\widehat{\eta}_{n,N,M,S_\eta}(t) := \sum_{|k| < S_\eta} \left(1 - \frac{|k|}{S_\eta}\right) c_k(\eta_{n,N,M})\, e^{\mathrm{i}kt}, \tag{3.18}$$

where S_η is a positive integer smaller than M, while $c_k(\eta_{n,N,M})$ is defined in (3.17).

[3] See, e.g., Chapter 6 in Mancino et al. (2017).

3.3.3 Fourier estimator of the vol-of-vol

The knowledge of the coefficients of the volatility process $\nu(t)$ also allows building an estimator of its quadratic variation, the vol-of-vol $\xi(t)$. In particular, an estimator of the coefficents of $\xi(t)$ is given by

$$c_k(\xi_{n,N,M}) := \frac{2\pi}{2M+1} \sum_{|j|\leq M} c_j(d\nu_{n,N})\, c_{k-j}(d\nu_{n,N}), \tag{3.19}$$

where, again, $c_j(d\nu_{n,N})$ is approximated with $\mathrm{i}jc_j(\nu_{n,N})$. Then the following theorem holds.

Theorem 3.3. *If $N/n \to 0$ and $M^4/N \to 0$ for $n,N,M \to \infty$, then the following convergence in probability holds*

$$\lim_{n,N,M\to\infty} c_k(\xi_{n,N,M}) = c_k(\xi),$$

where $c_k(\xi)$ is the k-th Fourier coefficient of the vol-of-vol process $\xi(t)$.

Proof. See Sanfelici et al. (2015). □

Finally, a consistent estimator of the the spot vol-of-vol $\xi(t)$ can be obtained as

$$\widehat{\xi}_{n,N,M,S_\xi}(t) := \sum_{|k|<S_\xi} \left(1 - \frac{|k|}{S_\xi}\right) c_k(\xi_{n,N,M})\, e^{\mathrm{i}kt} \tag{3.20}$$

where S_ξ is a positive integer smaller than M, while $c_k(\xi_{n,N,M})$ is defined in (3.19).

3.3.4 Fourier estimator of the price-leverage covariation

Similarly to what we have done for the volatility process $\nu(t)$, once its Fourier coefficients have been estimated, we can treat the second-order quantity $\eta(t)$ as an observable process and exploit the multivariate Fourier method to estimate the third-order quantity $\chi(t)$. The following asymptotic result is obtained.

Theorem 3.4. *If $N/n \to 1/2$ and $L^2M^2/N \to 0$ for $n,N,M,L \to \infty$[4], then the following convergence in probability holds*

$$\lim_{n,N,M,L\to\infty} c_k(\chi_{n,N,M,L}) = c_k(\chi),$$

[4] Note that these conditions also imply that $M^2/n \to 0$, satisfying the hypotheses of Theorem 3.2.

where, for L positive integer and smaller than M, we define

$$c_k\left(\chi_{n,N,M,L}\right) := \frac{2\pi}{2L+1} \sum_{|j|\leq L} c_j\left(dx_n\right) \mathrm{i} j c_{k-j}\left(\eta_{n,N,M}\right).$$

Proof. See Sanfelici and Mancino (2020). □

Accordingly, a consistent spot estimator of the process $\chi(t)$ is defined as

$$\widehat{\chi}_{n,N,M,L,S_\chi}(t) := \sum_{|k|<S_\chi} \left(1 - \frac{|k|}{S_\chi}\right) c_k\left(\chi_{n,N,M,L}\right) e^{\mathrm{i}kt}, \tag{3.21}$$

where S_χ is a positive integer smaller than L.

3.3.5 Fourier estimators of the derivatives of the leverage

The Continuous Mapping Theorem ensures that the ratio of the non-parametric Fourier estimators (3.21) and (3.16), i.e.,

$$\frac{\widehat{\chi}_{n,N,M,L,S_\chi}(t)}{\widehat{v}_{n,N,S_v}(t)} \tag{3.22}$$

is a consistent estimator of the derivative of the leverage process with respect to the log-price process under (3.1), as given in (3.6). Analogously, it also ensures that the ratio of the non-parametric Fourier estimators (3.21) and (3.18), i.e.,

$$\frac{\widehat{\chi}_{n,N,M,L,S_\chi}(t)}{\widehat{\eta}_{n,N,M,S_\eta}(t)}, \tag{3.23}$$

is a consistent estimator of the derivative of the leverage process with respect to the log-price process under (3.8), as given in (3.7)[5].

3.4 Simulation study

Given the availability of consistent Fourier estimators of the volatility, the leverage and the derivatives of the leverage with respect to the log-price or the volatility, a simple test to check with empirical data if the true model-free derivatives of the leverage match the corresponding model-dependent predictions under the CEV

[5]For n finite, estimators (3.22) and (3.23) are undefined if, respectively, $\widehat{v}_{n,N,S_v}(t)$ or $\widehat{\eta}_{n,N,M,S_\eta}(t)$ is equal to zero. However, the analysis conducted in this chapter is not affected by this potential problem, as estimators (3.22) and (3.23) are used only in equations (3.24) and (3.25), which are rewritten, to reduce numerical instabilities as, respectively, (3.26) and (3.27).

model, as given in equations (3.6) and (3.7), entails performing a linear regression between numerical approximations of the true derivatives, obtained via finite differences, and estimates of the corresponding theoretical derivatives, as given in (3.22) and (3.23). Formally, the test involves estimating the linear models

$$\frac{\widehat{\eta}_{n,N,M,S_\eta}(t+h)-\widehat{\eta}_{n,N,M,S_\eta}(t)}{x(t+h)-x(t)}=\alpha_1\frac{\widehat{\chi}_{n,N,M,L,S_\chi}(t)}{\widehat{\nu}_{n,N,S_\nu}(t)} \tag{3.24}$$

and

$$\frac{\widehat{\eta}_{n,N,M,S_\eta}(t+h)-\widehat{\eta}_{n,N,M,S_\eta}(t)}{\widehat{\nu}_{n,N,S_\nu}(t+h)-\widehat{\nu}_{n,N,S_\nu}(t)}=\alpha_2\frac{\widehat{\chi}_{n,N,M,L,S_\chi}(t)}{\widehat{\eta}_{n,N,M,S_\eta}(t)}. \tag{3.25}$$

If the estimation with empirical data yields statistically-significant estimates of the coefficients α_1 and α_2 that are close to the value of 1, then the predictions of the CEV model could be deemed as an accurate gauge of the true sensitivity of the leverage to changes in the price and the volatility. This in turn would suggest that empirical data support the interpretation of $\chi(t)$ as the process that captures the response of the leverage to changes in the price and the volatility, as implied by the CEV model via equations (3.6) and (3.7).

In order to obtain reliable results from the tests (3.24) and (3.25), it is not only crucial that finite-sample efficient Fourier-based estimates of the paths of the processes $\nu(t)$, $\eta(t)$ and $\chi(t)$ are used, but also that the step h for the differentiation procedure is carefully selected. Accordingly, the aim of the simulation study performed in this section is to provide guidance for the optimal selection of the step h.

For the simulation study, we generate price observations from the CEV model in equation (3.1), setting $\sigma = 0.3$ and $\delta = 0.5$. These parameter values are taken from the simulation study in Sanfelici and Mancino (2020). Further, the initial price value $X(0)$ is selected as $X(0) = 1$. Recall that a value of δ smaller than 1 reproduces the type of leverage effect usually observed on equity markets, that is, it yields a correlation between returns and volatility changes with negative sign. Based on these parameter values, we simulate a total of 100 days of 1-second observations. Each simulated day is 6.5-hour long.

Specifically, we simulate two scenarios: one where the efficient log-price $x(t)$ is observable and another, more realistic, where one can only observe the noisy price $\tilde{x}(t) := x(t) + \varepsilon(t)$, that is, the efficient price $x(t)$ contaminated by the presence of an i.i.d. zero-mean noise component $\varepsilon(t)$. For the simulation of $\varepsilon(t)$, we choose a Gaussian distribution, with standard deviation parameter equal to 10^{-4}.

In both simulated scenarios, we use all available data for the estimation of $\nu(t)$, $\eta(t)$ and $\chi(t)$, that is, we select $n = 23400$, which corresponds to the 1-second sampling frequency. Further, we make the following selections for the cutting frequencies. At the first level, to obtain spot volatility estimates, we select $N = n/2$ in the noise-free scenario, while in the noisy scenario we select N based on the noise-robust procedure proposed by Mancino and Sanfelici (2008); then, we select $S_\nu = n^{0.5}$. At the second level, we estimate the spot leverage by choosing $M = n^{0.5}$ and $S_\eta = 4n^{0.25}$ Finally, we select $L = 4n^{0.25}$ and $S_\chi = 6n^{0.125}$ at the third level, to obtain spot estimates of $\chi(t)$. All these selections, with the exception of N, are based on the numerical minimization of the mean-squared error (MSE). The estimated trajectories of $\nu(t)$, $\eta(t)$ and $\chi(t)$ in the absence and the presence of noise are plotted, along with the corresponding true values, in Figures 3.1 and 3.2 to demonstrate the accuracy of the estimation. Additionally, in Figures 3.1 and 3.2 we also show the accuracy of the estimated trajectory of $\xi(t)$, which will be used in the final part of the simulation study. Note that the selection of the cutting frequencies for the estimation of $\xi(t)$ is performed separately (see Theorem 3.3). In particular, first we select N either equal to $3n^{0.75}$ in the noise-free scenario or via the noise-robust approach by Mancino and Sanfelici (2008) in the noisy scenario, then we choose $M = 2n^{0.25}$ and $S_\xi = n^{0.25}$ based on the numerical optimization of the MSE.

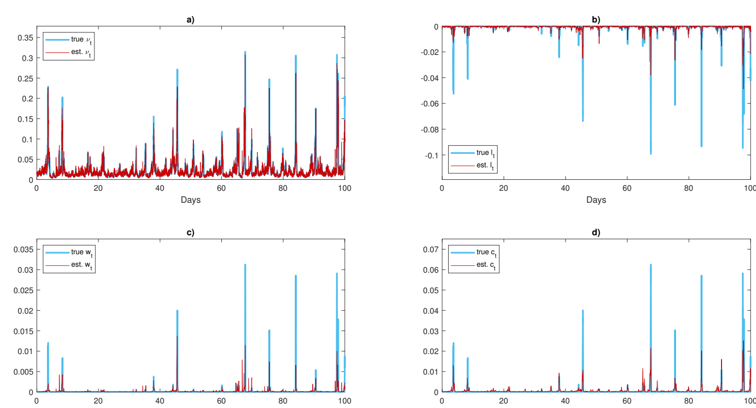

Fig. 3.1: Comparison, in the noise-free scenario, between the true and the estimated trajectories of $\nu(t)$ (panel a)), $\eta(t)$ (panel b)), $\xi(t)$ (panel c)) and $\chi(t)$ (panel d)). For each panel, the true and estimated trajectories are plotted on the equally-spaced grid of mesh size equal to 1 second.

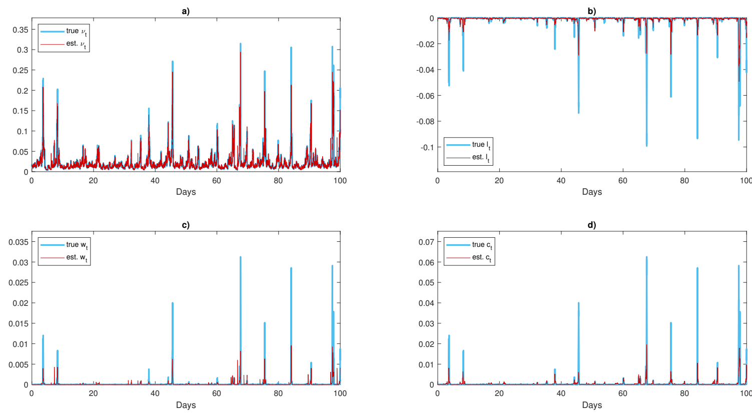

Fig. 3.2: Comparison, in the noisy scenario, between the true and the estimated trajectories of $\nu(t)$ (panel a)), $\eta(t)$ (panel b)), $\xi(t)$ (panel c)) and $\chi(t)$ (panel d)). For each panel, the true and estimated trajectories are plotted on the equally-spaced grid of mesh size equal to 1 second.

After having obtained accurate estimates of $\nu(t)$, $\eta(t)$ and $\chi(t)$, we perform tests (3.24) and (3.25). In this regard, to reduce the numerical instabilities related to the computation of ratios, we estimate α_1 and α_2 after rewriting (3.24) and (3.25) as, respectively,

$$\left[\widehat{\eta}_{n,N,M,S_\eta}(t+h)-\widehat{\eta}_{n,N,M,S_\eta}(t)\right]\widehat{\nu}_{n,N,S_\nu}(t)=\alpha_1\left[x(t+h)-x(t)\right]\widehat{\chi}_{n,N,M,L,S_\chi}(t), \tag{3.26}$$

and

$$\left[\widehat{\eta}_{n,N,M,S_\eta}(t+h)-\widehat{\eta}_{n,N,M,S_\eta}(t)\right]\widehat{l}_{n,N,M,S_\eta}(t)=\alpha_2\left[\widehat{\nu}_{n,N,S_\nu}(t+h)-\widehat{\nu}_{n,N,S_\nu}(t)\right]\widehat{\chi}_{n,N,M,L,S_\chi}(t) \tag{3.27}$$

The estimates of α_1 and α_2 in the noise-free and the noisy scenarios, obtained for different values of the step h, are plotted in Figures 3.3 - 3.6[6]. These figures show that the estimates of α_1 and α_2 fluctuate around the true level, that is, around 1, for values of h between 5 and 30 minutes[7]. This suggests that a more reliable

[6] Note that the estimation of α_1 and α_2 has been performed using the robust regression method with a bisquare weighting scheme to penalize outliers (see Holland and Welsch (1977)). The same holds for the estimation of α_1 and α_2 in the empirical study of the next section.

[7] For h smaller than 5 minutes, estimates of α_1 and α_2 tend to be biased and very noisy and thus are omitted from the plots.

estimate of α_1 or α_2 could be obtained by computing the average of point-wise estimates in correspondence of values of h between 5 and 30 minutes. Indeed, such averages, which are also plotted in Figures 3.3 - 3.6 (see the red dashed lines), appear to be very close to 1. The exact averages of the estimates of α_1 and α_2, along with average values of other relevant outputs of the estimation procedure, are reported in the Table 3.1. Note that these averages are quite accurate, that is, are quite close to 1. Also, note that point-wise coefficient estimates are all statistically significant, with a constant p-value equal to zero for both tests in both scenarios considered. Finally, note that the very large R^2 values confirm the accuracy of the Fourier estimates of $\nu(t)$, $\eta(t)$ and $\chi(t)$.

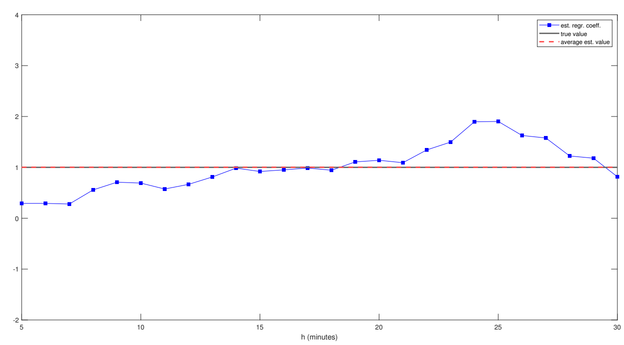

Fig. 3.3: Estimation of model (3.24) in the noise-free scenario: comparison of point-wise estimates of the coefficient α_1 in correspondence of different values of the step h (in blue) and their average (red dashed line) with the true value (grey line).

	coeff. est.	std. err.	t stat.	p-value	R^2
model (3.24), w/o noise	1.001 (0.550)	0.004 (0.001)	240.538 (110.971)	0 (0)	0.906 (0.138)
model (3.24), w/ noise	1.039 (0.654)	0.005 (0.002)	188.964 (81.688)	0 (0)	0.877 (0.190)
model (3.25), w/o noise	1.016 (0.587)	0.002 ($< 10^{-3}$)	590.663 (290.443)	0 (0)	0.955 (0.120)
model (3.25), w/ noise	1.064 (0.395)	0.002 ($< 10^{-3}$)	636.648 (223.603)	0 (0)	0.992 (0.005)

Tab. 3.1: Estimation results for models (3.24) and (3.25): average values of coefficient estimates, standard errors, t statistics, p-values and R^2, computed for h ranging between 5 and 30 minutes. Standard deviations are also reported in brackets. For each model, the lines “w/o noise” and “w/ noise” refer to, respectively, the simulated scenario without and with noise.

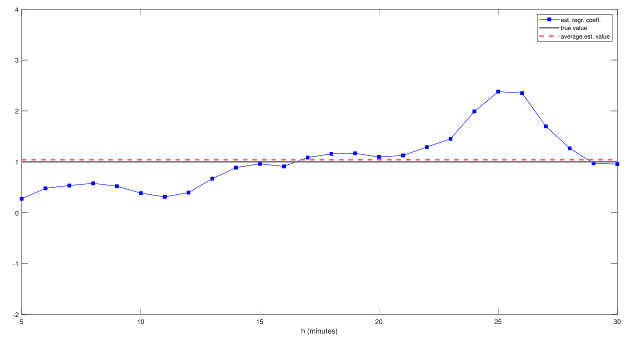

Fig. 3.4: Estimation of model (3.24) in the noisy scenario: comparison of point-wise estimates of the coefficient α_1 in correspondence of different values of the step h (in blue) and their average (red dashed line) with the true value (grey line).

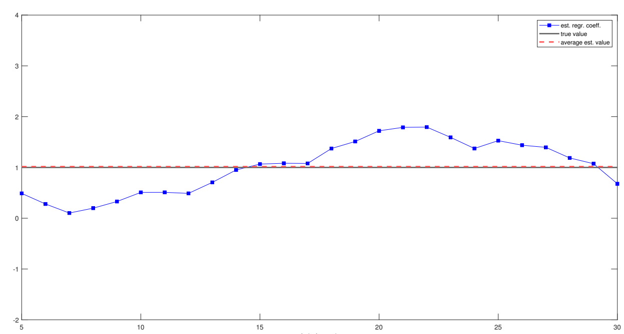

Fig. 3.5: Estimation of model (3.25) in the noise-free scenario: comparison of point-wise estimates of the coefficient α_2 in correspondence of different values of the step h (in blue) and their average (red dashed line) with the true value (grey line).

As mentioned in the previous section, based on equations (3.4) and (3.11), another aspect that could be investigated empirically is the existence of a linear link between $\chi(t)$ and the vol-of-vol $\xi(t)$. Specifically, the existence of such a link could be investigated by performing the linear regression

$$\widehat{\chi}_{n,N,M,L,S_\chi}(t) = \alpha_3 \widehat{\xi}_{n,N,M,S_\xi}(t). \tag{3.28}$$

A statistically-significant estimate of the coefficient α_3 would offer evidence,

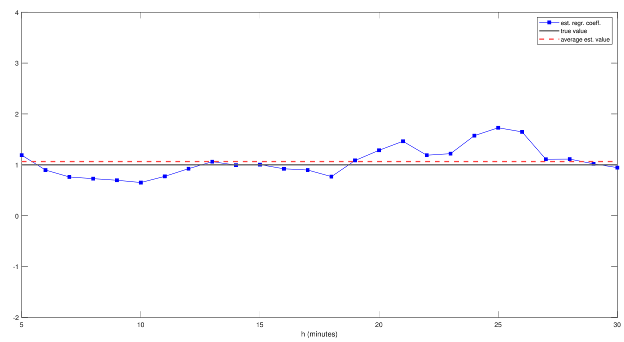

Fig. 3.6: Estimation of model (3.25) in the noisy scenario: comparison of point-wise estimates of the coefficient α_2 in correspondence of different values of the step h (in blue) and their average (red dashed line) with the true value (grey line).

in a model-free setting, that $\chi(t)$ is actually linear in $\xi(t)$, as predicted by the CEV model (and, more generally, by the large class of models represented by (3.8)). Interpreting $\chi(t)$ as the process that captures the response of the leverage to changes in the volatility or the price, this would mean that the latter is proportional to the uncertainty perceived by market operators about the actual riskiness of the asset of interest (i.e., the vol-of-vol $\xi(t)$).

As for (3.24) and (3.25), the accuracy of the regression (3.28) can also be tested on simulated observations from the CEV model, to obtain guidance for the selection of the optimal frequency for the sampling of Fourier estimates of $\chi(t)$ and $\xi(t)$. Estimates of α_3 in correspondence of different sampling frequencies between 5 and 30 minutes are shown in Figures 3.7 and 3.8. Note that point-wise estimates of α_3 appear to be very reliable, in that they are all very close to the true value of 2, at least for sampling frequencies smaller than 15 minutes. However, a conservative approach might suggest to adopt the average as a final estimate of α_3 also in this case (see the red dashed line in Figures 3.7 and 3.8). Average statistics of the regression are reported in Table 3.2[8]. Again, we obtain quite satisfactory R^2 values, which confirm the accuracy of the estimates of $\chi(t)$ and $\xi(t)$; also, all estimates of α_3 are significant, with constant p-values equal to zero.

Finally, note that we obtain comparable results in the noise-free and noisy scenarios for all three tests performed in this section, thereby confirming the robust-

[8]To account for auto-correlations in the residuals, we compute Newey-West standard errors, see Newey and West (1987). We do the same also when estimating α_3 in the empirical exercise of the next section.

ness of the Fourier methodology to the presence of noise.

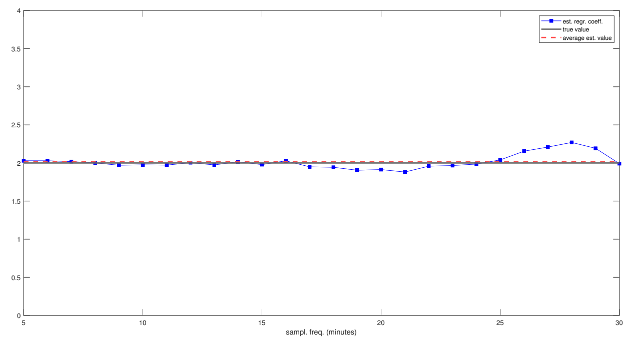

Fig. 3.7: Estimation of model (3.28) in the noise-free scenario: comparison of point-wise estimates of the coefficient α_3 in correspondence of different sampling frequencies (in blue) and their average (red dashed line) with the true value (grey line).

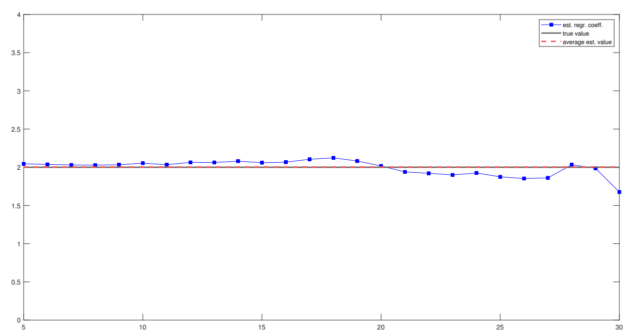

Fig. 3.8: Estimation of model (3.28) in the noisy scenario: comparison of point-wise estimates of the coefficient α_3 in correspondence of different sampling frequencies (in blue) and their average (red dashed line) with the true value (grey line).

3.5 Empirical study

In this section we perform tests (3.24), (3.25) and (3.28) on the series of 1-second S&P500 price observations over the period March, 2018 − April, 2018 (see Figure 3.9).

	coeff. est.	std. err.	t stat.	p-value	R^2
model (3.28), w/o noise	2.003 (0.126)	0.326 (0.045)	6.280 (1.177)	0 (0)	0.543 (0.032)
model (3.28), w/ noise	2.020 (0.157)	0.233 (0.041)	8.918 (1.671)	0 (0)	0.538 (0.038)

Tab. 3.2: Estimation results for models (3.24) and (3.25): average values of coefficient estimates, standard errors, t statistics, p-values and R^2, computed for sampling frequencies ranging between 5 and 30 minutes. Standard deviations are also reported in brackets. The lines "w/o noise" and "w/ noise" refer to, respectively, the simulated scenario without and with noise.

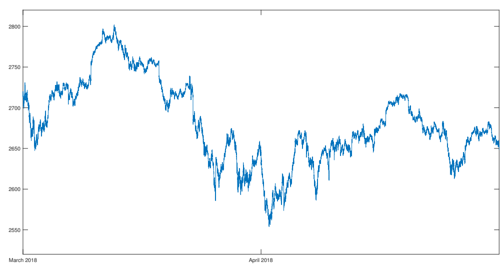

Fig. 3.9: S&P 500 1-second prices over the period March, 2018 − April, 2018.

To obtain Fourier estimates of the paths of $\nu(t)$, $\eta(t)$, $\chi(t)$ and $\xi(t)$, we use all data in the sample, that is, we select $n = 23400$, corresponding to the 1-second sampling frequency. Further, we select the cutting frequencies using as guidance the MSE-optimal values obtained via simulations in the noisy scenario of Section 3.4. Specifically, after choosing N via the noise-robust procedure given in Mancino and Sanfelici (2008), we select $S_\nu = M = n^{0.5}$, $S_\eta = L = 4n^{0.25}$ and $S_\chi = 6n^{0.125}$. For the estimation of $\xi(t)$, instead, we select $M = 2n^{0.25}$ and $S_\xi = n^{0.25}$. The estimated trajectories of $\nu(t)$, $\eta(t)$, $\xi(t)$ and $\chi(t)$ are plotted in Figure 3.10. Note that, before performing the estimation, we have removed days with price jumps from the 2-month sample, using the jump detection test by Corsi et al. (2010). In particular, the test at 99.9% confidence level detects only two days with jumps, namely March 20th and March 23rd. These two days are associated with market turbulence related to the so-called "trade war" between China and the US.

Using the reconstructed paths of $\nu(t), \eta(t), \chi(t)$ and $\xi(t)$ we then perform tests (3.24), (3.25) and (3.28)[9]. The results are summarized in Table 3.3 and Figures

[9]To avoid performing a spurious regression (see Granger and Newbold (1974)), we test for the

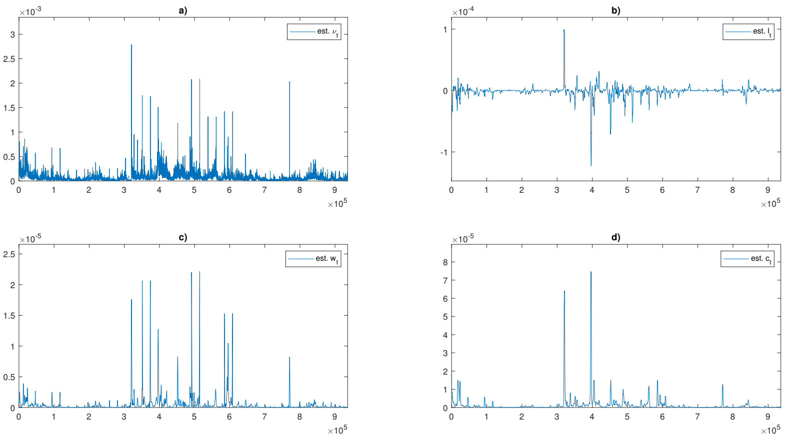

Fig. 3.10: Reconstructed 1-second trajectories of $\nu(t)$, $\eta(t)$, $\xi(t)$ and $\chi(t)$ for the S&P 500 index over the period March, 2018-April, 2018.

3.11 - 3.13.

Overall, the results of tests (3.24) and (3.25) support the interpretation of the process $\chi(t)$ as a process that captures the instantaneous response of the leverage to changes in the price and the level of market risk (i.e., the volatility). In fact, not only we obtain statistically-significant estimates of α_1 and α_2 for all values of h considered, but also these estimates fluctuate around average estimates which are close to 1, taking values equal, respectively, to 1.018 and 0.914

Additionally, the results of test (3.28) support the existence of a statistically-significant positive linear dependence between $\chi(t)$ and $\xi(t)$. This empirical result, if considered jointly with the results of tests (3.24) and (3.25), suggests that the sensitivity of the leverage to changes of the price or the volatility is larger when the uncertainty about the actual level of risk perceived on the market (i.e., the vol-of-vol) is larger. Additionally, note that point-wise estimates of α_3 are close to 2, the value predicted by the CEV model, with a final average estimate equal to 1.895.

Finally, note that we obtain R^2 values which are not far from the values obtained in simulations, for all three tests. This suggests that the tested models fit the sample data quite well.

null hypothesis of the presence of a unit root in the all the series of regressors and regressands involved, using the Augmented Dickey-Fuller test (see Dickey and Fuller (1979)). For all series, test results at the 99.9% confidence level reject the null hypothesis.

	coeff. est.	std. err.	t stat.	p-value	R^2
model (3.24)	1.018 (1.119)	0.006 (0.004)	163.605 (159.688)	0 (0)	0.821 (0.225)
model (3.25)	0.914 (0.660)	0.009 (0.002)	92.858 (73.879)	0 (0)	0.847 (0.159)
model (3.28)	1.895 (0.206)	0.625 (0.049)	3.037 (0.323)	0.004 (0.003)	0.409 (0.038)

Tab. 3.3: Estimation results for models (3.24), (3.25) and (3.28): average values of coefficient estimates, standard errors, t statistics, p-values and R^2, computed for values of h (models (3.24 and (3.25)) or sampling frequencies (model (3.28)) ranging between 5 and 30 minutes. Standard deviations are also reported in brackets.

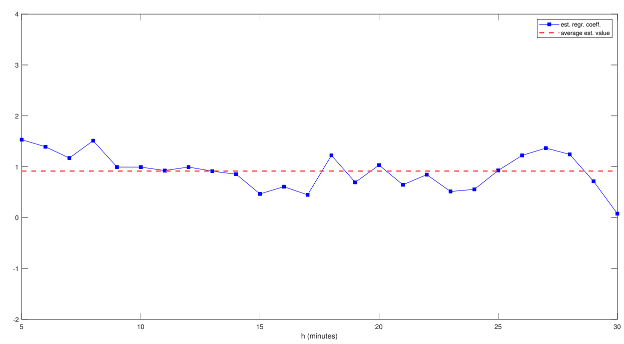

Fig. 3.11: Estimation of model (3.24): point-wise estimates of the coefficient α_1 in correspondence of different sampling frequencies (in blue), along with their average (red dashed line).

3.6 Conclusions

The main finding of this chapter is uncovering, both from an analytical and an empirical perspective, the relationship between the price-leverage covariation and the sensitivity of the leverage process to changes in the price or the volatility.

Indeed, first we show that under the CEV model, which is explicitly designed to capture the leverage effect, the derivatives of the leverage process with respect to the price and the volatility are equal to the price-leverage covariation scaled, respectively, by the volatility and the leverage itself. In this regard, we stress that a key analytical result we obtain is expressing the derivatives of a stochastic process (the leverage) as a function of objects that can be consistently estimated from sample prices over a fixed time horizon, that is, iterated covariances.

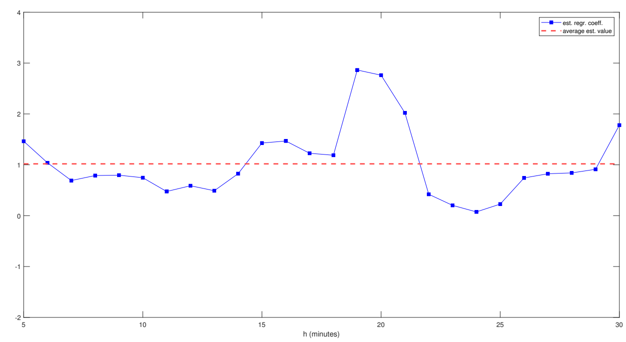

Fig. 3.12: Estimation of model (3.25): point-wise estimates of the coefficient α_2 in correspondence of different sampling frequencies (in blue), along with their average (red dashed line).

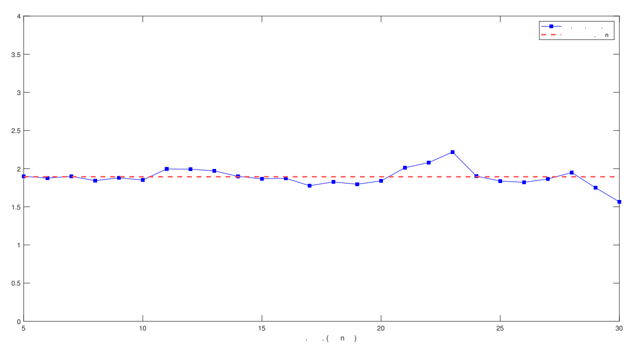

Fig. 3.13: Estimation of model (3.28): point-wise estimates of the coefficient α_3 in correspondence of different sampling frequencies (in blue), along with their average (red dashed line).

Then, after reconstructing the paths of the volatility, the leverage and the price-leverage covariation by means of the (non-parametric) Fourier methodology, we show, empirically, that these model-dependent predictions reproduce the model-free derivatives of the leverage quite accurately in the case of the S&P500 index over the period March, 2018-April, 2018.

Based on this empirical evidence, the price-leverage covariation could be understood by market operators as a gauge of the responsiveness of the leverage effect to the arrival of new information causing a change in the price level and/or the amount of market risk, that is, in the volatility.

Additionally, based on the existence of a linear link between the price-leverage covariation and the vol-of-vol under the CEV model, we also investigate the empirical dependence between these two quantities in model-free setting, that is, using non-parametric Fourier estimates of their paths. In this regard, empirical results support the existence of a statistically significant linear link, with a coefficient close to 2, the value predicted by the CEV model. This in turn suggests that the response of the leverage is stronger when the uncertainty about the actual level of risk perceived, i.e., the vol-of-vol, is larger (and viceversa).

Future research may verify the validity of the empirical findings presented in this chapter for different asset classes and/or more turbulent economic scenarios. Further, the findings of this chapter motivate the investigation of the asymptotic error distribution of the non-parametric Fourier estimator of the price-leverage covariation in future research. Additionally, future research may investigate the role of the price-leverage covariation in forecasting the leverage effect.

4. BIAS-OPTIMAL INTEGRATED VOL-OF-VOL ESTIMATION: THE ROLE OF WINDOW OVERLAPPING

4.1 Introduction

Estimating the volatility of asset volatility (hereinafter vol-of-vol) is relevant in many areas of mathematical finance, such as the calibration of stochastic volatility of volatility models (Barndorff-Nielsen and Veraart (2009), Sanfelici et al. (2015)), the hedging of portfolios against volatility of volatility risk (Huang et al. (2018)), the estimation of the leverage effect (Kalnina and Xiu (2017), Aït-Sahalia et al. (2017)), and the inference of future returns (Bollerslev et al. (2009)), along with spot volatilities (Mykland and Zhang (2009)).

The literature offers a number of consistent estimators for the integrated vol-of-vol. The first estimator to appear was the one proposed by Barndorff-Nielsen and Veraart (2009), termed Pre-estimated Spot-variance based Realized Variance (PSRV), which is, in fact, simply the realized variance of the unobservable spot variance, computed using estimates of the latter. Later, Vetter (2015) derived two sophisticated versions of the simple PSRV: one that allows for a central limit theorem with the optimal rate of convergence, but also for negative values, and another that preserves positivity at the expense of a slower rate of convergence. Note that the simple PSRV and its sophisticated versions are consistent when the price and volatility processes are continuous semimartingales, in the absence of microstructure noise contaminations. Further, Fourier-based estimators of the integrated vol-of-vol were introduced by Sanfelici et al. (2015) and Cuchiero and Teichmann (2015). In particular, the estimator by Sanfelici et al. (2015) is asymptotically unbiased in the presence of market microstructure noise, while the estimator by Cuchiero and Teichmann (2015) allows for a central limit theorem in the presence of jumps in the price and volatility processes.

The numerical studies in Aït-Sahalia et al. (2017) and Sanfelici et al. (2015) show that both realized and Fourier-based integrated vol-of-vol estimators may carry a substantial finite-sample bias unless the selection of the tuning parameters involved in their computation is carefully optimized. However, this is a rather unexplored issue, which we aim to explore. To do so, we focus on the simple PSRV, since it represents the most intuitive and easy-to-implement vol-of-vol estimator.

Furthermore, asymptotically-optimal estimators do not necessarily guarantee the best finite-sample performance, as pointed out in the extensive study by Gatheral and Oomen (2010) on integrated volatility estimators and confirmed for integrated vol-of-vol estimators by the numerical studies in Aït-Sahalia et al. (2017) and Sanfelici et al. (2015). Thus, there is no reason to expect a priori that the simple PSRV would show worse finite-sample performance than its sophisticated version with optimal rate of convergence.

As mentioned, the PSRV is the realized volatility of the unobservable spot volatility process, computed from discrete estimates of the latter. In other words, the PSRV is the sum of the squared increments of estimates of the unobservable spot volatility on a discrete grid. These estimates are obtained as local averages of the price realized volatility. Formally, the locally averaged realized variance and the PSRV are defined as follows.

Definition 4.1. ***Locally averaged realized variance***

Suppose that the log-price process x is observable on an equally-spaced grid of mesh size δ_N, with $\delta_N \to 0$ as $N \to \infty$. Also, let $k_N = O(\delta_N^b)$, $b \in (-1,0)$, be a sequence of positive integers such that $k_N \to \infty$ and define the local window $W_N := k_N \delta_N$ such that $W_N \to 0$ as $N \to \infty$. The locally averaged realized variance at time t is defined as

$$\hat{v}_N(t) := \frac{1}{k_N \delta_N} \sum_{j=1}^{k_N} \Big[x(\lfloor t/\delta_N \rfloor \delta_N - k_N \delta_N + j\delta_N) - x(\lfloor t/\delta_N \rfloor \delta_N - k_N \delta_N + (j-1)\delta_N) \Big]^2,$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

Definition 4.2. ***Pre-estimated Spot-variance based Realized Variance***

Suppose that the log-price process x is observable on an equally-spaced grid of mesh size δ_N, with $\delta_N \to 0$ as $N \to \infty$. The pre-estimated spot-variance based realized variance (PSRV) on the interval $[\tau, \tau + h]$ is defined as

$$PSRV_{[\tau,\tau+h],N} := \sum_{i=1}^{\lfloor h/\Delta_N \rfloor} \Big[\hat{v}_N(\tau + i\Delta_N) - \hat{v}_N(\tau + (i-1)\Delta_N) \Big]^2,$$

where:

- *$\hat{v}_N(\cdot)$ is the locally averaged realized variance in Definition 4.1, with $k_N = O(\delta_N^b)$, $b \in (-1,0)$;*

- *$\Delta_N = O(\delta_N^c)$, $c \in (0,1)$, is the locally averaged realized variance sampling frequency.*

The following propositions summarize the asymptotic properties of the locally averaged realized variance and the PSRV. For further details, see Chapter 8 in Aït-Sahalia and Jacod (2014).

Proposition 4.1. *Let the log-price process x be a continuous semimartingale and let the process ν denote its instantaneous volatility. Then $\hat{\nu}_N(t)$ is a consistent local estimator of $\nu(t)$ as $N \to \infty$.*

Proof. See Chapter 8.1 in Aït-Sahalia and Jacod (2014). □

Proposition 4.2. *Let the log-price process x and the spot volatility process ν be continuous semimartingales. Then the PSRV is a consistent estimator of the quadratic variation of the volatility process $\langle \nu, \nu \rangle_{[\tau, \tau+h]}$ if $b \in (-1/2, 0)$ and $c \in (0, -b/2)$.*

Proof. See Proposition 8 in Barndorff-Nielsen and Veraart (2009). □

Remark 4.1. *Note that the requirements for rates b and c that guarantee consistency imply that $\frac{W_N}{\Delta_N} \to 0$ as $N \to \infty$. Indeed, as one can easily verify, $-1/2 < b < 0$ and $0 < c < -b/2$ imply $c < 1 + b$, which, in turn, implies $\frac{W_N}{\Delta_N} \to 0$ as $N \to \infty$.*

In practical applications, when computing PSRV values, one has to select the spot volatility estimation grid. Moreover, since the spot volatility is estimated as an average of the price realized volatility over a local window, the length of the latter must also be selected. More specifically, the figure below details the different quantities involved in the computation of the PSRV: the time horizon h; the log-price sampling frequency $\delta_N := \frac{h}{N}$; the spot volatility sampling frequency $\Delta_N := \lambda_N \delta_N$, $\lambda_N = min(N, \lceil \lambda \delta_N^{c-1} \rceil)$, $\lambda > 0$, $c \in (0,1)$; the size of the local window to estimate the spot volatility $W_N = k_N \delta_N$, $k_N = \lceil \kappa \delta_N^b \rceil$, $\kappa > 0$, $b \in (-1,0)$; and the spot volatility estimates $\hat{\nu}(s)$, $s = \tau + j\Delta_N$, $j = 0, 1, ..., \lfloor h/\Delta_N \rfloor$. Note that $\lceil \cdot \rceil$ denotes the ceiling function.

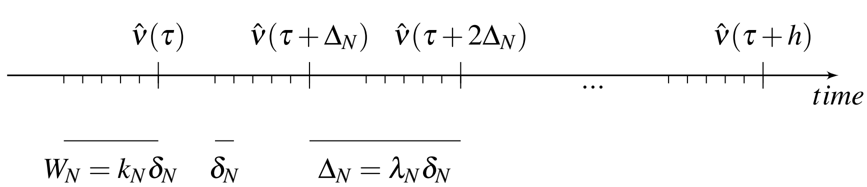

Fig. 4.1: Graphic representation of the quantities involved in the computation of the PSRV on the interval $[\tau, \tau+h]$.

As a consequence, for given values of the asymptotic rates b and c, the finite-sample performance of the PSRV (i.e., the performance of the PSRV for a fixed N) depends on the selection of two tuning parameters: λ, which determines the mesh of the spot volatility estimation grid and κ, which determines the length of the local window used to estimate the spot volatility. With regard to the selection of κ, note that the efficient computation of the spot volatility in finite samples may require the selection of fairly long local windows (see, e.g., Lee and Mykland (2008), Aït-Sahalia et al. (2013) and Zu and Boswijk (2014)). This in turn suggests that the finite-sample efficient implementation of the PSRV over a given period (e.g., one day) may require the use of price observations from the previous period(s) (e.g., day(s)). At the same time, this might imply that it is optimal to allow consecutive local windows to overlap in finite samples, that is, $W_N > \Delta_N$ for N finite. This aspect is confirmed by the numerical study in Sanfelici et al. (2015), which shows that it is optimal to select κ such that $W_N > \Delta_N$ in finite samples. The aim of this chapter is to gain insight into the bias-reducing effect due to window-overlapping from an analytical perspective. To do so, we follow an approach inspired by the one used in Aït-Sahalia et al. (2013) to solve the "leverage effect puzzle".

The "leverage effect puzzle" pertains to the absence of correlation between log-price and (estimated) volatility changes at high-frequencies, observed in empirical samples. Aït-Sahalia et al. (2013) solve this puzzle by showing analytically that a substantial bias masks the presence of correlation unless log-price and volatility estimates changes are computed on a suitably sparse grid. The aim is not to solve the problem of the efficient non-parametric estimation of the leverage at high-frequencies, but rather to obtain insight into the puzzle by solving it in a widely used parametric setting that allows for fully explicit computations. This chapter is written in the same spirit. In fact, we do not address the general problem of the efficient non-parametric vol-of-vol estimation from high-frequency prices, but, rather, our aim is to obtain insight from an analytical perspective into why the PSRV, the simplest and most natural vol-of-vol estimator, is plagued by a large bias in finite samples and investigate the role of window-overlapping as a tool for reducing such large bias.

To achieve this aim, we proceed as follows. In Section 4.2 we perform a preliminary numerical exercise that uncovers the crucial role of the local-window parameter κ in determining the finite-sample performance of the PSRV and, at the same time, shows that the latter is basically insensitive to the selection of the grid parameter λ. In particular, it is evident from simulations that the PSRV finite-sample bias is optimized by selecting κ such that consecutive local windows to estimate the spot volatility overlap. Numerical results of Section 4.2 confirm those by Sanfelici et al. (2015) and motivate the analytical study of Section 4.3.

In Section 4.3, we address the problem of the optimal selection of PSRV tuning parameters in finite samples from an analytical perspective. To do so, in the spirit of Aït-Sahalia et al. (2013), we assume a widely-used parametric form for the data-generating process, which allows us to obtain the full explicit PSRV finite-sample bias expression. Specifically, we assume the price to be a continuous semimartingale and the volatility to be a CIR process (see Cox et al. (1985)). In general, independently of the parametric assumption on the data-generating process, the PSRV finite-sample bias expression differs in case window overlapping is allowed, i.e., when $W_N > \Delta_N$, or not, i.e., $W_N \leq \Delta_N$. Consequently, in Section 4.3 we study both cases.

In the no-overlapping case we adopt a conventional approach and isolate the dominant term of the bias as $N \to \infty$, thereby showing that a value of κ that annihilates the dominant term of the bias does not always exist and, even when it does exist, its computation would be basically unfeasible in practice, as it depends on the drift parameters of the volatility, which cannot be reliably estimated on a fixed time horizon, due to the fact that their consistent estimation is possible only in the classic long-sample asymptotics setting (see, e.g., Kanaya and Kristensen (2016)). In addition, when the optimal value of κ exists, for typical orders of magnitude of the CIR parameters it actually satisfies the no-overlapping constraint only at ultra-high frequencies ($<$ 1 second), at which prices are typically not available.

In the overlapping case, instead, the natural expansion as $N \to \infty$ is precluded, as the consistency of the PSRV requires that consecutive windows do not overlap as the number of price observations grows to infinity (see Remark 4.1). Therefore, in this case we adopt a novel approach and expand sequentially the bias expression as the the tuning parameter λ and the time horizon h go to zero, based on the fact that, in practical applications, h and λ are typically very close to zero. This approach yields a dominant term of the bias which is independent of the tuning parameter λ and is annihilated by selecting the asymptotic rate of k_N as $b = -1/2$, the asymptotic rate of Δ_N as $c < 1/2$, and the local-window tuning parameter as $\kappa = 2\sqrt{\nu(\tau)}\gamma^{-1}$, where $\nu(t)$ and γ denote, respectively, the spot variance process at the initial time τ and the CIR diffusion parameter. This analytical result shows that, when overlapping is allowed, it is possible to select κ such that the bias is effectively optimized in practical applications and supports the numerical evidence on the bias-reducing effect due to window overlapping collected in Section 4.2 and in Sanfelici et al. (2015). However, this rule to select κ is unfeasible unless reliable estimates of $\nu(\tau)$ and γ are available. Accordingly, in the Appendix we detail a simple procedure to estimate $\nu(\tau)$ and γ.

In Section 4.3 we also address the problem of the bias-optimal implementation of the PSRV in the more realistic situation where the price process is contaminated by an i.i.d. microstructure noise process at high frequencies. Again, we distinguish between the overlapping and no-overlapping case and derive, for each case, the exact parametric expression of the extra bias term due to microstructure noise. However, in both cases it emerges that this extra term depends not only on some moments of the noise process but also on the drift parameters of the volatility process, which cannot be consistently estimated over a fixed time horizon (see, e.g., Kanaya and Kristensen (2016)). This precludes the possibility of efficiently subtracting the bias due to noise in small samples. As a solution, we suggest sampling prices on a suitably sparse grid as in the seminal paper by Andersen et al. (2001a), so that the presence of noise becomes negligible and the bias optimal rule to select the local-window parameter κ can still be applied. The efficiency of this solution is verified numerically in Section 4.5 for typical values of the noise-to-signal ratio. For completeness, we also analyze the noise bias expression in the no-overlapping case. In particular, we exploit this expression to derive the asymptotic rate of divergence of the PSRV bias as N tends to infinity.

Additionally, as a byproduct of the PSRV bias analysis, in Section 4.3 we quantify the bias reduction following the assumption that the initial value of the volatility process is equal to the long-term volatility parameter, in the case of both the PSRV and the locally averaged realized variance. This is a very common assumption in the literature, typically made in simulation studies where a mean-reverting process drives the spot volatility (see, e.g., among many others, Aït-Sahalia et al. (2013), Sanfelici et al. (2015), Vetter (2015)).

In Section 4.4 we use a heuristic approach based on dimensional analysis to generalize the rule for the selection of κ to the case of a volatility process belonging to the CKLS class (see Chan et al. (1992)). Specifically, we find that it is optimal, in terms of bias reduction, to select $\kappa = 2\frac{v(\tau)}{\sqrt{\xi(\tau)}}$, where $v(t)$ is the variance process and $\xi(t)$ is the vol-of-vol process, while τ is the initial time of the estimation horizon. Note that in the absence of price and volatility jumps (a condition required for the PSRV to be consistent), the semi-parametric stochastic volatility model where the price is a semimartingale and the volatility is a CKLS process represents a fairly flexible model. In fact, the CKLS framework encompasses a number of widely-used models for financial applications. Indeed, besides the CIR model, which determines the volatility dynamics in the popular Heston model (Heston (1993)) and its generalized version with stochastic leverage by Veraart and Veraart (2012), the CKLS family includes, e.g., the model by Brennan and Schwartz (1980) and the model by Cox et al. (1980), which appear, respectively, in the continuous-time GARCH stochastic volatility model by Nelson (1990) and 3/2 stochastic volatility model by Platen (1997).

In Section 4.5 we perform an extensive numerical study where we test the performance of the feasible rule to select κ derived in Section 4.3 and generalized in Section 4.4. The results confirm that this rule is effective in reducing the PSRV bias. We underline that this rule does not require the estimation of the drift parameters of the CIR process, which can not be consistently estimated on a fixed time horizon. Finally, in Section 4.6 we illustrate the results of an empirical study, in which we compute PSRV values from high-frequency S&P 500 prices, selecting κ based on the bias-optimal rule. Section 4.7 summarizes our conclusions. Finally, the Appendix (Section 4.8) contains the proofs and the illustration of the feasible procedure that we propose to select κ from sample prices.

4.2 Motivation

The finite-sample accuracy of the PSRV requires the careful selection of the tuning parameters κ and λ. In this section we gain some preliminary insight into this issue by performing a numerical study, whose result motivate the analytical investigation of Section 4.3. In particular, we simulate observations from the following data-generating process, where the volatility is a CIR process. Note that this data-generating process is also used for the analytical study in Section 4.3.

Assumption 4.1. ***Data-generating process***

For $t \in [0,T]$, $T > 0$, the dynamics of the log-price process x and the spot volatility process ν read

$$dx(t) = \sqrt{\nu(t)}dW(t)$$
$$d\nu(t) = \gamma\sqrt{\nu(t)}dZ(t) + \theta\Big(\alpha - \nu(t)\Big)dt$$

where: W and Z are two Brownian motions, with correlation parameter ρ, on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$, which satisfies the usual conditions; θ, α and γ are strictly positive constants (which denote, respectively, the speed of mean reversion, long-term mean and vol-of-vol parameters) such that $2\alpha\theta > \gamma^2$ to ensure that $\nu(t)$ is a.s. positive $\forall t \in [0,T]$.

In particular, we simulate one thousand 1-year trajectories of 1-second observations, with a year composed of 252 trading days of 6 hours each. We consider three scenarios determined by the following sets of model parameters: *Set 1*: $(\alpha,\theta,\gamma,\rho,\nu(0)) = (0.2,5,0.5,-0.2,0.2)$; *Set 2*: $(\alpha,\theta,\gamma,\rho,\nu(0)) = (0.02,10,0.25,-0.8,0.03)$; *Set 3*: $(\alpha,\theta,\gamma,\rho,\nu(0)) = (0.2,5,0.5,-0.2,0.4)$. For all three sets, we select $x(0) = ln(100)$.

The first set of parameters, *Set 1*, is used in Sanfelici et al. (2015) and Vetter (2015) and represents our baseline scenario. The second, *Set 2*, represents the opposite scenario. In fact, the volatility generated by *Set 2* is lower than the volatility generated by *Set 1*, since the long term mean, α, and the speed of mean reversion,

θ, are, respectively, much lower and much higher than in *Set 1*. The second scenario is also characterized by a lower volatility of the volatility, which is captured by the parameter γ and a more pronounced leverage effect, which is captured by the correlation parameter ρ. The third set of parameters, *Set 3*, differs from the first only in that the initial value of the volatility, $\nu(0)$, is twice the long term volatility, α. In this regard, note that if the initial volatility $\nu(0)$ is equal to α, the spot volatility has a constant unconditional mean over time under Assumption 4.1 (see Appendix A in Bollerslev and Zhou (2002)). Setting $\nu(0) = \alpha$ is a simplifying assumption typically adopted in numerical studies where a mean-reverting volatility process is used (see, e.g., among many others, Aït-Sahalia et al. (2013), Sanfelici et al. (2015), Vetter (2015)).

We estimate daily values of the PSRV in these three scenarios from simulated prices sampled with a 1-minute frequency. For the estimation, we set $b = -1/2$ and $c = 1/4$ [1] and study the sensitivity of the bias to different values of κ and λ. With respect to λ, we consider values in the set $(0.0002, 0.0004, 0.0006, 0.0010, 0.0019, 0.0029, 0.0057)$, which correspond to Δ_N equal to $1, 2, 3, 5, 10, 15, 30$ minutes, respectively, thereby preserving the high-frequency nature of the estimator. As for κ, we consider values in the set $(0.017, 0.033, 0.05, 0.1, 0.2, 0.4, 0.5, 1, 1.5, 2, 2.5, 3)$, which correspond to W_N equal to (approximately) $5, 10, 15, 30, 60, 120, 150, 300, 450, 600, 750, 900$ minutes, respectively. Overall, these sets of values for λ and κ allow us to consider both cases when window overlapping occurs, that is, when $W_N > \Delta_N$, and cases when it does not occur, that is, when $W_N \leq \Delta_N$. Figure 4.2 summarizes the results of the numerical exercise for values of κ that lead to a relative bias smaller than an absolute value of 1. This happens for $\kappa = 1.5, 2, 2.5, 3$. Instead, for κ smaller than 1.5, the relative bias rapidly explodes for all values of λ considered, as shown in Figure 4.3.

As one can easily verify, the values of κ in Figure 4.2 imply that local windows for estimating the spot volatility overlap, for all values of λ considered. Consequently, Figure 4.2 tells us that window overlapping is crucial in order to optimize the relative bias of the PSRV even when $\Delta_N >> \delta_N$. This confirms the numerical results in Sanfelici et al. (2015). Furthermore, one can also easily check that the combinations of λ and κ such that overlapping does not occur are all included in Figure 4.3, where the relative bias is larger than 1 and rapidly increases as κ becomes smaller, for any λ considered, reaching the order magnitude 10^3 when W_N equals 5 minutes.

Moreover, focusing on Figure 4.2, it is worth noting that the bias-optimal se-

[1] Note that this choice of b and c satisfies the constraints for asymptotic unbiasedness (see Theorem 4.2). Moreover, note that the selection $b = -1/2$ is also performed in the numerical exercises of Aït-Sahalia et al. (2017) and Sanfelici et al. (2015).

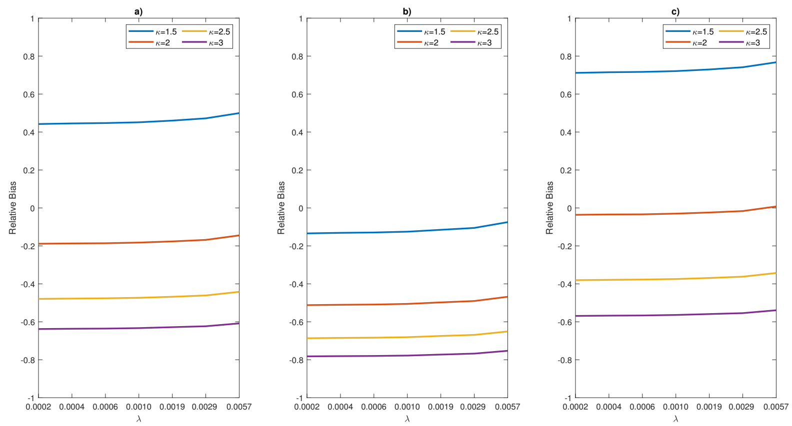

Fig. 4.2: Daily PSRV finite-sample relative bias as a function of λ for values of $\kappa \in (1.5, 2, 2.5, 3)$ and $\delta_N = 1$ minute, $b = -1/2$, $c = 1/4$. The values of λ on the x-axis correspond to Δ_N equal to $j\delta_N$, for $j = 1, 2, 3, 5, 10, 15, 30$. The panels refer to the following parameter sets: a) *Set 1*: $(\alpha, \theta, \gamma, \rho, \nu(0)) = (0.2, 5, 0.5, -0.2, 0.2)$; b) *Set 2*: $(\alpha, \theta, \gamma, \rho, \nu(0)) = (0.03, 10, 0.25, -0.8, 0.03)$; and c) *Set 3*: $(\alpha, \theta, \gamma, \rho, \nu(0)) = (0.2, 5, 0.5, -0.2, 0.4)$.

lection of κ is strongly dependent on the parameters of the data-generating process. In fact, the same value of κ may lead to very different values of the bias in the three scenarios considered: for instance, the selection $\kappa = 2$ leads to a relative bias of approximately -20% in scenario 1, -50% in scenario 2 and -3% in scenario 3. At the same time, Figure 4.2 also suggests that the bias is not very sensitive to the selection of λ. Eventually, for all values of λ considered, the bias-optimal value of κ is between 1.5 and 2 in the baseline scenario, slightly smaller than 1.5 in the second scenario, and around 2 in the third scenario. The indication for scenario 1 is in line with the numerical findings by Sanfelici et al. (2015), where, based on the same parameter set, the optimal value of κ is found to be approximately equal to 2.

In sum, our preliminary numerical study shows not only that allowing for window overlapping is critical to avoid obtaining highly biased vol-of-vol estimates, but also that the selection of κ is crucial for optimizing the PSRV finite-sample bias and, in particular, it is critical to uncover the dependence between the bias-optimal value of κ and the parameters of the data-generating process. Gaining a more in-depth understanding of these numerical findings is what motivates our analytical study in the next section.

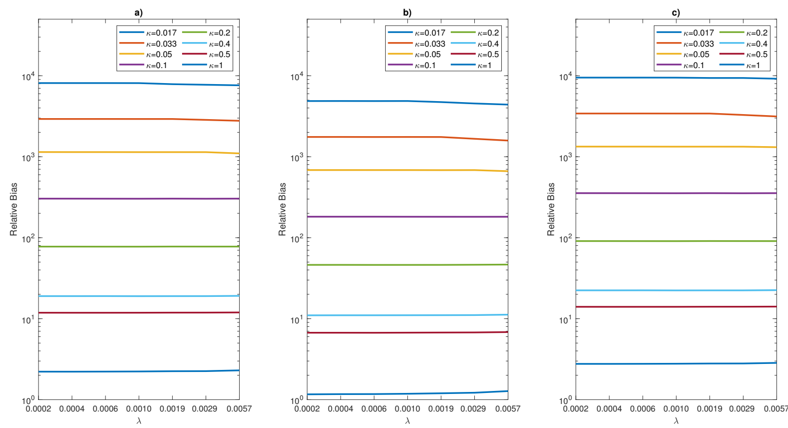

Fig. 4.3: Daily PSRV finite-sample relative bias as a function of λ for values of $\kappa \in (0.017, 0.033, 0.05, 0.1, 0.2, 0.4, 0.5, 1)$ and $\delta_N = 1$ minute, $b = -1/2$, $c = 1/4$. The values of λ on the x-axis correspond to Δ_N equal to $j\delta_N$, for $j = 1, 2, 3, 5, 10, 15, 30$. The panels refer to the following parameter sets: a) *Set 1*: $(\alpha, \theta, \gamma, \rho, \nu(0)) = (0.2, 5, 0.5, -0.2, 0.2)$; b) *Set 2*: $(\alpha, \theta, \gamma, \rho, \nu(0)) = (0.03, 10, 0.25, -0.8, 0.03)$; and c) *Set 3*: $(\alpha, \theta, \gamma, \rho, \nu(0)) = (0.2, 5, 0.5, -0.2, 0.4)$. The y-axis is expressed in log-scale.

4.3 Analytical results

In this section we analyze the PSRV finite-sample bias in a parametric setting, namely assuming that the volatility is a CIR process, so that a fully explicit formula of the latter can be obtained. We treat the overlapping case (i.e., the case when $W_N > \Delta_N$) and the no-overlapping case (i.e., $W_N \le \Delta_N$) separately as, in general, the finite-sample bias expression differs in the two cases, independently of the parametric model used. Lemma 4.1 details the explicit expression of the PSRV bias for N fixed.

Lemma 4.1. *Let Assumption 4.1 hold and let N be fixed. If $W_N \le \Delta_N$, the bias of the PSRV in Definition 4.2 reads*

$$E\left[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\right] = \gamma^2 \alpha h (A_N - 1) + \gamma^2 \Big(E[\nu(\tau)] - \alpha\Big) \frac{1 - e^{-\theta h}}{\theta} (B_N - 1) + C_N. \tag{4.1}$$

Instead, if $W_N > \Delta_N$, it reads

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \gamma^2\alpha h(A_N-1)+\gamma^2\Big(E[\nu(\tau)]-\alpha\Big)\frac{1-e^{-\theta h}}{\theta}(B_N-1)+C_N+O_N. \tag{4.2}$$

The parametric expressions of A_N, B_N, C_N and O_N are rather cumbersome and thus are reported in the Appendix (see, respectively, equations (4.13), (4.14), (4.15) and (4.16) in the proof to Lemma 4.1).

Proof. See the Appendix. □

Remark 4.2. *The bias in the case $W_N \leq \Delta_N$ differs from that in the case $W_N > \Delta_N$ for the presence of the extra term O_N, which appears due to the fact that the parametric expression of $E[RV(\tau+i\Delta_N, k_N\delta_N)RV(\tau+i\Delta_N-\Delta_N, k_N\delta_N)]$ differs in the two cases. See the proof of Lemma 4.1 for further details.*

In the next subsections we investigate the existence of a rule to select the tuning parameters κ and λ in both the cases $W_N > \Delta_N$ and $W_N \leq \Delta_N$. To do so, we first isolate the leading term of the bias in each case and then verify whether the latter can be canceled by an *ad hoc* selection of tuning parameters. We address overlapping case first, as it is the one relevant for practical applications, based on the results of the simulation studies in Section 4.2 and in Sanfelici et al. (2015).

4.3.1 The relevant case for practical applications: $W_N > \Delta_N$

When $W_N > \Delta_N$, the natural expansion of the bias as the number of sampled price observations N tends to infinity is precluded, because the consistency of the PSRV requires that $\frac{W_N}{\Delta_N} \to 0$ as $N \to \infty$. Thus, we determine the leading term of the bias through an alternative asymptotic expansion, which exploits some natural, non-restrictive constraints on the magnitude of the tuning parameter λ and the time horizon h. Specifically, we first regard the bias in equation (4.2) as a function of λ and we perform its Taylor expansion with base point $\lambda = 0$. Then, regarding each term of this expansion as a function of h, we perform their Taylor expansions with base point $h = 0$. The choice of the base point $\lambda = 0$ is supported by the fact that the largest feasible values of λ are very small, e.g., on the order of 10^{-3} when $c < 1/2$ and δ_N is equal to one minute (see Figure 4.2 for the case $c = 1/4$). Note that a value of λ is feasible if it satisfies $\Delta_N := \lambda\,\delta_N^c < h$. The choice of base point $h = 0$ is instead supported by the fact that in the literature on high-frequency econometrics, the typical time horizon used to estimate the integrated quantities is one trading day, i.e., $h = 1/252 \approx 4\cdot 10^{-3}$. The order of this sequential expansion is rather natural: intuitively, we first take the limit $\lambda \to 0$ to approximate the integral of the vol-of-vol in an infill-asymptotics sense, then take the limit $h \to 0$ to localize the estimate of the integral near the initial time τ. This approach leads to the following result.

Theorem 4.1. *Let Assumption 4.1 hold. Further, let $W_N > \Delta_N$. Then, as $\lambda \to 0, h \to 0$*

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \begin{cases} \left(\dfrac{4E[\nu(\tau)]^2}{\kappa^2\delta_N^{1+2b}} - \gamma^2 E[\nu(\tau)]\right)h + O(h^{1-b}) + O(\lambda) & \text{if} \quad b \geq -1/2, c < -b \\ -\gamma^2 E[\nu(\tau)]h + O(h^{-2b}) + O(\lambda) & \text{if} \quad b < -1/2, c < 1+b \end{cases} \tag{4.3}$$

Moreover, let $(\mathscr{F}_t^\nu)_{t\geq 0}$ be the natural filtration associated with the process ν. Then, for N fixed, as $\lambda \to 0, h \to 0$

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]} | \mathscr{F}_\tau^\nu\Big] = \begin{cases} \left(\dfrac{4\nu(\tau)^2}{\kappa^2\delta_N^{1+2b}} - \gamma^2 \nu(\tau)\right)h + O(h^{1-b}) + O(\lambda) & \text{if} \quad b \geq -1/2, c < -b \\ -\gamma^2 \nu(\tau)h + O(h^{-2b}) + O(\lambda) & \text{if} \quad b < -1/2, c < 1+b \end{cases} \tag{4.4}$$

Proof. See the Appendix. □

Remark 4.3. *The expansion in Theorem 4.1 is performed under the asymptotic constraints on rates b and c that ensure the asymptotic unbiasedness of the PSRV under Assumption 4.1 (see Theorem 4.2).*

Remark 4.4. *The conditional bias expansion in equation (4.4) allows the dominant term of the bias to be expressed in terms of $\nu(\tau)$ and γ, two quantities that can be consistently estimated over a fixed time horizon. This is crucial for the existence of a feasible procedure to select κ, as detailed below. Instead, the unconditional expression in equation (4.3) depends on $E[\nu(\tau)]$, whose parametric expression in turn depends on the drift parameters of the volatility and thus cannot be consistently estimated over a fixed time horizon (see, e.g., Kanaya and Kristensen (2016)). In particular, it holds $E[\nu(\tau)] = (\nu(0)-\alpha)e^{-\theta\tau} + \alpha$ (see equation (4) in Section 2.2.1 of Bollerslev and Zhou (2002)).*

Figure 4.4 compares the true finite-sample bias of the daily PSRV in equation 4.2 with the dominant term of the expansion in equation 4.3 as functions of the tuning parameter κ. Specifically, the panels refer to the three parameter sets already used in Section 4.2, that is, *Set 1* (panel a)), *Set 2* (panel b)), and *Set 3* (panel c)). Note that we have set $b = -1/2$, $c = 1/4$, $\lambda = 0.0006$, $h = 1/252$, and $N = 360$. The corresponding δ_N and Δ_N are equal to 1 minute and (approximately) 3 minutes, as we consider 6-hr trading days. The approximation of the true bias with the dominant term of the expansion is very accurate.

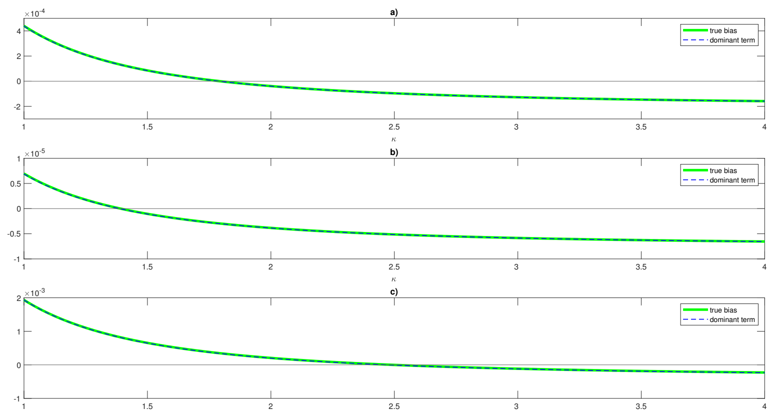

Fig. 4.4: Comparison between the true finite-sample bias of the daily PSRV and the dominant term of the expansion in Theorem 4.1 when $W_N > \Delta_N$ as functions of κ for $b = -1/2$, $c = 1/4$, $\lambda = 0.0006$, and $N = 360$. Panel a) refers to the parameter set $(\alpha, \theta, \gamma, \nu(0)) = (0.2, 5, 0.5, 0.2)$; panel b) to $(\alpha, \theta, \gamma, \nu(0)) = (0.03, 10, 0.25, 0.03)$, and panel c) to $(\alpha, \theta, \gamma, \nu(0)) = (0.2, 5, 0.5, 0.4)$. For panel c) we consider $\tau = 5$ days, while the bias terms in panels a) and b) are independent of τ. The PSRV finite-sample bias is independent of the correlation parameter ρ, which therefore does not appear.

Based on the conditional bias expansion in equation (4.4), we make the following considerations on the optimal selection of tuning parameters in finite samples. First, we note that the dominant term of the bias can be annihilated simply by suitably selecting κ for any feasible value of λ when $b \geq -1/2$, $c < -b$. Instead, when $b < -1/2$, $c < 1 + b$, the dominant term of the bias is independent of κ and λ. Specifically, when $b \geq -1/2$, $c < -b$, the suitable selection is

$$\kappa = \frac{2\sqrt{\nu(\tau)}}{\hat{\gamma}\delta_N^{b+1/2}}.$$

However, since κ is a tuning parameter, it is not allowed to depend on N. Therefore, the only admissible choice is $b = -1/2$ and $c < 1/2$, so that the suitable selection becomes

$$\kappa = \kappa^* := \frac{2\sqrt{\nu(\tau)}}{\gamma}. \tag{4.5}$$

Further, if $\nu(0) = \alpha$, then $E[\nu(\tau)] = \alpha$ (see equation (4) in Section 2.2.1 of Bollerslev and Zhou (2002)) and thus, based on equation (4.3), it is immediate to see that the bias-optimal value of κ reduces to $\frac{2\sqrt{\alpha}}{\gamma}$. Interestingly, this analytical result supports the optimal selections of b and κ determined numerically in the literature. Indeed, for the first parameter set in the numerical exercise in Section 4.2, *Set 1*, which is also used in Sanfelici et al. (2015), κ^* is equal to 1.79, a value compatible with the numerical result in Sanfelici et al. (2015), where the optimal κ is said to be approximately equal to 2. Note also that the numerical studies in Aït-Sahalia et al. (2017), Sanfelici et al. (2015) both select $b = -1/2$. Furthermore, the following remark regarding the selection $\kappa = \kappa^*$ is in order.

Remark 4.5. *The overlapping condition $W_N > \Delta_N$ implies a constraint on the price grid δ_N. In particular, if $\kappa = \kappa^*$, for $b = -1/2$, $c < 1/2$, $W_N > \Delta_N$ is equivalent to $\delta_N > \delta^* := \left(\frac{\kappa^*}{\lambda}\right)^{\frac{1}{c-1/2}}$. The threshold δ^* is very small for typical orders of magnitude of α, θ and γ, h corresponding to one trading day and any feasible value of λ. For example, for the values of the parameters in Set 1 (see Section 4.2), $\lambda = 0.0006$ and $c = 1/4$ (so that, if $\delta_N = 1$ minute, then $\Delta_N := \lambda \delta_N^c \approx 3$ minutes), we have $\delta^* = 7.5 \cdot 10^{-8}$ seconds and thus the constraint $\delta_N > \delta^*$ is largely satisfied at the most commonly available price sampling frequencies.*

However, Equation (4.5) implies that the bias-optimal selection $\kappa := \kappa^*$ is unfeasible unless reliable estimates of $\nu(\tau)$ and γ are available. In the Appendix we detail a simple feasible procedure to obtain κ^*. In a nutshell, the procedure is as follows. First, we estimate $\nu(\tau)$ using the Fourier spot volatility estimator by Malliavin and Mancino (2009). Then we estimate γ via a simple indirect inference method.

4.3.2 The case $W_N \leq \Delta_N$

The finite-sample bias expression for $W_N \leq \Delta_N$ in equation (4.1) is the starting point to derive the asymptotic constraints on rates b and c that ensure the asymptotic unbiasedness of the PSRV. In this regard, we obtain the following result, which is based on the asymptotic expansion of the bias in the limit $N \to \infty$.

Theorem 4.2. *Let Assumption 4.1 hold. Then, if $b \geq -1/2$ and $c < -b$ or $b < -1/2$ and $c < 1 + b$, $\frac{W_N}{\Delta_N} \to 0$ as $N \to \infty$ and the PSRV as given in Definition 4.2 is asymptotically unbiased, i.e.,*

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] \to 0 \quad as \quad N \to \infty.$$

In particular, as $N \to \infty$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu \rangle_{[\tau,\tau+h]}\Big] = a_1\Delta_N + a_2\frac{1}{k_N\Delta_N} + a_3\frac{k_N\delta_N}{\Delta_N} + o\Big(\Delta_N\Big) + o\Big(\frac{1}{k_N\Delta_N}\Big) + o\Big(\frac{k_N\delta_N}{\Delta_N}\Big), \quad (4.6)$$

where:

$$a_1 = -\frac{\theta}{2}\gamma^2\alpha h + \frac{\theta}{2}\gamma^2(E[\nu(\tau)]-\alpha)\frac{1-e^{-\theta h}}{\theta} + \frac{\theta}{2}(1-e^{-2\theta h})\Big[(E[\nu(\tau)]-\alpha)^2 + \frac{\gamma^2}{\theta}\Big(\frac{\alpha}{2} - E[\nu(\tau)]\Big)\Big],$$

$$a_2 = \frac{2}{\theta}\gamma^2\alpha h + \frac{4}{\theta}\gamma^2(E[\nu(\tau)]-\alpha)\frac{1-e^{-\theta h}}{\theta} + \frac{2}{\theta}(1-e^{-2\theta h})\Big[(E[\nu(\tau)]-\alpha)^2 + \frac{\gamma^2}{\theta}\Big(\frac{\alpha}{2} - E[\nu(\tau)]\Big)\Big] + 4\alpha^2 h + \frac{8\alpha(E[\nu(\tau)]-\alpha)(1-e^{-\theta h})}{\theta},$$

$$a_3 = -\gamma^2(E[\nu(\tau)]-\alpha)\frac{1-e^{-\theta h}}{\theta}.$$

Proof. See the Appendix. □

A bias-optimal rule for the selection of the tuning parameters κ and λ when $W_N \le \Delta_N$ is given in the following corollary to Theorem 4.2. Unfortunately, this bias-optimal rule is of little interest for practical applications, as explained in Remark 4.6.

Corollary 4.1. *The leading term of the PSRV finite-sample bias expansion in equation (4.6) can be canceled in the case $b=-1/2$ and $c=1/4$, provided that there exists a solution $(\tilde\kappa,\tilde\lambda) \in \mathbb{R}_{>0}\times\mathbb{R}_{>0}$ to the following system:*

$$\begin{cases} a_3\kappa^2 + a_1\lambda^2\kappa + a_2 = 0 \\ W_N \le \Delta_N \end{cases}.$$

If a solution $(\tilde\kappa,\tilde\lambda) \in \mathbb{R}_{>0}\times\mathbb{R}_{>0}$ exists, the corresponding bias-optimal selection of W_N and Δ_N reads

$$W_N = \tilde\kappa\delta_N^{1/2},\ \Delta_N = \tilde\lambda\delta_N^{1/4}.$$

Proof. See the Appendix. □

Remark 4.6. *For $b=-1/2$ and $c=1/4$, the no-overlapping condition $W_N \le \Delta_N$ is equivalent to $\delta_N \le (\lambda/\kappa)^4$. Assuming that a positive solution $\tilde\kappa(\lambda)$ to $a_3\kappa^2 + a_1\lambda^2\kappa + a_2 = 0$ exists for some $\lambda > 0$, we define the "no-overlapping" threshold for δ_N as $\delta^*(\lambda) := (\lambda/\tilde\kappa(\lambda))^4$. For the three sets of CIR parameters used in the numerical study in Section 4.2, Figure 4.5 plots the threshold $\delta^*(\lambda)$ as a function of $\lambda \in (0,\lambda^*]$, where λ^* is the largest admissible value of λ such that*

$\lambda \delta^*(\lambda)^{1/4} \leq h$, i.e., such that $\Delta_N \leq h$ when δ_N is equal to the "no-overlapping" threshold. Specifically, Figure 4.5 shows that the sampling frequency corresponding to $\delta^*(\lambda)$ is bounded by a value smaller than, respectively, 0.02 (see Panel a)), 0.05 (see Panel b)) and 0.125 seconds (see Panel c)). This suggests that for typical values of the CIR parameters, the system in Corollary 1 may be solved only for ultra-high frequencies at which prices of financial assets are typically not available. Also, note that the solution (if it exists) depends on a_1, a_2 and a_3, which in turn depend on the expected initial volatility and all CIR parameters, including the drift parameters, which can not be consistently estimated over a fixed time horizon.

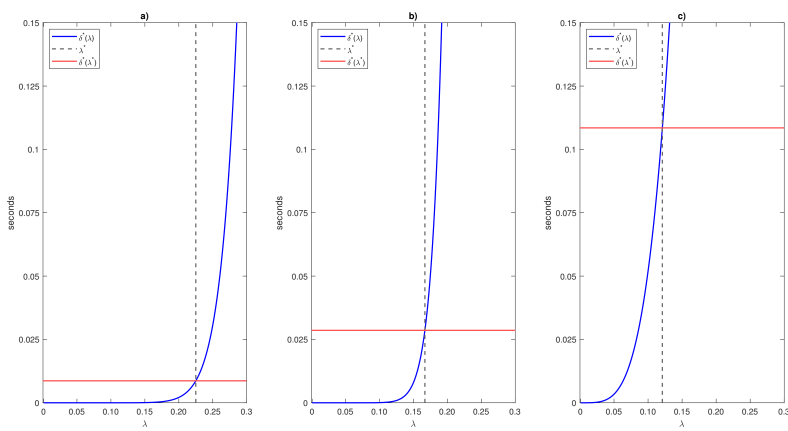

Fig. 4.5: The threshold $\delta^*(\lambda)$ is plotted in blue. The dotted gray vertical line corresponds to $\lambda = \lambda^*$. The dotted red horizontal line corresponds to $\delta^*(\lambda^*)$. The panels refer to the following sets of parameters: a) *Set 1*: $(\alpha, \theta, \gamma, \nu(0)) = (0.2, 5, 0.5, 0.2)$; b) *Set 2*: $(\alpha, \theta, \gamma, \nu(0)) = (0.03, 10, 0.25, 0.03)$; and c) *Set 3*: $(\alpha, \theta, \gamma, \nu(0)) = (0.2, 5, 0.5, 0.4)$. $\delta^*(\lambda)$ is independent of the correlation parameter ρ, which therefore does not appear. For panel c) we consider $\tau = 5$ days, while in panels a) and b) $\delta^*(\lambda)$ is independent of τ. We have assumed $h = 1/252$, corresponding to 6 hours (21600 seconds).

4.3.3 The impact of noise on the bias

In empirical applications one can only observe the noisy price $\tilde{x}(t)$, that is, the efficient price contaminated by a noise component that originates from market microstructure frictions, such as bid-ask bounce effects and price rounding. Here, we assume that the noise component is an i.i.d. process independent of the efficient price process, as in the seminal paper by Roll (1984). For a general discussion of the statistical models of microstructure noise, see Jacod et al. (2017).

Assumption 4.2. ***Data-generating process in the presence of noise***
The observable price process $\tilde{x}$ is given by

$$\tilde{x}(t) = x(t) + \varepsilon(t),$$

where $x(t)$ represents the efficient price process and evolves according to Assumption 4.1 while $\varepsilon(t)$ is a sequence of i.i.d. random variables independent of $x(t)$, such that $E[\varepsilon(t)] = 0$, $E[\varepsilon(t)^2] = V_\varepsilon < \infty$ and $E[\varepsilon(t)^4] = Q_\varepsilon < \infty \; \forall t$.

The presence of noise clearly changes the PSRV bias expression, introducing an extra term, as illustrated in the following lemma. Note that the parametric form of the extra bias term due to the presence of noise is different in the overlapping and no-overlapping cases.

Lemma 4.2. *Let Assumption 4.2 hold and let N be fixed. Moreover, let $\widetilde{PSRV}_{[\tau,\tau+h],N}$ denote the PSRV in Definition 4.2, computed from noisy price observations. If $W_N \leq \Delta_N$, then*

$$E\left[\widetilde{PSRV}_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]}\right] = \gamma^2\alpha h(A_N-1) + \gamma^2\Big(E[\nu(\tau)]-\alpha\Big)\frac{1-e^{-\theta h}}{\theta}(B_N-1)+C_N+D_N; \tag{4.7}$$

Instead, if $W_N > \Delta_N$, then

$$E\left[\widetilde{PSRV}_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]}\right] = \gamma^2\alpha h(A_N-1) + \gamma^2\Big(E[\nu(\tau)]-\alpha\Big)\frac{1-e^{-\theta h}}{\theta}(B_N-1)+C_N+O_N+D_N^*. \tag{4.8}$$

The parametric expressions of A_N, B_N, C_N and O_N are as in Lemma 4.1, while that of the extra term due to the presence of noise D_N (respectively, D_N^*) in the no-overlapping case (respectively, overlapping case) is as follows:

$$D_N = [4(Q_\varepsilon + V_\varepsilon^2) + 16\alpha V_\varepsilon \delta_N]h\frac{1}{k_N\delta_N^2\Delta_N} + \frac{8}{\theta}V_\varepsilon(\alpha - E[\nu(\tau)])(1-e^{-\theta h})\frac{(1+e^{-\theta\Delta_N})(1-e^{-\theta k_N\delta_N})}{(1-e^{-\theta\Delta_N})k_N^2\delta_N^2}; \tag{4.9}$$

$$\begin{aligned} D_N^* = {} & [4\left(Q_\varepsilon + V_\varepsilon^2\right) + 16\alpha V_\varepsilon\delta_N]h\frac{1}{k_N^2\delta_N^3} + \frac{8V_\varepsilon(\alpha - E[\nu(\tau)])(1-e^{-\theta h})}{\theta(1-e^{-\theta\Delta_N})k_N^2\delta_N^2}\Bigg\{\frac{(2+k_N)}{2k_N\delta_N} \\ & \times\left[\frac{(e^{\theta k_N\delta_N - \theta\Delta_N}-1)(k_N\delta_N+\Delta_N)}{k_N\delta_N-\Delta_N} + (e^{-\theta\Delta_N} - e^{\theta k_N\delta_N})\right] + \frac{k_N}{2\Delta_N}(1+e^{\theta k_N\delta_N})(1-e^{-\theta\Delta_N})\Bigg\}. \end{aligned} \tag{4.10}$$

Proof. See the Appendix. □

Remark 4.7. *From the proof of Theorem 4.3 in the Appendix, one can easily see that the expressions of D_N is the same for any continuous mean-reverting volatility model, as their computation only depends on the drift of ν in Assumption 4.1. The same holds also for D_N^* in the overlapping case.*

Ideally, in the overlapping case, if one could efficiently estimate the extra bias due to noise D_N^* and subtract it, then the bias-optimal rule to select κ could still be applied effectively. Unfortunately, D_N^* can not be consistently estimated over a fixed time horizon, as it depends on the drift parameters of the volatility α and θ, whose consistent estimation can not be achieved on a fixed time horizon[2]. As a solution, we suggest to sample prices on a suitably sparse grid, as done for the realized variance in the seminal paper by Andersen et al. (2001a), so that the extra bias term induced by the presence of noise becomes negligible and the bias optimal rule to select the local-window parameter κ can still be applied. The efficiency of this solution is verified numerically in Section 4.5.

Finally, for completeness, we also study the asymptotic behaviour of the additional bias due to noise in the no-overlapping case, D_N. More precisely, in the next theorem we derive its rate of divergence as $N \to \infty$.

Theorem 4.3. *Let Assumption 4.2 hold. Moreover, let $\widetilde{PSRV}_{[\tau,\tau+h],N}$ denote the PSRV in Definition 4.2, computed from noisy price observations. Then, if either $b \geq -\frac{1}{2}$ and $c < -b$ or $b < -\frac{1}{2}$ and $c < b+1$, $\frac{W_N}{\Delta_N} \to 0$ as $N \to \infty$ and $\widetilde{PSRV}_{[\tau,\tau+h],N}$ is asymptotically biased, i.e.,*

$$E\left[\widetilde{PSRV}_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle\right]_{[\tau,\tau+h],N} \to \infty, \quad as \quad N \to \infty,$$

since the bias term D_N in equation (4.7) of Lemma 4.2 diverges as $N \to \infty$. In particular, we have

$$k_N \delta_N^2 \Delta_N D_N = 4(Q_\varepsilon + V_\varepsilon^2)h + O(\delta_N) \quad and \quad k_N \delta_N^2 \Delta_N \to 0, \ \delta_N \to 0.$$

Proof. See the Appendix. □

[2] Note that Q_ε and V_ε can, instead, be estimated consistently for T fixed, see for instance Zhang et al. (2005)

4.3.4 The bias-reducing effect of the assumption $\nu(0) = \alpha$

As mentioned in Section 4.3.1, if $\nu(0) = \alpha$, then $E[\nu(\tau)] = \alpha$. Lemmas 4.1 and 4.2 quantify the bias reduction ensuing from assuming that $\nu(0) = \alpha$. Indeed, this assumption cuts off the entire source of bias B_N and part of the sources of bias D_N (see equation (4.9)) or D_N^* (see equation (4.10)). The finite-sample bias reduction ensuing from the assumption $\nu(0) = \alpha$ is not peculiar to the PSRV, though. In fact, this simplifying assumption is also beneficial for reducing the finite-sample bias of the locally averaged realized variance, as shown in the next theorem.

Theorem 4.4. *Let Assumption 4.1 hold. Moreover, let $\hat{\nu}(\tau)$ denote the locally averaged realized variance in Definition 4.1 at time τ. Then, if $b \in (-1,0)$, $\hat{\nu}(\tau)$ is asymptotically unbiased, i.e.,*

$$E[\hat{\nu}(\tau) - \nu(\tau)] = (\nu(0) - \alpha)e^{-\theta\tau}\frac{e^{\theta k_N \delta_N} - 1 - \theta k_N \delta_N}{\theta k_N \delta_N},$$

and, as $N \to \infty$, we have

$$E[\hat{\nu}(\tau) - \nu(\tau)] = \frac{\theta}{2}(\nu(0) - \alpha)e^{-\theta\tau}k_N \delta_N + o(k_N \delta_N), \quad k_N \delta_N \to 0.$$

Let Assumption 4.2 hold. Moreover, let $w(\tau)$ denote the locally averaged realized variance in Definition 4.1 at time τ computed from noisy price observations. Then, $\forall\, b \in (-1,0)$, $w(\tau)$ is asymptotically biased, i.e.,

$$E[w(\tau) - \nu(\tau)] = (\nu(0) - \alpha)e^{-\theta\tau}\frac{e^{\theta k_N \delta_N} - 1 - \theta k_N \delta_N}{\theta k_N \delta_N} + \frac{2V_\varepsilon}{\delta_N},$$

and, as $N \to \infty$, we have

$$E[w(\tau) - \nu(\tau)] = \frac{\theta}{2}(\nu(0) - \alpha)e^{-\theta\tau}k_N \delta_N + \frac{2V_\varepsilon}{\delta_N} + o(k_N \delta_N), \qquad k_N \delta_N \to 0.$$

Proof. See the Appendix. □

This theorem has two interesting implications. First, under Assumption 4.1, the locally averaged realized variance is unbiased in finite samples if and only if $\nu(0) = \alpha$. Second, under Assumption 4.2, if $\alpha > \nu(0)$, the presence of noise could actually compensate for the negative bias originating from the first term of the bias expression. This also holds for the PSRV finite-sample bias, provided that the term D_N (respectively, D_N^*) in Lemma 4.2 is of opposite sign with respect to the sum of the other terms in the bias expression.

4.4 Generalization via dimensional analysis

In this section we propose a heuristic approach, based on dimensional analysis, to generalize the rule for the bias-optimal selection of κ in equation (4.5), derived under the assumption that the volatility is a CIR process, to the more general case where the volatility follows a process in the CKLS class (see Chan et al. (1992)). Specifically, the stochastic volatility model we assume as the data-generating process is now as follows.

Assumption 4.3. ***Generalized data-generating process***

For $t \in [0,T]$, $T > 0$, the dynamics of the log-price process x and the spot volatility process ν follow

$$dx(t) = \sqrt{\nu(t)}dW(t) + \mu(t)dt$$
$$d\nu(t) = \gamma\nu(t)^{\beta}dZ(t) + \Big(\alpha - \nu(t)\Big)dt$$

where W and Z are two Brownian motions, with correlation parameter ρ, on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$, which satisfies the usual conditions, $\mu(t)$ is a continuous adapted process, $\beta \geq 1/2$, $\theta, \alpha, \gamma > 0$, and $2\alpha\theta > \gamma^2$ if $\beta = 1/2$.

The stochastic volatility model in Assumption 4.3 is quite flexible to reproduce empirical prices behaviour in the absence of price and volatility jumps. In fact it incorporates a number of widely-used stochastic volatility models with continuous price and volatility paths as special cases. For example, if $\beta = 1/2$, one obtains the model by Heston (1993); if $\beta = 1$ one finds the conitnuous-time Garch model by Nelson (1990); if $\beta = 3/2$, one gets the 3/2 model by Platen (1997). Further, by allowing for a stochastic correlation between W and Z, Assumption 4.3 includes also the generalized Heston model with stochastic leverage introduced by Veraart and Veraart (2012). Finally, note that Assumption 4.3 also includes a price drift. The numerical study in Section 4.5 confirms that the impact of the latter on the PSRV finite-sample bias is negligible.

We now use dimensional analysis to heuristically derive a rule for the bias-optimal selection of κ under Assumption 4.3. We test the efficacy of this rule in the numerical study of Section 4.5, with overwhelming results. Note that dimensional analysis is typically used in physics and engineering to make an educated guess about the solution to a problem without performing a full analytical study (see, e.g., Kyle and Obizhaeva (2017), Smith et al. (2003)).

The basic concept of dimensional analysis is that one can only add quantities with the same units[3]. Accordingly, when applying dimensional analysis, the first step entails identifying the units of the quantities appearing in the equations being studied. In this specific analysis, we start with the units of the quantities appearing in the model given in Assumption 4.3. Let $dim[q]$ denote the unit/dimension of the quantity q. The log-return $dx(t)$, $t > 0$, is a dimensionless quantity (i.e., a pure number) since it is the logarithm of a ratio of prices. Instead, the quadratic variation of the Wiener processes W and Z has the dimension of *time* since W and Z are continuous-time random walks. As a consequence, we have $dim[dW(t)] = dim[dZ(t)] = time^{1/2}$ (see, for example, Wilmott (2000) or the square-root-of-time rule in Danielsson and Zigrand (2006)). Now consider the dynamics of the log-price, bearing in mind that we cannot add or subtract quantities with different measurement units. The dimension of the left-hand side must then be equal to those of the addenda on the right-hand side, thereby implying that $dim[\mu] = 1/time$ and $\dim[\nu(t)] = 1/time$. Thus, from the dynamics of $\nu(t)$, we have $dim[\alpha] = 1/time$, $dim[\theta] = 1/time$ and $dim[\gamma\nu(t)^{\beta}\, dZ(t)] = 1/time$. The latter implies $dim[\gamma]dim[\nu(t)^{\beta}]dim[dZ(t)] = 1/time$. Therefore, bearing in mind that $dim[\nu(t)^{\beta}]dim[dZ(t)] = 1/time^{\beta-1/2}$, we obtain $dim[\gamma] = 1/time^{-\beta+3/2}$.

Now, without loss of generality, let $\nu(0) = \alpha$ and consider the dominant term in the expansion of Theorem 4.1, i.e., the term

$$\Big(\frac{4\alpha}{\kappa^2\delta_N^{1+2b}} - \gamma^2\alpha\Big)h.$$

Since the dominant term of the PSRV bias must clearly have the same dimension as the expected quadratic variation of ν over any generic interval of length h, i.e., $\gamma^2\alpha h$, we have

$$dim\Big[\Big(\frac{4\alpha}{\kappa^2\delta_N^{1+2b}} - \gamma^2\alpha\Big)h\Big] = dim[\gamma^2\alpha h] = 1/time^2,$$

and, as one can easily verify, this implies $dim[\kappa] = time^{-b}$ (alternatively, one can show that $dim[\kappa] = time^{-b}$ by simply noting that $k_N = \kappa\delta_N^b$ is dimensionless and $dim[\delta_N^b] = time^b$).

Now observe that the leading term of any expansion of the PSRV finite-sample bias must have dimension equal to $1/time^2$. Based on this observation, we conjecture that the leading term of the expansion in Theorem 4.1 under Assumption 4.3 is

[3] Dimensional analysis is also called a unit-factor method or a factor-label method, since a conversion factor is used to evaluate the units.

$$\Big(\frac{4E[\nu(\tau)]^2}{\kappa^2\delta_N^{1+2b}} - \gamma^2 E[\nu(\tau)^{2\beta}]\Big)h,$$

whose dimension is $1/time^2$, as one can easily check by recalling that $dim[\kappa] = time^{-b}$, $\dim[\nu(t)] = 1/time$ and $dim[\gamma] = 1/time^{-\beta+3/2}$. Accordingly, if one conditions the bias to the natural filtration of $\nu(t)$ up to time $t = \tau$, the generalized bias-optimal value of κ, for $b = -1/2$ and $c < 1/2$, reads

$$\kappa^{**} := 2\frac{\nu(\tau)^{1-\beta}}{\gamma}, \tag{4.11}$$

Note that equation (4.11) can be rewritten in non-parametric form as

$$\kappa^{**} = 2\frac{\nu(\tau)}{\sqrt{\xi(\tau)}},$$

where $\xi(t) := \gamma^2\nu(t)^{2\beta}$ is the vol-of-vol process. This result, while offering insight into the non-parametric solution to the problem of the bias-optimal selection of κ, is problematic in terms of feasibility as it requires the estimation of the spot vol-of-vol $\xi(t)$ at $t = \tau$, a challenging issue which has not been addressed so far in the literature to the best of our knowledge and goes beyond the scope of this paper.

Our conjecture is based on the origin of the two addenda in the leading term of the bias (see Theorem 4.1) in the CIR framework. In fact, bearing in mind the the leading term is

$$\left(\frac{4E[\nu(\tau)]^2}{k^2\delta_N^{1+2b}} - \gamma^2 E[\nu(\tau)]\right)h,$$

we note that the second addendum, i.e., $\gamma^2 E[\nu(\tau)]h$, comes from the expected quadratic variation of the volatility process. More specifically, it originates from the leading term of the following expansion:

$$E\Big[\langle\nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \gamma^2 E[\nu(\tau)]h + o(h), \quad h\to 0.$$

Instead, the first addendum, i.e., $\frac{4E[\nu(\tau)]^2}{k^2\delta_N^{1+2b}}$, is due to the drift of the volatility process.

Thus in the case of the CKLS model, the first addendum remains unchanged since the drift of the process is the same for any β, while the second addendum changes according to the expected quadratic variation of the volatility process, which, for small h, reads

$$E\Big[\langle\nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \gamma^2 E[\nu(\tau)^{2\beta}]h + o(h), \quad h\to 0.$$

since $E\Big[\langle\nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \gamma^2\int_\tau^{\tau+h} E[\nu(s)^{2\beta}]ds$.

4.5 Numerical results

4.5.1 Numerical results in the CIR setting

As detailed in Section 4.3, in the absence of microstructure noise and assuming $\nu(\tau)$ to be observable and γ to be known, the finite-sample bias of the PSRV is optimized, under Assumption 4.1 and for any $\nu(0)$, by selecting $b = -1/2$, $c < 1/2$ and $\kappa = \kappa^* := 2\sqrt{\nu(\tau)}\gamma^{-1}$. In this subsection, we give numerical confirmation of the optimality of this rule for the selection of κ in three progressively more realistic scenarios, where incremental sources of biases are added.

In the first scenario, we simulate log-price paths under Assumption 4.1 and compute daily PSRV values from noise-free price observations assuming that the CIR parameters are known and the initial volatility value $\nu(\tau)$ is observable. In this scenario, we use two price sampling frequencies, that is, $\delta_N = 1$ minute and $\delta_N = 5$ minutes. Results show that the bias generated by the price discrete sampling is relatively small, e.g., less than 5% if $\delta_N = 1$ minute when $\kappa = \kappa^*$ (see Table 4.1).

In the second scenario, we simulate log-price paths under Assumption 4.2 and compute PSRV values from noisy prices while assuming that the CIR parameters are known and the initial volatility value $\nu(\tau)$ is observable. As the PSRV is not robust to the presence of noise contaminations in the price process, here we only consider the sampling frequency $\delta_N = 5$ minutes, as recommended in the seminal paper by Andersen et al. (2001a), where the authors suggest that this sampling frequency reduces the impact of noise on returns while still falling within a high-frequency framework. Indeed, a comparison of the numerical results obtained in these first two scenarios shows that the impact of the price noise on the PSRV estimates is relatively small at the 5-minute sampling frequency, when $\kappa = \kappa^*$ is used.

In the third scenario, we still simulate the log-price path under Assumption 4.2, but the value of the initial volatility, $\nu(\tau)$, is now unobservable and the model parameter γ is unknown. Thus, we compute PSRV values from noisy prices by selecting $\kappa = \hat{\kappa}^* := \frac{2\sqrt{\hat{\nu}(\tau)}}{\hat{\gamma}}$. Here, $\hat{\nu}(\tau)$ and $\hat{\gamma}$ are obtained through the estimation procedure detailed in the Appendix. A comparison of the results obtained in these different scenarios shows that the PSRV finite-sample bias reduction obtained with the feasible selection $\kappa = \hat{\kappa}^*$ is very similar to the reduction obtained with the unfeasible selection $\kappa = \kappa^*$. Overall, for each scenario, we consider the three sets of parameters described in Section 4.2. For each parameter set, we simulate one thousand 1-year trajectories of 1-second observations.
The noise component ε in Assumption 4.2 is simulated as an i.i.d. Gaussian process, with noise-to-signal ratio ζ ranging from 0.5 to 3.5, as in the numerical exercise proposed in Sanfelici et al. (2015). We define the noise-to-signal ratio ζ as in Sanfelici et al. (2015), i.e., $\zeta := \frac{std(\omega)}{std(r)}$, where ω and r denote, respectively, the

increment of the noise process ε in Assumption 4.2 and the noise-free log-return, both computed at the maximum sampling frequency available in our numerical exercise, that is, 1 second.

From the simulated prices, we compute daily PSRV values, that is, we set a small time horizon h, i.e., $h = 1/252$. Recall that the bias-optimal rule for the selection of κ is valid when $b = -1/2$ and $c < 1/2$. Accordingly, we set $b = -1/2$ and $c = 1/4$ in our numerical study.

Tables 4.1–4.3 summarize the results of our numerical exercises and, to make the results related to the three parameter sets comparable, we report the values of the relative bias. Since we simulate 6-hr days, N is equal to 360 when $\delta_N = 1$ minute and 72 when $\delta_N = 5$ minutes. Note that the overlapping condition $W_N > \Delta_N$ is always satisfied for the values of Δ_N in Table 4.1. In particular, the average length of W_N is approximately equal to: 530 minutes for *Set 1*, 410 minutes for *Set 2* and 580 for *Set 3*, when $\delta_N = 1$ minute; 1200 minutes for *Set 1*, 930 minutes for *Set 2* and 1310 minutes for *Set 3*, when $\delta_N = 5$ minutes. These averages are computed over all simulated days and are stable across the three scenarios. Recall that the length of W_N varies by day, as it depends on κ^*, which in turn depends on the volatility value at the beginning of each day, i.e., $\nu(\tau)$ (in scenarios 1 and 2), or its estimate, i.e., $\hat{\nu}(\tau)$ (in scenario 3).

noise-to-signal ratio ζ	δ_N	Δ_N	λ	rel. bias (*Set 1*)	rel. bias (*Set 2*)	rel. bias (*Set 3*)
$\zeta = 0$	1 min.	δ_N (1 min.)	$2 \cdot 10^{-4}$	0.003	0.004	0.032
		$2\delta_N$ (2 min.)	$4 \cdot 10^{-4}$	0.006	0.006	0.033
		$3\delta_N$ (3 min.)	$6 \cdot 10^{-4}$	0.008	0.009	0.034
		$5\delta_N$ (5 min.)	$1 \cdot 10^{-3}$	0.011	0.013	0.036
		$10\delta_N$ (10 min.)	$1.9 \cdot 10^{-3}$	0.021	0.025	0.041
		$15\delta_N$ (15 min.)	$2.9 \cdot 10^{-3}$	0.031	0.037	0.047
$\zeta = 0$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.024	0.024	0.060
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.029	0.029	0.061
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.031	0.033	0.061
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.046	0.049	0.063

Tab. 4.1: Scenario 1: daily PSRV relative bias with $\kappa = \kappa^*$, $\zeta = 0$, γ known and $\nu(\tau)$ observable. Model parameters: $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.2$ (*Set 1*); $\alpha = 0.03$, $\theta = 10$, $\gamma = 0.25$, $\rho = -0.8$, $\nu(0) = 0.03$ (*Set 2*); $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.4$ (*Set 3*).

noise-to-signal ratio ζ	δ_N	Δ_N	λ	rel. bias (*Set 1*)	rel. bias (*Set 2*)	rel. bias (*Set 3*)
$\zeta = 0.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.025	0.024	0.062
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.030	0.029	0.062
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.032	0.036	0.064
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.047	0.052	0.065
$\zeta = 1.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.039	0.037	0.075
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.044	0.043	0.076
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.046	0.049	0.078
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.061	0.065	0.079
$\zeta = 2.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.064	0.064	0.102
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.069	0.070	0.103
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.075	0.075	0.105
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.091	0.091	0.107
$\zeta = 3.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.108	0.105	0.143
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.113	0.111	0.145
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.115	0.117	0.146
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.130	0.132	0.149

Tab. 4.2: Scenario 2: daily PSRV relative bias with $\kappa = \kappa^*$, $\zeta > 0$, γ known and $\nu(\tau)$ observable. Model parameters: $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.2$ (*Set 1*); $\alpha = 0.03$, $\theta = 10$, $\gamma = 0.25$, $\rho = -0.8$, $\nu(0) = 0.03$ (*Set 2*); $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.4$ (*Set 3*).

noise-to-signal ratio ζ	δ_N	Δ_N	λ	rel. bias (*Set 1*)	rel. bias (*Set 2*)	rel. bias (*Set 3*)
$\zeta = 0.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.059	0.011	0.046
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.059	0.011	0.047
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.060	0.013	0.047
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.060	0.017	0.047
$\zeta = 1.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.068	0.022	0.049
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.068	0.023	0.049
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.069	0.024	0.049
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.070	0.027	0.050
$\zeta = 2.5$	5 min.	δ_N (5 min.)	$6 \cdot 10^{-4}$	0.085	0.047	0.053
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.088	0.049	0.053
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.088	0.049	0.054
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.088	0.051	0.054
$\zeta = 3.5$	5 min.	δ_N (5 min)	$6 \cdot 10^{-4}$	0.112	0.083	0.058
		$2\delta_N$ (10 min.)	$1.3 \cdot 10^{-3}$	0.115	0.083	0.058
		$3\delta_N$ (15 min.)	$1.9 \cdot 10^{-3}$	0.117	0.084	0.059
		$6\delta_N$ (30 min.)	$3.8 \cdot 10^{-3}$	0.118	0.088	0.061

Tab. 4.3: Scenario 3: daily PSRV relative bias with $\kappa = \kappa^*$, $\zeta > 0$, γ unknown and $\nu(\tau)$ unobservable. Model parameters: $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.2$ (*Set 1*); $\alpha = 0.03$, $\theta = 10$, $\gamma = 0.25$, $\rho = -0.8$, $\nu(0) = 0.03$ (*Set 2*); $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.4$ (*Set 3*).

Table 4.1 shows that for $\delta_N = 1$ minute and $\Delta_N \leq 3$ minutes, the bias is almost negligible (i.e., less than 1%) when $\nu(0) = \alpha$, while it is slightly larger but still acceptable (i.e., between 3% and 4%) when $\nu(0) = 2\alpha$. This is in line with equation 4.2 in Lemma 4.1, where it is evident that the source of bias B_N is eliminated when $\nu(0) = \alpha$, which implies $E[\nu(\tau)] = \alpha$. With a price sampling frequency of five minutes, the bias is still acceptable, around 6% at worst. Additionally, Table 4.2 shows that in the presence of noise, price sampling at five-minute intervals to avoid microstructure frictions represents an acceptable compromise, as the bias is less than 15% even in the presence of very intense microstructure effects. Finally, Table 4.3 shows that the statistical error related to the estimation of γ and $\nu(\tau)$ could actually partially compensate for the bias due to the presence of noise, especially when the common assumption $\nu(0) = \alpha$ is violated.

4.5.2 Numerical results in the more general CKLS setting

We conclude this section by testing the efficacy of the generalized, conjecture-based, criterion for the bias-optimal selection of κ under Assumption 4.3, i.e., under the assumption that the volatility evolves as a CKLS model. In this case, the feasible version of the bias-optimal rule to select κ is given by $\hat{\kappa}^{**} = \frac{2\hat{\nu}(\tau)^{1-\beta}}{\hat{\gamma}}$, for $b = -1/2, c < 1/2$.

To test the efficacy of this criterion, we repeat the numerical exercise previously performed in scenario 1 under Assumption 4.1, considering three different values of β: $\beta = 1/2$, corresponding to the model by Heston (1993), which differs from the model of Assumption 4.1 only in the presence of a price drift; $\beta = 1$, corresponding to the continuous-time GARCH model by Nelson (1990); and $\beta = 3/2$, corresponding to the 3/2 model by Platen (1997). For all parameter sets, μ is constant and equal to 0.05. Tables 4,5 and 6 show that our general criterion for the bias-optimal selection of κ under Assumption 4.3 is effective, as it gives satisfactory results in terms of relative bias. Note that the case $\beta = 1/2$ is of interest only in that it confirms that the criterion for the bias-optimal selection of κ derived analytically under Assumption 4.1, i.e., $\kappa = \kappa^*$, is also effective in the presence of a price drift.

4.6 Empirical study

We conclude this chapter with an empirical analysis, where we apply the bias-optimal criterion for selecting κ in equation (4.11) to compute daily PSRV estimates. The dataset is composed of two 1-year samples of S&P 500 1-minute prices relative to the years 2016 and 2017, respectively. The two samples are analyzed

Model	δ_N	Δ_N	λ	rel. bias (*Set 1*)	rel. bias (*Set 2*)	rel. bias (*Set 3*)
$\beta = \frac{1}{2}$	1 min.	δ_N (1 min.)	$2 \cdot 10^{-4}$	0.014	0.012	0.027
		$2\delta_N$ (2 min.)	$4 \cdot 10^{-4}$	0.017	0.015	0.029
		$3\delta_N$ (3 min.)	$6 \cdot 10^{-4}$	0.020	0.016	0.029
		$5\delta_N$ (5 min.)	$1 \cdot 10^{-3}$	0.024	0.022	0.031
		$10\delta_N$ (10 min.)	$1.9 \cdot 10^{-3}$	0.034	0.033	0.036
		$15\delta_N$ (15 min.)	$2.9 \cdot 10^{-3}$	0.042	0.044	0.039

Tab. 4.4: $\beta = 1/2$: daily PSRV finite-sample relative bias with $\kappa = \kappa^{**}$, $\zeta = 0$, γ known and $\nu(\tau)$ observable. Model parameters: $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.2$ (*Set 1*); $\alpha = 0.03$, $\theta = 10$, $\gamma = 0.25$, $\rho = -0.8$, $\nu(0) = 0.03$ (*Set 2*); $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.4$ (*Set 3*). The price drift, μ, is constant and equal to 0.05.

Model	δ_N	Δ_N	λ	rel. bias (*Set 1*)	rel. bias (*Set 2*)	rel. bias (*Set 3*)
$\beta = 1$	1 min.	δ_N (1 min.)	$2 \cdot 10^{-4}$	0.003	0.002	0.011
		$2\delta_N$ (2 min.)	$4 \cdot 10^{-4}$	0.004	0.002	0.012
		$3\delta_N$ (3 min.)	$6 \cdot 10^{-4}$	0.005	0.002	0.014
		$5\delta_N$ (5 min.)	$1 \cdot 10^{-3}$	0.006	0.003	0.015
		$10\delta_N$ (10 min.)	$1.9 \cdot 10^{-3}$	0.008	0.005	0.017
		$15\delta_N$ (15 min.)	$2.9 \cdot 10^{-3}$	0.012	0.006	0.021

Tab. 4.5: $\beta = 1$: daily PSRV finite-sample relative bias with $\kappa = \kappa^{**}$, γ known and $\nu(\tau)$ observable. Model parameters: $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.2$ (*Set 1*); $\alpha = 0.03$, $\theta = 10$, $\gamma = 0.25$, $\rho = -0.8$, $\nu(0) = 0.03$ (*Set 2*); $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.4$ (*Set 3*). The price drift, μ, is constant and equal to 0.05.

Model	δ_N	Δ_N	λ	rel. bias (*Set 1*)	rel. bias (*Set 2*)	rel. bias (*Set 3*)
$\beta = \frac{3}{2}$	1 min.	δ_N (1 min.)	$2 \cdot 10^{-4}$	0.004	0.001	0.029
		$2\delta_N$ (2 min.)	$4 \cdot 10^{-4}$	0.004	0.002	0.031
		$3\delta_N$ (3 min.)	$6 \cdot 10^{-4}$	0.005	0.003	0.031
		$5\delta_N$ (5 min.)	$1 \cdot 10^{-3}$	0.006	0.006	0.037
		$10\delta_N$ (10 min.)	$1.9 \cdot 10^{-3}$	0.006	0.007	0.038
		$15\delta_N$ (15 min.)	$2.9 \cdot 10^{-3}$	0.008	0.009	0.041

Tab. 4.6: $\beta = 3/2$: daily PSRV finite-sample relative bias with $\kappa = \kappa^{**}$, γ known and $\nu(\tau)$ observable. Model parameters: $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.2$ (*Set 1*); $\alpha = 0.03$, $\theta = 10$, $\gamma = 0.25$, $\rho = -0.8$, $\nu(0) = 0.03$ (*Set 2*); $\alpha = 0.2$, $\theta = 5$, $\gamma = 0.5$, $\rho = -0.2$, $\nu(0) = 0.4$ (*Set 3*). The price drift, μ, is constant and equal to 0.05.

separately since the volatility of these two time series behaves very differently. In fact, the year 2016 is characterized by volatility spikes (due, e.g., to uncertainty pertaining to the so-called Brexit in the month of June or the U.S. presidential election in the month of November), while the year 2017 is characterized by low volatility, as one can see in Figure 4.6. Analyzing the two series separately allows for validation of the feasible rule for the selection of κ in two very different scenarios.

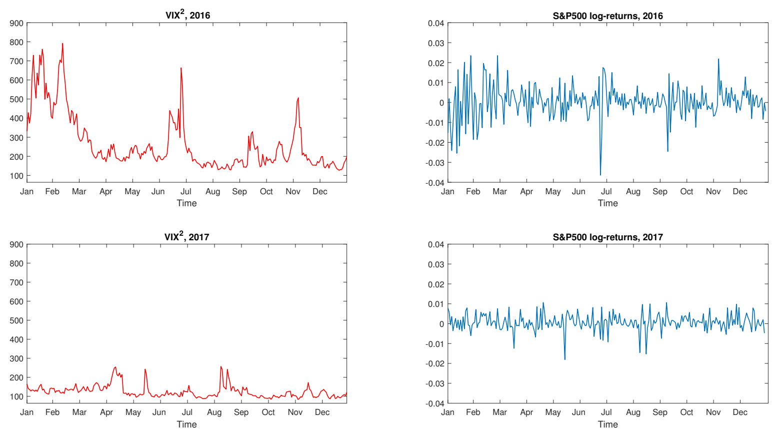

Fig. 4.6: Daily VIX^2 values (left) and daily S&P 500 log-returns (right) in the years 2016 and 2017.

We proceed as follows. First, through the method detailed in the Appendix, we obtain non-parametric Fourier estimates of the process v at the beginning of each day and estimates of γ under Assumption 4.3, for the three different values of β considered in the numerical exercise of Section 4.5. The results of the estimation of γ are shown in Table 4.7[4].

Then, based on R^2 values, we assume the Heston model ($\beta = 1/2$) as the data generating process for both samples. Consequently, we select $b = -1/2$, $c = 1/4$, $\kappa = 2\hat{\gamma}^{-1}\sqrt{\hat{v}(\tau)}$ and compute daily PSRV values from empirical prices sampled at the frequency $\delta_N = 5$ minutes. The resulting selection of W_N is approximately equal, on average, to 450 minutes in 2016 and 275 minutes in 2018. Note that the selection $\delta_N = 5$ minutes is justified by the fact that we assume the impact of microstructure contaminations to be negligible at that sampling frequency, based on the application of the Hausman test by Aït-Sahalia and Xiu (2019) for the presence

[4]The estimates of the process v at the beginning of each day are not reported for the sake of brevity. See Chapter 4 in Mancino et al. (2017) for a detailed study which demonstrates the finite-sample accuracy of the Fourier estimator of the spot volatility.

Model	Sample year	$\hat{\gamma}$	R^2
$\beta = \frac{1}{2}$	2016	0.7127	0.1383
	2017	0.4250	0.1736
$\beta = 1$	2016	6.1682	0.0725
	2017	5.9973	0.0901
$\beta = \frac{3}{2}$	2016	72.0358	0.1141
	2017	65.1084	0.0866

Tab. 4.7: Results of the estimation of γ under Assumption 4.3 for different values of β.

of noise, which tells that the impact of noise at the 5-minute frequency is negligible in our samples, confirming a well-known stylized fact (see Andersen et al. (2001a)). Note that we have also performed the jump-detection test by Corsi et al. (2010) on 5-minute returns (as the test is not robust to the presence of noise contaminations in the price process) and, based on its results, we have removed from the samples the days in which price jumps are detected. These days amount to 12.25% of the sample in 2016 and 8.30% of the sample in 2017[5].

The following figures show the PSRV values obtained for four different values of λ corresponding to a spot volatility estimation frequency Δ_N equal to 5, 10, 15, and 30 minutes, respectively.

Comparing the dynamics of the VIX^2 index in Figure 4.6 with those of the PSRV, one notices that when the VIX^2 spikes, the vol-of-vol also spikes (see, e.g., the behavior of the plots at the end of June 2016) and, viceversa, when the VIX^2 is low and stable (e.g., in 2017) the vol-of-vol is also low and stable. This evidence corroborates the goodness of our vol-of-vol estimates. Finally, note that for either of the two samples, the plots for different values of Δ_N are basically indistinguishable. With respect to the bias-optimal selection of λ (i.e., Δ_N), this evidence confirms what emerges from the analytical study in Section 4.3: the impact of the selection of λ (i.e., Δ_N) on PSRV values is marginal, if not negligible.

[5]Note that the analytical results in Section 4.3 are derived under the assumption of absence of jumps in the price and volatility. The literature on non-parametric jump tests provides large and robust empirical evidence, mainly based on US markets, that volatility jumps are accompanied by price jumps, consistent with the presence of a leverage effect (see, e.g., Jacod and Todorov (2010); Bandi and Renò (2016); Bibinger and Winkelmann (2018)). Thus removing days with price jumps basically also takes care of jumps in the volatility. After eliminating days with price jumps, the assumption of a model in the CKLS class for the volatility could provide a reasonable trade-off between accuracy in reproducing empirical features of prices and parsimony in terms of parameters to be estimated, as pointed out, e.g., in Christoffersen et al. (2010) and Goard and Mazur (2013).

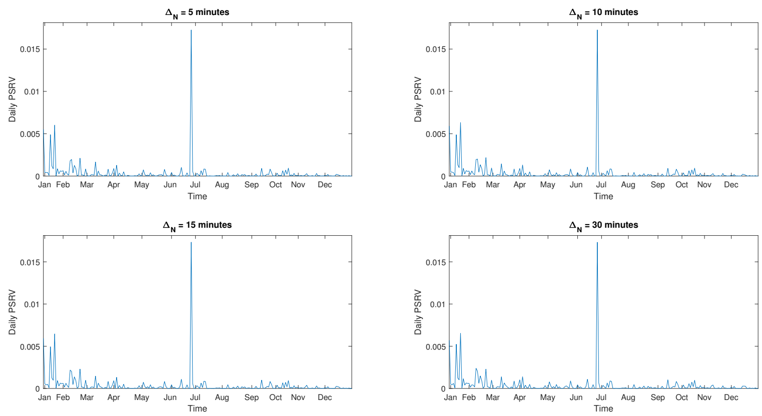

Fig. 4.7: Daily PSRV values in the year 2016.

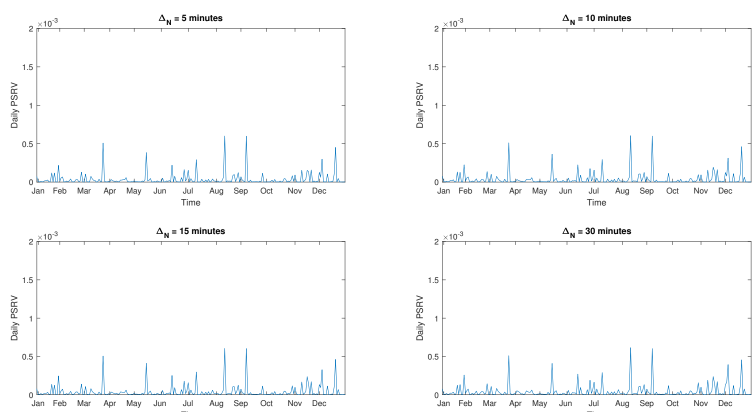

Fig. 4.8: Daily PSRV values in the year 2017.

4.7 Conclusions

The Pre-estimated Spot-variance based Realized Variance (PSRV) by Barndorff-Nielsen and Veraart (2009), the simplest and most natural consistent estimator of the integrated vol-of-vol, is typically affected by a substantial finite-sample bias. The main contribution of this chapter is to show, analytically, that local-window overlapping in finite samples effectively reduces this bias. This result confirms the findings of Sanfelici et al. (2015), based on simulations.

The chapter is written in the spirit of Aït-Sahalia et al. (2013). In Aït-Sahalia et al. (2013), a parametric data-generating process, namely the Heston model, is used to obtain a fully explicit bias expression for the price-volatility correlation, the most natural leverage estimator, which is very biased at high frequencies. Based on the full explicit knowledge of the bias, the authors are able to isolate the sources of bias that affect the simple leverage estimator and derive a feasible strategy to correct for them. In this chapter we follow a similar approach. Assuming that the volatility is a CIR process, we obtain the full explicit expression of the PSRV finite-sample bias. Crucially, we show that this expression differs in the overlapping case and the no-overlapping case and, most importantly, that a feasible bias-correction strategy for finite samples can be derived only in the overlapping case.

Further, using dimensional analysis, we generalize the feasible bias-correction strategy to hold under the assumption that the volatility process belongs to the more general CKLS class, which encompasses a number of widely-used parametric models. Numerical results corroborate the validity of the generalized rule in that nearly unbiased vol-of-vol estimates are obtained for two other models in the CKLS class, namely, the continuous-time GARCH model and the 3/2 model.

In the paper, the impact of microstructure noise on the PSRV bias is also investigated. First, we derive the exact analytical expression of the extra bias due to noise, which differs in the overlapping and no-overlapping cases. This extra bias can not be consistently estimated over a fixed time horizon and then subtracted, as it depends, other than on some moments of the noise process, on the drift of the volatility. As a solution, we propose to apply the feasible rule for the bias-optimal selection of the local-window parameter on sparsely-sampled prices, following Andersen et al. (2001a). Numerical evidence of the efficacy of this solution is provided.

We highlight that the analytical approach used in this chapter to study the PSRV finite-sample bias could be applied to analyze the finite-sample performance of other estimators of second-order quantities which require the pre-estimation of the spot volatility (e.g., estimators of the integral of the stochastic leverage, see Chapter 8 in Aït-Sahalia and Jacod (2014)).

Finally, as a byproduct of this analysis, we quantify, for both the PSRV and the locally averaged realized variance, the bias reduction ensuing from the assumption that the initial value of the volatility is equal to its long-term mean, which is very common in simulation studies found in the literature.

4.8 Appendix

4.8.1 Proofs

Proof of Lemma 4.1

From Definition 4.2 we have

$$PSRV_{[\tau,\tau+h],N} := \sum_{i=1}^{\lfloor h/\Delta_N \rfloor} \Big[\hat{\nu}(\tau+i\Delta_N) - \hat{\nu}(\tau+i\Delta_N-\Delta_N)\Big]^2,$$

where, for s taking values on the time grid of mesh-size δ_N:

- $\hat{\nu}(s) := RV(s,k_N\delta_N)(k_N\delta_N)^{-1}$,
- $RV(s,k_N\delta_N) := \sum_{j=1}^{k_N} \Delta x^2(s+j\delta_N-k_N\delta_N,\delta_N)$,
- $\Delta x(s,\delta_N) := x(s)-x(s-\delta_N)$.

Note that $E[PSRV_{[\tau,\tau+h],N}]$ can be rewritten as

$$E[PSRV_{[\tau,\tau+h],N}] = (k_N\delta_N)^{-2} \sum_{i=1}^{\lfloor h/\Delta_N \rfloor} E[RV^2(\tau+i\Delta_N,k_N\delta_N)] + E[RV^2(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]$$
$$-2E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]. \tag{4.12}$$

Therefore, under Assumption 4.1, the explicit formula for $E[PSRV_{[\tau,\tau+h],N}]$ can be obtained by deriving the analytical expression for $E[RV^2(\tau+i\Delta_N,k_N\delta_N)]$, $E[RV^2(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]$ and $E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]$. Note that the expression of the last term differs in the no-overlapping case $W_N \leq \Delta_N$ and the overlapping case $W_N > \Delta_N$.
We derive the exact expression of these terms separately as follows.

I) Analytical expression of $E[RV^2(\tau+i\Delta_N,k_N\delta_N)]$

To simplify the notation, let $a_{i,u,N}$ denote the quantity $a_{i,u,N} = \tau+i\Delta_N+(u-k_N)\delta_N$. Also, let $(\mathscr{F}_s^\nu)_{s\geq 0}$ be the natural filtration associated with the process ν. We then have:

$$E[RV^2(\tau+i\Delta_N,k_N\delta_N)]=\sum_{j=1}^{k_N}E[\Delta x^4(\tau+i\Delta_N+(j-k_N)\delta_N,\delta_N)]$$
$$+2\sum_{j=2}^{k_N}E[\Delta x^2(\tau+i\Delta_N+(j-k_N)\delta_N,\delta_N)\sum_{h=1}^{j-1}\Delta x^2(\tau+i\Delta_N+(h-k_N)\delta_N,\delta_N)]$$

$$=\sum_{j=1}^{k_N}E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^4\Big]+2\sum_{j=2}^{k_N}\sum_{h=1}^{j-1}E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^2\times\Big(\int_{a_{i,h-1,N}}^{a_{i,h,N}}\sqrt{v(s)}dW(s)\Big)^2\Big],$$

where:

- $\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)|\mathscr{F}^v_{a_{i,j,N}}\sim\mathscr{N}\Big(0,\int_{a_{i,j-1,N}}^{a_{i,j,N}}v(s)ds\Big)$ (see equation (2) in Section 2.1 of Andersen et al. (2001a)), which implies

$$E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^4\Big]=E\Big[E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^4|\mathscr{F}^v_{a_{i,j,N}}\Big]\Big]=3E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}v(s)ds\Big)^2\Big];$$

- for $h<j$ and $s<r$,

$$E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^2\Big(\int_{a_{i,h-1,N}}^{a_{i,h,N}}\sqrt{v(s)}dW(s)\Big)^2\Big]$$
$$=E\Big[E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^2\Big(\int_{a_{i,h-1,N}}^{a_{i,h,N}}\sqrt{v(s)}dW(s)\Big)^2|\mathscr{F}^v_{a_{i,j,N}}\Big]\Big]$$
$$=\int_{a_{i,j-1,N}}^{a_{i,j,N}}\int_{a_{i,h-1,N}}^{a_{i,h,N}}E[v(s)E[v(r)|\mathscr{F}^v_s]]dsdr.$$

Under Assumption 4.1 (see Appendix A in Bollerslev and Zhou (2002)), we also have:

- $E\Big[\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}v(s)ds\Big)^2\Big]=\frac{1}{\theta^2}(1-e^{-\theta\delta_N})^2\Big\{e^{-2\theta i\Delta_N-2\theta j\delta_N+2\theta(1+k_N)\delta_N}E[\nu(\tau)]^2$
$+\Big[\frac{\gamma^2}{\theta}(e^{-\theta i\Delta_N-\theta j\delta_N+\theta(k_N+1)\delta_N}-e^{-2\theta i\Delta_N-2\theta j\delta_N+2\theta(k_N+1)\delta_N})$
$+2\alpha e^{-\theta i\Delta_N-\theta j\delta_N+\theta(k_N+1)\delta_N}(1-e^{-\theta i\Delta_N-\theta j\delta_N+\theta(k_N+1)\delta_N})\Big]E[\nu(\tau)]$
$+\Big(\frac{\gamma^2\alpha}{2\theta}+\alpha^2\Big)(1+e^{-2\theta i\Delta_N-2\theta j\delta_N+2\theta(k_N+1)\delta_N}-2e^{-\theta_i\Delta_N-\theta j\delta_N+\theta(k_N+1)\delta_N})\Big\}$
$+\frac{\gamma^2}{\theta^2}\Big(\frac{1}{\theta}-2\delta_N e^{-\theta\delta_N}-\frac{1}{\theta}e^{-2\theta\delta_N}\Big)+2\frac{1}{\theta}(1-e^{-\theta\delta_N})\Big[\alpha\delta_N-\frac{\alpha}{\theta}(1-e^{-\theta\delta_N})\Big]$
$\times\Big[e^{-\theta i\Delta_N-\theta j\delta_N+\theta(k_N+1)\delta_N}E[\nu(\tau)]+\alpha(1-e^{-\theta i\Delta_N-\theta j\delta_N+\theta(k_N+1)\delta_N})\Big]$
$+\frac{\gamma^2}{\theta^2}\Big[\alpha\delta_N(1+2e^{-\theta\delta_N})+\frac{\alpha}{2\theta}(e^{-2\theta\delta_N}+4e^{-\theta\delta_N}-5)\Big]+\alpha^2\delta_N^2+\frac{\alpha^2}{\theta^2}(1-e^{-\theta\delta_N})^2-$
$2\frac{\alpha^2}{\theta}\delta_N(1-e^{-\theta\delta_N});$

- for $h<j$ and $s<r$,

$$\int_{a_{i,j-1,N}}^{a_{i,j,N}}\int_{a_{i,h-1,N}}^{a_{i,h,N}} E[\nu(s)E[\nu(r)|\mathscr{F}_s^\nu]]dsdr$$
$$= \left[\left(E[\nu(\tau)]-\alpha\right)^2+\frac{\gamma^2}{\theta}\left(\frac{\alpha}{2}-E[\nu(\tau)]\right)\right]\frac{1}{\theta^2}\cdot e^{-2\theta i\Delta_N-\theta j\delta_N-\theta h\delta_N+2\theta k_N\delta_N}(1-e^{\theta\delta_N})^2$$
$$-\frac{\gamma^2\alpha}{2\theta^3}e^{-\theta j\delta_N+\theta h\delta_N}(2-e^{-\theta\delta_N}-e^{\theta\delta_N})+\alpha^2\delta_N^2-\frac{\gamma^2}{\alpha\theta}(E[\nu(\tau)]-\alpha)\delta_N(1-e^{\theta\delta_N})e^{-\theta i\Delta_N-\theta j\delta_N+\theta k_N\delta_N}$$
$$-\frac{\alpha}{\theta}(E[\nu(\tau)]-\alpha)\delta_N(1-e^{\theta\delta_N})e^{-\theta i\Delta_N-\theta h\delta_N+\theta k_N\delta_N}.$$

Finally, putting everything together, we obtain the following expression for $E[RV^2(\tau+i\Delta_N,k_N\delta_N)]$:

$$E[RV^2(\tau+i\Delta_N,k_N\delta_N)]$$
$$=(1-e^{-2\theta k_N\delta_N})(1-e^{-2\theta\delta_N})^{-1}e^{-2i\Delta_N+2\theta k_N\delta_N}(1-e^{-\theta\delta_N})^2\frac{3}{\theta^2}\left[(E[\nu(\tau)]-\alpha)^2+\frac{\gamma^2}{\theta}\left(\frac{\alpha}{2}-E[\nu(\tau)]\right)\right]$$
$$+(1-e^{-\theta k_N\delta_N})(1-e^{-\theta\delta_N})^{-1}e^{-i\Delta_N+\theta k_N\delta_N}\left\{\frac{\gamma^2}{\theta}(E[\nu(\tau)]-\alpha)\frac{3}{\theta^2}(1-e^{-\theta\delta_N})^2\right.$$
$$\left.+\frac{\gamma^2}{\theta}\left(\frac{1}{\theta}-2e^{-\theta\delta_N}\delta_N-\frac{1}{\theta}e^{-2\theta\delta_N}\right)\frac{3}{\theta}(E[\nu(\tau)]-\alpha)+\left[6\frac{\alpha}{\theta}\delta_N(1-e^{-\theta\delta_N})\right](E[\nu(\tau)]-\alpha)\right\}$$
$$+\frac{\gamma^2}{\theta}k_N\left[\frac{3\alpha}{2\theta^2}(1-e^{-\theta\delta_N})^2+3\frac{\alpha}{\theta}\left(\frac{1}{\theta}-2e^{-\theta\delta_N}\delta_N-\frac{1}{\theta}e^{-2\theta\delta_N}\right)+3\frac{\alpha}{\theta}\delta_N(1+2e^{-\theta\delta_N})\right.$$
$$\left.+\frac{3\alpha}{2\theta^2}(e^{-2\theta\delta_N}+4e^{-\theta\delta_N}-5)\right]+3\alpha^2\delta_N^2k_N$$
$$+2\left[\frac{\gamma^2}{\theta}\left(\frac{\alpha}{2}-E[\nu(\tau)]\right)+(E[\nu(\tau)]-\alpha)^2\right]\frac{1}{\theta^2}e^{2\theta k_N\delta_N-2\theta i\Delta_N-\theta\delta_N}$$
$$\times(1-e^{-\theta(k_N-1)\delta_N}+e^{-\theta(k_N+i)\delta_N}-e^{-\theta\delta_N}+e^{-\theta(2k_N-1)\delta_N}-e^{-2\theta k_N\delta_N})(1-e^{-2\theta\delta_N})^{-1}$$
$$+2\alpha(E[\nu(\tau)]-\alpha)\frac{1}{\theta}\delta_Ne^{\theta k_N\delta_N-\theta i\Delta_N}(e^{\theta\delta_N}-1)^{-1}$$
$$\times[e^{-\theta k_N\delta_N}(e^{\theta k_N\delta_N}-1+k_N-k_Ne^{\theta\delta_N})+k_N(e^{\theta\delta_N}-1)+e^{\theta\delta_N}(e^{-\theta k_N\delta_N}-1)]$$
$$+2\frac{\gamma^2}{\theta}(E[\nu(\tau)]-\alpha)\frac{1}{\theta}\delta_Ne^{-\theta i\Delta_N}(e^{\theta k_N\delta_N}-1+k_N-k_Ne^{\theta\delta_N})(e^{\theta\delta_N}-1)^{-1}$$
$$+\gamma^2\alpha\frac{1}{\theta^3}(e^{-\theta k_N\delta_N}-1+k_N-k_Ne^{-\theta\delta_N})+\alpha^2\delta_N^2(k_N^2-k_N).$$

II) Analytical expression of $E[RV^2(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]$

The analytical expression of $E[RV^2(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]$ under Assumption 4.1 is easily obtained by replacing i with $i-1$ in the explicit expression of $E[RV^2(\tau+i\Delta_N,k_N\delta_N)]$ derived in **I)**.

IIIa) Analytical expression of $E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+i\Delta_N-\Delta_N\Delta_N,k_N\delta_N)]$ for $W_N\le\Delta_N$

Assume that we are in the no-overlapping case $W_N\le\Delta_N$. Then

$$
\begin{aligned}
&E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+(i-1)\Delta_N,k_N\delta_N)] \\
&= E\Big[E\Big[\sum_{j=1}^{k_N}\Big(\int_{a_{i,j-1,N}}^{a_{i,j,N}}\sqrt{v(s)}dW(s)\Big)^2\sum_{j=1}^{k_N}\Big(\int_{a_{i-1,j-1,N}}^{a_{i-1,j,N}}\sqrt{v(s)}dW(s)\Big)^2|\mathscr{F}^v_{a_{i,k_N,N}}\Big]\Big] \\
&= E\Big[\sum_{j=1}^{k_N}\int_{a_{i,j-1,N}}^{a_{i,j,N}}v(s)ds\sum_{j=1}^{k_N}\int_{a_{i-1,j-1,N}}^{a_{i-1,j,N}}v(s)ds\Big] = \int_{a_{i,0,N}}^{a_{i,k_N,N}}\int_{a_{i-1,0,N}}^{a_{i-1,k_N,N}}E[v(s)E[v(r)|\mathscr{F}^v_s]]dsdr, s<r.
\end{aligned}
$$

Under Assumption 4.1 (see, again, Appendix A in Bollerslev and Zhou (2002)),

$$
\begin{aligned}
&E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+(i-1)\Delta_N,k_N\delta_N)] \\
&= \frac{1}{\theta^2}e^{\theta\Delta_N}\left(1-e^{\theta k_N\delta_N}\right)^2 e^{-2\theta i\Delta_N}\left[(E[\nu(\tau)]-\alpha)^2+\frac{\gamma^2}{\theta}\left(\frac{\alpha}{2}-E[\nu(\tau)]\right)\right] \\
&-e^{-\theta\Delta_N}\left(2-e^{\theta k_N\delta_N}-e^{-\theta_{k_N}\delta_N}\right)\frac{\gamma^2\alpha}{2\theta^3}-\frac{1}{\theta}k_N\delta_N\left(1-e^{\theta k_N\delta_N}\right)e^{-\theta i\Delta_N}\left[\left(\frac{\gamma^2}{\theta}+\alpha\right)(E[\nu(\tau)]-\alpha)\right] \\
&-\frac{1}{\theta}k_N\delta_N\left(1-e^{\theta k_N\delta_N}\right)e^{\theta\Delta_N-\theta i\Delta_N}[\alpha(E[\nu(\tau)]-\alpha)]+\alpha^2\left(k_N\delta_N\right)^2.
\end{aligned}
$$

IIIb) Analytical expression of $E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+i\Delta_N-\Delta_N\Delta_N,k_N\delta_N)]$ for $W_N>\Delta_N$

Assume now that we are in the overlapping case $W_N>\Delta_N$. Then the parametric expression of $E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+i\Delta_N-\Delta_N,k_N\delta_N)]$ can be decomposed $E[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+(i-1)\Delta_N,k_N\delta_N)]$ into the sum of four components, that is,

$$
\begin{aligned}
&E\Big[RV(\tau+i\Delta_N,k_N\delta_N)RV(\tau+(i-1)\Delta_N,k_N\delta_N)\Big] \\
&= E\Big[\Big(RV(\tau+i\Delta_N,\Delta_N)+RV(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)\Big)\Big(RV(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)+RV(\tau+i\Delta_N-k_N\delta_N,\Delta_N)\Big)\Big] \\
&= E\Big[RV(\tau+i\Delta_N,\Delta_N)RV(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)\Big]+E\Big[(RV(\tau+i\Delta_N,\Delta_N)RV(\tau+i\Delta_N-k_N\delta_N,\Delta_N)\Big]+E\Big[RV^2(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)\Big] \\
&+E\Big[RV(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)RV(\tau+i\Delta_N-k_N\delta_N,\Delta_N)\Big].
\end{aligned}
$$

We then obtain the parametric expressions of these four components, which we term O_1, O_2, O_3 and O_4, respectively (we omit the intermediate steps, as are they are analogous to those followed in **I)** and **IIIa)**):

- $O_1 := E\Big[(RV(\tau+i\Delta_N,\Delta_N)RV(\tau+i\Delta_N-k_N\delta_N,\Delta_N)\Big]$
$= \alpha^2\Delta_N^2 - (E[\nu(\tau)]-\alpha)\left(\frac{\gamma^2}{\theta}+\alpha\right)\Delta_N\frac{1}{\theta}e^{-\theta i\Delta_N}\left(1-e^{\theta\Delta_N}\right)$
$-\alpha\,(E[\nu(\tau)]-\alpha)\,\Delta_N\frac{1}{\theta}e^{-\theta i\Delta_N+\theta k_N\delta_N}\left(1-e^{\theta\Delta_N}\right)-\dfrac{\gamma^2\alpha}{2\theta^3}e^{-\theta k_N\delta_N}(2-e^{-\theta\Delta_N}-e^{\theta\Delta_N})$
$+\left[\dfrac{\gamma^2}{\theta}\left(\dfrac{\alpha}{2}-E[\nu(\tau)]\right)+(E[\nu(\tau)]-\alpha)^2\right]\dfrac{1}{\theta^2}e^{-2\theta i\Delta_N}(1-e^{\theta\Delta_N})^2e^{\theta k_N\delta_N}$;

- $O_2 := E\Big[RV(\tau+i\Delta_N,\Delta_N)RV(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)\Big]$
$= \alpha^2\Delta_N(k_N\delta_N-\Delta_N)+(E[\nu(\tau)]-\alpha)\left(\frac{\gamma^2}{\theta}+\alpha\right)(k_N\delta_N-\Delta_N)\frac{1}{\theta}e^{-\theta i\Delta_N}(e^{\theta\Delta_N}-1)$
$+\alpha(E[\nu(\tau)]-\alpha)\Delta_N\frac{1}{\theta}e^{-\theta(i-1)\Delta_N}(e^{\theta(k_N\delta_N-\Delta_N)}-1)-\dfrac{\gamma^2\alpha}{2\theta^3}(1-e^{\theta\Delta_N})e^{-\theta k_N\delta_N}(e^{\theta(k_N\delta_N-\Delta_N)}-1)$
$-\left[\dfrac{\gamma^2}{\theta}(\dfrac{\alpha}{2}-E[\nu(\tau)])+(E[\nu(\tau)]-\alpha)^2\right]\dfrac{1}{\theta^2}e^{-2\theta i\Delta_N}(1-e^{\theta\Delta_N})e^{\theta\Delta_N}(e^{\theta(k_N\delta_N-\Delta_N)}-1)$;

- $O_3 := E\Big[RV(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)RV(\tau+i\Delta_N-k_N\delta_N,\Delta_N)\Big]$
$= \alpha^2\Delta_N\left(k_N\delta_N-\Delta_N\right)+(E[\nu(\tau)]-\alpha)\left(\frac{\gamma^2}{\theta}+\alpha\right)\Delta_N\frac{1}{\theta}e^{-\theta(i-1)\Delta_N}(e^{\theta(k_N\delta_N-\Delta_N)}-1)$
$+\alpha(E[\nu(\tau)]-\alpha)(k_N\delta_N-\Delta_N)\frac{1}{\theta}e^{-\theta i\Delta_N+\theta k_N\delta_N}(e^{\theta\Delta_N}-1)$
$+\dfrac{\gamma^2\alpha}{2\theta^3}e^{\theta\Delta_N}(e^{\theta(k_N\delta_N-\Delta_N)}-1)e^{-\theta k_N\delta_N}(1-e^{-\theta\Delta_N})$
$-\left[\dfrac{\gamma^2}{\theta}\left(\dfrac{\alpha}{2}-E[\nu(\tau)]\right)+(E[\nu(\tau)]-\alpha)^2\right]\dfrac{1}{\theta^2}e^{-2\theta i\Delta_N+\theta k_N\delta_N}(1-e^{\theta\Delta_N})e^{\theta\Delta_N}(e^{\theta(k_N\delta_N-\Delta_N)}-1)$;

- $O_4 := E\Big[RV^2(\tau+(i-1)\Delta_N,k_N\delta_N-\Delta_N)\Big]$
$= (k_N\delta_N-\Delta_N)\dfrac{1}{\delta_N}\Big[\dfrac{3\alpha\gamma^2}{4\theta^4}(1-e^{-\theta\delta_N})^2+\dfrac{1}{\theta}-2\delta_Ne^{-\theta\delta_N}-\dfrac{1}{\theta}e^{-2\theta\delta_N}+\delta_N(1+2e^{-\theta\delta_N})$
$+\dfrac{1}{2\theta}(e^{-2\theta\delta_N}+4e^{-\theta\delta_N}-5)+3\alpha^2\delta_N^2\Big]$
$+e^{-2\theta\delta_N}(1-e^{-2\theta(k_N\delta_N-\Delta_N)})(1-e^{-2\theta\delta_N})^{-1}\Big[3\dfrac{1}{\theta^2}(1-e^{-\theta\delta_N})^2e^{-2\theta i\Delta_N+2\theta(1+k_N)\delta_N}$
$\times\left(E[\nu(\tau)]^2-\dfrac{\gamma^2}{\theta}E[\nu(\tau)]-2\alpha E[\nu(\tau)]+\dfrac{\gamma^2\alpha}{2\theta}+\alpha^2\right)\Big]$
$+\dfrac{3}{\theta^2}e^{-\theta\delta_N}(1-e^{-\theta(k_N\delta_N-\Delta_N)})(1-e^{-\theta\delta_N})^{-1}e^{-\theta i\Delta_N+\theta(1+k_N)\delta_N}\Big[\dfrac{\gamma^2}{\theta}(E[\nu(\tau)]-\alpha)(1-e^{-\theta\delta_N})^2$
$+\gamma^2(E[\nu(\tau)]-\alpha)\left(\dfrac{1}{\theta}-2\delta_Ne^{-\theta\delta_N}-\dfrac{1}{\theta}e^{-2\theta\delta_N}\right)+2\alpha(E[\nu(\tau)]-\alpha)(1-e^{-\theta\delta_N})\theta\delta_N\Big]$

$$
\begin{aligned}
&+\alpha^2(k_N\delta_N-\Delta_N)^2-\alpha^2\delta_N(k_N\delta_N-\Delta_N)\\
&-\frac{2}{\theta^2}e^{-2\theta i\Delta_N+2\theta k_N\delta_N-2\theta\delta_N}(1-e^{\theta\delta_N})(1-e^{-\theta\delta_N})^{-1}(1-e^{-2\theta\delta_N})^{-1}(1-e^{-\theta(k_N\delta_N-\Delta_N)})\\
&\times\Big[\tfrac{\gamma^2}{\theta}(E[\nu(\tau)]-\alpha)^2+\Big(\tfrac{\alpha}{2}-E[\nu(\tau)]\Big)\Big]\\
&\times\Big[-e^{\theta\delta_N}e^{-\theta(k_n\delta_N-\Delta_N)}+\Big(1+e^{-\theta(k_N\delta_N-\Delta_N)}\Big)-e^{-\theta\delta_N}\Big]\\
&+2\frac{\gamma^2\alpha}{2\theta^3}\Big((e^{-\theta(k_N\delta_N-\Delta_N)}-1)+(k_N\delta_N-\Delta_N)\delta_N^{-1}(1-e^{-\theta\delta_N})\Big)\\
&+\frac{2}{\theta}\Big(\frac{\gamma^2}{\theta}+\alpha\Big)(E[\nu(\tau)]-\alpha)\delta_N e^{-\theta i\Delta_N+\theta\Delta_N}(e^{\theta\delta_N}-1)^{-1}\\
&\times\Big((e^{\theta(k_N\delta_N-\Delta_N)}-1)+(k_N\delta_N-\Delta_N)\delta_N^{-1}(1-e^{\theta\delta_N})\Big)\\
&+\frac{2\alpha}{\theta}(E[\nu(\tau)]-\alpha)\delta_N e^{-\theta i\Delta_N+\theta k_N\delta_N}(e^{\theta\delta_N}-1)^{-1}\\
&\times\Big[(k_N\delta_N-\Delta_N)\delta_N^{-1}(e^{\theta\delta_N}-1)+e^{\theta\delta_N}(e^{-\theta(k_N\delta_N-\Delta_N)}-1)\Big].
\end{aligned}
$$

The contribution to the PSRV finite-sample bias due to the overlapping of consecutive local windows to estimate the spot volatility (i.e., due to assuming that $W_N = k_N\delta_N > \Delta_N$) is mainly due to the terms O_2, O_3, and O_4. In fact, when $k_N\delta_N = \Delta_N$ (i.e., $W_N = \Delta_N$), the terms O_2, O_3, and O_4 are equal to zero. Interestingly, the terms O_2, O_3, and O_4 are functions of the quantity $(k_N\delta_N-\Delta_N)$ (i.e., $W_N-\Delta_N$) and, in particular, are $O(k_N\delta_N-\Delta_N)$ as $(k_N\delta_N-\Delta_N)\to 0^+$ as $N\to\infty$, as one can check focusing on the terms $(k_N\delta_N-\Delta_N)$ and $(e^{\theta(k_N\delta_N-\Delta_N)}-1)$.

After plugging the explicit expressions obtained in **I)**, **II)** and **IIIa)** (respectively, **IIIb)**) into equation (4.12), simple but tedious calculations yield the parametric expression of $E[PSRV_{[\tau,\tau+h],N}]$ under Assumption 4.1, which can be expressed in the following compact form:

$$
E[PSRV_{[\tau,\tau+h],N}]=\gamma^2\alpha hA_N+\gamma^2\Big(E[\nu(\tau)]-\alpha\Big)\frac{1-e^{-\theta h}}{h}B_N+C_N \quad if \quad W_N\le\Delta_N,
$$

$$
E[PSRV_{[\tau,\tau+h],N}]=\gamma^2\alpha hA_N+\gamma^2\Big(E[\nu(\tau)]-\alpha\Big)\frac{1-e^{-\theta h}}{h}B_N+C_N+O_N \quad if \quad W_N>\Delta_N,
$$

where:

$$A_N = (k_N\delta_N)^{-2}\Delta_N^{-1}\Big\{\frac{2}{\theta}k_N\Big[\frac{3}{2\theta^2}(1-e^{-\theta\delta_N})^2 + 3\frac{1}{\theta}\Big(\frac{1}{\theta} - 2e^{-\theta\delta_N}\delta_N - \frac{1}{\theta}e^{-2\theta\delta_N}\Big)$$
$$+3\frac{1}{\theta}\delta_N(1+2e^{-\theta\delta_N}) + \frac{3}{2\theta^2}(e^{-2\theta\delta_N}+4e^{-\theta\delta_N}-5)\Big] + \frac{2}{\theta^3}(e^{-\theta k_N\delta_N}-1+k_N-k_Ne^{-\theta\delta_N})$$
$$+\frac{1}{\theta^3}e^{-\theta\Delta_N}(2-e^{\theta k_N\delta_N}-e^{-\theta k_N\delta_N})\Big\}; \tag{4.13}$$

$$B_N = (k_N\delta_N)^{-2}e^{-\theta\Delta_N}(1-e^{-\theta\Delta_N})^{-1}\Big\{(1+e^{\theta\Delta_N})\Big[\frac{3}{\theta^2}(e^{\theta k_N\delta_N}-1)(1-e^{-\theta\delta_N})$$
$$+\frac{3}{\theta}(e^{\theta k_N\delta_N}-1)(1-e^{-\theta\delta_N})^{-1}\Big(\frac{1}{\theta}-2e^{-\theta\delta_N}\delta_N-\frac{1}{\theta}e^{-2\theta\delta_N}\Big)$$
$$+2\frac{1}{\theta}\delta_N(e^{\theta\delta_N}-1)^{-1}(k_N-1+e^{\theta k_N\delta_N}-k_Ne^{\theta\delta_N})\Big] + \frac{2}{\theta}k_N\delta_N(1-e^{\theta k_N\delta_N})\Big\}; \tag{4.14}$$

$$C_N = (k_N\delta_N)^{-2}\Bigg\{e^{-2\theta\Delta_N}(1-e^{-2\theta h})(1-e^{-2\theta\Delta_N})^{-1}\frac{1}{\theta^2}\Big[(E[\nu(\tau)]-\alpha)^2+\frac{\gamma^2}{\theta}\Big(\frac{\alpha}{2}-\nu(\tau)\Big)\Big]$$
$$\times\Big\{(1+e^{2\theta\Delta_N})(1-e^{-2\theta\delta_N})^{-1}\Big[3(e^{2\theta k_N\delta_N}-1)(1-e^{-\theta\delta_N})^2+2(1-e^{-\theta\delta_N})$$
$$+2e^{\theta k_N\delta_N}(e^{-2\theta\delta_N}-1)+2e^{2\theta k_N\delta_N-\theta\delta_N}(1-e^{-\theta\delta_N})\Big]-2e^{\theta\Delta_N}(1-e^{\theta k_N\delta_N})^2\Big\}+(6\alpha^2\delta_N^2k_N-2\alpha^2k_N\delta_N^2)h\Delta_N^{-1}$$
$$+e^{-\theta\Delta_N}(1-e^{-\theta h})(1-e^{-\theta\Delta_N})^{-1}\Big\{\Big[6\frac{\alpha}{\theta}\delta_N(E[\nu(\tau)]-\alpha)(e^{\theta k_N\delta_N}-1)+2\frac{\alpha}{\theta}\delta_N(E[\nu(\tau)]-\alpha)(e^{\theta\delta_N}-1)^{-1}$$
$$\times[(e^{\theta k_N\delta_N}-1+k_N-k_Ne^{\theta\delta_N})+k_Ne^{\theta k_N\delta_N}(e^{\theta\delta_N}-1)+e^{\theta\delta_N}(1-e^{\theta k_N\delta N})]\Big](1+e^{\theta\Delta_N})$$
$$+\frac{2\alpha}{\theta}k_N\delta_N(E[\nu(\tau)]-\alpha)(1+e^{\theta\Delta_N})(1-e^{\theta k_N\delta_N})\Big\}\Bigg\}; \tag{4.15}$$

$$
\begin{aligned}
O_N &= (k_N\delta_N)^{-2}\Bigg\{4\alpha^2 h\delta_N \\
&+\gamma^2\alpha h\theta^{-3}\frac{e^{-\theta(\delta_N+k_N\delta_N+\Delta_N)}}{\Delta_N}\Big[e^{\theta\delta_N}-2e^{\theta\delta_N(1+k_N)}+e^{\theta\delta_N(1+2k_N)}-2e^{\theta(\delta_N+\Delta_N)} \\
&-4e^{\theta(\delta_N k_N+\Delta_N)}k_N+e^{\theta(\delta_N+\delta_N k_N+\Delta_N)}(2+k_N(4-6\theta\delta_N))\Big] \\
&+2\gamma^2\alpha h\theta^{-3}\frac{e^{-\theta\delta_N(1+k_N)}}{\Delta_N}\Big[e^{\theta\delta_N}+2e^{\theta\delta_N k_N}k_N-e^{\theta\delta_N(1+k_N)}(1-k_N(3\theta\delta_N-2))\Big] \\
&-\gamma^2\alpha h\theta^{-3}\frac{e^{-\theta(1+2k_N)\delta_N}}{\delta_N\Delta_N}\Big[-4e^{2\theta\delta_N k_N}(1-e^{\theta\delta_N})\Delta_N+6\theta e^{\theta\delta_N(1+2k_N)}k_N\delta_N^2 \\
&+\delta_N\Big(2e^{\theta(\delta_N+\delta_N k_N-\Delta_N)}-2e^{\theta(\delta_N+2\delta_N k_N-\Delta_N)}+4e^{2\theta\delta_N k_N}k_N-2e^{\theta\delta_N(1+2k_N)}(2k_N+3\theta\Delta_N)\Big)\Big] \\
&-\frac{1}{\theta^2(1-e^{-2\theta\Delta_N})}\Big[(-2e^{-\theta\Delta_N})(1-e^{-2\theta h})(-1+e^{\theta k_N\delta_N})^2 \\
&+\frac{(-3+e^{\theta\delta_N}-e^{\theta k_N\delta_N}+3e^{\theta\delta_N(1+k_N)})(1+e^{2\theta\Delta_N})}{1+e^{\theta\delta_N}}\Big(\frac{\gamma^2(\alpha-2E[\nu(\tau)])}{2\theta}+(\alpha-E[\nu(\tau)])^2\Big)\Big] \\
&-2\theta^{-2}(\alpha-E[\nu(\tau)])(1-e^{-\theta h})\frac{(e^{\theta k_N\delta_N}-1)(\gamma^2+\alpha\theta(1+e^{\theta\Delta_N}k_N))\delta_N}{(-1+e^{\theta\Delta_N})} \\
&+2\theta^{-3}(\alpha-E[\nu(\tau)])\frac{e^{-\theta(\delta_N+k_N\delta_N-\Delta_N)}(-e^{-\theta h}+1)}{(-1+e^{\theta\delta_N})(-1+e^{\theta\Delta_N})}\Big[\alpha\theta^2 e^{2\theta\delta_N(1+k_N)}k_N\delta_N-\alpha\theta^2 e^{\theta\delta_N(1+2k_N)}k_N\delta_N \\
&-\theta(\gamma^2+\alpha\theta)\Big[(e^{\theta\delta_N}-1)e^{\theta(\delta_N+k_N\delta_N-\Delta_N)}\Big]k_N\delta_N-e^{\theta\delta_N(2+k_N)}\Big(\alpha\theta^2(4+k_N)\delta_N+\gamma^2(6+\theta k_N\delta_N-\theta\Delta_N)\Big) \\
&+e^{\theta\delta_N(1+k_N)}\Big(\alpha\theta^2(4+k_N)\delta_N+\gamma^2(6+\theta\delta_N(4+k_N)-\theta\Delta_N)\Big) \\
&+e^{\theta(2\delta_N(1+k_N)-\Delta_N)}\Big(\alpha\theta^2(4+k_N)\delta_N+\gamma^2(6+\theta\Delta_N)\Big) \\
&-e^{\theta(\delta_N+2k_N\delta_N-\Delta_N)}\Big(\alpha\theta^2(4+k_N)\delta_N+\gamma^2(6+4\theta\delta_N+\theta\Delta_N)\Big)\Big] \\
&-\frac{2\alpha^2\theta+\alpha(\gamma^2-4\theta E[\nu(\tau)])+2E[\nu(\tau)](\theta E[\nu(\tau)]-\gamma^2)}{2\theta^3(1+e^{\theta\delta_N})(-1+e^{2\theta\Delta_N})}\Big[(1-e^{-2\theta h})(-1+e^{\theta k_N\delta_N}) \\
&\times(3-e^{\theta\delta_N}+e^{\theta k_N\delta_N}-3e^{\theta(1+k_N)\delta_N})(1+e^{2\theta\Delta_N})\Big] \\
&-\frac{\theta^{-3}(2\alpha^2\theta+\alpha(\gamma^2-4\theta E[\nu(\tau)]))+2E[\nu(\tau)](-\gamma^2+\theta E[\nu(\tau)])(1-e^{-2h\theta})}{(1+e^{\theta\delta_N})(-1+e^{2\theta\Delta_N})} \\
&\times\Big[-2e^{2\theta k_N\delta_N}+2e^{\theta(1+2k_N)\delta_N}+e^{\theta\Delta_N}+2e^{2\theta\Delta_N}+e^{\theta(\delta_N+\Delta_N)} \\
&-2e^{\theta(k_N\delta_N+\Delta_N)}-2e^{\theta(\delta_N+k_N\delta_N+\Delta_N)}+e^{\theta(2k_N\delta_N+\Delta_N)}+e^{\theta(\delta_N+2k_N\delta_N+\Delta_N)}-2e^{\theta(\delta_N+2\Delta_N)}\Big]\Bigg\}.
\end{aligned}
\tag{4.16}
$$

The proof is complete.

Proof of Theorem 4.1

Consider the exact parametric expression for the PSRV bias under Assumption 4.1 in the case $W_N > \Delta_N$, given in Lemma 4.1. By expanding it sequentially, first as $\lambda \to 0$, and then as $h \to 0$, we obtain:

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \begin{cases} \left(\dfrac{4E[\nu(\tau)]^2}{\kappa^2\delta_N^{1+2b}} - \gamma^2 E[\nu(\tau)]\right)h + O(h^{1-b}) + O(\lambda) \quad \text{if} \quad b \geq -1/2, c < -b \\ \\ -\gamma^2 E[\nu(\tau)]h + O(h^{-2b}) + O(\lambda) \quad \text{if} \quad b < -1/2, c < 1+b \end{cases},$$

as $\lambda \to 0$, $h \to 0$.

The sequential expansions as $h \to 0$, $\lambda \to 0$ are performed using the software *Mathematica*. The code is available on request.

Furthermore, let $(\mathscr{F}_t^\nu)_{t\geq 0}$ denote the natural filtration associated with the process ν. It is straightforward to see that

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]} | \mathscr{F}_\tau^\nu\Big] = \begin{cases} \left(\dfrac{4\nu(\tau)^2}{\kappa^2\delta_N^{1+2b}} - \gamma^2 \nu(\tau)\right)h + O(h^{1-b}) + O(\lambda) \quad \text{if} \quad b \geq -1/2, c < -b \\ \\ -\gamma^2 \nu(\tau)h + O(h^{-2b}) + O(\lambda) \quad \text{if} \quad b < -1/2, c < 1+b \end{cases},$$

as $\lambda \to 0$, $h \to 0$.

Proof of Theorem 4.2

Consider the exact parametric expression for the PSRV bias under Assumption 4.1 in the case $W_N \leq \Delta_N$, given in Lemma 4.1. Then recall that for $N \to \infty$, $\Delta_N = O(\delta_N^c)$, $c \in (0,1)$, and $k_N = O(\delta_N^b)$, $b \in (-1,0)$. Moreover, note that for $b \geq -1/2$ and $c < -b$ or $b < -1/2$ and $c < 1+b$, we have

$$\lim_{N\to+\infty} \frac{1}{k_N \Delta_N} = 0$$

and

$$\lim_{N\to+\infty} \frac{k_N \delta_N}{\Delta_N} = 0.$$

Expanding A_N, B_N, and C_N as $N \to \infty$, one obtains

- $A_N \sim 1 + \frac{2}{\theta k_N \Delta_N} + \frac{\theta (k_N\delta_N)^2}{4\Delta_N} - \frac{\theta^2 (k_N\delta_N)^2}{4} - \frac{\theta\Delta_N}{2} + \frac{\theta^3 (k_N\delta_N)^2 \Delta_N}{8}$;

- $B_N \sim 1 + \frac{4}{\theta k_N \Delta_N} + \frac{2\delta_N}{\Delta_N} + \frac{2}{k_N} - \frac{4\delta_N}{k_N\Delta_N} - \frac{k_N\delta_N}{\Delta_N} - \frac{2\theta\delta_N^2}{\Delta_N} + \frac{1}{2}\theta\delta_N - \frac{2\theta\delta_N}{k_N} + \frac{1}{2}\theta\Delta_N + \frac{\theta\Delta_N}{k_N} + \frac{\theta\delta_N^2 k_N}{2\Delta_N} - \frac{\theta^2\delta_N\Delta_N}{k_N} - \theta^2\delta_N^2 + \frac{\theta^2\delta_N\Delta_N}{4} - \frac{\theta^3\delta_N^2\Delta_N}{2}$;
- $C_N \sim \frac{1-e^{-2\theta h}}{2\theta^3}\Big[(E[\nu(\tau)]-\alpha)^2 + \frac{\gamma^2}{\theta}(\frac{\alpha}{2} - E[\nu(\tau)])\Big]\Big[\frac{4\theta^2}{k_N\Delta_N} + \frac{4\theta^3\delta_N}{\Delta_N} + \theta^4\Delta_N + \frac{4\theta^4\Delta_N}{k_N} + 3\theta^5\Delta_N\delta_N\Big] + \frac{4\alpha^2 h}{k_N\Delta_N} + \frac{8\alpha(E[\nu(\tau)]-\alpha)(1-e^{-\theta h})}{\theta k_N\Delta_N}$,

from which we get equation (4.6).

Based on the corresponding asymptotic expansions, one can easily check that as $N\to\infty$, if $b \geq -1/2$ and $c < -b$ or, alternatively, $b < -1/2$ and $c < 1+b$, then $A_N \to 1$, $B_N \to 1$ and $C_N \to 0$. This implies that as $N\to\infty$, if $b \geq -1/2$ and $c < -b$ or, alternatively, $b < -1/2$ and $c < 1+b$, then $E[PSRV_{[\tau,\tau+h],N}] = \gamma^2\alpha h A_N + \gamma^2(E[\nu(\tau)]-\alpha)\frac{1-e^{-\theta h}}{h}B_N + C_N$ converges to $E[\langle\nu,\nu\rangle_{[\tau,\tau+h]}] = \gamma^2\alpha h + \gamma^2(E[\nu(\tau)]-\alpha)\frac{1-e^{-\theta h}}{\theta}$, where the equivalence $E[\langle\nu,\nu\rangle_{[\tau,\tau+h]}] = \gamma^2\alpha h + \gamma^2(E[\nu(\tau)]-\alpha)\frac{1-e^{-\theta h}}{\theta}$ is obtained from Appendix A in Bollerslev and Zhou (2002).

In particular, one can easily verify that, as $N\to\infty$:

- for $b \geq -1/2$ and $c < -b$,

$$A_N - 1 = O(\Delta_N),\ B_N - 1 = O(\Delta_N),\ C_N = O(\Delta_N) \quad \textit{if} \quad c < -b/2, \tag{4.17}$$
$$A_N - 1 = O\Big(\frac{1}{k_N\Delta_N}\Big),\ B_N - 1 = O\Big(\frac{1}{k_N\Delta_N}\Big),\ C_N = O\Big(\frac{1}{k_N\Delta_N}\Big)\ \textit{if} -b/2 \leq c < -b; \tag{4.18}$$

- for $-2/3 \leq b < -1/2$ and $c < 1+b$,

$$A_N - 1 = O(\Delta_N),\ B_N - 1 = O(\Delta_N),\ C_N = O(\Delta_N) \qquad \textit{if} \quad c < (1+b)/2, \tag{4.19}$$
$$A_N - 1 = O(\Delta_N),\ B_N - 1 = O\Big(\frac{k_N\delta_N}{\Delta_N}\Big),\ C_N = O(\Delta_N) \quad \textit{if} \quad (1+b)/2 \leq c < -b/2, \tag{4.20}$$
$$A_N - 1 = O\Big(\frac{1}{k_N\Delta_N}\Big), B_N - 1 = O\Big(\frac{k_N\delta_N}{\Delta_N}\Big), C_N = O\Big(\frac{1}{k_N\Delta_N}\Big)\ \textit{if} -b/2 \leq c < 1+b; \tag{4.21}$$

- for $b < -2/3$ and $c < 1+b$,

$$A_N - 1 = O(\Delta_N),\ B_N - 1 = O(\Delta_N), \quad C_N = O(\Delta_N) \quad \textit{if} \quad c < (1+b)/2, \tag{4.22}$$

$$A_N - 1 = O(\Delta_N),\ B_N - 1 = O\Big(\frac{k_N \delta_N}{\Delta_N}\Big), \quad C_N = O(\Delta_N)\ \textit{if}\,(1+b)/2 \le c < 1+b. \tag{4.23}$$

The proof is complete.

Proof of Corollary 1

Based on equation (4.6) and the asymptotic rates of A_N, B_N and C_N (see equations (4.17) – (4.23)), we observe that:

- for $b \geq -1/2$, $c < -b/2$ or $b < -1/2$, $c < (1+b)/2$ or $b = -2/3$, $c < 1/6$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] = a_1 \lambda \delta_N^c + o(\delta_N^c);$$

- for $b > -1/2, c \in (-b, -b/2)$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] = a_2 \frac{1}{\kappa\lambda} \delta_N^{-b-c} + o(\delta_N^{-b-c});$$

- for $b \in (-2/3, -1/2), c \in ((1+b)/2, 1+b)$ *or* $b < -2/3, c \in ((1+b)/2, 1+b)$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] = a_3 \frac{\kappa}{\lambda} \delta_N^{1+b-c} + o(\delta_N^{1+b-c});$$

- for $b = -2/3$, $1/6 < c < 1/3$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] = a_3 \frac{\kappa}{\lambda} \delta_N^{1/3-c} + o(\delta_N^{1/3-c});$$

- for $b = -1/2$, $c = 1/4$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] = \frac{1}{\lambda} \delta_N^{1/4} (a_1 \lambda^2 + a_2 \kappa^{-1} + a_3 \kappa) + o(\delta_N^{1/4});$$

- for $b = -1/2, c > 1/4$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu, \nu \rangle_{[\tau,\tau+h]}\Big] = \frac{1}{\lambda} \delta_N^{1/2-c} (a_2 \kappa^{-1} + a_3 \kappa) + o(\delta_N^{1/2-c});$$

- for $b = -2/3$, $c = 1/6$,

$$E\Big[PSRV_{[\tau,\tau+h],N} - \langle \nu,\nu\rangle_{[\tau,\tau+h]}\Big] = \delta_N^{1/6}(a_1\lambda + a_3\kappa\lambda^{-1}) + o(\delta_N^{1/6}).$$

Thus, it is possible to select κ and λ such that the dominant term of the bias expansion is canceled only when $b = -1/2$ and $c \geq 1/4$ or $b = -2/3$ and $c = 1/6$, provided that the selected values of κ and λ verify the condition $W_N \leq \Delta_N$, which is equivalent to $\kappa\delta_N^{1+b} \leq \lambda\delta_N^c$.

The case $b = -1/2$ and $c = 1/4$ is of particular interest, as it may allow to cancel the dominant term under the usual assumption $\nu(0) = \alpha$, which is equivalent to $E[\nu(\tau)] = \alpha$. In fact, if $E[\nu(\tau)] = \alpha$, then $a_3 = 0$ and it is not possible to cancel the leading term of the bias expansion through the selection of κ and λ when $b = -1/2$ and $c > 1/4$ or $b = -2/3$ and $c = 1/6$.

Specifically, the leading term of the bias expansion in equation (4.6) can be canceled in the case $b = -1/2$ and $c = 1/4$ if there exists a solution $(\tilde{\kappa},\tilde{\lambda}) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ to the following system

$$\begin{cases} a_3\kappa^2 + a_1\lambda^2\kappa + a_2 = 0 \\ W_N \leq \Delta_N \end{cases},$$

where $W_N = \kappa\delta_N^{1/2}$ and $\Delta_N = \lambda\delta_N^{1/4}$. If a solution $(\tilde{\kappa},\tilde{\lambda}) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ exists, the corresponding bias-optimal selection of W_N and Δ_N reads

$$W_N = \tilde{\kappa}\delta_N^{1/2}, \quad \Delta_N = \tilde{\lambda}\delta_N^{1/4}.$$

Proof of Lemma 4.2

Let Assumption 4.2 hold and consider the estimator:
$$\widetilde{PSRV}_{[\tau,\tau+h],N} := \sum_{i=1}^{\lfloor h/\Delta_N \rfloor} \Big[w(\tau + i\Delta_N) - w(\tau + i\Delta_N - \Delta_N)\Big]^2,$$
where, for s taking values on the time grid of mesh-size δ_N:

- $w(s) := \widetilde{RV}(s, k_N\delta_N)(k_N\delta_N)^{-1}$,
- $\widetilde{RV}(s, k_N\delta_N) := \sum_{j=1}^{k_N} \Delta\tilde{x}^2(s + j\delta_N - k_N\delta_N, \delta_N)$,
- $\Delta\tilde{x}(s, \delta_N) := \tilde{x}(s) - \tilde{x}(s - \delta_N)$.

To simplify the notation, we replace $\Delta x(\tau+i\Delta_N+(j-k_N)\delta_N,\delta_N)$ with $r(i,j,N)$ and $\Delta\varepsilon(\tau+i\Delta_N+(j-k_N)\delta_N,\delta_N)$ with $\omega(i,j,N)$. Then we decompose $E[\widetilde{PSRV}_{[\tau,\tau+h],N}]$ as:

$$E[\widetilde{PSRV}_{[\tau,\tau+h],N}] = E[PSRV_{[\tau,\tau+h],N}] \tag{4.24}$$

$$+(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor} E\left[\sum_{j=1}^{k_N}\Big(\omega^2(i,j,N)-\omega^2(i-1,j,N)\Big)\right]^2 \tag{4.25}$$

$$+4(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor} E\left[\sum_{j=1}^{k_N}\Big(r(i,j,N)\omega(i,j,N)-r(i-1,j,N)\omega(i-1,j,N)\Big)\right]^2 \tag{4.26}$$

$$+2(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor}\left\{E\Big[\sum_{j=1}^{k_N}\Big(r^2(i,j,N)-(r^2(i-1,j,N)\Big)\Big]E\Big[\sum_{j=1}^{k_N}\Big(\omega^2(i,j,N)-\omega^2(i-1,j,N)\Big)\Big]\right\} \tag{4.27}$$

$$+4(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor} E\left[\sum_{j=1}^{k_N}\Big(r^2(i,j,N)-r^2(i-1,j,N)\Big)\sum_{j=1}^{k_N}\Big(r(i,j,N)\omega(i,j,N)-r(i-1,j,N)\omega(i-1,j,N)\Big)\right] \tag{4.28}$$

$$+4(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor} E\left[\sum_{j=1}^{k_N}\Big(\omega^2(i,j,N)-\omega^2(i-1,j,N)\Big)\sum_{j=1}^{k_N}\Big(r(i,j,N)\omega(i,j,N)-r(i-1,j,N)\omega(i-1,j,N)\Big)\right]. \tag{4.29}$$

Note that under Assumption 4.2, r is zero-mean and ω is a zero-mean stationary process independent of r. Therefore components (4.27), (4.28) and (4.29) are equal to zero. Moreover, note that the analytical expression of (4.24) is already given in Lemma 4.1. Thus, in order to obtain the analytical expression of $E[\widetilde{PSRV}_{[\tau,\tau+h],N}]$ under Assumption 4.2, we only have to compute the analytical expressions of (4.25) and (4.26).

We start with (4.25). We have:

$$(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor} E\left[\sum_{j=1}^{k_N}\Big(\omega^2(i,j,N)-\omega^2(i-1,j,N)\Big)\right]^2$$
$$=(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor}\left\{\sum_{j=1}^{k_N}E\left[\omega^4(i,j,N)+\omega^4(i-1,j,N)-2\omega^2(i,j,N)\omega^2(i-1,j,N)\right]\right.$$

$$+2\sum_{j=2}^{k_N}\sum_{h=1}^{j-1}E\Bigg[\Big(\omega^2(i,j,N)\omega^2(i,h,N)-\omega^2(i,j,N)\omega^2(i-1,h,N)$$
$$-\omega^2(i-1,j,N)\omega^2(i,h,N)+\omega^2(i-1,j,N)\omega^2(i-1,h,N)\Big)\Bigg]\Bigg\}$$
$$=(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor}\left\{\sum_{j=1}^{k_N}2E[\omega^4(i,j,N)]-2E[\omega^2(i,j,N)]^2\right\}=\frac{4\Big(Q_\varepsilon+V_\varepsilon^2\Big)h}{k_N\delta_N^2\Delta_N},$$

since ω^2 is an i.i.d. process such that $E[\omega^2(i,j,N)]=2V_\varepsilon$ and $E[\omega^4(i,j,N)]=2Q_\varepsilon+6V_\varepsilon^2$, as one can easily check.

Then we move on to (4.26). We have:

$$4(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor}E\left[\sum_{j=1}^{k_N}\Big(r(i,j,N)\omega(i,j,N)-r(i-1,j,N)\omega(i-1,j,N)\Big)\right]^2$$
$$=4(k_N\delta_N)^{-2}\sum_{i=1}^{\lfloor h/\Delta_N\rfloor}\left\{\sum_{j=1}^{k_N}E\Bigg[r^2(i,j,N)\omega^2(i,j,N)+r^2(i-1,j,N)\omega^2(i-1,j,N)\right.$$
$$-2r(i,j,N)r(i-1,j,N)\omega(i,j,N)\omega(i-1,j,N)\Bigg]+2\sum_{j=2}^{k_N}\sum_{h=1}^{j-1}E\Bigg[r(i,j,N)r(i,h,N)\omega(i,j,N)\omega(i,h,N)$$
$$-r(i,j,N)r(i-1,h,N)\omega(i,j,N)\omega(i-1,h,N)-r(i-1,j,N)r(i,h,N)\omega(i-1,j,N)\omega(i,h,N)$$
$$\left.+r(i-1,j,N)r(i-1,h,N)\omega(i-1,j,N)\omega(i-1,h,N)\Bigg]\right\}$$
$$=\frac{8V_\varepsilon(\alpha-E[\nu(\tau)])(1+e^{-\theta\Delta_N})(1-e^{\theta k_N\delta_N})(1-e^{-\theta h})}{\theta(1-e^{-\theta\Delta_N})k_N^2\delta_N^2}+\frac{16\alpha V_\varepsilon h}{k_N\delta_N\Delta_N},$$

due to the stationariy of ω and the fact that r is zero-mean and independent of ω.

Finally, putting everything together, we have

$$E[\widetilde{PSRV}_{[\tau,\tau+h],N}]=E[PSRV_{[\tau,\tau+h],N}]+D_N,$$

where

$$D_N:=[4(Q_\varepsilon+V_\varepsilon^2)+16\alpha V_\varepsilon\delta_N]h\frac{1}{k_N\delta_N^2\Delta_N}+\frac{8}{\theta}V_\varepsilon(\alpha-E[\nu(\tau)])(1-e^{-\theta h})\frac{(1+e^{-\theta\Delta_N})(1-e^{-\theta k_N\delta_N})}{(1-e^{-\theta\Delta_N})k_N^2\delta_N^2}.$$

Analogous calculations in the overlapping case $W_N>\Delta_N$ lead to

$$D_N^* = [4\left(Q_\varepsilon + V_\varepsilon^2\right) + 16\alpha V_\varepsilon \delta_N]h\frac{1}{k_N^2\delta_N^3} + \frac{8V_\varepsilon(\alpha - E[\nu(\tau)])(1-e^{-\theta h})}{\theta(1-e^{-\theta\Delta_N})k_N^2\delta_N^2}$$
$$\times\left\{\frac{(2+k_N)}{2k_N\delta_N}\left[\frac{(e^{\theta k_N\delta_N - \theta\Delta_N}-1)(k_N\delta_N+\Delta_N)}{k_N\delta_N-\Delta_N} + (e^{-\theta\Delta_N} - e^{\theta k_N\delta_N})\right] + \frac{k_N}{2\Delta_N}(1+e^{\theta k_N\delta_N})(1-e^{-\theta\Delta_N})\right\}.$$

The proof is complete.

Theorem 4.3

Consider D_N in Lemma 4.2. Then recall that as $N\to\infty$, $\Delta_N = O(\delta_N^c)$, $c\in(0,1)$, and $k_N = O(\delta_N^b)$, $b\in(-1,0)$. Moreover, note that:

- as $N\to\infty$, $E[PSRV_{[\tau,\tau+h],N}] \to \langle\nu,\nu\rangle_{[\tau,\tau+h]}$ if $b\geq -1/2$ and $c<-b$ or $b<-1/2$ and $c<1+b$ (see Theorem 4.2);
- as $N\to\infty$, $D_N \sim 4(Q_\varepsilon+V_\varepsilon^2)h\frac{1}{k_N\delta_N^2\Delta_N} + 16\alpha V_\varepsilon h\frac{1}{k_N\delta_N\Delta_N} + 8V_\varepsilon(\alpha - E[\nu(\tau)])(1-e^{-\theta h})(1+e^{-\theta\Delta_N})\frac{1}{k_N\delta_N\Delta_N}$, thus $D_N\to\infty$ as $N\to\infty$ for any $(b,c)\in(-1,0)\times(0,1)$. In particular, as $N\to\infty$, D_N is $O(\frac{1}{k_N\delta_N^2\Delta_N})$ for any $(b,c)\in(-1,0)\times(0,1)$.

Therefore, as $N\to\infty$, if $b\geq -1/2$ and $c<-b$ or $b<-1/2$ and $c<1+b$, then $E[\widetilde{PSRV}_{[\tau,\tau+h],N}] = E[PSRV_{[\tau,\tau+h],N}] + D_N$ diverges, with rate $\frac{1}{k_N\delta_N^2\Delta_N}$. The proof is complete.

Theorem 4.4

Recall from Definition 4.1 that for τ with values on the price-sampling grid of mesh size δ_N:

$\hat{\nu}(\tau) := (k_N\delta_N)^{-1}\sum_{j=1}^{k_N}\left[x(\tau - k_N\delta_N + j\delta_N) - x(\tau - k_N\delta_N + (j-1)\delta_N)\right]^2.$

Moreover, from Appendix A in Bollerslev and Zhou (2002), we have, under Assumption 4.1:

$$E\left[\int_{\tau-\Delta}^{\tau}\nu(t)dt\right] = \alpha\Delta + (\nu(0)-\alpha)\theta^{-1}e^{-\theta\tau}(e^{\theta\Delta}-1)$$

and

$$E[\nu(\tau)] = \alpha + (\nu(0)-\alpha)e^{-\theta\tau}.$$

Therefore, under Assumption 4.1,

$$
\begin{aligned}
E[\hat{\nu}(\tau)-\nu(\tau)] &= (k_N\delta_N)^{-1}E\Big[\sum_{j=1}^{k_N}\Big[x(\tau-k_N\delta_N+j\delta_N)-x(\tau-k_N\delta_N+(j-1)\delta_N)\Big]^2\Big] \\
&-\Big[\alpha+(\nu(0)-\alpha)e^{-\theta\tau}\Big] \\
&= (k_N\delta_N)^{-1}E\Big[\sum_{j=1}^{k_N}\Big[\int\limits_{\tau-k_N\delta_N+(j-1)\delta_N}^{\tau-k_N\delta_N+j\delta_N}\sqrt{\nu(t)}dW(t)\Big]^2\Big]-\Big[\alpha+(\nu(0)-\alpha)e^{-\theta\tau}\Big] \\
&= (k_N\delta_N)^{-1}E\Big[\int_{\tau-k_N\delta_N}^{\tau}\nu(t)dt\Big]-\Big[\alpha+(\nu(0)-\alpha)e^{-\theta\tau}\Big] \\
&= (\nu(0)-\alpha)e^{-\theta\tau}\Big[(\theta k_N\delta_N)^{-1}(e^{\theta k_N\delta_N}-1)-1\Big].
\end{aligned}
$$

Expanding this as $N\to\infty$, we can rewrite $E[\hat{\nu}(\tau)-\nu(\tau)]=(\nu(0)-\alpha)e^{-\theta\tau}\frac{1}{2}\theta k_N\delta_N+o(k_N\delta_N)$. Furthermore, recall that $k_N\delta_N=O(\delta_N^{b+1})$ and $b\in(-1,0)$. Therefore, under Assumption 4.1, $E[\hat{\nu}(\tau)-\nu(\tau)]$ converges to zero as $N\to\infty$, with rate $k_N\delta_N$.

Now let Assumption 4.2 hold and replace x with $\tilde{x}$ in the definition of the locally averaged realized variance, i.e., consider the estimator $w(\tau):=(k_N\delta_N)^{-1}\sum_{j=1}^{k_N}\Big[\tilde{x}(\tau-k_N\delta_N+j\delta_N)-\tilde{x}(\tau-k_N\delta_N+(j-1)\delta_N)\Big]^2$. Simple calculations lead to:

$$
\begin{aligned}
E[w(\tau)-\nu(\tau)] &= E[\hat{\nu}(\tau)-\nu(\tau)]+(k_N\delta_N)^{-1}\sum_{j=1}^{k_N}E\Big[\Big[\varepsilon(\tau-k_N\delta_N+j\delta_N) \\
&-\varepsilon(\tau-k_N\delta_N+j\delta_N)\Big]^2\Big]=E[\hat{\nu}(\tau)-\nu(\tau)]+2V_\eta\delta_N^{-1}.
\end{aligned}
$$

Therefore, under Assumption 4.2, $E[w(\tau)-\nu(\tau)]$ diverges as $N\to\infty$, with rate $\frac{1}{\delta_N}$. The proof is complete.

4.8.2 Indirect inference method for the feasible bias-optimal selection of the local-window tuning parameter

The feasible selection of κ in equation (4.11) requires, for a given β, the knowledge of the volatility process $\nu(t)$ at the instant $t=\tau$ and the vol-of-vol parameter γ. A simple and computationally-efficient indirect inference method to obtain estimates of those quantities is as follows.

First, one estimates the spot volatility path using the fast Fourier transform algorithm, following the procedure detailed in Appendix B.5 of Mancino et al. (2017). In particular, from a given sample of log-price observations, one obtains estimates $\hat{v}_{n,N,S}$ of the spot volatility[6] on the grid $t_i = \frac{i-1}{2S+1}$, $i = 1,...,2S+1$, of mesh size $\Delta_S := \frac{1}{2S+1}$.

Then, using the reconstructed volatility path $\hat{v}_{n,N,S}(t_i)$, one infers the value of the parameter γ by applying the following zero-intercept multivariate regression, based on the discretization of the CKLS process in Assumption 4.3:

$$\hat{v}_{n,N,S}(t_i)^{-\beta}[\hat{v}_{n,N,S}(t_{i+1}) - \hat{v}_{n,N,S}(t_i)] = \alpha\theta\,\Delta_M \hat{v}_{n,N,S}(t_i)^{-\beta} - \theta\Delta_S \hat{v}_{n,N,S}(t_i)^{1-\beta} + \gamma\sqrt{\Delta_S}Z(t_i), \tag{4.30}$$

where $Z(t_i)$ is i.i.d. standard normal.

Denoting by $\hat{\psi}$ the estimate of the standard deviation of the disturbance term, obtained from the regression residuals, we have $\hat{\gamma} = \hat{\psi}/\sqrt{\Delta_M}$.

An estimate of $v(\tau)$ is simply given by the Fourier estimate of volatility in correspondence of the beginning of the period of interest.

Finally, note that comparing the R^2 of the regression (4.30) for different values of β allows deciding which model under Assumption 4.3 fits the data better.

[6] Note that n denotes the sample size, while N and S denote, respectively, the cutting frequencies for the computation of the Fourier coefficients of the volatility and the reconstruction of the spot volatility path. See Chapter 4 in Mancino et al. (2017) for guidance on the efficient selection of N and S for a given n.

5. IS THE VARIANCE SWAP RATE AFFINE IN THE SPOT VARIANCE? EVIDENCE FROM S&P500 DATA

5.1 Introduction

The class of the exponential affine processes, introduced in the seminal paper by Duffie et al. (2000) and characterized by Filipovic (2001), has received large consensus in the quantitative finance literature, based on its main advantages in terms of analytical tractability and empirical flexibility. The classic example of an exponential affine process, and the only one with continuous paths, is the CIR diffusion, see Cox et al. (1985). The related stochastic volatility model, studied by Heston (1993), is considered as a reference model by scholars and practitioners. Kallsen et al. (2011) have studied the valuation of options written on the quadratic variation of the asset price within the exponential affine stochastic volatility framework. In particular, they have proved, analytically, the existence of an affine link between the expected cumulated variance, i.e., the variance swap rate, and the spot variance. Note that the class of stochastic volatility models considered in Kallsen et al. (2011) allows for jumps and leverage effects, but fails to include some popular stochastic volatility models, e.g., the models by Beckers (1980); Platen (1997); Hagan et al. (2002); Grasselli (2016).

The variance swap is possibly the most plain vanilla contingent claim written on the realized variance. Indeed, it can be seen, to some extent, as the forward of the integrated variance of log-returns (see, for instance, Carr and Sun (2007); Carr and Wu (2008); Kallsen et al. (2011); Filipovic et al. (2016); Bernis et al. (2019); Jiao et al. (2019)). Volatility derivatives appear nowadays with a vast demand, especially after the global financial crisis of 2008, which induced large fluctuations in the volatility and other indicators of market stress. The large demand for volatility derivatives has resulted in a major increase in their liquidity, and thus in the reliability of their prices (see, for instance, Carr and Wu (2008)).

Based on the study by Kallsen et al. (2011), two natural questions arise: (i) could we analytically identify a wider class of models which admits an affine link between the variance swap rate and the spot variance? (ii) is it possible to test if empirical data satisfy a given link (e.g., affine, quadratic) between the variance swap rate and the unobservable spot variance? This chapter contributes to answering both questions. Regarding question (i), we prove that a larger class of models exhibits a linear link between the variance swap rate and the spot variance; also, we show that, within the class of polynomial models (see Cuchiero (2011) and Cuchiero et al. (2012)), a quadratic (respectively, affine) link appears between the variance swap rate and the multidimensional stochastic process characterizing the model in the presence (respectively, absence) of jumps. Regarding question (ii), we set up a simple testing procedure, based on Ordinary-Least-Squares (OLS), in which the unobservable spot variance is replaced with efficient estimates thereof. Then, we apply it to S&P500 empirical data over the period 2006-2018.

In particular, our first result is showing that a model exhibits the affine link between the variance swap rate and the spot variance if the stochastic differential equation satisfied by the latter is the sum of an affine drift and a zero-mean stochastic process. We term this class exponential mean-reverting variance models. This class is fairly large. In fact, it contains not only exponential affine processes with jumps (see, e.g., Bates (1996); Barndorff-Nielsen and Shephard (2001, 2002b); Duffie et al. (2003); Benth et al. (2007); Jiao et al. (2017, 2019)), but also, under suitable conditions (see Cuchiero et al. (2012); Ackerer et al. (2018)), polynomial processes. Moreover, it also contains some models based on the fractional Brownian motion (see, for instance, Abi Jaber et al. (2019)). However, it is worth noting that many popular models, e.g., the CEV model (Beckers (1980)), the SABR model (Hagan et al. (2002)), the $3/2$ and $4/2$ models (Platen (1997); Grasselli (2016)), fail to verify the affine link (see, for instance, the analysis in Section 4 of Jarrow et al. (2013) for the 3/2 model). Further, we consider the class of stochastic volatility models based on polynomial processes, introduced in Cuchiero (2011) and Cuchiero et al. (2012). The exponential affine models by Kallsen et al. (2011), which exhibit an affine link between the variance swap rate and the spot variance, are included in the polynomial class, as a special case, see Example 3.1 in Cuchiero et al. (2012). In the polynomial framework, we prove the existence, in the presence of jumps, of a quadratic correction in the link between the theoretical variance swap rate and the spot variance.

On financial markets, traded variance swaps are actually written on the realized variance, that is, the finite sum of squared log-returns sampled over a discrete grid. Instead, the corresponding theoretical pricing formulae use the continuous time approximation given by the quadratic variation of the log-price, in virtue of higher mathematical tractability. Thus, we also study the case where the theoretical variance swap rate, i.e., the expected future quadratic variation, is replaced by its empirical counterpart, namely the expected future realized variance. In this regard, we show that polynomial processes exhibit a quadratic link between the expected future realized variance and the (multidimensional) stochastic process characterizing the model. The pricing error related to this approximation has been investigated by

Broadie and Jain (2008), who conclude that the approximation works quite well, based on simulated data obtained from four different models (the Black-Scholes model, the Heston stochastic volatility model, the Merton jump-diffusion model and the Bates stochastic volatility and jump model).

Based on these results, our second contribution is testing, using OLS, if an affine or a quadratic link is satisfied by actual financial data, namely S&P 500 daily data. This may allow us to determine which class of models, affine or polynomial, provides a better fit for empirical data. Clearly, such a test requires the availability, in addition to daily prices, of the daily time-series of variance swap rate and spot variance observations. For what concerns S&P500 variance swap rates, a reliable proxy for the maturity equal to one month is quoted on the market in the form of the (squared) VIX index (see Carr and Wu (2008); CBOE (2019)). The use of the VIX squared as a proxy of the variance swap rate is present in, e.g., Todorov (2010); Martin (2011); Aït-Sahalia et al. (2018).

The spot variance is instead a latent process. Thus, the main hurdle impeding the testing of the affine/quadratic link is the latent nature of the variance process. To overcome this hurdle, the spot variance is estimated by means of the Fourier method proposed in Malliavin and Mancino (2009) and extended to jump-diffusions in Cuchiero and Teichmann (2015). Note that, although reconstructed from historical data, spot volatility estimates do not depend on the equivalent probability measure considered. This allows us to conduct our empirical tests under the risk neutral probability measure, coherently with the use of the (squared) VIX, a risk neutral object, as a proxy of the variance swap rate. Furthermore, in order to test the robustness of our findings, we also perform our empirical study replacing Fourier estimates of the spot volatility with estimates obtained through the localized realized estimators by Zu and Boswijk (2014) and Bandi and Renò (2018). Furthermore, with the same aim of strengthening our results, we test the possible relations between the VIX index squared and reconstructed the spot variance using both opening and closing values of the former.

The findings of our empirical tests are summarized as follows. First, we obtain overwhelming empirical evidence supporting the use of exponential affine models. Exponential affine models imply the existence of an affine link between the variance swap rate and the spot variance, with strictly positive coefficients. The test of the affine link over the period 2006-2018 is coherent with this prediction, in that it yields statistically significant positive coefficients and an R^2 larger than 0.95, both when Fourier or realized spot variance estimates are used. Instead, the test of the quadratic link between the variance swap rate and the spot variance over the period 2006-2018 yields a non-significant quadratic coefficient, using either Fourier or realized spot variance estimates. This result may shed light on the negligibility of the discrete sampling effect affecting the variance-swap pricing formula. In fact, the absence of a significant quadratic coefficient confirms that the daily sampling used to compute the VIX index is enough to match the continuous-time approximation of the latter, i.e., the expected future quadratic variation. This empirical finding, which is achieved in a non-parametric fashion, i.e., without assuming any parametric form for the price evolution, supports the numerical findings by Broadie and Jain (2008).

The affine and quadratic tests are performed also on yearly subsamples, to investigate the sensitivity of the results to different economic scenarios, that is, to different volatility regimes (see, e.g., Goutte et al. (2017)). Test results on yearly subsamples are more nuanced. In particular, the intercept in the affine test is not significant in 2008 and 2011, two years characterized by extreme volatility spikes. This suggests that S&P500 data in 2008 and 2011 are consistent only with the broader assumption of an exponential mean-reverting variance framework, which does not put any restrictions on the sign of the intercept (see, e.g., the model by Hull and White (1987)). Moreover, the quadratic test yields significant quadratic corrections in the years characterized by a relatively high number of price jumps. This findings support the use of polynomial models with jumps in periods when jumps are frequent. In general, our empirical analysis reveals that jumps play a non-negligible role, as we detect price-jumps in approximately 10% of days of our 13-year sample. This result is in accordance with a large literature, see, e.g., Bates (1996); Bakshi et al. (1997); Barndorff-Nielsen and Shephard (2002a); Eraker (2004). Perhaps surprisingly, high-volatility periods and periods with a larger number of jumps do not necessarily coincide. For example, in 2007, 2010 and 2013, in spite of a relatively low VIX index, the number of days with jumps is relatively large. Again, we remark that results on yearly subsamples are robust to the use of different volatility estimators.

The chapter is organized as follows. In Section 5.2 we describe the analytical framework of the paper, illustrating the exponential affine model, the exponential mean-reverting variance model and the polynomial model. In Section 5.3 we detail the spot variance estimation method and perform empirical tests to investigate if S&P500 daily data over the period 2006-2018 are consistent with the affine or the quadratic link. Section 5.4 concludes. The proofs are in the Appendix (Section 5.5).

5.2 Variance swap rate and model set-up

In this section we introduce the problem of the variance swap valuation and investigate the types of models under which an affine link between the variance swap rate and the spot variance exists.

Based on the fundamental theorem of asset pricing by Delbaen and Schachermayer (1994), we assume that the time evolution of the logarithmic asset price x follows a square-integrable semimartingale model, that is,

$$x(t) = A(t) + M(t) \tag{5.1}$$

where M is a square-integrable martingale and A is a finite-variation process on a filtered space $(\Omega, \mathscr{F}, P)$. Being interested in the pricing problem, asset price dynamics are specified under a risk neutral measure along the paper. Moreover, in the chapter we denote by $[x]_t$ the quadratic variation of the process x up to time t. The semimartingale hypothesis assures that the $[x]_t$ is finite for all times t and coincides with the quadratic variation of the martingale M, if the finite-variation process A has continuous paths.

A classical result proves that the quadratic variation can be obtained as the limit of the realized variance. More precisely, letting $\pi_n := \{0 = t_0 < t_1 < \ldots < t_n = \tau\}$ be a partition of a generic interval $[0, \tau]$ and $|\pi_n| := \sup_{k=1,\ldots,n} (t_k - t_{k-1})$ be the step of the partition, the realized variance is defined as

$$RV^n_{[0,\tau]} = \sum_{k=1}^{n} \Big(x(t_k) - x(t_{k-1}) \Big)^2. \tag{5.2}$$

Then, the following convergence holds in probability

$$[x]_\tau = \lim_{|\pi_n| \to 0} RV^n_{[0,\tau]}. \tag{5.3}$$

A financial product, called *variance swap*, has been introduced to hedge volatility risk.

Definition 5.1. ***Variance swap***
A variance swap is a financial derivative characterized by two legs, one paying the mean realized variance over an interval $[t, t+\tau]$, the other paying a fixed amount, generally called the rate or strike. Variance swap buyer pays the fix amount and receives the realized variance $RV^n_{[t,t+\tau]}$, with the convention that $t_k - t_{k-1}$ is one day, $t_0 = t$ and $t_n = t + \tau$. The strike VS^τ_t reads

$$VS^\tau_t = \tau^{-1} E\left[RV^n_{[t,t+\tau]} \,|\, \mathscr{F}_t \right]. \tag{5.4}$$

Based on higher mathematical tractability, the finite-sample realized variance (5.2) is replaced, in the theoretical variance swap pricing formula, by its continuous-time approximation, the quadratic variation $[x]_\tau$. As a consequence, the strike of the variance swap (5.4), under the continuous-time limit, reads

$$VS^\tau_t = \tau^{-1} E\left[[x]_{t+\tau} - [x]_t \,|\, \mathscr{F}_t \right]. \tag{5.5}$$

The simulation study by Broadie and Jain (2008), based on four different models (the Black-Scholes model, the Heston stochastic volatility model, the Merton jump-diffusion model and the Bates stochastic volatility and jump model), suggests that the continuous-time approximation for the variance swap pricing formula works quite well. A more general analysis of discrete variance swaps for a general time-homogeneous stochastic volatility model is provided by Bernard and Cui (2014).

A model-free pricing method, used to compute the VIX index (see CBOE (2019)), has been also proposed by Carr and Lee (2008). This method exploits the fact that the variance swap can be perfectly statically replicated through vanilla Puts and Calls, as pointed by the next result (see Carr and Wu (2006) for the proof). For a more general result, see Baldeaux and Rutkowski (2010).

Proposition 5.1. *Assuming that the underlying asset log-price x has continuous paths, then the variance swap can be statically replicated by a weighted position on vanilla Puts and Calls, that reads*

$$VS_0^\tau = \frac{2}{\tau}\mathrm{e}^{r\tau}\left(\int_0^F \frac{1}{K^2}P(K)dK + \int_F^\infty \frac{1}{K^2}C(K)dK\right), \tag{5.6}$$

where F, τ and r denote, respectively, the forward of the underlying, the maturity and the risk-free interest rate, which is assumed to be constant. The prices of the Call and Put options with strike K and maturity τ are denoted, respectively, by $C(K)$ and $P(K)$.
Moreover, in the presence of jumps in the log-price process x and assuming that the jump measure is exponentially integrable, the formula (5.6) is subject to the correction ε, which depends only on the jump measure and reads

$$\varepsilon = -\frac{2}{\tau}E\left[\int_0^\tau \int \left(\mathrm{e}^x - 1 - x - \frac{1}{2}x^2\right)\lambda(dt,dx)\right], \tag{5.7}$$

where $\lambda(dt,dx)$ denotes the compensated Levy measure of the jump process.

Note that, based on this static replication procedure, variance swap rates are intrinsically risk neutral objects. As a consequence, the empirical findings we detail in Section 5.3.2 hold under the risk neutral measure.

As far as equity models are concerned, in this work we focus on a two-dimensional framework, where the first process is the logarithm of asset price as in (5.1) and the second, called variance process, is the variance of the martingale part in (5.1) or a function of the latter. More precisely, in the rest of the chapter we consider various model specifications within the following general class for the price evolution

$$\begin{aligned} dx(t) &= \sqrt{v(t)}dW(t) + dJ(t) + a(t)dt \\ dv(t) &= dZ(t) + b(t)dt \end{aligned} \tag{5.8}$$

where W is a Brownian motion, J is a compensated jump process characterized by the Levy measure λ and Z is an integrable zero-mean stochastic process satisfying some technical conditions to preserve the non-negativity of v (see Definition 5.3). Note that the process Z is not required to be neither a semimartingale nor a Markov process. This allows us to include also some models based on the fractional Brownian motion (see, for instance, Abi Jaber et al. (2019)).

The class of models (5.8) and its extension to multidimensional volatility processes are extremely large and include almost all stochastic volatility models commonly used in finance.

5.2.1 Exponential affine model

With pricing and forecasting applications in mind, researchers have focused on some subclasses of (5.8), which are able to capture equity stylized facts while still remaining parsimonious. For instance, during the last two decades, a large literature, started by Duffie et al. (2000), has focused on exponential affine models, which are defined as follows, see Definition 2.1 in Duffie et al. (2003).

Definition 5.2. ***Exponential affine stochastic volatility model***

A Markov process (x,v) is called affine if the characteristic function of the process has an exponential affine dependence on the initial condition. That is, for every $0 \leq u < t$, there exists functions $(\psi^x_{(a,b)}(t,u), \psi^v_{(a,b)}(t,u), \phi_{(a,b)}(t,u))$ such that

$$E\left[e^{ax(t)+bv(t)}|\mathscr{F}_u\right] = \exp\left\{x(u)\psi^x_{(a,b)}(t-u) + v(u)\psi^v_{(a,b)}(t-u) + \phi_{(a,b)}(t,u)\right\}.$$

Under natural financial hypotheses, we have $\psi^x_{(a,b)}(t-u) = 1$. Moreover, Duffie et al. (2000) show that $\psi^v_{(a,b)}$ satisfies a generalized first order non-linear differential equation of Riccati type and $\phi_{(a,b)}$ is a primitive of a functional of $\psi^v_{(a,b)}$.

The most popular exponential affine model, and the only one with continuous paths, is the model by Heston (1993), which reads

$$\begin{aligned} dx(t) &= \sqrt{v(t)}dW(t) + (r - \frac{1}{2}v(t))dt \\ dv(t) &= \gamma\sqrt{v(t)}dB(t) + \theta(\alpha - v(t))dt \end{aligned}, \tag{5.9}$$

where W and B are correlated Brownian motions. Moreover, it is easy to verify that, under the Heston model, the variance swap strike (5.5) has the following expression:

$$VS^\tau_t = \alpha + (v(t) - \alpha)\frac{1 - e^{-\theta\tau}}{\theta\tau}. \tag{5.10}$$

The class of exponential affine models is wide, including also jumps processes, and has been extensively investigated, see for instance Bates (1996); Filipovic and Mayerhofer (2009); Benth (2011); Keller-Ressel (2011); Hubalek et al. (2017); Bernis et al. (2019); Horst and Xu (2019); Jiao et al. (2019). In this regard, we highlight the results by Keller-Ressel et al. (2011) and Cuchiero and Teichmann (2013), who show that exponential affine processes are regular. Note that, in the exponential affine framework, the variance process v needs to be driven by a martingale Z (see (5.8)) with finite quadratic variation. Moreover, the drift process b and the Levy measure of the jump process J in (5.8) need to be affine with respect to the variance process v.

Kallsen et al. (2011) show that the affine link between the spot variance and the expected integrated variance holds for any exponential affine stochastic volatility model. Their result is presented in the following proposition.

Proposition 5.2. *Let (x,v) be an exponential affine stochastic volatility model. Then, the triplet $(x(t),v(t),[x]_t)$ is a Markov exponential affine process. Moreover, the process $[x]_t$ has the following characteristic function*

$$E\left[e^{u[x]_{t+\tau}}|\mathscr{F}_t\right]=\exp\left\{u[x]_t+v(t)\Psi_u^V(\tau)+\Phi_u^V(\tau)\right\},$$

where Ψ_u^V satisfies a couple of first order non-linear differential equations of Riccati type and Φ_u^V is a primitive of a functional of Ψ_u^V. More precisely, using the parameter notation for the exponential affine model introduced in Lemma 4.2 of Kallsen et al. (2011), they satisfy

$$\frac{\partial\Psi_u^V}{\partial t}(t)=\frac{1}{2}\gamma_1^{11}\left(\Psi_u^V(t)\right)^2+\beta_1^1\Psi_u^V(t)+\gamma_1^{22}u+\int_{\mathbb{R}^+\times\mathbb{R}}\left(e^{x_1\Psi_u^V(t)+ux_2^2}-1-\Psi_u^V(t)h(x_1)\right)\kappa_1(dx),$$

$$\Psi_u^V(0)=0,$$

$$\Phi_u^V(t)=\int_0^t\left[\beta_0^1\Psi_u^V(s)+\gamma_0^{22}u+\int_{\mathbb{R}^+\times\mathbb{R}}\left(e^{x_1\Psi_u^V(s)+ux_2^2}-1-\Psi_u^V(s)h(x_1)\right)\kappa_0(dx)\right]ds.$$

Moreover, assuming that $\int_0^\tau E[v(t)]dt<\infty$, then

$$VS_t^\tau=v(t)\Psi(\tau)+\Phi(\tau), \tag{5.11}$$

where $\Psi(\tau)$ (respectively, $\Phi(\tau)$) is the partial derivative of $\Psi_u^V(\tau)$ (respectively, $\Phi_u^V(\tau)$) with respect to u, at $u=0$.

Note that this affine link is not satisfied by all stochastic volatility models with an explicit Laplace transform. For instance, it is not satisfied by the $3/2$ model of Platen (1997) and by the $4/2$ model of Grasselli (2016) (see the analysis in Section 4.3 of Jarrow et al. (2013)).

In the following proposition we complete the result by Kallsen et al. (2011), showing that the functions $\Psi(\tau)$ and $\Phi(\tau)$ are strictly positive. This additional result is interesting in view of our empirical study of section 5.3, where we test if S&P data are coherent with the exponential affine framework, based on the significance of the estimates of the coefficients in (5.11). The proof of this additional result crucially relies on the characterization of exponential affine models by Filipovic (2001), who shows, under mild conditions (mainly the non-negativity of v), that the volatility process v has to be a continuous-state branching processes with immigration in the exponential affine framework. Note that the explicit stochastic differential equation satisfied by a generic continuous-state branching process with immigration is provided by Dawson and Li (2006) and Li and Ma (2008), who also detail the conditions to have a stationary distribution for the variance process. The existence of a stationary distribution is usually considered as a natural property of the variance process.

Proposition 5.3. *Let (x,v) follow an exponential affine stochastic volatility model and assume that the variance process v admits a non-degenerate stationary distribution. Then $\Psi(t) > 0$ and $\Phi(t) > 0$ for all $t > 0$.*

Based on Proposition 5.3, the exponential affine framework could be rejected by empirical data if any of the coefficient estimates is not strictly positive. In that event, it could be worth investigating the adequacy of the more general exponential mean-reverting variance and polynomial frameworks, respectively detailed in Subsection 5.2.2 and 5.2.3.

5.2.2 Exponential mean-reverting variance model

In this subsection we introduce a more general subclass of the stochastic volatility models included in (5.8), which we name *exponential mean-reverting variance models*. Moreover, we show that, under this paradigm, an affine relationship between the variance swap rate and the spot variance holds.

Definition 5.3. ***Exponential mean-reverting model***

The stochastic volatility model (x,v) is an exponential mean-reverting variance model if (x,v) satisfies

$$\begin{aligned} dx(t) &= \sqrt{v(t)}dW(t) + dJ(t) \\ dv(t) &= dZ(t) + \theta(\alpha - v(t))dt \end{aligned} \tag{5.12}$$

where $\theta > 0$, the jump process J is square-integrable, with Levy measure affine in the volatility process, and Z is an integrable zero-mean stochastic process without negative jumps and such that $dZ(t)$ is zero and the volatility drift is non-zero for all t such that $v(t) = 0$, in order to preserve the non-negativity of v (see, for instance, Feller (1951) and Tanaka (1979) for more details about reflecting conditions on $\mathbb{R}^+$).

This subclass includes not only exponential affine processes but also some non-Markovian processes driven by a fractional Brownian motion. Relevant examples with rough behavior inside this subclass are provided by Abi Jaber et al. (2019). The next result shows that the expected quadratic variation of an exponential mean-reverting variance model is affine in the spot variance.

Proposition 5.4. *Let (x, v) be an exponential mean-reverting variance model, as defined in (5.12). Then the expectation of the quadratic variation $[x]$ of the log-price is an affine function of the spot variance v, i.e., there exist deterministic functions Ψ and Φ such that*

$$E\left[[x]_\tau\right] = v(0)\Psi(\tau) + \Phi(\tau). \tag{5.13}$$

Differently from the case of the exponential affine framework, in this case the coefficients $\Psi(\tau)$ and $\Phi(\tau)$ are not strictly positive. A first example of a model satisfying Definition 5.3 but not Definition 5.2 is the stochastic volatility model by Hull and White (1987), under which the volatility is log-normal. In particular, the Hull-White model fits Definition 5.3 for $\alpha = 0$ and $dZ(t) = \gamma v(t) dW(t)$, where W is a Brownian motion and $\gamma > 0$. A straightforward computation shows that the variance swap rate is linear with respect to the spot volatility. Moreover, note that this model only admits a degenerate steady-state distribution, namely a Dirac delta on zero.

Some other interesting examples are given by models defined via their Laplace transform as, for instance, those based on stochastic Volterra equations (see Abi Jaber et al. (2019) and references therein). The main mathematical difficulty inherent to these models is that the volatility is not a Markov process. One could overcome this problem by taking an infinite-dimensional point of view. The initial value of the variance process $v(0)$ is then replaced by a function g_0 that takes into account the initial conditions. Thus, under the infinite-dimensional viewpoint, the link between the variance swap rate and the initial variance is the functional linear link between the variance swap rate and the function g_0. Finally, note that it is not possible to work with the function g_0 empirically, unless this function is assigned a parametric form.

In Section 5.3.2 we show that empirical subsamples related to the years 2008 and 2011, where, respectively, the outbreak of the global financial crisis and the

Euro-zone debt crisis took place, exhibit a non-significant intercept parameter. This result can tilt the balance in favor of log-normal models like Hull-White during crisis periods.

5.2.3 Polynomial model

In this section we consider the class of stochastic volatility models based on polynomial processes, introduced in Cuchiero (2011) and Cuchiero et al. (2012). As pointed out in Cuchiero et al. (2012), exponential affine processes are polynomial processes. Moreover, under suitable restrictions, the polynomial class could be considered as a sub-class of (5.8), see Cuchiero (2018).

Let $\mathscr{P}_k$ denote the vector space of polynomials up to degree k. In the bi-dimensional case, we have the following definition of a polynomial process.

Definition 5.4. ***Polynomial process***

A time-homogeneous Markov process (x,v) is said m-polynomial, if, for all $k \in \{0,\dots,m\}$, $f \in \mathscr{P}_k$, (r,s) in the state space and $t > 0$, it holds that

$$(r,s) \to E^{(r,s)}\left[f(x(t),v(t))\right] \in \mathscr{P}_k, \tag{5.14}$$

where, for any $0 \le u < t$, we adopt the standard notation $E^{(r,s)}\left[f(x(t),v(t))\right] = E[f(x(t),v(t))|x(u) = r, v(u) = s]$. Also, the semigroup is assumed to be strongly continuous. Moreover, if (x,v) is m-polynomial for all $m \in \mathbb{R}$, then (x,v) is said polynomial.

A relevant non-affine example in this class is the Jacobi stochastic volatility model, see Ackerer et al. (2018). Other examples and applications of polynomial process can be found in Cuchiero (2011); Filipovic et al. (2016); Callegaro et al. (2017); Cuchiero (2018); Cuchiero et al. (2018); Ackerer and Filipovic (2020).

The following proposition allows us to investigate the existence of a quadratic correction in the link between theoretical variance swap rates and spot variance in the polynomial framework.

Proposition 5.5. *Let (x,v) be a 2-polynomial process describing a stochastic volatility model, then the expected quadratic variation of x belongs to $\mathscr{P}_2$ in (r,s). Moreover, if (x,v) has continuous paths, then the expected quadratic variation of x is affine in (r,s).*

This result suggests that the presence of a statistically significant quadratic correction could be explained by the presence of jumps in the underlying. In fact, the empirical analysis in Section 5.3.2 seems to support this finding. In particular, in Section 5.3.2, we point out that a quadratic coefficient is statistically significant in the years with a higher frequency of price jumps.

We conclude the section by discussing the effects of discrete sampling on the functional form of the variance swap rate. Indeed, the actual price of traded variance swaps relies on the computation of the realized variance in place of its asymptotic approximation, given by the quadratic variation (see Definition 5.1). In this regard, the following result holds.

Proposition 5.6. *If (x,v) is a 2-polynomial process describing a stochastic volatility model, then the expected realized variance of x belongs to $\mathscr{P}_2$.*

Based on Proposition 5.6, the variance swap rate for a polynomial stochastic volatility model is at most quadratic in (x,v), that is, there exist coefficients $p_{i,j}(\cdot)$, $i,j=0,1,2$, such that

$$VS_t^\tau = p_{0,0}(\tau)+p_{1,0}(\tau)\,x(t)+p_{0,1}(\tau)\,v(t)+p_{2,0}(\tau)\,x(t)^2+p_{1,1}(\tau)\,x(t)\,v(t)+p_{0,2}(\tau)\,v(t)^2\,. \tag{5.15}$$

This result is interesting in that it may help collect empirical evidence supporting the result by Broadie and Jain (2008). The authors show, for some well known models, that the expected quadratic variation provides an efficient approximation of the actual VIX index, whose computation is based on a daily sampling scheme (see Carr and Wu (2006); CBOE (2019)). In other words, non-significant estimates of the quadratic coefficients in (5.15) may represent empirical evidence that the continuous-time approximation works well enough.

Finally, note that, based on Proposition 5.5, the expression (5.15) is also implied by the assumption that the data-generating process is a polynomial stochastic volatility model with jumps. Section 5.3.2 analyses if a second order correction fits S&P500 data better than the affine link implied by the affine framework (5.11).

5.3 Empirical study

In this Section we perform an empirical study to investigate if S&P500 daily data over the period $2006-2018$ are consistent with the affine framework (see Paragraph 5.3.2) or the polynomial framework (see Paragraphs 5.3.2 and 5.3.2), based on the statistical significance of the estimates of the coefficients $p_{i,j}(\cdot), i,j=0,1,2$, in (5.15). To perform this study, we use the daily series of variance swap rate and log-price observations, plus a daily series of estimates of the unobservable spot variance. Accordingly, this Section begins with the description of the data used for the empirical exercise, while Section 5.3.1 describes the methods employed to reconstruct the spot variance path on a daily grid from the series of high-frequency prices.

The data we use, ranging over the period 2006-2018, are the series of S&P500 prices, recorded at the 1-minute sampling frequency (see panel a) in Figure 5.1), and the series of daily VIX index values (see panel b) in Figure 5.1). The period 2006-2018 encompasses a number of volatility peaks, corresponding to historical financial events such as the global financial crisis of 2008, the flash-crash of May 2010, the Eurozone debt crisis of 2011, the Brexit events of 2016 and the US-China ‘trade war’ of 2018. For a detailed description of the events that have affected the US stock market since the 1990s, see Horst and Xu (2019).

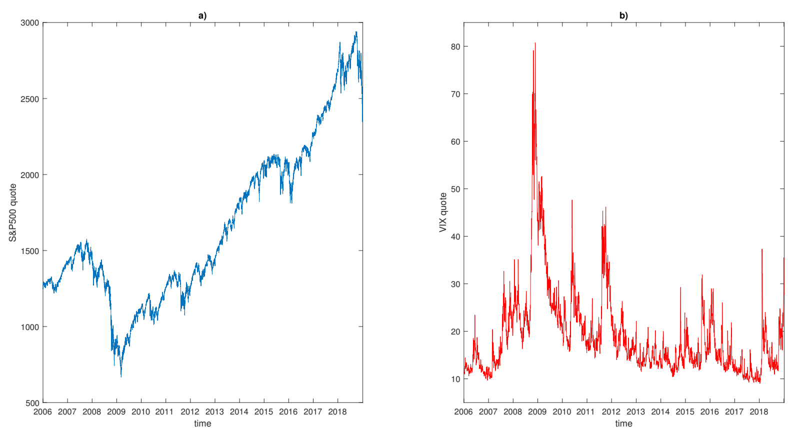

Fig. 5.1: Panel a): 1-minute S&P500 trade prices over the period 2006-2018; panel b): Opening quotes of the VIX index over the period 2006-2018.

5.3.1 Spot variance estimation

The latent spot variance is reconstructed from 1-minute prices using, as a reference, the Fourier methodology, which is naturally suited to globally reconstruct the spot volatility trajectory from high-frequency prices (see Chapter 4 in Mancino et al. (2017)). However, as a robustness check, we also use localized realized variance estimators (see Chapter 8 in Aït-Sahalia and Jacod (2014)) in our empirical study. For the reader’s convenience, we briefly recall the definition and main properties of both Fourier and realized spot variance estimators.

Fourier estimators

Let $[0,2\pi]$ be the time horizon[1] and consider the time price-sampling grid $\{0 = t_0 < \ldots < t_n = 2\pi\}$. For any integer $k, |k| \leq 2N$, the k-th (discrete) Fourier coefficient

[1] In applications, we can always assume that the price process $x(t)$ is observed on $[0,2\pi]$ by re-scaling the actual time interval.

of the log-return process is given by

$$c_k(dx_n) := \frac{1}{2\pi}\sum_{j=0}^{n-1} e^{-\mathrm{i}kt_j}\Big(x(t_{j+l}) - x(t_j)\Big), \tag{5.16}$$

where the symbol i denotes the imaginary unit, that is, $\mathrm{i} = \sqrt{-1}$.

Then, for $|k| \leq N$, the k-th (discrete) Fourier coefficient of the volatility is defined as

$$c_k(v_{n,N}) := \frac{2\pi}{2N+1}\sum_{|s|\leq N} c_s(dx_n)\, c_{k-s}(dx_n), \tag{5.17}$$

The convolution (5.17) contains the identity relating the Fourier transform of the log-price process $x(t)$ to the Fourier transform ofthe variance $v(t)$. In this regard, it is important to note that the computation of the Fourier coefficients of the variance process $v(t)$ is not affected by the presence of a price drift, as proved in Lemma 2.2 of Malliavin and Mancino (2009). Therefore, Fourier volatility estimates obtained from historical prices coincide with those obtained under the risk neutral measure. The same invariance holds for realized estimators (see Chapter 6 in Aït-Sahalia and Jacod (2014)). Also, as pointed out in Proposition 5.1, variance swap rates are derived via a static replication based on vanilla options and thus are risk neutral objects. Consequently, the empirical results we show in Section 5.3.2 hold under the risk neutral measure.

By (5.17) we gather all the Fourier coefficients of the variance function by means of the Fourier transform of the log-returns. Then, the reconstruction of the variance function V_t from its Fourier coefficients is obtained through the Fourier-Fejér summation, i.e., the Fourier estimator of the spot variance is defined the random function of time

$$\widehat{v}_{n,N,S}(t) := \sum_{|k|<S}\left(1 - \frac{|k|}{S}\right) c_k(v_{n,N})\, e^{\mathrm{i}kt}, \tag{5.18}$$

where S is a positive integer smaller than N. The uniform consistency of this estimator is proven in Malliavin and Mancino (2009)[2].

Note that the definition of the estimator $\widehat{v}_{n,N,S}(t)$ depends on three parameters, the number of data n and the two cutting frequencies N and S. An appropriate choice of the cutting frequencies is needed to filter out microstructure noise effects. In fact, on one side the estimation of the instantaneous volatility benefits from the availability of a large amount of data, at least from a statistical point of view. On the other side, high-frequency data are affected by microstructure noise

[2]This result is also reported in Chapter 3 of this thesis

effects deriving from, e.g., bid-ask bounces, infrequent trading and price discreteness. Therefore, it is necessary to employ volatility estimators which are able to filter out microstructure noise contaminations. The estimator of the spot variance $v(t)$ by means of the Fourier method has been designed to this aim, and is robust to the presence of different types of noise contaminations in the price process, see Mancino and Sanfelici (2008).

The Fourier method to estimate the spot variance has been extended to the case where jumps are present in the price and the volatility processes by Cuchiero and Teichmann (2015). The procedure has two steps. First, an estimate of the Fourier coefficients of a continuous invertible function $\rho(v)$ of the instantaneous variance is obtained. The estimator of the k-th Fourier coefficient takes the form

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathrm{e}^{-\mathrm{i}kt_j} g(\sqrt{n}(x(t_{j+1}) - x(t_j))), \tag{5.19}$$

where the function g can assume different specifications. We will consider $\rho_g(v(t)) = \mathrm{e}^{-v(t)/2}$, that is, we choose here $g(x) = \cos x$. Second, we invoke the Fourier-Fejér inversion formula as in (5.18) to reconstruct the path of the process $\rho(v)$ as follows:

$$\frac{1}{2\pi}\frac{1}{n}\sum_{j=0}^{n-1} F_S(t - t_j)\ g(\sqrt{n}(x(t_{j+1}) - x(t_j))), \tag{5.20}$$

where $F_S(\cdot)$ is the Fejér kernel. Note that also (5.18) can be re-written by means of $F_S(\cdot)$, see Mancino and Recchioni (2015). Finally, this is translated into an estimator of the spot variance by inverting the function $\rho(\cdot)$. The obtained estimator of the instantaneous variance is consistent and asymptotically efficient in the absence of microstructure noise (see Cuchiero and Teichmann (2015)).

In order to assess whether the characteristics of the S&P500 1-minute prices data require either the use of the jump-robust Fourier estimator of spot volatility or not, we perform the following tests.

First, we split the sample into daily subsamples and apply the test by Aït-Sahalia and Xiu (2019) for the null hypothesis that the price is an Itô semimartingale. Test results at the 95% confidence level, illustrated in the Figure 5.2, panel a), show, consistently with the literature Andersen et al. (2001a), that the impact of microstructure noise on prices can be considered negligible at the 5-minute frequency. This finding is consistent with the behavior of the volatility signature (Figure 5.2, panel b)), which shows that the total realized variance of the sample stabilises, around 0.4, from the 5-minute frequency and downwards.

Secondly, after downsampling the log-price series at the 5-minute frequency, we apply the jump detection test by Corsi et al. (2010) for the null hypothesis that the price is a continuous semimartingale. Test results at the 99.9% confidence level show that jumps are detected in 10.35% of the daily subsamples over the period 2006-2018. Figure 5.3 shows the values of the test statistic computed from daily subsamples (panel a)) and the ensuing percentage of days with jumps per year (panel b)). The percentage of jumps detected per year is compatible with the empirical results in Corsi et al. (2010).

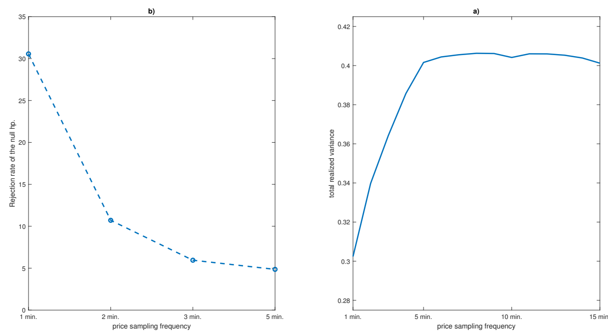

Fig. 5.2: Panel a): rejection rate of the null hypothesis of the noise-detection test by Aït-Sahalia and Xiu (2019), performed on daily subsamples of S&P500 prices over the period 2006-2018, for different sampling frequencies; panel b): volatility signature plot, i.e., total S&P500 realized variance for the period 2006-2018 as a function of the price sampling frequency.

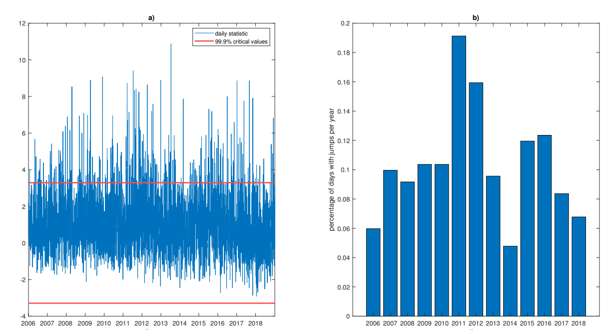

Fig. 5.3: Panel a): values of the statistic of the jump-detection test by Corsi et al. (2010) computed over the period 2006-2018; panel b): ensuing percentage of days with jumps per year.

Based on the results of the two tests, in order to obtain spot variance estimates, we proceed as follows. On the consecutive days in which the hypothesis of absence of jumps is not rejected (amounting approximately to 90% of the sample), the Fourier estimator (5.18) is applied with all prices recorded at 1-minute frequency. Instead, for the sparse days in which jumps are detected (amounting approximately 10% of the sample) we use spot variance estimates obtain through the Fourier estimator (5.20), applied to sparsely sampled 5-minute prices[3].

In the case of the estimator (5.18), the cutting frequencies have been selected as $N = n^{2/3}/2$ and $S = n^{2/3}/(16\pi)$, according to Mancino and Recchioni (2015), who find these cutting frequencies to be optimal in the presence of different types of noise and noise intensities. For the estimator (5.20), instead, the frequency S is selected as $S = (n/4)^{2/3}$, in accordance to Cuchiero and Teichmann (2015).

Local realized estimators

As mentioned, in order to test the robustness of the results of our empirical study, we replicate the latter using local realized volatility estimators in place of Fourier estimators. More precisely, in the absence of jumps, we apply the localized version of the noise-robust two-scale realized variance of Zhang et al. (2005) to 1-minute prices, while, when jumps are detected, we apply the localized version of the jump-robust threshold realized variance of Mancini (2009) to 5-minute prices. These estimators are defined as follows.

Consider the time window $[t-h,t]$ and assume that $n+1$ prices are sampled on the grid $\{t-h = t_0 < t_1 \dots < t_n = t\}$. The localized two-scale realized variance and threshold realized variance at time t are given, respectively, by

$$\hat{v}_{n,K,h}(t) = \frac{1}{Kh}\sum_{j=K}^{n}(x(t_j)-x(t_{j-K}))^2 - \frac{\bar{n}}{n}\frac{1}{h}\sum_{j=1}^{n}(x(t_j)-x(t_{j-1}))^2, \tag{5.21}$$

where $\bar{n} = \frac{nh-K+1}{Kh}$, and

$$\hat{v}_{n,\theta,h}(t) = \frac{1}{h}\sum_{j=1}^{n}(x(t_j)-x(t_{j-1}))^2 \mathbf{1}_{\{(x(t_j)-x(t_{j-1}))^2 \,\leq\, \theta\}}, \tag{5.22}$$

where θ is a vanishing threshold. Consistency and asymptotic normality of (5.21) and (5.22) are proved, respectively, by Zu and Boswijk (2014) and Bandi and Renò (2018). The finite-sample implementation of (5.21) (respectively, (5.22)) requires

[3]Note that the estimator by Cuchiero and Teichmann (2015) is also consistent in the presence of jumps in the volatility, which typically occur jointly to jumps in the price, coherently with the so-called leverage effect (see, e.g., the empirical studies on US markets in Jacod and Todorov (2010); Bandi and Renò (2016); Bibinger and Winkelmann (2018)).

the selection of K and h (respectively θ and h). In the case of (5.21), we select $K = 3$ and $h = 4$ hours based on the feasible method illustrated in Section 3.4 of Zu and Boswijk (2014). In the case of (5.22), following Section 6 of Bandi and Renò (2018), we select the threshold θ based on the data-driven procedure suggested by Corsi et al. (2010) and maintain $h = 4$ hours to have enough data points to perform the estimation effectively.

5.3.2 Empirical test results

We now focus on testing the empirical link between the rate of the variance swap with time to maturity equal to one month, i.e., the VIX index squared, and the couple spot variance - log price. If estimates of the unobservable spot variance are available, equation (5.15) can be rewritten, in the case of the S&P500 index, as follows. Let $L = 3267$ denote the number of days in our sample and let $\tau = 1/12$. Then, for $t_i = \frac{i-1}{252}$, $i = 1, 2, ..., L$[4], we write

$$\mathrm{VIX}^2_{t_i} = p_{0,0}(\tau) + p_{1,0}(\tau)\, x_{t_i} + p_{0,1}(\tau)\, \widehat{v}^n_{t_i} + p_{2,0}(\tau)\, x^2_{t_i} + p_{1,1}(\tau)\, x_{t_i}\, \widehat{v}^n_{t_i} + p_{0,2}(\tau)\, (\widehat{v}^n_{t_i})^2, \tag{5.23}$$

where:

- VIX_{t_i} denotes the opening (respectively, closing) quote of the VIX index on the i-th day;

- x_{t_i} denotes the opening (respectively, closing) log-price of the S&P500 index on the i-th day;

- $\widehat{v}^n_{t_i}$ denotes the estimated spot variance at the beginning (respectively, end) of the i-th day, obtained from a sample of size n through a consistent estimator.

Some comments are needed. First, based on the results of Propositions 5.2 and 5.5, the presence of jumps does not spoil the affine/polynomial structure, thus the regression coefficients $p_{i,j}(\tau),\ i,j = 0,1,2$, include the potential contribution of jumps. In the following, we drop the argument τ from $p_{i,j}(\tau)$ as we always consider monthly coefficients. Secondly, the consistency of the spot variance estimators (5.18) and (5.20) (respectively, (5.21) and (5.22)) allows us to neglect the finite-sample error related to $\hat{v}^n_{t_i}$.

We aim at testing the significance of the estimates of the coefficients in equation (5.23) in three progressively broader frameworks: the affine framework, introduced in Definition 5.2 and extended in Definition 5.3 (hereafter, *affine framework*); the polynomial framework of Definition 5.4, where the variance swap rate is first assumed to be a quadratic function of v only (hereafter, *quadratic framework*) and

[4] This holds for opening quotes. When closing quotes are used, then $t_i = \frac{i}{252}$, $i = 1, ..., L$.

then is assumed to be a polynomial function of the couple (x, v) (hereafter, *polynomial framework*). Each test is performed using both Fourier and realized spot variance estimators and both VIX opening and closing quotes to check the robustness of the results.

Affine framework

In this paragraph we consider the exponential affine and exponential mean-reverting variance models, which both imply the existence of an affine relationship between the variance swap rate and the spot variance. Note that in the affine framework equation (5.23) reduces to

$$\mathrm{VIX}_{t_i}^2 = p_{0,0} + p_{0,1}\,\widehat{v}_{t_i}^n. \tag{5.24}$$

Recall that the main discriminant factor between the exponential affine model and its extension to the exponential mean-reverting variance class is that the former implies the coefficients $p_{0,0}$ and $p_{0,1}$ in equation (5.24) are strictly positive. Thus, we are not only interested in testing if the affine dependence between the variance swap rate and the spot variance is satisfied by empirical data, but also in verifying if both parameter estimates are significantly different from zero, as this would allow us to accept or reject the exponential affine framework.

The coefficients in (5.24) are estimated using OLS. In order to avoid performing a spurious regression (see Granger and Newbold (1974)), we first test for the null hypothesis of the presence of a unit root in the VIX squared series and in the Fourier (respectively, realized) spot volatility estimates series, using the Augmented Dickey-Fuller test (see Dickey and Fuller (1979)). For all series, test results at the 99% confidence level reject the null hypothesis. Thus, the series are assumed to be stationary in the rest of the analysis. The results of the OLS estimation are overwhelming: we obtain an R^2 larger than 0.95 and significant coefficients estimates, independently of the spot variance estimator and the VIX sampling time used, as shown in Table 5.1. In particular, the fact that estimates of both coefficient are always significant and positive suggests that an exponential affine framework is a suitable fit for the S&P500 data over the period 2006-2018. Note that the regression standard errors have been computed using the Newey-West methodology (see Newey and West (1987)), to account for the presence of heteroskedasticity and autocorrelation in the residuals.

A natural question that arises is whether the coefficients in (5.24) are sensitive to events of distress, such as the global financial crisis of 2008, or are stable over time, instead. To investigate this aspect, the coefficients of (5.24) are estimated on yearly subsamples. Estimation results are detailed in Table 5.5 in the Appendix

	coeff.	estimate	std. err.	t stat.	p value	R^2
Fourier spot var. estimates &	$p_{0,0}$	0.013	0.003	4.142	0	0.958
VIX opening quotes	$p_{0,1}$	0.975	0.129	7.560	0	
Fourier spot var. estimates &	$p_{0,0}$	0.013	0.003	3.904	0	0.958
VIX closing quotes	$p_{0,1}$	0.989	0.132	7.516	0	
Realized spot var. estimates &	$p_{0,0}$	0.013	0.004	3.804	0	0.950
VIX opening quotes	$p_{0,1}$	1.228	0.196	6.267	0	
Realized spot var. estimates &	$p_{0,0}$	0.013	0.003	3.849	0	0.951
VIX closing quotes	$p_{0,1}$	1.251	0.205	6.095	0	

Tab. 5.1: Affine framework (5.24): estimation from S&P500 data over the period 2006-2018 for different combinations of spot variance estimation methods and VIX sampling times (p-values $\leq 10^{-4}$ are reported as zero).

and offer interesting insights. For brevity, here and in the following paragraphs, for yearly subsamples we report only results for the case when the spot variance is estimated with Fourier estimators and the VIX opening quotes are used, as results for other combinations are basically the same for each framework analysed.

In periods of distress, like 2008, when the global financial crisis broke out, or 2011, when the financial turmoil related to sovereign debt crisis in the Euro area took place, the intercept estimates are not significant at the 95% confidence level. Thus, based on Proposition 5.3, S&P500 data in 2008 and 2011 look consistent only with the broader assumption of an exponential mean-reverting variance data-generating process, which poses no restrictions on the sign of the coefficients. In other words, results on yearly subsamples tilt the balance in favor of the use, during crisis periods, of models that imply a linear relationship between the variance swap rate and the spot variance, such as the model by Hull and White (1987) (see the discussion in Section 5.2.2). Finally, note that the empirical analysis suggests that the drift of the variance process is affine in the variance itself. As a consequence, models with stronger mean reversion, e.g., the $3/2$ model, are not coherent with our empirical findings.

Quadratic framework

In this paragraph we extend the analysis to take into account a possible quadratic link between variance swap rates and spot variance. Through the following, the term *quadratic* is used to refer to polynomial models that are not exponential affine, where no risk of confusion exists. According to Propositions 5.5 and 5.6, when polynomial models with jumps are considered and/or discrete-sampling effects can not be neglected, the variance swap rate is a bivariate polynomial in X and V. Based on these results, the rest of our empirical study is aimed to investigating the possible existence of a quadratic link between the variance swap rate and the spot variance and/or the log-price. In particular, the analysis is split into two parts. In this paragraph, devoted to what we have called the *quadratic framework*, we check if a quadratic form with respect to the spot variance fits the data better than an affine form. In the next paragraph we will deal with the general form as in (5.23). In the quadratic framework, equation (5.23) reads

$$\mathrm{VIX}_{t_i}^2 = p_{0,0} + p_{0,1}\,\widehat{v}_{t_i}^n + p_{0,2}\left(\widehat{v}_{t_i}^n\right)^2. \tag{5.25}$$

However, we measure the sample linear correlation between the reconstructed spot variance and its square and find it to be larger than 0.8 for both Fourier and realized estimates, thus signalling the presence of collinearity, which represents a violation of the OLS hypotheses. The problem of collinearity is typical of polynomial regressions and can be solved by transforming the regressors in equation (5.25) through the use of orthogonal polynomials, i.e., by performing an orthogonal polynomial regression (see Narula (1979)). This way we are able to isolate the actual additional contribution of the square of variance estimates to the dynamics of the VIX index squared, if any. Accordingly, using the Gram-Schmidt algorithm, we transform the vector of spot variance estimates and the vector of their squared values into orthogonal vectors and estimate the following regression model:

$$\mathrm{VIX}_t^2 = q_{0,0} + q_{0,1}\,z_t^{(1)} + q_{0,2}\,z_t^{(2)}, \tag{5.26}$$

where $z^{(1)}$ and $z^{(2)}$ denote, respectively, the orthogonal transformations of the vector of the spot variance estimates and the vector of the squared spot variance estimates. Clearly, the coefficients in equation (5.26) are not comparable to those those in (5.25). However, this is not relevant for our study, as we aim only at assessing the significance of the additional contribution of the squared variance estimates to the dynamics of the VIX squared, not at making inference of the coefficients in equation (5.25).

The results of the OLS estimation of the coefficients in (5.26) over the period 2006-2018 are reported in Table 5.2. These results point out that the additional contribution of the squared spot variance is not statistically significant, independently of the spot variance estimator and VIX sampling time used. In order to interpret these results, we first need to recall that, from one side the class of polynomial models includes exponential affine one as a subclass, from the other, in the presence of jumps, while the polynomial model gives rise to a quadratic correction (see Proposition 5.5), the exponential affine model still ensures an affine link between the variance swap rate and the spot variance. Therefore, as in Paragraph 5.3.2, we have already ascertained that the exponential affine framework is a suitable fit for S&P500 data over the period 2006-2018, we deduce that the results in Table 5.2 confirm the adequacy of the exponential affine framework in capturing the empirical features of S&P500 data. In other words, Table 5.2 points towards the fact that the extension to the quadratic framework is not necessary to capture the empirical link between the variance swap rate and the spot variance.

Furthermore, Table 5.2 offers additional interesting insight. Recall that the computation of the VIX index is based on a daily sampling scheme. Thus, it is natural to ask whether the VIX index squared is adequately approximated by its asymptotic counterpart, namely the future expected quadratic variation. If this were not the case, one would observe a significant quadratic correction due to the discrete sampling, see Proposition 5.6. As this is not the case, we infer that the continuous limit represents a very good approximation, thus providing empirical support to the numerical result by Broadie and Jain (2008).

	coeff.	estimate	std. err.	t stat.	p value	R^2
Fourier spot var. estimates & VIX opening quotes	$q_{0,0}$	0.040	0.001	32.150	0	0.959
	$q_{0,1}$	2.959	0.394	7.507	0	
	$q_{0,2}$	0.019	0.566	0.034	0.973	
Fourier spot var. estimates & VIX closing quotes	$q_{0,0}$	0.040	0.001	35.507	0	0.959
	$q_{0,1}$	2.713	0.316	8.586	0	
	$q_{0,2}$	0.020	0.378	0.054	0.957	
Realized spot var. estimates & VIX opening quotes	$q_{0,0}$	0.041	0.001	35.378	0	0.951
	$q_{0,1}$	4.324	0.353	12.257	0	
	$q_{0,2}$	0.018	0.268	0.067	0.946	
Realized spot var. estimates & VIX closing quotes	$q_{0,0}$	0.042	0.001	36.699	0	0.952
	$q_{0,1}$	4.546	0.330	13.780	0	
	$q_{0,2}$	0.019	0.262	0.073	0.942	

Tab. 5.2: Quadratic framework (5.26): estimation from S&P500 data over the period 2006-2018 for different combinations of spot variance estimation methods and VIX sampling times (p-values $\leq 10^{-4}$ are reported as zero).

As in the affine framework (see Paragraph 5.3.2), we also analyse yearly subsamples in order to evaluate if coefficient estimates are sensitive to events of distress. The ensuing estimation results are shown in Table 5.6 in the Appendix. Note that the quadratic term is not statistically significant in 2006, 2008, 2012, 2014 and 2018. Based on Figure 5.1, panel b), and the detailed analysis in Horst and

Xu (2019), these years appear truly different, in terms of the state of the financial market. For instance, during 2008 and 2018, the VIX exhibits spikes, related, respectively, to the global financial crisis and the 'China-US trade war'. In contrast, 2006 and 2012 do not experience relevant economic events, see Horst and Xu (2019). The year 2014 represents an intermediate situation, where the VIX is almost flat until the end of October, when a cluster of spikes arises due to the end of quantitative easing policy by the Federal Reserve in the US. Thus, it is hard to attribute the statistical significance of the quadratic terms in Table 5.6 to events of financial distress.

Focusing on the frequency of price jumps in Figure 5.3, panel b), we highlight that 2006, 2014, and 2018 show a relatively low percentage of days with jumps. Thus, keeping in mind the result in Proposition 5.5, the non-significance of the quadratic coefficient in these years could be linked to the low percentage of days with jumps. The number of jumps in 2012 seems not coherent with this interpretation of the results in Table 5.6. Indeed, the quadratic term is not statistically significant in 2012, despite the fact that the percentage of days with jumps in 2012 is the second largest after 2011. However, 2012 can be deemed as an atypical year, in terms of market liquidity. In 2012 a series of important expansionary monetary policies were started by central banks to respond to the Euro-zone debt crisis and its international ramifications. These include the decision by the European Central Bank to cut its rates in multiple steps and to start a long-term refinancing operation (LTRO) during the first trimester of 2012, and the decision by the US Federal Open Market Committee to start a quantitative easing in September and to increase it in December of 2012. Thus, the year 2012 is characterized by an atypical number of positive jumps in response to this new paradigm of 'Infinity Quantitative Easing', that massively increased the market liquidity.

Polynomial framework

In the last paragraph, the polynomial framework (5.23) is analysed. Before fitting this model, we examine the sample correlation matrix of the regressors, which is shown in Table 5.3 for the case when the spot variance is reconstructed using the Fourier method and log-price daily opening values are used. Other combinations of daily opening/closing log-price series and Fourier/realized spot variance estimates yield sample correlation matrices which are almost identical to that in Table 5.3.

Table 5.3 provides empirical evidence of the existence of an almost perfect linear dependence between the log-price and its square, and between the spot variance and the product of the log-price and the spot variance. Moreover, the analysis conducted in Section 5.3.2 has already shown that the additional contribution of the squared spot variance estimates to the dynamics of the squared VIX is not signifi-

	$\widehat{V}_t$	$\widehat{V}_t^2$	X_t	X_t^2	$X_t\widehat{V}_t$
$\widehat{V}_t$	1				
$\widehat{V}_t^2$	0.839	1			
X_t	-0.370	-0.181	1		
X_t^2	-0.362	-0.177	0.999	1	
$X_t\widehat{V}_t$	0.999	0.834	-0.355	-0.347	1

Tab. 5.3: Sample correlation matrix of the regressors of the polynomial form (5.23) over the period 2006-2018.

cant. Thus, it remains only to evaluate the additional contribution of the log-price. This polynomial framework could then be associated with a *fully affine* form in both the log-return and the spot variance.

With regard to daily (opening and closing) log-prices series, recall that it is a well-known stylized fact that daily asset price series are non-stationary. Indeed, the Augmented Dickey Fuller test, performed at the 90% confidence level, confirms that our opening and closing log-price series have a unit root. To cope with non-stationarity, we estimate the coefficients in equation (5.23) after replacing log-prices with their detrended values, i.e., their values minus their sample mean. The estimation results are summarized in Table 5.4. Based on Table 5.4, the contribution of the log-price is not statistically significant at the 95% confidence level, but only at 90% level, independently of the spot variance estimator and the VIX sampling time used. Overall, this result confirms that the affine framework is sufficient to adequately fit our sample.

Finally, it is worth evaluating if the additional contribution of the price in explaining the dynamics of the VIX index squared is statistically significant on yearly subsamples, i.e., under different economic scenarios. The results of the year-by-year estimation are summarized in Table 5.7 in the Appendix and are in line with the whole-sample results.

	coeff.	estimate	std. err.	t stat.	p value	R^2
Fourier spot var. estimates & VIX opening quotes	$p_{0,0}$	0.015	0.004	4.089	0	0.959
	$p_{0,1}$	0.931	0.142	6.549	0	
	$p_{1,0}$	-0.018	0.010	-1.774	0.076	
Fourier spot var. estimates VIX closing quotes	$p_{0,0}$	0.016	0.003	5.132	0	0.960
	$p_{0,1}$	0.864	0.125	6.898	0	
	$p_{1,0}$	-0.018	0.010	-1.783	0.075	
Realized spot var. estimates VIX opening quotes	$p_{0,0}$	0.018	0.003	6.405	0	0.952
	$p_{0,1}$	0.899	0.114	7.909	0	
	$p_{1,0}$	-0.019	0.010	-1.875	0.061	
Realized spot var. estimates VIX closing quotes	$p_{0,0}$	0.017	0.004	4.686	0	0.953
	$p_{0,1}$	1.042	0.156	6.699	0	
	$p_{1,0}$	-0.018	0.010	-1.818	0.069	

Tab. 5.4: Polynomial (fully affine) framework: estimation from S&P500 data over the period 2006-2018 for different combinations of spot variance estimation methods and VIX sampling times (p-values $\leq 10^{-4}$ are reported as zero; $p_{1,0}$ indicates the coefficient of the detrended price).

5.4 Conclusions

This chapter provides empirical evidence that S&P500 data over the period 2006-2018 are coherent with the exponential affine framework, introduced by Kallsen et al. (2011), who analytically prove the existence of an affine relationship between the expected future variance, i.e., the variance swap rate, and the spot variance. This chapter collects empirical evidence that this affine relationship fits the data overwhelmingly well, with statistically significant coefficients and an R^2 larger than 0.95. Further, this chapter provides empirical evidence that the daily sampling used to compute the actual variance swap rates is frequent enough to erase the quadratic correction due to discrete sampling. The quadratic correction is expected within the polynomial framework, which includes the exponential affine framework as a special case. This empirical non-parametric result confirms the result by Broadie and Jain (2008), which was obtained on data simulated from four parametric models belonging to the exponential affine class. In general, note that the empirical findings of the chapter are basically insensitive to the volatility estimator used and the VIX sampling time, and this clearly robustifies their validity.

The chapter focuses also on yearly subsamples, in order to evaluate the sensitivity to events of financial distress. Empirical results on yearly subsamples are more nuanced. In particular, it emerges that the exponential affine framework could be rejected in 2008 and 2011. These two years include the outbreaks of two global financial crisis sparked, respectively, by the American housing market and the sovereign debt in the Euro area. Models in the exponential mean-reverting variance framework seem more adequate to capture the features of empirical data in those two years of financial distress. Farther, the significance of the quadratic coefficients in years with a relatively large number of price jumps supports the use of polynomial models in the presence of more frequent jumps.

5.5 Appendix

5.5.1 Proofs

Proof of Proposition 5.3

Here we adopt, where no ambiguity arises, the parameter notation introduced in Lemma 4.2 of Kallsen et al. (2011) for the exponential affine model. Using the equation for Ψ_u^V and Φ_u^V in Proposition 5.2 and differentiating the two equations with respect to u, we have

$$\frac{\partial^2 \Psi_u^V}{\partial u \partial t}(t) = \gamma_1^{11} \Psi_u^V(t) \frac{\partial \Psi_u^V(t)}{\partial u} + \beta_1^1 \frac{\partial \Psi_u^V(t)}{\partial u} + \gamma_1^{22} + \int_{\mathbb{R}^+ \times \mathbb{R}} \left(e^{x_1 \Psi_u^V(t) + u x_2^2} \left(x_1 \frac{\partial \Psi_u^V(t)}{\partial u} + x_2^2 \right) - \frac{\partial \Psi_u^V(t)}{\partial u} h(x_1) \right) \kappa_1(dx),$$

$$\frac{\partial \Phi_u^V(t)}{\partial u} = \int_0^t \left[\beta_0^1 \frac{\partial \Psi_u^V(s)}{\partial u} + \gamma_0^{22} + \int_{\mathbb{R}^+ \times \mathbb{R}} \left(e^{x_1 \Psi_u^V(s) + u x_2^2} \left(x_1 \frac{\partial \Psi_u^V(s)}{\partial u} + x_2^2 \right) - \frac{\partial \Psi_u^V(s)}{\partial u} h(x_1) \right) \kappa_0(dx) \right] ds\,.$$

Taking $u = 0$ and recalling that $\Psi_0^V(t) = 0$ for all t, we obtain the relations satisfied by (Ψ, Φ), that read

$$\frac{\partial \Psi}{\partial t}(t) = \beta_1^1 \Psi(t) + \gamma_1^{22} + \int_{\mathbb{R}^+ \times \mathbb{R}} \left(x_1 \Psi(t) + x_2^2 - \Psi(t) h(x_1) \right) \kappa_1(dx),$$

$$\Phi(t) = \int_0^t \left[\beta_0^1 \Psi(s) + \gamma_0^{22} + \int_{\mathbb{R}^+ \times \mathbb{R}} \left(x_1 \Psi(s) + x_2^2 - \Psi(s) h(x_1) \right) \kappa_0(dx) \right] ds\,.$$

Note that in our case $\gamma_0^{22} = 0$, see (5.8). Then, splitting the integrals and recalling that $h(x)$ is a truncating function, there exists non-negative parameters $(\widetilde{\beta}_1^1, \widetilde{\beta}_0^1)$, that is, the parameters associated with the truncating function $h(x) = x$, such that:

$$\frac{\partial \Psi}{\partial t}(t) = \widetilde{\beta}_1^1 \Psi(t) + \gamma_1^{22} + \int_{\mathbb{R}^+ \times \mathbb{R}} x_2^2 \, \kappa_1(dx),$$

$$\Phi(t) = \int_0^t \left[\widetilde{\beta}_0^1 \Psi(s) + \int_{\mathbb{R}^+ \times \mathbb{R}} x_2^2 \, \kappa_0(dx) \right] ds \,.$$

Note that Ψ solves a non-homogeneous linear differential equation with non-negative external term $\gamma_1^{22} + \int_{\mathbb{R}^+ \times \mathbb{R}} x_2^2 \, \kappa_1(dx)$. This term is zero if and only if $\gamma_1^{22} = 0$ and $\kappa_1(dx) = 0$, that is, in the case of the exponential Levy model. We deduce that $\Psi(s) > 0$ for all positive s, except for the exponential Levy model, for which volatility is constant and thus the stationary distribution is degenerate. We now turn to Φ and assume $\Psi(s) > 0$. Using the integral representation of Φ, we easily obtain that $\Phi > 0$ if and only if $\widetilde{\beta}_0^1 \neq 0$ or $\kappa_0(dx) \neq 0$, see also Filipovic (2001).

This is equivalent to assuming that the process V is a continuous-state branching process with immigration. Instead, in the case $\widetilde{\beta}_0^1 = 0$ and $\kappa_0(dx) = 0$, the process V is a continuous-state branching process without immigration. Without-immigration continuous-state branching processes do not have a stationary distribution, see Theorem 3.20 and Corollary 3.21 in Li (2011).

Proof of Proposition 5.4

The quadratic variation of x is rewritten as $[x]_t = [x^c]_t + \sum_{s \le t} (\Delta x_s)^2$, where x^c denotes the continuous part of the log-price x. According to (5.12), we have $[x^c]_t = \int_0^t v(s)ds$. It is easy to show that the variance process v is integrable using Gronwall's lemma, since the drift of the variance process is affine and Z is integrable by hypothesis. We now focus on the jump contribution, $\sum_{s \le t} (\Delta x_s)^2 = \sum_{s \le t} (\Delta J_s)^2$. Recalling that the process J is square-integrable, we obtain that the optional version of the quadratic variation $[x]_t$ is finite almost surely. Introducing the predictable version $\langle x \rangle_t$ of the quadratic variation and recalling that the optional and the predictable version of the quadratic variation differ by a martingale, we obtain that

$$E\left[[x]_\tau\right] = E\left[\langle x \rangle_\tau\right] = E\left[\int_0^\tau v(s)ds + \int \int_0^\tau \zeta^2 \lambda(d\zeta, ds)\right].$$

Considering first the jump term, exploiting that λ is affine in the variance process v, we deduce that the jump part is affine in the expectation of the integral of the variance process. Focusing now on the term $E\left[\int_0^\tau v(s)ds\right]$, we consider the integral version of the stochastic differential equation (5.12), i.e.,

$$v(t) - v(0) = Z_t + \theta \int_0^t (\alpha - v(s))ds.$$

Taking the expectation, we obtain that $E\left[\int_0^\tau v(s)ds\right] = \theta^{-1} \left(\alpha\theta\tau + v(0) - E\left[v(\tau)\right]\right)$ and $E\left[v(t)\right]$ satisfies a linear ODE. This result, combined with the previous result, proves that the expectation of the quadratic variation is an affine function of the initial spot variance.

Proof of Proposition 5.5

According to the characterization in Proposition 2.12 of Cuchiero et al. (2012), if (x,v) is a 2-polynomial process then

$$[x,x]_t = \int_0^t v(s)ds + \int_0^t \int \zeta^2 \lambda(d\zeta, ds) =: \int_0^t a(x(s), v(s))ds,$$

where $a \in \mathscr{P}_2$. Then, the result for $E^{(r,s)}[[x,x]_t]$ follows from Theorem 3.2 in Cuchiero (2011) and the application of the stochastic Fubini theorem.
In particular, if we consider the quadratic variation of x^c, together with the evolution (5.8), we have that $[x^c]_t = \int_0^t v(s)ds$. Taking the expectation and applying the stochastic Fubini theorem, we obtain $E\left[[x^c]_t\right] = \int_0^t E\left[v(s)\right]ds$. Now, using the hypothesis that (x,v) is 2-polynomial, we see that the function $f(r,s) = s \in \mathscr{P}_1$, and integrating we obtain the result.

Proof of Proposition 5.6

Using the definition of realized variance $RV^n_{[0,\tau]}$ given in (5.2), we take the expected value of RV^n. Due to the finiteness of the sum, the expected value of $RV^n_{[0,\tau]}$ gives

$$E^{(r,s)}\left[RV^n_{[0,\tau]}\right] = \sum_{k=1}^{n} E^{(r,s)}\left[\left(x(t_k) - x(t_{k-1})\right)^2\right].$$

Noting that the function inside the expectation belongs to $\mathscr{P}_2$, we have that $E\left[RV^n_{[0,\tau]}\right]$ belongs to $\mathscr{P}_2$.

5.5.2 Results of tests on yearly subsamples

year	coeff.	estimate	std. err.	t stat.	p value	R^2
2006	$p_{0,0}$	0.011	0.001	9.422	0	0.441
	$p_{0,1}$	0.473	0.173	2.728	0.006	
2007	$p_{0,0}$	0.014	0.004	3.944	0	0.659
	$p_{0,1}$	0.858	0.205	4.178	0	
2008	$p_{0,0}$	0.020	0.016	1.358	0.175	0.980
	$p_{0,1}$	0.932	0.177	5.293	0	
2009	$p_{0,0}$	0.031	0.008	3.847	0	0.863
	$p_{0,1}$	1.315	0.161	8.188	0	
2010	$p_{0,0}$	0.025	0.008	3.152	0.002	0.824
	$p_{0,1}$	0.936	0.314	2.986	0.003	
2011	$p_{0,0}$	0.014	0.013	1.091	0.275	0.911
	$p_{0,1}$	1.434	0.413	3.470	0.001	
2012	$p_{0,0}$	0.023	0.002	9.383	0	0.319
	$p_{0,1}$	0.550	0.124	4.461	0	
2013	$p_{0,0}$	0.016	0.001	13.223	0	0.495
	$p_{0,1}$	0.351	0.083	4.226	0	
2014	$p_{0,0}$	0.015	0.001	12.823	0	0.541
	$p_{0,1}$	0.486	0.215	2.264	0.024	
2015	$p_{0,0}$	0.015	0.006	2.467	0.014	0.887
	$p_{0,1}$	0.695	0.332	2.094	0.036	
2016	$p_{0,0}$	0.014	0.001	11.834	0	0.817
	$p_{0,1}$	0.599	0.058	10.303	0	
2017	$p_{0,0}$	0.011	0.001	9.000	0	0.248
	$p_{0,1}$	0.245	0.060	4.076	0	
2018	$p_{0,0}$	0.012	0.001	10.055	0	0.912
	$p_{0,1}$	0.753	0.061	12.178	0	

Tab. 5.5: Affine framework (5.24): estimation from S&P500 data over yearly subsamples when Fourier spot variance estimates and VIX opening quotes are used (p-values $\leq 10^{-4}$ are reported as zero).

year	coeff.	estimate	std. err.	t stat.	p value	R^2
2006	$q_{0,0}$	0.020	0.008	2.429	0.015	0.448
	$q_{0,1}$	8.286	4.572	1.812	0.070	
	$q_{0,2}$	-6.765	6.979	-0.969	0.332	
2007	$q_{0,0}$	0.028	0.001	23.010	0	0.670
	$q_{0,1}$	2.956	0.584	5.064	0	
	$q_{0,2}$	-4.384	0.652	-6.722	0	
2008	$q_{0,0}$	0.052	0.036	1.433	0.152	0.980
	$q_{0,1}$	2.842	0.348	8.159	0	
	$q_{0,2}$	0.723	1.237	0.585	0.559	
2009	$q_{0,0}$	0.056	0.002	23.510	0	0.872
	$q_{0,1}$	1.850	0.613	3.020	0.003	
	$q_{0,2}$	-3.859	0.799	-4.832	0	
2010	$q_{0,0}$	0.044	0.001	37.123	0	0.838
	$q_{0,1}$	0.309	0.142	2.178	0.029	
	$q_{0,2}$	-2.235	0.673	-3.323	0.001	
2011	$q_{0,0}$	0.044	0.036	1.233	0.218	0.930
	$q_{0,1}$	0.512	0.197	2.598	0.009	
	$q_{0,2}$	-3.944	1.206	-3.271	0.001	
2012	$q_{0,0}$	0.038	0.010	3.998	0	0.320
	$q_{0,1}$	1.842	0.405	4.552	0	
	$q_{0,2}$	0.119	2.924	0.041	0.968	
2013	$q_{0,0}$	0.013	0.004	3.667	0	0.496
	$q_{0,1}$	4.462	1.483	3.008	0.003	
	$q_{0,2}$	-3.784	1.047	-3.614	$< 10^{-3}$	
2014	$q_{0,0}$	0.064	0.017	3.806	0	0.739
	$q_{0,1}$	15.423	7.114	2.168	0.030	
	$q_{0,2}$	9.245	4.889	1.891	0.059	
2015	$q_{0,0}$	0.030	0.001	25.012	0	0.890
	$q_{0,1}$	0.340	0.042	8.086	0	
	$q_{0,2}$	-1.104	0.234	-4.713	0	
2016	$q_{0,0}$	0.028	0.001	23.007	0	0.867
	$q_{0,1}$	1.301	0.489	2.661	0.008	
	$q_{0,2}$	-2.629	1.039	-2.531	0.011	
2017	$q_{0,0}$	0.065	0.022	3.024	0.003	0.330
	$q_{0,1}$	33.004	8.664	3.809	0	
	$q_{0,2}$	-22.721	5.791	-3.923	0	
2018	$q_{0,0}$	0.035	0.004	9.667	0	0.914
	$q_{0,1}$	2.903	1.430	2.029	0.042	
	$q_{0,2}$	0.453	1.245	0.364	0.716	

Tab. 5.6: Quadratic framework (5.26): estimation from S&P500 data over yearly subsamples when Fourier spot variance estimates and VIX opening quotes are used (p-values $\leq 10^{-4}$ are reported as zero).

year	coeff.	estimate	std. err.	t stat.	p value	R^2
2006	$p_{0,0}$	0.011	0.005	2.248	0.024	0.601
	$p_{0,1}$	0.288	0.126	2.284	0.022	
	$p_{1,0}$	-0.052	0.023	-2.261	0.024	
2007	$p_{0,0}$	0.019	0.005	3.997	0	0.670
	$p_{0,1}$	0.855	0.196	4.365	0	
	$p_{1,0}$	0.065	0.037	1.741	0.082	
2008	$p_{0,0}$	0.023	0.013	1.726	0.084	0.980
	$p_{0,1}$	0.602	0.191	3.148	0.002	
	$p_{1,0}$	-0.387	0.101	-3.830	0	
2009	$p_{0,0}$	0.046	0.016	2.921	0.004	0.909
	$p_{0,1}$	0.874	0.226	3.869	0	
	$p_{1,0}$	-0.197	0.049	-3.997	0	
2010	$p_{0,0}$	0.024	0.008	2.855	0.004	0.829
	$p_{0,1}$	0.514	0.179	2.870	0.004	
	$p_{1,0}$	-0.180	0.052	-3.486	0.001	
2011	$p_{0,0}$	0.067	0.040	1.696	0.090	0.934
	$p_{0,1}$	0.681	0.302	2.257	0.024	
	$p_{1,0}$	-0.483	0.133	-3.619	$< 10^{-3}$	
2012	$p_{0,0}$	0.017	0.006	2.798	0.005	0.394
	$p_{0,1}$	0.259	0.082	3.174	0.002	
	$p_{1,0}$	0.004	0.004	0.999	0.317	
2013	$p_{0,0}$	0.016	0.001	12.989	0	0.496
	$p_{0,1}$	0.356	0.082	4.349	0	
	$p_{1,0}$	0.004	0.005	0.749	0.453	
2014	$p_{0,0}$	0.017	0.005	3.497	0.001	0.552
	$p_{0,1}$	0.461	0.225	2.052	0.040	
	$p_{1,0}$	-0.014	0.024	-0.600	0.549	
2015	$p_{0,0}$	0.121	0.019	6.307	0	0.894
	$p_{0,1}$	0.141	0.036	3.897	0	
	$p_{1,0}$	-0.366	0.069	-5.346	0	
2016	$p_{0,0}$	0.048	0.010	4.996	0	0.900
	$p_{0,1}$	0.420	0.048	8.743	0	
	$p_{1,0}$	-0.106	0.033	-3.257	0.001	
2017	$p_{0,0}$	0.022	0.004	5.995	0	0.470
	$p_{0,1}$	0.186	0.058	3.227	0.001	
	$p_{1,0}$	-0.025	0.007	-3.497	0.001	
2018	$p_{0,0}$	0.065	0.025	2.569	0.010	0.913
	$p_{0,1}$	0.587	0.089	6.603	0	
	$p_{1,0}$	-0.090	0.043	-2.082	0.037	

Tab. 5.7: Polynomial (fully affine) framework: estimation from S&P500 data over yearly subsamples when Fourier spot variance estimates and VIX opening quotes are used (p-values $\leq 10^{-4}$ are reported as zero; $p_{1,0}$ indicates the coefficient of the detrended price).

6. CONCLUSIONS: MAIN RESULTS OF THE THESIS

The main findings of this thesis can be summarized as follows. For what regards analytical results, the main contribution of this thesis is three-fold. First, we derive a rate-efficient central limit theorem for the non-parametric Fourier estimator of the integrated leverage (see Chapter 2). Secondly, in a level-dependent setting, we derive explicit expressions for the derivatives of the leverage process with respect to the price and the volatility as the ratios of quantities that can be consistently estimated from sample prices over a fixed time horizon, that is, as the ratios of the price-leverage covariation and, respectively, the volatility and the leverage (see Chapter 3). Finally, we uncover the sources of finite-sample bias that affect the simple realized vol-of-vol estimator, the PSRV by Barndorff-Nielsen and Veraart (2009), and provide a feasible rule to optimize such bias, based on the overlapping of consecutive local windows for pre-estimating the spot volatility (see Chapter 4), when the volatility is a process in the CKLS class (see Chan et al. (1992)).

Additionally, this thesis provides three contributions to the empirical literature on financial econometrics, based on the study of S&P500 data over the period 2006-2018. First, we show that the inclusion of an extra term that accounts for the leverage effect increases the explanatory power of the HAR volatility model by Corsi (2009) in a statistically significant manner, thereby extending and robustifying the empirical results in Mykland and Wang (2014), which are based on a simple auto-regressive model of order 2 (see Chapter 2). Further, we collect empirical evidence that suggests to interpret the price-leverage covariation as a measure of the responsiveness of the leverage to price and volatility changes (see Chapter 3). Finally, we collect overwhelming empirical evidence that supports the use affine stochastic volatility models in financial applications (see Chapter 5).

BIBLIOGRAPHY

Abi Jaber, E., Cuchiero, C., Larsson, M., and Pulido, S. (2019). A weak solution theory for stochastic volterra equations of convolution type. *arXiv preprint arXiv:1909.01166*.

Ackerer, D. and Filipovic, D. (2020). Linear credit risk models. *Finance and Stochastics*, 24(1):169–214.

Ackerer, D., Filipovic, D., and Pulido, S. (2018). The jacobi stochastic volatility model. *Finance and Stochastics*, 22(3):667–700.

Aït-Sahalia, Y., Fan, J., Laeven, R. J. A., Wang, C. D., and Yang, X. (2017). Estimation of the continuous and discontinuous leverage effects. *Journal of the American Statistical Association*, 112(520):1744–1758.

Aït-Sahalia, Y., Fan, J., and Li, Y. (2013). The leverage effect puzzle: disentangling sources of bias at high frequency. *Journal of Financial Economics*, 109:224–249.

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517.

Aït-Sahalia, Y. and Jacod, J. (2014). High-frequency financial econometrics. *Princeton University Press*.

Aït-Sahalia, Y., Karaman, M., and Mancini, L. (2018). The term structure of variance swaps and risk premia. *Swiss Finance Institute Research Paper n.18-37*.

Aït-Sahalia, Y. and Xiu, D. (2019). A hausman test for the presence of noise in high frequency data. *Journal of Econometrics*, 211(1):176–205.

Alvarez, A., Panloup, F., Pontier, M., and Savy, N. (2011). Estimation of the instantaneous volatility. *Statistical Inference for Stochastic Processes*, 15:27–59.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905.

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

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

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

Araneda, A. (2020). The fractional and mixed-fractional cev model. *Journal of Computational and Applied Mathematics*, 363(1):106–123.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5):2003–2049.

Baldeaux, J. and Rutkowski, M. (2010). Static replication of forward-start claims and realized variance swaps. *Applied Mathematical Finance*, 17(2):99–131.

Bandi, F. and Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics*, 169(1):94–113.

Bandi, F. M. and Renò, R. (2016). Price and volatility co-jumps. *Journal of Financial Economics*, 119(1):107–146.

Bandi, F. M. and Renò, R. (2018). Nonparametric stochastic volatility. *Econometric Theory*, 34(6):1–49.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008a). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(7):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008b). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society*, 63(2):167–241.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. E. and Veraart, A. (2009). Stochastic volatility of volatility in continuous time. *CREATES Research Paper*, No. 2009-25.

Barucci, E., Malliavin, P., Mancino, M. E., Renó, R., and Thalmaier, A. (2003). The price volatility feedback rate: an implementable mathematical indicator of market stability. *Mathematical Finance*, 13(1):17–35.

Barucci, E. and Mancino, M. E. (2010). Computation of volatility in stochastic volatility models with high-frequency data. *International Journal of Theoretical and Applied Finance*, 13(5):767–787.

Bates, D. S. (1996). Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69–107.

Beckers, S. (1980). The constant elasticity of variance model and its implications for option pricing. *The Journal of Finance*, 35(3):661–673.

Benth, F. E. (2011). The stochastic volatility model of barndorff-nielsen and shephard in commodity markets. *Mathematical Finance*, 21(4):595–625.

Benth, F. E., Groth, M., and Kufakunesu, R. (2007). Valuing volatility and variance swaps for a non-gaussian ornstein–uhlenbeck stochastic volatility model. *Applied Mathematical Finance*, 14(4):347–363.

Bernard, C. and Cui, Z. (2014). Prices and asymptotics for discrete variance swaps. *Applied Mathematical Finance*, 21(2):140–173.

Bernis, G., Brignone, R., Scotti, S., and Sgarra, C. (2019). A gamma ornstein-uhlenbeck model driven by a hawkes process. *SSRN paper n.3370304*.

Bibinger, M. and Winkelmann, L. (2018). Common price and volatility jumps in noisy high-frequency data. *Electronic Journal of Statistics*, 12(1):2018–2073.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association*, pages 171–181.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11):4463–4492.

Bollerslev, T. and Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109:33–65.

Brennan, M. and Schwartz, E. (1980). Analyzing convertible securities. *Journal of Financial and Quantitative Analysis*, 15(4):907–929.

Broadie, M. and Jain, A. (2008). The effects of jumps and discrete sampling on volatility and variance swaps. *International Journal of Theoretical and Applied Finance*, 11(8):761– 797.

Callegaro, G., Fiorin, L., and Pallavicini, A. (2017). Quantization goes polynomial. *ArXiv paper n.1710.11435*.

Carr, P. and Lee, R. (2008). Robust replication of volatility derivatives. *SSRN paper n.1108429*.

Carr, P. and Sun, J. (2007). A new approach for option pricing under stochastic volatility. *Review of Derivatives Research*, 10(2):87–150.

Carr, P. and Wu, L. (2006). A tale of two indices. *The Journal of Derivatives*, 13(3):13–29.

Carr, P. and Wu, L. (2008). Variance risk premiums. *The Review of Financial Studies*, 22(3):1311–1341.

CBOE (2019). Vix white paper. *Available at: www.cboe.com/micro/vix/vixwhite.pdf*.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47(3):1209–1227.

Christoffersen, P., Jakobs, K., and Mimouni, K. (2010). Volatility dynamics for the s&p500: evidence from realized volatility, daily returns and option prices. *The Review of Financial Studies*, 23(8):3141–3189.

Clement, E. and Gloter, A. (2011). Limit theorems in the fourier transform method for the estimation of multivariate volatility. *Stochastic Processes and their Applications*, 121(5):1097–1124.

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

Corsi, F., Pirino, D., and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*, 30(3):368–380.

Cox, J., Ingersoll, J., and Ross, S. (1980). An analysis of variable rate loan contracts. *The Journal of Finance*, 35:389–403.

Cox, J., Ingersoll, J., and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Cuchiero, C. (2011). Affine and polynomial processes. *Ph.D. thesis (ETH Zurich)*.

Cuchiero, C. (2018). Polynomial processes in stochastic portfolio theory. *Forthcoming in Stochastic processes and their applications*.

Cuchiero, C., Keller-Ressel, M., and Teichmann, J. (2012). Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740.

Cuchiero, C., Larsson, M., and Svaluto-Ferro, S. (2018). Polynomial jump-diffusions on the unit simplex. *Annals of Applied Probability*, 28(4):2451–2500.

Cuchiero, C. and Teichmann, J. (2013). Path properties and regularity of affine processes on general state spaces. *in Séminaire de Probabilités XLV, Springer*.

Cuchiero, C. and Teichmann, J. (2015). Fourier transform methods for pathwise covariance estimation in the presence of jumps. *Stochastic Processes and Their Applications*, 125(1):116–160.

Curato, I. (2019). Estimation of the stochastic leverage effect using the fourier transform method. *Stochastic Processes and their Applications*, 129(9):3207–3238.

Curato, I. and Sanfelici, S. (2019). Stochastic leverage effect in high-frequency data: a fourier-based analysis. *Working paper available at arXiv:1910.06660*.

Curato, I. V., Mancino, M. E., and Recchioni, M. C. (2018). Spot volatility estimation using the laplace transform. *Econometrics and Statistics*, 6:22–43.

Danielsson, J. and Zigrand, J. P. (2006). On time-scaling of risk and the square-root-of-time rule. *Journal of Banking and Finance*, 30:2701–2713.

Dawson, D. A. and Li, Z. (2006). Skew convolution semigroups and affine markov processes. *The Annals of Probability*, 34(3):1103–1142.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1):463–520.

Derman, E. and Kani, I. (1994). Riding on the smile. *RISK*, 7:32–39.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–443.

Duffie, D., Filipovic, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, 68(6):1343–1376.

Dupire, B. (1994). Pricing with a smile. *RISK*, 7:18–20.

Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of Finance*, 59(3):1367–1403.

Fan, J. and Wang, Y. (2008). Spot volatility estimation for high-frequency data. *Statistics and Its Interface*, 1:279–288.

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182.

Filipovic, D. (2001). A general characterization of one factor affine term structure models. *Finance and Stochastics*, 5:389–412.

Filipovic, D., Gourier, E., and Mancini, L. (2016). Quadratic variance swap models. *Journal of Financial Economics*, 119(1):44–68.

Filipovic, D. and Mayerhofer, E. (2009). Affine diffusion processes: theory and applications. *Advanced financial modelling*, 8:1–40.

Foster, D. P. and Nelson, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica*, 64(1):139–174.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19:3–29.

Gatheral, J. and Oomen, R. C. A. (2010). Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14(2):249–283.

Goard, J. and Mazur, M. (2013). Stochastic volatility models and the pricing of vix options. *Mathematical Finance*, 23(3):439–458.

Goutte, S., Ismail, A., and Pham, H. (2017). Regime-switching stochastic volatility model: estimation and calibration to vix options. *Applied Mathematical Finance*, 24(1):38–75.

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120.

Grasselli, M. (2016). The 4/2 stochastic volatility model: a unified approach for the heston and the 3/2 model. *Mathematical Finance*, 27(4):1013–1034.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. *The Best of Wilmott*, 1:249–296.

Hasbrouck, J. (2007). Empirical market microstructure: the institutions, economics, and econometrics of securities trading. *Oxford University Press*.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Hobson, D. and Rogers, L. (1998). Complete models with stochastic volatility. *Mathematical Finance*, 8:27–48.

Holland, W. P. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6(9):813–827.

Horst, U. and Xu, W. (2019). The microstructure of stochastic volatility models with self exciting jump dynamics. *ArXiv paper n.1911.12969*.

Huang, D., Schlag, C., Shaliastovich, I., and Thimme, J. (2018). Volatility-of-volatility risk. *Journal of Financial and Quantitative Analysis*, page 1–63.

Hubalek, F., Keller-Ressel, M., and Sgarra, C. (2017). Geometric asian option pricing in general affine stochastic volatility models with jumps. *Quantitative Finance*, 17(6):873–888.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300.

Jacod, J. (1994). Limit of random measures associated with the increments of a brownian semimartingale. *Manuscript, Institute de Mathematiques de Jussieu, Universite Pierre et Marie Curie, Paris*, 1.

Jacod, J. (1997). On continuous conditional gaussian martingales and stable convergence in law. *Seminaire de Probabilites XXXI, Springer*, 1:232–246.

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

Jacod, J., Li, Y., and Zheng, X. (2017). Statistical properties of microstructure noise. *Econometrica*, 85:1133–1174.

Jacod, J. and Todorov, V. (2010). Do price and volatility jump together? *Annals of Applied Probability*, 20(4):1425–1469.

Jarrow, R., Kchia, Y., Larsson, M., and Protter, P. (2013). Discretely sampled variance and volatility swaps versus their continuous approximations. *Finance and Stochastics*, 17(2):305–324.

Jiao, Y., Ma, C., and Scotti, S. (2017). Alpha-cir model with branching processes in sovereign interest rate modeling. *Finance and Stochastics*, 21(3):789–813.

Jiao, Y., Ma, C., Scotti, S., and Zhou, C. (2019). The alpha-heston stochastic volatility model. *https://arxiv.org/abs/1812.01914*.

Kallsen, J., Muhle-Karbe, J., and Voss, M. (2011). Pricing options on variance in affine stochastic volatility models. *Mathematical Finance*, 21(4):627–641.

Kalnina, I. and Xiu, D. (2017). Nonparametric estimation of the leverage effect: a trade-off between robustness and efficiency. *Journal of the American Statistical Association*, 112(517):384–399.

Kanaya, S. and Kristensen, D. (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory*, 32:861–916.

Keller-Ressel, M. (2011). Moment explosions and long-term behavior of affine stochastic volatility models. *Mathematical Finance*, 21(1):73–98.

Keller-Ressel, M., Schachermayer, W., and Teichmann, J. (2011). Affine processes are regular. *Probability Theory and Related Fields*, 151(3-4):591–611.

Kyle, A. S. and Obizhaeva, A. (2017). Dimensional analysis and market microstructure invariance. *Working Paper w0234, New Economic School (NES)*.

Lee, S. and Mykland, P. A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563.

Li, Z. (2011). Measure-valued branching markov processes. *Springer*.

Li, Z. and Ma, C. (2008). Catalytic discrete state branching models and related limit theorems. *Journal of Theoretical Probability*, 21(4):936–965.

Malliavin, P. and Mancino, M. E. (2002). Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6(1):49–61.

Malliavin, P. and Mancino, M. E. (2009). A fourier transform method for nonparametric estimation of multivariate volatility. *The Annals of Statistics*, 37(4):1983–2010.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36:270–296.

Mancino, M. E. and Recchioni, M. C. (2015). Fourier spot volatility estimator: asymptotic normality and efficiency with liquid and illiquid high-frequency data. *PLOS ONE*, 10(9).

Mancino, M. E., Recchioni, M. C., and Sanfelici, S. (2017). Fourier-malliavin volatility estimation: Theory and practice. *Spinger*.

Mancino, M. E. and Sanfelici, S. (2008). Robustness of fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics and Data Analysis*, 52(6):2966–2989.

Martin, I. (2011). Simple variance swaps. *NBER Working Paper n.16884*.

Mykland, P. and Zhang, L. (2006). Anova for diffusions. *Annals of Statistics*, 34(4):1931–1963.

Mykland, P. A. and Wang, D. (2014). The estimation of leverage effect with high-frequency data. *Journal of the American Statistical Association*, 109(505):197–215.

Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445.

Narula, S. C. (1979). Orthogonal polynomial regression. *International Statistical Review*, 47(1):31–36.

Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics*, 45(2):7–38.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Platen, E. (1997). A non-linear stochastic volatility model. *Financial Mathematics Research Report No.FMRR005-97, Center for Financial Mathematics, Australian National University*.

Podolskij, M. and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39:1127–1139.

Sanfelici, S., Curato, I. V., and Mancino, M. E. (2015). High frequency volatility of volatility estimation free from spot volatility estimates. *Quantitative Finance*, 15(8):1331–1345.

Sanfelici, S. and Mancino, M. E. (2020). Identifying financial instability conditions using high frequency data. *Journal of Economic Interaction and Coordination*, 15:221–242.

Smith, E., Farmer, D. J., Gillemot, L., and Krishnamurthy, S. (2003). Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514.

Tanaka, H. (1979). Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Mathematical Journal*, 9(1):163–177.

Todorov, V. (2010). Variance risk-premium dynamics: the role of jumps. *The Review of Financial Studies*, 23(1):345–383.

Todorov, V. and Tauchen, G. (2012). The realized laplace transforms of volatility. *Econometrica*, 80(3):1105–1127.

Veraart, A. E. D. and Veraart, L. A. M. (2012). Stochastic volatility and stochastic leverage. *Annals of Finance*, 8(23):205–223.

Vetter, M. (2015). Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. *Bernoulli*, 21(4):2393–2418.

Wilmott, P. (2000). Derivatives. *John Wiley and Sons*.

Zhang, L., Mykland, P. A., and 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(472):1394–1411.

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