SCUOLA
NORMALE
SUPERIORE

Stochastic models for financial time series: modelling, estimation and option pricing

PhD Thesis in
in
Financial Mathematics

Presented by
Giulia Livieri

Supervisors
Giacomo Bormetti
Stefano Marmi

Introduction

A European option is a financial instrument that gives the right, but not the obligation, to buy a certain amount of an underlying instrument at a given price, the strike price, at a pre-specified future date, the exercise time. The underlying instrument can be a security such as a bond or an equity share, an exchange rate, a commodity, or sometimes another financial product. The pay-off of a European option – which is specified in the contract – may depend on the value of the underlying instrument either at the exercise time or over the entire life of the option. In the latter case the option is named path-dependent. The theory of option pricing aims at determining the fair price of the contract. Pricing a financial derivative typically consists in solving four main issues: (i) The choice of a dynamics for the underlying instrument under the physical measure $\mathbb{P}$; (ii) the estimation of the parameters of the chosen dynamics; (iii) the definition of a pricing mechanism described by a stochastic discounting[1]; (iv) the development of a fast, accurate, and possibly flexible numerical method to determine the fair price of the instrument.

In modelling the dynamics of the underlying instruments under the physical measure $\mathbb{P}$, a common dilemma arises as models which describe the historical dynamics of the asset price with adequate realism are usually unable to precisely describe the entire dynamics of the implied volatility surface.

Traditionally, due to the mathematical tractability, the option pricing problem has been approached by describing the asset price dynamics in continuous time in terms of a Stochastic Differential Equation (SDE). In this SDE the diffusive coefficient, i.e. the volatility – which is the most important ingredient of any underlying instrument's dynamics – can be either a constant (Black and Scholes, 1973), a deterministic function of time and price (Dupire, 1994) as in local volatility models, or a stochastic process itself (Heston, 1993; Bates, 1996b; Gatheral et al., 2014) as in stochastic volatility models. Although the legitimate use of local volatility models for the description of the asset dynamics is highly questionable, they have the merit of being flexible and allowing for the exact reproduction of the volatility smiles implied by the market. For this reason, local volatility models are also employed in practice to evaluate path-dependent options whenever smile risk dominates path-dependency risk, e.g. when pricing Asian options[2]. On the contrary, standard[3]

[1] We note that it is possible to approach the option pricing problem by starting with the description of the dynamics of the underlying instrument either under the historical measure $\mathbb{P}$ or the risk-neutral one $\mathbb{Q}$. This choice depends, for instance, on the modelling needs. In particular, only in the former case it is necessary to describe a stochastic discounting. In this thesis we start with the description of the dynamics under the measure $\mathbb{P}$.

[2] For instance, the pay-off of a fixed strike Asian call is given by $\max(A(0,T) - K, 0)$, where $A(0,T)$ is the arithmetic average of the values of the underlying instrument over $[0,T]$, and T and K are the maturity and the strike of the option, respectively.

[3] In this thesis we will name standard a stochastic volatility model in which the driving noise of the SDE describing the volatility dynamics is a standard Brownian Motion.

stochastic volatility models are able to reproduce well-established stylized facts like the negative correlation of returns and volatility, and volatility smile. However, drawbacks arise since standard stochastic volatility models fail to calibrate the steep smile shapes usually found at short maturities together with their slow decrease for longer maturities, unless we complicate them with *ad hoc* time-dependent and/or state-dependent parameter choices, as in stochastic local volatility models (Ren et al., 2007) or in stochastic jump diffusion models (Bates, 1996b). Recently (see, for instance Gatheral et al., 2014; Bayer et al., 2016), it has been shown that rough[4] stochastic volatility models are able to reproduce both most of the well-stylized facts of financial time series and of option prices data, one of the most important being the implied volatility term structure.

However, models for the underlying price dynamics under the historical (or physical) measure $\mathbb{P}$ have mainly been constructed in discrete time. Here, the latent conditional variance of the returns can either be a function of past lagged values of the conditional variance itself and-or of squared returns as in the time-varying volatility models of the ARCH-GARCH families (Engle, 1982; Bollerslev, 1986; Glosten et al., 1993; Nelson, 1991), or made observable through a non-parametric measure of volatility constructed with high-frequency data as in the so called Realized Volatility (RV) approach, or a stochastic process itself (Taylor, 1994, 2007). Recently, either within a GARCH framework (Shephard and Sheppard, 2010; Hansen et al., 2012) or in a stochastic volatility one (Takahashi et al., 2009; Dobrev and Szerszen, 2010; Koopman and Scharth, 2013), another approach has been developed, consisting in adding a measurement equation relating the realized measure of volatility to the conditional variance of returns. It has been shown (see, for instance Hansen et al., 2012, and references therein) that this leads to a substantial improvement in the empirical fit of financial time series over the relative incomplete models, i.e. the models with only one measurement equation. Regarding the fitting of the dynamics of the volatility surface, a discrete-time volatility model should be able to incorporate volatility persistence – periods of high volatility tend to be followed by periods of high volatility– and leverage effect – the mechanism producing the asymmetric impact of positive and negative past returns on future volatility. In particular, the persistence of the volatility process is caused by the heterogeneity of agents acting in the market – the fact that different types of agents operate in the market, with different beliefs, risk profiles and degree of information (Müller et al., 1997). Moreover, it plays a crucial role in the correct fitting of the long-term part of the implied volatility surface.
It turns out that option pricing models based on realized volatility provide good performance (Corsi et al., 2013; Christoffersen et al., 2014; Majewski et al., 2015). On the other hand, stochastic volatility models offer increased flexibility in incorporating the above cited features with respect to the other class of models since they assume separate innovations for the conditional mean and the conditional variance of the returns. However, little work has been devoted to combine the stochastic volatility literature with that on option pricing to construct stochastic volatility option pricing models. In particular, we are not aware of any analytical discrete-time option pricing model combining stochastic volatility with realized measures.

Regarding the second issue (i.e. the estimation of the parameters of the chosen dynamics), in a continuous-time setting the parameters of the model are usually calibrated directly to the option prices. In a discrete-time framework, instead, once the volatility is filtered (as in GARCH and in stochastic volatility models) or proxied non-parametrically

[4]In this thesis we will name rough a stochastic volatility model in which the driving noise of the SDE describing the volatility dynamics is a fractional Brownian motion with Hurst parameter $H < 1/2$.

(as in RV models), the parameters are estimated directly from the time series of observed quantities. This is true even when the above cited features are taken into account, which has been shown to be crucial for an option pricing perspective (see, among others Majewski et al., 2015). The autoregressive structure characterizing GARCH-ARCH specifications has led these models to guide the way of measuring and forecasting volatility, since it facilitates maximum-likelihood estimation. Realized Volatility models are even more straightforward to estimate since the volatility is an observable quantity and thus there is no need of any filtering procedure. In stochastic volatility models, instead, the presence of variance-specific disturbances comes to the use of latent variables and this calls for the usage of suitable inference tools such as stochastic filtering and simulation based inference. So, in order to take advantage of the flexibility offered from a stochastic volatility model, it is crucial to have effective numerical procedures for its estimation.

Regarding the third issue, a pricing mechanism should take into account the uncertainty about the future level of asset volatility (see, for instance, Schwert, 2011). It is not so surprising that an investor will demand a compensation for bearing this risk. In financial literature, this reward is named variance risk premium (Bakshi and Kapadia, 2003) and quantifies the difference between conditional expectations of future volatility computed under the physical measure $\mathbb{P}$ and the risk-neutral one $\mathbb{Q}$. The need for a general pricing kernel incorporating a variance premium has been well documented in Christoffersen et al. (2013). Indeed, this premium is able to explain several puzzles concerning the consistency between the time series properties of the underlying asset prices and the distributions implicit in option prices (this unconsistency was considered by Bates, 1996a, one of central empirical issue in option research). Importantly, the pricing kernel in Christoffersen et al. (2013) is monotonic in returns and also in volatility, and its projection onto the stock price return alone is U-shaped, as the one observed in market data. In order to take under consideration the variance risk premium, the stochastic discounting should be multi-dimensional. An important class of multi-dimensional stochastic discount factor (SDF) is - in the state variables - the exponential affine one (see Bertholon et al., 2008; Gagliardini et al., 2011; Corsi et al., 2013; Alitab et al., 2015; Majewski et al., 2015), where the different sources of risk are clearly identified and compensated with separated risk-premia. Remarkably, the exponential-affine specification is well suited to the usage of the Laplace Transform (see Darolles et al., 2006; Gourieroux et al., 2006; Gouriéroux et al., 2009, to cite only a few) which permits to express the change of measure from $\mathbb{P}$ to $\mathbb{Q}$ in closed form.

Finally, concerning the problem of the development of fast, accurate, and flexible numerical method to determine the fair price of a financial derivative, three are the main methodologies that are employed in practice: i) Numerical methods based on Fourier transform techniques, ii) numerical schemes for solving the pricing Partial Differential Equation (PDE), iii) Monte Carlo simulations. The first methodology represents a very effective tool when the characteristic function of the considered underlying dynamics is available. It turns out that a popular method based on Fourier-cosine expansion for pricing non path-dependent European options is the COS one (Fang and Oosterlee, 2008). The PDE approach (see Wilmott et al., 1993; Alziary et al., 1997; Zvan et al., 2000; Pascucci, 2011, to cite only a few), which requires the definition of a pay-off specific pricing equation, is computationally efficient especially if we deal with a PDE in one space dimension[5].

[5]We note that, in general, to determine the price of an Asian option a PDE in two space dimensions has to be solved (see Wilmott et al., 1993). In Rogers and Shi (1995) a variable reduction was used to find a PDE in one space dimension for the value of an Asian claim.

Nonetheless, some options with exotic pay-off and exercise rules are subtle to price even within the Black, Scholes, and Merton framework (see Dewynne and Shaw, 2008; Siyanko, 2012). The Monte Carlo approach (Glasserman, 2004) is a very flexible method to simulate random trajectories of the price dynamics which permits to achieve any degree of precision – at least in theory – by simply increasing the number of generated paths. Nevertheless, it suffers from some ineffectiveness – especially when pricing out-of-the-money (OTM) options – because a relevant number of sampled paths may not contribute to the option pay-off. So, in order to take advantage of the flexibility offered from the Monte Carlo approach, it is important to combine several numerical techniques to reduce the variance of the price estimator.

The rest of the thesis is organized as follows.

Chapter 1 In this chapter we give an overview on the continuous and discrete time approaches to asset prices modelling.

Chapter 2 Here we introduce in more detail the themes and research questions that we address in this thesis. We provide some context and we highlight the most relevant contributions.

Chapter 3 This chapter contains the first original contribution of this thesis. We build on the work of Gatheral et al. (2014) to revisit their findings that spot volatilities can be very well modelled by rough stochastic volatility models with an Hurst exponent around 0.10. Using implied-volatility based approximations of the spot volatility – instead of high-frequency volatility estimations from historical price data as in Gatheral et al. (2014) – we find an Hurst parameter of order 0.30. We show analytically that this upward bias is due to a smoothing effect intrinsic to the chosen proxies. So, we give an additional argument for the usage of rough volatility models for applications, notably for option pricing and hedging (Gatheral et al., 2014; Bayer et al., 2016).

Chapter 4 In this chapter we turn to the modelling of discrete time financial time series, in particular high-frequency returns and Realized Volatilities. We introduce a new family of flexible and tractable discrete-time stochastic volatility models which allows to filter the latent conditional variance process from both observed returns and realized volatility measures. We develop an effective Bayesian estimation procedure for both parameters and latent quantities. We show that the proposed modelling approach accurately reproduces well-established stylized fact observed in financial time series, while preserving closed-form formulas for option prices and closed-form filtering and smoothing for the latent process. So, we propose a procedure that is not only innovative from both modelling and estimation perspectives but it is also of substantial interest in financial applications such as option pricing.

Chapter 5 This last chapter contributes to the construction of fast, accurate, and possible flexible numerical method to determine the fair price of a financial instrument. We develop a novel algorithm – named Backward Monte Carlo – to exotic option pricing which relies on the construction of a discrete multinomial tree. We use two ways to characterize the tree. The first one is inspired by the finite difference approximation of the diffusion infinitesimal generator (Albanese, 2007). The second method relies on the concept of optimal state-partitioning of a random variable and employs the recent technique termed Recursive Marginal Quantization Algorithm (RMQA) of Pagés and Sagna (2015). First, we fix some flaws of the RMQA which have been highlighted in the existing

literature. Second, we show that our proposal may be relevant from an applied perspective by assessing its reliability with respect to competitor Monte Carlo methods.

The last three chapters contain the original contributions of this thesis. Each of them is self-contained and in principle can be read separately.

Contents

Introduction **3**

List of publications **11**

1 Price dynamics model in continuous and discrete time **15**
1.1 Price dynamics models in continuous time 15
1.2 Price dynamics model in discrete time 18

2 Presented research **21**
2.1 Outline 22
2.2 Rough volatility: evidence from option prices 23
2.2.1 A proxy for the spot volatility 23
2.2.2 Regularity of the volatility 23
2.2.3 On the upward bias of the Hurst parameter 25
2.3 A discrete-time stochastic volatility framework for pricing options with realized measure 26
2.3.1 Dynamics under the physical measure $\mathbb{P}$ 26
2.3.2 Estimation of the model under the physical measure $\mathbb{P}$ 27
2.3.3 Risk neutral dynamics and financial applications 28
2.4 A backward Monte Carlo algorithm 29
2.4.1 Main ideas of the backward Monte Carlo algorithm 29
2.4.2 Recovering the transition probabilities 30
2.4.3 Financial applications 30

3 Rough volatility: evidence from option prices **33**
3.1 Introduction 33
3.2 A simple spot volatility proxy 35
3.2.1 Data description 35
3.2.2 Scaling property 35
3.2.3 Distribution of log-volatility increments 37
3.3 A refined spot volatility proxy 38
3.3.1 Data description and processing 39
3.3.2 The Medvedev-Scaillet correction formula 40
3.3.3 The scaling property revisited 41
3.4 On the upward bias of the Hurst parameter 43
3.4.1 Monte Carlo study 43
3.4.2 Analytical illustration of the upward bias 46
3.5 Conclusion 48

4 A discrete-time stochastic volatility framework for pricing options with realized measure 49
4.1 Introduction . . . 49
4.2 The model . . . 52
4.2.1 Dynamics under physical probability $\mathbb{P}$ . . . 52
4.2.2 Dynamics under risk-neutral probability $\mathbb{Q}$ . . . 57
4.3 Particular cases . . . 59
4.3.1 The SV-ARG model . . . 59
4.3.2 The SV-LHARG model . . . 60
4.4 Estimation of models under physical probability measure . . . 61
4.5 Simulation study . . . 64
4.5.1 Experiment design . . . 64
4.5.2 Results . . . 65
4.6 Financial application . . . 71
4.6.1 Data description . . . 71
4.6.2 Estimation of the models under $\mathbb{P}$ . . . 71
4.6.3 Option pricing: performance assessment . . . 72
4.7 Conclusions . . . 80
A Appendix . . . 81
A.1 Proof of the results in Section 4.2 . . . 81
A.2 Definition of distributions used in this chapter . . . 82
A.3 Proof of the results in Section 4.3 . . . 84
A.4 Background material in Markov Chain Monte Carlo . . . 89
A.5 Computational details for the SV-ARG . . . 91
A.6 Computational details for the SV-LHARG . . . 92
A.7 Further details for the financial application . . . 95
A.8 Further details on simulation experiments . . . 96

5 A backward Monte Carlo approach to exotic option pricing 111
5.1 Introduction . . . 111
5.2 The backward Monte Carlo algorithm . . . 113
5.3 Recovering the transition probabilities . . . 117
5.3.1 A quantization based algorithm . . . 118
5.3.2 The Large Time Step Algorithm . . . 122
5.4 Financial applications . . . 126
5.4.1 Model, pay-off specifications and variance reduction techniques . . . 126
5.4.2 Numerical results and discussion . . . 128
5.5 Conclusions . . . 138
B Appendix . . . 138
B.1 The distortion function and companion parameters . . . 138
B.2 Lloyd I method within the RMQA . . . 139
B.3 Robustness checks . . . 140
B.4 Construction of the Markov generator $\mathcal{L}_\Gamma$ and sensitivity analysis . . . 143

Conclusions 147

Bibliography 162

List of publications

Published articles :

- Fischer, M. and G. Livieri, (2016). Continuous time mean-variance portfolio optimization through the mean field approach, *ESAIM: Probability and Statistics 20*, 30-44.
- Bormetti, G., G. Callegaro, G. Livieri, and A. Pallavicini, (2017). A backward Monte Carlo approach to exotic option pricing, *European Journal of Applied Mathematics*, DOI: `https://doi.org/10.1017/S0956792517000079`.

Working papers available on-line :

- Bormetti, G., R. Casarin, F. Corsi and G. Livieri, (2016). Smile at errors: A discrete-time stochastic volatility framework for pricing options with realized measure *Available at* SSRN 28114304.
- Livieri, G., S. Mouti, A. Pallavicini and M. Rosenbaum, (2017). Rough volatility: evidence from options prices, *Available at* SSRN 2914086.

Other works and collaborations :

- Donadelli, M., G. Livieri, A. Paradiso, (2017). A stochastic Solow growth model ("plus" cycle).

Acknowledgments

Primarily and foremost I want to express my profound gratitude to my advisors. Thank you Giacomo for your enthusiastic guidance during the last three years. I am sure that without your support, exchange of ideas and training this thesis would not have been possible. Thank you Stefano for your enthusiastic guidance, patience, discussions, ideas, and funding to make my work more efficient. I am definitely indebted to both of you.

I am very grateful to Andrea Pallavicini for the research presented in Chapter 3 and 5. I have acquired many skills and a lot of knowledge while working with him. For the research in Chapter 3 I am also grateful to my co-authors Saad Mouti and Mathieu Rosenbaum and I thank Jim Gatheral and Olivier Scaillet for fruitful discussions.

I wish to express my deepest gratitude to Fulvio Corsi and Roberto Casarin for the research presented in Chapter 4, in particular for their suggestions, ideas and discussions. Moreover I would like to thank the Department of Economics Ca' Foscari University of Venice for the great hospitality and specifically to Monica Billio and my co-authors Michael Donadelli and Antonio Paradiso. I thank Dario Alitab for support during the development of the pricing code, and Drew D. Creal, Mark Jensen, Christian P. Robert, Francesco Ravazzolo, Roberto Renó, Herman K. van Dijk for helpful comments and fruitful discussions.

I also wish to express my gratitude to my co-author Giorgia Callegaro for the research presented in Chapter 5. Working with her was an inspiring and fruitful experience.

I am very thankful to Fabrizio Lillo for all the lectures during the first year of the Ph.D and for his comments, ideas and time. Moreover I am thankful to Maria Elvira Mancino.

I am also grateful to Davide Pirino, Marcello Rambaldi, Daniele Regoli, Michael Schneider, Damian Eduardo Taranto, Daniele Tantari and to all the other dear colleagues of the Quantitative Finance group at Scuola Normale Superiore in Pisa.

Last but not least I thank my mother and sister for all the support.

Chapter 1

Price dynamics model in continuous and discrete time

1.1 Price dynamics models in continuous time

Louis Bachelier's PhD thesis *Théorie de la Spéculation* (Bachelier, 1900) introduced financial mathematics to the world. Among the many original contributions of Bachelier's remarkable thesis, the most impressive are the mathematical foundation for the modern theory of *Brownian motion* and the introduction of the concept of *martingale*. Precisely, he established the probability law for the price fluctuations that the market admits at a given instant. Remarkably, having obtained the increments of the price process as independent *Gaussian* random variables, Bachelier wrote down what we would now call the Chapman-Kolmogorov equation and from this derives the connection with the heat equation. Moreover, when computing the price of a barrier option[1] he worked with the so called true price, i.e. a price for which the mathematical expectation of the speculator is null. In particular, Bachelier is saying that the true price is a martingale. However, the economic aspect of Bachelier's work was not completely investigated until the 1960s.

Samuelson (1965), to overcome the major drawback of Bachelier (1900) (i.e. the asset price modelled with Brownian motion can generate negative values while the price of an asset cannot), proposed to model the asset price as a geometric Brownian motion, which is a stochastic process described by a SDE characterized by a log-normal marginal distribution for the asset price. Eight years later, Black and Scholes (1973) introduced another watershed event in financial economics by deriving, through a non arbitrage argument[2], the closed-form European call option price formula assuming that the asset price dynamics is given by a geometric Brownian motion. The Black-Scholes model has become a market standard of quoting options because of its simplicity and tractability. If option prices in the market were compatible with the Black-Scholes formula, all the market implied volatilities corresponding to various options written on the same asset would coincide with the constant volatility parameter of the underlying asset. However, in today's reality[3] the Black-Scholes implied volatility heavily depends on the calendar time, the time to ma-

[1] The pay-off of a barrier option depends on whether the underlying asset price reaches a trigger point, the so called barrier price.

[2] The first two sentences of Black and Scholes (1973)'s abstract work read as follows: *If options are correctly priced in the market, it should not be possible to make sure profits by creating portfolios of long and short positions in options and their underlying stocks. Using this principle, a theoretical valuation formula for options is derived.*

[3] Precisely, after the so called Black Monday (October 19th, 1987). Before this market crash, implied volatility surfaces were compatible with those implied by the Black-Scholes model.

turity, and the moneyness of the option. If one plots implied volatility calculated from market quotes against different strikes for a fixed time to maturity, he will see a U-shaped pattern, the so-called smile. Indeed, the Black-Scholes model is not able to reproduce these dependences and this is considered its main limitation.

Since the late 1980s different approaches for the introduction of the smile in option pricing have been developed. The first one is by allowing for a time varying volatility, as in the case of stochastic volatility models (Hull and White, 1987; Scott, 1987; Heston, 1993) and local volatility models (Derman and Kani, 1996; Dupire, 1994). The second one consists in allowing for jumps in the dynamics of the underlying asset's price (see Merton, 1976; Bates, 1996b; Geman et al., 2001; Kou, 2002, among others). In this way, the dynamics is described as a Lévy process, thus obtaining skewness and non-zero excess kurtosis in the log-returns' distribution[4]. In local-volatility models, volatility becomes a deterministic function of time and price state. As said in the introduction, local volatility models have the merit of being flexible and enabling to the exact reproduction of the volatility smiles implied by the market. So, they are also used in practice to price exotic instruments whenever the smile risk dominates path-dependency risk e.g. when pricing Asian options. However, local volatility models do not reproduce the well-established stylized fact of the negative correlation of returns and volatility. Stochastic volatility models, which are defined in terms of price and a volatility process both driven by standard Brownian motion, in contrast, are able to incorporate this feature. Many others empirical features can be reproduced by using these type of models. For instance, a symmetric smile is well explained by assuming independence between the driving noise of the asset price and of the volatility (see Renault and Touzi, 1996). The correlation between these two driving noises, instead, reproduces the so-called leverage effect (see Hull and White, 1987). However stochastic volatility models fail to calibrate the steep smile shapes usually found at short maturities (i.e. empirically for maturities less than one months) *together* with their decrease for longer maturities (i.e. empirically for maturities ranging from one to two years), unless one complicates them either with *ad hoc* time-dependent and/or state-dependent parameter choices, as in stochastic-local volatility models (Ren et al., 2007), or with jumps (Bates, 1996b). Gatheral et al. (2014) argues that it is the generated term structure of the at-the-money (ATM) volatility skew[5] that really allows to say more about an option pricing stochastic volatility model. Standard stochastic volatility models generate an ATM volatility skew which is constant for short maturities (cfr. Fukasawa, 2011, for a theoretical corroboration of this fact) and decreases faster than it goes according to empirical data when time to maturity increases. In particular, inability of accurately reproducing the term structure of the ATM volatility skew at both short and long time-to-maturity is considered as the main limitation of standard stochastic volatility models. For this reason, new models for the joint dynamics of the price and its volatility

[4]For an introduction to time varying volatility models and jump-diffusion models see Tankov (2003) and Gatheral (2011), respectively

[5]The term structure of the ATM volatility skew is defined as

$$\psi(\tau) \doteq \left| \frac{\partial \sigma^{\mathrm{MKT}}(\tau, K)}{\partial K} \right|_{\mathrm{ATM}},$$

where τ is the option time-to-maturity, and σ^{MKT} is the volatility quoted by the market. In particular, Equity markets usually quotes call or put option prices in term of the volatility parameter to be put in the Black-Scholes model to match the option price. Empirically — at least in equity market — an ATM volatility skew proportional to $1/\tau^{\alpha}$ for some $0 < \alpha < 1/2$ is observed over a very wide range of expirations (cfr. for example Bayer et al., 2016).

have been introduced: We name these models fractional stochastic volatility models. The name comes from the fact that the driving noise of the volatility process is a fractional Brownian motion (fBM) (Mandelbrot and Van Ness, 1968), which is characterized by a single parameter $H \in (0,1)$, the so called Hurst parameter[6]. In particular, if $H > 1/2$ there is long memory in the dynamics of the fBM, in the sense that the covariance between two of its increments is not an integrable function.

The first and most celebrated fractional stochastic volatility model is the model in Comte and Renault (1998) where the dynamics of price follows a diffusion process with the logarithm of the volatility following a long memory, i.e. $H > 1/2$, fractional Ornstein-Uhlenbeck process with a strong mean reversion α. Formally

$$d\log(\sigma_t) = \nu dW_t^H + \alpha\left(m - \log\left(\sigma_t\right)\right) dt.$$

Around this model a large literature has subsequently been developed (Cheridito et al., 2003; Rosenbaum, 2008; Comte et al., 2012) to capture the widely accepted stylized fact of the long memory property of the volatility process (Ding et al., 1993; Andersen and Bollerslev, 1997; Andersen et al., 2001). However there is still a lot of ambiguity about statistical tests verifying its existence and over time it seems that this term has acquired a more precise meaning. First, the autocorrelation function is not integrable (Beran, 1994). Second, the autocorrelation function has a power-law decay with an exponent less than one. Since it is not possible to estimate the asymptotic behaviour of the covariance function without assuming a specific form, much of the most recent literature assumes long memory in volatility in this second, more technical, sense (see, for instance, Chen et al., 2006). Nonetheless, it turns out that there exist processes having autocorrelation function without the previous mentioned power law decay, which are classified as long memory using a standard statistical procedure (Andersen et al., 2001, 2003) aimed at detecting such a characteristic. In particular, the recent fractional stochastic volatility model of Gatheral et al. (2014) belongs to this class of models. In this model the dynamics of the price follows a diffusion process with the logarithm of the volatility following a short memory – H of order 0.1 – fractional Ornstein-Uhlenbeck process with a very small mean reversion α. Its use is suggested from a statistical investigation of recent prices and options data[7]. Importantly, it reproduces the explosive behaviour of the ATM volatility skew at short time to maturity, without using jumps (see Bayer et al., 2016). Inspired by this ground-breaking work of Gatheral et al. (2014) several studies on short-memory fractional volatility model have been developed (see, for instance Bennedsen et al., 2016, 2015; El Euch et al., 2016; El Euch and Rosenbaum, 2016; Funahashi and Kijima, 2015; Guennoun et al., 2014; Jaisson and Rosenbaum, 2016; Neuenkirch and Shalaiko, 2016). We finally mention the work of Corlay et al. (2014), in which the logarithm of the volatility follows a multi-fractional Ornstein-Uhlenbeck process: The constant parameter H is replaced with a time-dependent function h with value in $(0,1)$. Estimated on their dataset the function h is often smaller that $1/2$. In this thesis we integrate this last strand of literature revisiting the finding in Gatheral et al. (2014) and by studying implied volatility based approximations of the spot volatility.

[6]The Hurst parameter of a standard Brownian motion is equal to $1/2$.

[7]In Gatheral et al. (2014), the dataset consists of daily non-parametric estimates of volatility of the S&P and NASDAQ indices – including Realized Variance (RV) and Realized Kernel (RK) estimates – from January 3, 2000 to March 31, 2014 (`http://realized.oxford-man.ox.ac.uk`)

1.2 Price dynamics model in discrete time

Models for asset dynamics in discrete time have primarily been developed under the physical measure $\mathbb{P}$. Before the 1980s traditional econometric models assumed a constant one-period forecast variance of the returns. To overcame this implausible assumption, a new class of models named AutoRegressive Conditional Heteroschedastic (ARCH) has been introduced by Engle (1982). In these models the conditional variance of the returns at each fixed time is a linear function of the past squared returns. Four years later Bollerslev (1986) introduced the Generalized AutoRegressive Conditional Heteroscedastic (GARCH) model, where not only the squared lagged returns but also the lagged conditional variances influence the current value of the latter. Importantly, these time-varying volatility models are able to reproduce different regularities of returns and volatility. First, the fact that large observations, in absolute value, occur more often than expected from a normal variable, i.e. they have leptokurtic distributions. Second, the fact that volatile periods, characterized by large returns in absolute value, tend to alternate with more quiet periods of smaller returns, i.e. the so called volatility clustering (see Mandelbrot, 1997). Other parametrizations of volatility, which can capture these stylized facts of asset returns have subsequently been proposed (e.g. Higgins and Bera, 1992; Sentana, 1995, to cite only a few). Nontheless, in their original formulation, they are not able to reproduce other empirical regularities of asset returns such as the the negative skewness due to the leverage effect. Asymmetric ARCH (AARCH) by Engle et al. (1990), Exponential GARCH model (EGARCH) of Nelson (1991), Threshold ARCH model (TARCH) proposed by Zakoian (1994) and its modified version of Glosten et al. (1993) (GJR) are able to capture this predictable asymmetric effect. However, in all the previous models, endogenous shocks for the conditional variance process are not taken in consideration.

Taylor (1994) introduced a basic alternative to ARCH type models by allowing the conditional variance of the returns to be an unobserved random process. The main advantage of using Stochastic Volatility models is that they provide greater flexibility in describing the above-cited empirical regularities (see, for instance Ruiz, 1994; Harvey et al., 1994). However, the use of variance-specific disturbances comes to the use of latent variables and this calls for the use of suitable inference tools such as stochastic filtering and simulation based inference (e.g So et al., 1998, 2002; Jacquier et al., 2004, to cite only a few). In this thesis we introduce a new family of discrete time stochastic volatility models, named SV-LHARG(p) for the joint modelling of returns and realized measures of volatility. Importantly, we are able to propose an effective Markov Chain Monte Carlo (MCMC) algorithm to infer both the parameters and the latent variables.

Recently, thanks to the availability of high frequency data, another well-established discrete time volatility modelling approach is the so called Realized Volatility (RV) approach [8], where the RV is used as proxy for volatility related to the hours of open market. In particular, this implies that no filtering procedures are required and the model can be estimated directly using the observed RV. Moreover, by its very nature – it is constructed from the intra-day price movements – RV changes rapidly according to the market's movement, and this characteristic turns out to be crucial in improving volatility forecast. Indeed,

[8] Merton (1980) introduced the idea of RV measures in his seminal work. He showed that the integrated variance of a Brownian motion can be approximated by the sum of a large number of intra-day squared returns. This original intuition has been recently formalized and generalized by (Comte and Renault, 1998; Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2001; Barndorff-Nielsen, 2002; Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2005)

many studies have documented that incorporating realized measures in volatility models leads to a large economic and statistical gain (see, for instance Dobrev and Szerszen, 2010; Maheu and McCurdy, 2011; Hansen et al., 2012; Christoffersen et al., 2014, 2015). As regards the modelling, the Heterogeneous Autoregressive multi-component model by Corsi (2009) has become one of the standard models for describing and forecasting the dynamics of RV. Building on this model, different extensions have been proposed (Corsi et al., 2013; Majewski et al., 2015; Alitab et al., 2015) to better describe the characteristics observed in financial data both under the historical measure $\mathbb{P}$ and the risk-neutral one $\mathbb{Q}$. However, RV are affected by measurement errors and overnight effect.

A way to take advantage of the informative content of the realized measures of volatility and, at the same time, filter them out from measurement errors is to add to the model a measurement equation for the latent conditional variance of the returns. Recently , this approach has been followed either within a GARCH framework (Shephard and Sheppard, 2010; Hansen et al., 2012), a MEM framework (Engle and Gallo, 2006; Gallo and Otranto, 2015) or in a stochastic volatility one (Takahashi et al., 2009; Dobrev and Szerszen, 2010; Koopman and Scharth, 2013). However – to the best of our knowledge – no work has been devoted to the integration of Stochastic Volatility (SV) models incorporating RV measures with the literature on analytical option pricing. In this thesis, our SV-LHARG(p) fills-in this gap.

It turns out that an important class of process in finance is the affine one. A stochastic process is called affine if the logarithm of characteristic function of its transition distribution is affine with respect to the initial state. The importance of the affine property in finance has been acknowledged in many studies (see Duffie et al., 2000, 2003; Majewski et al., 2015; Alitab et al., 2015) and in particular for an option pricing application. Indeed, it allows for closed-form solutions for many of pricing problems. One problem that arises when addressing the problem of option pricing in a discrete-time framework is the determination of a risk-neutralization procedure. This theme has been treated for the first time by Duan (1995) in the context of a GARCH asset return process by using equilibrium argument and proposing a locally risk-neutral valuation procedure. Since then, many example of option pricing discrete-time volatility models have been proposed both in a GARCH-based (see Heston and Nandi, 2000; Gourieroux and Monfort, 2007; Christoffersen et al., 2008; Gagliardini et al., 2011, among others) and RV approaches (see Stentoft, 2008b; Corsi et al., 2013; Christoffersen et al., 2015; Majewski et al., 2015; Alitab et al., 2015, among others). In Bertholon et al. (2008); Gagliardini et al. (2011); Corsi et al. (2013); Majewski et al. (2015); Alitab et al. (2015) the risk-neutralization is performed by employing the so called Stochastic Discount Factor approach, which permits to incorporate multiple factor-dependent risk premia. Additionally, in order to have a good option pricing performance is essential to give a reliable description of the dynamics of the conditional variance of the returns and a smooth measure of it, since otherwise the noise in the latter is carried on the option prices. In this thesis, through our SV-LHARG(p), we corroborate this fact empirically. Indeed the fact that we are able – in an effective way – to filter out measurement errors of the realized measures of volatility turns out to be crucial in inflating the persistence of the latent conditional variance of the returns – the crucial parameter for the effective pricing of Standard and Poor's 500 Index options.

Chapter 2

Presented research

The goal of this thesis is to shed light on some of the issues that a researcher has to examine when he/she has to determine the fair price of a financial derivative: (i) The choice of a dynamics for the underlying instrument under the physical measure $\mathbb{P}$, (ii) the estimation of the parameters of the chosen dynamics, (iii) the definition of a pricing mechanism described by a stochastic discounting, (iv) the development of fast, accurate, and possibly flexible numerical methods to determine the fair price of the instrument.

The work presented in Chapter 3 is about the first issue. As pointed out, here, a trade-off arises as models which describe the historical dynamics of the asset price with adequate realism are usually unable to precisely describe the entire dynamics of the implied volatility surface. Recently, Gatheral et al. (2014) have proposed a new volatility model, named Rough Fractional Stochastic Volatility (RFSV), in which increments of log-volatility are assumed to follow a fractional Brownian Motion (fBm) with an estimated Hurst exponent around 0.1. They have shown how this model is able to reproduce several important features of empirical time-series and option prices. Remarkably, they reproduce the smoothness and the empirical autocorrelation structure of the volatility, and the exploding term structure when maturity goes to zero of the at-the-money skew (i.e., the derivative of the implied volatility with respect to strike). However, the volatility is a latent, unobservable variable. In Gatheral et al. (2014), recent estimation methods based on high-frequency price data have been used to estimate the spot volatility[1]. A natural question arise:
Question 1. What will be the value of the estimated Hurst exponent if one replicates the analysis of Gatheral et al. (2014) by using implied volatility based approximations of the spot volatility?

The work presented in Chapter 4 is mainly about the first two issues. In this Chapter, we move to the modelling of discrete time financial time series, in particular high frequency returns and realized volatilities. In a recent work, Majewski et al. (2015) have proposed a fully analytical realized volatility option pricing model featuring multiple-component structure in both volatility and leverage, and a flexible pricing kernel with multiple risk premia. On one hand, realized volatility, by its very nature, has the capability of changing rapidly according to the market's movement and this feature is essential in order to track the dynamics of the short-term implied volatility surface. On the other hand, realized measures provide noisy information about the latent conditional variance of the returns.

[1]The dataset consists of daily non-parametric estimates of volatility of the S&P and NASDAQ indices – including Realized Variance (RV) and Realized Kernel (RK) estimates – from January 3, 2000 to March 31, 2014 (`http://realized.oxford-man.ox.ac.uk`)

For this reason, realized volatility models fail to reproduce the high-level persistence of the volatility observed in the data and this fact turns out be crucial for the effective pricing of medium-long time to maturity options. We thus ask to ourselves the following question:
Question 2. Is it possible to build up a stochastic volatility model with realized measures of volatility able to: (i) reproduce the well-established stylized fact of financial time series, and in particular volatility persistence (ii) be effective both in estimation and calibration, (iii) filter out the measurement errors of the realized volatility, and finally (iv) preserve closed-form solutions for options prices?

The work presented in Chapter 5 is about the fourth issue and, specifically, about the construction of numerical techniques to price financial derivatives products whose pay-off may depend on the whole path followed by underlying asset, i.e. path-dependent options. The price of path-dependent derivative products is usually determined by employing two classes of numerical methods. The first approach consists in solving numerically a Partial Differential Equation (PDE) (Wilmott et al., 1993; Alziary et al., 1997; Zvan et al., 2000; Pascucci, 2011). The second one consists in using standard Monte Carlo methods(Glasserman, 2004). In particular, Monte Carlo is a very flexible method to generate random trajectories from the underlying asset price. It permits to achieve any degree of precision – at least in theory – by simply increasing the number of generated paths. However, this approach suffers from some ineffectiveness – especially when pricing out-of-the-money (OTM) options – since a relevant number of sampled paths does not contribute to the pay-off of the option. For this reason, several numerical techniques have been introduced to reduce the variance of the Monte Carlo price (see Clewlow and Strickland, 1996; Glasserman, 2004). We ask to ourself the following question:
Question 3 Is it possible to put together several tools or algorithms in order to obtain an effective numerical scheme based on Monte Carlo simulation to price exotic options?

2.1 Outline

Each question presented above corresponds to a part of the thesis.

In order to answer **Question 1**, we use two implied volatility based approximation of the spot volatility. Firstly, the implied volatility of an at-the-money liquid option with short maturity. Secondly, a refined version of it due to Medvedev and Scaillet (2007). We perform the same empirical analysis as in Gatheral et al. (2014) and we confirm that the increments of the log-volatility can be modelled as a fBm with Hurst exponent H less than $1/2$. However, the Hurst parameter found in our analysis, of order 0.30, is slightly larger than the one obtained in Gatheral et al. (2014). Nonetheless, we are able to provide both a numerical investigation and a quantitative understanding of this upward bias, thus confirming their results, i.e. the dynamics of the log-volatility can be described as follows

$$d\log(\sigma_t) = \alpha(m - \log(\sigma_t))\,dt + \nu dW_t^H, \tag{2.1}$$

with H of order 0.1 and mean reversion time scale very large, compared to the time scales of interest.[2]

[2]We stress that in spite of the apparent similarities between this model and the pioneering model of Comte and Renault (1998), they are conceptually very different. In Comte and Renault (1998), H has been taken larger than $1/2$ to reproduce the supposed long memory of the volatility while α large to capture the mean reversion of the volatility, (i.e. upward movements of the volatility tend to be followed by downward movements of the volatility). In Gatheral et al. (2014), instead, H has been taken very small to reproduce

To answer **Question 2**, we introduce a new family of discrete-time stochastic volatility models for the joint modelling of returns and realized measures of volatility. The novelty in our model is the presence of two measurement equations[3] which relate both the observed returns and realized measures to the latent conditional variance. Regarding the estimation part, first we extend the work of Creal (2015) and we provide analytical filtering and smoothing recursions for the basic (i.e. the model without a heterogeneous autoregressive structure and leverage component) version of the model, henceforth SV-ARG. Then, we develop an effective MCMC algorithm for richer variants. Building on Majewski et al. (2015), we develop a fully analytical option pricing framework. We show the importance of the filtering and smoothing of realized volatility in an option prices exercise on Standard and Poor's 500 Index options.

Finally, to answer to **Question 3**, we propose a new algorithm – named Backward Monte Carlo – which relies on the construction of a discrete multinomial tree. The crucial feature of our algorithm is that each random path runs backward from a terminal fixed point to the initial spot price. To characterize the tree (i.e. to determine the discrete points of the tree and the different probabilities of moving from one point to another one), we use two different methodologies. First, a recent technique termed Recursive Marginal Quantization (Pagés and Sagna, 2015). Second, an approach inspired by the finite difference approximation of the diffusion's infinitesimal generator (Albanese, 2007). We assess the reliability of the new methodology comparing the performance of both approaches and benchmarking them with competitor Monte Carlo methods.

We now rapidly review the main results of this thesis.

2.2 Rough volatility: evidence from option prices

2.2.1 A proxy for the spot volatility

We use two implied-volatility based approximation of the spot volatility: (i) The implied volatility of an at-the-money liquid option with short maturity. This idea can be justified by the fact that – in most models – the at-the-money implied volatility tends to the spot volatility as maturity goes to zero (see, among others, Mühle-Karbe and Nutz, 2011). (ii) An approximation formula due to Medvedev and Scaillet (2007). This correction formula allows to compute a proxy for the spot volatility starting from an at-the-money implied volatility with any – short – maturity. A drawback of Medvedev-Scaillet's formula is that it is proved to be valid only within a restricted class of stochastic volatility models, which does not include rough volatility models. However, our goal in this part is to see whether a proxy obtained from a standard volatility model still exhibits a rough behaviour.

2.2.2 Regularity of the volatility

In order to study the regularity of the volatility, we follow the methodology of Gatheral et al. (2014) and we look at the behaviour of the so-called empirical structure (or partition)

the exact roughness of the volatility – which correspond to some sort of mean reversion – and α small to capture the positive autocorrelation, or persistence, of the volatility.

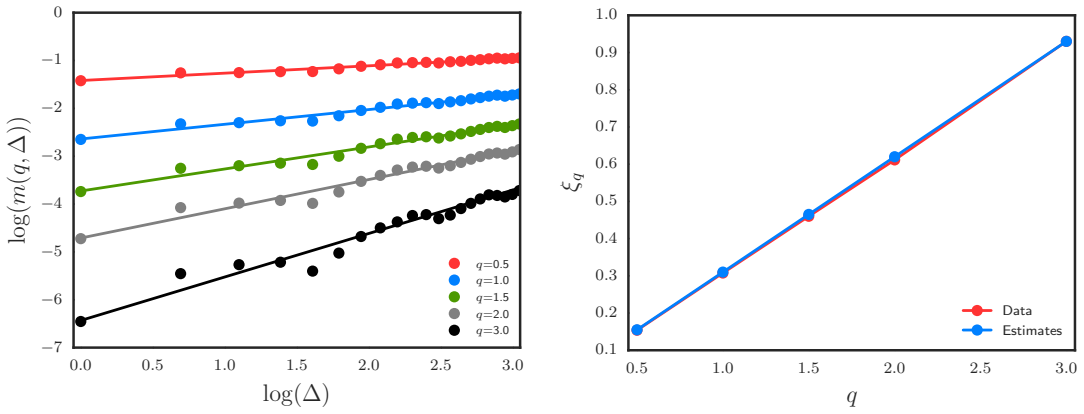
[3] A measurement equation ties the observable quantities to the latent conditional variance.

function $m(q, \Delta)$:

$$m(q,\Delta) = \frac{1}{N} \sum_{k=0}^{\lfloor (N-1)/\Delta \rfloor} \left| \log(\sigma_{t_{(k+1)\Delta}}) - \log(\sigma_{t_{(k)\Delta}}) \right|^q,$$

for different values of $q > 0$ and lags $\Delta > 0$. If the empirical financial process is scaling, then $m(q, \Delta) \sim c_q \Delta^{\xi(q)}$. We estimate $\xi(q)$ by regressing $m(q, \Delta)$ on Δ in log-log plots for different values of q. Figure 2.1 – left-panel – for instance, shows the structure function when using the at-the-money implied volatility with short maturity as proxy for the spot volatility[4]. For every q and for a wide range of Δ the points with coordinates $(\log(\Delta), \log(m(q, \Delta))$ are almost perfectly on the same line, thus confirming that the process is scaling. Additionally, if the empirical financial process is mono-fractal, then $\xi(q) = qH$. Figure 2.1 – right-panel – shows the points with coordinates $(q, \xi(q))$, with $\xi(q)$ the slope of the line in the left-panel of the same Figure, corresponding to the power q, and the points with coordinates $(q, 0.32\, q)$: The two graphs can hardly be distinguished. As last check, we verify that the increments of the log-volatility present a Gaussian behaviour. Figure 2.2 shows the histograms of log-volatility increments over different time intervals, together with a Gaussian density fit and the Gaussian density associated to the increments of a fBm with Hurst parameter equal to 0.32. From these graphs, we obtain that empirical distributions of log-volatility are reasonably approximated by Gaussian law[5]. We can conclude that – with our proxy – increments of log-volatility are well modelled with a fBm with Hurst parameter of order 0.30.

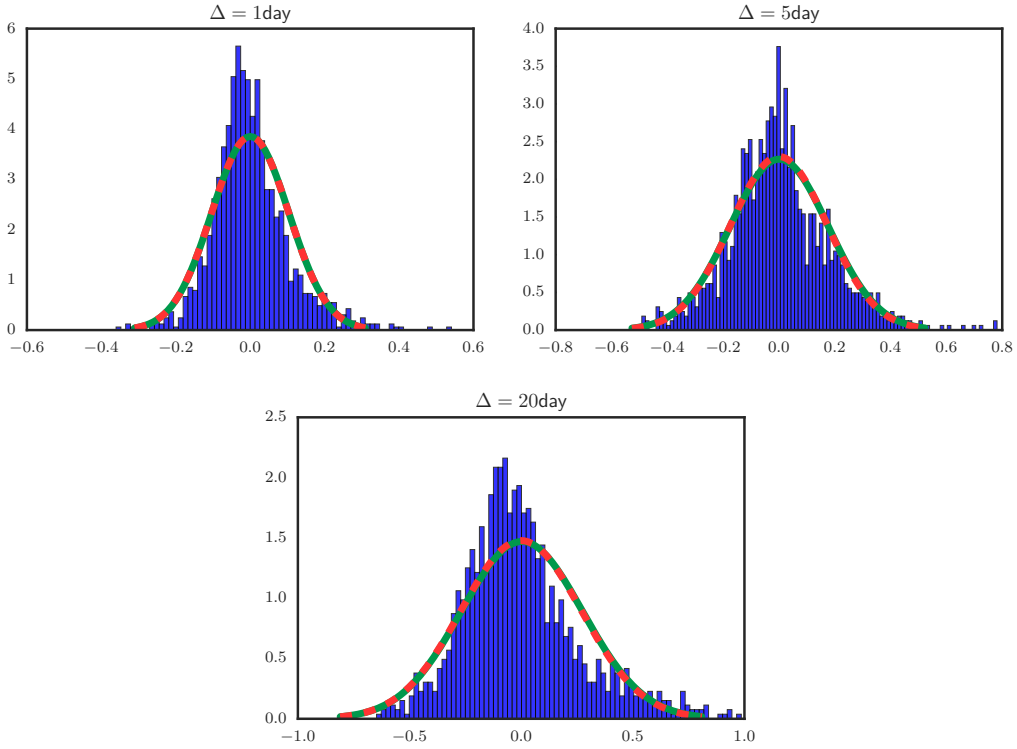
Figure 2.1: Scaling property of the log-volatility increments when using the implied volatility of an at-the-money liquid option with short maturity



NOTES: We use a data set from Bloomberg (Data obtained from AXA Group Risk Management) made of daily observations of the implied volatility of the option with maturity one month on the S&P500 index, from January 5, 2006 to May 5 2011.

[4]Results with the Medvedev-Scaillet formula are very similar.
[5]We remark that the empirical distributions are slightly more concentrated around their center.

Figure 2.2: Histograms of the log-volatility increments when using the implied volatility of an at-the-money liquid option with short maturity



NOTES: The Gaussian fit in blue and the density associated to the increments of a fBm with Hurst parameter equal to 0.32 is in red. Data set: see NOTES in Figure 2.1.

2.2.3 On the upward bias of the Hurst parameter

We ask to ourselves the reason of this upward bias between our estimated Hurst exponent and the one found in Gatheral et al. (2014). We provide an explanation for this bias, both numerically and theoretically. Numerically, we perform a Monte Carlo study. We simulate option prices in a rough volatility model with $H = 0.04$ and we estimate – through the same methodology described in the previous section – the Hurst parameter of the at-the-money implied volatility as a function of time to maturity. Figure 2.3 displays sample paths of the spot volatility and of the at-the-money implied volatility at maturity $\tau = 5$ and $\tau = 20$ days: At the visual level, it is evident that implied volatility trajectories are not as rough as that of the spot volatility and that the longer the time to maturity, the larger the smoothing effect. Quantitatively, we show that the extent of the bias on a simplified, although representative, setting. Precisely, we suppose that the at-the-money implied variance at time t of an option with time to maturity $\tau > 0$, denoted by $\hat{v}^\tau(t)$, is given by

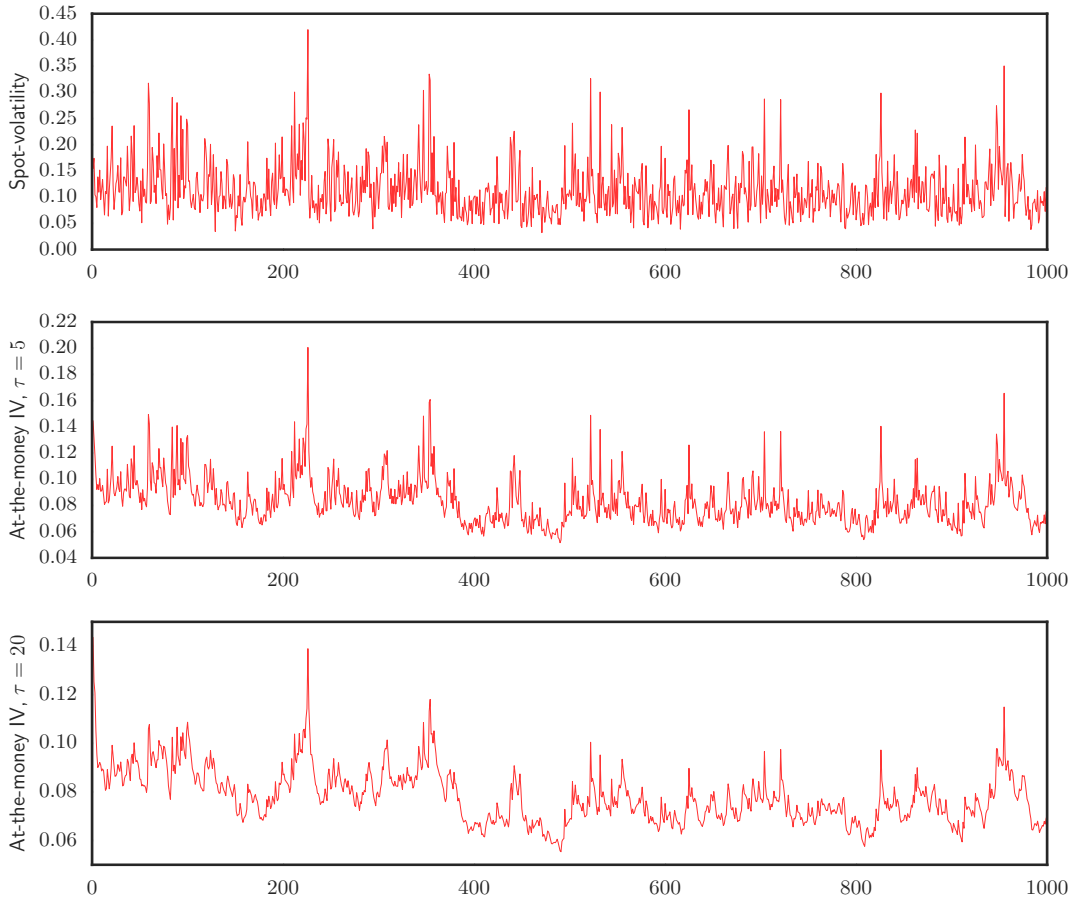
$$\hat{v}^\tau(t) = \frac{1}{\tau}\int_t^{t+\tau} \mathbb{E}_t[v_u]du,$$

where v_u is the spot variance at time u and $\mathbb{E}_t[.]$ the conditional expectation operator with respect to information up to time t. Furthermore, we take a simplified rough volatility model assuming that for $u > 0$,

$$v_u = v_0 + \nu W_u^H,$$

for some $v_0 > 0$ and $\nu > 0$. At this point, we compute a quantity very related to $m(2, \Delta)$, namely $\hat{m}^\tau(2, \Delta) = \mathbb{E}[(\hat{v}^\tau(\Delta) - \hat{v}^\tau(0))^2]$, and we show that the same scaling relationship as that associated to the spot volatility – i.e. $\widehat{m}^\tau(2, \Delta) \sim \Delta^{2H}$ – is approximately satisfied only if one considers implied volatilities with small enough times to maturity. Otherwise, this scaling property is disrupted, thus implying biased estimations for the Hurst parameter.

Figure 2.3: Sample paths of spot volatility and at-the-money implied volatility for $\tau = 5$ and $\tau = 20$ days when the underlying dynamics is a rough volatility model with $H = 0.04$.



2.3 A discrete-time stochastic volatility framework for pricing options with realized measure

2.3.1 Dynamics under the physical measure $\mathbb{P}$

We describe our new family of discrete-time stochastic volatility model in terms of the following distributional representation

$$
\begin{aligned}
r_t|\widetilde{\mathcal{F}}_t^{\mathrm{CV}} &\overset{d}{\sim} \mathcal{N}(\mu + \boldsymbol{\kappa}_1^{'}\boldsymbol{x}_{1t} + \gamma \mathrm{CV}_t, \mathrm{CV}_t), \\
\mathrm{RV}_t|\widetilde{\mathcal{F}}_t^{\mathrm{CV}} &\overset{d}{\sim} \mathcal{G}\left(\alpha e^{-\boldsymbol{\kappa}_2^{'}\boldsymbol{x}_{2t}}, \mathrm{CV}_t e^{\boldsymbol{\kappa}_2^{'}\boldsymbol{x}_{2t}}\right), \\
\mathrm{CV}_t|z_t &\overset{d}{\sim} \mathcal{G}\left(\nu + z_t, c\right), \\
z_t|\mathrm{CV}_{t-1}, \boldsymbol{l}_{t-1} &\overset{d}{\sim} \mathcal{P}o\left(\varphi(\mathrm{CV}_{t-1}, \boldsymbol{l}_{t-1})\right),
\end{aligned}
$$

$t = 1, \dots, T$. In previous equations, r_t denotes the geometric log-return of a risky asset computed from closing prices, RV_t is the realized volatility and CV_t the latent conditional variance of the log-return – which is described by a general and flexible Heterogeneous Autoregressive Gamma process with leverage effect (Majewski et al., 2015). With respect to the realized volatility model of Majewski et al. (2015), we have two measurement densities instead of one: a Gaussian density for the daily returns and a Gamma density for the realized volatility measure. The latter is intended to adjust measurement errors and overnight bias of the realized volatility. It has an intuitive interpretation: The conditional variance of the returns corresponds to the RV plus a random innovation, with the parameter α accounting for the overnight bias[6]. Importantly, the distributional representation above permits to derive analytically several quantities characterizing the model, e.g. the recursive formula for the computation of the conditional moment generating function (MGF) under $\mathbb{P}$.

2.3.2 Estimation of the model under the physical measure $\mathbb{P}$

Because of the presence of the latent variable CV_t, we employ a Bayesian approach for the inference, since it provides a natural way to include simulation methods such as Markov Chain Monte Carlo (MCMC) in the estimation process. We use a two-steps procedure. First, we extend the work of Creal (2015) and we derive analytical filtering and smoothing recursion for the latent variables of the SV-ARG. For instance, we derive the following

Proposition 2.3.1. *For the* SV-ARG *model the conditional likelihood,* $p(r_t, \mathrm{RV}_t | z_t, \boldsymbol{x}_t, \boldsymbol{x}_{2t}; \boldsymbol{\theta})$*, the Markov transition,* $p(z_t | z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})$*, and the initial distribution of* z_t*,* $p(z_1; \boldsymbol{\theta})$*, are respectively given by:*

$$
\begin{aligned}
&p(r_t, \mathrm{RV}_t | z_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = 2\eta(z_t, \mathrm{RV}_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) K_{\lambda(z_t)}\left(\sqrt{\psi\chi^{(t)}}\right)\left(\sqrt{\frac{\chi^{(t)}}{\psi}}\right)^{\lambda(z_t)}, \\
&p(z_t | z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) \propto \mathcal{S}\left(\lambda(z_{t-1}), \chi^{(t-1)}\frac{\phi^{(d)}}{c}, \psi\frac{c}{\phi^{(d)}}\right), \\
&p(z_1; \boldsymbol{\theta}) \propto \mathcal{NB}\left(\nu, \phi^{(d)}\right),
\end{aligned}
$$

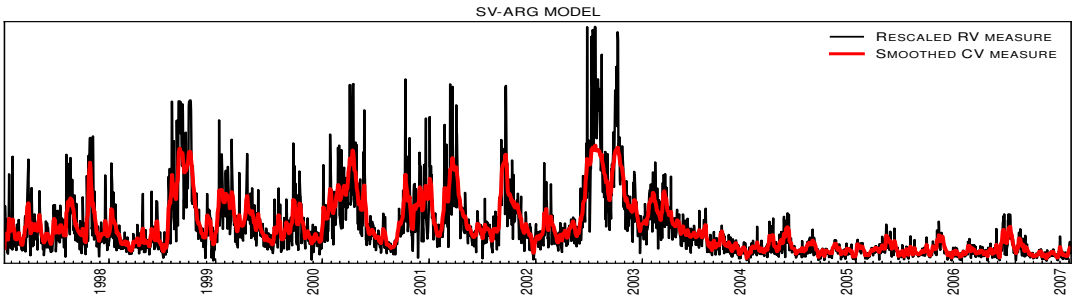
for suitable function η*,* $\lambda(z_t)$*,* ψ *and* $\chi^{(t)}$*. In the previous expression* $\mathcal{S}$ *indicates a Sichel distribution whereas* $\mathcal{NB}$ *a negative binomial distribution. A formal definition of these distributions can be found in Chapter 4, Appendix A.2),*

which permits to evaluate the likelihood of the SV-ARG exactly. Second, we provide a Bayesian inference procedure for estimating the parameters and the latent variables given a set of observations. By assuming a quadratic loss function, the Bayesian estimator is given by the mean of the posterior distribution, which typically is a multiple-dimensional integral with respect to a density function that is not tractable and which is known up to normalizing constant. Thus, we apply Monte Carlo methods to approximate the posterior distribution and the posterior mean. More specifically, we propose a Markov chain Monte Carlo (see Casella and Robert, 2004) algorithm for approximating the posterior distribution of the parameters and the latent variables. The parameters are sampled by applying a Metropolis-Hastings algorithm. As regards to the latent variables, we exploit the analytical filtering and smoothing results to develop an effective Forward Filtering Backward Sampling (FFBS) (see Frühwirth-Schnatter, 2006) for the SV-ARG model and

[6] We remind that the expected value of a Gamma distribution is given by the product between the two parameters.

an effective Metropolis-Hastings algorithm for the general SV-LHARG(p) using the SV-ARG as an auxiliary model. Figure 2.4, for instance, displays the output of the smoothed conditional variance estimates (*red line*) together with the RV rescaled to the whole sample estimates of the overnight factor on S&P 500 Futures. At a visual level it is evident the inflation in the volatility persistence due to filtering and smoothing.

Figure 2.4: Output of the smoothed conditional variance estimates for SV-ARG model, on S&P 500 Futures, 1997-2007



2.3.3 Risk neutral dynamics and financial applications

To determine the risk-neutral dynamics, we follow, among others, Majewski et al. (2015), and we employ an exponential affine Stochastic Discount Factor (SDF), with two risk premia. Formally

$$M_{t,t+1} = \frac{e^{-\nu_1 \mathrm{CV}_{t+1} - \nu_2 r_{t+1}}}{\mathbb{E}^{\mathbb{P}}\left[e^{-\nu_1 \mathrm{CV}_{t+1} - \nu_2 r_{t+1}} | \widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t\right]},$$

where ν_1 is the conditional variance premium and ν_2 is the standard equity premium. Building on Majewski et al. (2015), we derive the dynamics of the model under the risk-neutral measure showing that the conditional variance dynamics under $\mathbb{Q}$ is obtained from the dynamics under $\mathbb{P}$ through a deterministic transformation of the parameters estimated under the historical measure. We evaluate the model by benchmarking its performance with respect to two competitor models taken from both the GARCH and the RV literature in an option pricing exercise. Table 2.1 reports the relative Root Mean Square Error on the percentage IV[7] of the SV-LHARG with respect to the realized volatility model of Majewski et al. (2015) for different moneyness m and maturities τ: The amelioration for longer maturities and deep-out-the-money options is on average 15%. Figure 2.5, instead, compares the level [8] dynamics of the SV-LHARG (*red-line*) and the two-component GARCH (*blue line*) model introduced in Christoffersen et al. (2008). It is evident that SV-LHARG reacts more dynamically to changes in the volatility level.

[7]The Root Mean Square Error on the percentage IV ($\mathrm{RMSE}_{\mathrm{IV}}$) is defined as

$$\mathrm{RMSE}_{\mathrm{IV}} = \sqrt{\sum_{i=1}^{N} \frac{\left(\mathrm{IV}_i^{\mathrm{MOD}} - \mathrm{IV}_i^{\mathrm{MKT}}\right)^2}{N}},$$

where N is the number of options, and $\mathrm{IV}_i^{\mathrm{MOD}}$ and $\mathrm{IV}_i^{\mathrm{MKT}}$ are the model and the market implied volatility, respectively.

[8]The level is the average implied volatility of at-the-money options and maturity at the shortest available on a given day

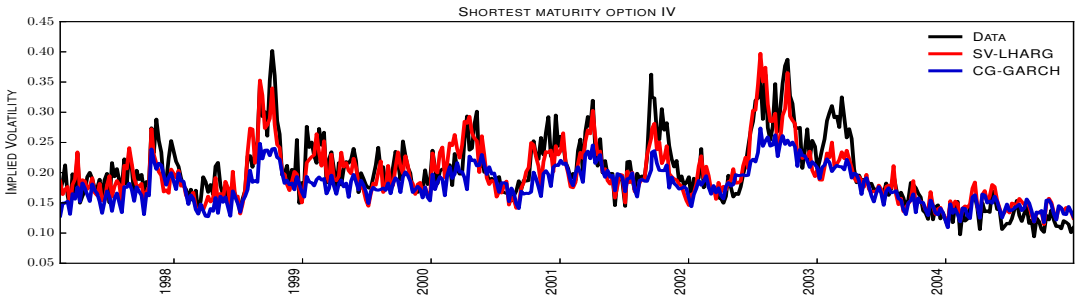
Table 2.1: Model comparison: option pricing performance on S&P500 out-of-the-money options from February 12, 1997 to December 29, 2004

Option Pricing Performance

Moneyness	Maturity $\tau \leq 50$	$50 < \tau \leq 90$	$90 \leq \tau \leq 160$	$160 < \tau$
Panel B	SV-LHARG/P-LHARG RMSE			
$0.8 \leq m \leq 0.9$	0.913	0.857	0.865	0.897
$0.9 < m \leq 0.98$	0.937	0.928	0.936	0.956
$0.98 \leq m \leq 1.02$	0.993	1.004	1.005	0.992
$1.02 \leq m \leq 1.1$	0.962	0.918	0.947	0.969
$1.1 \leq m \leq 1.2$	1.166	0.900	0.828	0.893

Notes: Panel A: The historical data are given by the daily RV computed on tick-by-tick data for the S&P500 Futures. The estimation period ranges from 8-January-1997 to 8-January-2007.

Figure 2.5: Shortest maturity option IV, SV-LHARG and CGARCH



Notes: Level dynamic from February 12, 1997 to December 31, 2004.

2.4 A backward Monte Carlo approach to exotic option pricing

2.4.1 Main ideas of the backward Monte Carlo algorithm

We start from a Markovian dynamics described by a generic Stochastic Differential Equation (SDE)

$$\begin{cases} dX_t = b\left(t, X_t\right) \, dt + \sigma\left(t, X_t\right) \, dW_t & t \in [0, T] \\ X_0 = x_0, \end{cases}$$

and we assume that it admits a unique strong solution. We split the time interval $[0, T]$ into n equally spaced subintervals $[t_k, t_{k+1}]$, $k \in \{0, \cdots, n-1\}$, with $t_0 = 0$ and $t_n = T$, and we approximate the SDE with an Euler-Maruyama scheme. Then, we assume that each random variable $\bar{X}_{t_k}$, $k \in \{1, \cdots, n\}$ of the Euler scheme takes values in $\Gamma_k \doteq \left\{\gamma_1^k, \cdots, \gamma_N^k\right\}$, with $\Gamma_0 = \gamma^0 = x_0$, i.e. we build up a multinomial tree. At this point, we construct the transition probabilities from node at time t_k to a node at time t_{k+1}, $k \in \{0, \cdots, n-1\}$

$$\Pi_{i,j}^{k,k+1} \doteq \mathbb{P}(\gamma_j^{k+1} | \gamma_i^k), \quad \gamma_i^k \in \Gamma_k, \ \gamma_j^{k+1} \in \Gamma_{k+1}, \text{ and } i, j \in \{1, \ldots, N\}$$

and analogously we define

$$\Pi^{k+1,k}_{\ell,m} \doteq \mathbb{P}(\gamma^k_m|\gamma^{k+1}_\ell), \quad \gamma^k_m \in \Gamma_k,\ \gamma^{k+1}_\ell \in \Gamma_{k+1},\ \text{and } l,m \in \{1,\dots,N\}.$$

We use a simple idea and we apply Bayes' theorem to recover the backward probabilities from the forward one. Formally,

$$\Pi^{k+1,k}_{i,j} = \frac{\Pi^{k,k+1}_{j,i}\ \mathrm{P}^k_j}{\mathrm{P}^{k+1}_i}$$

where, $\mathrm{P}^k_i \doteq \mathbb{P}(\widehat{\widehat{X}}_{t_k} = \gamma^k_i|\widehat{\widehat{X}}_{t_0} = x_0)$, for $i = 1,\dots,N$, $k = 1,\dots,n$. Using these probabilities one can go through the multinomial tree in a backward way, from each of terminal points γ^n_j to the initial node x_0. Indeed, we first select in a deterministic way only those points $\widetilde{\Gamma}_n$ of the final grid Γ_n for which the pay-off of our financial derivative is different from zero, then we evaluate the expectation of the pay-off with fixed initial and terminal points. For instance, for an up-and-out barrier call option[9] $\widetilde{\Gamma}_n = \{\gamma^n_i \in \Gamma_n\ :\ K \leq \gamma^n_i \leq B\}$.

2.4.2 Recovering the transition probabilities

We use two methodologies to recover the multinomial tree and the transition probabilities. First, we use an algorithm due to Pagés and Sagna (2015) termed Recursive Marginal Quantization Algorithm (RMQA). The idea of the RMQA is to optimally approximate each $\bar{X}_{t_k}$ of the Euler scheme, in a least-squares sense, by a discrete random variable $\widehat{\widehat{X}}_{t_k} : \Omega \to \Gamma_k$, where Γ_k is a finite set of elements in $\mathbb{R}$, in a recursive way. This algorithm gives as output also the transition probabilities[10]. Second, we use an algorithm which we term Large Time Step Algorithm (LTSA). The LTSA works with an *a priori* user-specified grid and requires less computational burden than the RMQA when pricing financial derivatives products whose underline instrument has a piecewise time-homogeneous local volatility dynamics and the pay-off requires the observation of the underlying instrument on a predefined finite set of dates. The LTSA consists in approximating the transition probability matrix of the (discretized) solution of the backward Kolmogorov equation (see Kijima, 1997; Karatzas and Shreve, 2012) associated to the initial diffusion.
Once that transition probabilities are recovered, we select appropriately the final domain $\widetilde{\Gamma}_n$ and we price financial derivatives simulating the paths backward.

2.4.3 Financial applications

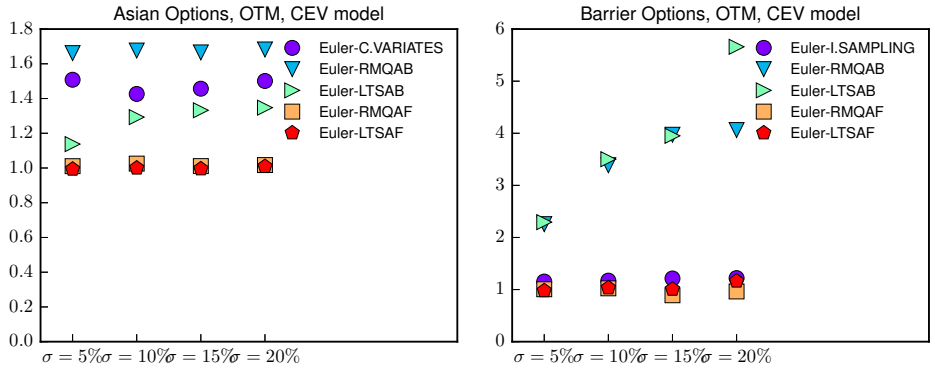
We assess the reliability of the Backward Monte Carlo in an option pricing exercise by benchmarking its performance with respect to different variance reduction methods taken from the literature (Glasserman, 2004). For instance, we apply the control variates method when pricing Asian options, and the importance sampling technique when pricing Barrier options. We also compute the prices of Asian and Barrier options through a forward Monte Carlo simulation on top of the multinomial tree. Figure 2.6 shows the so called *Error ratio* (i.e. the ratio between the error associated to the Monte Carlo estimator

[9] The pay-off of a up-and-out barrier call option is given by $\max(X_T - K, 0)1_{\{\tau>T\}}$, $\tau = \inf\{t \geq 0 : X_t \geq B\}$, B is the barrier and and T and K are the maturity and the strike of the option, respectively.

[10] As a novel contribution of this thesis, we also fix some numerical instabilities of the RMQA. We refer to Chapter 5 for an exhaustive explanation.

when the price is computed via a standard Monte Carlo procedure on the process $\bar{X}$, and the one associated to the other techniques) when we price an out-the-money Asian (resp. up-and-out barrier option) in a Constant Elasticity of Variance (CEV) model for different values of the volatility parameter: the strategy of reverting the Monte Carlo paths and simulating them from maturity back to starting date is an effective alternative to the pure Euler Monte Carlo or to Euler Monte Carlo combined with different variance reduction techniques.

Figure 2.6: Plot of the *Error ratio* as a function of the volatility parameter σ for CEV model when pricing Asian option and up-and-out barrier options.



NOTES: The initial spot price is $\bar{X}_0 = 1.36$, the strike is 1.37 and the barrier is 1.39.

Chapter 3

Rough volatility: evidence from option prices

The material of this chapter is taken from Livieri et al. (2017).

3.1 Introduction

Since the seminal work of Black and Scholes (1973), the most classical way to model the behaviour of the price S_t of a financial asset is to use a continuous semi-martingale dynamics of the form

$$d \log S_t = \mu_t dt + \sigma_t dW_t,$$

with μ_t a drift process and W_t a Brownian motion. The coefficient σ_t is referred to as the volatility process and it is the key ingredient in the model when one is interested in derivatives pricing and hedging.
Historically, following the pioneering approach of Black and Scholes (1973), practitioners have first considered the case where the process σ_t is constant or deterministic, that is the Black and Scholes model. However, in the late eighties, it became clear that such specification for the volatility is inadequate. In particular, the Black and Scholes model is inconsistent with observed prices for liquid European options. Indeed the implied volatility, that is the volatility parameter that should be plugged into the Black-Scholes formula to retrieve a market option price, depends in practice on the strike and maturity of the considered option, whereas it is constant in the Black-Scholes framework.
Hence more sophisticated models have been introduced. A first possible extension, proposed by Dupire (1994) and Derman and Kani (1996), is to take σ_t as a deterministic function of time and asset price. Such models, called local volatility models, enable us to perfectly reproduce a given implied volatility surface. However, their dynamic is usually quite unrealistic under local volatility. Another approach is to consider the volatility σ_t itself as an Ito process driven by an additional Brownian motion, typically correlated to W. Doing so one obtains less accurate static fits for the implied volatility surface but more suitable dynamics. Among the most famous of these stochastic volatility models are the Hull and White (Heston, 1993) and the SABR model (Hagan et al., 2002). More recent market practice is to use so-called local-stochastic volatility models which both fit the market exactly and generate reasonable dynamics.
In all the Brownian volatility models mentioned above, the smoothness of the sample path of the volatility is the same as that of a Brownian motion, namely $1/2 - \varepsilon$ Hölder continuous, for any $\varepsilon > 0$. However, it has been recently shown in Gatheral et al. (2014) that in

practice, spot volatility is much rougher than this. This result in Gatheral et al. (2014) is based on a statistical analysis of historical data using sophisticated high frequency estimation methods. More precisely, it is established in Gatheral et al. (2014) that the dynamic of the log-volatility process is very close to that of a fractional Brownian motion with Hurst parameter smaller than 1/2. We recall the following (Mandelbrot and Van Ness, 1968)

Definition 3.1.1. *A fractional Brownian Motion with Hurst parameter $H \in (0,1)$ is a centred self-similar Gaussian process with stationary increments satisfying for any $t \in \mathbb{R}$, $\Delta \geq 0$, $q > 0$:*

$$\mathbb{E}\left[\left|W^H_{t+\Delta} - W^H_t\right|^q\right] = \tilde{c}_q \Delta^{qH},$$

with $\tilde{c}_q$ the moment of order q of the absolute value of a standard Gaussian variable. For $H = 1/2$, one retrieves the classical Brownian motion.

The fBM is thus parametrized by the single parameter H, which has a well precise meanings. First, when $H > 1/2$, the increments of the fBM are positively correlated and exhibit long memory in the sense that

$$\sum_{t=0}^{+\infty} \mathbb{C}\text{ov}\left[W^H_1, W^H_t - W^H_{t-1}\right] = +\infty.$$

Indeed, $\mathbb{C}\text{ov}\left[W^H_1, W^H_t - W^H_{t-1}\right]$ is of order t^{2H-2} as $t \to \infty$[1]. Second, it gives exactly the smoothness of the sample path of fBM. Formally one has that $\forall \epsilon > 0$, W^H is $(H - \epsilon)-$Hölder continuous almost surely. So, there is a one to one correspondence between regularity and long memory through the Hurst parameter H.

Models where the volatility is driven by a fractional Brownian motion with $H < 1/2$ are called rough volatility models. In addition to fitting almost perfectly historical volatility time series, rough volatility models enable to reproduce important stylized facts of liquid option prices that local/stochastic volatility models typically fail to generate. In particular, the exploding term structure when maturity goes to zero of the at-the-money skew (the derivative of the implied volatility with respect to strike) is readily obtained, see Bayer et al. (2016); Fukasawa (2016). Other developments about rough volatility models can be found in Bennedsen et al. (2015, 2016); El Euch et al. (2016); El Euch and Rosenbaum (2016); Forde and Zhang (2015); Funahashi and Kijima (2015); Guennoun et al. (2014); Jaisson and Rosenbaum (2016); Neuenkirch and Shalaiko (2016).

The goal of the present work is to revisit the finding in Gatheral et al. (2014) using implied volatility data. Indeed in Gatheral et al. (2014), the authors work with historical price data from underlyings to estimate spot volatility. Here we use a spot volatility proxy which is not based on historical data, but on implied volatility. More precisely, we approximate the spot volatility by the implied volatility of an at-the-money liquid option with short maturity (or a refined version of it). This idea can be justified by the fact that in most models, the at-the-money implied volatility tends to the spot volatility as maturity goes to zero, see for example Mühle-Karbe and Nutz (2011). Our main result is a confirmation of that in Gatheral et al. (2014): When using alternative spot volatility measurement methods based on option prices, we can still conclude that volatility is rough.

The rest of the chapter is organized as follows. We investigate in Section 3.2 the roughness

[1] It holds true that

$$\text{Cov}[W^H_t, W^H_s] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right).$$

of time series of spot volatility approximations given by implied volatilities of at-the-money options on the S&P500 index, with maturity one month. In Section 3.3, instead of using raw implied volatilities, we compute spot volatilities from implied ones through a correction formula due to Medvedev and Scaillet (Medvedev and Scaillet, 2007). We then carry the same analysis as in Section 3.2. The results in Sections 3.2 and 3.3 are very similar to those in Gatheral et al. (2014). However, the estimated values for the Hurst parameter, although smaller than 1/2, are actually larger than those obtained in Gatheral et al. (2014). We show numerically and analytically in Section 3.4 that this upward bias comes from a regularizing effect due to the remaining time to maturity of the considered options.

3.2 At-the-money implied volatility with short maturity as spot volatility proxy

As explained in the introduction, our goal is to study the behavior of the spot volatility and to show that it is well approximated by a rough process. Of course this is a difficult task since volatility is a latent, unobserved variable. In Gatheral et al. (2014), the authors use recent estimation methods based on ultra high frequency price data to estimate spot volatility. In this work, instead, we wish to use option price data. This idea is reasonable if we use at-the-money options for which the time to maturity is short. Indeed, it is well-known that in most models, the at-the-money implied volatility converges to the spot volatility as maturity goes to zero, see for example Mühle-Karbe and Nutz (2011).

3.2.1 Data description

In this section, we use a data set from Bloomberg[2], made of daily observations of the implied volatility of the option with maturity one month on the S&P500 index, from January 5, 2006 to May 5, 2011[3]. Note that the data are in fact already extrapolated internally by the data provider (using quoted options at 4 PM) and do not necessarily exactly correspond to transaction data, see Bloomberg (2008). In Section 3.3, we present a method enabling us to derive spot volatilities from observed option prices with various maturities. Here we rely on the data provider approach to get option prices with the same maturity. This is not an issue since our aim in this work is to show that a rough dynamic for the volatility is obtained from any reasonable spot volatility proxy.

3.2.2 Scaling property

Let $\sigma^{imp}_{t_0}, ..., \sigma^{imp}_{t_N}$ be the time series of implied volatilities extracted from our data base. Here for $i \geq 0$, $t_{i+1} - t_i$ corresponds to one business day. In the spirit of Gatheral et al. (2014), we wish to review the behaviour of the so-called structure function $m(q, \Delta)$ given by

$$m(q, \Delta) = \frac{1}{N} \sum_{k=0}^{\lfloor (N-1)/\Delta \rfloor} | \log(\sigma^{imp}_{t_{(k+1)\Delta}}) - \log(\sigma^{imp}_{t_{k\Delta}}) |^q$$

for various $q > 0$ and lags Δ going from 1 to about 40 days[4]. Through the quantity $m(q, \Delta)$, our goal is to revisit the finding in Gatheral et al. (2014) that the (spot) log-

[2]Data obtained from AXA Group Risk Management.

[3]Data around the third Friday of each month (settlement date) are removed from the data base. We have 1166 points in total.

[4]Of course when computing $m(q, \Delta)$ we in fact also average over the possible starting points $t_0, ..., t_{\Delta-1}$.

volatility is well approximated by a fractional Brownian motion with Hurst parameter H smaller than $1/2$. In this case, assuming spot and implied volatilities coincide, we should observe the following relationship:

$$m(q, \Delta) \sim c_q \Delta^{qH}, \tag{3.1}$$

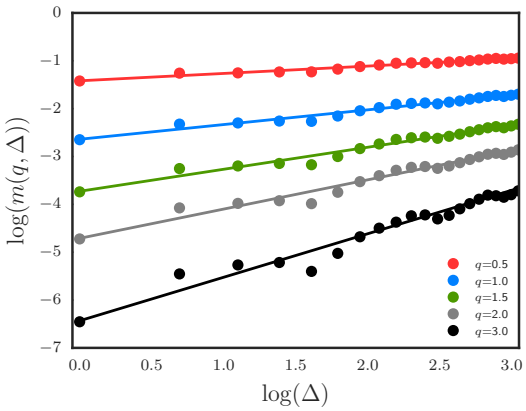
with c_q a constant depending on q. Indeed (cfr. Definition 3.1.1), we have for $t \geq 0$ and $\Delta > 0$

$$\mathbb{E}[|W^H_{t+\Delta} - W^H_t|^q] = \tilde{c}_q \Delta^{qH},$$

with $\tilde{c}_q$ the absolute moment of order q of a standard Gaussian random variable.

To investigate the validity of (3.1), we plot in Figure 3.1 the logarithm of $m(q, \Delta)$ against the logarithm of Δ, for several values of q.

Figure 3.1: Scaling property of log-volatility increments of the at-the-money implied volatility with short maturity



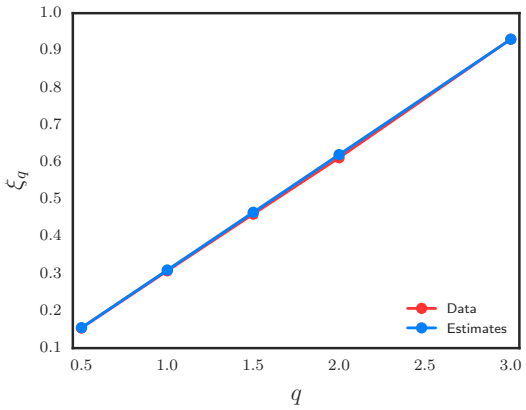
NOTES: Data set is from Bloomberg made of daily observations of the implied volatility of the option with maturity one month on the S&P500 index, from January 5, 2006 to May 5 2011.

For every q, the points with coordinates $(\log(\Delta), \log(m(q, \Delta)))$ are almost perfectly on the same line, and this for a wide range of Δ. Figure 3.1 is actually very similar to that obtained from historical volatility measurements in Gatheral et al. (2014). Thus we can deduce that indeed, for a given q,

$$m(q, \Delta) \sim c_q \Delta^{\zeta(q)},$$

for some $\zeta(q)$.
Now we want to check whether $\zeta(q)$ can be taken of the form qH for some H, as suggested in Gatheral et al. (2014). This would lead to the same mono-fractal scaling as that of the fractional Brownian motion with Hurst parameter H. To answer this, we plot in Figure 3.2 the points with coordinates $(q, \zeta(q))$, where $\zeta(q)$ is taken as the slope of the line in Figure 3.1 corresponding to the power q, and the points with coordinates $(q, 0.32q)$.

Figure 3.2: Mono-fractal scaling of the at-the-money implied volatility with short maturity



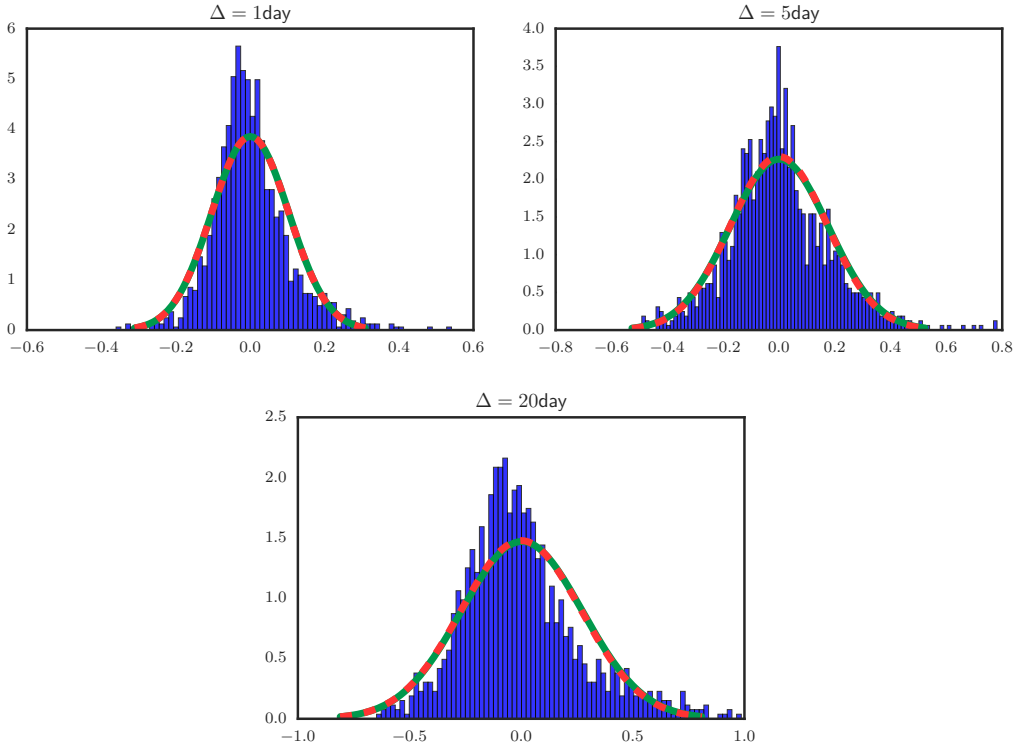
NOTES: Data set is from Bloomberg made of daily observations of the implied volatility of the option with maturity one month on the S&P500 index, from January 5, 2006 to May 5 2011.

We see that the two graphs on Figure 3.2 can hardly be distinguished. This means that (3.1) almost perfectly holds, with H around 0.32. Note that such value for H corresponds to rough volatility since it is smaller than $1/2$. However, it is larger than those reported in Gatheral et al. (2014). This is actually due to the fact that our options have a significant remaining time to maturity of one month. This induces a smoothing phenomenon in the estimation of the Hurst parameter. This effect is of the same nature as that described and explained in Gatheral et al. (2014) caused by the discrepancy between spot and integrated volatility over a short time interval. We quantify this measurement bias numerically and analytically in Section 3.4.

3.2.3 Distribution of log-volatility increments

Recall that it is suggested in Gatheral et al. (2014) that the log-volatility process is well modeled by a fractional Brownian motion with Hurst parameter smaller than $1/2$. This implies monofractal scaling as investigated above but also a Gaussian behavior of the log-volatility increments. This feature is indeed satisfied when using historical estimates as measurements for spot volatility, see Gatheral et al. (2014). Here we wish to study whether such property also holds when the volatility proxies are given by our short term at-the-money implied volatilities. To this end, we display in Figure 3.3 histograms of log-volatility increments over different time intervals, together with a Gaussian density fit and the Gaussian density associated to the increments of a fractional Brownian motion with Hurst parameter equal to 0.32.

Figure 3.3: Histograms of the log-volatility increments when using the implied volatility of an at-the-money liquid option with short maturity



NOTES: The Gaussian fit in blue and the density associated to the increments of a fBm with Hurst parameter equal to 0.32 is in red. Data set: see NOTES in Figures above.

From these graphs, we obtain that the empirical distributions of log-volatility increments are reasonably approximated by Gaussian laws. However, we can remark that the empirical distributions are slightly more concentrated around their center. Finally, the Gaussian fits almost exactly coincide with those associated to the fractional Brownian motion with Hurst parameter equal to 0.32.

In conclusion, using at-the-money implied volatilities with maturity one month as spot volatility proxies, we obtain that log-volatility is well approximated by a rough fractional Brownian motion. This confirms the finding in Gatheral et al. (2014).

3.3 A refined implied volatility based proxy for the spot volatility

In this section, we wish to study the robustness of the results obtained in Section 3.2. To do so, we work with another spot volatility proxy based on at-the-money options with short maturity. More precisely, we use the approximation formula from Medvedev and Scaillet (2007). This correction formula enables us to compute a spot volatility proxy from an at-the-money implied volatility with any (short) maturity. This is an advantage compared to what is done in Section 3.2 where only options with one month maturity

are considered[5]. The drawback of Medvedev-Scaillet formula is that it is proved to be valid only within a restricted class of stochastic volatility models, which does not include rough volatility models. However our goal here is to see whether a proxy obtained from a Brownian volatility model still exhibits a rough behaviour.

3.3.1 Data description and processing

Here our data set is provided by OptionMetrics and consists in daily close bid/ask prices of European puts and calls on the S&P500 index, from September 5, 2001 to January 31, 2012, for various strikes and maturities, together with the daily traded volumes. We discard options with price less than 2.5 cents of dollar or with zero trading volume. Besides, as in Section 3.2, prices corresponding to settlement dates are removed, so as obvious outliers.

We then want to compute the market implied volatilities from put and call prices. These can be obtained by inverting the Black-Scholes-Merton formula given the market prices of call or put options. For the sake of clarity, let us consider a day t in the dataset and focus on call prices. The market implied volatility is defined so that the following pricing equation is satisfied with the price C given by the market.

$$C(\tau, K) = D(\tau)\left(F(\tau)N(d_1(\mathrm{IV}(\tau);\tau,K) - KN(d_2(\mathrm{IV}(\tau);\tau,K))\right)$$

where $D(\tau)$ is the zero-coupon bond price maturing at τ, while we define

$$d_1(\mathrm{IV}(\tau);\tau,K) \doteq \frac{1}{\mathrm{IV}(\tau)\sqrt{\tau}}\log\frac{F(\tau)}{K} + \frac{1}{2}\mathrm{IV}(\tau)\sqrt{\tau}$$

and

$$d_2(\mathrm{IV}(\tau);\tau,K) \doteq d_1(\mathrm{IV}(\tau);\tau,K) - \mathrm{IV}(\tau)\sqrt{\tau}.$$

We can notice that to invert the above formula we need also the zero-coupon bond prices and the asset forward prices. On the other hand, we know for each time-to-maturity τ call and put bid/ask prices for a whole range of strike levels. Thus, we are able to implement a best-fit procedure to jointly calibrate for each τ all the quantities: IV, D and F. We implement an optimization procedure for each time t and time-to-maturity τ. First, we construct two portfolios: (i) one long call and one short put with the same strike K, (ii) one short call and one long put again with the same strike K. The prices of these portfolios can be easily calculated by using the put-call parity when bid/ask spreads are null, since we have

$$C(\tau, K) - P(\tau, K) = D(\tau)\left(F(\tau) - K\right)$$

In our case the market is quoting bid/ask prices, so that we search for the optimal solution

$$\{D^\star, F^\star\} = \underset{D,F}{\arg\min}\left\{\sum_i w_i\left(\frac{1}{2}\left(C_i^{\mathrm{a}} - P_i^{\mathrm{b}}\right) + \frac{1}{2}\left(C_i^{\mathrm{b}} - P_i^{\mathrm{a}}\right) - D(\tau)\left(F(\tau) - K_i\right)\right)\right\}$$

where $C_i^{\mathrm{a,b}}$ and $P_i^{\mathrm{a,b}}$ are respectively the call and put market prices (the apices stand for ask and bid) quoted at strike level K_i, and the weights are given by

$$w_i \doteq \frac{\sqrt{\min\{V_i^C, V_i^P\}}}{\frac{1}{2}(C_i^{\mathrm{a}} - C_i^{\mathrm{b}}) + \frac{1}{2}(P_i^{\mathrm{a}} - P_i^{\mathrm{b}})},$$

[5]Mixing various maturities without any correction would have been very arguable.

with V_i^C and V_i^P standing for the trading volumes of call and put options quoted at strike level K_i. Finally, our implied volatility is taken as that of a call whose price would be the mid-price between the bid and ask prices. Recall that for our approximations to be valid, we focus on at-the-money implied volatilities with short maturity. Following Medvedev and Scaillet (2007), we only select implied volatilities of options with time to maturity ranging from 15 to 60 days. Shorter term options are discarded because quotes can be noisy. Moreover, we restrict our data to log-forward moneyness – here[6] defined as $X = \log(S\, e^{r\,\tau}/K)$ – with τ the time-to-maturity and K the strike price – belonging to the interval $[-0.03, 0.03]$. Such procedure yields a total number of 34842 implied volatilities over 2569 days.

3.3.2 The Medvedev-Scaillet correction formula

In Medvedev and Scaillet (2007), the authors consider a general modelling framework encompassing most of the classical parametric price models. They use a two factors jump-diffusion stochastic volatility model of the form

$$\begin{cases} dS_t = (r - \mu(\sigma_t))\, S_t dt + \sigma_t S_t\, dZ_t + S_t\, dJ_t \\ d\sigma_t = a(\sigma_t)dt + b(\sigma_t)\left(\rho\, dZ_t + \sqrt{1-\rho^2}\, dW_t\right), \end{cases} \tag{3.2}$$

where Z_t and W_t are two independent Brownian motions and J_t is a Poisson-type jump process, independent of Z_t and W_t. Both r and the correlation coefficient ρ are assumed to be constant. The expected jump size $\mathbb{E}[\Delta J]$ is also constant, but the jump intensity $\lambda(\sigma_t)$ may depend on the volatility in a deterministic way. Here, as in the numerical experiments in Medvedev and Scaillet (2007), we consider the following parametric forms:

$$b(\sigma_t) = \beta \sigma_t^{\phi}\,, \quad \lambda(\sigma_t) = \lambda_0 \sigma_t^{\psi},$$

for some non-negative constants β, ϕ, λ_0 and ψ.

Let σ be the spot volatility and $\hat{\sigma} = \hat{\sigma}(\tau)$ be the at-the-money implied volatility of an option with time to maturity τ. Following Medvedev and Scaillet (2007), we build our option-based spot volatility proxy in two steps. First, the chosen model is calibrated from the approximation formula in Proposition 7 in Medvedev and Scaillet (2007) using all our option prices over the entire time period. Precisely, if we set $\theta = X/(\sigma\sqrt{\tau})$ (resp. $\hat{\theta} = X/(\hat{\sigma}\sqrt{\tau})$) the moneyness degree and $\widehat{\mathrm{IV}}\left(\hat{\theta}, \tau; \hat{\sigma}\right)$ the market implied volatility we use the following formula to calibrate the model

$$\begin{aligned} \widehat{\mathrm{IV}}\left(\hat{\theta}, \tau; \hat{\sigma}\right) = \hat{\sigma} &+ \left(\mathrm{I}_1(\hat{\theta}, \hat{\sigma}) - \mathrm{I}_1(0, \hat{\sigma})\right)\sqrt{\tau} \\ &+ \left(\mathrm{I}_1(0,\hat{\sigma})\left(\frac{\partial \mathrm{I}_1(0,\hat{\sigma})}{\partial \sigma} - \frac{\partial \mathrm{I}_1(\hat{\theta},\hat{\sigma})}{\partial \sigma} + \frac{\hat{\theta}}{\hat{\theta}}\frac{\partial \mathrm{I}_1(\hat{\theta},\hat{\sigma})}{\partial \sigma}\right)\right. \\ &\left. + \mathrm{I}_2(\hat{\theta}, \hat{\sigma}) - \mathrm{I}_2(0, \hat{\sigma})\right)\tau + O(\tau\sqrt{\tau}), \end{aligned} \tag{3.3}$$

[6] We note that this is not the definition of forward moneyness usually found in the literature, i.e. $-X$. In the reminder of the chapter we will use X as in (Medvedev and Scaillet, 2007) and we will make clear the precise definition used whenever relevant

where the function I_1 and I_2 are given by

$$\begin{aligned}\mathrm{I}_1(\hat{\theta},\hat{\sigma}) &= -\frac{b\rho}{2}\hat{\theta} - \mu g \\ \mathrm{I}_2(\hat{\theta},\hat{\sigma}) &= -\frac{\mu^2}{2\hat{\sigma}}\hat{\theta}^2 g^2 + \left(-\frac{\mu b\rho}{2\sigma}\hat{\theta}^3 - \frac{\mu\hat{\sigma}}{2}\hat{\theta} - \hat{\sigma}\lambda\hat{\theta}\right) g + \mathcal{P}\left(\hat{\theta};\hat{\sigma}\right),\end{aligned} \tag{3.4}$$

where $\mathcal{P}$ is a quadratic function in $\hat{\theta}$[7]:

$$\begin{aligned}\mathcal{P}(\hat{\theta},;\hat{\sigma}) = &\left(-\frac{5}{12}\frac{\rho^2 b^2}{\hat{\sigma}} + \frac{1}{6}\frac{b^2}{\hat{\sigma}} + \frac{1}{6}\rho^2 bb' - \frac{1}{2}\frac{\mu b\rho}{\hat{\sigma}}\right)\hat{\theta}^2 \\ &+ \frac{\rho b\hat{\sigma}}{4} + \frac{\rho b\mu}{2\hat{\sigma}} + \frac{1}{24}\frac{\rho^2 b^2}{\hat{\sigma}} + \frac{1}{12}\frac{b^2}{\hat{\sigma}} \\ &- \frac{1}{6}\rho^2 bb' + \frac{\mu^2}{2\hat{\sigma}} - \frac{\hat{\sigma}\mu}{2} - \lambda\hat{\sigma},\end{aligned} \tag{3.5}$$

with $g = N(\hat{\theta})/n(\hat{\theta})$, $h = 1/n(\hat{\theta})$, and $N(\hat{\theta})$ and $n(\hat{\theta})$ are the p.d.f and the c.d.f of the standard normal distribution. Besides, $\mu = \mu(\hat{\sigma})$, $b = b(\hat{\sigma})$, and b' denotes the derivative of b w.r.t. σ evaluated in $\hat{\sigma}$.
Then, to retrieve the proxy for the spot volatility, we then consider the following expansion as τ goes to zero shown in Medvedev and Scaillet (2007):

$$\begin{aligned}\sigma = \hat{\sigma} &- \mathrm{I}_1(0,\hat{\sigma})\sqrt{\tau} \\ &+ \left(\mathrm{I}_1(0,\hat{\sigma})\frac{\partial \mathrm{I}_1(0,\hat{\sigma})}{\partial\sigma} - \mathrm{I}_2(0,\hat{\sigma}) + \frac{1}{2}\rho b(\hat{\sigma})\mathbb{E}[\Delta_J]\frac{\partial\lambda(\hat{\sigma})}{\partial\sigma}\right)\tau + O(\tau\sqrt{\tau}).\end{aligned} \tag{3.6}$$

Remark 1. *Formula* (3.3) *does not involve the unobserved spot volatility. As a consequence, it can be used to calibrate any parametric specification of model* (3.2) *to a set of option prices across calendar dates simultaneously. Moreover, since all the parameters are calibrated with the exception of the spot volatility, we are able to filter out the skew dynamics effect from the time series* (3.6)*. This is important for the purpose of our analysis.*

3.3.3 The scaling property revisited

We now wish to study the scaling property of spot volatility proxies based on the approximation formula (3.6). We consider two cases: The Heston case, where $\phi = 0$ and $\lambda_0 = 0$, and the general case, where all the parameters are calibrated. The calibration results are given in Table 3.1

[7]We note that in the original paper of Medvedev and Scaillet (2007) $\mathcal{P}\left(\hat{\theta};\hat{\sigma}\right)$ depends also from the volatility drift. However, in practice, the dataset is filtered to select only options with short maturities and it cannot be inferred from data.

Table 3.1: Parameters calibrated on quoted S&P500 option prices, from September 5, 2001 to January 31, 2012.

PARAMETER	HESTON	GENERAL CASE
$\beta\rho$	-0.18 (0.00)	-3.27 (0.08)
ρ	-0.48 (0.00)	-0.39 (0.00)
ϕ	0	1.79 (0.02)
$\lambda_0\mathbb{E}(\Delta J)$	0	-0.6924 (0.03)
$\mathbb{E}(\Delta J)$	$--$ $--$	-0.17 (0.00)
ψ	$--$ $--$	1.11 (0.01)

NOTES: The square root of the average of squared errors in implied volatilities is equal to 0.0065 for the Heston model and 0.0035 for the General case.

Once the parameters are obtained, we can implement Equation (3.6) to compute everyday a spot volatility proxy. Note that in Equation (3.6), we take for $\hat{\sigma}$ the implied volatility with shortest time to maturity. Then we conduct the same analysis as in Section 3.2.2. The results are given in Figure 3.4 for the Heston model and Figure 3.5 for the general case (notations are the same as in Section 3.2.2).

Figure 3.4: Scaling property of log-volatility increments when based on Heston proxy. In the second graph H is taken equal to 0.33.

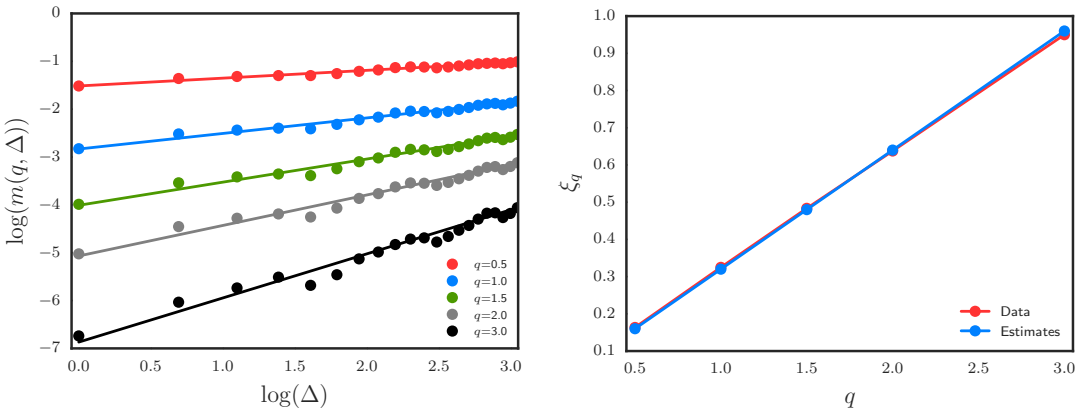
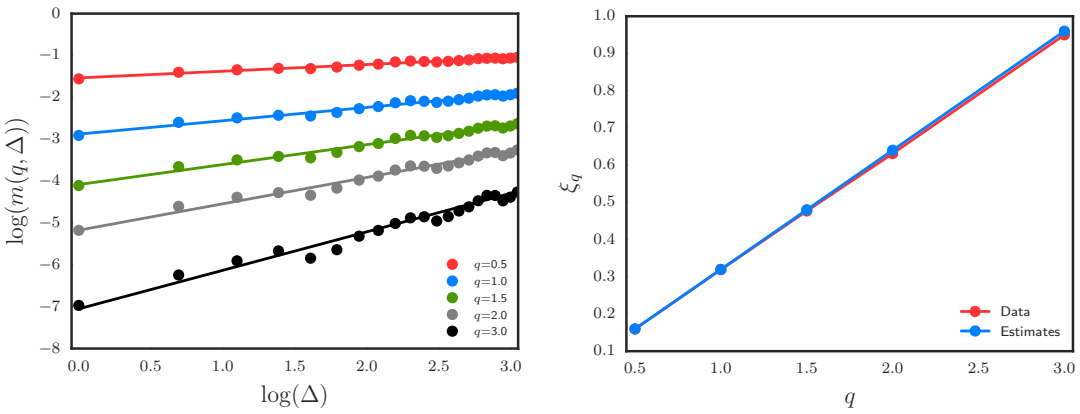


Figure 3.5: Scaling property of log-volatility increments when based on the general case proxy. In the second graph H is taken equal to 0.34.



The results are very similar to those in Section 3.2.2. Here again we can confirm the fact that volatility is rough. This is obtained even though in the models in which the proxies are computed volatility is of Brownian type and therefore not rough.

3.4 On the upward bias when estimating the Hurst parameter

We explain in this section why using implied volatility measures as spot volatility proxies induces an upward bias in the estimation of the Hurst parameter. We start with a numerical investigation of this phenomenon.

3.4.1 Monte Carlo study

To understand the extent of the bias when estimating the Hurst parameter, we simulate option prices in a rough volatility model. Then we compute the Hurst parameter based on these simulated data.
Let $T > 0$. We consider the following model without leverage effect over the time interval $[0, T]$:

$$d \log S_t = \sigma_t dZ_t, \quad d \log \sigma_t = \eta dW_t^H.$$

Here Z_t is a Brownian motion, W_t^H a fractional Brownian motion independent of Z_t and $\eta > 0$.

Simulation of fractional Brownian motion

We consider a time interval $[0, T]$ and fix an equidistant partition $0 = t_0 < t_1 < ... < t_n = T$. We first wish to simulate $(W_{t_1}^H, \dots, W_{t_n}^H)$. For $i, j \in \{1, ..., n\}$, we have

$$\mathbb{E}[W_{t_i}^H W_{t_j}^H] = \frac{1}{2} \left(t_i^{2H} + t_j^{2H} - \mid t_i - t_j \mid^{2H} \right).$$

Then we can use the Cholesky decomposition of the covariance matrix Σ of $(W_{t_1}^H, \dots, W_{t_n}^H)$: $\Sigma = LL^T$, where $L = (l_{ij})_{i,j \in \{1,n\}}$ is lower-triangular. Thus simulating a sample path of the fractional Brownian motion at times (t_i) can be done generating a vector $X = (X_1, ..., X_n)$ of independent standard Gaussian random variables and setting $(W_{t_1}^H, ..., W_{t_n}^H) = LX$.

Simulating option prices under rough volatility

We place ourselves at time $t_i > 0$ and assume past spot volatilities and prices have been observed at times $t_1, \dots, t_i$. We want to compute the price at time t_i of an option with expiration date $t_k = t_i + \tau$ for some $\tau > 0$. The procedure goes as follows:

- We generate M paths of the volatility process on the interval $[t_{i+1}, t_k]$. This is done simulating $(W_{t_j}^H)_{t_{i+1} \leq t_j \leq t_k}$ conditional on past information, that is the filtration generated by $(X_{t_1}, ..., X_{t_i})$. Using the lower triangular form of L, these new values for the fractional Brownian motion at times $t_{i+1} \leq t_j \leq t_k$ can be obtained writing

$$W_{t_j}^H = \sum_{p=1}^{i} l_{jp} X_p + \sum_{p=i+1}^{j} l_{jp} X_p.$$

  The i first variables X_p are those used to simulate the fractional Brownian motion up to time t_i, whereas $(X_{i+1}, \dots, X_j)$ is a sample of independent standard Gaussian

random variables, independent from past values. Taking the exponential, we get our spot volatility sample path. We write σ^m for the m-th volatility trajectory.

- The price at time t_i of an at-the-money option with time to maturity τ is obtained computing

$$\frac{1}{M}\sum_{m=1}^{M} C_{BS}\left(S_{t_i}, \tau, \sqrt{\frac{1}{\tau}\sum_{p=i+1}^{k}(\sigma_{t_p}^m)^2}\right),$$

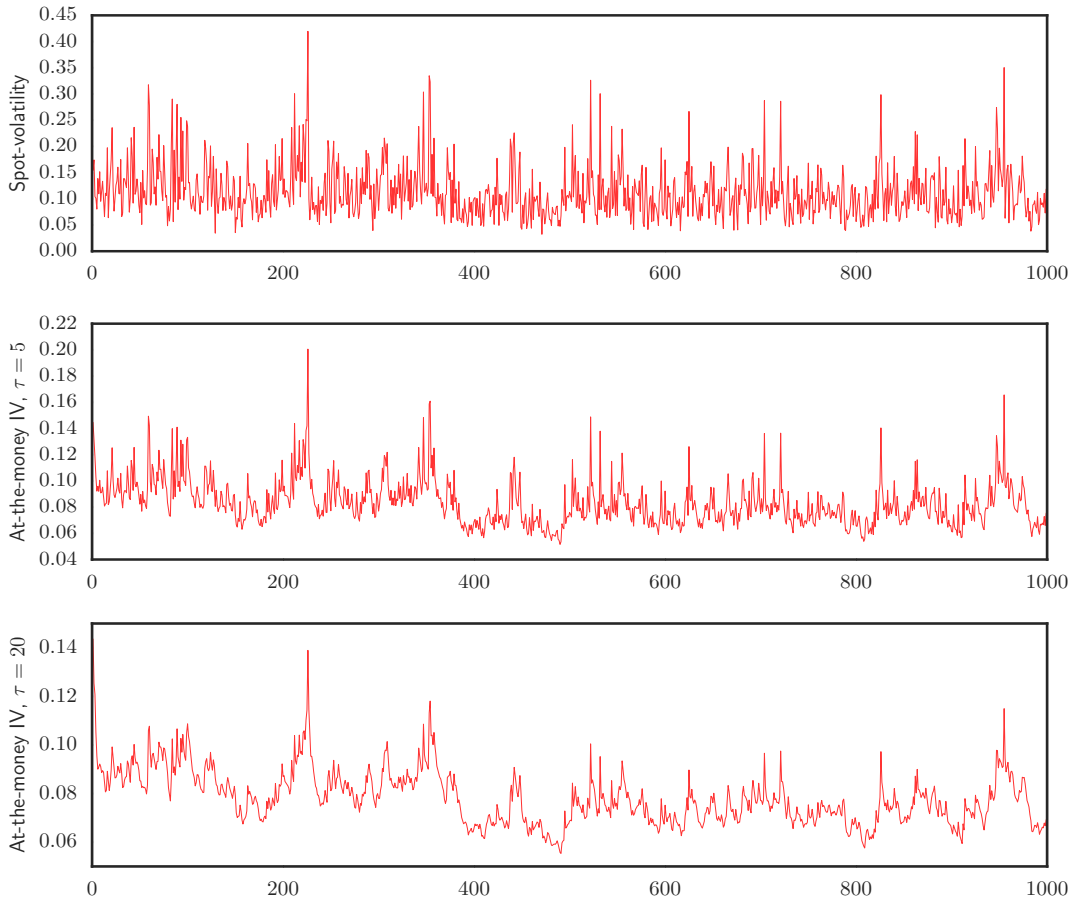
  where $C_{BS}(S_{t_i}, \tau, \sigma)$ is the price of an at-the-money option with time to maturity τ in a Black-Scholes model with volatility σ, zero interest rate, and underlying value S_{t_i}.

- Eventually we invert Black-Scholes formula to obtain the implied volatility.

Results

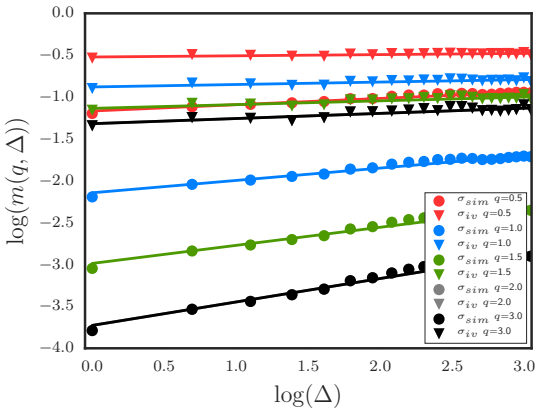
We consider the following set of parameters: $H = 0.04$, $\eta = 1.0$ and $T = 1000$ days. Such parameters are consistent with Bayer et al. (2016); Gatheral et al. (2014). We take $\tau \in \{1, \dots, 20\}$ days and run $M = 10^4$ simulations. Figure 3.6 displays the sample path of the spot volatility together with those of the implied volatilities associated to 5 and 20 days.

Figure 3.6: Sample paths of spot volatility and at-the-money implied volatility for $\tau = 5$ and $\tau = 20$ days when the underlying dynamics is a rough volatility model with $H = 0.04$.



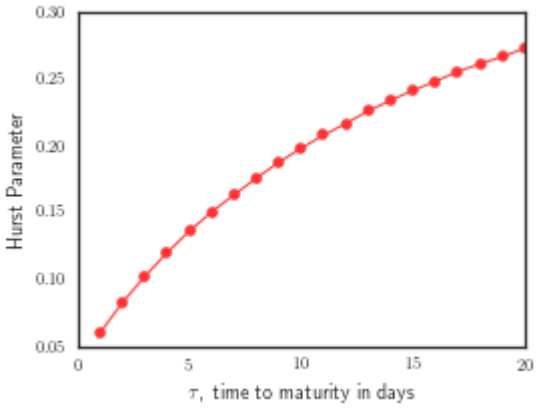
At the visual level, it is already clear that implied volatility trajectories are not as rough as that of the spot volatility. Furthermore, the longer the time to maturity, the larger the smoothing effect.
As in Sections 3.2 and 3.3, we now consider Equation (3.1). Based on our simulation, for several values of q, we plot in Figure 3.7 the logarithm of $m(q, \Delta)$ against the logarithm of Δ. This is done in two cases: when m is obtained from spot volatility values and when m is derived from implied volatility values, with $\tau = 5$ days.

Figure 3.7: Scaling property of log-volatility increments: spot volatility and implied volatility with $\tau = 5$.



We see that for a given q, when $m(q, \Delta)$ is computed from implied volatilities, the points with coordinates $(\log(\Delta), \log(m(q, \Delta)))$ remain on the same line. However, the slope of this line is larger than that obtained when $m(q, \Delta)$ is computed from spot volatilities (which provides the true underlying H up to small statistical error). Hence there is indeed a smoothing effect due to the remaining time to maturity of the considered options.
Finally, we give in Figure 3.8 the estimated values of H when using implied volatilities from the simulation, for different times to maturity.

Figure 3.8: Estimated values of the Hurst parameter using implied volatilities as a function of time to maturity.



Under our simulation framework, we see that using options with maturity 1 day, we obtain a quite accurate value for H of 0.06, while the true parameter is equal to 0.04. Taking longer maturities leads to an increasing bias. With 20 days maturity, one gets an estimated

Hurst parameter of about 0.27. These results are in line with those in Sections 3.2 and 3.3.

3.4.2 Analytical illustration of the upward bias

In the spirit of Appendix C in Gatheral et al. (2014), we finally want to provide a more quantitative understanding of the observed upward bias when estimating the Hurst parameter from implied volatilities. To do so, we consider a very crude approximation. Indeed we suppose that the at-the-money implied variance at time t of an option with time to maturity $\tau > 0$, denoted by $\hat{v}^\tau(t)$, is given by

$$\hat{v}^\tau(t) = \frac{1}{\tau}\int_t^{t+\tau} \mathbb{E}_t[v_u]du,$$

where v_u is the spot variance at time u and $\mathbb{E}_t[.]$ the conditional expectation operator with respect to information up to time t. Furthermore, we take a simplified rough volatility model assuming that for $u > 0$,

$$v_u = v_0 + \nu W_u^H,$$

for some $v_0 > 0$ and $\nu > 0$. These approximations are actually probably enough to shed light on the bias phenomenon. Indeed it is due to the effects of the conditional expectation and integral operators appearing in the implied volatility.
In this simplified setting, our goal is to illustrate the smoothing effect leading to the upward bias. To do so, we compute a quantity very related to $m(2, \Delta)$, namely

$$\hat{m}^\tau(2, \Delta) = \mathbb{E}[(\hat{v}^\tau(\Delta) - \hat{v}^\tau(0))^2].$$

Indeed, under our assumptions, if the implied volatility were equal to the spot one, this quantity would be proportional to Δ^{2H}. However, we now show that because of the use of implied volatility in $\hat{m}(2, \Delta)$, this relationship no longer holds, particularly for large τ/Δ. We recall the Mandelbrot and Van Ness representation of fractional Brownian motion:

$$W_t^H = c_H\left(\int_0^t (t-s)^{H-1/2}dW_s + \int_{-\infty}^0 \left((t-s)^{H-1/2} - (-s)^{H-1/2}\right)dW_s\right),$$

where W_t is a two-sided Brownian motion and c_H is so that the variance of W_1^H is equal to 1. We easily have

$$\begin{aligned}\hat{v}^\tau(\Delta) = v_0 &+ \frac{\nu}{\tau}c_H\int_0^\tau\int_{-\infty}^0 \left((\Delta+u-s)^{H-1/2} - (-s)^{H-1/2}\right)dW_s du \\ &+ \frac{\nu}{\tau}c_H\int_0^\tau\int_0^\Delta (\Delta+u-s)^{H-1/2}dW_s du.\end{aligned}$$

Using stochastic Fubini theorem, this gives

$$\begin{aligned}\hat{v}^\tau(\Delta) - \hat{v}^\tau(0) = &\frac{\nu}{\tau}c_H\int_{-\infty}^0\int_0^\tau \left((\Delta+u-s)^{H-1/2} - (u-s)^{H-1/2}\right)du dW_s \\ &+ \frac{\nu}{\tau}c_H\int_0^\Delta\int_0^\tau (\Delta+u-s)^{H-1/2}du dW_s.\end{aligned}$$

Hence we easily deduce from Ito isometry that

$$\hat{m}^\tau(2, \Delta) = A\left(h_1(\Delta, \tau) + h_2(\Delta, \tau)\right),$$

with

$$A = \frac{c_H^2 \nu^2}{(H+1/2)^2},$$

$$h_1(\Delta,\tau) = \frac{1}{\tau^2}\int_{-\infty}^{0}\Big((\Delta+\tau-s)^{H+1/2} - (\Delta-s)^{H+1/2} - (\tau-s)^{H+1/2} + (-s)^{H+1/2}\Big)^2 ds,$$

$$h_2(\Delta,\tau) = \frac{1}{\tau^2}\int_{0}^{\Delta}\Big((\Delta+\tau-s)^{H+1/2} - (\Delta-s)^{H+1/2}\Big)^2 ds.$$

We write $h_1(\Delta,\tau)$ under the form

$$\frac{1}{\tau^2}\Delta^{2H+2}\int_{-\infty}^{0}\Big((1+\frac{\tau}{\Delta}-s)^{H+1/2} - (1-s)^{H+1/2} - (\frac{\tau}{\Delta}-s)^{H+1/2} + (-s)^{H+1/2}\Big)^2 ds.$$

Setting $\theta = \tau/\Delta$, we obtain

$$h_1(\Delta,\tau) = \Delta^{2H} f_1(\theta),$$

where

$$f_1(\theta) = \frac{1}{\theta^2}\int_{-\infty}^{0}\Big((1+\theta-s)^{H+1/2} - (1-s)^{H+1/2} - (\theta-s)^{H+1/2} + (-s)^{H+1/2}\Big)^2 ds.$$

Similarly, we have

$$h_2(\Delta,\tau) = \Delta^{2H} f_2(\theta),$$

where

$$f_2(\theta) = \frac{1}{\theta^2}\int_{0}^{1}\Big((1+\theta-s)^{H+1/2} - (1-s)^{H+1/2}\Big)^2 ds.$$

So

$$\hat{m}^\tau(2,\Delta) = A\Delta^{2H}\left(f_1(\theta) + f_2(\theta)\right).$$

Now remark that

$$\lim_{\theta\to 0} f_1(\theta) = (H+1/2)^2\int_{-\infty}^{0}\Big((1-s)^{H-1/2} - (-s)^{H-1/2}\Big)^2 ds$$

and

$$\lim_{\theta\to 0} f_2(\theta) = (H+1/2)^2\int_{0}^{1}(1-s)^{2H-1}.$$

Consequently,

$$\lim_{\theta\to 0}(f_1(\theta) + f_2(\theta)) = (H+1/2)^2\frac{1}{c_H^2}.$$

Thus, when θ is small,

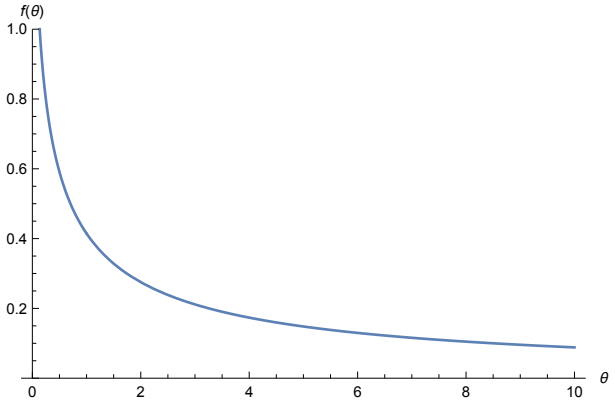
$$\hat{m}^\tau(2,\Delta) \sim \nu^2\Delta^{2H}.$$

This means that the same scaling relationship as that associated to the spot volatility is approximately satisfied when considering implied volatilities with small enough times to maturity. Otherwise, one should add the multiplicative factor

$$f(\theta) = \frac{c_H^2}{(H+1/2)^2}\left(f_1(\theta) + f_2(\theta)\right)$$

on the right hand side of the above relationship. This disrupts the scaling property and implies biased estimations for the Hurst parameter. We draw in Figure 3.9 the graph of the function f for $H = 0.04$.

Figure 3.9: The function f for $H = 0.04$.



For fixed τ (as in Section 3.2), the function f is increasing with Δ. Therefore, when doing a regression analysis of the cloud of points with coordinates $(\log(\Delta), \log(\hat{m}^\tau(2, \Delta)))$, this implies an upward bias in the estimation of H due to a higher slope.

3.5 Conclusion

Motivated by the recent strand of literature on rough fractional volatility models, in this chapter we used two implied-volatility based approximations for the daily spot volatility and we show that the resulting time-series enjoy two simple regularities. First, that the distributions of increments of log-volatility are approximatively Gaussian. Second, we found a mono-fractal scaling relationship, i.e.

$$\mathbb{E}\left[|\log(\sigma_\Delta) - \log(\sigma_0)|^q\right] \sim \Delta^{q\,H}$$

where H is of order 0.30 and represents a measure of smoothness of the volatility process. We justify both numerically and quantitatively that, actually, this value for H is upward bias. Thus, we show that rough volatility models seem to be indeed compatible with observed prices of liquid options providing another argument for the use of rough volatility models for financial applications, especially for option pricing and hedging.

Chapter 4

A discrete-time stochastic volatility framework for pricing options with realized measure

The material of this chapter is taken from Bormetti et al. (2016).

4.1 Introduction

Thanks to the availability of high frequency data, the literature has recently introduced a number of realized measures of daily volatility, (see Andersen and Bollerslev, 1998; Andersen et al., 2001; Barndorff-Nielsen, 2002; Barndorff-Nielsen and Shephard, 2004; Andersen et al., 2008; Hansen and Horel, 2009, and references therein). Any of these Realized Volatility (RV) measures is far more informative about the Conditional Volatility (CV) of daily returns than the ones obtained only from the daily squared returns. On one hand, it is well established that incorporating realized measures in volatility models leads to a large economic and statistical gains (see, for instance Dobrev and Szerszen, 2010; Maheu and McCurdy, 2011; Hansen et al., 2012; Christoffersen et al., 2014, 2015). On the other hand, on empirical data, realized measures of volatility are still affected by measurement errors and overnight bias. These two facts have motivated the development of several strands of literature in which the RV measures are used in a measurement equation for the latent CV, either within a GARCH framework (see Shephard and Sheppard, 2010; Hansen et al., 2012), a MEM framework (see Engle and Gallo, 2006; Gallo and Otranto, 2015), or in a stochastic volatility one (see Takahashi et al., 2009; Dobrev and Szerszen, 2010; Koopman and Scharth, 2013). However, to the best of our knowledge, no work has been devoted to the integration of Stochastic Volatility (SV) models incorporating RV measures with the literature on analytical option pricing: The work presented in the present chapter fills this gap.

The goal of the work presented in this chapter is to introduce a new family of flexible and tractable discrete-time SV models which allows to filter the latent CV process from both observed returns and RV measures. The contemporaneous use of the two sources of information reduces measurement errors and allows to recover the higher persistence of the true CV dynamics. The proposed modelling approach accurately reproduces well-established stylized fact observed in financial time series, while preserving closed-form formulas for option prices and closed-form filtering and smoothing for the latent process.

In particular, the proposed model has the two following measurement densities: a

Gaussian density for the daily returns and a Gamma density for the RV measure. The dynamics of the latent CV is assumed to follow the flexible class of Heterogeneous Autoregressive Gamma process with Leverage RV-LHARG(p) recently introduced by Majewski et al. (2015) for the dynamics of RV.
We then term the general version of the presented model as SV-LHARG(p)[1]. A parsimonious instance of the latter is able to capture the widely observed stylized facts of volatility. It is well known, in fact, that volatility is a clustered and a highly persistent process with a multi-factor structure under both the physical measure $\mathbb{P}$ (see Müller et al., 1997; Barndorff-Nielsen and Shephard, 2001; Bollerslev and Wright, 2001; Calvet and Fisher, 2004) and the risk-neutral measure $\mathbb{Q}$ (see Bates, 2000; Adrian and Rosenberg, 2008; Christoffersen et al., 2008; Li and Zhang, 2010). Moreover, equity and stock-index volatilities show significant asymmetric response to past returns – an effect commonly referred to as leverage effect (see Christie, 1982; Glosten et al., 1993; Bollerslev et al., 2006). In modelling the latter effect, recent literature advocates the need of a multi-factor leverage structure under both $\mathbb{P}$ and $\mathbb{Q}$ (see Corsi and Renò, 2012; Scharth and Medeiros, 2009).

The first main contribution of the work presented in this chapter is to develop a fully analytical option pricing framework for discrete-time SV models featuring multiple component structure in both volatility and leverage. The option pricing literature in discrete time traces back to the seminal affine model of Heston and Nandi (2000). However, the analytical tractability of discrete-time option pricing is guaranteed only for rather specific types of models. Noticeable exceptions include GARCH-based (see Christoffersen et al., 2008, 2013; Bormetti et al., 2015) and realized volatility approaches (see Stentoft, 2008a; Corsi et al., 2013; Christoffersen et al., 2014; Majewski et al., 2015) later extended to separately deal with the continuous and discontinuous components of realized measures (see Christoffersen et al., 2015; Alitab et al., 2015). Moreover, the option pricing literature has recently acknowledged the need for a flexible pricing kernel incorporating a variance-dependent risk premium, in addition to the common equity risk premium (see Christoffersen et al., 2013).

Our general SV-LHARG(p) inherits all the analytical features characterizing the RV-LHARG(p). In particular, we are able to: (i) derive the recursive formulae to compute the Moment Generating Function (MGF) under $\mathbb{P}$, (ii) derive the change of measure using a general and flexible exponentially affine Stochastic Discount Factor (SDF) featuring a variance risk premium, (iii) characterize analytically the no-arbitrage conditions, (iv) obtain the recursive relation to compute the MGF under $\mathbb{Q}$, (v) derive an explicit one-to-one mapping between the parameters of the latent process dynamics under $\mathbb{P}$ and $\mathbb{Q}$, and, finally, (vi) have closed-form option prices.

The class of SV models offers increased flexibility over GARCH-type specifications since they assume separate innovation processes for the conditional mean and the conditional variance of the observables. Seminal contributions in this area date back to Taylor (1994) and Taylor (2007). The use of variance-specific disturbances comes to the use of latent variables and this calls for the use of suitable inference tools such as stochastic filtering and simulation based inference. A Bayesian approach to inference provides a natural way to include simulation procedures, such as MCMC, in the estimation process. **?** propose a suitable Bayesian inference procedure for Gaussian SV models based on MCMC technique for posterior approximation. Extensions within the Bayesian univariate SV framework include the non-Gaussian SV with leverage Jacquier et al. (2004), the Markov-switching SV

[1]SV-LHARG(p) stands for Stochastic Volatility Heterogeneous Autoregressive model with Leverage effect.

So et al. (1998), an the threshold autoregressive SV So et al. (2002) models. The univariate SV model has been successfully extended to the multivariate context in order to capture dependencies and spillover effects between the volatility of different variables (e.g. see Harvey et al., 1994). Bayesian multivariate SV models which have Gamma autoregressive processes as special univariate case include for example Philipov and Glickman (2006); Asai and McAleer (2009); Casarin et al. (2016). See also Asai et al. (2006) for a review. Other extensions of the SV models related to the work presented in this chapter are Takahashi et al. (2009); Shirota et al. (2014); Bekierman and Gribisch (2016) for the univariate case and Jin and Maheu (2012) for the multivariate case, which both propose to augment the state space SV model with an RV equation. In spite of the flexibility of these classes of models in modelling volatility and forecasting return, they do not admit either exact filtering or analytical pricing formulae.

The second main contribution of this work is to provide analytical filtering and smoothing for the latent variables of the SV-LHARG(p) in the case of $p = 1$ and no leverage components. This basic instance of our new family of SV models, that we denote SV-ARG constitutes a significant contribution to the nonlinear and non-Gaussian filtering literature. In fact, it is well known (see, e.g. Harvey (1989)) that only for the classes of Gaussian and linear state-space and finite-state state-space models analytical filtering and smoothing recursions are available. For the other classes of models, with some exceptions (e.g., see Smith and Miller (1986), Shephard (1994), Ferrante and Vidoni (1998), Vidoni (1999), Deschamps (2011), de Pinho et al. (2016)), filtering and smoothing are obtained through analytical or numerical approximation techniques (see, e.g. Tanizaki (1996) and Doucet et al. (2001)). More specifically, our framework extends the results given in Creal (2015) for non-linear and non-Gaussian models with one measurement equation to the case of two-measurement equations. Indeed, for the SV-ARG, the analytical filtering and smoothing recursions allows for evaluating the likelihood exactly and developing effective inference procedures.

The third contribution of the work is to provide a Bayesian inference approach for both parameters and latent CV of a SV-LHARG(p) and an efficient MCMC procedure for posterior approximation. The analytical filtering results for the SV-ARG model allow us to develop an effective Forward Filtering Backward Sampling FFBS algorithm for the SV-ARG and an effective Metropolis-Hastings algorithm for a general SV-LHARG(p) model by using the SV-ARG as an auxiliary model in combination with a blocking sampling strategy introduce by Shephard and Pitt (1997) for state-space models and then successfully employed in combination with MH for latent variable estimation (e.g., see So (2006), Casarin et al. (2011) and Billio et al. (2016)). Following the literature on simulation-based inference for SV models we test the effectiveness of the MCMC algorithm in three different scenarios of low, medium and high persistence (e.g., see Chib et al. (2002) and Casarin et al. (2009)). Our extensive experimentation shows that the inefficiency factor and the mean square estimation errors of our MCMC procedure are in line with the results obtained for other SV models in Chib et al. (2002). Finally, by applying our fully tractable SV-LHARG model on a large sample of S&P500 index options, we show its superior option pricing performances in comparison with competitor models taken from both the GARCH and RV option pricing literature. In fact, the filtering and smoothing procedure in the SV-LHARG leads to an higher persistence of the latent volatility process, thus allowing the SV-LHARG to over-perform the realized volatility model RV-LHARG in pricing medium to long maturity options.

The remainder of the chapter proceeds as follows. Section 4.2 introduces the general SV-LHARG(p). Section 4.3 discusses in detail two nested specifications of the SV-

LHARG(p), the SV-ARG and SV-LHARG. Section 4.4 describes the Bayesian inference procedure and Section 4.5 shows the effectiveness of the estimation methodology on simulated time series. Finally, Section 4.6 presents a financial application where we benchmark the performance of the SV-LHARG model with that of competitor models in an option pricing exercise.

4.2 The model

Here we introduce the SV-LHARG(p) model in a general framework. We present some properties and discuss some special cases.

4.2.1 Dynamics under physical probability $\mathbb{P}$

We consider a risky asset with price S_t and geometric log-return $r_t = \log(S_{t+1}/S_t)$ computed from closing prices. Besides, we indicate with CV_t a latent volatility process and with $\boldsymbol{x}_{it}$, $i \in \{1, 2\}$, two vectors of exogenous variables. Let $\mathcal{F}_t \doteq \sigma(r_t, \mathrm{RV}_t, \boldsymbol{x}_{1,t+1}, \boldsymbol{x}_{2,t+1})$ the σ-algebra containing the information about observable quantities (log-return, realized variance RV_t, and exogenous variables) available at time t, and $\widetilde{\mathcal{F}}_t^{\mathrm{CV}} \doteq \sigma(\mathcal{F}_{t-1}, \mathrm{CV}_t)$. We assume the following model for the dynamics of the log-returns

$$r_t = \mu + \boldsymbol{\kappa}_1' \boldsymbol{x}_{1t} + \gamma \mathrm{CV}_t + \sqrt{\mathrm{CV}_t}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \tag{4.1}$$

$t = 1, \ldots, T$, where: (i) μ is the risk-free rate, (ii) $\boldsymbol{\kappa}_1$ is a n_1-dimensional vector of real parameters, (iii) γ is the market price of risk. $\mathcal{N}(m, \sigma^2)$ indicates the univariate normal distribution with mean m and variance σ^2. The covariates vector $\boldsymbol{x}_{1t}$ can be used to capture a jump component in the volatility process, in order to improve the fitting properties of the model under the physical measure $\mathbb{P}$ and the flexibility under risk neutral one (see, for instance Alitab et al., 2015). We refer to Equation (4.1) as *return equation*. The dynamics in Equation (4.1) has a well-precise justification. In a continuous time setting the log-returns dynamics is described by the following SDE

$$d\mathrm{R}_t = (\mu + \gamma\sigma_t^2)\, dt + \sigma_t\, dW_t, \tag{4.2}$$

with μ and γ as above, W_t a standard Brownian motion and σ_t a stochastic process indicating the volatility log-returns. Ané and Geman (2000) showed that the dynamics described in Equation (4.2) can be interpreted as a Brownian motion (with drift) with a change of time of the type $t \to \mathrm{IV}_t$, where IV_t is the integrated variance, i.e.

$$\mathrm{IV}_t = \int_0^t \sigma_s^2\, ds.$$

A well-known result in financial econometrics literature is that the daily log-return has a marginal distribution which is not consistent with a Gaussian one. Precisely, Clark (1973) theoretically argues that, for an underlying continuous-time diffusion process, the standard Gaussian distribution can be recovered by rescaling the log-return by an appropriate measure of the market activity, such as the integrated variance. When dealing with continuous-time diffusion, the integrated variance corresponds to the so called quadratic variation QV_t, defined as

$$\mathrm{QV}_t = \lim_{\|M_n\| \to 0} \sum_{i=1}^{n} \left(R_{t_i} - R_{t_{i-1}}\right)^2,$$

where $\|M_n\|$ is the length of the longest sub-interval of an equally spaced partition of the interval $[0, t]$. So, we have that

$$\mathrm{R}_t|\mathrm{QV}_t \sim \mathcal{N}\left(\mu\, t + \gamma \mathrm{QV}_t, \mathrm{QV}_t\right).$$

In particular, Equation (4.1) can be recovered by identifying CV_t with QV_t. We stress that the dynamics in Equation (4.1) differs from that employed in Corsi et al. (2013); Majewski et al. (2015) for daily log-returns inasmuch in these works authors consider as driving process for returns a realized measure of volatility, i.e. the realized measure of volatility and the true conditional variance are used as synonymous.

A realized measure of volatility contains information on the latent conditional variance process. Nontheless, on empirical data, they are affected from measurement errors and overnight bias. In spirit of Hansen and Lunde (2006); Engle and Gallo (2006); Shephard and Sheppard (2010); Takahashi et al. (2009), we add a second measurement equation to the model, which relates the RV_t to the latent CV_t. Precisely, we assume

$$\mathrm{RV}_t|\mathrm{CV}_t \overset{i.d}{\sim} \mathcal{G}\left(\alpha e^{-\boldsymbol{\beta_2'} \boldsymbol{x_{2t}}}, \mathrm{CV}_t e^{\boldsymbol{\beta_2'} \boldsymbol{x_{2t}}}\right), \tag{4.3}$$

where $\alpha \in \mathbb{R}_+$ is constant, and $\boldsymbol{\kappa}_2$ is a n_2-dimensional vector of real parameters. In the previous equation, $\mathcal{G}(k, \vartheta)$ denotes a Gamma distribution with positive shape, k, and scale parameter, ϑ (see Appendix A.2 for the definition of the Gamma distribution). The covariates vector $\boldsymbol{x}_{2t}$ could be employed to describe possible structural changes in the parameter α. A more thorough motivation on Equation (4.3) is given in the following. The idea of RV goes back to the seminal work of Merton (1980), in which he showed that the integrated variance of a Brownian motion can be approximated by the sum of a large number of intra-day squared returns. This idea has been recently formalized and generalized by Andersen et al. (2001) – using the quadratic variation theory – to the class of finite mean semi-martingales. Precisely, the sum of intra-day squared returns converges, as the maximal temporal length of returns goes to zero, to the integrated volatility of the prices, where the convergence is intended in probability. So, theoretically, one can build up an estimate of the actual daily volatility without any error. However, in practice, RV presents some biases. Indeed, for an equally spaced returns series, the RV over a time interval of one day (i.e., $t = 1$ day) is defined as (see Andersen et al., 2001; Barndorff-Nielsen, 2002; Barndorff-Nielsen and Shephard, 2001; Barndorff-Nielsen, 2002; Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2005, among others)

$$\mathrm{RV}_t^{\Delta_n} = \sum_{i=0}^{n-1} r_{t-j\Delta_n}^2, \tag{4.4}$$

where $\Delta_n = \frac{t}{n}$ and $r_{t-j\Delta}$ indicates the continuously compounded Δ_n-frequency returns, that is, intra-day returns sampled at time interval Δ_n. It is clear from (4.4) that the realized volatility is constructed only from available intra-day returns, i.e. the volatility during the close market period is missed from the definition. So far, the literature on stock market realized volatility has adopted several approaches to deal with this overnight bias. Corsi et al. (2008) and Wu (2011) simply ignore the overnight period. Becker et al. (2007), Pooter et al. (2008), Blair et al. (2010) and Bollerslev et al. (2009) calculate the overnight return by subtracting each day close from the next day open value of the stock and add this squared return to the summation (4.4). Martens (2002) and Koopman et al. (2005) calculate realized volatility by ignoring the overnight period, but then scaling the

resulting value upward so that the volatility estimate covers an entire 24-hour day. In this work, instead, we want to adjust this overnight bias within the model.

Moreover, the distribution theory for the RV in relation to IV has been derived in Barndorff-Nielsen (2002). They have shown that the law of the scaled difference between the realized volatility and the integrated one has a mixed Gaussian limit,

$$\sqrt{n}\,(\mathrm{RV}_t - \mathrm{IV}_t) \xrightarrow{d} \mathrm{MN}\,(0, 2\mathrm{IQ}_t) \qquad n \to \infty, \tag{4.5}$$

where IQ_t is the integrated quarticity

$$\mathrm{IQ}_t = \int_0^t \sigma_s^4\, ds.$$

It is intuitive that RV is less precise when σ is high since the size of the error bounds in Equation (4.5) is positively related to σ. Equation (4.3) is intended to take charge also of this bias[2]. This is better understood if we write down the conditional moments of the realized variance, which can be readily obtained from the distributional assumption on RV_t. In particular, the first two conditional moments are

$$\mathbb{E}^{\mathbb{P}}\left[\mathrm{RV}_t|\widetilde{\mathcal{F}}_t^{\mathrm{CV}}\right] = \alpha\mathrm{CV}_t, \qquad \mathbb{V}^{\mathbb{P}}\left[\mathrm{RV}_t|\widetilde{\mathcal{F}}_t^{\mathrm{CV}}\right] = \alpha\exp\left(\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)\mathrm{CV}_t^2.$$

The conditional mean of RV_t suggests that α adjusts the overnight bias of the RV estimator. In particular, in our financial application we use data for the S&P 500 index which is quoted from 9:30 AM to 4:00 PM Eastern time, so we should expect $\alpha < 1$. The conditional variance, instead, indicates that the error of the RV estimator is proportional to the square of the CV, in agreement with the distribution theory for the RV (Barndorff-Nielsen, 2002).

Finally, we define the dynamics for the CV process. It has been proved in the literature that a stochastic volatility model does not price correctly options with long or short maturity and out-of-the-money options. This is primarily due to the fact that a single factor of volatility running on a single time scale is not sufficient to completely describe the dynamics of the volatility (see Müller et al., 1997, among others). In particular, two are the main approaches that have been used so far to account for dependencies among volatilities at different time-scales. The first one is to decompose the daily volatility into a short- and a long-run volatility component and to describe them independently (see, for instance Christoffersen et al., 2008; Fouque and Lorig, 2011). The second approach is to define factors as an average of past volatilities over different time scales (see, for instance Corsi, 2009; Majewski et al., 2015). In the present work, we follow the second approach. Additionally, we adopt a multi-component structure also for the leverage, whose importance has been stressed by Corsi and Renò (2012). Regarding the functional form, we follow Majewski et al. (2015) and we assume that the leverage corresponds to a quadratic function of the conditional variance

$$l_{t-1} = \left(\epsilon_{t-1} - \lambda\sqrt{\mathrm{CV}_t}\right)^2 \qquad \lambda > 0.$$

This structure for the leverage is similar to the one in Heston and Nandi (2000). It captures the fact that large positive idiosyncratic component ϵ_{t-1} has a smaller impact on CV_t that large negative one. Introducing the notation $\boldsymbol{l}_{t-1} = (l_{t-1}, \ldots, l_{t-p})'$,

[2] Actually, RV is altered also by the market micro-structure noise – caused by infrequent trading, bid-ask spread, and rounding effects. Various methods are available in the literature to mitigate this distortion (see, for example Hansen and Lunde, 2005; Zhang et al., 2005; Bandi and Russell, 2006, 2008; Barndorff-Nielsen et al., 2008). In this work, we adopt the Two Scale estimator of Zhang et al. (2005)

$\mathbf{CV}_{t-1} = (\mathrm{CV}_{t-1}, \ldots, \mathrm{CV}_{t-p})'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$, we assume that CV_t follows an autoregressive gamma process (see Gouriéroux and Jasiak, 2006) with transition distribution:

$$\mathrm{CV}_t | \widetilde{\mathcal{F}}_{t-1}^{\mathrm{CV}}, r_{t-1} \overset{d}{\sim} \bar{\mathcal{G}}(\nu, \varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1}), c). \tag{4.6}$$

In the previous equation, $\bar{\mathcal{G}}(\nu, \varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1}), c)$ denotes the non-central gamma distribution with shape $\nu > 0$, scale $c > 0$ and non-centrality $\varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1})$. The non-centrality parameter is given by

$$\varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1}) = \sum_{i=1}^{p} \beta_i \mathrm{CV}_{t-i} + \sum_{i=1}^{p} \alpha_i l_{t-i},$$

where $\beta_i \in \mathbb{R}^+$ and $\alpha_i \in \mathbb{R}^+$ are the autoregressive coefficients and leverage coefficients, respectively. Here, we use an alternative re-parametrization of the non-centrality equation in the following way

$$\varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1}) = \sum_{i=1}^{p} \frac{\phi_i}{c} \mathrm{CV}_{t-i} + \sum_{i=1}^{p} \alpha_i l_{t-i},$$

where $\phi_i = \beta_i c$[3]. Using the Poisson mixture representation for the non-central gamma distribution (see Gouriéroux and Jasiak, 2006, for more details), we rewrite Equation (4.6) as

$$\begin{aligned} &\mathrm{CV}_t | z_t \overset{i.d}{\sim} \mathcal{G}(\nu + z_t, c), \\ &z_t | \mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1} \overset{i.d}{\sim} \mathcal{P}(\varphi(\mathbf{CV_{t-1}}, \boldsymbol{l}_{t-1}), \end{aligned}$$

where, in general, $\mathcal{P}o(v)$ indicates the Poisson distribution with intensity parameter v. The latter representation is useful for both the characterization of CV_t and the inference procedure. The conditional mean and variance of the process CV_t are given by:

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_t | \widetilde{\mathcal{F}}_{t-1}^{\mathrm{CV}}, r_{t-1}\right] &= c\nu + c\left(\sum_{i=1}^{p} \beta_i \mathrm{CV}_{t-i} + \sum_{i=1}^{p} \alpha_i l_{t-i}\right), \\ \mathbb{V}^{\mathbb{P}}\left[\mathrm{CV}_t | \widetilde{\mathcal{F}}_{t-1}^{H}, r_{t-1}\right] &= c^2\nu + 2c^2\left(\sum_{i=1}^{p} \beta_i \mathrm{CV}_{t-i} + \sum_{i=1}^{p} \alpha_i l_{t-i}\right). \end{aligned} \tag{4.7}$$

In particular, the first and second order conditional moments are affine functions of the lagged values of the volatility process and of the leverage.

Summarising, the SV-LHARG(p) is described by the following distributional representation:

$$\begin{aligned} &r_t | \widetilde{\mathcal{F}}_t^{\mathrm{CV}} \overset{d}{\sim} \mathcal{N}(\mu + \boldsymbol{\kappa}_1' \boldsymbol{x}_{1t} + \gamma \mathrm{CV}_t, \mathrm{CV}_t), \\ &\mathrm{RV}_t | \widetilde{\mathcal{F}}_t^{\mathrm{CV}} \overset{d}{\sim} \mathcal{G}\left(\alpha e^{-\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}}, \mathrm{CV}_t e^{\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}}\right), \\ &\mathrm{CV}_t | z_t \overset{d}{\sim} \mathcal{G}(\nu + z_t, c), \\ &z_t | \mathrm{CV}_{t-1}, \boldsymbol{l}_{t-1} \overset{d}{\sim} \mathcal{P}o(\varphi(\mathrm{CV}_{t-1}, \boldsymbol{l}_{t-1})), \end{aligned} \tag{4.8}$$

[3]We use interchangeably the two parametrizations in the reminder of the chapter and we make clear the precise version used whenever relevant.

$t = 1, \dots, T$ with initial conditions $\mathrm{CV}_0, \dots, \mathrm{CV}_{p+1}$. Because of our empirical applications, in this section we set $n_1 = n_2 = 1$ and impose $x_{1t} = 0$ and $x_{2t} = 1$ for all $t \in \{1, \cdots, T\}$. We present now the properties of SV-LHARG(p)[4].

First, thanks to the relations in Equation (4.7), it is possible to compute analytically the invariant unconditional mean and variance of both the latent and the observable variable. The following relations hold true:

$$
\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_1\right] = \frac{c\left(\nu + \sum_{i=1}^{p} \alpha_i\right)}{1 - c\left(\sum_{i=1}^{p} \beta_i + \lambda^2 \sum_{i=1}^{p} \alpha_i\right)}, \\
&\mathbb{V}^{\mathbb{P}}\left[\mathrm{CV}_1\right] = \frac{c^2\left(\nu + 2\sum_{i=1}^{p} \alpha_i + 2\sum_{i=1}^{p} \alpha_i^2\right) + 2c^2\mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_1\right]\left(\sum_{i=1}^{p} \beta_i + \lambda^2 \sum_{i=1}^{p} \alpha_i + 4\lambda^2 \sum_{i=1}^{p} \alpha_i^2\right)}{1 - c^2\left(\sum_{i=1}^{p} \beta_i^2 + \lambda^4 \sum_{i=1}^{p} \alpha_i^2\right)}, \\
&\mathbb{E}^{\mathbb{P}}\left[r_1\right] = \mu + \gamma\mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_1\right], \\
&\mathbb{V}^{\mathbb{P}}\left[r_1\right] = \mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_1\right] + \gamma^2\mathbb{V}^{\mathbb{P}}\left[\mathrm{CV}_1\right], \\
&\mathbb{E}^{\mathbb{P}}\left[\mathrm{RV}_1\right] = \alpha\mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_1\right], \\
&\mathbb{V}^{\mathbb{P}}\left[\mathrm{RV}_1\right] = \left(\alpha\, e^{\kappa_2} + \alpha^2\right)\mathbb{V}^{\mathbb{P}}\left[\mathrm{CV}_1\right] + \alpha\, e^{\kappa_2}\mathbb{E}^{\mathbb{P}}\left[\mathrm{CV}_1\right].
\end{aligned}
\tag{4.9}
$$

Second, the SV-LHARG(p) satisfies the affine property. We remind that a stochastic process is called affine if the logarithm of the characteristic function of its transition distribution is affine with respect to the initial state. Moreover, the importance of the affine property in finance has been acknowledged in many studies (see, for instance Duffie et al., 2000; Darolles et al., 2006; Majewski et al., 2015). Its advantage is manifold. It allows us to provide an exhaustive probabilistic description of both log-return and CV dynamics. In particular, we obtain a closed-form expression for the conditional moment generating function of the SV-LHARG(p) under the physical and risk-neutral measure, as well as we derive an explicit one-to-one mapping between the parameters of the SV-LHARG(p) under the measures $\mathbb{P}$ and $\mathbb{Q}$.

Proposition 4.2.1. *For the* SV-LHARG(p) *process the following relation holds*

$$
\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[e^{\bar{z}r_{t+1} + \bar{b}\mathrm{CV}_{t+1} + \bar{c}l_{t+1} + \bar{d}\mathrm{RV}_{t+1}} | \widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t, l_t\right] \\
&\quad = \exp\left(\mathcal{A}\left(\bar{z}, \chi, \bar{c}\right) + \sum_{i=1}^{p} \mathcal{B}_i\left(\bar{z}, \chi, \bar{c}\right)\mathrm{CV}_{t+1-i} + \sum_{i=1}^{p} \mathcal{C}_i\left(\bar{z}, \chi, \bar{c}\right) l_{t+1-i}\right),
\end{aligned}
$$

where the functions $\mathcal{A} : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\mathcal{B}_i : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, *and* $\mathcal{C}_i : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $i = 1, \dots, p$, *are defined as:*

$$
\begin{aligned}
&\mathcal{A}(\bar{z}, \chi, \bar{c}) = \bar{z}\mu - \frac{1}{2}\log\left(1 - 2\bar{c}\right) - \nu\mathcal{W}(x(\bar{z}, \chi, \bar{c}), c), \\
&\mathcal{B}_i(\bar{z}, \chi, \bar{c}) = \mathcal{V}(x(\bar{z}, \chi, \bar{c}), c)\beta_i, \\
&\mathcal{C}_i(\bar{z}, \chi, \bar{c}) = \mathcal{V}(x(\bar{z}, \chi, \bar{c}), c)\alpha_i,
\end{aligned}
$$

with

$$
\begin{aligned}
&\mathcal{V}(x, c) = \frac{cx}{1 - cx}, \\
&\mathcal{W}(x, c) = \log\left(1 - cx\right), \\
&x(\bar{z}, \chi, \bar{c}) = \bar{z}\gamma + \chi + \frac{\frac{1}{2}\bar{z}^2 + \lambda^2\bar{c} - 2\bar{c}\bar{z}\lambda}{1 - 2\bar{c}}, \\
&\chi = \bar{b} - e^{\kappa_2}\log(1 - \alpha e^{-\kappa_2}\bar{d}).
\end{aligned}
$$

[4]The properties of the SV-LHARG(p) largely follow from the results in Majewski et al. (2015) and so we will refer to the original work of Majewski et al. (2015) for the complete proofs when necessary.

Proof. See Appendix A.1. □

Then, following the procedure in Majewski et al. (2015), Appendix C, one can easily derive the moment generating function of the SV-LHARG(p) under the physical measure $\mathbb{P}$.

Proposition 4.2.2. *Under* $\mathbb{P}$*, the* MGF *for* SV-LHARG(p) *model has the following form*

$$\varphi^{\mathbb{P}}(t,T,\bar{z}) = \mathbb{E}^{\mathbb{P}}\left[e^{\bar{z}\, r_{t,T}} | \widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t, l_t\right] = \exp\left(\bar{\mathrm{a}}_t + \sum_{i=1}^{p} \bar{\mathrm{b}}_{t,i}\mathrm{CV}_{t+1-i} + \sum_{i=1}^{p} \bar{\mathrm{c}}_{t,i}\mathrm{l}_{t+1-i}\right) \quad (4.10)$$

where

$$\begin{aligned}
\bar{\mathrm{a}}_s &= \bar{\mathrm{a}}_{s+1} + \bar{z}\mu - \frac{1}{2}\log\left(1 - 2\bar{\mathrm{c}}_{s+1,1}\right) - \nu\mathcal{W}\left(x_{s+1}, c\right) \\
\bar{\mathrm{b}}_{s,i} &= \begin{cases} \bar{\mathrm{b}}_{s+1,i+1} + \mathcal{V}\left(x_{s+1}, c\right)\beta_i & for\ 1 \le i \le p-1 \\ \mathcal{V}\left(x_{s+1}, c\right)\beta_i & for\ i = p \end{cases} \\
\bar{\mathrm{c}}_{s,i} &= \begin{cases} \bar{\mathrm{c}}_{s+1,i+1} + \mathcal{V}\left(x_{s+1}, c\right)\alpha_i & for\ 1 \le i \le p-1 \\ \mathcal{V}\left(x_{s+1}, c\right)\alpha_i & for\ i = p \end{cases}
\end{aligned}$$

with

$$x_{s+1} = \bar{z}\gamma + \bar{\mathrm{b}}_{s+1,1} + \frac{\frac{1}{2}\bar{z}^2 + \lambda^2\bar{\mathrm{c}}_{s+1,1} - 2\bar{\mathrm{c}}_{s+1,1}\bar{z}\lambda}{1 - 2\bar{\mathrm{c}}_{s+1,1}},$$

and the terminal conditions read $\bar{\mathrm{a}}_T = \bar{\mathrm{b}}_{T,i} = \bar{\mathrm{c}}_{T,i} = 0$ *for* $i = 1, \cdots, p$.

Mimicking the reasoning in Gouriéroux and Jasiak (2006), Appendix F, we also derive the stationary condition for CV_t process.

Proposition 4.2.3. *The* SV-LHARG(p) *in Equation* (4.8) *is stationary if the following condition*

$$c\left(\sum_{i=1}^{p}\beta_i + \lambda^2\sum_{i=1}^{p}\alpha_i\right) < 1$$

is satisfied.

Proof. See Appendix A.1. □

4.2.2 Dynamics under risk-neutral probability $\mathbb{Q}$

We turn now to analyse the risk-neutral dynamics of the SV-LHARG(p). In order to preserve the analytical tractability of the model, we risk-neutralize it by employing a Stochastic Discount Factor (SDF) within the exponential affine family, whose high flexibility allows to incorporate multiple factor-dependent risk premia. This approach has been extensively used in literature (see Bertholon et al., 2008; Gagliardini et al., 2011; Corsi et al., 2013; Majewski et al., 2015; Alitab et al., 2015, among others). We specify the following SDF

$$\mathrm{M}_{t,t+1} = \frac{e^{-\nu_1\mathrm{CV}_{t+1} - \nu_2 r_{t+1}}}{\mathbb{E}^{\mathbb{P}}\left[e^{-\nu_1\mathrm{CV}_{t+1} - \nu_2 r_{t+1}} | \widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t\right]}, \quad (4.11)$$

which represents the Esscher transform from the physical log-return density to the risk neutral one (see, for instance Gerber et al., 1994; Bühlmann et al., 1996). The main advantage of the SDF (4.11) is to clearly identify the sources of risk and explicitly compensate

them with separated risk premia. Specifically, this form allows for both a conditional variance premium, ν_1, and the standard equity premium, ν_2. The latter has to satisfy the following no-arbitrage condition[5].

Proposition 4.2.4. *The* SV-LHARG(p) *model defined by Equation* (4.8) *with* SDF *given by* (4.11) *satisfies the no-arbitrage condition if and only if*

$$\nu_2 = \gamma + \frac{1}{2}.$$

Proof of Proposition 4.2.4. See Majewski et al. (2015), Appendix C. □

The MGF under the risk-neutral measure $\mathbb{Q}$ for SV-LHARG(p) reads

Proposition 4.2.5. *Under* $\mathbb{Q}$*, the MGF for SV-LHARG(p) model has the following form*

$$\varphi^{\mathbb{Q}}(t,T,\bar{z}) = \exp\left(\bar{\mathrm{a}}^*_t + \sum_{i=1}^{p} \bar{\mathrm{b}}^*_{t,i}\mathrm{CV}_{t+1-i} + \sum_{i=1}^{p} \bar{\mathrm{c}}^*_{t,i}\mathrm{l}_{t+1-i}\right) \tag{4.12}$$

where

$$\begin{aligned}
\bar{\mathrm{a}}^*_s &= \bar{\mathrm{a}}^*_{s+1} + \bar{z}\mu - \frac{1}{2}\log\left(1-2\bar{\mathrm{c}}^*_{s+1,1}\right) - \nu\mathcal{W}\left(x^*_{s+1},c\right) + \nu\mathcal{W}\left(y^*_{s+1},c\right)\\
\bar{\mathrm{b}}^*_{s,i} &= \begin{cases} \bar{\mathrm{b}}^*_{s+1,i+1} + \left(\mathcal{V}\left(x^*_{s+1},c\right) - \mathcal{V}\left(y^*_{s+1},c\right)\right)\beta_i & for\ 1 \le i \le p-1\\ \left(\mathcal{V}\left(x^*_{s+1},c\right) - \mathcal{V}\left(y^*_{s+1},c\right)\right)\beta_i & for\ i = p \end{cases}\\
\bar{\mathrm{c}}^*_{s,i} &= \begin{cases} \bar{\mathrm{c}}^*_{s+1,i+1} + \left(\mathcal{V}\left(x^*_{s+1},c\right) - \mathcal{V}\left(y^*_{s+1},c\right)\right)\alpha_i & for\ 1 \le i \le p-1\\ \left(\mathcal{V}\left(x^*_{s+1},c\right) - \mathcal{V}\left(y^*_{s+1},c\right)\right)\alpha_i & for\ i = p \end{cases}
\end{aligned}$$

with

$$y^*_{s+1} = -\nu_2\gamma - \nu_1 + \frac{1}{2}\nu_2^2,$$

and the terminal conditions read $\bar{\mathrm{a}}^*_T = \bar{\mathrm{b}}^*_{T,i} = \bar{\mathrm{c}}^*_{T,i} = 0$ *for* $i = 1, \cdots, p$.

In this work, we derive the price of vanilla options. In this case it is sufficient to know only the MGF under the risk-neutral measure $\mathbb{Q}$, which has been given in the previous proposition. Precisely, the comparison of the two MGFs, under $\mathbb{P}$ and $\mathbb{Q}$, permits to derive the one-to-one mapping of the parameters under the historical measure and the risk-neutral one. This mapping ensures that the risk-neutral log-return dynamics is still governed by a SV-LHARG(p) process.

Proposition 4.2.6. *Under risk-neutral measure* $\mathbb{Q}$ *the conditional variance follows a* SV-LHARG(p) *process with parameters*

$$\begin{aligned}
&\beta^{(*)} = \frac{1}{1-cy^*}\beta, && \alpha^{(*)} = \frac{1}{1-cy^*}\alpha && c^* = \frac{1}{1-cy^*}c,\\
&\nu^* = \nu, && \lambda^* = \lambda + \gamma + \frac{1}{2}, &&
\end{aligned}$$

where $y^* = -\frac{\gamma^2}{2} - \nu_1 + \frac{1}{8}$.

Proof of Proposition 4.2.6. See Majewski et al. (2015), Appendix D. □

[5]We remind that the no-arbitrage condition is derived by imposing

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[\mathrm{M}_{t,t+1}|\widetilde{\mathcal{F}}^{\mathrm{CV}}_t, r_t\right] = 1 && for\, t \in \mathbb{Z}_+\\
&\mathbb{E}^{\mathbb{P}}\left[\mathrm{M}_{t,t+1}e^{r_{t+1}}|\widetilde{\mathcal{F}}^{\mathrm{CV}}_t, r_t\right] = e^{\mu} && for\, t \in \mathbb{Z}_+
\end{aligned}$$

4.3 Particular cases

We now discuss two specifications of SV-LHARG(p) which will be used in our empirical application. The first one is a SV-LHARG(p) model without leverage and heterogeneous structure and it will be labelled SV-ARG. The SV-ARG turns out to be crucial in the estimation phase and it extends the class of non-Gaussian state-space models introduced in Creal (2015). The second specification is a SV-LHARG model, which is an heterogeneous autoregressive model for the CV with leverage term. This type of parametrization was introduced in Majewski et al. (2015) to describe the dynamics of the RV.

Before presenting the models, let us introduce some notations. We define with $\boldsymbol{r}_{s:t} = (r_s, \dots, r_t)' \in \mathbb{R}^{t-s+1}$, $\mathbf{RV}_{s:t} = (\mathrm{RV}_s, \dots, \mathrm{RV}_t)' \in \mathbb{R}_+^{t-s+1}$, $\mathbf{CV}_{s:t} = (\mathrm{CV}_s, \dots, \mathrm{CV}_t)' \in \mathbb{R}_+^{t-s+1}$, $\boldsymbol{z}_{s:t} = (z_s, \dots, z_t)' \in \mathbb{R}_+^{t-s+1}$ and $\boldsymbol{x}_{s:t} = ((x_{1s}, x_{2s}), \dots, (x_{1t}, x_{2t}))' \in \mathbb{R}^{(s-p+1)\times(n_1+n_2)}$ the collections, from time s to time t, of the log-daily returns, the realized variance, the conditional variance, the state and the exogenous variables respectively. Besides we denote with: (i) $\mathcal{S}(\lambda, \chi, \psi)$ the Sichel distribution with parameters $\lambda \in \mathbb{R}$, $\chi \in \mathbb{R}_+$, $\psi \in \mathbb{R}_+$, (ii) $\mathcal{NB}(\omega, p)$ the Negative Binomial distribution with parameters $\omega \in \mathbb{R}_+$ and $p \in (0,1)$, (iv) $\mathcal{G}ig(\lambda, \chi, \psi)$ the Generalized Inverse Gaussian distribution with parameters $\lambda \in \mathbb{R}$, $\chi \in \mathbb{R}_+$, $\psi \in \mathbb{R}_+$, (v) $K_\lambda(x)$ the modified Bessel function of the second kind. See Appendix A.2 for a definition of these distributions and functions.

4.3.1 The SV-ARG model

The SV-ARG is obtained from the SV-LHARG(p) in Equation (4.8) setting $p = 1$ and the centrality parameter $\varphi(\mathbf{CV_{t-1}}, \boldsymbol{l_{t-1}}) = \varphi(\mathrm{CV}_{t-1}) = \beta_1 \mathrm{CV}_{t-1} \doteq \beta^{(d)} \mathrm{CV}_{t-1}$, and is described by the following distributional representation

$$
\begin{aligned}
&r_t | \widetilde{\mathcal{F}}_t^{\mathrm{CV}} \overset{d}{\sim} \mathcal{N}(\mu + \boldsymbol{\kappa}_1^{'} \boldsymbol{x}_{1t} + \gamma \mathrm{CV}_t, \mathrm{CV}_t), \\
&\mathrm{RV}_t | \widetilde{\mathcal{F}}_t^{\mathrm{CV}} \overset{d}{\sim} \mathcal{G}\left(\alpha e^{-\boldsymbol{\kappa}_2^{'} \boldsymbol{x}_{2t}}, \mathrm{CV}_t e^{\boldsymbol{\kappa}_2^{'} \boldsymbol{x}_{2t}}\right), \\
&\mathrm{CV}_t | z_t \overset{d}{\sim} \mathcal{G}(\nu + z_t, c), \\
&z_t | \mathrm{CV}_{t-1} \overset{d}{\sim} \mathcal{P}o\left(\beta^{(d)} \mathrm{CV}_{t-1}\right),
\end{aligned}
\tag{4.13}
$$

$t = 1, \cdots, T$ with initial condition CV_0. We denote by $\boldsymbol{\theta} = \left(\mu, \boldsymbol{\kappa}_1', \gamma, \alpha, \boldsymbol{\kappa}_2', \nu, c, \beta^{(d)}\right)'$ the $(6 + n_1 + n_2)$-dimensional vector of parameters.

Following a similar argument as in Creal (2015) we are able to to provide analytical expressions for the: (i) conditional likelihood, (ii) Markov transition, (iii) initial distribution of z_t, (iv) filtering and the smoothing of the latent process CV_t.
For the SV-ARG the following two propositions hold true.

Proposition 4.3.1. *For the* SV-ARG *model described in Equation* (4.13) *the conditional likelihood,* $p(r_t, \mathrm{RV}_t | z_t, \boldsymbol{x}_t, \boldsymbol{x}_{2t}; \boldsymbol{\theta})$*, the Markov transition,* $p(z_t | z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})$*, and the initial distribution of* z_t*,* $p(z_1; \boldsymbol{\theta})$*, are respectively given by:*

$$
\begin{aligned}
&p(r_t, \mathrm{RV}_t | z_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = 2\eta(z_t, \mathrm{RV}_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) K_{\lambda(z_t)}\left(\sqrt{\psi \chi^{(t)}}\right) \left(\sqrt{\frac{\chi^{(t)}}{\psi}}\right)^{\lambda(z_t)}, \\
&p(z_t | z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) \propto \mathcal{S}\left(\lambda(z_{t-1}), \chi^{(t-1)} \frac{\phi^{(d)}}{c}, \psi \frac{c}{\phi^{(d)}}\right), \\
&p(z_1; \boldsymbol{\theta}) \propto \mathcal{NB}\left(\nu, \phi^{(d)}\right),
\end{aligned}
$$

with

$$\eta(z_t, \mathrm{RV}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = \frac{\exp(\gamma\mu_{1t})}{\sqrt{2\pi}} \frac{\mathrm{RV}^{\alpha_t - 1}}{\Gamma(\alpha_t)\left(\exp\left(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right)\right)^{\alpha_t}} \frac{1}{\Gamma(\nu + z_t) c^{\nu + z_t}},$$
$$\mu_{1t} = r_t - \mu - \boldsymbol{\kappa}_1' \boldsymbol{x}_{1t},$$
$$\alpha_t = \alpha e^{-\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}},$$
$$\lambda(z_t) = \nu + z_t - \alpha_t - 1/2,$$
$$\chi^{(t)} = \mu_{1t}^2 + 2\mu_{2t},$$
$$\mu_{2t} = \frac{\mathrm{RV}_t}{\exp(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t})},$$
$$\psi = \gamma^2 + \frac{2}{c}.$$

Proof. See Appendix A.3. □

In other words, the variable CV_t can be integrated out analytically leaving the auxiliary variable z_t as the only state variable remaining in the model. The original SV-ARG is thus reformulated as a Markov-switching model with state variable z_t and known transition distribution. Moreover, conditional on the discrete variable z_t and the data RV_t and r_t, the state variable CV_t is independent of itself at other periods. This allows us to compute the filtered $p(\mathrm{CV}_t | \boldsymbol{r}_{1:t}, \mathbf{RV}_{1:t}, \boldsymbol{z}_{1:t}, \boldsymbol{x}_{1:t}; \boldsymbol{\theta})$ and smoothed $p(\mathrm{CV}_t | \boldsymbol{r}_{1:T}, \mathbf{RV}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}; \boldsymbol{\theta})$ distributions.

Proposition 4.3.2. *Let $\lambda(z_t)$, $\chi^{(t)}$ and ψ be the quantities defined in Proposition 4.3.1. The marginal filtered, $p(\mathrm{CV}_t | \boldsymbol{r}_{1:t}, \mathbf{RV}_{1:t}, \boldsymbol{z}_{1:t}, \boldsymbol{x}_{1:t}; \boldsymbol{\theta})$, and smoothed, $p(\mathrm{CV}_t | \boldsymbol{r}_{1:T}, \mathbf{RV}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}; \boldsymbol{\theta})$ distributions are*

$$p(\mathrm{CV}_t | \boldsymbol{r}_{1:t}, \mathbf{RV}_{1:t}, \boldsymbol{z}_{1:t}, \boldsymbol{x}_{1:t}; \boldsymbol{\theta}) \propto \mathcal{G}ig\left(\lambda(z_t), \chi^{(t)}, \psi\right),$$
$$p(\mathrm{CV}_t | \boldsymbol{r}_{1:T}, \mathbf{RV}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}; \boldsymbol{\theta}) \propto \mathcal{G}ig\left(\lambda(z_t) + z_{t+1}, \chi^{(t)}, \psi + 2\frac{\phi^{(d)}}{c}\right),$$

$t = 1, \cdots, T$.

Proof. See Appendix A.3. □

4.3.2 The SV-LHARG model

The SV-LHARG is obtained from the SV-LHARG(p) in Equation (4.8) setting $p = 22$ and making a parsimonious, but effective choice for the centrality parameter of the non-central gamma distribution.
Specifically, the non-centrality $\varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1}) \doteq \varphi^{(H)}(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1})$ is given by:

$$\varphi^{(H)}(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1}) = \beta^{(d)}\mathrm{CV}_{t-1}^{(d)} + \beta^{(w)}\mathrm{CV}_{t-1}^{(w)} + \beta^{(m)}\mathrm{CV}_{t-1}^{(m)} + \alpha^{(d)} l_{t-1}^{(d)} + \alpha^{(w)} l_{t-1}^{(w)} + \alpha^{(m)} l_{t-1}^{(m)},$$

with

$$\mathrm{CV}_{t-1}^{(d)} = \mathrm{CV}_{t-1}, \qquad l_{t-1}^{(d)} = l_{t-1},$$
$$\mathrm{CV}_{t-1}^{(w)} = \frac{1}{4}\sum_{i=2}^{5} \mathrm{CV}_{t-i}, \qquad l_{t-1}^{(w)} = \frac{1}{4}\sum_{i=2}^{5} l_{t-i},$$
$$\mathrm{CV}_{t-1}^{(m)} = \frac{1}{17}\sum_{i=6}^{22} \mathrm{CV}_{t-i}, \qquad l_{t-1}^{(m)} = \frac{1}{17}\sum_{i=6}^{22} l_{t-i}.$$

In the SV-LHARG, lagged terms of conditional variance and of leverage are collected in three different non-overlapping factors. A daily factor d (or short-term volatility factor), a weekly factor w (or medium-term volatility factor), and a monthly factor m (or long-term volatility factor). This specification permits to capture the desired properties of volatility – i.e. the memory persistence observed in financial data and the multi-component structure of volatility and leverage – while preserving parameters parsimony. The SV-LHARG is so described by the following distributional representation

$$\begin{aligned} r_t|\widetilde{\mathcal{F}}_t^{\mathrm{CV}} &\overset{d}{\sim} \mathcal{N}(\mu + \boldsymbol{\kappa}_1^{'}\boldsymbol{x}_{1t} + \gamma \mathrm{CV}_t, \mathrm{CV}_t), \\ \mathrm{RV}_t|\widetilde{\mathcal{F}}_t^{\mathrm{CV}} &\overset{d}{\sim} \mathcal{G}\left(\alpha e^{-\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}}, \mathrm{CV}_t e^{\boldsymbol{\kappa}_2^{'}\boldsymbol{x}_{2t}}\right), \\ \mathrm{CV}_t|z_t &\overset{d}{\sim} \mathcal{G}(\nu + z_t, c), \\ z_t|\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1} &\overset{d}{\sim} \mathcal{P}o\left(\varphi^{(H)}(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1})\right). \end{aligned} \tag{4.14}$$

We denote by $\boldsymbol{\theta} = \left(\mu, \boldsymbol{\kappa}_1', \gamma, \alpha, \boldsymbol{\kappa}_2', \nu, c, \beta^{(d)}, \beta^{(w)}, \beta^{(m)}, \alpha^{(d)}, \alpha^{(w)}, \alpha^{(m)}, \lambda\right)'$ the $(12 + n_1 + n_2)$-dimensional vector of parameters.

4.4 Estimation of models under physical probability measure

In this section, we present the estimation procedure for the SV-LHARG(p). Because of our empirical application we consider $p = 22$, although the methodology presented can be easily adapted to the SV-LHARG(p) without restrictions.

As commonly happens in latent variable models for time series analysis – such as SV models – the likelihood function of the SV-LHARG(p) is a high-dimensional integral that has not a closed-form solution. We thus apply a data-augmentation principle (see Tanner and Wong, 1987), thus obtaining a so called complete-data likelihood function which includes also the latent variables. Regarding the initial p values of the SV-LHARG(p), we follow Vermaak et al. (2004) and Casarin et al. (2012) and consider a pseudo-likelihood approach by assuming that the observations start at $t = p + 1$. We denote with $\boldsymbol{\xi}$ and $\boldsymbol{w}_t$ the two $(2+n_1)$-dimensional vectors $\boldsymbol{\xi} = (\mu, \boldsymbol{\kappa}_1', \gamma)'$ and $\boldsymbol{w}_t = (1, \boldsymbol{x}_{1t}', \mathrm{CV}_t)'$. The complete-data pseudo-likelihood is

$$\begin{aligned} \mathcal{L}\left(r_{p+1:T}, \mathbf{RV}_{p+1:T}, \mathbf{CV}_{p+1:T}, z_{p+1:T}|\boldsymbol{\theta}, \boldsymbol{x}_{p+1:T}\right) &= \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi \mathrm{CV}_t}} \exp\left(-\frac{1}{2}\frac{(r_t - \boldsymbol{\xi}'\boldsymbol{w}_t)^2}{\mathrm{CV}_t}\right) \\ \cdot \prod_{t=p+1}^{T} \frac{1}{\Gamma(\alpha_t)(\mathrm{CV}_t \exp(\boldsymbol{\kappa}_2^{'}\boldsymbol{x}_{2t}))^{\alpha_t}} \mathrm{RV}_t^{\alpha_t - 1} \exp\left(-\frac{\mathrm{RV}_t}{\mathrm{CV}_t \exp(\boldsymbol{\kappa}_2^{'}\boldsymbol{x}_{2t})}\right) \\ \cdot \prod_{t=p+1}^{T} \frac{1}{\Gamma(\nu + z_t)c^{\nu + z_t}} \mathrm{CV}_t^{\nu + z_t - 1} \exp\left(-\frac{\mathrm{CV}_t}{c}\right) \\ \cdot \prod_{t=p+1}^{T} \frac{1}{z_t!} \left(\varphi(\mathbf{CV}_{t-1}, \boldsymbol{l}_{t-1})\right)^{z_t} \exp\left(-\varphi(\mathbf{CV}, \boldsymbol{l}_{t-1})\right). \end{aligned}$$

In what follows, we describe the algorithm employed to compute the likelihood of the model, as well as the filtered and smoothed estimates of the latent variables. Specifically, we used a Bayesian inference. In a Bayesian framework, the first step is the specification of a prior distribution (i.e. a beliefs about the unknown parameters of the model) $\pi(\boldsymbol{\theta})$

on $\boldsymbol{\theta}$. Let $\boldsymbol{\beta} = \left(\beta^{(d)}, \beta^{(w)}, \beta^{(m)}\right)$ and $\boldsymbol{\alpha} = \left(\alpha^{(d)}, \alpha^{(w)}, \alpha^{(m)}\right)$, we assume:

$$\pi\left(\boldsymbol{\theta}\right) \propto \mathbb{I}_{\mathbb{R}^{2+n_1}}\left(\boldsymbol{\xi}\right)\mathbb{I}(\alpha)_{\mathbb{R}_+}\mathbb{I}_{\mathbb{R}^{n_2}}\left(\boldsymbol{\kappa_2}\right)\mathbb{I}_{\mathbb{R}_+}\left(\nu\right)\mathbb{I}_{\mathbb{R}_+}\left(\boldsymbol{\beta}\right)\mathbb{I}_{\mathbb{R}_+}\left(\boldsymbol{\alpha}\right)\mathbb{I}_{\mathbb{R}_+}\left(\lambda\right)\mathbb{I}_{\mathcal{A}_{\boldsymbol{\theta}}}\left(\boldsymbol{\theta}\right)$$

where $\mathbb{I}_{\Theta}\left(\theta\right)$ is the indicator function which takes value 1 if $\theta \in \Theta$ and 0 otherwise. In Equation (4.4) $\mathcal{A}_{\boldsymbol{\theta}}$ indicates the set of parameters values which satisfies the stationarity condition in Proposition 4.2.3. At this point, the second step is to write down the parameters and latent variables joint posterior distribution, which results to be proportional to the product between the prior distribution $\pi\left(\boldsymbol{\theta}\right)$ and the complete-data pseudo-likelihood function, i.e.

$$\begin{aligned}&\pi\left(\boldsymbol{\theta}, \mathbf{CV}_{p+1:T}, \boldsymbol{z}_{p+1:T} | \boldsymbol{r}_{p+1:T}, \mathbf{RV}_{p+1:T}, \boldsymbol{x}_{p+1:T}\right)\\ &\propto \pi(\boldsymbol{\theta})\mathcal{L}\left(\boldsymbol{r}_{p+1:T}, \mathbf{RV}_{p+1:T}, \mathbf{CV}_{p+1:T}, \boldsymbol{z}_{p+1:T} | \boldsymbol{\theta}, \boldsymbol{x}_{p+1:T}\right).\end{aligned}$$

In our case, this distribution is not tractable. As commonly happens in the literature, we rely our Bayesian estimation procedure on a MCMC algorithm and we develop a Gibbs sampling algorithm to generate random draws from the posterior distribution and to approximate all posterior quantities of interest (Casella and Robert, 2004). In Appendix A.4 a background material in MCMC is provided. In a data-augmentation framework, the estimation of the parameters under the physical measure $\mathbb{P}$ involves an extra computational cost due to the estimation of the latent variables. However, despite this difficulty in the computation of the likelihood, we use the analytical filtering and smoothing recursions derived for the SV-ARG model (see Subsection 4.3.1) in order to design an effective computational method for posterior approximation.

Our proposed MCMC algorithm for the SV-LHARG model iterates over the following steps:

- **1st step:** Initialize $\boldsymbol{\theta}$ – the set of parameters – and $\boldsymbol{z}_{p+1:T}$ and $\mathrm{CV}_{p+1:T}$ – the latent variables;
- **2st step:** Sample $\boldsymbol{\theta}$ given $(\boldsymbol{r}_{p+1:T}, \mathrm{RV}_{p+1:T}, \mathrm{CV}_{p+1:T}, \boldsymbol{z}_{p+1:T})$;
- **3st step:** Sample $(\boldsymbol{z}_{p+1:T}, \mathrm{CV}_{p+1:T})$ given $(\boldsymbol{r}_{p+1:T}, \mathrm{RV}_{p+1:T}, \boldsymbol{\theta})$;
- **4st step:** Go to the 2st step.

We now give a more thoroughly description of previous steps. Concerning the **2st step**, we consider the following eight blocks of parameters

$$\begin{aligned}&\left\{\left(\mu, \gamma, \boldsymbol{\kappa}_1'\right), \alpha, \boldsymbol{\kappa}_2', \nu, c\left(\beta^{(d)}, \beta^{(w)}, \beta^{(m)}\right), \left(\alpha^{(d)}, \alpha^{(w)}, \alpha^{(m)}\right), \lambda\right\}\\ &\doteq \left\{\boldsymbol{\theta}_1, \alpha, \boldsymbol{\kappa}_2', \nu, c, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \lambda\right\},\end{aligned}$$

where, for instance, the notation $(\mu, \gamma, \boldsymbol{\kappa}_1')$ means that μ, γ and $\boldsymbol{\kappa}_1'$ are sampled in one block, conditioned on the remaining blocks. These steps are important for reducing the serial dependence in the MCMC output. The second step of the Gibbs sampler given above is blocked further as follows:

- **2.1** Sample $\boldsymbol{\theta}_1$ from $\pi\left(\boldsymbol{\theta}_1 | \alpha, \boldsymbol{\kappa}_2, \nu, c, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \lambda\right)$.
- **2.2** Sample α from $\pi\left(\alpha | \boldsymbol{\theta}_1, \boldsymbol{\kappa}_2, \nu, c, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \lambda\right)$.
- **2.3** Sample $\boldsymbol{\kappa}_2$ from $\pi\left(\boldsymbol{\kappa}_2 | \boldsymbol{\theta}_1, \alpha, \nu, c, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \lambda\right)$.

- **2.4** Sample ν from $\pi\left(\nu|\boldsymbol{\theta}_1, \alpha, \boldsymbol{\kappa}_2, c, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \lambda\right)$.
- **2.5** Sample c from $\pi\left(c|\boldsymbol{\theta}_1, \alpha, \boldsymbol{\kappa}_2, \nu, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \lambda\right)$.
- **2.6** Sample $\boldsymbol{\theta}_2$ from $\pi\left(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \alpha, \boldsymbol{\kappa}_2, \nu, c, \boldsymbol{\theta}_3, \lambda\right)$.
- **2.7** Sample $\boldsymbol{\theta}_3$ from $\pi\left(\boldsymbol{\theta}_3|\boldsymbol{\theta}_1, \alpha, \boldsymbol{\kappa}_2, \nu, c, \boldsymbol{\theta}_2, \lambda\right)$.
- **2.8** Sample λ from $\pi\left(\lambda|\boldsymbol{\theta}_1, \alpha, \boldsymbol{\kappa}_2, \nu, c, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3\right)$.

Details on the full conditional distributions and the sampling methods for the parameters of the SV-LHARG (resp. SV-ARG) model are reported in Appendix A.6 Section A.6 (resp. A.3 Section A.5).

Then, regarding the **3st step** of the Gibbs sampler, we exploit the analytical filtering and smoothing results in Propositions 4.3.1 and 4.3.2 for the SV-ARG model and we construct a two-step procedure. Precisely, first we develop a Forward Filtering Backward Sampling (FFBS) procedure for the SV-ARG model (Frühwirth-Schnatter, 2006). Then, we develop an effective Metropolis-Hastings algorithm for the general SV-LHARG(p) by using the SV-ARG as an auxiliary model.

In the first step, we follow the strategy used in Creal (2015) and we reformulate the original SV-ARG model as a Markov-switching model with state variable z_t. The regime z_t is the outcome of an unobserved N-state Markov chain. Assuming $z_t \in \{0, \ldots, N-1\}$, we evaluate the transition probabilities $p_{l,k}^{(t)} \propto p(z_t = k|z_{t-1} = l, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_t; \boldsymbol{\theta})$ by using the Sichel distribution given in Proposition 4.3.1, $t \in \{1, \ldots, T\}$. We collect these probabilities in an $(N \times N)$-dimensional matrix $P^{(t)}$ whose (l,k)-th entry is equal to $p_{l,k}^{(t)}$. Note that the rows of the matrix have to sum up to one, thus for $N < \infty$ the transition given by the Sichel kernel needs to be normalized to get the $p_{l,k}^{(t)}$. Then, we apply the Forward-Filtering Backward-Sampling (FFBS) procedure. To do this, we denote with: (i) $\boldsymbol{\eta}_t$ the N-dimensional vector of *conditional likelihood* whose l-th element is given by $p(r_t, \mathrm{RV}_t|z_t = l, \boldsymbol{x}_t; \boldsymbol{\theta})$ (see Proposition 4.3.1), and with (ii) $\boldsymbol{\xi}_{t|\tau}$ the N-dimensional vector whose l-th element contains the filtered ($\tau = t$), predicted ($\tau = t-1$) and smoothed ($\tau > t$) distributions of the latent process z_t, $p\left(z_t = l|\boldsymbol{r}_{1:\tau}, \mathrm{RV}_{1:\tau}, \boldsymbol{x}_{1:\tau}; \boldsymbol{\theta}\right)$, $l \in \{0, \ldots, N-1\}$ and $t \in \{1, \ldots, T\}$. The FFBS procedure is based on $P^{(t)}$ and $\boldsymbol{\eta}_t$. More precisely, in order to sample a realization of $\boldsymbol{z}_{p+1:T}$ from $p(\boldsymbol{z}_{p+1:T}|\boldsymbol{r}_{p+1:T}, \mathrm{RV}_{p+1:T}, \boldsymbol{x}_{p+1:T}; \boldsymbol{\theta})$, we first apply the Hamilton filter algorithm forward in time, t from $p+1$ to T, to find $p\left(z_t = l|\boldsymbol{r}_{1:t}, \mathrm{RV}_{1:t}, \boldsymbol{x}_{1:t}, \boldsymbol{\theta}\right)$, then the Kim algorithm backward in time, t from T to $p+1$, to find $p\left(z_t = l|\boldsymbol{r}_{1:T}, \mathrm{RV}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{\theta}\right)$ (we refer to Hamilton, 1994; Frühwirth-Schnatter, 2006, and references therein for an exhaustive description of these two algorithms). A draw of z_t is obtained by multinomial sampling with multinomial probabilities $\boldsymbol{\xi}_{t|T}$, $t \in \{1, \ldots, T\}$. Once that the trajectory for the state variable z_t has been generated, we sample a realization of $\mathbf{CV}_{p+1:T}$ from the $\mathcal{G}ig$ distribution as in Proposition4.3.2, $t \in \{p+1, \ldots, T\}$.

In the second step we design a multi-move proposal distribution assuming an auxiliary SV-ARG. The multi-move proposal works as follows. At the $j-th$ iteration of the Gibbs sampler, we apply the FFBS to generate $\boldsymbol{z}_{\tau:\tau+\delta}^{(*)}$ and $\mathbf{CV}_{\tau:\tau+\delta}^{(*)}$, $\tau \in \{1, \ldots, T\}$ and $\delta \in \mathbb{N}_+$ such that $\tau+\delta \leq T$. Then, this proposal is accepted or rejected according to the acceptance log-probability $\rho\left((\boldsymbol{z}_{\tau:\tau+\delta}^{(j-1)}, \mathbf{CV}_{\tau:\tau+\delta}^{(j-1)}), (\boldsymbol{z}_{\tau:\tau+\delta}^{(*)}, \mathbf{CV}_{\tau:\tau+\delta}^{(*)})\right)$ (see Appendix A.6, Section A.6 for further details) of the Metropolis-Hastings algorithm. The multi-move proposal allows for a rapid mixing of the MCMC chain. To improve further the mixing, we use a random block updating. See Takahashi et al. (2009); Fiorentini et al. (2014); Billio et al. (2016).

Before moving to our financial application, we perform an intensive simulation exercise to study both the efficiency and effectiveness of the MCMC algorithm.

4.5 Simulation study

Several methodologies have been proposed in the literature to test the efficiency and the effectiveness of a MCMC algorithm. In this thesis we use three different measures to evaluate the mixing of the MCMC chain: (i) The acceptance rate (ACC) of the Metropolis-Hastings steps, (ii) the inefficiency factor (INEFF) of the estimation of the posterior mean, (iii) the convergence diagnostic statistics (CD).
The INEFF is defined as

$$\text{INEFF} \doteq 1 + 2\sum_{k=1}^{\infty} \rho(k)$$

where $\rho(k)$ is the autocorrelation at lag k for a set of samples of the parameter of interest. Intuitively, this measure captures the quantity of information that we actually have about a parameter. Indeed, if there are some correlation between successive samples, then our sample has not revealed as much information of the posterior distribution of our parameter as we could have gotten if the samples were independent. We point out that in Geweke et al. (1991) the inverse of the inefficiency factor INEFF is used.

The ACC is defined in Appendix A.4. The CD (Geweke et al., 1991) compares the location of the sampled parameter on two different time interval of the chain. If the mean values of the parameter within these two time intervals are close to each other, then one can assume that the two samples come from the same distribution. We clarify that, using the same notation of Geweke et al. (1991), we set $p = 10,000$, $p_{\text{A}} = 0.1\,p$ and $p_{\text{B}} = 0.5\,p$.

Besides, we measure the efficiency of the FFBS procedure through the Normalized Root Mean Square Error (NRMSE) averaged over the iterations of the Gibbs sampler. Specifically, indicating with M the effective number of the MCMC draws and with T the length of the sample, the NRMSE is defined as

$$\text{NRMSE} \doteq \frac{1}{\max(\widehat{\text{CV}}_{1:T}^{(i)}) - \min(\widehat{\text{CV}}_{1:T}^{(i)})} \frac{1}{\text{M}} \sum_{i=1}^{M} \sqrt{\frac{1}{T}\sum_{t=1}^{\text{M}} \left(\widehat{\text{CV}}_t^{(i)} - \text{CV}_t\right)^2}. \tag{4.15}$$

In previous equation, we denote with $\widehat{\text{CV}}_t^{(i)}$ the estimate of the latent variable CV_t on the i-th iteration of the Gibbs sampler and with $\max(\widehat{\text{CV}}_{1:T}^{(i)})$ (resp. $\min(\widehat{\text{CV}}_{1:T}^{(i)})$) the maximum (resp. the minimum) value of $\widehat{\text{CV}}_t$ on the i-th iteration of the Gibbs sampler.

Some practical considerations for the implementation of the inference algorithm arise. Specifically, the initialization of the Hamilton filter algorithm and the choice of the number of states for the approximating Markov Chain of the process z_t. To initialize the Hamilton filter algorithm, we set $\hat{\boldsymbol{\xi}}_{0|0} = \boldsymbol{\rho}$, where $\boldsymbol{\rho} = N^{-1}\boldsymbol{\iota}$ ($\boldsymbol{\iota}$ being the unit vector of dimension N). The choice of the number of states, instead, is a model-specific issue. We discuss this aspect when presenting empirical results.

4.5.1 Experiment design

In our simulation design, data are generated from the following four models:

(i) SV-ARG: This is the model in Equation (4.13).

(ii) SV-LARG : This is the model in Equation (4.14) with $\beta^{(w)} = \beta^{(m)} = 0$ and $\alpha^{(w)} = \alpha^{(m)} = 0$.

(iii) SV-HARG: This is the model in Equation (4.14) with $\alpha^{(d)} = \alpha^{(w)} = \alpha^{(m)} = 0$.

(iv) SV-LHARG : This is the model in Equation (4.14).

For each model we simulate 50 data-series of 1,000 observations. For each data-series we run the Gibbs sampler for 100,000 iterations, discard the first 20,000 draws to avoid dependence from initial conditions, and finally apply a thinning procedure to reduce the dependence between consecutive draws. We devote particular attention to the SV-ARG because it is used as proposal for the subsequent models. After setting $c = 1$, we follow, for instance, Chib et al. (2002); Casarin et al. (2009) and test the efficiency of the algorithm in three different scenarios: Low-Persistence ($\beta^{(d)} = 0.3$), MEDIUM PERSISTENCE ($\beta^{(d)} = 0.6$), and finally, HIGH PERSISTENCE ($\beta^{(d)} = 0.9$). The true values for the other parameters used in the simulations are reported in Table 4.1 and 4.2 for the SV-ARG and other models respectively. For all the models, at each fixed time t, we set $x_{1t} = 0$ and $x_{2t} = 0.2\mathbf{1}_{\{t<0.5T\}} + 0.01\epsilon_t$ with $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$.

4.5.2 Results

Tables 4.1 and 4.3 contain the grand average of the parameter posterior means for the SV-ARG model along with their standard deviations and efficiency indicators, i.e. INEFF, CD and ACC. In particular, Table 4.1 indicates the accuracy of the MCMC scheme is remarkable for all the scenarios (LOW PERSISTENCE, MEDIUM PERSISTENCE, HIGH PERSISTENCE). As regards the efficiency, the magnitudes of INEFF in Table 4.3 – after applying a thinning procedure – are below ten. Precisely, we use a thinning of 20 for the LOW and MEDIUM PERSISTENCE scenario and of 50 for the HIGH PERSISTENCE one. The high efficiency is due to the analytical filtering and smoothing recursive formula for the latent variables (see Proposition 4.3.2). The p-values of the CD statistics indicate that the null hypothesis that the sub-samples associated with the Markov chain have the same distribution is accepted. Actually, some of these values are not too high. However, Table 4.3 displays only the average p-values over the last 10,000 iterations using the same values for p_{A} and p_{B} across all the parameters. Finally, the average ACCs are in line with the ideal values of $[0.25, 0.50]$ suggested in Roberts et al. (1997), thanks to the careful tuning of the proposal density in the Metropolis-Hastings steps to the target densities.

Table 4.2 confirms that the accuracy of the estimate of the parameters is remarkable also for the SV-LARG, SV-HARG and SV-LHARG model. The INEFFs factors in Table A.4 – after applying a thinning procedure with a factor of 50 – are slightly higher respect to those of the SV-ARG model, although fully satisfactory. This loss of efficiency is due to the use of a Metropolis-Hastings step for the latent variables $\boldsymbol{z}_{1:T}$ and $\mathbf{CV}_{1:T}$. Nevertheless, only the inefficiency measure for the $\beta^{(m)}$ and $\alpha^{(m)}$ parameters in the SV-LHARG are above 20 and both the p-values of the CD statistics and the ACCs indicate convergence of the MCMC chains.

Our extensive experimentation shows that the careful implementation of the Metropolis-Hastings step for the latent variables is crucial for both the magnitude of the inefficiency factor and the number of sweeps of the Gibbs sampler necessary to obtain representative samples from the posterior distributions of interest (see also Chib et al., 2002). As regards the efficiency of the FFBS procedure, the grand average of the NRMSEs as in Equation (4.15) is given by 0.1147, 0.0941 and 0.0911 for the LOW, MEDIUM and HIGH PERSISTENCE

scenario of the SV-ARG. It is equal to 0.0758, 0.1253 and 0.0782 for the SV-LARG, SV-HARG and SV-LHARG model respectively[6].

[6] A more extensive description of the diagnostic before and after applying thinning to the MCMC samples is available in the Appendix A.8.

Table 4.1: Summary output of the parameter estimates for the SV-ARG model

	Low Persistence		Medium Persistence		High Persistence	
$\boldsymbol{\theta}$	True	Estimate	Std	Estimate	Std	Estimate	Std
μ	0.0	0.0018	0.0118	−0.0051	0.0177	−0.0074	0.0358
γ	1.0	1.0552	0.0738	1.0523	0.0720	1.0685	0.0784
κ_1	0.0	−0.0221	0.0906	0.0081	0.1157	−0.0062	0.2610
α	0.8	0.8428	0.0572	0.8327	0.0575	0.8474	0.0647
κ_2	−1.0	−0.9538	0.1224	−0.9652	0.1119	−0.9540	0.1005
ν	0.8	0.8033	0.0371	0.7981	0.0394	0.8182	0.0576
c	1.0	0.9654	0.0938	0.9706	0.0909	0.9395	0.0790
$\beta^{(d)}$		0.3118	0.0595	0.6376	0.0746	0.9702	0.0839

Notes: Reported are the parameters estimation results for three different scenarios of the SV-ARG model. The results are averages over a set of 50 independent MCMC experiments on 50 independent datasets of 1000 observations. On each dataset we ran the proposed MCMC algorithm for 100000 iterations and then discard the first 20000 iterations. A thinning procedure with a factor of 20 (low and medium persistence scenario) and 50 (high persistence scenario) is applied. We report the true value of parameters (second column) and in each setting (Low Persistence, Medium Persistence, and High Persistence) its mean (left column) and standard deviation (right column). The value of the parameter $\beta^{(d)}$ is 0.3, 0.6 and 0.90 in the low, medium and high persistence scenario respectively.

Table 4.2: Summary output of the parameter estimates for SV-LARG, SV-HARG, SV-LHARG models

		SV-LARG		SV-HARG		SV-LHARG	
$\boldsymbol{\theta}$	True	Estimate	Std	Estimate	Std	Estimate	Std
μ	0.0	−0.0240	0.04913	0.3038	0.13521	−0.5057	0.3100
γ	1.0	1.038	0.03738	0.8251	0.05820	1.0034	0.01492
κ_1	0.0	0.0080	0.03506	−0.0853	0.06324	0.1259	0.12642
α	0.8	0.8650	0.03112	0.6929	0.03996	0.8773	0.03155
κ_2	−1.0	−0.9358	0.06742	−1.2126	0.09249	0.1199	0.05230
ν	1.5	1.3254	0.07160	1.5383	0.1574	1.6004	0.26841
c	1.0	1.0197	0.04495	1.0763	0.07340	0.9842	0.01553
$\beta^{(d)}$	0.30	0.3108	0.06183	0.3169	0.03998	0.2918	0.0227
$\beta^{(w)}$	0.20	−−	−−	0.2245	0.03951	0.2295	0.03177
$\beta^{(m)}$	0.10	−−	−−	0.0601	0.03648	0.08389	0.03516
$\alpha^{(d)}$	0.15	0.1119	0.03776	−−	−−	0.1688	0.01467
$\alpha^{(w)}$	0.10	−−	−−	−−	−−	0.0883	0.02497
$\alpha^{(m)}$	0.05	−−	−−	−−	−−	0.0449	0.02579
λ	1.0	1.1608	0.47378	−−	−−	1.0007	0.01850

Notes: Reported are the parameters estimation results for the SV-LARG, SV-HARG and SV-LHARG model. The results are averages over a set of 50 independent MCMC experiments on 50 independent data set of 1000 observations. On each dataset we ran the proposed MCMC algorithm for 100,000 iterations and then discard the first 20,000. A thinning procedure with a factor of 50 is applied. We report the true value of the parameter (*second column*) and for each model its mean (*left column*) and standard deviation (*right column*).

Table 4.3: Summary output of the efficiency indicators of parameter estimates for the SV-ARG model

	PANEL A Low Persistence			PANEL B Medium Persistence			PANEL C High Persistence		
$\boldsymbol{\theta}$	INEFF	(CD, p)	ACC	INEFF	(CD, p)	ACC	INEFF	(CD, p)	ACC
μ	1.16	$(0.38, 0.36)$	$--$	1.20	$(-0.98, 0.16)$	$--$	1.28	$(-0.48, 0.31)$	$--$
γ	6.03	$(-1.18, 0.12)$	$--$	6.89	$(1.08, 0.14)$	$--$	7.30	$(1.59, 0.06)$	$--$
κ_1	1.13	$(-0.77, 0.22)$	$--$	1.12	$(1.74, 0.04)$	$--$	1.24	$(0.02, 0.49)$	$--$
α	7.86	$(-1.34, 0.09)$	0.20	7.88	$(1.28, 0.10)$	0.22	1.19	$(1.58, 0.06)$	0.20
κ_2	6.86	$(-0.56, 0.29)$	0.25	6.61	$(1.24, 0.11)$	0.32	6.22	$(0.80, 0.21)$	0.17
ν	1.68	$(0.98, 0.16)$	0.20	1.39	$(1.74, 0.04)$	0.23	1.29	$(1.36, 0.08)$	0.22
c	4.97	$(-1.11, 0.13)$	$--$	6.13	$(-0.58, 0.28)$	$--$	7.03	$(-1.99, 0.03)$	$--$
$\beta^{(d)}$	2.87	$(1.97, 0.04)$	$--$	5.07	$(0.50, 0.30)$	$--$	6.78	$(1.40, 0.08)$	$--$

Notes: Reported are efficiency indicators of the parameters estimation results for three different scenarios of the SV-ARG model. The results are averages over a set of 50 independent MCMC experiments on 50 independent datasets of 1,000 observations. On each dataset we ran the proposed MCMC algorithm for 100,000 iterations and then discard the first 20,000 iterations. A thinning procedure with a factor of 20 (low and medium persistence scenario) and 50 (high persistence scenario) is applied. In each PANEL: inefficiency factor (INEFF); convergence diagnostic statistic ((CD,p)); average acceptance rate of the Metropolis-Hastings algorithm steps (ACC). True values of the parameters are reported in Table 4.1

Table 4.4: Summary output of the efficiency indicators of parameter estimates for the SV-LARG, SV-HARG, SV-LHARG models.

	PANEL A SV-LARG			PANEL B SV-HARG			PANEL C SV-LHARG		
$\boldsymbol{\theta}$	INEFF	(CD, p)	ACC	INEFF	(CD, p)	ACC	INEFF	(CD, p)	ACC
μ	5.71	$(1.55, 0.06)$	$--$	5.77	$(0.35, 0.36)$	$--$	1.12	$(-0.65, 0.26)$	$--$
γ	13.73	$(-1.22, 0.11)$	$--$	16.34	$(0.72, 0.23)$	$--$	1.11	$(0.42, 0.34)$	$--$
κ_1	9.52	$(-1.16, 0.12)$	$--$	1.67	$(-1.17, 0.12)$	$--$	1.06	$(0.19, 0.42)$	$--$
α	16.85	$(-1.44, 0.07)$	0.21	17.05	$(0.34, 0.37)$	0.28	1.23	$(-1.10, 0.14)$	0.26
κ_2	8.58	$(1.25, 0.05)$	0.16	12.89	$(-0.01, 0.50)$	0.26	1.11	$(-0.82, 0.21)$	0.23
ν	7.84	$(-1.16, 0.12)$	0.22	13.10	$(-1.04, 0.15)$	0.21	1.29	$(-0.77, 0.22)$	0.27
c	13.72	$(1.33, 0.09)$	$--$	17.44	$(-0.88, 0.19)$	$--$	1.17	$(0.50, 0.31)$	$--$
$\beta^{(d)}$	8.44	$(1.11, 0.13)$	0.19	12.51	$(0.31, 0.37)$	0.24	2.67	$(0.99, 0.16)$	0.17
$\beta^{(w)}$	$--$	$--$	$--$	11.90	$(1.48, 0.07)$	0.24	7.22	$(0.95, 0.17)$	0.17
$\beta^{(m)}$	$--$	$--$	$--$	18.19	$(-1.17, 0.12)$	0.24	22.62	$(1.49, 0.06)$	0.17
$\alpha^{(d)}$	11.9	$(1.56, 0.06)$	0.30	$--$	$--$	$--$	1.79	$(-0.65, 0.26)$	0.32
$\alpha^{(w)}$	$--$	$--$	$--$	$--$	$--$	$--$	10.87	$(0.26, 0.40)$	0.32
$\alpha^{(m)}$	$--$	$--$	$--$	$--$	$--$	$--$	22.57	$(0.94, 0.17)$	0.32
λ	10.12	$(-1.65, 0.05)$	0.27	$--$	$--$	$--$	1.02	$(0.16, 0.43)$	0.29

NOTES: Reported are efficiency indicators of the parameters estimation results for the SV-LARG (PANEL A), SV-HARG (PANEL B) and SV-LHARG (PANEL C) on a dataset of 1000 observations. For each dataset we ran the proposed MCMC algorithm for 100,000 iterations and then discard the first 20,000 iterations. A thinning procedure with a factor of 50 is applied. In each PANEL: inefficiency factor (INEFF); convergence diagnostic statistic ((CD,p)); average acceptance rate of the Metropolis-Hastings algorithm steps (ACC). True values of the parameters are reported in Table 4.2

4.6 Financial application

4.6.1 Data description

In this section we present empirical results using daily log-returns and realized variance for the S&P500 Futures[7]. We employ the RV computed from tick-by-tick data, from 8-January-1997 to 8-January-2007. In particular, the RV is constructed in the following way. Although we acknowledge the importance of the jump contribution in log-returns and RV, this component is not included in the class of models considered in this thesis. In order to exclude the effect of jump on log-return and volatility process from the empirical analysis we employ the same methodology adopted by Corsi et al. (2013); Majewski et al. (2015). Precisely: (i) we estimate the total variation of the log-price process using the Two-Scales estimator proposed by Zhang et al. (2005) (with a fast scale of two ticks and a slower one of 20 ticks). This proxy includes jumps in both returns and volatility; (ii) we purify it from the jump component in prices by means of the Threshold Bi-power variation method introduced in Corsi et al. (2010) with a significance level of 99%; (iii) we remove the most extreme observations, seemingly due to volatility jumps, in the volatility series employing a threshold-based jumps detection method.

4.6.2 Estimation of the models under physical probability measure

We pick out the FED Fund rate as proxy for the risk-free rate μ in all considered models and we estimate them following the procedure specified in Section 4.4.

Table 4.5 contains posterior quantities for the physical processes developed above. As in Majewski et al. (2015), both log-returns and volatilities are on a daily and decimal basis. The posterior quantities are computed from 10,000 draws of the MCMC algorithm, collected after an initial burn-in period of 5,000 iterations and after applying a thinning procedure with a factor of 10. We now analyse the results.

The ν parameters (*fourth row*) do not have standard errors as they are computed by variance targeting thus exactly matching the observed sample variance of returns[8] (see, for instance Christoffersen et al., 2015). For all the four models, the risk premium parameter γ is estimated to be negative and significant. This fact implies that the distribution of returns are negatively skewed and is in line with the finding of Creal (2015). The estimates of the overnight factor α (*second row*) suggest that volatility during the open period amounts to about 65% of daily volatility. We comment the estimates of the complete model SV-LHARG. The impact of past lags on the present value of the CV is given by the the partial autocorrelation coefficients $c(\beta^{(d)} + \beta^{(w)} + \beta^{(m)} + (\alpha^{(d)} + \alpha^{(w)} + \alpha^{(m)})\lambda^2)$. According to our estimates (considering also the leverage effect), the sensitivity of the conditional mean of CV_t on CV_{t-1} is $c(\beta^{(d)} + \lambda^2\alpha^{(d)}) = 0.52$, whereas the sensitivity on $\mathrm{CV}^{(w)}_{t-1}$ and $\mathrm{CV}^{(m)}_{t-1}$ is $c(\beta^{(w)} + \lambda^2\alpha^{(w)}) = 0.35$ and $c(\beta^{(m)} + \lambda^2\alpha^{(m)}) = 0.06$, respectively. All parameters are statistically significant except the monthly leverage component $\alpha^{(m)}$. The CV coefficients show a decreasing impact of past lags on the present value of the conditional variance (see also the estimates of the SV-HARG model). This fact has been already documented in the realized volatility counterpart version of our model (see Corsi et al., 2013; Majewski et al., 2015). Actually, observing the estimates for the SV-HARG we note that in this model the conditional mean of CV_t is mainly influenced by CV_{t-1}. We attribute this fact

[7] Future contracts are chosen because of their high liquidity.

[8] Actually, assuming $\gamma = 0$ in the inference algorithm we match the sample variance of the returns to the invariant unconditional mean of the process CV_t (see Equation4.9).

to the smoothing procedure of the latent variable CV_t. This phenomena is not so evident when introducing the leverage component. In this case also the idiosyncratic component ϵ_t need to be filtered and so, some noise is introduced. Finally, the last row reports for each model the value of the risk-premium parameter ν_1 fitted on option prices (see next Section). Finally, the persistence of the CV process under the physical measure $\mathbb{P}$, defined as $c(\beta^{(d)} + \beta^{(w)} + \beta^{(m)} + (\alpha^{(d)} + \alpha^{(w)} + \alpha^{(m)})\lambda^2)$, is equal to: (i) 0.9596 for the SV-ARG, (ii) 0.8665 for the SV-LARG, (iii) 0.9666 for the SV-HARG, and (iv) 0.8901 for the SV-LHARG. In particular, the inclusion of the leverage component into SV-ARG and SV-HARG has the effect of reducing the persistence of the volatility process. Evidence of the persistence reduction can be found from comparing the CV smoothed estimates (*red line*) and the RV rescaled to the whole sample estimate of the overnight factor (*black line*) in Figure 4.1. This fact is also documented in Corsi et al. (2013) and Majewski et al. (2015). In Appendix A.7 further computational details and investigations are reported.

4.6.3 Option pricing: performance assessment

We perform our analysis on European out-of-the-money (OTM) options on S&P500 index for each Wednesday from from February 12, 1997 to December 29, 2004 (data are provided by OptionMetrics). We first apply a standard filter removing options with maturity less than 10 days or more than 365 days, implied volatility larger than 70% and prices less than 5 cents (see Barone-Adesi et al., 2008; Corsi et al., 2013; Majewski et al., 2015). Using K/S_t as definition of moneyness, we filter out deep OTM options with moneyness larger than 1.2 for call options and less than 0.8 for put options. This choice yields a total number of 36,377 observations. For our purposes, put options are identified as Deep OTM if their moneyness is between $0.8 \leq m \leq 0.9$ and OTM if $0.9 < m \leq 0.98$. On the other hand, call options are said to be DOTM if $1.1 < m \leq 1.2$ and OTM if $1.02 < m \leq 1.1$. Options are called at-the-money (ATM) if $0.98 < m \leq 1.02$. As far as the time to maturity τ is concerned, we identify options as short maturity ($\tau \leq 50$ days), short-medium maturity ($50 < \tau \leq 90$ days), long-medium maturity ($90 < \tau \leq 160$ days), and long maturity ($\tau > 160$ days). In order to derive the risk-neutral dynamics, the value for risk premia parameters (ν_1, ν_2) vector needs to be identified. According to Proposition 4.2.4, in our framework ν_2 is fixed by the no-arbitrage condition, while ν_1 remains a free parameter that needs to be calibrated from option prices. In particular, we follow the same reasoning of Corsi et al. (2013); Majewski et al. (2015) performing the unconditional calibration of ν_1 such that the model generated and the average market IV for a one-year time to maturity at-the-money option coincide. In order to compute the option prices – and associated implied volatilities – we employ the option pricing numerical method termed COS, introduced by Fang and Oosterlee (2008), and which has been proven to be efficient. The method is based on Fourier-cosine expansions and is available as long as the characteristic function of log-returns is known.

To sum up, we proceed pricing options following four steps: (i) estimation under the physical measure $\mathbb{P}$, (ii) unconditional calibration of the parameter ν_1 (iii) mapping of the parameters of the model estimated under $\mathbb{P}$ into the parameters under $\mathbb{Q}$, and (iv) approximation of option prices by the COS method using the MGF of the model with parameters under the risk-neutral measure.

Thanks to Proposition 4.2.6, the persistence of the CV process under the risk-neutral measure $\mathbb{Q}$, defined as $c^* \left(\beta^{(d,*)} + \beta^{(w,*)} + \beta^{(m,*)} + \left(\alpha^{(d,*)} + \alpha^{(w,*)} + \alpha^{(m,*)}\right)\lambda^{*2}\right)$ is equal to: (i)0.9682 for the SV-ARG, (ii)0.8927 for the SV-LARG, (iii) 0.9756 for the SV-HARG and (iv) 0.9107 for the SV-LHARG.

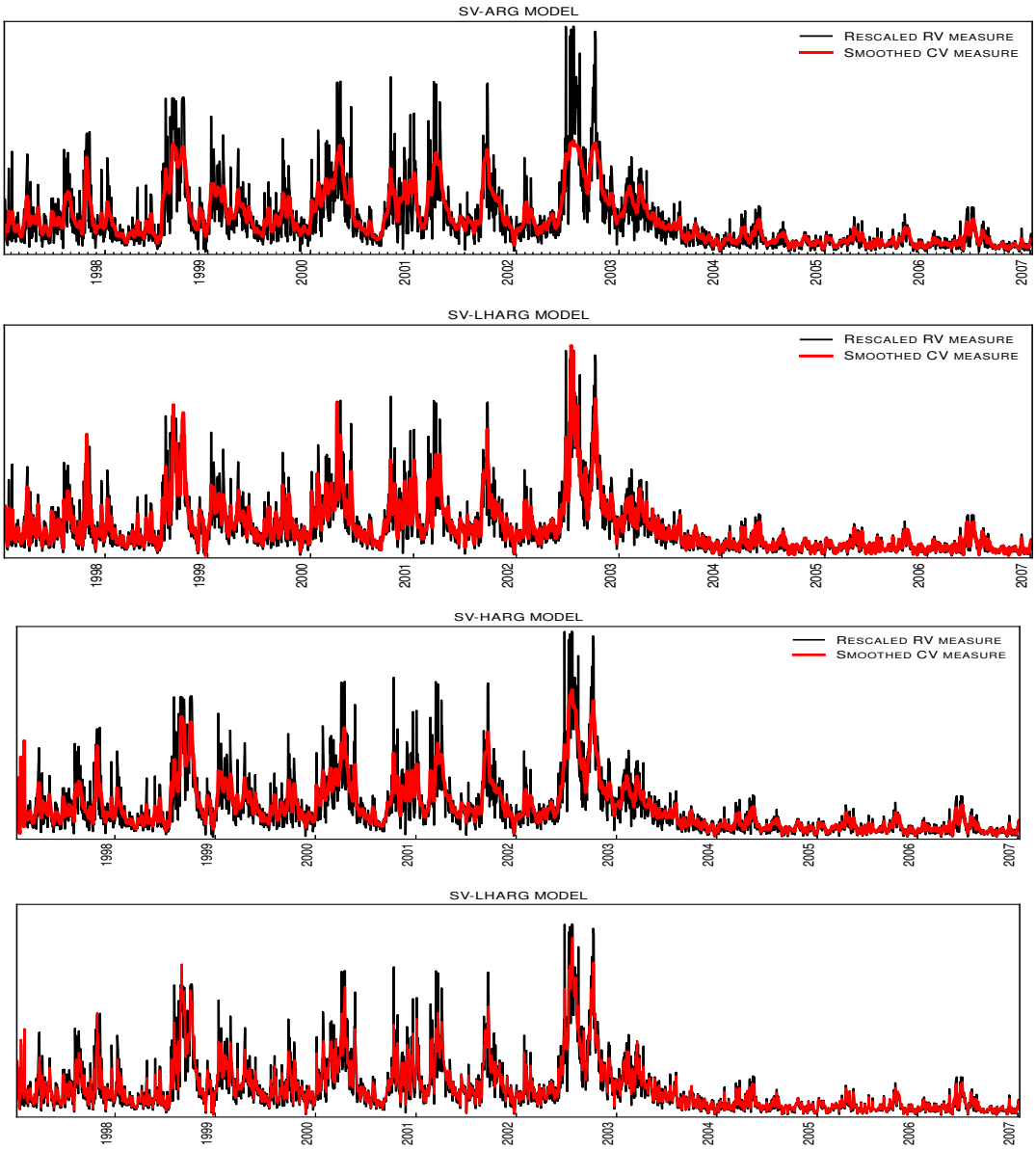
Table 4.5: Summary output of the parameter estimates of SV-ARG, SV-LARG, SV-HARG, SV-LHARG models, on S&P 500 Futures, 1997-2007

PANEL A

	SV-ARG		SV-LARG		SV-HARG		SV-LHARG	
PARAMETERS	ESTIMATE	STD	ESTIMATE	STD	ESTIMATE	STD	ESTIMATE	STD
γ	-0.129	$3.2e-02$	-0.117	$2.86e-02$	-0.130	$3.0e-02$	-0.086	$3.2e-02$
α	0.6492	$2.7e-02$	0.6422	$3.47e-02$	0.6501	$4.44e-02$	0.6305	$3.93e-02$
κ_2	-2.561	$3.5e-02$	-2.962	$4.6e-02$	-2.764	$3.75e-02$	-3.199	$9.12e-02$
ν	3.2499	$--$	1.9201	$--$	1.7448	$--$	1.2175	$--$
c	$1.320e-06$	$2.5e-08$	$7.313e-06$	$7.3e-08$	$2.701e-06$	$4.5e-08$	$6.937e-06$	$5.2e-07$
$\beta^{(d)}$	$54.808e+04$	$14.29e+03$	$11.318e+04$	$1.535e+03$	$31.934e+04$	$1.026e+03$	$3.92e+04$	$1.2e+03$
$\beta^{(w)}$	$--$	$--$	$--$	$--$	$2.38e+04$	$5.9e+03$	$1.92e+04$	$7.6e+03$
$\beta^{(m)}$	$--$	$--$	$--$	$--$	$1.46e+04$	$2.9e+03$	$6.46e+03$	$4.4e+03$
$\alpha^{(d)}$	$--$	$--$	0.653	$7.8e-02$	$--$	$--$	0.517	$7.1e-02$
$\alpha^{(w)}$	$--$	$--$	$--$	$--$	$--$	$--$	0.452	$7.9e-02$
$\alpha^{(m)}$	$--$	$--$	$--$	$--$	$--$	$--$	0.0459	$9.1e-02$
λ	$--$	$--$	90.1	11.3	$--$	$--$	265.78	21.81
ν_1	$-2,539$		$-1,994$		$-1,517$		$-1,508$	

NOTES: Reported are the parameters estimation results for the SV-ARG, SV-LARG, SV-HARG and SV-LHARG on S&P 500 Futures, from 8-January-1997 to 8-January-2007. We ran the proposed MCMC algorithm for 15,000 iterations and then discard the first 5,000 iterations. A thinning procedure with a factor of 10 is applied. For each parameter we report the mean (ESTIMATE) and the posterior standard deviations (STD). The parameter ν_1 for each model has been calibrated on option prices.

Figure 4.1: Output of the smoothed conditional variance estimates for SV-ARG, SV-LARG, SV-HARG, SV-LHARG models, on S&P 500 Futures, 1997-2007



NOTES: Reported are the average of the smoothed estimates of the CV (*red line*), and the RV rescaled to the overnight factor estimates, $\hat{\alpha}$, (*black line*). The top panel refers to the SV-ARG model, the second panel refers to the SV-LARG model, while the third and the bottom panels refer to the SV-HARG and SV-LHARG, respectively. The parameter estimates are taken from Table A.2. The sample starts January 8, 1997 and ends January 8, 2007.

We analyse now the static properties of the option pricing results. In particular, as customary in the literature (see Renault, 1997; Corsi et al., 2013; Majewski et al., 2015), we employ the Root Mean Square Error on the percentage IV (RMSE_{IV}) as performance measure:

$$\text{RMSE}_{\text{IV}} = \sqrt{\sum_{i=1}^{\text{N}} \frac{(\text{IV}_i^{\text{MOD}} - \text{IV}_i^{\text{MKT}})^2}{\text{N}}}.$$

In previous equation, N is the number of options, and IV^{MOD} and IV^{MKT} are the model and the market implied volatility, respectively. In particular, in Table 4.6 we report the global option pricing performance on S&P500 out-of-the-money options from February 12, 1997 to December 31, 2004, computed with the CV measure estimated from 1997 to 2007. The first row shows the absolute RMSE_{IV} for the SV-ARG in the range of moneyness $0.8 < m < 1.2$ and $0.9 < m < 1.1$, while the remaining rows display the SV-ARG relative performances with respect to the other models in the same ranges of moneyness.

At first glance, the SV-LHARG model outperforms all the others. A closer inspection shows the following. (i) The SV-LARG improves of about 10% both the SV-ARG and SV-HARG in the range of moneyness $0.8 < m < 1.2$, while in the range $0.9 < m < 1.1$ the performance is comparable. This result confirms that the inclusion of the leverage component is essential for the pricing of OTM-DOTM options. (ii) The price performance of SV-ARG and SV-HARG is very similar. This is coherent with the observation done about the sensitivity of CV_{t-1} on the first moment of CV_t. (ii) The SV-LHARG outperforms all the other models by about 19% in the range of moneyness $0.9 < m < 1.1$ and $24\% - 30\%$ in the range $0.8 < m < 1.2$.

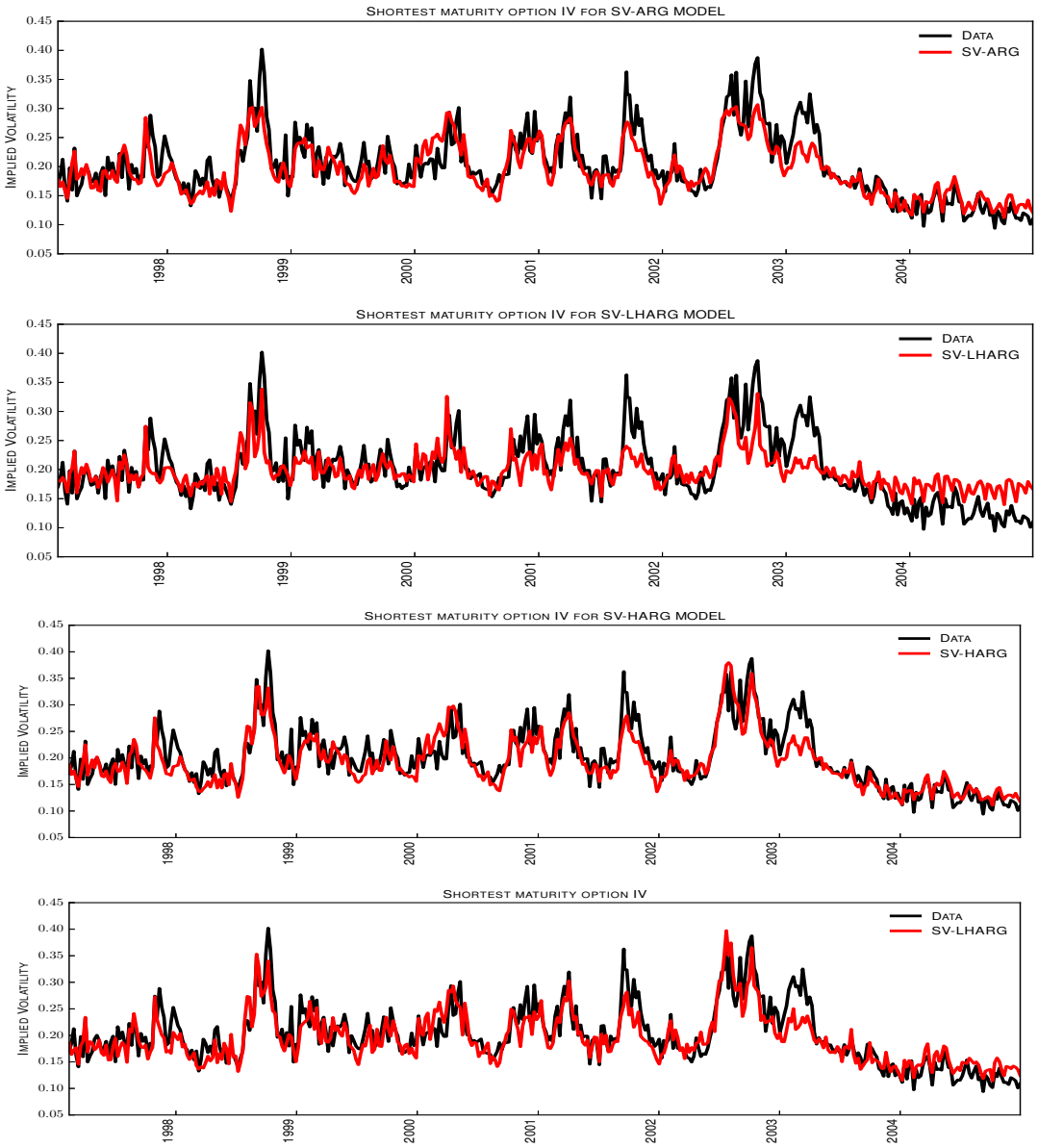
To gain a deeper understanding of the pricing performance, in Table 4.7 we report the results in terms of RMSE_{IV} disaggregated for different maturities, τ, and moneyness, m. Panel B-D confirm that the main advantage deriving from the inclusion of the leverage component is the ability to capture the volatility smile, especially at the put side and for small maturities. Panel D confirms the superiority of the SV-LHARG model with respect all the others. While the performance for the ATM options is comparable, for OTM and DOTM options the improvement is remarkable. On the put side the improvement varies from 11% to 28%, while on the call side the improvement is even more apparent (neglecting the DOTM call option at the shortest maturity), varying from 8% to 39%. This analysis suggests that the various ingredients (heterogeneity, leverage and persistence) are necessary to accurately price options across different maturities and moneyness. This is the reason why the SV-LHARG consistently shows the best option pricing performance among the considered models

In addition to the static analysis just carried out, we here investigate the ability of the different models to describe the dynamic evolution of the IV surface, focusing our attention on the dynamics of the short-end of the IV surface. Specifically, Figure 4.2 represents the level (i.e., the average IV of short-term ATM options) dynamic from February 12, 1997 to December 31, 2004 for the four models analysed. This graphical analysis confirms that the SV-LHARG produces the most adaptive dynamics to volatility changes among the four model considered.

We conclude this Section by comparing the option pricing performance of the SV-LHARG with two models present in the literature.

The first one is the P-LHARG model introduced in Majewski et al. (2015). It comes

Figure 4.2: Shortest maturity option IV



Notes: level dynamic from February 12, 1997 to December 31, 2004. level is the average implied volatility of at-the-money options (with moneyness $m = K/S_t$ between 0.95 and 1.05, where K and S are the strike and underlying price, respectively) and maturity at the shortest available on a given day. In each panel, the black line represents the data, the red line, the model. The top panel illustrates the performance of the SV-ARG, the second panel refers to the SV-LARG model, while the third and the bottom panels refer to the SV-HARG and SV-LHARG, respectively. The parameter estimates are taken from Table A.2.

Table 4.6: Global option pricing performance on S&P 500 out-of-the-money options from February 12, 1997 to December 29, 2004.

Model	Moneyness	
	$0.9 < m < 1.1$	$0.8 < m < 1.2$
SV-ARG	4.217	5.948
SV-LARG/SV-ARG	1.017	0.908
SV-HARG/SV-ARG	0.987	1.000
SV-LHARG/SV-ARG	0.817	0.755

Notes: Global option pricing performance on S&P 500 out-of-the-money options from February 1, 1997 to December 31, 2004, computed with the CV measure estimated from 1997 to 2007. We use the parameter estimates from Table A.2. First row: percentage Implied Volatility Root Mean Squared Error (RMSE_{IV}) of the SV-ARG model for different moneyness range. Second and subsequent rows: relative RMSE_{IV} of the other models with respect to SV-ARG model as benchmark.

from the class of the realized volatility models and it is described by the following relations

$$\begin{aligned} r_t &= \mu + \gamma \text{RV}_t + \sqrt{\text{RV}_t}\epsilon_t \\ \text{RV}_t|\mathcal{F}_{t-1} &\overset{d}{\sim} \bar{\mathcal{G}}\left(\nu, \varphi(\mathbf{RV}_{t-1}, \boldsymbol{l}_{t-1}), c\right). \end{aligned}$$

where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and $\mathcal{F}_t = \sigma(r_t, \text{RV}_t)$. In particular, the conditional distribution of r_t is taken as that of CV_t in the SV-LHARG. Thus, we refer to Sub-section 4.3.2 for an accurate description of it.

The second one, instead, comes from the class of GARCH-type option pricing models. We consider the two-component GARCH (in the following, CGARCH) introduced in Christoffersen et al. (2008) with a variance-dependent pricing kernel. The CGARCH model is given by:

$$\begin{aligned} r_t &= \mu + \gamma \text{CV}_t + \sqrt{\text{CV}_t}\epsilon_t \\ \text{CV}_t &= q_t + b_s\left(\text{CV}_{t-1} - q_{t-1}\right) + a_s\left(\left(\epsilon_{t-1} - c_s\sqrt{\text{CV}_{t-1}}\right)^2 - \left(1 + c_s^2 q_{t-1}\right)\right), \\ q_t &= \omega + \rho q_{t-1} + \varphi\left(\left(\epsilon_{t-1}^2 - 1\right) - 2c_l\sqrt{\text{CV}_{t-1}}\epsilon_{t-1}\right), \end{aligned}$$

where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and $(\text{CV}_{t-1} - q_{t-1})$ and q_t represent the short- and long-run persistent components, respectively. Both the P-LHARG and the CGARCH have a SDF comparable to the one used in our research.

Table 4.8, Panel A, reports the parameter estimates of both the P-LHARG and the CGARCH along with their standard error (in brackets). Panel B, instead, reports the relative RMSE_{IV} of SV-LHARG to its direct competitor model, the P-LHARG, disaggregated for the same maturities and moneyness considered in Table 4.7. While the performances of the SV-LHARG and P-LHARG in the ATM region are quite similar, SV-LHARG outperform P-LHARG by about $5-7\%$ in the OTM region at both the put and call side. The amelioration for short maturities and DOTM options reaches about 10%. For longer maturities and DOTM options the gain is even stronger: we obtain on average 15% smaller RMSE_{IV}. This improvement is due to the gain in persistence. Precisely, the persistence under the historical dynamics of the P-LHARG is 0.8383, whereas

Table 4.7: Option pricing performance on S&P500 out-of-the-money options from February 12, 1997 to December 29, 2004.

Moneyness	Maturity $\tau \leq 50$	$50 < \tau \leq 90$	$90 \leq \tau \leq 160$	$160 < \tau$
Panel A	SV-ARG CV RMSE			
$0.8 \leq m \leq 0.9$	13.429	9.001	7.677	6.281
$0.9 < m \leq 0.98$	6.156	4.999	4.520	4.389
$0.98 \leq m \leq 1.02$	2.798	3.019	3.131	3.622
$1.02 \leq m \leq 1.1$	3.370	3.171	3.046	3.478
$1.1 \leq m \leq 1.2$	4.249	3.839	3.684	3.687
Panel B	SV-LARG/SV-ARG CV RMSE			
$0.8 \leq m \leq 0.9$	0.788	0.890	0.939	0.975
$0.9 < m \leq 0.98$	0.846	1.008	1.023	1.041
$0.98 \leq m \leq 1.02$	1.258	1.274	1.189	1.111
$1.02 \leq m \leq 1.1$	1.082	1.261	1.244	1.157
$1.1 \leq m \leq 1.2$	0.858	0.902	0.972	1.030
Panel C	SV-HARG/SV-ARG CV RMSE			
$0.8 \leq m \leq 0.9$	1.001	1.022	1.008	0.988
$0.9 < m \leq 0.98$	1.021	1.033	1.016	0.967
$0.98 \leq m \leq 1.02$	0.979	0.984	0.942	0.904
$1.02 \leq m \leq 1.1$	0.935	0.885	0.841	0.858
$1.1 \leq m \leq 1.2$	1.032	1.036	1.000	0.928
Panel D	SV-LHARG/SV-ARG CV RMSE			
$0.8 \leq m \leq 0.9$	0.727	0.705	0.729	0.793
$0.9 < m \leq 0.98$	0.723	0.795	0.832	0.898
$0.98 \leq m \leq 1.02$	0.957	0.965	0.969	0.975
$1.02 \leq m \leq 1.1$	0.828	0.804	0.875	0.921
$1.1 \leq m \leq 1.2$	1.174	0.643	0.618	0.749

Notes: Reported is the option pricing performance on S&P500 out-of-the-money options from February 12, 1997 to December 29, 2004, computed with the smoothed CV estimated from January 8, 1997 to January 8, 2007. Panel A: percentage Implied Volatility Root Mean Square Error (RMSE_{IV}) of the SV-ARG model sorted by moneyness and maturities. Panels B to D: relative RMSE_{IV} sorted by moneyness and maturity. The parameter estimates are taken from Table A.2.

Table 4.8: Model comparison: option pricing performance on S&P500 out-of-the-money options from February 12, 1997 to December 29, 2004

Panel A

Parameter Estimates

P-LHARG			CGARCH		
Parameter	Estimates		Parameter	Estimates	
$\gamma^{(*)}$	0.9	(1.7)	γ	2.9	(1.6)
ν	1.2157		b_s	0.670	(6.1e-02)
c	1.2863e-05	(7.3e-08)	a_s	1.49e-06	(6.6e-07)
$\beta^{(d)}$	1.81e+04	(1.9e+03)	c_s	4.2e+02	(1.7e+02)
$\beta^{(w)}$	1.49e+04	(2.5e+03)	ω	1.27e-06	(1.9e-07)
$\beta^{(m)}$	9.73e+03	(2.7e+03)	b_l	0.9861	(2.1e-03)
$\alpha^{(d)}$	0.262	(1.3e-02)	a_l	2.45e-06	(2.8e-07)
$\alpha^{(w)}$	0.193	(5.4e-03)	c_l	88	(15)
$\alpha^{(m)}$	0.0417	(3.9e-02)			
λ	214	(24)			
ν_1	-1,175		ν_1	-77,418	

Option Pricing Performance

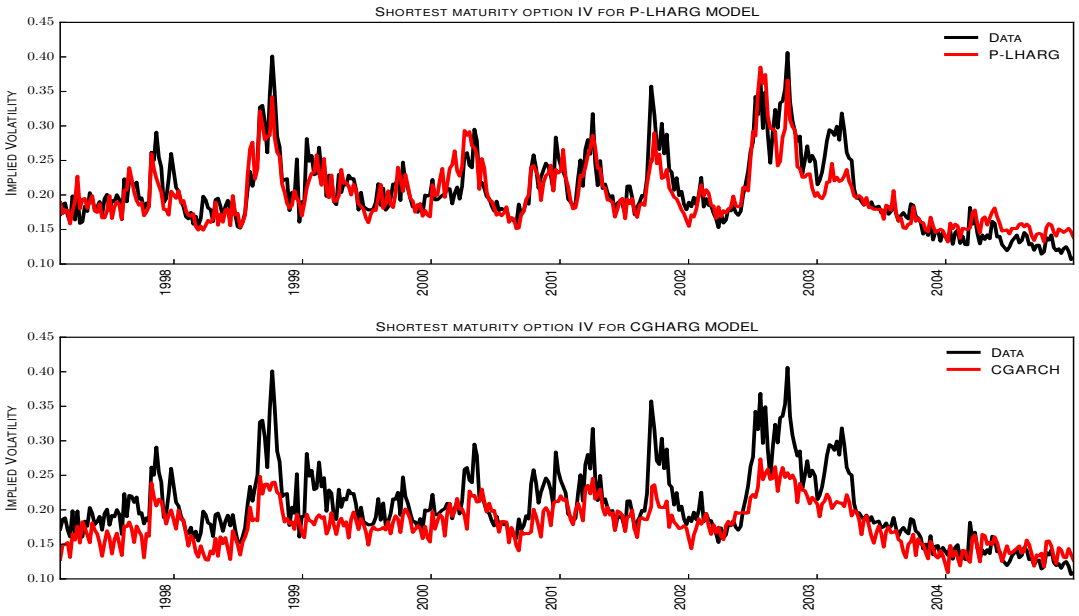
Moneyness	Maturity			
	$\tau \leq 50$	$50 < \tau \leq 90$	$90 \leq \tau \leq 160$	$160 < \tau$
Panel B	SV-LHARG/P-LHARG RMSE			
$0.8 \leq m \leq 0.9$	0.913	0.857	0.865	0.897
$0.9 < m \leq 0.98$	0.937	0.928	0.936	0.956
$0.98 \leq m \leq 1.02$	0.993	1.004	1.005	0.992
$1.02 \leq m \leq 1.1$	0.962	0.918	0.947	0.969
$1.1 \leq m \leq 1.2$	1.166	0.900	0.828	0.893

Notes: Panel A: Maximum likelihood estimates with the relative standard errors of the P-LHARG and CGARCH. The historical data for the P-LHARG model are given by the daily RV computed on tick-by-tick data for the S&P500 Futures. The estimation period ranges from 8-January-1997 to 8-January-2007. The parameter ν_1 is calibrated on option prices.
$(*)$ The estimation of the market price of risk γ is performed regressing the centred and normalized log-returns on the RV (see Equation (18) in Corsi et al., 2013). Panel B: relative RMSE_{IV} sorted by moneyness and maturity. The parameter estimates are taken from Panel a.

that under the risk-neutral one is 0.8742. For the SV-LHARG it is equal to 0.8901 and 0.9107 under $\mathbb{P}$ and $\mathbb{Q}$ respectively.

Finally, Figure 4.3 depicts the Level dynamics from February 12, 1997 to December 31, 2004 for the P-LHARG (*top panel*) and the CGARCH (*bottom panel*) model. The evolution of the IV level for P-LHARG is quite similar to that of the SV-LHARG (see 4.2), albeit the former seems to be less reactive especially for low IV. This is because the (rescaled) RV is a noisier version of the CV, and part of the noise is transmitted to option prices. The CGARCH, instead, tends to reproduce the empirical level dynamics with some delay.

Figure 4.3: Shortest maturity option IV, P-LHARG and CGARCH model



Notes: level dynamic from February 12, 1997 to December 31, 2004. level is the average implied volatility of at-the-money options (with moneyness $m = K/S_t$ between 0.95 and 1.05, where K and S are the strike and underlying price, respectively) and maturity at the shortest available on a given day. In each panel, the black line represents the data, the red line, the model. The top panel illustrates the performance of the P-LHARG, the second panel refers to the CGARCH model.

4.7 Conclusions

Motivated by the presence of measurement errors in the empirically computed realized volatility measures we introduce a new family of discrete-time SV option pricing models, named SV-LHARG(p). The SV-LHARG(p) model is characterized by two *measurement equations* (one extracting information from the daily returns and the other from the RV measure) and a transition equation for the latent states CV described by a general and flexible Heterogeneous Autoregressive Gamma process with leverage effects.

The SV-LHARG(p) represents the first fully analytical option pricing framework for discrete-time SV models incorporating realized measures of volatility. Indeed, it inherits several analytical features from Majewski et al. (2015). Specifically: (i) the recursive formula for the computation of the conditional MGF under $\mathbb{P}$, (ii) the explicit change of measure using a general and flexible exponentially affine SDF, (iii) the characterization of the no-arbitrage condition in terms of risk premia, (iv) the explicit one-to-one mapping between the parameters of the latent process under $\mathbb{P}$ and $\mathbb{Q}$, (v) the recursive formula for the computation of the conditional MGF under $\mathbb{Q}$. In addition, building on Creal (2015), we derive the analytical filtering and smoothing for the basic specification of the SV-LHARG(p) with $p = 1$ and no leverage effect, the so called SV-ARG. Then, we employ this novel results to design an effective Bayesian inference procedure for both the parameters and the latent factor of the general model SV-LHARG(p) by sampling the analytically tractable SV-ARG through a M-H step. The key feature of the Stochastic Volatility model is that, thanks to the filtering and smoothing procedures, the dynamics of the estimated CV is less noisy and more persistent than the dynamics of the Realized

Volatility measure.

The estimation methodology is extensively tested and validated on simulated data and then applied to daily returns data on the S&P 500 Future index. Finally, in an option pricing exercise, we benchmark the performance of our SV-LHARG model with that of competitor models taken from both the GARCH and RV option pricing literature. The proposed model tracks the dynamics of the short-end of the implied volatility surface with remarkable realism. As already documented in Corsi et al. (2013) the CGARCH tends to reproduce the empirical level with some delay (especially during periods of high volatility), whereas the SV model reacts more dynamically to changes in the volatility level. The higher persistence of the latent volatility in the SV-LHARG model is the crucial feature which allows to over-perform the P-LHARG model when pricing medium to long maturity options.

A Appendix

A.1 Proof of the results in Section 4.2

Proof of Proposition 4.2.1. First, we need to show that the one-step ahead conditional moment generating function is exponential-affine in the state variables (returns, conditional variance and leverage).

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[e^{\bar{z}r_{t+1}+\bar{b}\mathrm{CV}_{t+1}+\bar{c}l_{t+1}+\bar{d}\mathrm{RV}_{t+1}}|\widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t, l_t\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[e^{\bar{z}r_{t+1}+\bar{b}\mathrm{CV}_{t+1}+\bar{c}l_{t+1}}\mathbb{E}^{\mathbb{P}}\left[e^{\bar{d}\mathrm{RV}_{t+1}}|r_{t+1}, h_{t+1}, l_{t+1}\right]|\widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t, l_t\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[e^{\bar{z}r_{t+1}+(\bar{b}-e^{\kappa_2}\log(1-\alpha e^{-\kappa_2}\bar{d}))\mathrm{CV}_{t+1}+\bar{c}l_{t+1}}|\widetilde{\mathcal{F}}_t^{\mathrm{CV}}, r_t, l_t\right].
\end{aligned}$$

Then, the explicit form of the scalar functions $\mathcal{A}$, $\mathcal{B}_i$ and $\mathcal{C}_i$ follows from Appendix C in Majewski et al. (2015). □

Proof of Proposition 4.2.3. Let us define the function $x : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as

$$x(\bar{z}, \bar{b}, \bar{c}) = \bar{z}\lambda + \bar{b} + \frac{\frac{1}{2}\bar{z}^2 + \lambda^2\bar{c} - 2\bar{c}\bar{z}\lambda}{1 - 2\bar{c}},$$

and $x_1 = x(0, u_1, v_1)$. Reasoning as in Appendix F of Gouriéroux and Jasiak (2006), we have to find the conditions which ensure that the solution of the $2p$-dimensional recursive system:

$$\begin{aligned}
u_{1,t} &= \frac{cx_{1,t-1}}{1 - cx_{1,t-1}}\beta_1 + u_{2,t-1}, & v_{1,t} &= \frac{cx_{1,t-1}}{1 - cx_{1,t-1}}\alpha_1 + v_{2,t-1} \\
&\vdots & &\vdots \\
u_{p-1,t} &= \frac{cx_{1,t-1}}{1 - cx_{1,t-1}}\beta_{p-1} + u_{p,t-1}, & v_{p-1,t} &= \frac{cx_{1,t-1}}{1 - cx_{1,t-1}}\alpha_{p-1} + v_{p,t-1} \\
u_{p,t} &= \frac{cx_{1,t-1}}{1 - cx_{1,t-1}}\beta_p, & v_{p,t} &= \frac{cx_{1,t-1}}{1 - cx_{1,t-1}}\alpha_p
\end{aligned}$$

tends to $(0, \ldots, 0)'$ when t tends to infinity, for any non-negative initial values $(u_{1,0}, \ldots, u_{p,0})'$

and $(v_{1,0},\dots,v_{p,0})'$. System above is equivalent to the following one

$$\begin{aligned} u_{1,t} &= \frac{\beta_1}{\beta_p}u_{p,t} + u_{2,t-1}, & v_{1,t} &= \frac{\alpha_1}{\alpha_p}v_{p,t} + v_{2,t-1} \\ &\vdots & &\vdots \\ u_{p-1,t} &= \frac{\beta_{p-1}}{\beta_p}u_{p,t} + u_{p,t-1}, & v_{p-1,t} &= \frac{\alpha_{p-1}}{\alpha_p}v_{p,t} + v_{p,t-1} \\ u_{p,t} &= \frac{\beta_p}{1-cx_{1,t-1}} - \beta_p, & v_{p,t} &= \frac{\alpha_p}{1-cx_{1,t-1}} - \alpha_p. \end{aligned}$$

From the latter, it follows that $u_{i,t}$ and $v_{j,t}$ take non negative values for all $i,j \in \{1,\dots,p\}$ and that $u_{p,t}$ and $v_{p,t}$ are always larger than $-\beta_p$ and $-\alpha_p$, respectively. Moreover, the sequence $(u_{p,t})$ satisfies the non-linear recursive equation:

$$u_{p,t} = \frac{\beta_p}{1-\frac{c}{\beta_p}\left(\sum_{i=1}^{p} u_{p,t-i}\beta_i + \lambda^2\beta_p\frac{\sum_{j=1}^{p} v_{p,t-j}\alpha_j}{1-2\sum_{j=1}^{p} v_{p,t-j}\alpha_j}\right)} - \beta_p. \tag{A.1}$$

In light of the relation $v_{p,t}\beta_p = u_{p,t}\alpha_p$, all $t \geq 0$, we rewrite previous Equation (A.1) as follows:

$$u_{p,t} = \frac{\beta_p}{1-\frac{c}{\beta_p}\left(\sum_{i=1}^{p} u_{p,t-i}\beta_i + \lambda^2\frac{\sum_{j=1}^{p} u_{p,t-j}\alpha_j}{1-\frac{2}{\beta_p}\sum_{j=1}^{p} u_{p,t-j}\alpha_j}\right)} - \beta_p.$$

Thus, a possible limiting value l for the rescaled sequence $(u_{p,t}\beta_p^{-1})$ satisfies:

$$l = \frac{1}{1-lc\left(\sum_{i=1}^{p}\beta_i + \lambda^2\frac{\sum_{j=1}^{p} a_j}{1-2l\sum j=1^p\alpha_j}\right)} - 1 \doteq \frac{1}{1-lc\left(\|\boldsymbol{\beta}\| + \lambda^2\frac{\|\boldsymbol{\alpha}\|}{1-2l\|\boldsymbol{\alpha}\|}\right)} - 1.$$

Therefore, the admissible values are $l = 0$ and $l = \frac{1}{c(\|\boldsymbol{\beta}\|+\lambda^2\|\boldsymbol{\alpha})} - 1$.
If $c(\|\boldsymbol{\beta}\| + \gamma^2\|\boldsymbol{\alpha}\|) = \sum_{j=1}^{p}\phi_j + c\lambda^2\sum_{j=1}^{p}\alpha_j < 1$ the rescaled sequence $(u_{p,t}\beta_p^{-1})$ takes values in the compact set $[-1,0]$ and the unique solution is $l = 0$. Given the relation $v_{p,t} = u_{p,t}\alpha_p/\beta_p$, the same conclusion holds for $v_{p,t}$. □

A.2 Definition of distributions used in this chapter

Generalized Inverse Gaussian distribution

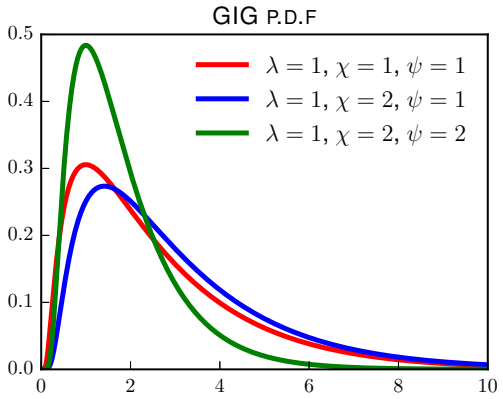
A Generalized Inverse Gaussian (GIG) random variable $X \overset{d}{\sim} \mathcal{G}ig(\lambda,\chi,\psi)$ has probability distribution function (p.d.f) given by:

$$p(x|\lambda,\chi,\psi) = \left(\sqrt{\frac{\psi}{\chi}}\right)^{\lambda}\frac{1}{2K_\lambda(\sqrt{\chi\psi})}x^{\lambda-1}\exp\left(-\frac{1}{2}\left(\chi\frac{1}{x}+\psi x\right)\right),$$

The GIG distribution has the Gamma distribution as special cases. More specifically, the Gamma distribution $\mathcal{G}(k,\vartheta)$ with shape $k > 0$ and scale $\vartheta > 0$ can be obtained setting $\lambda = k$, $\psi = 2/\vartheta$ ad $\chi = 0$ in a $\mathcal{G}ig(\lambda,\chi,\psi)$. The non-central moments of order δ of a GIG distribution are defined as

$$\mathbb{E}\left(X^\delta\right) = \left(\sqrt{\frac{\chi}{\psi}}\right)^{\delta}\frac{K_{\lambda+\delta}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}.$$

Figure A.1: Generalized Inverse Gaussian distribution



NOTES: reported is the probability density function (p.d.f) of a Generalized Inverse Gaussian (GIG) random variable for three different parameter setting (λ, χ, ψ).

Figure A.1 represents the p.d.f of a GIG random variable for three different parameters setting.

Sichel and Negative Binomial distribution

A Sichel ($\mathcal{S}$) random variable $Z \overset{d}{\sim} \mathcal{S}(\lambda, \chi, \psi)$ is obtained by taking a Poisson ($\mathcal{P}o$) random variable $Z \overset{d}{\sim} \mathcal{P}o(X)$ and allowing its mean X to be a random draw from a GIG ($\mathcal{G}ig$) distribution (see Subsection A.2 above), $X \overset{d}{\sim} \mathcal{G}ig(\lambda, \chi, \psi)$. A Sichel random variable has mass function given by:

$$p(z|\lambda, \chi, \psi) = \left(\sqrt{\frac{\psi}{\psi+2}}\right)^{\lambda} \left(\frac{\chi}{\psi+2}\right)^{z} \frac{1}{z!} \frac{K_{\lambda+z}\left(\sqrt{\chi(\psi+2)}\right)}{K_{\lambda}(\sqrt{\chi\psi})}, \; z \geq 0.$$

The first two moments of a Sichel random variable are defined as:

$$\mathbb{E}(Z) = \left(\sqrt{\frac{\chi}{\psi}}\right) \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})},$$
$$\mathbb{E}(Z^2) = \left(\sqrt{\frac{\chi}{\psi}}\right) \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})} + \left(\frac{\chi}{\psi}\right) \frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}.$$

Figure A.2 represents the p.d.f. function of a Sichel random variable for three different parameters setting.

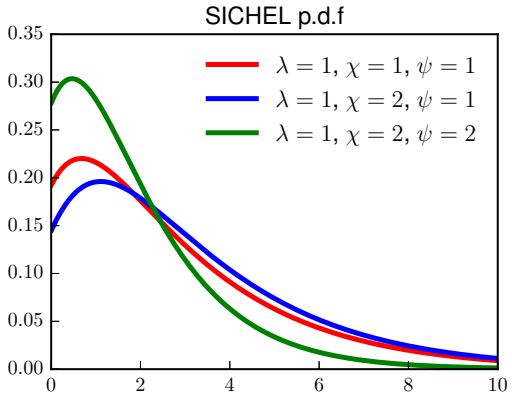
A negative binomial ($\mathcal{NB}$) random variable $K \overset{d}{\sim} \mathcal{NB}(\omega, p)$ is a special case of a $\mathcal{S}(\lambda, \chi, \psi)$ random variable as χ tends to zero. In particular, A $\mathcal{NB}$ has mass function given by:

$$p(k|\omega, p) = \frac{\Gamma(\omega+k)}{\Gamma(\omega)\Gamma(k+1)} p^k (1-p)^{\omega}.$$

Computation of the modified Bessel function of the second kind

In order to calculate the (logarithm of the) modified Bessel function of the second kind $\log(K_{\nu+z_t}(x))$ in a computationally stable manner, we implement the same algorithm

Figure A.2: Sichel distribution



NOTES: reported is the probability density function (p.d.f) of a Sichel random variable for three different parameter setting (λ, χ, ψ).

as in Creal (2015). More precisely, we first define the ratio of Bessel functions $R_{\nu+j}$ as $R_{\nu+j} = \frac{K_{\nu+j}(x)}{K_{\nu+j-1}(x)}$, then we start by evaluating $K_{\nu+j}(x)$ at the stable values $j = 0$ and $j = 1$ and by computing $R_{\nu+1}(x)$. Finally, we apply the following recursion:

$$
\begin{aligned}
R_{\nu+j}(x) &= \frac{1}{R_{\nu+j-1}(x)} + \frac{2}{x}(\nu + j - 1), \\
\log\left(K_{\nu+j}(x)\right) &= \log\left(R_{\nu+j}(x)\right) + \log\left(K_{\nu+j-1}(x)\right),
\end{aligned}
$$

$j = 2, \ldots, z_t$, which gives $K_{\nu+z_t}$ for any z_t.

A.3 Proof of the results in Section 4.3

Proof of Proposition 4.3.1. First, we solve the following integrals:

$$p(r_t, \mathrm{RV}_t | z_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = \int_0^\infty p(r_t, \mathrm{RV}_t | \mathrm{CV}_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) p(\mathrm{CV}_t | z_t; \boldsymbol{\theta})\, d\mathrm{CV}_t \tag{A.2}$$

$$p(z_t | z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = \int_0^\infty p(z_t | \mathrm{CV}_{t-1}; \boldsymbol{\theta}) p(\mathrm{CV}_{t-1} | r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})\, d\mathrm{CV}_{t-1} \tag{A.3}$$

$$p(z_1; \boldsymbol{\theta}) = \int_0^\infty p(z_1 | h_0; \boldsymbol{\theta}) p(h_0; \boldsymbol{\theta}) dh_0, \tag{A.4}$$

and then we compute the *conditional likelihood*, the *Markov transition* and the *initial distribution* of z_1, respectively.

Define the following quantities:

$$
\begin{aligned}
&\mu_{1t} = r_t - \mu - \boldsymbol{\beta}_1' \boldsymbol{x}_{1t}, \\
&\mu_{2t} = \frac{\mathrm{RV}_t}{\exp\left(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right)}, \\
&\alpha_t = \alpha \exp\left(-\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right), \\
&\eta(z_t, \mathrm{RV}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = \frac{\exp\left(\gamma \mu_{1t}\right)}{\sqrt{2\pi}} \frac{\mathrm{RV}^{\alpha_t - 1}}{\Gamma(\alpha_t) \left(\exp\left(\boldsymbol{\kappa_2'} \boldsymbol{x}_{2t}\right)\right)^{\alpha_t}} \frac{c^{-\nu - z_t}}{\Gamma(\nu + z_t)}, \\
&\lambda(z_t) = \nu + z_t - \alpha_t - 1/2, \\
&\chi^{(t)} = \mu_{1t}^2 + 2\mu_{2t}, \\
&\psi = \gamma^2 + \frac{2}{c},
\end{aligned}
$$

then, the *conditional likelihood* is

$$
\begin{aligned}
&p(r_t, \mathrm{RV}_t | z_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = \int_0^\infty p(r_t, \mathrm{RV}_t, \mathrm{CV}_t | z_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})\, d\mathrm{CV}_t = \\
&= \int_0^\infty p(r_t | \mathrm{CV}_t, \boldsymbol{x}_{1t}; \boldsymbol{\theta}) p(\mathrm{RV}_t | \mathrm{CV}_t, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) p(\mathrm{CV}_t | z_t)\, d\mathrm{CV}_t \\
&= \int_0^\infty (2\pi \mathrm{CV}_t)^{-1/2} \exp\left(-\frac{1}{2}\left(\left(r_t - \mu - \boldsymbol{\kappa}_1' \boldsymbol{x}_{1t}\right)^2 \frac{1}{\mathrm{CV}_t} + \gamma^2 \mathrm{CV}_t\right)\right) \exp\left(\gamma\left(r_t - \mu - \boldsymbol{\kappa}_1' \boldsymbol{x}_{1t}\right)\right) \\
&\quad \cdot \frac{\mathrm{RV}_t^{\alpha_t - 1}}{\Gamma(\alpha_t) \left(\exp\left(\boldsymbol{\beta}_2' \boldsymbol{x}_{2t}\right) \mathrm{CV}_t\right)^{\alpha_t}} \exp\left(-\frac{y_t}{\exp(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t})} \frac{1}{\mathrm{CV}_t}\right) \\
&\quad \cdot \frac{c^{-\nu - z_t}}{\Gamma(\nu + z_t)} \mathrm{CV}_t^{\nu + z_t - 1} \exp\left(-\frac{\mathrm{CV}_t}{c}\right) d\mathrm{CV}_t \\
&\doteq \int_0^\infty (2\pi)^{-1/2} \mathrm{CV}_t^{-1/2} \exp\left(-\frac{1}{2}\left(\mu_{1t}^2 \frac{1}{\mathrm{CV}_t} + \gamma^2 \mathrm{CV}_t\right)\right) \exp\left(\gamma \mu_{1t}\right) \\
&\quad \cdot \frac{y_t^{\alpha_t - 1}}{\Gamma(\alpha_t) \left(\exp\left(\boldsymbol{\beta}_2' \boldsymbol{x}_{2t}\right)\right)^{\alpha_t}} \mathrm{CV}_t^{-\alpha_t} \exp\left(-\mu_{2t} \frac{1}{\mathrm{CV}_t}\right) \\
&\quad \cdot \frac{c^{-\nu - z_t}}{\Gamma(\nu + z_t)} \mathrm{CV}_t^{\nu + z_t - 1} \exp\left(-\frac{\mathrm{CV}_t}{c}\right) d\mathrm{CV}_t \\
&= \frac{\exp\left(\gamma \mu_{1t}\right)}{\sqrt{2\pi}} \frac{\mathrm{RV}_t^{\alpha_t - 1}}{\Gamma(\alpha_t) \left(\exp\left(\boldsymbol{\kappa_2'} \boldsymbol{x}_{2t}\right)\right)^{\alpha_t}} \frac{c^{-\nu - z_t}}{\Gamma(\nu + z_t)} \\
&\quad \cdot \int_0^\infty \mathrm{CV}_t^{(\nu + z_t - \alpha_t - 1/2) - 1} \exp\left(-\frac{1}{2}\left(\left(\mu_{1t}^2 + 2\mu_{2t}\right) \frac{1}{\mathrm{CV}_t} + \left(\gamma^2 + \frac{2}{c}\right) \mathrm{CV}_t\right)\right) d\mathrm{CV}_t \\
&\doteq \eta\left(z_t, \mathrm{RV}_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{\theta}\right) \int_0^\infty \mathrm{CV}_t^{\lambda(z_t) - 1} \exp\left(-\frac{1}{2}\left(\left(\mu_{1t}^2 + 2\mu_{2t}\right) \frac{1}{\mathrm{CV}_t} + \psi \mathrm{CV}_t\right)\right) d\mathrm{CV}_t \\
&= 2\eta\left(z_t, \mathrm{RV}_t, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{\theta}\right) K_{\lambda(z_t)}\left(\sqrt{\psi \chi^{(t)}}\right) \left(\sqrt{\frac{\chi^{(t)}}{\psi}}\right)^{\lambda(z_t)}.
\end{aligned}
$$

The last equality follows from the definition of the kernel of the Generalized Inverse Gaussian distribution in Equation (A.2) with parameters $\lambda(z_t)$, $\chi^{(t)}$ and ψ.

Markov transition

$$
\begin{aligned}
&p(z_t|z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta}) = \int_0^\infty p(z_t, \mathrm{CV}_{t-1}|z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})\, d\mathrm{CV}_{t-1} = \\
&= \int_0^\infty p(z_t|\mathrm{CV}_{t-1}) p(\mathrm{CV}_{t-1}|z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})\, d\mathrm{CV}_{t-1} = \\
&= \int_0^\infty p(z_t|\mathrm{CV}_{t-1}) \frac{p(\mathrm{CV}_{t-1}, z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})}{p(z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})}\, d\mathrm{CV}_{t-1} = \\
&\propto \int_0^\infty p(z_t|\mathrm{CV}_{t-1}) p(\mathrm{CV}_{t-1}, z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}; \boldsymbol{\theta})\, d\mathrm{CV}_{t-1} \\
&\propto \int_0^\infty p(z_t|\mathrm{CV}_{t-1}; \boldsymbol{\theta}) p(r_{t-1}|\mathrm{CV}_{t-1}, \boldsymbol{x}_{1t-1}; \boldsymbol{\theta}) p(\mathrm{RV}_{t-1}|\mathrm{CV}_{t-1}, \boldsymbol{x}_{2t-1}; \boldsymbol{\theta}) p(\mathrm{CV}_{t-1}|z_{t-1}; \boldsymbol{\theta})\, d\mathrm{CV}_{t-1} = \\
&\propto \int_0^\infty \frac{1}{z_t!} \left(\frac{\phi^{(d)}}{c}\mathrm{CV}_{t-1}\right)^{z_t} \exp\left(-\frac{\phi^{(d)}}{c}\mathrm{CV}_{t-1}\right) p(r_{t-1}|\mathrm{CV}_{t-1}, \boldsymbol{x}_{1t-1}, \boldsymbol{\theta}) \\
&\qquad \cdot p(\mathrm{RV}_{t-1}|\mathrm{CV}_{t-1}, \boldsymbol{x}_{2t-1}, \boldsymbol{\theta}) p(\mathrm{CV}_{t-1}|z_{t-1}; \boldsymbol{\theta})\, d\mathrm{CV}_{t-1}.
\end{aligned}
\tag{A.5}
$$

Similarly to the computation of the *conditional likelihood* above we have:

$$
\begin{aligned}
&p(r_{t-1}|\mathrm{CV}_{t-1}, \boldsymbol{x}_{1t-1}, \boldsymbol{\theta}) p(\mathrm{RV}_{t-1}|\mathrm{CV}_{t-1}, \boldsymbol{x}_{2t-1}, \boldsymbol{\theta}) p(\mathrm{CV}_{t-1}|z_{t-1}; \boldsymbol{\theta}) = \\
&= \mathrm{CV}_{t-1}^{(\nu+z_{t-1}-\alpha_t-1/2)-1} \exp\left(-\frac{1}{2}\left(\left((r_{t-1} - \mu - \boldsymbol{\kappa}_1' \boldsymbol{x}_{1t-1})^2\right.\right.\right. \\
&\qquad \left.\left.\left. + 2\frac{\mathrm{RV}_{t-1}}{\exp\left(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t-1}\right)}\right) \frac{1}{\mathrm{CV}_{t-1}} + \left(\gamma^2 + \frac{2}{c}\right)\mathrm{CV}_{t-1}\right)\right) = \\
&= \mathrm{CV}_{t-1}^{\lambda(z_{t-1})-1} \exp\left(-\frac{1}{2}\left(\left(\mu_{1t}^2 + 2\mu_{2t}\right)\frac{1}{\mathrm{CV}_{t-1}} + \psi \mathrm{CV}_{t-1}\right)\right) \doteq \\
&\doteq \mathrm{CV}_{t-1}^{\lambda(z_{t-1})-1} \exp\left(-\frac{1}{2}\left(\chi^{(t-1)}\frac{1}{\mathrm{CV}_{t-1}} + \psi \mathrm{CV}_{t-1}\right)\right).
\end{aligned}
$$

Equation (A.5) becomes:

$$
\begin{aligned}
&p(z_t|z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{\theta}) \propto \\
&\propto \int_0^\infty \frac{1}{z_t!}\left(\frac{\phi^{(d)}}{c}\mathrm{CV}_{t-1}\right)^{z_t} \exp\left(-\frac{\phi^{(d)}}{c}\mathrm{CV}_{t-1}\right) \mathrm{CV}_{t-1}^{\lambda(z_{t-1})-1} \\
&\qquad \cdot \exp\left(-\frac{1}{2}\left(\chi^{(t-1)}\frac{1}{\mathrm{CV}_{t-1}} + \psi \mathrm{CV}_{t-1}\right)\right) d\mathrm{CV}_{t-1} = \\
&= \frac{1}{z_t!}\left(\frac{\phi^{(d)}}{c}\right)^{z_t} \int_0^\infty \mathrm{CV}_{t-1}^{(\lambda(z_{t-1})+z_t)-1} \exp\left(-\frac{1}{2}\left(\chi^{(t-1)}\frac{1}{\mathrm{CV}_{t-1}}\right.\right. \\
&\qquad \left.\left. + \left(\psi + 2\frac{\phi^{(d)}}{c}\right)\mathrm{CV}_{t-1}\right)\right) d\mathrm{CV}_{t-1} \\
&= \frac{1}{z_t!}\left(\frac{\phi^{(d)}}{c}\right)^{z_t} 2K_{\lambda(z_{t-1})+z_t}\left(\sqrt{\chi^{(t-1)}\left(\psi + 2\frac{\phi^{(d)}}{c}\right)}\right)\left(\sqrt{\frac{\chi^{(t-1)}}{(\psi + 2\phi^{(d)}/c)}}\right)^{\lambda(z_{t-1})+z_t} \\
&\propto \frac{1}{z_t!}\left(\sqrt{\frac{(\phi^{(d)})^2}{c^2}\frac{\chi^{(t-1)}}{(\psi + 2\phi^{(d)}/c)}}\right)^{z_t}\left(\sqrt{\frac{\chi^{(t-1)}}{(\psi + 2\phi^{(d)}/c)}}\right)^{\lambda(z_{t-1})} K_{\lambda(z_{t-1})+z_t}\left(\sqrt{\chi^{(t-1)}\left(\psi + 2\frac{\phi^{(d)}}{c}\right)}\right).
\end{aligned}
\tag{A.6}
$$

If we define

$$
\begin{aligned}
\bar{\chi}^{(t-1)} &\doteq \chi^{(t-1)}\frac{\phi^{(d)}}{c}, \\
\bar{\psi} &\doteq \psi\frac{c}{\phi^{(d)}},
\end{aligned}
$$

then the *Markov transition* can be re-written as:

$$
\begin{aligned}
&p(z_t|z_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_{1t}, \boldsymbol{x}_{2t}, \boldsymbol{\theta}) \propto \\
&\propto \frac{1}{z_t!}\left(\sqrt{\frac{\bar{\chi}^{(t-1)}}{\bar{\psi}+2}}\right)^{z_t}\left(\sqrt{\frac{\bar{\psi}}{(\bar{\psi}+2)}}\right)^{\lambda(z_{t-1})} K_{\lambda(z_{t-1})+z_t}\left(\sqrt{\bar{\chi}^{(t-1)}\left(\bar{\psi}+2\right)}\right) \\
&\propto \mathcal{S}\left(\lambda(z_{t-1}), \bar{\chi}^{(t-1)}, \bar{\psi}\right) \\
&\propto \mathcal{S}\left(\nu + z_{t-1} - \alpha_t - 1/2, \chi^{(t-1)}\frac{\phi^{(d)}}{c}, \psi\frac{c}{\phi^{(d)}}\right).
\end{aligned}
$$

where $\mathcal{S}\left(\lambda(z_{t-1}), \bar{\chi}^{(t-1)}, \bar{\psi}\right)$ is the Sichel distribution with parameters $\lambda(z_{t-1})$, $\bar{\chi}^{(t-1)}$ and $\bar{\psi}$.

Initial distribution

$$
\begin{aligned}
p(z_1;\boldsymbol{\theta}) &= \int_0^\infty p(z_1, \mathrm{CV}_{t-1};\boldsymbol{\theta}) d\mathrm{CV}_{t-1} \\
&= \int_0^\infty p(z_1|\mathrm{CV}_{t-1};\boldsymbol{\theta}) p(\mathrm{CV}_{t-1};\boldsymbol{\theta}) d\mathrm{CV}_{t-1} \\
&= \int_0^\infty \frac{1}{z_1!}\left(\frac{\phi^{(d)}}{c}\mathrm{CV}_{t-1}\right)^{z_1} \exp\left(-\frac{\phi^{(d)}}{c}\mathrm{CV}_{t-1}\right) \frac{1}{\Gamma(\nu)}\mathrm{CV}_{t-1}^{\nu-1}\left(\frac{1-\phi^{(d)}}{c}\right)^{\nu} \\
&\qquad \cdot \exp\left(-\mathrm{CV}_{t-1}\left(\frac{1-\phi^{(d)}}{c}\right)\right) d\mathrm{CV}_{t-1} \\
&= \frac{1}{z_1!}\left(\frac{\phi^{(d)}}{c}\right)^{z_1}\left(\frac{1-\phi^{(d)}}{c}\right)^{\nu} \frac{1}{\Gamma(\nu)} \int_0^\infty \mathrm{CV}_{t-1}^{\nu+z_1-1} \exp\left(-\frac{\mathrm{CV}_{t-1}}{c}\right) d\mathrm{CV}_{t-1} \\
&= \frac{1}{z_1!}\left(\frac{\phi^{(d)}}{c}\right)^{z_1}\left(\frac{1-\phi^{(d)}}{c}\right)^{\nu} \frac{\Gamma(\nu+z_1)}{\Gamma(\nu)} c^{\nu+z_1} \\
&= \frac{1}{z_1!}\left(\phi^{(d)}\right)^{z_1}(1-\phi^{(d)})^{\nu}\frac{\Gamma(\nu+z_1)}{\Gamma(\nu)\Gamma(z_1+1)} \\
&\propto \mathcal{NB}\left(\nu, \phi^{(d)}\right),
\end{aligned}
$$

where $\mathcal{NB}\left(\nu, \phi^{(d)}\right)$ is the Negative Binomial distribution with parameters ν and $\phi^{(d)}$. □

Proof of Proposition 4.3.2. Marginal filtered distribution

$$
\begin{aligned}
&p(\mathrm{CV}_t|\boldsymbol{r}_{1:t}, \mathbf{RV}_{1:t}, \boldsymbol{z}_{1:t}, \boldsymbol{x}_t;\boldsymbol{\theta}) \propto \\
&\propto p(r_t|\mathrm{CV}_t, \boldsymbol{x}_{1t};\boldsymbol{\theta}) p(\mathrm{RV}_t|\mathrm{CV}_t, \boldsymbol{x}_{2t};\boldsymbol{\theta}) p(\mathrm{CV}_t|z_t;\boldsymbol{\theta}) \propto \\
&\propto \mathrm{CV}_t^{-1/2}\exp\left(-\frac{1}{2}\left(\left(r_t-\mu-\boldsymbol{\kappa}_1'\boldsymbol{x}_{1t}\right)^2\frac{1}{\mathrm{CV}_t}+\gamma^2\mathrm{CV}_t\right)\right)\mathrm{CV}_t^{-\alpha_t} \\
&\qquad \cdot \exp\left(-\frac{\mathrm{RV}_t}{\exp(\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t})}\frac{1}{\mathrm{CV}_t}\right)\mathrm{CV}_t^{\nu+z_t-1}\exp\left(-\frac{\mathrm{CV}_t}{c}\right) \\
&= \mathrm{CV}_t^{(\nu+z_t-\alpha_t-1/2)-1}\exp\left(-\frac{1}{2}\left(\left((r_t-\mu-\boldsymbol{\kappa}_1'\boldsymbol{x}_{1t})^2+2\frac{\mathrm{RV}_t}{\exp\left(\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)}\right)\frac{1}{\mathrm{CV}_t}\right.\right. \\
&\qquad \left.\left.+\left(\gamma^2+\frac{2}{c}\right)\mathrm{CV}_t\right)\right) d\mathrm{CV}_t \\
&= \mathrm{CV}_t^{\lambda(z_t)-1}\exp\left(-\frac{1}{2}\left(\left(\mu_{1t}^2+2\mu_{2t}\right)\frac{1}{\mathrm{CV}_t}+\psi\mathrm{CV}_t\right)\right) \\
&= \mathrm{CV}_t^{\lambda(z_t)-1}\exp\left(-\frac{1}{2}\left(\chi^{(t)}\frac{1}{\mathrm{CV}_t}+\psi\mathrm{CV}_t\right)\right) \\
&\propto \mathcal{G}ig(\lambda(z_t), \chi^{(t)}, \psi).
\end{aligned}
$$

Marginal smoothed distribution

$$
\begin{aligned}
&p(\mathrm{CV}_t|\boldsymbol{r}_{1:T}, \mathbf{RV}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{x}_{1:T}; \boldsymbol{\theta}) \propto \\
&\propto p(\mathrm{CV}_t|\boldsymbol{r}_{1:t}, \mathbf{RV}_{1:t}, \boldsymbol{z}_{1:t}, \boldsymbol{x}_{1:t}; \boldsymbol{\theta}) p(z_{t+1}|\mathrm{CV}_t, \boldsymbol{\theta}) \propto \\
&\propto \mathrm{CV}_t^{\lambda(z_t)-1} \exp\left(-\frac{1}{2}\left(\chi^{(t)}\frac{1}{\mathrm{CV}_t} + \psi \mathrm{CV}_t\right)\right) \mathrm{CV}_t^{z_{t+1}} \exp\left(-\frac{\phi^{(d)}}{c}\mathrm{CV}_t\right) = \\
&= \mathrm{CV}_t^{\lambda(z_t)+z_{t+1}-1} \exp\left(-\frac{1}{2}\left(\chi^{(t)}\frac{1}{\mathrm{CV}_t} + \left(\psi + 2\frac{\phi^{(d)}}{c}\right)\mathrm{CV}_t\right)\right) \\
&\propto \mathcal{G}ig\left(\lambda(z_t) + z_{t+1}, \chi^{(t)}, \psi + 2\frac{\phi^{(d)}}{c}\right).
\end{aligned}
$$

□

A.4 Background material in Markov Chain Monte Carlo

Essentials for MCMC

Most of the material in this section is from Casella and Robert (2004), Chapters 6, 7 and 10. In what follows, we give a brief introduction to Markov Chain Monte Carlo (MCMC) methods and formally define the Metropolis-Hastings algorithm.

Broadly speaking, a Markov Chain Monte Carlo (MCMC) method for the simulation of a distribution π, defined on a state space $\mathcal{X}$, is any method producing an ergodic Markov chain, $X_1, X_2, \ldots, X_n, \ldots$ whose stationary probability measure is π. Precisely, let

$$\mathfrak{I} = \mathbb{E}_\pi\left[h(x)\right] = \int_{\mathcal{X}} h(x) d\pi(x),$$

then, the ergodic theorem guarantees the almost sure convergence to the quantity $\mathfrak{I}$ of the empirical average

$$\widehat{\mathfrak{I}}_N = \frac{1}{N}\sum_{n=1}^{N} h(X_n). \tag{A.7}$$

Thus, a sequence produced by a MCMC algorithm can be employed just as an i.i.d. sample.

In the set-up of MCMC algorithms, Markov chains are defined by a transition kernel. Precisely, the following definition holds true, where we indicate with $\mathcal{B}(\mathcal{X})$ the Borel sets of $\mathcal{X}$.

Definition A.1. *A transition kernel is a function K defined on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ such that*

- $\forall\, x \in \mathcal{X}$, $K(x, \cdot)$ *is a probability measure;*
- $\forall\, A \in \mathcal{B}(\mathcal{X})$, $K(\cdot, A)$ *is measurable.*

Once that a transition kernel is given, the definition of a Markov Chain is readily obtained.

Definition A.2. *Given a transition kernel K, a sequence $X_0, X_1, \cdots, X_n, \cdots$ of random variables is a Markov chain if, for any t, the conditional distribution of X_t given $X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \cdots, X_0 = x_0$ is the same as the distribution of X_t given x_{t-1}. Formally,*

$$\mathbb{P}\left(X_{k+1} \in A | X_0 = x_0, X_1 = x_1, \ldots, X_k = x_k\right) = \mathbb{P}\left(X_{k+1} \in A | X_k = x_k\right) = \int_A K(x_k, dx).$$

Moreover a chain is time homogeneous if the distribution of $(X_{t_1}, \cdots, X_{t_k})$ given $X_0 = x_0$ is the same as the distribution of $(X_{t_1-t_0}, X_{t_2-t_0}, \cdots, X_{t_k-t_0})$ given $X_0 = x_0$ for every k and every $(k+1)$-uplet $t_0 \leq t_1 \leq \cdots \leq t_k$.

It turns out that the chains encountered in MCMC settings has a stationary probability distribution π by construction; that is, a distribution π such that if $X_n \sim \pi$, then $X_{n+1} \sim \pi$. This means that the kernel K allows for free moves all over the state space. In the theory of Markov Chain this freedom is called irreducibility. The irreducibility property ensures that both the average number of visit to an arbitrary set A is infinite – the chain is recurrent – and the probability of an infinite number of returns to A is one – the chain is Harris recurrent.

The stationary distribution is also a limiting distribution in the sense that the limiting distribution of X_{n+1} is π under the total variation norm, independently from the initial value X_0. Importantly, this implies the convergence as in Equation (A.7). Besides, when the chain is reversible (cfr. Definition A.4) – i.e. the kernel is symmetric – a Central Limit Theorem (CLT) also holds for the empirical average above. We now give the essentials definitions, lemmas and theorems to apply this CLT.
Firstly, we define the kernel for n $(n > 1)$ transitions, $K^n(x, A)$ with $K^1(x, A) = K(x, A)$

$$K^n(x, A) = \int_{\mathcal{X}} K^{n-1}(y, A)\, K(x, dy),$$

then we report the following result, which provides convolution formulas of the type $K^{m+n} = K^m * K^n$.

Lemma A.1 (Chapman-Kolmogorov equation)**.** *For every* $(m, n) \in \mathbb{N}^2$, $x \in \mathcal{X}$, $A \in \mathcal{B}(\mathcal{X})$,

$$K^{m+n}(x, A) = \int_{\mathcal{X}} K^n(y, A)\, K^m(x, dy).$$

We give now the definition of invariant measure.

Definition A.3. *A σ-finite measure π is invariant for the transition kernel $K(\cdot, \cdot)$ and for the associated chain if*

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\, \pi(dx), \qquad \forall B \in \mathcal{B}(\mathcal{X}).$$

If π is a probability measure the invariant distribution is also referred to as stationary. In particular, the chain is stationary in distribution. We turn now to the definition of reversible Markov chain.

Definition A.4. *A stationary Markov chain (X_n) is reversible if the distribution of X_{n+1} conditionally on $X_{n+2} = x$ is the same as the distribution of X_{n+1} conditionally on $X_n = x$.*

The following definition and result provide a sufficient condition for a measure to be the stationary measure of a kernel.

Definition A.5. *A Markov chain with transition kernel K satisfies the detailed balance condition if there exists a function f satisfying*

$$K(y, x) f(y) = K(x, y) f(x)$$

for every (x, y).

Theorem A.1. *Suppose that a Markov chain with transition function K satisfies the detailed balance condition with π a probability density function. Then*

- *The density π is the invariant density of the chain.*
- *The chain is reversible.*

The Metropolis-Hastings algorithm

Within MCMC methods the Metropolis-Hastings (MH) algorithms have the advantage of imposing minimal requirements on the target density π and allowing for a wide choice of possible implementations. It requires the choice of a conditional density q, called proposal density, defined with respect to the dominating measure for the target. Beside, it assumes that the ratio $\pi(y)/q(y|x)$ is known up to a constant independent of x. The transition from the value of the Markov chain at time t and its value at time $t+1$ proceeds via the following transition steps. Given $X_t = x_t$,

- **1st step:** Generate $Y_t \sim q\,(y|x_t)$;
- **2st step:** Take

$$X_{t+1} = \begin{cases} Y_t & \text{with probability} \quad \rho(x_t, Y_t) \\ x_t & \text{with probability} \quad 1 - \rho(x_t, Y_t), \end{cases}$$

  where

$$\rho(x, y) = \min\left\{\frac{\pi(y)}{\pi(x)}\frac{q(x|y)}{q(y|x)}, 1\right\},$$

  is called the Metropolis-Hastings acceptance probability.

The transition kernel of the M-H algorithm is

$$K(x, y) = \rho(x, y)q(y|x) + (1 - r(x))\delta_x(y),$$

where $r(x) = \int \rho(x, y)q(y|x)\,dy$ and δ_x denotes the Dirac mass in x. It is straightforward to verify that the M-H chain satisfies the detailed balance condition, that is

$$\begin{aligned} &\rho(x, y)q(y|x)\pi(x) = \rho(y, x)q(x|y)\pi(y) \\ &(1 - r(x))\delta_x(y)\pi(x) = (1 - r(y))\delta_y(x)\pi(y). \end{aligned}$$

In particular, Theorem A.1 holds true.

A.5 Computational details for the SV-ARG

Sampling the parameters

For the SV-ARG we have to sample from $\boldsymbol{\xi}$, α, $\boldsymbol{\kappa}_2$, ν, c and $\beta^{(d)}$. We refer the reader to Subsection A.6 for details on the sampling of the parameters $\boldsymbol{\xi}$, α, $\boldsymbol{\kappa}_2$, ν and c because their full conditional distributions are the same used for the SV-LHARG.

Instead, the full conditional distribution of $\beta^{(d)}$ is the following one:

$$\begin{aligned}\pi\left(\beta^{(d)}|\boldsymbol{r},\mathbf{RV},\mathbf{CV},\boldsymbol{z},\boldsymbol{\theta}_{(\beta^{(d)})}\right) &\propto \pi\left(\beta^{(d)}|\mathbf{CV},\boldsymbol{z}\right)\\ &\propto \prod_{t=p+1}^{T}\left(\beta^{(d)}\right)^{z_t}\exp\left(-\beta^{(d)}\mathrm{CV}_{t-1}\right)\\ &\propto \left(\beta^{(d)}\right)^{\sum_{t=p+1}^{T}z_t}\exp\left(-\beta^{(d)}\sum_{t=p+1}^{T}\mathrm{CV}_{t-1}\right)\\ &\propto \mathcal{G}\left(k_{\beta^{(d)}},\theta_{\beta^{(d)}}\right),\end{aligned}$$

where $\mathcal{G}(k,\theta)$ indicates a Gamma random variable with shape $k>0$ and scale $\theta>0$; $k_{\beta^{(d)}}=\sum_{t=p+1}^{T}z_t-1$ and $\theta_{\beta^{(d)}}=\left(\sum_{t=p+1}^{T}\mathrm{CV}_{t-1}\right)^{-1}$.

A.6 Computational details for the SV-LHARG

Sampling the parameters

In what follows, we denote with $\odot$ and $\oslash$ the element-by-element multiplication and division respectively, with $\boldsymbol{r}\doteq r_{p+1:T}$, $\mathbf{RV}\doteq y_{p+1:T}$, $\boldsymbol{z}\doteq z_{p+1:T}$, $\mathbf{CV}\doteq \mathrm{CV}_{p+1:T}$, and with $\boldsymbol{\theta}_{(*)}$ the vector $\boldsymbol{\theta}$ deprived by the parameter $(*)$. We use the parametrization in $\left(\left[\beta^{(d)},\beta^{(w)},\beta^{(m)}\right],c\right)$ for the centrality parameter of the non-central gamma distribution. The convergence behaviour of the MCMC chain crucially depends on the parametrization of the latent process (see Bernardo et al., 2003; Frühwirth-Schnatter, 2004; Roberts et al., 2004, for more details). Extensive experimentation shows that the latter parametrization further improves the mixing of the MCMC chain.

The full conditional distribution of the $(2+n_1)$-dimensional vector $\boldsymbol{\xi}$ is

$$\begin{aligned}\pi(\boldsymbol{\xi}|\boldsymbol{r},\mathbf{RV},\mathbf{CV},\boldsymbol{z},\boldsymbol{x},\boldsymbol{\theta}_{(\boldsymbol{\xi})}) &\propto \pi(\boldsymbol{\xi}|\boldsymbol{r},\mathbf{CV},\boldsymbol{x})\\ &\propto \prod_{t=p+1}^{T}\exp\left(-\frac{1}{2}\frac{(r_t-\boldsymbol{\xi}'\boldsymbol{w}_t)^2}{\mathrm{CV}_t}\right)\\ &\propto \exp\left(-\frac{1}{2}(\boldsymbol{r}-W\boldsymbol{\xi})'\Upsilon_{\boldsymbol{r}}^{-1}(\boldsymbol{r}-W\boldsymbol{\xi})\right)\\ &\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\xi}'W'\Upsilon_{\boldsymbol{r}}^{-1}W\boldsymbol{\xi}-2\boldsymbol{\xi}'W'\Upsilon_{\boldsymbol{r}}^{-1}\boldsymbol{r}\right)\right)\\ &\propto \mathcal{N}_{2+n_1}(\mu_{\boldsymbol{\xi}},\Upsilon_{\boldsymbol{\xi}}),\end{aligned}$$

where we indicate with $\boldsymbol{W}$ the $(T-p)\times(2+n_1)$-dimensional matrix $\boldsymbol{W}=(\boldsymbol{w}_{p+1}';\ldots;\boldsymbol{w}_T')$ and with $\Upsilon_{\boldsymbol{r}}$ the $(T-p)\times(T-p)$-dimensional diagonal matrix $\Upsilon_{\boldsymbol{r}}=\mathrm{diag}(\boldsymbol{r})$. For the parameter $\boldsymbol{\xi}$ the full conditional distribution can be sample exactly.

The full conditional distribution of α is

$$\begin{aligned}&\pi(\alpha|\boldsymbol{r},\mathbf{RV},\mathbf{CV},\boldsymbol{z},\boldsymbol{x},\boldsymbol{\theta}_{(\alpha)}) \propto \pi\left(\alpha|\mathbf{RV},\mathbf{CV},\boldsymbol{x},\boldsymbol{\kappa}_2\right)\\ &\propto \prod_{t=p+1}^{T}\frac{1}{\Gamma\left(\alpha\exp\left(-\boldsymbol{\beta}_2'\boldsymbol{x}_{2t}\right)\right)}\frac{1}{\left(\mathrm{CV}_t\exp\left(\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)\right)^{\alpha\exp\left(-\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)}}\mathrm{RV}_t^{\alpha\exp\left(-\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)}\\ &\propto \exp\left(-\sum_{t=p+1}^{T}\log(\Gamma\left(\alpha\exp\left(-\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)\right))-\sum_{t=p+1}^{T}\alpha\exp\left(-\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}\right)\left(\mathrm{CV}_t\boldsymbol{\kappa}_2'\boldsymbol{x}_{2t}-\log(\mathrm{RV}_t)\right)\right).\end{aligned}$$

To simulate from this distribution we employ a Metropolis-Hastings step. We consider a Gamma random walk proposal and, at the j-th iteration of the algorithm, given the previous value $\alpha^{(j-1)}$ of the chain, we simulate

$$\alpha^{(*)} \overset{d}{\sim} \mathcal{G}\left(\alpha^{(j-1),2}/\xi_\alpha, \xi_\alpha/\alpha^{(j-1)}\right),$$

where ξ_α represents the scale of the random walk. The proposal value generated from this density is then accepted or rejected according to the acceptance ratio of the Metropolis-Hastings algorithm.

The full conditional distribution of $\boldsymbol{\kappa}_2$ is

$$\begin{aligned}
&\pi(\boldsymbol{\kappa}_2|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\theta}_{(\boldsymbol{\kappa}_2)}) \propto \pi\left(\alpha|\mathbf{RV}, \mathbf{CV}, \boldsymbol{x}, \alpha\right) \\
&\propto \prod_{t=p+1}^{T} \frac{1}{\Gamma\left(\alpha \exp\left(-\boldsymbol{\beta}_2' \boldsymbol{x}_{2t}\right)\right)} \frac{1}{\left(\mathrm{CV}_t \exp\left(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right)\right)^{\alpha \exp\left(-\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right)}} \mathrm{RV}_t^{\alpha \exp\left(-\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right)} \times \\
&\exp\left(-\frac{\mathrm{RV}_t}{\mathrm{CV}_t \exp\left(\boldsymbol{\kappa}_2' \boldsymbol{x}_{2t}\right)}\right).
\end{aligned}$$

To simulate from this distribution we employ a Metropolis-Hastings step. Specifically, we consider a Normal random walk proposal and, at the j-th iteration of the algorithm, given the previous value $\boldsymbol{\kappa}_2^{(j-1)}$ of the chain, we simulate

$$\boldsymbol{\kappa}_2^{(*)} \overset{d}{\sim} \mathcal{N}_{n_2}\left(\boldsymbol{\kappa}_2^{(j-1)}, \xi_{\boldsymbol{\kappa}_2} \mathrm{I}_{n_2}\right),$$

where $\xi_{\boldsymbol{\kappa}_2}$ represents the scale of the random walk. The proposal value generated from this density is then accepted or rejected according to the acceptance ratio of the Metropolis-Hastings algorithm.

The full conditional distribution of ν is

$$\begin{aligned}
&\pi(\nu|\boldsymbol{r}, \mathbf{RV}, \boldsymbol{z}, \boldsymbol{\theta}_{(\nu)}) \propto \pi(\nu|\mathbf{CV}, \boldsymbol{z}, c) \propto \prod_{t=p+1}^{T} \frac{1}{\Gamma(\nu + z_t)} \left(\frac{1}{c}\right)^{\nu} \mathrm{CV}_t^{\nu} \\
&\propto \exp\left(-\sum_{t=p+1}^{T} \log(\Gamma(\nu + z_t)) - \nu(T-p)\log(c) + \nu \sum_{t=p+1}^{T} \log(\mathrm{CV}_t)\right).
\end{aligned}$$

Similarly to what was done for the parameter α, in the Metropolis-Hastings step, we consider a Gamma random walk proposal with scale ξ_ν. The proposal value generated from this density is the accepted or rejected according to the acceptance ratio of the Metropolis-Hastings algorithm.

The full conditional distribution of c is

$$\begin{aligned}
&\pi\left(c|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\theta}_{(c)}\right) \propto \pi\left(c|\mathbf{CV}, \boldsymbol{z}, \nu\right) \propto \prod_{t=p+1}^{T} \left(\frac{1}{c}\right)^{\nu+z_t} \exp\left(-\frac{\mathrm{CV}_t}{c}\right) \\
&\propto \left(\frac{1}{c}\right)^{(T-p)\nu + \sum_{t=p+1}^{T} z_t} \exp\left(-\frac{1}{c} \sum_{t=p+1}^{T} \mathrm{CV}_t\right) \\
&\propto \mathcal{IG}\left(\bar{k}_c, \bar{\theta}_c\right),
\end{aligned}$$

where $\mathcal{IG}(\bar{k}, \bar{\theta})$ indicates an Inverse Gamma random variable with shape $\bar{k} > 0$ and scale $\bar{\theta} > 0$; $\bar{k}_c = (T-p)\nu + \sum_{t=p+1}^{T} z_t - 1$ and $\bar{\theta}_c = \sum_{t=p+1}^{T} \mathrm{CV}_t$.

In order to sample from $\boldsymbol{\beta}'$, $\boldsymbol{\alpha}'$ and λ, we introduce now the following (3×22)-dimensional matrix

$$\mathrm{E} = \begin{pmatrix} 1 & \mathbf{0}_4' & \mathbf{0}_{17}' \\ 0 & \frac{1}{4}\boldsymbol{\iota}_4' & \mathbf{0}_{17}' \\ 0 & \mathbf{0}_4' & \frac{1}{17}\boldsymbol{\iota}_{17}' \end{pmatrix},$$

where $\boldsymbol{\iota}_n$ and $\mathbf{0}_n$ indicate the n-dimensional unit and null vector. Besides, we indicate with $\mathbf{CV_{t-1}}$ and $\boldsymbol{l_{t-1}}$ the 22-dimensional vectors $\mathbf{CV_{t-1}} = (\mathrm{CV}_{t-1}, \ldots, h_{t-22})'$ and $\boldsymbol{l_{t-1}} = (l_{t-1}, \ldots, l_{t-22})'$, respectively.

The full conditional distribution of $\boldsymbol{\beta} = \left(\beta^{(d)}, \beta^{(w)}, \beta^{(m)}\right)'$ is

$$\pi\left(\boldsymbol{\beta}|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\theta}_{(\boldsymbol{\beta})}\right) \propto \pi\left(\boldsymbol{\beta}|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\alpha}, \lambda\right) \propto$$
$$\propto \prod_{t=p+1}^{T} \left(\left(\mathrm{E}\mathbf{CV_{t-1}}\right)'\boldsymbol{\beta} + \left(\mathrm{E}\boldsymbol{l}_{t-1}\right)'\boldsymbol{\alpha}\right)^{z_t} \exp\left(-\left(\mathrm{E}\mathbf{CV}_{t-1}\right)'\boldsymbol{\beta}\right).$$

To simulate from this distribution we employ a Metropolis-Hastings step. We consider a Normal random walk proposal and, at the j-th iteration of the algorithm, given the previous value $\boldsymbol{\beta}^{(j-1)}$ of the chain, we simulate

$$\boldsymbol{\beta} \overset{d}{\sim} \mathcal{N}_3\left(\boldsymbol{\beta}^{(j-1)}, \Upsilon_{\boldsymbol{\beta}}\right),$$

where $\Upsilon_{\boldsymbol{\beta}}$ is a (3×3)-dimensional diagonal matrix with diagonal given by $\left(\xi_{\beta^{(d)}}, \xi_{\beta^{(w)}}, \xi_{\beta^{(m)}}\right)$. The latter represent the scale of the random walk.

The full conditional distribution of $\boldsymbol{\alpha} = \left(\alpha^{(d)}, \alpha^{(w)}, \alpha^{(m)}\right)$ is

$$\pi\left(\boldsymbol{\alpha}|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\theta}_{(\boldsymbol{\alpha})}\right) \propto \pi\left(\boldsymbol{\alpha}|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\beta}, \lambda\right) \propto$$
$$\propto \prod_{t=p+1}^{T} \left(\left(\mathrm{E}\mathbf{CV_{t-1}}\right)'\boldsymbol{\beta} + \left(\mathrm{E}\boldsymbol{l}_{t-1}\right)'\boldsymbol{\alpha}\right)^{z_t} \exp\left(-\left(\mathrm{E}\boldsymbol{l}_{t-1}\right)'\boldsymbol{\alpha}\right).$$

Similarly to what was done for the parameter $\boldsymbol{\beta}$, in the Metropolis-Hastings step, we consider a Normal random walk proposal with scales $(\xi_{\alpha^{(d)}}, \xi_{\alpha^{(w)}}, \xi_{\alpha^{(m)}})$.

Finally, the full conditional distribution of λ is

$$\pi\left(\lambda|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\theta}_{(\lambda)}\right) \propto \pi\left(\boldsymbol{\alpha}|\boldsymbol{r}, \mathbf{RV}, \mathbf{CV}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\alpha}\right)$$
$$\propto \prod_{t=p+1}^{T} \left(\left(\mathrm{E}\mathbf{CV_{t-1}}\right)'\boldsymbol{\beta} + \left(\mathrm{E}\boldsymbol{l}_{t-1}\right)'\boldsymbol{\alpha}\right)^{z_t} \exp\left(-\left(\mathrm{E}\boldsymbol{l}_{t-1}\right)'\boldsymbol{\alpha}\right),$$

where, we remind that $l_{t-i} = \left(\epsilon_{t-i} - \lambda\sqrt{\mathrm{CV}_{t-i}}\right)^2$, $i \in \{1, \ldots, 22\}$. To simulate from this distribution we employ a Metropolis-Hastings algorithm with a proposal distribution that makes use of the information on the structure of the leverage component. In particular, we would like to capture the asymmetric influence of shock: a large positive idiosyncratic component has a smaller impact on the CV than a large negative one. We consider a gamma random walk proposal and, at the j-th iteration of the algorithm, given the previous value $\lambda^{(j-1)}$ of the chain, we simulate

$$\lambda^{(*)} \overset{d}{\sim} \mathcal{G}\left(\lambda^{(j-1),2}/\xi_\lambda, \xi_\lambda/\lambda^{(j-1)}\right),$$

where ξ_λ represents the scale of the random walk. The proposal value generated from this density is then accepted or rejected according to the acceptance ratio of the Metropolis-Hastings algorithm.

Sampling the latent variables

The acceptance log-probability $\rho\left((\boldsymbol{z}^{(j-1)}_{\tau:\tau+\delta}, \mathbf{CV}^{(j-1)}_{\tau:\tau+\delta}), (\boldsymbol{z}^{(*)}_{\tau:\tau+\delta}, \mathbf{CV}^{(*)}_{\tau:\tau+\delta})\right)$ is given by the minimum between one and the exponential transform of

$$
\begin{aligned}
&\log\left(\boldsymbol{\mathcal{L}}\left(r_{\tau+1:\tau+\delta}, \mathrm{RV}_{\tau+1:\tau+\delta}, \mathrm{CV}^{(*)}_{\tau+1:\tau+\delta}, z^{(*)}_{\tau+1:\tau+\delta} | \boldsymbol{x}_{\tau+1:\tau+\delta}, \boldsymbol{\theta}\right)\right) \\
&- \log\left(\boldsymbol{\mathcal{L}}\left(r_{\tau+1:\tau+\delta}, \mathrm{RV}_{\tau+1:\tau+\delta}, \mathrm{CV}^{(j-1)}_{\tau+1:\tau+\delta}, z^{(j-1)}_{\tau+1:\tau+\delta} | \boldsymbol{x}_{\tau+1:\tau+\delta}, \boldsymbol{\theta}\right)\right) \\
&+ \sum_{t=\tau+1}^{\tau+\delta-1} \log\left(p\left(z^{(j-1)}_t | z^{(j-1)}_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_t, \boldsymbol{\theta}\right)\right) \\
&+ \log\left(p\left(\mathrm{CV}^{(j-1)}_t | r_{p+1:T}, \mathrm{RV}_{p+1:T}, z^{(j-1)}_{p+1:T}, \boldsymbol{x_{p+1:T}}, \boldsymbol{\theta}\right)\right) \\
&- \log\left(p\left(z^{(*)}_t | z^{(*)}_{t-1}, r_{t-1}, \mathrm{RV}_{t-1}, \boldsymbol{x}_t, \boldsymbol{\theta}\right)\right) - \log\left(p\left(\mathrm{CV}^{(*)}_t | r_{p+1:T}, \mathrm{RV}_{p+1:T}, z^{(*)}_{p+1:T}, \boldsymbol{x_{p+1:T}}, \boldsymbol{\theta}\right)\right),
\end{aligned}
$$

A.7 Further details for the financial application

For the SV-ARG model, in order to get an appropriate guess for the initial value of $\boldsymbol{\theta} = \left(\mu, \gamma, \alpha, \kappa_2, \nu, c, \beta^{(d)}\right)$, we use the following argumentations. The risk-free rate μ is fixed to $\mu = (2.5/252)\%$. The market price of risk γ is set to 0.0. To fix the initial value of α, we estimate it with a rolling-window analysis over the period of estimation using as proxy $\hat{\alpha}$ introduced in Hansen and Lunde (2005). Precisely

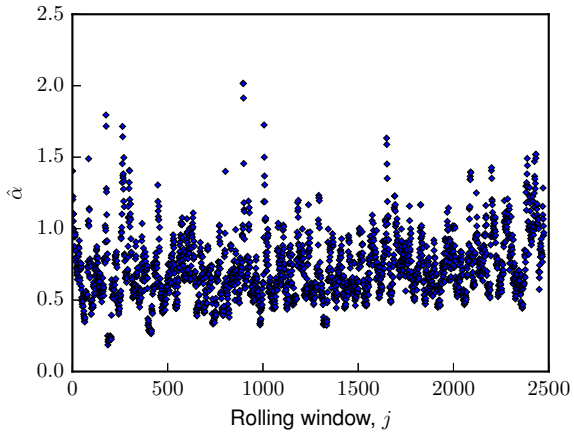
$$\hat{\alpha} = \frac{\sum_{t=1}^{m} (r_t - \bar{r})^2}{\sum_{t=1}^{m} \mathrm{RV}_t},$$

where r_t is the daily close-to-close return and $\bar{r}$ is its sample average over the m-days sample period. We set the rolling window size equals to $m = 20$ over the estimation period (from 8-Jan-1997 to 8-Jan-2007) and then we average the $T - m + 1$ estimates to fix the initial guess. Specifically $\alpha \approx 0.66$. The following Figure A.3 displays the result of the rolling window analysis.

Starting values for the parameters $\phi^{(d)}$, c (so for $\beta^{(d)}$) and ν are obtained by matching the unconditional mean, variance and persistence of average squared returns to the marginal invariant distribution of the ARG(1) process (see also Creal, 2015). Finally, initial value for κ_2 is set by matching the unconditional variance of our RV estimator to the marginal invariant distribution of the RV process. In order to fix the initial trajectory for $\mathbf{CV}_{1:T}$, we make use of the physical interpretation of the CV, setting it to $\mathbf{CV}_{1:T} = \hat{\alpha}\mathbf{RV}_{1:T}$. With this values for parameters and CV we simulate $\boldsymbol{z}_{1:T}$ according with the SV-ARG dynamics. The maximum number of states for the approximating Markov Chain of the process z_t is fixed to 300. As done in Creal (2015), the Feller condition, $\nu > 1$, as well as the stationary constraint, $0 < \beta^{(d)} c < 1$, are imposed throughout the estimation.

For the SV-LARG model, in order to get an appropriate guess for the initial value of $\boldsymbol{\theta} = \left(\mu, \gamma, \alpha, \beta_2, \nu, c, \beta^{(d)}, \alpha^{(d)}, \lambda\right)$, we use the following arguments. For parameters $\left\{\mu, \gamma, \alpha, \beta_2, c, \beta^{(d)}\right\}$ we use the same strategy as above. Instead, because of the distributional properties of the idiosyncratic component of the daily log-return, $\alpha^{(d)}$ and λ are set

Figure A.3: Rolling-window analysis for the initial guess of the parameter α



NOTES: Reported is the 20 days rolling window estimates of the parameter α over the estimation period, from 8-January-1997 to 8-January-2007, computed with the proxy introduced in Hansen and Lunde (2005). Average value $\alpha \approx 0.66$.

to $\alpha^{(d)} = (1/T)\sum_{t=1}^{T} \hat{CV}_t$ and $\lambda = 1/\sqrt{\alpha^{(d)}}$, being $\mathbf{CV}_{1:T} = \hat{\alpha}\mathbf{RV}_{1:T}$. The parameter ν is derived by variance targeting. $\boldsymbol{z}_{1:T}$ is simulated according with the SV-LARG dynamics. We fix $N = 300$, in order to have a meaningful comparison among models. Also in this case, both the Feller condition and the stationary constraint, $0 < c(\beta^{(d)} + \lambda^2\alpha^{(d)}) < 1$, are imposed throughout the estimation.

For the SV-HARG model, in order to get an appropriate guess for the initial value of $\boldsymbol{\theta} = \left(\mu, \gamma, \alpha, \beta_2, \nu, c, \left[\beta^{(d)}, \beta^{(w)}, \beta^{(m)}\right]\right)$, we set $\{\mu, \gamma, \alpha, \beta_2, c\}$ as for the SV-ARG and the autoregressive component to $\left[0.5\hat{\beta}^{(d)}, 0.3\hat{\beta}^{(d)}, 0.20\hat{\beta}^{(d)}\right]$, with $\hat{\beta}^{(d)}$ the SV-ARG estimate. The parameter ν is derived by variance targeting. Initial guesses for $\mathbf{CV}_{1:T}$ and $\boldsymbol{z}_{1:T}$ are obtained as before; $N = 300$.

Finally, a similar strategy is followed for the SV-LHARG. In particular, $\left[\alpha^{(d)}, \alpha^{(w)}, \alpha^{(m)}\right] = \left[0.5\hat{\alpha}^{(d)}, 0.3\hat{\alpha}^{(d)}, 0.20\hat{\alpha}^{(d)}\right]$, being $\hat{\alpha}^{(d)}$ the SV-LARG estimate.

Table A.1 reports efficiency indicators of the parameters estimation results for the four estimated models.

A.8 Further details on simulation experiments

Description of the experiments

To test the efficiency of the proposed MCMC algorithm we simulate 50 independent data series of 1000 observations from each of the considered models SV-ARG, SV-LARG, SV-HARG, SV-LHARG. On each set of data we run the proposed MCMC algorithm for 100,000 iterations, discard the first 20,000, and apply a thinning procedure with an appropriate factor. The posterior quantities reported in the subsequent sections are the grand averages over the 50 experiments. To fix the number of states of the approximating Markov chain for the process z_t we use the following strategy. Since we know the invariant unconditional mean of an ARG, LARG, HARG, LHARG model, then we approximate the unconditional distribution of z_t with a Poisson random variable z having constant intensity equals to the invariant unconditional mean, i.e. $z_t \approx z\ \forall t$. Finally, we determine N in such a way that $\mathbb{P}(z_t \geq N) \leq \delta$, with δ a constant as small as desired. As second

Table A.1: Summary output of the efficiency indicators of parameter estimates for the SV-LARG, SV-HARG, SV-LHARG model on S&P 500, 1997-2007

	PANEL A SV-ARG			PANEL B SV-LARG			PANEL C SV-HARG			PANEL D SV-LHARG		
$\boldsymbol{\theta}$	INEFF	(CD, P)	ACC	INEFF	(CD, P)	ACC	INEFF	(CD, P)	ACC	INEFF	(CD, P)	ACC
γ	1.15	(0.41, 0.34)	−−	1.86	(0.66, 0.26)	−−	1.78	(-0.89, 0.19)	−−	3.77	(0.42, 0.34)	−−
α	12.52	(0.16, 0.44)	0.18	7.69	(-0.23, 0.42)	0.32	19.26	(-0.22, 0.43)	0.37	4.25	(0.76, 0.22)	0.32
κ_2	2.42	(-0.75, 0.23)	0.39	12.23	(-0.25, 0.40)	0.32	4.88	(-0.84, 0.20)	0.33	15.20	(0.72, 0.24)	0.38
ν	−−	−−	−−	−−	−−	−−	−−	−−	−−			
c	36.76	(0.67, 0.25)	0.27	19.29	(0.77, 0.22)	0.24	18.52	(0.75, 0.23)	0.37	31.77	(-0.80,0.21)	0.51
$\beta^{(d)}$	37.69	(-1.53, 0.06)	0.18	21.10	(0.04, 0.48)	0.20	29.39	(0.90, 0.18)	0.18	26.22	(-0.95,0.10)	0.42
$\beta^{(w)}$	−−	−−	−−	−−	−−	−−	31.98	(0.92, 0.16)	−−	27.42	(-0.83,0.20)	−−
$\beta^{(m)}$	−−	−−	−−	−−	−−	−−	32.59	(0.85, 0.19)	−−	33.04	(-1.54,0.05)	−−
$\alpha^{(d)}$	−−	−−	−−	7.37	(-0.29, 0.39)	0.25	−−	−−	−−	31.09	(0.93, 0.13)	0.23
$\alpha^{(w)}$	−−	−−	−−	−−	−−	−−	−−	−−	−−	28.84	(0.84, 0.20)	−−
$\alpha^{(m)}$	−−	−−	−−	−−	−−	−−	−−	−−	−−	35.74	(1.51, 0.07)	−−
λ	−−	−−	−−	2.95	(0.21, 0.42)	0.29	−−	−−	−−	33.46	(-0.30, 0.38)	0.18

NOTES: Reported are efficiency indicators of the parameters estimation results for the SV-ARG (PANEL A), SV-LARG (PANEL B), SV-HARG (PANEL C), SV-LHARG (PANEL D on S&P 500 Futures, from 8-January-1997 to 8-January-2007. We ran the proposed MCMC algorithm for 15,000 iterations and then discard the first 5,000. A thinning procedure with a factor of 10 is applied. In each PANEL: inefficiency factor (IN.); convergence diagnostic statistic ((CD,P)); average acceptance rate of the Metropolis-Hastings algorithm steps (ACC).

Table A.2: True values for SV-LARG, SV-HARG, SV-LHARG parameters, simulated data

	Model		
Parameter	SV-LARG	SV-HARG	SV-LHARG
μ	0.0	0.0	0.0
γ	1.0	1.0	1.0
κ_1	0.0	0.0	0.0
α	0.8	0.8	0.8
κ_2	-1.0	-1.0	-1.0
ν	1.5	1.5	1.5
c	1.0	1.0	1.0
$\beta^{(d)}$	0.30	0.30	0.30
$\beta^{(w)}$	0.0	0.20	0.20
$\beta^{(m)}$	0.0	0.10	0.10
$\alpha^{(d)}$	0.15	0.0	0.15
$\alpha^{(w)}$	0.0	0.0	0.10
$\alpha^{(m)}$	0.0	0.0	0.05
λ	1.0	0.0	1.0

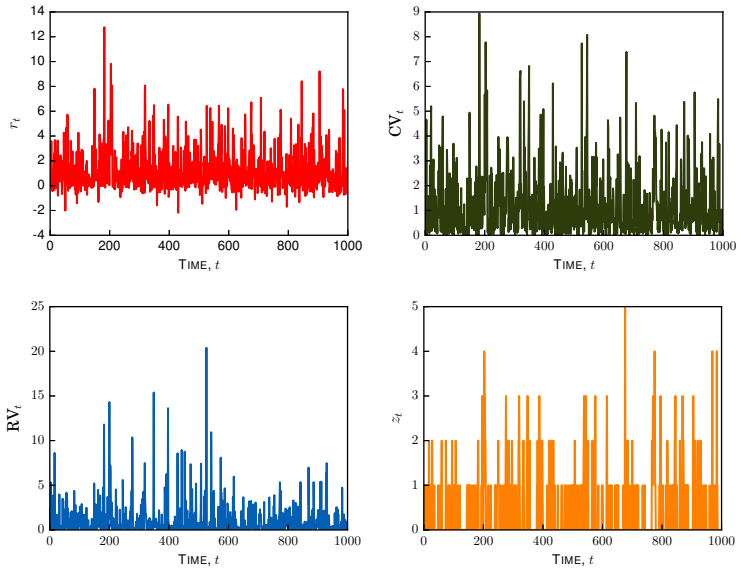
Notes: Reported are the values of the parameters used to simulate from the SV-LARG (*second column*, SV-HARG (*third column*, SV-LHARG (*fourth column*) in order to test the efficiency of the MCMC algorithm.

alternative, a model selection procedure can be used to select the optimal number of states given a set of observations. In our research we opt for the first approach.

We simulate from the SV-ARG by using the following values for the parameters: $\mu = 0.0$, $\gamma = 1.0$, $\kappa_1 = 0.0$, $\alpha = 0.8$, $\kappa_2 = -1.0$, $\nu = 0.8$, $c = 1.0$, $\beta^{(d)} \in \{0.3, 0.6, 0.9\}$. Besides, at each fixed time t, we use the following values for the exogenous variables: $x_{1t} = 0$ and $x_{2t} = 0.2\mathbf{1}_{\{t<0.5T\}} + 0.01\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$. We test the efficiency of the MCMC algorithm in three different scenarios: *low-persistence*, $\beta^{(d)} = 0.3$, *medium-persistence*, $\beta^{(d)} = 0.6$, and finally, *high persistence*, $\beta^{(d)} = 0.9$, scenario. The values for the parameters for the other models used in the simulation procedure are reported in the Table A.2

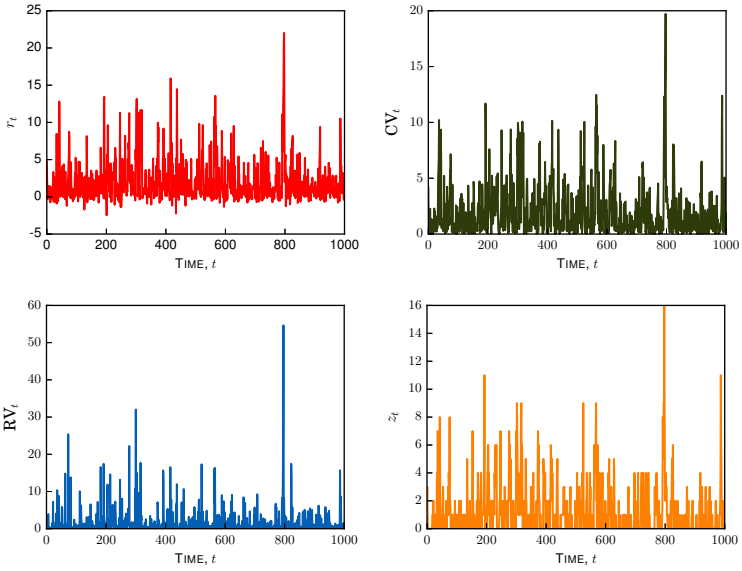
Besides, at each fixed time t, we use the following values for the exogenous variables: $x_{1t} = 0$ and $x_{2t} = 0.2\mathbf{1}_{\{t<0.5T\}} + 0.01\epsilon_t$ with $\epsilon_t \overset{d}{\sim} \mathcal{N}(0,1)$. Figures A.4, A.5 and A.6 represent simulated trajectories of a SV-ARG model in the *low persistence* (1), *medium-persistence* (2), *high-persistence* (3) scenario. Simulated trajectories of a SV-LARG, SV-HARG, SV-LHARG model are reported in Figure A.7, A.8, A.9 respectively.

Figure A.4: Simulated trajectories of a SV-ARG in the *low-persistence* scenario



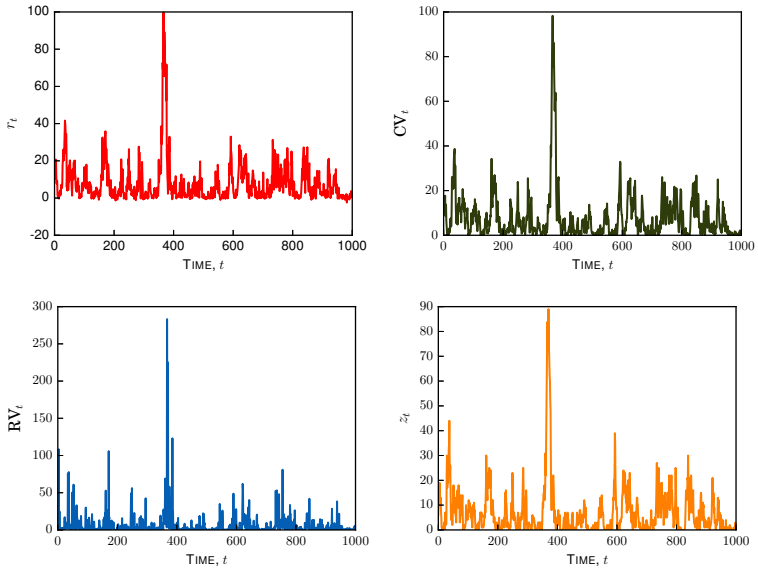
NOTES: Reported are simulated trajectories of a SV-ARG model in the *low-persistence* scenario. The values for the parameters are the following: $\mu = 0.0$, $\gamma = 1.0$, $\beta_1 = 0.0$, $\alpha = 0.8$, $\beta_2 = -1.0$, $\nu = 0.8$, $c = 1.0$, $\beta^{(d)} = 0.3$. *Top-left*: daily log-returns. *Top-right*: conditional variance of the daily log-returns. *Bottom-left*: realized variance. *Bottom-right*: state variable.

Figure A.5: Simulated trajectories of a SV-ARG in the *medium-persistence* scenario



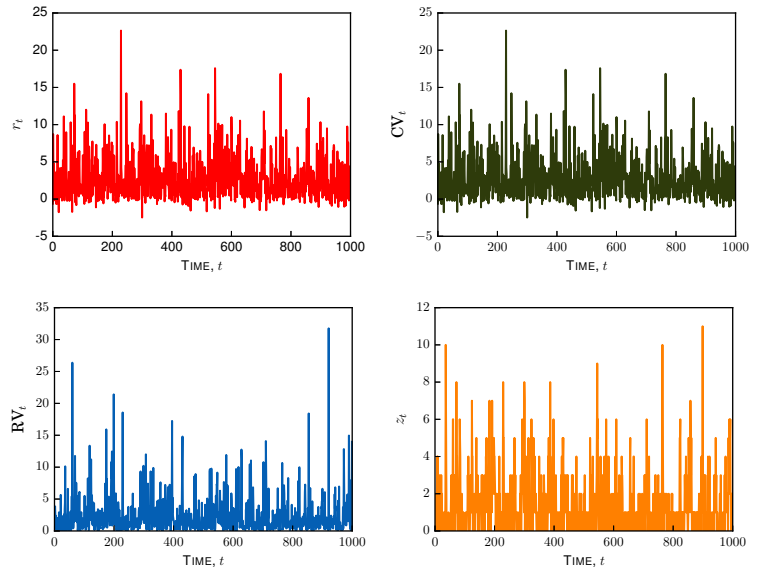
NOTES: Reported are simulated trajectories of a SV-ARG model in the *medium-persistence* scenario. The values for the parameters are the following: $\mu = 0.0$, $\gamma = 1.0$, $\beta_1 = 0.0$, $\alpha = 0.8$, $\beta_2 = -1.0$, $\nu = 0.8$, $c = 1.0$ and $\beta^{(d)} = 0.6$. *Top-left*: daily log-returns. *Top-right*: conditional variance of the daily log-returns. *Bottom-left*: realized variance. *Bottom-right*: state variable.

Figure A.6: Simulated trajectories of a SV-ARG in the *high-persistence* scenario



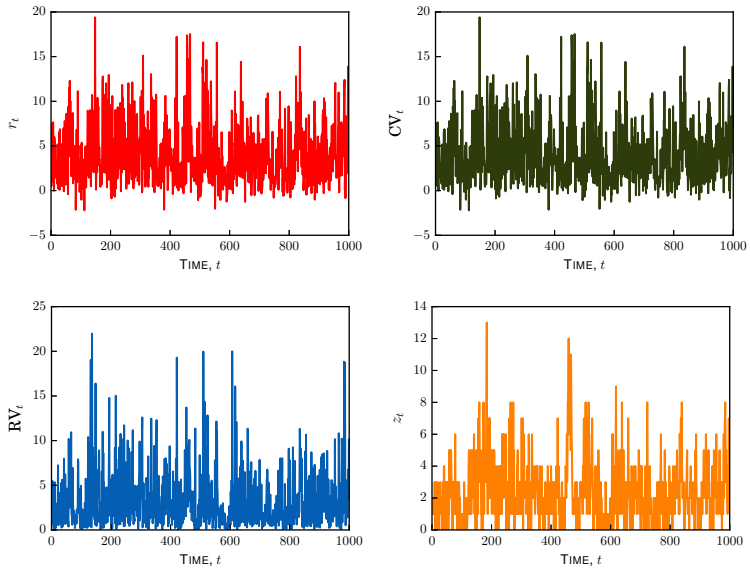
NOTES: Reported are simulated trajectories of a SV-ARG model in the *high-persistence* scenario. The values for the parameters are the following: $\mu = 0.0$, $\gamma = 1.0$, $\beta_1 = 0.0$, $\alpha = 0.8$, $\beta_2 = -1.0$, $\nu = 0.8$, $c = 1.0$ and $\beta^{(d)} = 0.9$. *Top-left*: daily log-returns. *Top-right*: conditional variance of the daily log-returns. *Bottom-left*: realized variance. *Bottom-right*: state variable.

Figure A.7: Simulated trajectories of a SV-LARG



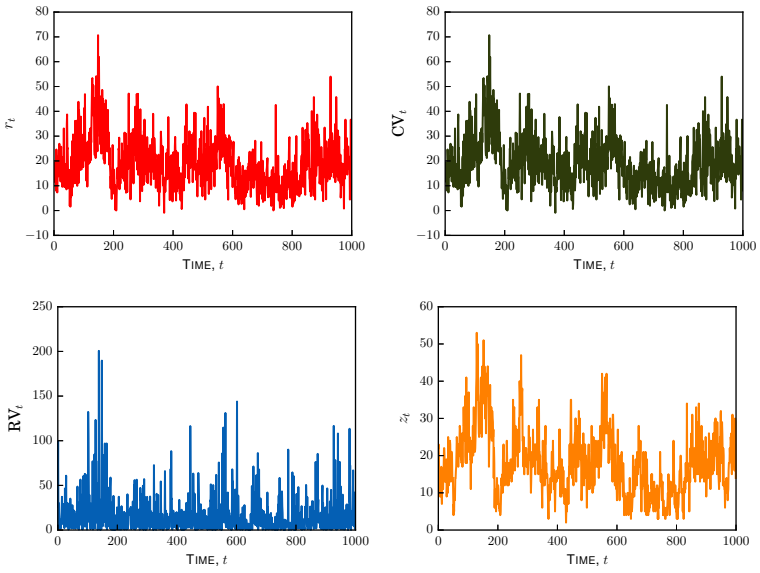
NOTES: Reported are simulated trajectories of a SV-LARG model. The values for the parameters are those in Table A.2, *first row*. *Top-left*: daily log-returns. *Top-right*: conditional variance of the daily log-returns. *Bottom-left*: realized variance. *Bottom-right*: state variable.

Figure A.8: Simulated trajectories of a SV-HARG



NOTES: Reported are simulated trajectories of a SV-HARG model. The values for the parameters are those in Table A.2, *first row*. *Top-left*: daily log-returns. *Top-right*: conditional variance of the daily log-returns. *Bottom-left*: realized variance. *Bottom-right*: state variable.

Figure A.9: Simulated trajectories of a SV-LHARG



NOTES: Reported are simulated trajectories of a SV-LHARG model. The values for the parameters are those in Table A.2, *first row*. *Top-left*: daily log-returns. *Top-right*: conditional variance of the daily log-returns. *Bottom-left*: realized variance. *Bottom-right*: state variable.

Parameter posterior approximation

We report additional evidences about the efficiency of the proposed MCMC algorithm through the *inefficiency factor* (INEFF) of the chain computed before and after that a thinning procedure with an appropriate factor is applied. For the SV-ARG we report also

the estimated Autocorrelation Function ACFs of the chain before and after the thinning procedure. In particular, Table A.3 contains the diagnostic summaries about the INEFF for the SV-ARG. Figures A.10-A.11, A.12-A.13, A.14-A.15, instead, contain the analysis of the ACFs(j), up to lag $j = 50$, of the MCMC output in the *low-persistence* (7-8), *medium-persistence* (9-10), *high-persistence* (11-12) scenario for the SV-ARG. More precisely, on the left-side of each figure we report the ACFs estimated on the raw output, whereas on the right-side the ACFs estimated after that the thinning procedure on the raw output is applied. Finally, Table A.4 reports the comparison between the INEFFs before and after the thinning for SV-LARG, SV-HARG and SV-LHARG.

Table A.3: Summary output of the INEFF estimates for 50 replications from the SV-ARG model in the low, medium and high persistence scenario

	Low Persistence		Medium Persistence		High Persistence	
$\boldsymbol{\theta}$	INEFF I	INEFF II	INEFF I	INEFF II	INEFF I	INEFF II
μ	2.41	1.13	2.73	1.19	3.05	1.25
γ	26.24	5.99	30.00	6.97	36.45	7.24
κ_1	2.36	1.10	2.26	1.10	2.76	1.20
α	35.06	7.81	35.16	7.97	37.24	7.23
κ_2	31.99	6.82	30.38	6.69	32.70	6.21
ν	18.27	1.67	14.30	1.37	13.16	1.16
c	27.58	4.94	31.10	6.15	37.35	6.94
$\beta^{(d)}$	23.45	2.81	29.33	5.07	37.13	6.94

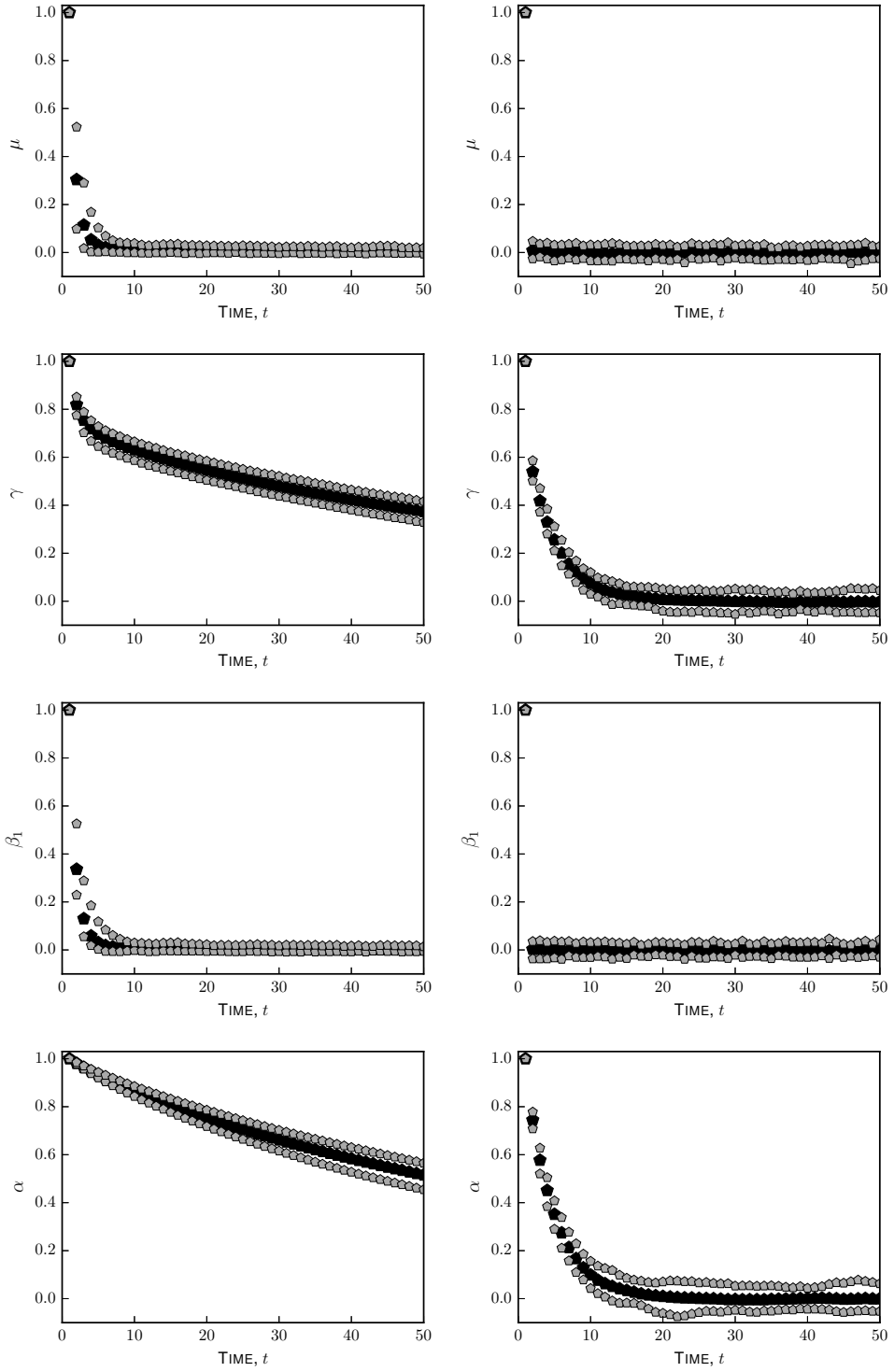
Notes: Reported are the *inefficiency factors* (INEFF) estimates for the three different scenarios of the SV-ARG model. The true values for the parameters are the following: $\mu = 0.0$, $\gamma = 1.0$, $\beta_1 = 0.0$, $\alpha = 0.8$, $\beta_2 = -1.0$, $\nu = 0.8$, $c = 1.0$, $\beta^{(d)} \in \{0.3, 0.6, 0.9\}$. Result are averages over a set of 50 independent MCMC experiments on 50 independent dataset on 1000 observations. On each dataset we ran the proposed MCMC algorithm for 100000 iterations and then discard the first 20000 iterations. We report for each setting (Low Persistence, Medium Persistence, and High Persistence) the INEFF estimated on: (i) the raw output (*left column*, INEFF I), (ii) the raw output when a thinning procedure with a factor of 20 (Low Persistence and Medium Persistence) or 50 (High Persistence) is applied (*right column*, INEFF II).

Table A.4: Summary output of the INEFF estimates for 50 replications from the SV-LARG, SV-HARG and SV-LHARG

	SV-LARG		SV-HARG		SV-LHARG	
$\boldsymbol{\theta}$	INEFF I	INEFF II	INEFF I	INEFF II	INEFF I	INEFF II
μ	11.66	5.71	19.36	5.77	1.01	1.11
γ	24.32	13.72	33.60	16.34	1.04	1.10
κ_1	14.04	9.51	11.03	1.67	0.99	1.07
α	29.79	16.84	35.55	17.05	14.31	1.23
κ_2	25.88	8.58	30.35	12.89	15.00	1.11
ν	24.58	7.04	32.33	13.10	22.26	1.29
c	25.78	13.72	35.81	17.44	18.97	1.17
$\beta^{(d)}$	34.41	8.44	34.90	12.51	29.22	2.66
$\beta^{(w)}$	−−	−−	36.01	11.90	35.35	7.22
$\beta^{(m)}$	−−	−−	39.59	18.19	39.47	22.62
$\alpha^{(d)}$	35.81	11.09	−−	−−	26.42	1.79
$\alpha^{(w)}$	−−	−−	−−	−−	37.51	10.87
$\alpha^{(m)}$	−−	−−	−−	−−	39.77	22.57
λ	38.41	10.12	−−	−−	9.71	1.02

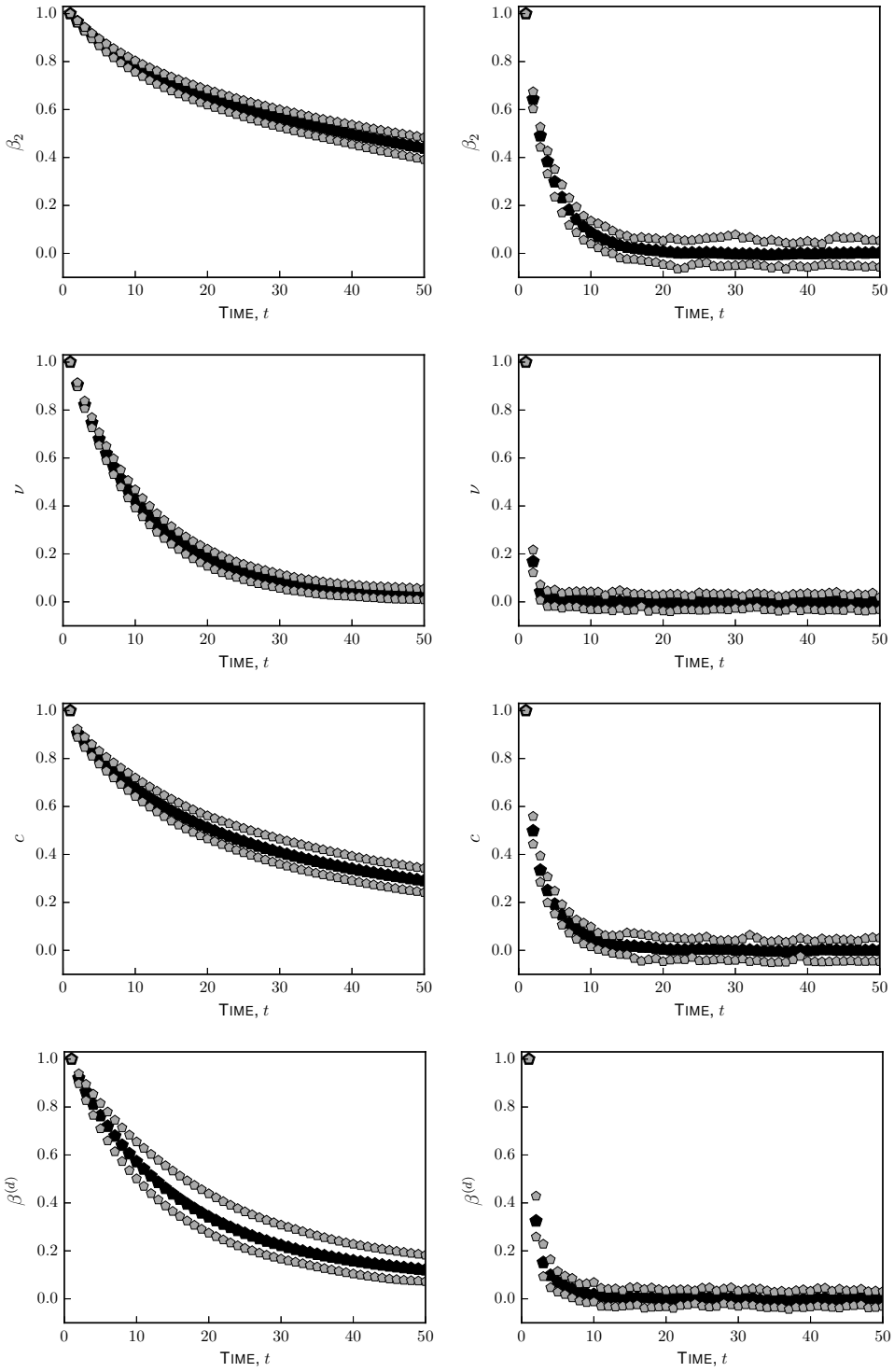
NOTES: Reported are the *inefficiency factors* (INEFF) estimates for the SV-LARG, SV-HARG, SV-LHARG. The true values for the parameters are reported in Table A.2. Results are averages over a set of 50 independent MCMC experiments on 50 independent dataset on 1000 observations. On each dataset we ran the proposed MCMC algorithm for 100000 iterations and then discard the first 20000 iterations. We report for each model the INEFF estimated on: (i) the raw output (*left column*, INEFF I), (ii) the raw output when a thinning procedure with a factor of 20 (LOW PERSISTENCE and MEDIUM PERSISTENCE) or 50 (HIGH PERSISTENCE) is applied (*right column*, INEFF II).

Figure A.10: ACFs of the MCMC output for the parameters μ, γ, κ_1, α of the SV-ARG model in the *low-persistence* scenario.



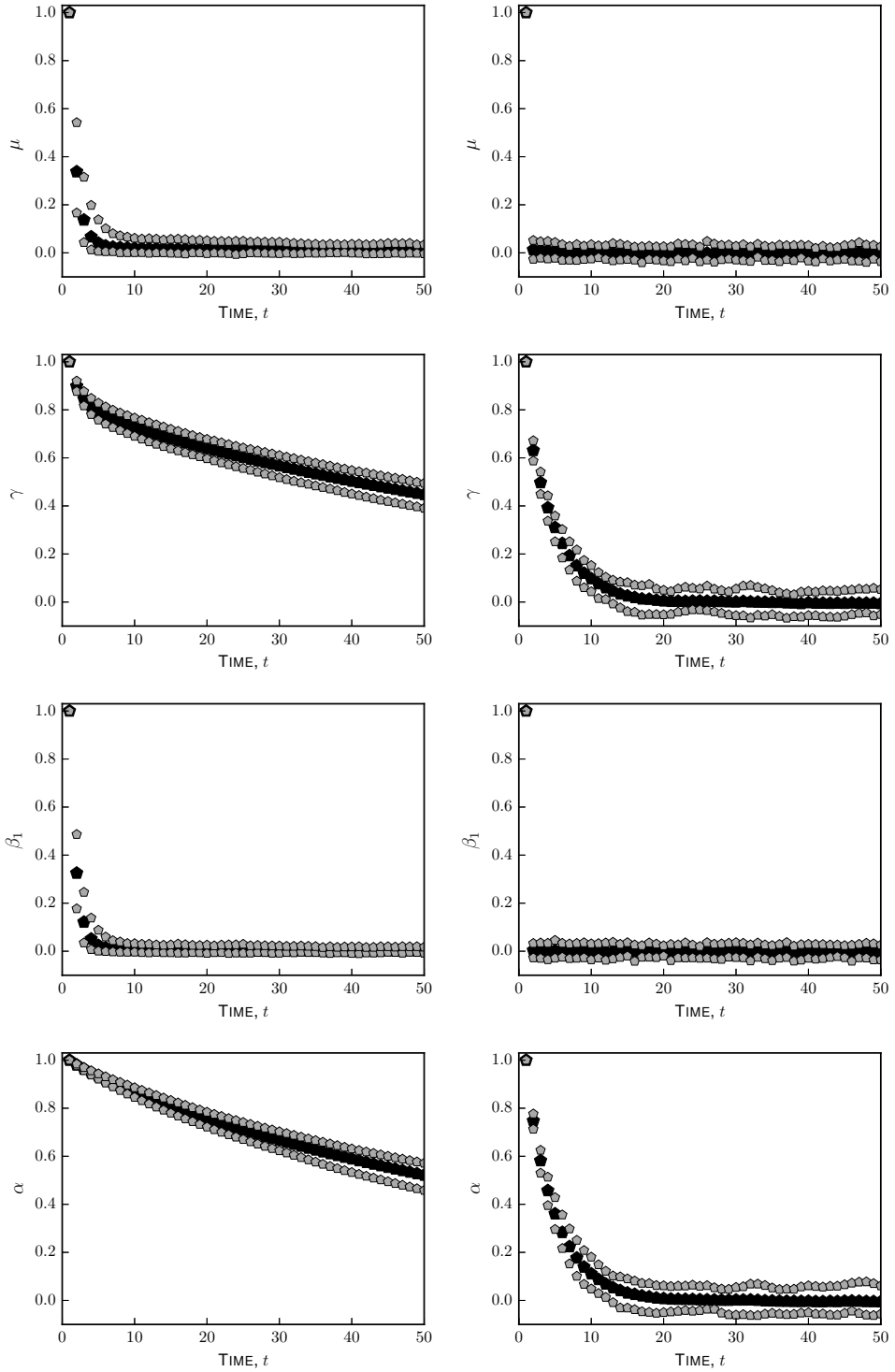
NOTES: Reported are the ACFs(j) up to lag $j = 50$ of the MCMC output for the parameters μ, γ, κ_1 and α of the SV-ARG model in the *low-persistence* scenario. For computational details see NOTES in Table A.3. *Left side:* ACFs estimated on the raw output. *Right side:* ACFs estimated when a thinning procedure with a factor of 20 is applied on the raw output.

Figure A.11: ACFs of the MCMC output for the parameters κ_2, ν, c, $\beta^{(d)}$ of the SV-ARG model in the *low-persistence* scenario.



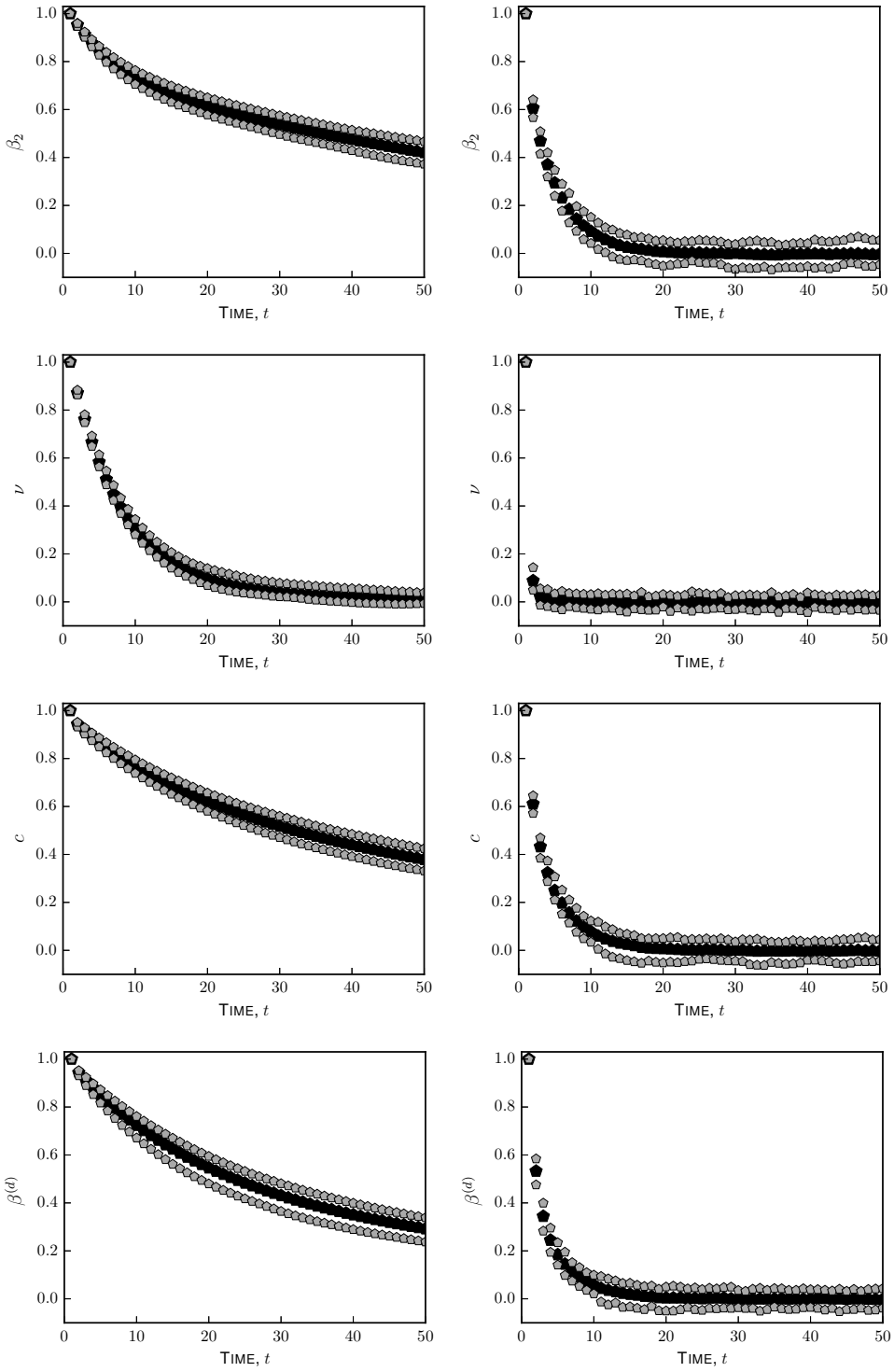
NOTES: Reported are the ACFs(j) up to lag $j = 50$ of the MCMC output for the parameters κ_2, ν, c, $\beta^{(d)}$ of the SV-ARG model in the *low-persistence* scenario. For computational details see NOTES in Table A.3. *Left side:* ACFs estimated on the raw output. *Right side:* ACFs estimated when a thinning procedure with a factor of 20 is applied on the raw output.

Figure A.12: ACFs of the MCMC output for the parameters μ, γ, κ_1, α of the SV-ARG model in the *medium-persistence* scenario.



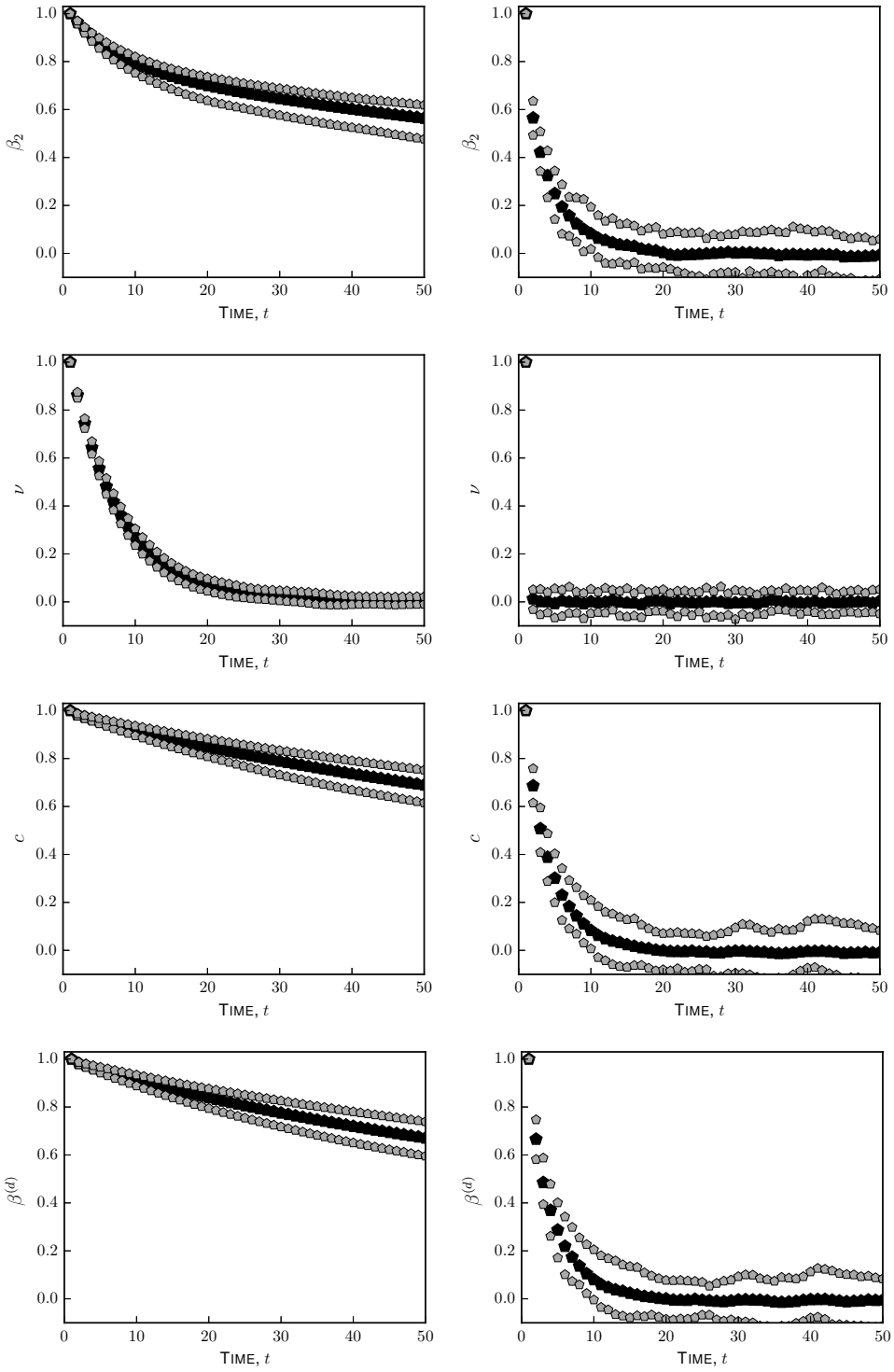
NOTES: Reported are the ACFs(j) up to lag $j = 50$ of the MCMC output for the parameters μ, γ, κ_1 and α of the SV-ARG model in the *medium-persistence* scenario. For computational details see NOTES in Table A.3. *Left side:* ACFs estimated on the raw output. *Right side:* ACFs estimated when a thinning procedure with a factor of 20 is applied on the raw output.

Figure A.13: ACFs of the MCMC output for the parameters κ_2, ν, c, $\beta^{(d)}$ of the SV-ARG model in the *medium-persistence* scenario.



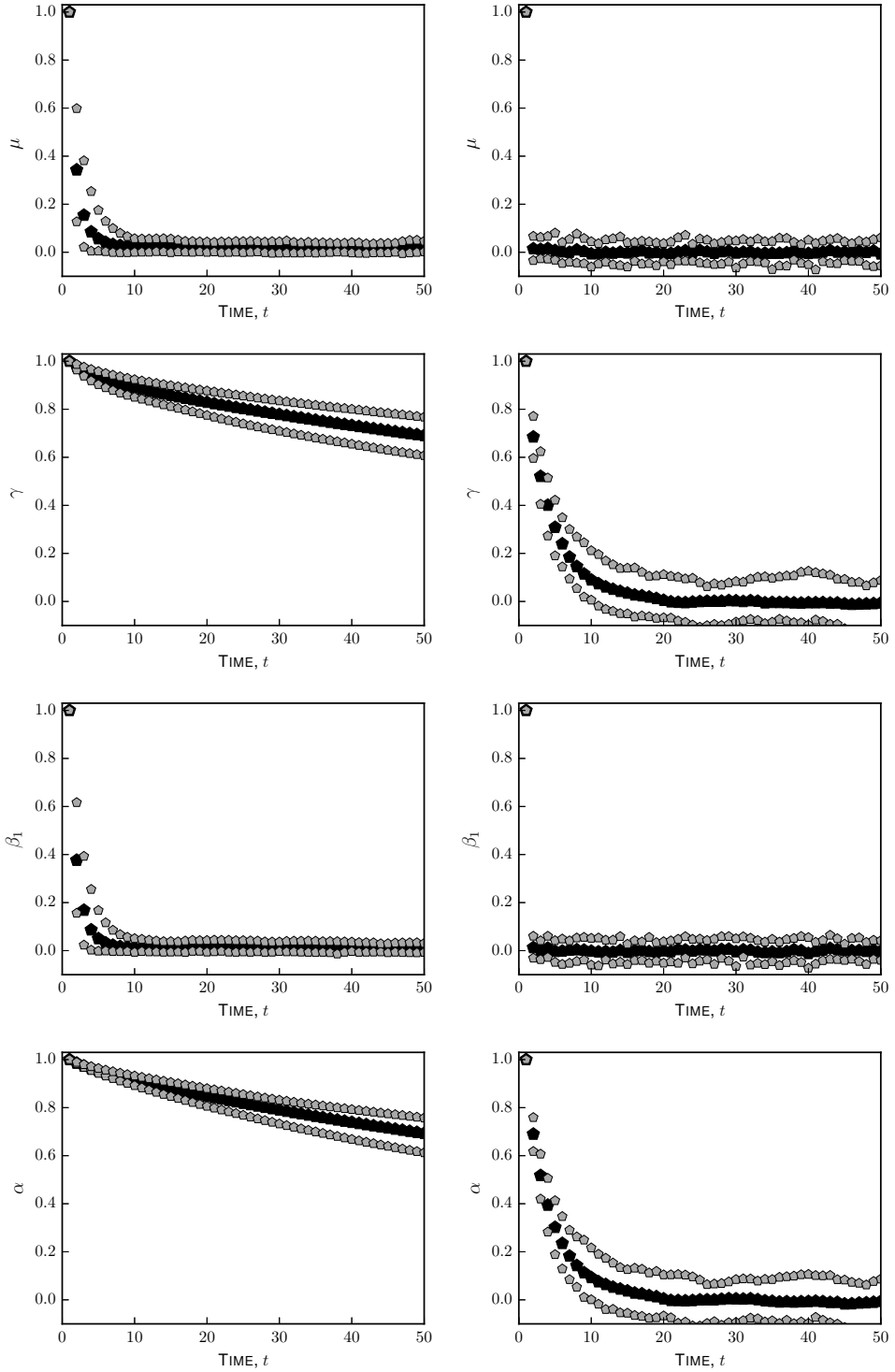
NOTES: Reported are the ACFs(j) up to lag $j = 50$ of the MCMC output for the parameters κ_2, ν, c, $\beta^{(d)}$ of the SV-ARG model in the *medium-persistence* scenario. For computational details see NOTES in Table A.3. *Left side:* ACFs estimated on the raw output. *Right side:* ACFs estimated when a thinning procedure with a factor of 20 is applied on the raw output.

Figure A.15: ACFs of the MCMC output for the parameters κ_2, ν, c, $\beta^{(d)}$ of the SV-ARG model in the *high-persistence* scenario.



NOTES: Reported are the ACFs(j) up to lag $j = 50$ of the MCMC output for the parameters κ_2, ν, c, $\beta^{(d)}$ of the SV-ARG model in the *high-persistence* scenario. For computational details see NOTES in Table A.3. *Left side:* ACFs estimated on the raw output. *Right side:* ACFs estimated when a thinning procedure with a factor of 50 is applied on the raw output.

Figure A.14: ACFs of the MCMC output for the parameters μ, γ, κ_1, α of the SV-ARG model in the *high-persistence* scenario.



NOTES: Reported are the ACFs(j) up to lag $j = 50$ of the MCMC output for the parameters μ, γ, κ_1 and α of the SV-ARG model in the *high-persistence* scenario. For computational details see NOTES in Table A.3. *Left side:* ACFs estimated on the raw output. *Right side:* ACFs estimated when a thinning procedure with a factor of 50 is applied on the raw output.

Chapter 5

A backward Monte Carlo approach to exotic option pricing

The material of this chapter is taken from Bormetti et al. (2017).

5.1 Introduction

Among the issues that a researcher has to solve when pricing a financial derivatives, two are the main ones. In the first place, the choice of a flexible model for the stochastic evolution of the underlying asset price. At this point, a common trade-off arises as models which describe the historical dynamics of the asset price with adequate realism are usually unable to precisely match volatility smiles observed in the option market (see, for instance Bouchaud and Potters, 2003; Gatheral, 2011). Secondly, once a reasonable candidate has been identified, there is the need to develop fast, accurate, and possibly flexible numerical methods (see Wilmott et al., 1993; Clewlow and Strickland, 1996; Hull, 2006) to price financial derivatives. As regards the former point, local volatility (LV) models have become very popular since their introduction by Dupire (see Dupire, 1994), and Derman and co-authors (see Derman and Kani, 1996). Even though the legitimate use of LV models for the description of the asset dynamics is highly questionable, the ability to self-consistently reproduce volatility smiles implied by the market motivates their widespread diffusion among practitioners. Since calibration *à la Dupire* (see Dupire, 1994) of LV models assumes the unrealistic availability of a continuum of vanilla option prices across different strikes and maturities (see Kahalé, 2004), recent years have seen the emergence of a growing strand of literature dealing with this problem (see, for instance Coleman et al., 1999; Kahalé, 2004; Andreasen and Huge, 2011; Lipton and Sepp, 2011; Reghai et al., 2012). However, in the present work we fix the calibration following the latest achievements and we solely focus on the latter issue. Specifically, our goal is to design a novel pricing algorithm based on the Monte Carlo approach able to achieve a sizeable variance reduction with respect to competitor approaches.

The main result of this work is the development of a flexible and efficient pricing algorithm – termed the *backward Monte Carlo algorithm* – which runs backward on top of a multinomial tree. The flexibility of this algorithm permits to price generic payoffs without the need of designing tailor-made solutions for each payoff specification. This feature is inherited directly from the Monte Carlo approach (see Glasserman, 2004, for an almost exhaustive survey of Monte Carlo methods in finance). The effectiveness, instead, is linked primarily to the backward movement on the multinomial tree. Indeed, our approach

combines both advantages of stratified sampling Monte Carlo and the Brownian Bridge construction (see, for instance Glasserman, 2004; Bormetti et al., 2006), extending them to more general financial-asset dynamics than the simplistic assumptions of Black, Scholes, and Merton (see Black and Scholes, 1973; Merton, 1973). The second purpose of this work – minor in relative terms with respect to the first one – is to investigate an alternative scheme for the implementation of the Recursive Marginal Quantization Algorithm (henceforth RMQA). The RMQA is a recursive algorithm which allows to approximate a continuous time diffusion by means of a discrete-time Markov Chain defined on a finite grid of points. The alternative scheme, employed at each step of the RMQA, is based on the Lloyd I method (see Kieffer, 1982) in combination with the Anderson acceleration Algorithm (see Anderson, 1965; Walker and Ni, 2011) developed to solve fixed-point problems. The accelerated scheme permits us to speed up the linear rate of convergence of the Lloyd I method (see Kieffer, 1982), and more importantly, to fix some flaws of previous RMQA implementations highlighted in Callegaro et al. (2015).

In more detail, a discrete-time Markov Chain approximation of the asset price dynamics can be achieved by introducing at each time step two quantities: (i) a grid for the possible values that the price can take, and (ii) the transition probabilities to propagate from one state to another one. Among the approaches discussed in the literature for computing these quantities, in the present work we analyse and extend two of them. The first approach quantizes via the RMQA the Euler-Maruyama approximation of the Stochastic Differential Equation (SDE) modelling the underlying asset price. The RMQA has been introduced in Pagés and Sagna (2015) to compute vanilla call and put options prices in a pseudo Constant Elasticity of Variance (CEV) LV model. In Callegaro et al. (2015) authors employ it to calibrate a Quadratic Normal LV model. The alternative approach, instead, discretises in an appropriate way the infinitesimal Markov generator of the underlying diffusion by means of a finite difference scheme (see Albanese and Mijatovic, 2007; Kushner and Dupuis, 2001, for a detailed discussion of theoretical convergence results). We name the latter approach Large Time Step Algorithm, henceforth LTSA. In Reghai et al. (2012) authors implement a modified version of the LTSA to price discrete look-back options in a CEV model, whereas in Albanese et al. (2009) they employ the LTSA idea to price a particular class of path-dependent pay-off termed Abelian pay-off. More specifically, they incorporate the path-dependency feature – in the specific case whether or not the underlying asset price hits a specified level over the life of the option – within the Markov generator. The joint transition probability matrix is then recovered as the solution of a pay-off specific matrix equation. The RMQA and LTSA present two major differences which can be summarized as follows: (i) the RMQA permits to recover the optimal – according to a specific criterion (see Printems et al., 2005) – multinomial grid, whereas the RMQA works with an *a priori* user-specified grid, (ii) the LTSA necessitates less computational burden than the RMQA when pricing financial derivatives products whose pay-off requires the observation of the underlying on a predefined finite set of dates. Unfortunately, this result holds only for a piecewise time-homogeneous LV dynamics.

The usage in both equity and foreign exchange (FX) markets of LV models is largely motivated by the flexibility of the approach which allows the exact calibration to the whole volatility surface. Moreover, the accurate re-pricing of plain vanilla instruments and of most liquid European options, together with the stable computation of the option sensitivity to model parameters and the availability of specific calibration procedures, make the LV modelling approach a popular choice. The LV models are also employed in practice to evaluate Asian options and other path-dependent options, although more sophisticated stochastic local volatility models are adopted when the path-dependency risk dominates

the smile risk (see, for instance Predota, 2005)[1]. We refer to (Ren et al., 2007) for more details on calibration and on pricing with LV models. The price of path-dependent derivative products is then usually computed either solving numerically a Partial Differential Equation (PDE) or via Monte Carlo methods. The PDE approach is computationally efficient but it requires the definition of a pay-off specific pricing equation (see Wilmott et al., 1993, for an extensive survey on PDE approach in financial contexts). Moreover, some options with exotic pay-offs and exercise rules are subtle to price even within the Black, Scholes, and Merton framework (Dewynne and Shaw, 2008; Siyanko, 2012). On the other hand, standard Monte Carlo method suffers from some ineffectiveness – especially when pricing out-of-the-money (OTM) options – since a relevant number of sampled paths does not contribute to the option pay-off. However, the Monte Carlo approach is extremely flexible and several numerical techniques have been introduced to reduce the variance of the Monte Carlo estimator (see Clewlow and Strickland, 1996; Glasserman, 2004). The backward Monte Carlo algorithm pursues this task.

In this work we consider the FX market, where we can trade spot and forward contracts along with vanilla and exotic options. In particular, we model the EUR/USD rate using a LV dynamics. The calibration procedure is the one employed in Reghai et al. (2012) for the equity market. Specifically, we calibrate the stochastic dynamics for the EUR/USD rate in order to reproduce the observed implied volatilities with a one basis point tolerance while the extrapolation to implied volatilities for maturities not quoted by the market is achieved by means of a piecewise time-homogeneous LV model. In order to show the competitive performances of the backward Monte Carlo algorithm, we compute the price of different kinds of options. We do not price basket options, but we only focus on derivatives written on a single underlying asset considering Asian calls, up-out barrier calls, and auto-callable options. We show that these instruments can be priced more effectively by simulating the discrete-time Markov Chain approximation of the diffusive dynamics from the maturity back to the initial date. In these cases, indeed, the backward Monte Carlo algorithm leads to a significant reduction of the Monte Carlo variance.

The remaining of the chapter is structured as follows. We start with the introduction of the key ideas of the backward Monte Carlo algorithm on a multinomial tree in Section 5.2. Section 5.3 presents the alternative schemes of implementation based on the RMQA and LTSA, and details the numerical investigations testing the performance of both approaches. Section 5.4 presents a piecewise time-homogeneous LV model for the FX market and reports the pricing performances of the backward Monte Carlo algorithm, benchmarking them with different Monte Carlo algorithms. We conclude in Section 5.5.

5.2 The backward Monte Carlo algorithm

First of all, let us introduce our working framework. We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a given time horizon $T > 0$ and a stochastic process $X = (X_t)_{t\in[0,T]}$ describing the evolution of the underlying asset price. We suppose that the market is complete, so that under the unique risk-neutral probability measure $\mathbb{Q}$, X has a Markovian dynamics described by the following SDE

$$\begin{cases} \mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t\,, \\ X_0 = x_0 \in \mathbb{R}\,, \end{cases} \tag{5.1}$$

[1]We clarify that the algorithms defined in this Chapter are presented for sake of simplicity in a LV framework in view of studying stochastic local volatility models.

where $(W_t)_{t\in[0,T]}$ is a standard one-dimensional $\mathbb{Q}$-Brownian motion, and $b : [0,T]\times\mathbb{R} \to \mathbb{R}$ and $\sigma : [0,T] \times \mathbb{R} \to \mathbb{R}_+$ are two measurable functions satisfying the usual conditions ensuring the existence and uniqueness of a (strong) solution to the SDE (5.1). Besides, we consider deterministic interest rates. Henceforth, we will always work under the risk-neutral probability measure $\mathbb{Q}$, since we focus on the pricing of derivative securities written on X. Specifically, we are interested in pricing financial derivative products whose pay-off may depend on the whole path followed by the underlying asset, i.e. path-dependent options.

Let us now motivate the introduction of our novel pricing algorithm. Even in the classical Black, Scholes, and Merton (see Black and Scholes, 1973; Merton, 1973) framework, when pricing financial derivatives the actual analytical tractability is limited to plain vanilla call and put options and to few other cases (for instance, see the discussion in Hui et al., 2000; Vecer and Xu, 2004). Such circumstances motivate the quest for general and reliable pricing algorithms able to handle more complex contingent claims in more realistic stochastic market models. In this respect the Monte Carlo (MC) approach represents a natural candidate. Nevertheless, a general purpose implementation of the MC method is known to suffer from a low rate of convergence. In particular, in order to increase its numerical accuracy, it is either necessary to draw a large number of paths or to implement tailor-made variance reduction techniques. Moreover, the standard MC estimator is strongly ineffective when considering OTM options, since a relevant fraction of sampled paths does not contribute to the pay-off function. For these reasons, in this Chapter, we present a novel MC methodology which allows to effectively reduce the variance of the estimated price. To this end, we proceed as follows. First we introduce a discrete-time and discrete-space process $\widehat{\bar{X}}$ approximating the continuous time (and space) process X in Equation (5.1). Then, we propose a MC approach – the backward Monte Carlo algorithm – to sample paths from $\widehat{\bar{X}}$ and to compute derivative prices.

In particular, we first split the time interval $[0,T]$ into n equally-spaced subintervals $[t_k, t_{k+1}]$, $k \in \{0,\dots,n-1\}$, with $t_0 = 0$ and $t_n = T$ and we approximate the SDE in Equation (5.1) with an Euler-Maruyama scheme as follows:

$$\begin{cases} \bar{X}_{t_{k+1}} = \bar{X}_{t_k} + b(t_k, \bar{X}_{t_k})\Delta t + \sigma(t_k, \bar{X}_{t_k})\sqrt{\Delta t}\; Z_k\,, \\ \bar{X}_{t_0} = X_0 = x_0\,, \end{cases} \tag{5.2}$$

where $(Z_k)_{0\le k\le n-1}$ is a sequence of *i.i.d.* standard Normal random variables and $\Delta t \doteq t_{k+1} - t_k = T/n$. Then, we assume that $\forall k \in \{1,\dots,n\}$ each random variable $\bar{X}_{t_k}$ in Equation (5.2) can be approximated by a discrete random variable taking values in $\Gamma_k \doteq \{\gamma_1^k,\dots,\gamma_N^k\}$, whereas for t_0 we have $\Gamma_0 = \gamma^0 = x_0$. We denote by $\widehat{\bar{X}}_{t_k}$ the discrete-valued approximation of the random variable $\bar{X}_{t_k}$. In this way, we constrain the discrete-time Markov process $(\widehat{\bar{X}}_{t_k})_{1\le k\le n}$ to live on a multinomial tree. Notice that, by definition $|\Gamma_k| = N$, all $k \in \{1,\dots,n\}$. Nevertheless, this is not the most general setting. For instance, within the RMQA framework authors in Pagés and Sagna (2015) perform numerical experiments letting the number of points in the space discretisation grids vary over time. However, they underline how the complexity in the time varying case becomes higher as N increases, although the difference in the results is negligible.
In order to define our pricing algorithm, we need the transition probabilities from a node at time t_k to a node at time t_{k+1}, $k \in \{0,\dots,n-1\}$, so that in the next section we provide a detailed description of two different approaches to consistently approximate them. For the moment, we describe the backward Monte Carlo algorithm assuming the knowledge of both the multinomial tree $(\Gamma_k)_{0\le k\le n}$ and the transition probabilities.

As aforementioned, our final target is the computation at time t_0 of the fair price of a path-dependent option with maturity $T > 0$. We denote by F its general discounted pay-off function. In particular, it is a function of a finite number of discrete observations. We are not going to make precise F at this point, we only recall here that in this paper we will focus on Asian options, up-and-out barrier options and auto-callable options. According to the arbitrage pricing theory (see, for instance Björk, 2009), the price is given by the conditional expectation of the discounted pay-off under the risk-neutral measure $\mathbb{Q}$, given the information available at time t_0. By means of the Euler-Maruyama discretisation, we can approximate the option price as follows:

$$\mathbb{E}_{t_0}\left[\mathrm{F}(x_0, \bar{X}_{t_1}, \ldots, \bar{X}_{t_n})\right] = \int_{\mathbb{R}^n} \mathrm{F}(x_0, x_1, \ldots, x_n)\; p(x_0, x_1, \ldots, x_n)\; \mathrm{d}x_1 \cdots \mathrm{d}x_n\,,$$

where $p(x_0, x_1, \ldots, x_n)$ is the joint probability density function (PDF) of $(\bar{X}_0, \bar{X}_{t_1}, \ldots, \bar{X}_{t_n})$. The previous expression can be further approximated exploiting the process $\widehat{\bar{X}}$ and its discrete nature (recall that $\Gamma_k = \{\gamma_1^k, \ldots, \gamma_N^k\}$):

$$\begin{aligned}\mathbb{E}_{t_0}\left[\mathrm{F}(x_0, \bar{X}_{t_1}, \ldots, \bar{X}_{t_n})\right] &\simeq \mathbb{E}_{t_0}\left[\mathrm{F}(x_0, \widehat{\bar{X}}_{t_1}, \ldots, \widehat{\bar{X}}_{t_n})\right] \\ &= \sum_{i_1=1}^{N} \cdots \sum_{i_n=1}^{N} \mathrm{F}(x_0, \gamma_{i_1}^1 \ldots, \gamma_{i_n}^n)\; \mathbb{P}(x_0, \gamma_{i_1}^1, \ldots, \gamma_{i_n}^n),\end{aligned} \tag{5.3}$$

where

$$\mathbb{P}(x_0, \gamma_{i_1}^1, \ldots, \gamma_{i_n}^n) \doteq \mathbb{P}(\widehat{\bar{X}}_{t_0} = x_0, \widehat{\bar{X}}_{t_1} = \gamma_{i_1}^1, \ldots, \widehat{\bar{X}}_{t_n} = \gamma_{i_n}^n).$$

Exploiting the Markovian nature of $\widehat{\bar{X}}$ and using Bayes' theorem, we rewrite the right hand side of Equation (5.3) in the following, equivalent, way:

$$\sum_{i_1=1}^{N} \cdots \sum_{i_n=1}^{N} \mathrm{F}(x_0, \gamma_{i_1}^1, \ldots, \gamma_{i_n}^n)\; \mathbb{P}(\gamma_{i_1}^1 | x_0) \cdots \mathbb{P}(\gamma_{i_n}^n | \gamma_{i_{n-1}}^{n-1}), \tag{5.4}$$

where

$$\mathbb{P}(\gamma_{i_{k+1}}^{k+1} | \gamma_{i_k}^k) \doteq \mathbb{P}(\widehat{\bar{X}}_{t_{k+1}} = \gamma_{i_{k+1}}^{k+1} | \widehat{\bar{X}}_{t_k} = \gamma_{i_k}^k), \tag{5.5}$$

all $\gamma_{i_k}^k \in \Gamma_k$ and all $k \in \{1, \ldots, n-1\}$

In order to compute the expression in Equation (5.4), a straightforward application of the standard MC theory would require the simulation of N_{MC} paths all originating from x_0 at time $t_0 = 0$. The same aforementioned arguments about the lack of effectiveness of the MC estimator for the case of a continuum of state-spaces still hold for the discrete case. However, forcing each random variable $\bar{X}_{t_k}$, $1 \leq k \leq n$, to take at most N values leads in general to a reduction of the variance of the Monte Carlo estimator.

For each $t_k \in \{t_1, \ldots, t_{n-1}\}$ we denote by $\Pi^{k,k+1}$ the $(N \times N)$-dimensional matrix whose elements are the transition probabilities:

$$\Pi_{i,j}^{k,k+1} \doteq \mathbb{P}(\gamma_j^{k+1} | \gamma_i^k), \quad \gamma_i^k \in \Gamma_k,\ \gamma_j^{k+1} \in \Gamma_{k+1},\ \text{and } i, j \in \{1, \ldots, N\}$$

and analogously we define

$$\Pi_{\ell,m}^{k+1,k} \doteq \mathbb{P}(\gamma_m^k | \gamma_\ell^{k+1}), \quad \gamma_m^k \in \Gamma_k,\ \gamma_\ell^{k+1} \in \Gamma_{k+1},\ \text{and } l, m \in \{1, \ldots, N\}.$$

The key idea behind the backward Monte Carlo algorithm is to express $\Pi_{i,j}^{k+1,k}$ as a function of $\Pi_{j,i}^{k,k+1}$ by applying Bayes' theorem:

$$\Pi_{i,j}^{k+1,k} = \frac{\Pi_{j,i}^{k,k+1}\, \mathrm{P}_j^k}{\mathrm{P}_i^{k+1}} \tag{5.6}$$

where, $\mathrm{P}_i^k \doteq \mathbb{P}(\widehat{\bar{X}}_{t_k} = \gamma_i^k | \widehat{\bar{X}}_{t_0} = x_0)$, for $i = 1, \dots, N$, $k = 1, \dots, n$.
Iteratively, we recover all the backward transition probabilities. These allow us to go through the multinomial tree in a backward way, from each of the terminal points γ_j^n to the initial node x_0. In particular, relation in Equation (5.6) permits us to rewrite the joint probability appearing in Equation (5.3) and then in Equation (5.4) as

$$\mathbb{P}(x_0, \gamma_{i_1}^1, \dots, \gamma_{i_n}^n) = \mathbb{P}(\gamma_{i_1}^1 | x_0) \cdots \mathbb{P}(\gamma_{i_n}^n | \gamma_{i_{n-1}}^{n-1}) = \left(\prod_{k=0}^{n-1} \Pi_{i_{k+1},i_k}^{k+1,k} \right) \mathrm{P}_{i_n}^n.$$

Consistently, we obtain the following proposition, containing the core of our pricing algorithm:

Proposition 5.2.1. *The price at time t_0 of a path-dependent option with discounted pay-off* F*,*
$\mathbb{E}_{t_0}\left[\mathrm{F}(x_0, \bar{X}_{t_1}, \dots, \bar{X}_{t_n})\right]$*, can be approximated by:*

$$\begin{aligned}\mathbb{E}_{t_0}\left[\mathrm{F}(x_0, \widehat{\bar{X}}_{t_1}, \dots, \widehat{\bar{X}}_{t_n})\right] &= \sum_{i_n=1}^{N} \mathrm{P}_{i_n}^n \sum_{i_1=1}^{N} \cdots \sum_{i_{n-1}=1}^{N} \left(\prod_{k=0}^{n-1} \Pi_{i_{k+1},i_k}^{k+1,k} \right) \mathrm{F}(x_0, \gamma_{i_1}^1, \dots, \gamma_{i_n}^n) \\ &\doteq \sum_{i_n=1}^{N} \mathrm{P}_{i_n}^n \mathcal{F}(x_0, \gamma_{i_n}^n),\end{aligned}$$

where $\mathcal{F}(x_0, \gamma_{i_n}^n)$ is the expectation of the pay-off function F *with respect to all paths starting at x_0 and terminating at $\gamma_{i_n}^n$.*

The expectation $\mathcal{F}(x_0, \gamma_{i_n}^n)$ can be computed sampling $N_{MC}^{i_n}$ MC paths from the conditional law of $(\widehat{\bar{X}}_{t_1}, \dots, \widehat{\bar{X}}_{t_{n-1}})$ given x_0 and $\gamma_{i_n}^n$, thus obtaining, at the same time, the error σ_{i_n} associated to the MC estimator. By virtue of the Central Limit Theorem the errors scale with the square root of $N_{MC}^{i_n}$, so that the larger $N_{MC}^{i_n}$ is, the smaller the error. In particular, if we indicate with $\widetilde{\Gamma}_n = \{\tilde{\gamma}_1^n, \dots, \tilde{\gamma}_{N^+}^n\}$, with $N^+ \leq N$, those points of Γ_n for which the pay-off F is different from zero, we can estimate the boundary values corresponding to the 95% confidence interval for the derivative price as

$$\sum_{i_n=1}^{N^+} \mathrm{P}_{i_n}^n \widehat{\mathcal{F}(x_0, \tilde{\gamma}_{i_n}^n)} \pm 1.96 \sqrt{\sum_{i_n=1}^{N^+} (\mathrm{P}_{i_n}^n \sigma_{i_n})^2}. \tag{5.7}$$

In the previous Equation $\widehat{\mathcal{F}(x_0, \tilde{\gamma}_{i_n}^n)}$ corresponds to the Monte Carlo estimator of $\mathcal{F}(x_0, \tilde{\gamma}_{i_n}^n)$. It is worth noticing that the error in Equation (5.7) does not take into account the effect of the finiteness of $\widetilde{\Gamma}_n$ and of the time discretisation. The sensitivity of the price to the finite size N, of the grids and to the choice of the boundary conditions are investigated in Appendices B.3 and B.4, respectively. We point out that the numerical results presented in this Chapter are computed with time and price grids such that the error due to the finite granularity is negligible with respect to the statistical error due to the finiteness of

the Monte Carlo sample.
A sizeable variance reduction results from having split the n sums in Equation (5.4) into the external summation over the points of the deterministic grid $\widetilde{\Gamma}_n$ and the evaluation of an expectation of the pay-off with fixed initial and terminal points. This procedure corresponds to the variance reduction technique known as stratified sampling Monte Carlo (see Glasserman, 2004). In particular, in Glasserman (2004) the author proves analytically that the variance of the MC estimator without stratification is always greater than or equal to that of the stratified one. Besides, he points out that stratified sampling involves consideration of two issues: (i) the choice of the points in Γ_n and the allocation $N_{MC}^{i_n}$, $i_n \in \{1, \dots, N\}$, (ii) the generation of samples from $\widehat{\widetilde{X}}$ conditional on $\widehat{\widetilde{X}}_{t_n} \in \Gamma_n$ and on $\widehat{\widetilde{X}}_{t_0} = \gamma_0$. Our procedure resolves both these points. Precisely, once selected $\widetilde{\Gamma}_n$, the backward Monte Carlo algorithm allows us to choose the number of paths from all the points in $\widetilde{\Gamma}_n$, independently on the value of $P_{i_n}^n$.

At this point, two are the main ingredients needed in order to compute the quantities in Equation (5.7): (i) the transition probabilities, (ii) the fast backward simulation of the process $\widehat{\widetilde{X}}$. For both purposes, we introduce ad-hoc numerical procedures. As regards the former point, we analyse and extend two approaches already present in the literature. The first one is based on the concept of optimal state-partitioning of a random variable (called stratification in Barraquand and Martineau (1995) and quantization in Bally et al. (2003)) and employs the RMQA (see Pagés and Sagna, 2015; Callegaro et al., 2015). The second approach provides a recipe to compute in an effective way the transition probability matrix between any two arbitrary dates for a piecewise time-homogeneous process (see Albanese, 2007; Reghai et al., 2012). More details on these two methods will be given in Section 5.3.

For what concerns the backward simulation, we employ the Alias method introduced in Kronmal and Peterson Jr (1979). More specifically, for every k from $n-1$ to 1, the (backward) simulation of $\widehat{\widetilde{X}}_{t_k}$ conditional on $\{\widehat{\widetilde{X}}_{t_{k+1}} = \gamma_j^{k+1}\}$ is equivalent to sampling at each time t_{k+1} from a discrete non-uniform distribution with support Γ_k and probability mass function equal to the j-th row of $\Pi^{k+1,k}$. Given the discrete distribution, a naïve simulation scheme consists in drawing a uniform random number from the interval $[0, 1]$ and recursively search over the cumulative sum of the discrete probabilities. However, in this case the corresponding computational time grows linearly with the number N of states. The Alias method, instead, reduces this numerical complexity to $O(1)$ by cleverly pre-computing a table – the Alias table – of size N. We base our implementation on this method, which enables a large reduction of the MC computation time. A more detailed description of the Alias method can be found at `www.keithschwarz.com`.

5.3 Recovering the transition probabilities

We present the two approaches used for the approximation of the transition probabilities of a discrete-time Markov Chain. The RMQA is described and extended in Section 5.3.1. In particular, we first provide a brief overview on optimal quantization of a random variable, then we propose an alternative implementation of the RMQA. The LTSA is presented in Section 5.3.2, where we also provide a brief introduction on Markov processes and on Markov generators.

5.3.1 A quantization based algorithm

Optimal quantization

We present here the concept of optimal quantization of a random variable by emphasizing its practical features, without providing all the mathematical details behind it. A more extensive discussion can be found in Graf and Luschgy (2000); Pagès et al. (2004); Printems et al. (2005); Pagès (2014). Let $\bar{X}$ be a one-dimensional continuous random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}_{\bar{X}}$ the measure induced by it. The quantization of $\bar{X}$ consists in approximating it by a one-dimensional discrete random variable $\widehat{\bar{X}}$. In particular, this approximation is defined by means of a quantization function q_N of $\bar{X}$, that is to say $\widehat{\bar{X}} \doteq q_N(\bar{X})$, defined in such a way that $\widehat{\bar{X}}$ takes $N \in \mathbb{N}^+$ finitely many values in $\mathbb{R}$. The finite set of values for $\widehat{\bar{X}}$, denoted by $\Gamma \equiv \{\gamma_1, \dots, \gamma_N\}$, is the quantizer of $\bar{X}$, while the image of the function q_N is the related quantization. The components of Γ can be used as generator points of a Voronoi tessellation $\{C_i(\Gamma)\}_{i=1,\dots,N}$. In particular, one sets up the following tessellation with respect to the absolute value in $\mathbb{R}$

$$C_i(\Gamma) \subset \{\gamma \in \mathbb{R} : |\gamma - \gamma_i| = \min_{1 \le j \le N} |\gamma - \gamma_j|\},$$

and the associated quantization function q_N is defined as follows:

$$q_N(\bar{X}) = \sum_{i=1}^{N} \gamma_i \mathbb{1}_{C_i(\Gamma)}(\bar{X}).$$

Notice that in our setting, we are going to quantize the random variables $(\bar{X}_{t_k})_{0 \le k \le n}$ introduced in Equation (5.2).

Such a construction rigorously define a probabilistic setting for the random variable $\widehat{\bar{X}}$, by exploiting the probability measure induced by the continuous random variable $\bar{X}$. The approximation of $\bar{X}$ through $\widehat{\bar{X}}$ induces an error, whose L^2 version – called L^2-mean quantization error – is defined as

$$\|\bar{X} - q_N(\bar{X})\|_2 \doteq \sqrt{\mathbb{E}\left[\min_{1 \le i \le N} |\bar{X} - \gamma_i|^2\right]}. \tag{5.8}$$

The expected value in Equation (5.8) is computed with respect to the probability measure which characterizes the random variable $\bar{X}$. The purpose of the optimal quantization theory is finding a quantizer indicated by Γ^*, which minimizes the error in Equation (5.8) over all possible quantizers with size at most N. We remind that in one dimension the uniqueness of the optimal N quantizer is guaranteed if the distribution of $\bar{X}$ is absolutely continuous with a log-concave density function (see, for instance Pagès, 2014). From the theory (see Graf and Luschgy, 2000, among others) we know that the mean quantization error vanishes as the grid size N tends to infinity. Besides, its rate of convergence is ruled by Zador theorem. However, computationally, finding explicitly Γ^* can be a challenging task. This has motivated the introduction of sub-optimal criteria linked to the notion of stationary quantizer (see Pagès, 2014):

Definition 5.3.1. *A quantizer $\Gamma = \{\gamma_1, \dots, \gamma_N\}$ inducing the quantization q_N of the random variable $\bar{X}$ is said to be stationary if*

$$\mathbb{E}\left[\bar{X} | q_N(\bar{X})\right] = q_N(\bar{X}). \tag{5.9}$$

Remark 2. *An optimal quantizer is stationary, the vice-versa does not hold true in general (see Pagès, 2014).*

In order to compute optimal (or sub-optimal) quantizers, one first introduces a notion of distance between a random variable $\bar{X}$ and a quantizer Γ as follows

$$d(\bar{X}, \Gamma) \doteq \min_{1 \leq i \leq N} |\bar{X} - \gamma_i| \,,$$

then one considers the so called distortion function

$$D(\Gamma) \doteq \mathbb{E}\left[d(\bar{X}, \Gamma)^2\right] = \mathbb{E}\left[\min_{1 \leq i \leq N} |\bar{X} - \gamma_i|^2\right] = \sum_{i=1}^{N} \int_{C_i(\Gamma)} |\xi - \gamma_i|^2 \, \mathrm{d}\mathbb{P}_{\bar{X}}(\xi) \,. \tag{5.10}$$

It can be shown (see Pagès, 2014), that the distortion function is continuously differentiable as a function of Γ. In particular, it turns out that stationary quantizers are critical points of the distortion function, that is, a stationary quantizer Γ is such that $\nabla D(\Gamma) = 0$. Several numerical approaches have been proposed in order to find stationary quantizers (for a review see Pagès et al., 2004). These approaches can be essentially divided into two categories: gradient-based methods and fixed-point methods. The former class includes the Newton-Raphson algorithm, whereas the second category includes the Lloyd I algorithm. More specifically, the Newton-Raphson algorithm requires the computation of the gradient, $\nabla D(\Gamma)$, and of the Hessian matrix, $\nabla^2 D(\Gamma)$, of the distortion function. The Lloyd I algorithm, on the other hand, does not require the computation of the gradient and Hessian and it consists in a fixed-point algorithm based on the stationary Equation (5.9).

The Recursive Marginal Quantization Algorithm

The RMQA is a recursive algorithm, which has been recently introduced in Pagés and Sagna (2015). It consists in quantizing the stochastic process X in Equation (5.1) by working on the (marginal) random variables $\bar{X}_{t_k}$, all $k \in \{1, \ldots, n\}$ in (5.2). The key idea behind the RMQA is that the discrete-time Markov process $\bar{X} = (\bar{X}_{t_k})_{0 \leq k \leq n}$ in Equation (5.2) is completely characterized by the initial distribution of $\bar{X}_{t_0}$ and by the transition probability densities. We indicate by $\widehat{\bar{X}}_{t_k}$ the quantization of the random variable $\bar{X}_{t_k}$ and by $\bar{D}(\Gamma_k)$ the associated distortion function.

Remark 3. *The process $\widehat{\bar{X}} = (\widehat{\bar{X}}_{t_k})_{0 \leq k \leq n}$ is not, in general, a discrete-time Markov Chain. Nevertheless, it is known (see, for instance Pagès et al., 2004) that there exists a discrete-time Markov Chain, $\widehat{\bar{X}}^c \doteq (\widehat{\bar{X}}^c_{t_k})_{0 \leq k \leq n}$, with initial distribution and transition probabilities equal to those of $\widehat{\bar{X}}$. Hence, throughout the rest of the paper, when we will write "discrete-time Markov Chain" within the Recursive Marginal Quantization framework we will refer, by tacit agreement, to the process $\widehat{\bar{X}}^c$.*

Here we give a quick drawing of the RMQA. First of all, one introduces the Euler operator associated to the Euler scheme in Equation (5.2):

$$\mathcal{E}_k(x, \Delta t; Z) \doteq x + b(t_k, x)\Delta t + \sigma(t_k, x)\sqrt{\Delta t}\; Z$$

where $Z \sim \mathcal{N}(0, 1)$, so that, from (5.2), $\bar{X}_{t_{k+1}} = \mathcal{E}_k(\bar{X}_{t_k}, \Delta t; Z_k)$.

Lemma 1. *Conditionally on the event $\{\bar{X}_{t_k} = x\}$, the random variable $\bar{X}_{t_{k+1}}$ is a Gaussian random variable with mean $m_k(x) = x + b(t_k, x)\Delta t$ and standard deviation $v_k(x) = \sqrt{\Delta t}\; \sigma(t_k, x)$, all $k = 1, \ldots, n-1$.*

Proof. It follows immediately from the equality $\bar{X}_{t_{k+1}} = \mathcal{E}_k(\bar{X}_{t_k}, \Delta t; Z_k)$, given that Z_k is a standard Normal random variable. □

At this point, one writes down the following crucial equalities:

$$\begin{aligned} \bar{D}(\Gamma_{k+1}) &= \mathbb{E}\left[d(\bar{X}_{t_{k+1}}, \Gamma_{k+1})^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[d(\bar{X}_{t_{k+1}}, \Gamma_{k+1})^2 | \bar{X}_{t_k}\right]\right] \\ &= \mathbb{E}\left[d(\mathcal{E}_k(\bar{X}_{t_k}, \Delta t; Z_k), \Gamma_{k+1})^2\right] , \end{aligned} \tag{5.11}$$

where $(Z_k)_{0\leq k\leq n}$ is a sequence of *i.i.d.* one-dimensional standard Normal random variables. As said, stationary quantizers are zeros of the gradient of the distortion function. By definition, the distortion function $\bar{D}(\Gamma_{k+1})$ depends on the distribution of $\bar{X}_{t_{k+1}}$, which is, in general, unknown. Nevertheless, thanks to Lemma 1, the distortion in Equation (5.11) can be computed explicitly. Equation (5.11) is the essence of the RMQA. More precisely, one starts setting the quantization of $\bar{X}_{t_0}$ to x_0, namely $q_N(\bar{X}_{t_0}) = x_0$. Then, one approximates $\bar{X}_{t_1}$ with $\widetilde{X}_{t_1} \doteq \mathcal{E}_0(x_0, \Delta t; Z_1)$ and the distortion function associated to $\bar{X}_{t_1}$ with that associated to $\widetilde{X}_{t_1}$, namely $\bar{D}(\Gamma_1) \approx \widetilde{D}(\Gamma_1) \doteq \mathbb{E}\left[d(\mathcal{E}_0(x_0, \Delta t; Z_1), \Gamma_1)^2\right]$. Then, one looks for a stationary quantizer Γ_1 by searching for a zero of the gradient of the distortion function, using either Newton-Raphson or Lloyd I method. The procedure is applied iteratively at each time step t_k, $1 \leq k \leq n$, leading to the following sequence of stationary (marginal) quantizers:

$$\begin{aligned} &\widehat{\widetilde{X}}_{t_0} \doteq \bar{X}_{t_0} , \\ &\widehat{\widetilde{X}}_{t_k} = q_N(\widetilde{X}_{t_k}) \quad \text{and} \quad \widetilde{X}_{t_{k+1}} = \mathcal{E}_k(\widehat{\widetilde{X}}_{t_k}, \Delta t; Z_{k+1}) , \\ &(Z_k)_{1\leq k\leq n} \quad \textit{i.i.d.} \text{ Normal random variables independent from } \bar{X}_{t_0} . \end{aligned}$$

In Pagés and Sagna (2015) the authors give an estimation of the (quadratic) error bound $\|\bar{X}_{t_k} - \widehat{\widetilde{X}}_{t_k}\|_2$, for fixed $k = 1, \ldots, n$.
At this point, the approximated transition probabilities (termed companion parameters in Pagés and Sagna, 2015) are obtained instantaneously given the quantization grids and Lemma 1. In particular:

$$\Pi_{i,j}^{k,k+1} = \mathbb{P}(\gamma_j^{k+1} | \gamma_i^k) \approx \mathbb{P}\left(\widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}) | \widetilde{X}_{t_k} \in C_i(\Gamma_k)\right) . \tag{5.12}$$

In Appendix B.1 we provide the explicit expressions of the distortion function $\widetilde{D}(\Gamma_{k+1})$ and of the approximated transition probabilities $\mathbb{P}\left(\widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}) | \widetilde{X}_{t_k} \in C_i(\Gamma_k)\right)$.

In order to compute numerically the sequence of stationary quantizers $(\Gamma_k)_{1\leq k\leq n}$ in Pagés and Sagna (2015) and Callegaro et al. (2015) authors employ the Newton-Raphson algorithm. However, as pointed out in Callegaro et al. (2015), it may become unstable when $\Delta t \to 0$ due to the ill-condition number of the Hessian matrix $\nabla^2 D(\Gamma)$. An alternative approach is based on fixed-point algorithms, such as the Lloyd I method, even though such method converges to the optimal solution with a smaller rate of convergence (see Kieffer, 1982, for a discussion). For these reasons and as original contribution we combine it with a particular acceleration scheme, called Anderson acceleration.

The Anderson accelerated procedure

The acceleration scheme, called *Anderson acceleration*, was originally discussed in Anderson (1965), and outlined in Walker and Ni (2011) together with some practical considerations for implementations. For completeness, in Appendix B.2 we give some details

on how the Lloyd I method works when employed in the recursive marginal quantization setting.

Now, we discuss the major differences between a general fixed-point algorithm – and its associated fixed-point iterations – and the same fixed-point method coupled with the Anderson acceleration. We outline the practical features without giving all the technical details concerning the numerical implementation of the accelerated scheme (please refer to Walker and Ni, 2011, for an extensive discussion on this issue). A general fixed-point problem – also known as Picard problem – and its associated fixed-point iteration are defined as follows:

$$\begin{aligned}
&\textbf{Fixed-point problem}: \text{Given } g:\mathbb{R}^N \to \mathbb{R}^N, \quad \text{find } \Gamma \in \mathbb{R}^N \text{ s.t. } \quad \Gamma = g(\Gamma).\\
&\qquad \textbf{\textit{Algorithm}} \textbf{ (Fixed Point Iteration)}\\
&\qquad\qquad \text{Given } \quad \Gamma^0,\\
&\qquad\qquad \text{for } l \geq 0, l \in \mathbb{N}\\
&\qquad\qquad\qquad \text{set } \Gamma^{l+1} = g(\Gamma^l)\,.
\end{aligned} \tag{5.13}$$

The same problem coupled with the Anderson acceleration scheme is modified as follows:

$$\begin{aligned}
&\textbf{\textit{Algorithm}} \textbf{ (Anderson acceleration)}\\
&\text{Given } \Gamma^0 \text{ and } m \geq 1\ , m \in \mathbb{N},\\
&\text{set } \Gamma^1 = g(\Gamma^0),\\
&\text{for } l \geq 1, l \in \mathbb{N}\\
&\qquad \text{set } m_l = \min(m, l)\\
&\qquad \text{set } F_l = (f_{l-m_l}, \ldots, f_l), \text{ where } f_i = g(\Gamma^i) - \Gamma^i\\
&\qquad \text{determine } \alpha^{(l)} = (\alpha_0^{(l)}, \ldots, \alpha_{m_l}^{(l)})^T \text{ that solves}\\
&\qquad\qquad \min_{\alpha^{(l)} \in \mathbb{R}^{m_l+1}} \|F_l \alpha^{(l)}\|_2 \text{ s.t. } \sum_{i=0}^{m_l} \alpha_i^{(l)} = 1\\
&\qquad \text{set } \Gamma^{l+1} = \sum_{i=0}^{m_l} \alpha_i^{(l)} g(\Gamma_{l-m_l+i}).
\end{aligned} \tag{5.14}$$

The Anderson acceleration algorithm stores (at most) m user-specified previous function evaluations and computes the new iterate as a linear combination of those evaluations with coefficients minimising the Euclidean norm of the weighted residuals. In particular, with respect to the general fixed-point iteration, Anderson acceleration exploits more information in order to find the new iterate.

In Equation (5.14) Anderson acceleration algorithm allows to monitor the conditioning of the least squares problem. In particular, we follow the strategy used in Walker and Ni (2011) where the constrained least squares problem is first transformed in an unconstrained one, then solved using a QR decomposition. The usage of the QR decomposition to solve the unconstrained least square problem represents a good balance of accuracy and efficiency. Indeed, if we name $\mathcal{F}_l$ the least squares problem matrix, it is obtained from its predecessor $\mathcal{F}_{l-1}$ by adding a new column on the right. The QR decomposition of $\mathcal{F}_l$ can be efficiently attained from that of $\mathcal{F}_{l-1}$ in $O(m_l N)$ arithmetic operations using standard QR factor-updating techniques (see Golub and Van Loan, 2012). The Anderson acceleration scheme speeds up the linear rate of convergence of the general fixed-point problem without increasing its computational complexity. More importantly, it does not

suffer the extreme sensitivity of the Newton-Raphson method to the choice of the initial point (grid). We refer to the numerical experiments in Appendix B.3 for an illustration of both the improvement of the Anderson acceleration with respect to the convergence speed of the fixed point iteration and of the over-performance of Lloyd I method with respect to the stability of the Newton-Raphson algorithm. Appendix B.3 is by no means intended to be exhaustive, since it illustrates the performance of the Anderson acceleration algorithm in some examples.

5.3.2 The Large Time Step Algorithm

The LTSA is employed to recover the transition probability matrix associated to a time and space discretisation of a LV model. We start here by recalling some known results about Markov processes, that will be used in what follows. We work under the following assumption:

Assumption 5.3.1. *The asset price process X follows the dynamics in Equation* (5.1)*, where the drift and diffusion coefficients b and σ are piecewise-constant functions of time.*

Let us consider the Markov process X in Equation (5.1) and let us denote by $p(t^{'}, \gamma^{'}|t, \gamma)$, with $0 \leq t < t' \leq T$ and $\gamma, \gamma' \in \mathbb{R}$, the transition probability density from state γ at time t to state γ' at time t'. Under some non stringent assumptions, it is known that p, as a function of the backward variables t and γ, satisfies the backward Kolmogorov equation (see Kijima, 1997; Karatzas and Shreve, 2012):

$$\begin{aligned} &\frac{\partial p}{\partial t}(t', \gamma'|t, \gamma) + (\mathcal{L}p)(t', \gamma'|t, \gamma) = 0 \quad \text{for} \quad (t, \gamma) \in (0, t') \times \mathbb{R}\,, \\ &p(t, \gamma'|t, \gamma) = \delta(\gamma - \gamma^{'})\,, \end{aligned} \tag{5.15}$$

where δ is the Dirac delta and $\mathcal{L}$ is the infinitesimal operator associated with the SDE (5.1), namely a second order differential operator acting on functions $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ belonging to the class $C^{1,2}$ and defined as follows:

$$(\mathcal{L}f)(t, \gamma) = b(t, \gamma)\frac{\partial f}{\partial \gamma}(t, \gamma) + \frac{1}{2}\sigma^2(t, \gamma)\frac{\partial^2 f}{\partial \gamma^2}(t, \gamma)\,. \tag{5.16}$$

The solution to Equation (5.15) can be formally written as

$$p(t^{'}, \gamma^{'}|t, \gamma) = \mathrm{e}^{(t^{'}-t)\mathcal{L}}p(t, \gamma). \tag{5.17}$$

The LTSA consists in approximating the transition probabilities relative to a discrete-time finite-state Markov chain approximation of X using Equation (5.17). We report now a simple example to clarify how the LTSA works.

Example 1. *Consider for example the case when b and σ in Equation* (5.1) *are defined as:*

$$\begin{aligned} b(t, X_t) &= b_1(X_t)\mathbb{1}_{[0,T_1]}(t) + b_2(X_t)\mathbb{1}_{[T_1,T_2]}(t)\,, \\ \sigma(t, X_t) &= \sigma_1(X_t)\mathbb{1}_{[0,T_1]}(t) + \sigma_2(X_t)\mathbb{1}_{[T_1,T_2]}(t)\,, \end{aligned}$$

where T_1 and $T_2 = T$ are two target maturities and $b_1, b_2, \sigma_1, \sigma_2$ suitable functions. The transition probabilities in this case are explicitly given. In particular, if we denote by $\mathcal{L}^1_\Gamma$ and $\mathcal{L}^2_\Gamma$ the infinitesimal Markov generators of the Markov chain approximation of X in

$[0, T_1]$ and $[T_1, T_2]$ respectively, the transition probabilities between any two arbitrary dates t and t' are given by:

$$\begin{aligned} &\mathrm{e}^{(t'-t)\mathcal{L}_\Gamma^1} && \text{for} \quad 0 \le t \le t' \le T_1\,, \\ &\mathrm{e}^{(T_1-t)\mathcal{L}_\Gamma^1}\mathrm{e}^{(t'-T_1)\mathcal{L}_\Gamma^2} && \text{for} \quad 0 \le t \le T_1 \le t' \le T_2\,, \\ &\mathrm{e}^{(t'-t)\mathcal{L}_\Gamma^2} && \text{for} \quad T_1 \le t \le t' \le T_2\,. \end{aligned}$$

In real market situations the above assumption on b and on σ is not at all restrictive, as we are going to see in Section 5.4.

Let us now give more details on the algorithm. First of all, once a time discretisation grid $\{u_0, u_1, \ldots, u_m\}$ has been chosen (think for example to the calibration pillars or to the expiry dates of the calibration dates of vanilla options), we need to obtain the space discretisation grids $\Gamma_k, 0 \le k \le m$. As a major difference with respect to the RMQA, the LTSA grids do not stem from the minimization of any distortion function. In particular, they can be defined with a large degree of arbitrariness (see for instance Appendix B.4) as follows:

$$\Gamma_0 \equiv x_0 \quad \text{and} \quad \Gamma_k \equiv \Gamma \doteq \{\gamma_1, \ldots, \gamma_N\}, \quad k = 1, \ldots, m\,,$$

with the only restriction that the grid $\Gamma_k \equiv \Gamma$ for all $k = 1, \ldots, m$. Actually, to ensure the applicability of the LTSA (see also Remark 4) it is essential to keep Γ_k constant within each time interval where the underlying process is time homogeneous. This apparent disadvantage of the LTSA is indeed a crucial positive feature when numerical simulations have to be implemented on Graphic Processing Units to exploit fast exponentiation.

Then, the method consists in discretising, opportunely, the Markov generator $\mathcal{L}$ and in calculating, in an effective and accurate way, the transition probabilities. As regards the discretisation, authors in Albanese and Mijatovic (2007) give a recipe to construct the discrete counterpart of $\mathcal{L}$ – denoted by $\mathcal{L}_\Gamma$ – so that the Markov chain approximation of X converges to the continuous limit process in a weak or distributional sense (see Kushner and Dupuis, 2001). In particular, $\mathcal{L}_\Gamma$ corresponds to the discretisation of Equation (5.16) through an explicit Euler finite difference approximation of the derivatives (see Mitchell and Griffiths, 1980). In Appendix B.4 we provide more details on the discretisation of $\mathcal{L}$.

Once $\mathcal{L}_\Gamma$ is constructed, one writes a (matrix) Kolmogorov equation for the transition probability matrix. In particular, using operator theory (Albanese, 2007), the transition probability matrix between any two arbitrary dates u_k and $u_{k'}$ with $0 \le u_k < u_{k'} \le T$ can be expressed as a matrix exponential.

Remark 4. *The piecewise time-homogeneous feature of the process X plays a crucial role as regards the computational burden required to compute the transition probability matrix. Indeed, in case of time-dependent drift and volatility coefficient, it can no longer be expressed, in a straightforward way, as the exponential of the (time-dependent) Markov generator $\mathcal{L}_\Gamma$ (see, for instance Blanes et al., 2009).*

The LTSA is computationally convenient with respect to the RMQA when pricing path-dependent derivatives whose pay-off specification requires the observation of the asset price on a pre-specified set of dates, for example, $\{u_0, u_1, \ldots, u_m\}$. Indeed, in this case we first calculate off-line the m transition matrices as in Equation (1), then we price the derivative products via MC with coarse-grained resolution. In Figure 5.1 we plot an example of a possible path corresponding to the case $m = 3$. This major difference between RMQA and LTSA becomes more evident looking at Figures 5.2 and 5.3, where we plot, respectively, a MC simulation connecting the initial point x_0 with a random final

Figure 5.1: Example of one Monte Carlo path sampled with the LTSA with $u_1 = t_1$, $u_2 = t_3$, and $u_3 = t_5$.

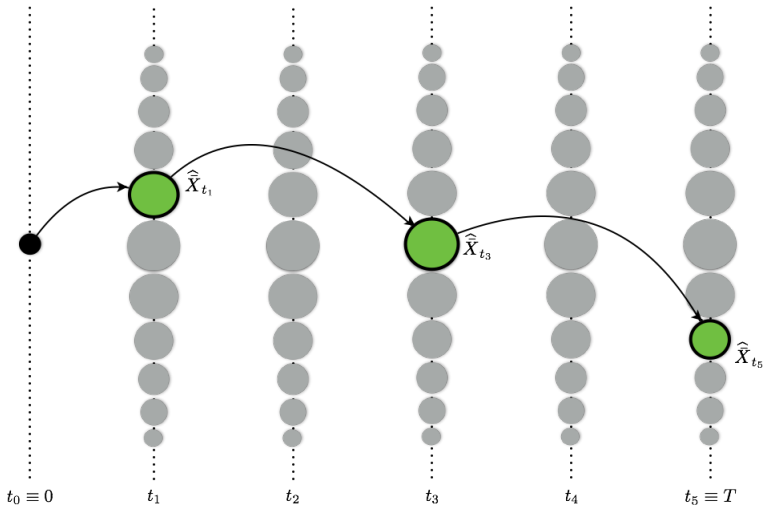
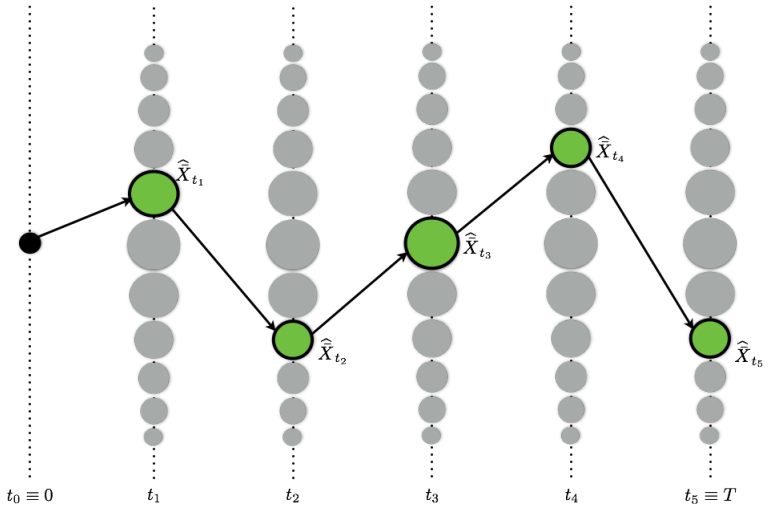


Figure 5.2: Example of one MC path sampled with RMQA over a time-grid computed with six time buckets.



point $\widehat{\widehat{X}}_{t_5}$, and a direct jump to date simulation to random points $\widehat{\widehat{X}}_{t_k}$ with $k = 1, \ldots, 5$, respectively.

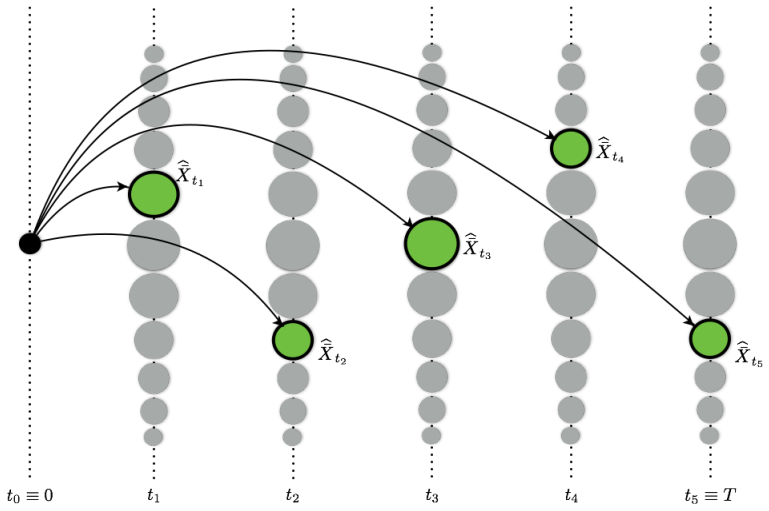
We underline that, both the RMQA and the LTSA enable the computation of the price of vanilla options by means of a straightforward scalar product. Indeed, the price of a vanilla call option with strike K and maturity $T = t_m$ can be computed as follows

$$\sum_{i=1}^{N} \mathbb{P}(\gamma_i^m | x_0)(\gamma_i^m - K)^+ . \tag{5.18}$$

LTSA implementation: More technical details

In order to compute effectively and accurately a matrix exponential and so recover the transition probability for the LTSA, we use the so called Scaling and Squaring method along with Padè approximation. In particular, we implement the version of the method proposed by Higham (2005) because it outperforms, both in efficiency and in accuracy,

Figure 5.3: Example of direct transitions from the initial point x_0 to random points $\widehat{\widehat{X}}_{t_k}$ with $k = 1, \ldots, 5$ computed with the RMQA.



previous implementations proposed in Sidje (1998) and Ward (1977). We now give a brief drawing of the algorithm implemented in Higham (2005), outlining its practical features but without giving all the mathematical details behind it. For a more extensive analysis on the method we refer to the original paper. Besides, we refer to Baker and Graves-Morris (1996) for an extensive description of Padè approximation.

The Scaling and Squaring algorithm exploits the following trivial property of the exponential function:

$$\mathrm{e}^{\widetilde{\mathcal{L}}_\Gamma} = \left(\mathrm{e}^{\frac{\widetilde{\mathcal{L}}_\Gamma}{\beta}}\right)^\beta , \tag{5.19}$$

where $(t_{\tilde{k}} - t_k)\mathcal{L}_\Gamma \doteq \widetilde{\mathcal{L}}_\Gamma$, together with the fact that $\mathrm{e}^{\widetilde{\mathcal{L}}_\Gamma}$ is well approximated by a Padè approximation near the origin, that is for small $\|\widetilde{\mathcal{L}}_\Gamma\|$, where $\|\cdot\|$ is any subordinate matrix norm. In particular, Padè approximation estimates $\mathrm{e}^{\widetilde{\mathcal{L}}_\Gamma}$ with the ratio between two (matrix) polynomials of degree 13. The mathematical elegance of the Padè approximation is enhanced by the fact that the two approximating polynomial are known explicitly.
The main hint of the Scaling and Squaring method is to choose β in Equation (5.19) as an integer power of 2, $\beta = 2^n$ say, so that $\widetilde{\mathcal{L}}_\Gamma/\beta$ has norm of order 1, to approximate $\mathrm{e}^{\widetilde{\mathcal{L}}_\Gamma/\beta}$ by a Padè approximation, and then to compute $\mathrm{e}^{\widetilde{\mathcal{L}}_\Gamma}$ by repeating squaring n times. In particular, we define $\delta t \doteq (t_{k'} - t_k)/2^n$. We use Padè approximation because of the usage of the explicit Euler Scheme for the discretisation of $\mathcal{L}$ (see Appendix B.4). Indeed, one needs to impose the so called Courant condition for the matrix $\delta t\mathcal{L}_\Gamma$. The Courant condition requires that $\|\delta\mathcal{L}_\Gamma\|_\infty < 1$. This translates into the following stringent condition for δt

$$\delta t < \frac{1}{2}\frac{(\Delta\gamma)^2}{\sigma(\gamma_i)}, \ \forall\, 1 \leq i \leq N\,.$$

The usage of the Padè approximation permits to relax the last constraint. In particular, the implementation in Higham (2005) allows $\|\delta\mathcal{L}_\Gamma\|_\infty$ to be much larger.

Remark 5. *The Backward Monte Carlo algorithm can be applied also to price financial derivatives in stochastic volatility models. We do not deal with this aspect in this paper, leaving it for future research. Indeed, the computational effort necessary to recover grids*

and transition probabilities increases substantially and to effectively compute matrix-matrix multiplications we need to switch to an alternative technology based on Graphic Processing Units.

5.4 Financial applications

In this final section we present and discuss how results achieved in the previous Sections 5.2 and 5.3 can be applied to finance, and in particular to option pricing in the FX market, where spot and forward contracts, along with vanilla and exotic options are traded (see Reiswich and Uwe, 2012, for a broad overview on FX market). In particular, we consider the following types of path-dependent options: (i) Asian calls, (ii) up-and-out barrier calls, (iii) automatic callable (or auto-callable). We choose two different models for the underlying EUR/USD FX rate: a LV model as a benchmark and the CEV (see Cox, 1975), coming from the academic literature.

5.4.1 Model, pay-off specifications and variance reduction techniques

Let us first introduce some notations relative to the LV model and to the CEV model. Recall that we have assumed in Section 5.2 deterministic interest rates. Moreover, we indicate by X_t the spot price at time t of one EUR expressed in USD and by $X_t(T)$ the corresponding forward price at time t for delivery at time T. We introduce the so-called normalized spot exchange rate $x_t^{LV} \doteq X_t/X_0(t)$, all $t \in [0,T]$ (where the superscript "LV" clearly stands for Local Volatility), and we suppose that the process $x^{LV} \doteq (x_t^{LV})_{t\in[0,T]}$ follows the SDE

$$\begin{cases} \mathrm{d}x_t^{LV} = x_t^{LV}\eta(t, x_t^{LV})\,\mathrm{d}W_t\,, \\ x_0^{LV} = 1\,. \end{cases}$$

Hence, in this LV model, x^{LV} corresponds to the underlying process X introduced in Section 5.2, and besides making reference to Equation (5.1) we have $b(t, x_t^{LV}) = 0$ and $\sigma(t, x_t^{LV}) = \eta(t, x_t^{LV})x_t^{LV}$, where $\eta : [0,T] \times \mathbb{R} \to \mathbb{R}_+$ corresponds to the local volatility function. Specifically, it is a cubic monotone spline for fixed $t \in [0,T]$ (see Fritsch and Carlson, 1980, for an overview on interpolation technique) with flat extrapolation. The set of points to be interpolated is determined numerically during the calibration procedure. Precisely, we use the calibration procedure proposed in Reghai et al. (2012). This procedure is particularly robust. Indeed, the resulting local volatility surface is ensured to be a smooth function of the spot. In particular, this procedure leads to a piecewise time-homogeneous dynamics for the process x^{LV}.

The data set used in the calibration is available upon requests.

As a second example we consider the CEV, i.e., we assume that the asset price process X follows a CEV dynamics

$$\begin{cases} \mathrm{d}X_t = rX_t\,\mathrm{d}t + \sigma X_t^{\alpha}\,\mathrm{d}W_t\,, \\ X_0 = x_0 \in \mathbb{R}_+\,, \end{cases} \tag{5.20}$$

where $r \doteq r_d - r_f \in \mathbb{R}$ is the risk-free interest rate with r_d the domestic interest rate and r_f the foreign interest rate, $\sigma \in \mathbb{R}_+$ is the volatility, and $\alpha \in [1/2, 1)$ is a constant parameter. Even though coefficients in Eq. (5.20) do not satisfy the usual Lipschitz conditions assumed in Section 5.2, it is well known that in our setting the SDE (5.20) admits a unique strong solution. Moreover, Jeanblanc et al. (2009) (see Lemma 6.4.4.1) show that – for the above choice of the parameter alpha – the solution is non-negative, in the

sense that zero is an absorbing state. As far as the discretisation scheme is concerned, in order to prevent $\bar{X}_{t_{k+1}}$ to take negative values, the usual Euler-Maruyama recipe has to be slightly adjusted. Labbé et al. (2011) review various possible modifications discussed in the literature. For the general CEV model under consideration here, a non-negative definite Euler-Maruyama scheme and its weak and strong convergence has been studied in Bossy and Diop (2007) and Berkaoui et al. (2008), respectively.

Then, given a time discretisation grid $\{0 = t_0, t_1, \ldots, t_n = T\}$ on $[0, T]$ as in Section 5.2 and making reference to the Euler-Maruyama scheme in Equation (5.2) we consider the unidimensional pay-off specifications below. In particular, we compute the price at time $t_0 = 0$.

i) **Asian calls.** The discounted pay-off function of a discretely monitored Asian call option is

$$\mathrm{F_A}(\bar{X}_0, \ldots, \bar{X}_{t_n}) \doteq e^{-r(t_n - t_0)} \max\left(\frac{1}{n+1}\sum_{i=0}^{n} \bar{X}_{t_i} - K, 0\right),$$

where K is the strike price and $T > 0$ the maturity. As benchmark price for an Asian call, we consider the price computed with control variates (see, for instance Glasserman, 2004, Chapter 4). Specifically, under the risk-neutral measure the discounted asset price is a martingale and we can employ it as the control variate variable. The comparison among the different results is postponed to the next section.

ii) **Up-and-out barrier calls.** We consider barrier options of European style. The discounted pay-off at maturity $T > 0$ of an up-and-out barrier call is given by

$$e^{-rT} \max(X_T - K, 0) \mathbb{1}_{\{\tau > T\}}, \tag{5.21}$$

where K is the strike price, $\tau \doteq \inf\{t \geq 0 : X_t \geq B\}$ and B is the upper barrier. The simplest method to price such an option is to consider the discrete version of the continuous time discounted pay-off in Equation (5.21)

$$\mathrm{F_B}(\bar{X}_{t_0}, \ldots, \bar{X}_{t_n}) \doteq \mathrm{e}^{-r(t_n - t_0)} \max(\bar{X}_{t_n} - K, 0) \prod_{k=0}^{n} \mathbb{1}_{\{\bar{X}_{t_k} < B\}}.$$

Proceeding this way, however, the price of the option is overestimated because X can cross the barrier at some time t between two grid points $k\Delta t$ and $(k+1)\Delta t$, $0 \leq k \leq n-1$, while being never above the barrier at the observation dates $t_0, t_1, \ldots t_n$. In Gobet (2000) (see, also Glasserman, 2004), the author proposes a strategy to obtain a better approximation of the price of the option in Equation (5.21) when employing MC simulation. It relies on the observation that the probability of the Brownian motion breaching the barrier is computable from the knowledge of the volatility and of the value of the spot X at the two end points of the observation interval. Namely, given the points $\bar{X}_{t_k}$ and $\bar{X}_{t_{k+1}}$, $0 \leq k \leq n-1$, the probability that X has reached the barrier B from below in (t_k, t_{k+1}) conditioned on the fact that it is below the barrier at both $\bar{X}_{t_k}$ and $\bar{X}_{t_{k+1}}$ is given by

$$\bar{p}_k \doteq 1 - \exp\left[-\frac{2}{\sigma_k \Delta t}(B - \bar{X}_{t_k})(B - \bar{X}_{t_{k+1}})\right],$$

where σ_k is the diffusive coefficient of the underlying asset price in (t_k, t_{k+1}) (assuming B is greater than both $\bar{X}_{t_k}$ and $\bar{X}_{t_{k+1}}$). So, at each $t_k = k\Delta t$, $1 \leq k \leq n$, and

for all the MC paths ℓ, $1 \leq \ell \leq N_{MC}$, after verifying that $\bar{X}_{t_k}^{(\ell)} < B$, we investigate whether the simulated path has reached the barrier B over the interval (t_{k-1}, t_k). In particular, we draw a random variable from a Bernoulli distribution with parameter $(1 - \bar{p}_{k-1})$: If the outcome is favorable then the barrier has been reached over this interval and the price associated to the path is zero.
The variance of the MC estimator can be further reduced by applying an importance sampling technique. Given the value of $\bar{X}_{t_{k-1}}$, each $\bar{X}_{t_k}$ is distributed as a Gaussian random variable with mean $\bar{\mu}_{t_k} \doteq \bar{X}_{t_{k-1}} + b(t_{k-1}, \bar{X}_{t_{k-1}})\Delta t$ and standard deviation $\bar{\sigma}_{t_k} \doteq \sigma(t_{k-1}, \bar{X}_{t_{k-1}})\sqrt{\Delta t}$. Then, one modifies this density so that values of $\bar{X}_{t_k}$ greater than B are not sampled. The procedure to draw a Gaussian below B is to sample a uniform random variable U and set $\bar{X}_{t_k} = F^{-1}_{\bar{X}_{t_k}}(\theta U)$, with $\theta = \mathbb{P}\left(\bar{X}_{t_k} < B\right)$. The final value is multiplied by θ. This algorithm is implemented in Joshi and Leung (2011) to price continuous barrier options in a jump-diffusion model.
As for the Asian option case, we postpone the comparison among the various techniques to Section 5.4.2.

iii) **Automatic callable (or auto-callable)** The discounted pay-off of an auto-callable option (they were first issued in the U.S. by BNP Paribas in August 2003 as cited for instance in Deng et al. (2011)) with unitary notional is given by

$$\begin{cases} \mathrm{e}^{-r(t_i^c - t_0)} Q_i & \text{if } \bar{X}_{t_j^c} < X_0 \mathrm{b} \leq \bar{X}_{t_i^c} \quad \text{for all } j < i\,, \\ \mathrm{e}^{-r(t_n - t_0)} \frac{\bar{X}_{t_n}}{X_0} & \text{if } \bar{X}_{t_i^c} < X_0 \mathrm{b} \quad \text{for all } i = 1, \ldots, m\,, \end{cases}$$

where $\{t_1^c, \ldots, t_m^c\}$ is a set of pre-fixed call dates, $\mathrm{b} > X_0$ is a pre-fixed barrier level, and $\{Q_1, \ldots, Q_m\}$ is a set of pre-fixed coupons. The set of call dates $\{t_1^c, \ldots, t_m^c\}$ does not coincide with the set of times of the Euler scheme discretisation $\{t_0, \ldots, t_n\}$. In particular, the latter has finer time resolution grid.

We show in Section 5.4.2 that all previous pay-offs can be priced efficiently by using our novel algorithm, i.e., by reverting the Monte Carlo paths and simulating them from maturity back to the initial date.

5.4.2 Numerical results and discussion

Let us introduce some terminology that we will use in the summary Tables of our numerical results. In particular, we will denote by: (i) *Euler Scheme* the prices obtained via a standard MC procedure on the process $\bar{X}$, (ii) *Euler Scheme C.V.* the prices obtained via a MC procedure on the process $\bar{X}$ when using the variance reduction technique Control Variates, (iii) *Euler Scheme I.S.* the prices obtained via a MC procedure on the process $\bar{X}$ when using the variance reduction technique Importance Sampling, (iv) *Forward RMQA* the prices obtained via a forward MC procedure on $\widehat{\bar{X}}$ from the starting date to the maturity, when grids and transition probabilities are computed through the RMQA, (v) *Forward LTSA* the prices obtained via a forward MC procedure on $\widehat{\bar{X}}$ from the starting date to the maturity, when grids and transition probabilities are computed through the LTSA, (vi) *Backward RMQA* the prices computed through the backward Monte Carlo algorithm on $\widehat{\bar{X}}$ when grids and transition probabilities are computed through the RMQA, (vii) *Backward LTSA* the prices computed through the backward Monte Carlo algorithm on $\widehat{\bar{X}}$ when grids and transition probabilities are computed through the LTSA, (viii) *Benchmark* the price

computed either as an *Euler Scheme* price (in case of Asian and up-and-out barrier call options) or as a *Forward LTSA* price (in case of auto-callable option) with an intensive Monte Carlo simulation in order to reduce the pricing error below the last significant digit. Besides, in brackets we will report the numerical error corresponding to one standard deviation.

Let us now stress some aspects related to the implementation of the backward Monte Carlo algorithm along with the procedures described in Sections 5.3.1 and 5.3.2.

In order to have a meaningful comparison between *Euler Scheme*-, *Forward*-, and *Backward*-type prices, for each of the N^+ points $\tilde{\gamma}^n_{i_n} \in \widetilde{\Gamma}_n$ we generate $N^{i_n}_{MC}$ random paths in such a way that $N^{i_n}_{MC} \times N^+ = N_{MC}$, where N_{MC} indicates the number of simulations employed to compute either *Euler Scheme*- or *Forward*-type prices. Besides, $\widetilde{\Gamma}_n \doteq \{\gamma^n_i \in \Gamma_n : \ K \leq \gamma^n_i \leq B\}$ when pricing up-and-out call barrier options and $\widetilde{\Gamma}_n = \Gamma_n$ when pricing In-The-Money (ITM), At-The-Money (ATM), Out-The-Money Asian call options and auto-callable options.

Concerning the granularity of the state-space, we fix it in such a way that the error on vanilla call option prices, computed as

$$|\sigma^{mkt} - \sigma^{alg}| \tag{5.22}$$

is less than or equal to 5 bps (recall that 1 bp $= 10^{-4}$). In Equation (5.22), σ^{mkt} is the market implied volatility, whereas σ^{alg} is the volatility implicitly obtained by using Equation (5.18). In particular we set the cardinality of the quantizers Γ_k, $1 \leq k \leq n$, to 100. We refer to Appendix B.3 and B.4 for precise considerations on this aspect.

The stopping criteria for the RMQA corresponds to $\|\Gamma^{l+1}_k - \Gamma^l_k\| \leq 10^{-5}$, $1 \leq k \leq n$, where Γ^l_k is the quantizer computed by the algorithm at time $t_k \in \{t_1, \ldots, t_n\}$ at the l-th iteration. Moreover, in the Backward Monte Carlo algorithm case, for each point in $\widetilde{\Gamma}_n$ we generate $N^{i_n}_{MC} = N_{MC}/|N^+| = 10^4/|\widetilde{\Gamma}_n|$ random paths. Let us now come to the discussion of the numerical results.

In Tables 5.1 and 5.2 we report up-and-out barrier call option prices for both LV and CEV, as well as their relative pricing errors. In order to test the performances of our algorithm we price ITM, ATM and OTM options. In particular, for both models the initial spot price is $\bar{X}_0 = 1.36$. This value corresponds to the value of the EUR/USD exchange rate as of date 23-June-2014. Besides, for both dynamics the value of the pair strike-barrier, (K, B), is set to $(1.35, 1.39)$, $(1.36, 1.39)$ and to $(1.37, 1.39)$ for ITM, ATM and OTM up-and-out call barrier options respectively. The maturity T is 6 months in the case of CEV model. In the case of LV model the maturity is $T = 1$ month, $T = 6$ months or $T = 12$ months. The number of Euler steps is $n = 9$, $n = 51$ and $n = 100$ for $T = 1$ month, $T = 6$ months or $T = 12$ months respectively. For the CEV model we fix $\alpha = 0.5$ and $r = r_d - r_f = 0.32\%$ (the latter corresponds to the value of the 6 months domestic interest rates implied by the forward USD curve at pricing date). As regards the parameter σ we vary it from $\sigma = 5\%$ to $\sigma = 20\%$ with steps $\Delta\sigma = 5\%$.

In Tables 5.1 and 5.2 we compare the Euler Scheme Monte Carlo price and error with those of the other methods for the LV and CEV dynamics, respectively. For both models the Backward Monte Carlo algorithm exhibits the best performances. To better visualize this fact Figures 5.4 and 5.5 report the ratio between the Euler Scheme MC error and that of the other techniques, henceforth *Error ratio*, for the LV and the CEV respectively. Some observations arise. The two methods employed to recover grids and transition probabilities lead to similar results. Actually, in the case of CEV model with $\sigma = 20\%$, the algorithm performs better when grids and probabilities are recovered through the LTSA. Figure 5.5

suggests that the gain in efficiency is more evident if we increase the value of the parameter σ. Intuitively, this happens because the probability for the price paths to hit the barrier B over the life of the option increases with σ. Moreover, for a fixed value of σ the gain in efficiency is more evident when pricing OTM options. This happens because for OTM options a relevant number of forward paths do not contribute to the pay-off. In order to increase the pricing accuracy of the Euler Scheme MC, it would be necessary to force paths to sample the region in which the pay-off is different from zero, namely between the strike K and the barrier B. Remind that in the *Euler Scheme* price we have already taken into account the possibility that the price path hits the barrier over a time interval (t_{k-1}, t_k), $1 \leq k \leq n$.

Table 5.1: Summary output of the results when pricing an up-and-out barrier call option in a LV dynamics

	Up-and-out barrier call		
Algorithm	**ITM**	**ATM**	**OTM**
		Local Volatility model	
$T = 1$ **month**			
Euler Scheme	$5.156E-3(7.3E-5)$	$2.316E-3(4.5E-5)$	$6.90E-4(2.1E-5)$
Euler Scheme I.S.	$5.089E-3(6.4E-5)$	$2.338E-3(4.1E-5)$	$6.95E-4(1.9E-5)$
Forward RMQA	$4.950E-3(6.4E-5)$	$2.073E-3(3.8E-5)$	$5.74E-4(1.6E-5)$
Forward LTSA	$4.929E-3(6.4E-5)$	$2.081E-3(3.8E-5)$	$5.82E-4(1.6E-5)$
Backward RMQA	$5.159E-3(2.7E-5)$	$2.291E-3(1.4E-5)$	$6.78E-4(5E-6)$
Backward LTSA	$5.118E-3(2.7E-5)$	$2.252E-3(1.4E-5)$	$6.65E-4(5E-6)$
Benchmark	$5.150E-3$	$2.308E-3$	$6.94E-4$
$T = 6$ **months**			
Euler Scheme	$1.089E-3(3.7E-5)$	$4.47E-4(1.9E-5)$	$1.37E-4(9E-6)$
Euler Scheme I.S.	$1.082E-3(3.2E-5)$	$4.40E-4(1.7E-5)$	$1.30E-4(8E-6)$
Forward RMQA	$9.90E-4(3.45-5)$	$4.11E-4(1.9E-5)$	$1.24E-4(3E-6)$
Forward LTSA	$9.92E-4(3.3E-5)$	$4.02E-4(1.8E-5)$	$1.16E-4(3E-6)$
Backward RMQA	$1.046E-3(1.7E-5)$	$4.59E-4(9E-6)$	$1.38E-4(8E-6)$
Backward LTSA	$1.053E-3(1.7E-5)$	$4.54E-4(9E-6)$	$1.38E-4(7E-6)$
Benchmark	$1.069E-3$	$4.56E-4$	$1.41E-4$
$T = 12$ **months**			
Euler Scheme	$3.01E-4(2.0E-5)$	$1.28E-4(1.1E-5)$	$3.84E-5(5.2E-6)$
Euler Scheme I.S.	$2.70E-4(1.6E-5)$	$1.37E-4(1.0E-5)$	$3.09E-5(3.8E-6)$
Forward RMQA	$2.91E-4(1.9E-5)$	$1.20E-4(1.05-5)$	$3.05E-5(3.5E-6)$
Forward LTSA	$2.81E-4(1.8E-5)$	$1.17E-4(1.0E-5)$	$3.27E-5(5.4E-6)$
Backward RMQA	$2.884-4(8E-6)$	$1.17E-4(4E-6)$	$3.64E-5(1.3E-6)$
Backward LTSA	$2.91E-4(8E-6)$	$1.16E-4(4E-6)$	$3.47E-5(1.3E-6)$
Benchmark	$2.769E-4$	$1.18E-4$	$3.72E-5$

NOTES: Numerical values for *Euler Scheme*, *Euler Scheme I.S.*, *Forward RMQA*, *Forward LTSA*, *Backward RMQA*, *Backward LTSA*, *Benchmark* prices for an up-and-out barrier call option and LV dynamics. *Errors* (in brackets) correspond to one standard deviation. The initial spot price is $\bar{X}_0 = 1.36$, whereas the pair strike-barrier is set to $(1.35, 1.39)$, $(1.36, 1.39)$, $(1.37, 1.39)$ for ITM, ATM and OTM options respectively. The maturity T is 1 month, 6 months or 12 months.

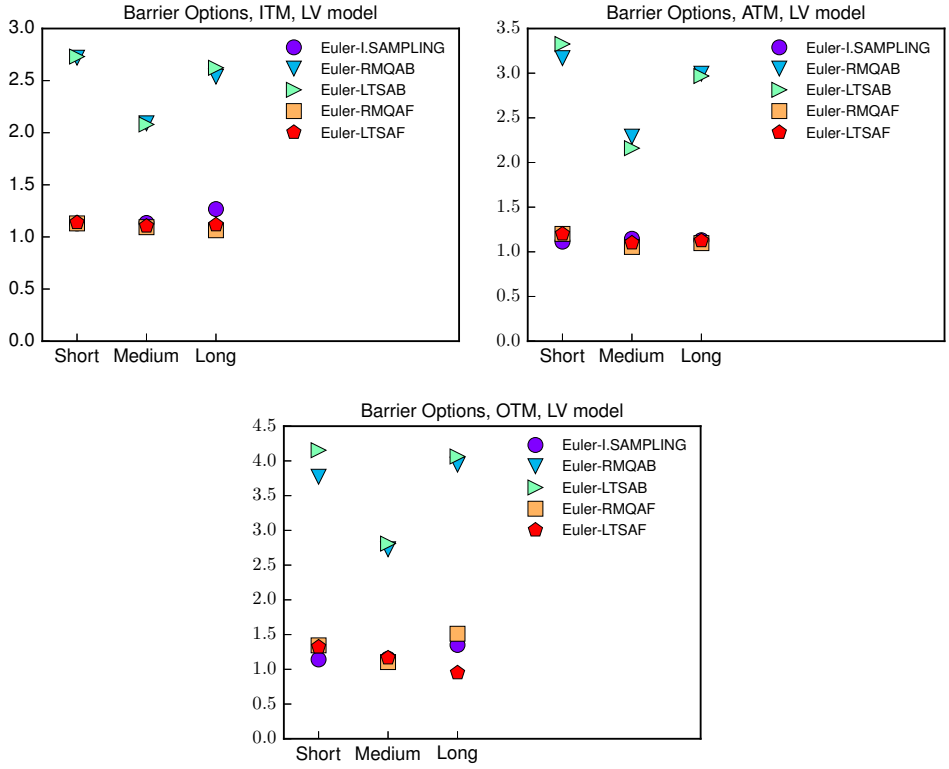
In Table 5.3 and 5.4 we compare performances of the Euler Scheme Monte Carlo with those of the Backward Monte Carlo when pricing Asian call options, for both LV and CEV model. We set $\bar{X}_0 = 1.36$. As done for up-and-out barrier calls we test the efficiency of our novel algorithm when pricing ITM ($K = 1.35$), ATM ($K = 1.36$) and OTM ($K = 1.37$) options. The maturity T is 6 months in the case of CEV model. In the case of LV model the

Table 5.2: Summary output of the results when pricing an up-and-out barrier call option in a CEV dynamics

Up-and-out barrier call			
Algorithm	**ITM**	**ATM**	**OTM**
		CEV model	
$\sigma = 5\%$			
Euler Scheme	$2.491E-3(6.8E-5)$	$1.102E-3(3.9E-5)$	$3.34E-4(1.8E-5)$
Euler Scheme I.S.	$2.457E-3(6.0E-5)$	$1.087E-3(3.4E-5)$	$3.52E-4(1.6E-5)$
Forward RMQA	$2.417E-3(6.7E-5)$	$1.088E-3(3.9E-5)$	$3.40E-4(1.8E-5)$
Forward LTSA	$2.680E-3(7.1E-5)$	$1.109E-3(4.0E-5)$	$3.32E-4(1.9E-5)$
Backward RMQA	$2.426E-3(3.4E-5)$	$1.087E-3(1.7E-5)$	$3.45E-4(7E-6)$
Backward LTSA	$2.525E-3(3.3E-5)$	$1.114E-3(1.7E-5)$	$3.50E-4(7E-6)$
Benchmark	$2.509E-3$	$1.117E-3$	$3.46E-4$
$\sigma = 10\%$			
Euler Scheme	$4.06E-4(2.8E-5)$	$1.71E-4(1.5E-5)$	$5,63E-5(7.4E-6)$
Euler Scheme I.S.	$4.22E-4(2.4E-5)$	$1.57E-4(1.2E-5)$	$5.82E-5(6.3E-6)$
Forward RMQA	$5.07E-4(3.25-5)$	$1.90E-04(1.7E-5)$	$5.26E-5(7.3E-6)$
Forward LTSA	$4.25E-4(2.9E-5)$	$1.86E-04(1.7E-5)$	$5.28E-5(7.2E-6)$
Backward RMQA	$4.16E-4(1.2E-5)$	$1.86E-4(6E-6)$	$5.24E-5(2.2E-6)$
Backward LTSA	$4.13E-4(1.1E-5)$	$1.71E-4(5E-6)$	$5.02E-5(2.1E-6)$
Benchmark	$4.08E-4$	$1.73E-4$	$5.12E-5$
$\sigma = 15\%$			
Euler Scheme	$1.64E-4(1.8E-5)$	$5.26E-5(8.8E-6)$	$1.75E-5(4.2E-6)$
Euler Scheme I.S.	$1.36E-4(1.2E-5)$	$5.81E-5(7.1E-6)$	$2.08E-5(3.4E-6)$
Forward RMQA	$1.44E-4(1.7E-5)$	$5.48E-5(9.25-5)$	$1.96E-5(4.7E-6)$
Forward LTSA	$1.35E-4(1.6E-5)$	$5.28E-5(9.0E-5)$	$1.78E-5(4.2E-6)$
Backward RMQA	$1.24E-4(5E-6)$	$5.22E-5(2.7E-6)$	$1.69E-5(1.1E-6)$
Backward LTSA	$1.28E-4(5E-6)$	$5.77E-5(2.7E-6)$	$1.93E-5(1.1E-6)$
Benchmark	$1.31E-4$	$5.54E-5$	$1.65E-5$
$\sigma = 20\%$			
Euler Scheme	$6.20E-5(1.2E-5)$	$2.38E-5(5.8E-6)$	$5.24E-6(2.2E-06)$
Euler Scheme I.S.	$5.49E-5(8E-6)$	$3.14E-5(5.4E-6)$	$6.22E-6(1.8E-06)$
Forward RMQA	$6.71E-5(1.2E-5)$	$2.34E-5(5.9E-6)$	$5.93E-6(2.3E-06)$
Forward LTSA	$5.73E-5(9E-6)$	$2.45E-5(4.4E-6)$	$4.27E-6(1.9E-06)$
Backward RMQA	$5.75E-5(3E-6)$	$2.22E-5(1.4E-6)$	$6.50E-6(6E-07)$
Backward LTSA	$5.61E-5(3E-6)$	$2.06E-5(1.2E-6)$	$6.28E-6(4E-07)$
Benchmark	$5.66E-5$	$2.38E-5$	$7.10E-6$

NOTES: Numerical values for *Euler Scheme*, *Euler Scheme I.S.*, *Forward RMQA*, *Forward LTSA*, *Backward RMQA*, *Backward LTSA*, *Benchmark* prices for an up-and-out barrier call options and CEV dynamics. *Errors* (in brackets) correspond to one standard deviation. The initial spot price is $\bar{X}_0 = 1.36$, whereas the pair strike-barrier is set to $(1.35, 1.39)$, $(1.36, 1.39)$, $(1.37, 1.39)$ for ITM, ATM and OTM options respectively. The maturity T is 6 months. σ varies from 5% to 20% with step of 5%.

Figure 5.4: Plot of the *Error ratio* as a function of the maturity T for LV model when pricing up-and-out barrier call option. The initial spot price is $\bar{X}_0 = 1.36$, whereas the pair strike-barrier is set to $(1.35, 1.39)$, $(1.36, 1.39)$, $(1.37, 1.39)$ for ITM, ATM and OTM options respectively.



maturity is $T = 1$ month, $T = 6$ months or $T = 12$ months. The number of Euler steps is $n = 9$, $n = 51$ and $n = 100$ for $T = 1$ month, $T = 6$ months or $T = 12$ months respectively. Also in this case, for CEV model we fix the value of the risk-free rate $r = r_d - r_f = 0.32\%$ and of $\alpha = 0.5$, and we vary the value of σ from 5% to 20% with steps $\Delta\sigma = 5\%$. Results for the LV dynamics are reported in Table 5.3, whereas Table 5.4 reports the results for the CEV. Table 5.3 suggests that the strategy of reverting the MC paths and simulating them from maturity back to starting date is an effective alternative to the pure Euler MC or to Euler MC combined with a control variates variance reduction technique. In this case the improvement in efficiency derives from the fact that with Backward MC we decide the number of paths to sample from each of the final points in Γ_n, sampling efficiently also those regions that are infrequently explored by the price process because of its diffusive behaviour. Besides, we note that grids and transition probabilities recovered through the RMQA perform better than LTSA. This is because the grids in the RMQA are obtained as local minima of the distortion function, i.e. they well approximate the law of the stochastic process at any discretisation date. Figures 5.6 and 5.7 support our conclusions, especially when pricing OTM options. The *Error ratio* is almost constant across the value of σ for a fixed scenario (ITM, ATM or OTM).

In Table 5.5 we report prices of auto-callable options for LV dynamics. In this case, we compare only the *Backward LTSA* with the *Forward LTSA*. The Euler Scheme MC is ineffective for pay-offs specifications which depend on the observation of the underlying on a pre-specified set of dates (such as auto-callable and European). The multinomial tree

Table 5.3: Summary output of the results when pricing an Asian option in a LV dynamics

	Asian option		
Algorithm	**ITM**	**ATM**	**OTM**
	Local Volatility model		
$T = 1$ **month**			
Euler Scheme	$8.669E-3(8.0E-5)$	$3.920E-3(5.9E-5)$	$1.321E-3(3.1E-5)$
Euler Scheme C.V.	$8.669E-3(4.1E-5)$	$3.988E-3(3.7E-5)$	$1.379E-3(2.9E-5)$
Forward RMQA	$8.860E-3(8.2E-5)$	$4.062E-3(6.1E-5)$	$1.433E-3(3.7E-5)$
Forward LTSA	$8.773E-3(8.2E-5)$	$3.988E-3(6.0E-5)$	$1.430E-3(3.7E-5)$
Backward RMQA	$8.725E-3(4.2E-5)$	$3.948E-3(3.1E-5)$	$1.362E-3(1.7E-5)$
Backward LTSA	$8.690E-3(5.8E-5)$	$3.957E-3(4.3E-5)$	$1.373E-3(2.3E-5)$
Benchmark	$8.750E-3$	$3.978E-3$	$1.374E-3$
$T = 6$ **months**			
Euler Scheme	$1.3656E-2(1.64E-4)$	$9.408E-3(1.41E-4)$	$6.149E-3(1.17E-4)$
Euler Scheme C.V.	$1.3980E-2(1.06E-4)$	$9.707E-3(0.98E-4)$	$6.402E-3(0.88E-4)$
Forward RMQA	$1.3698E-2(1.64E-4)$	$9.446E-3(1.40E-4)$	$6.197E-3(1.16E-4)$
Forward LTSA	$1.3535E-2(1.61E-4)$	$9.228E-3(1.37E-4)$	$5.792E-3(1.12E-4)$
Backward RMQA	$1.3536E-2(1.09E-4)$	$9.305E-3(0.92E-4)$	$6.068E-3(0.74E-4)$
Backward LTSA	$1.3567E-2(1.43E-4)$	$9.321E-3(1.23E-4)$	$5.852E-3(0.93E-4)$
Benchmark	$1.3724E-2$	$9.4663E-3$	$6.196E-3$
$T = 12$ **months**			
Euler Scheme	$1.8160E-2(2.33E-4)$	$1.4019E-2(2.10E-4)$	$1.0548E-2(1.86E-4)$
Euler Scheme C.V.	$1.7840E-2(1.59E-4)$	$1.3662E-2(1.50E-4)$	$1.0203E-2(1.39E-4)$
Forward RMQA	$1.8085E-2(2.27E-4)$	$1.3912E-2(2.04E-4)$	$1.0413E-2(1.79E-4)$
Forward LTSA	$1.7709E-2(2.28E-4)$	$1.3527E-2(2.06E-4)$	$1.0084E-2(1.83E-4)$
Backward RMQA	$1.8025E-2(1.66E-4)$	$1.3862E-2(1.48E-4)$	$1.0341E-2(1.29E-4)$
Backward LTSA	$1.7898E-2(2.10E-4)$	$1.3685E-2(1.87E-4)$	$1.0183E-2(1.62E-4)$
Benchmark	$1.8111E-2$	$1.3939E-2$	$1.0457E-2$

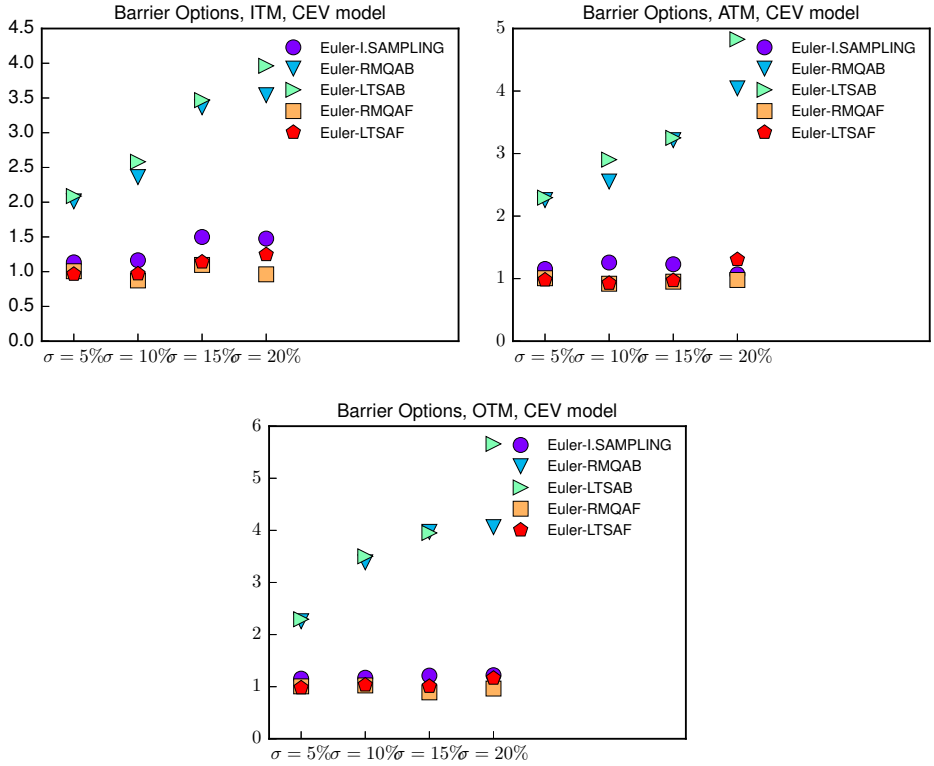
NOTES: Numerical values for *Euler Scheme*, *Euler Scheme C.V.*, *Forward RMQA*, *Forward LTSA*, *Backward RMQA*, *Backward LTSA*, *Benchmark* prices for an Asian option and LV dynamics. *Errors* (in brackets) correspond to one standard deviation. The initial spot price is $\bar{X}_0 = 1.36$, whereas the value of the strike is set to 1.35, 1.36 and 1.37 for ITM, ATM and OTM options respectively. The maturity T is 1 month, 6 months or 12 months. The Euler step is $\Delta t = 0.01$ ($n = 9$, $n = 51$ and $n = 100$ when T is 1 month, 6 months or 12 months respectively).

Table 5.4: Summary output of the results when pricing an Asian option in a CEV dynamics

	Asian option		
Algorithm	**ITM**	**ATM**	**OTM**
		CEV model	
$\sigma = 5\%$			
Euler Scheme	$1.5942E-2(1.75E-4)$	$9.962E-3(1.43E-4)$	$5.651E-3(1.09E-4)$
Euler Scheme C.V.	$1.6284E-2(1.04E-4)$	$1.0245E-2(9.5E-5)$	$5.859E-3(8.0E-5)$
Forward RMQA	$1.5194E-2(1.74E-4)$	$9.969E-3(1.42E-4)$	$5.649E-3(1.07E-4)$
Forward LTSA	$1.6202E-2(1.77E-4)$	$1.0021E-2(1.44E-4)$	$5.727E-3(1.10E-4)$
Backward RMQA	$1.5844E-2(1.05E-4)$	$9.867E-3(6.5E-5)$	$5.542E-3(6.5E-5)$
Backward LTSA	$1.5941E-2(1.53E-4)$	$1.0021E-2(9.3E-5)$	$5.649E-3(9.3E-5)$
Benchmark	$1.5964E-2$	$9.985e-3$	$5.662E-3$
$\sigma = 10\%$			
Euler Scheme	$2.4825E-2(3.17E-4)$	$1.9362E-2(2.84E-4)$	$1.474E-2(2.50E-4)$
Euler Scheme C.V.	$2.5449E-2(1.99E-4)$	$1.9929E-2(1.89E-4)$	$1.5245E-2(1.75E-4)$
Forward RMQA	$2.4634E-2(3.12-4)$	$1.9202E-2(2.79E-4)$	$1.4611E-2(2.44E-4)$
Forward LTSA	$2.4972E-2(3.17E-4)$	$1.9487E-2(2.84E-4)$	$1.4847E-2(2.50E-4)$
Backward RMQA	$2.4588E-2(1.92E-4)$	$1.9135E-2(1.71E-4)$	$1.4506E-2(1.49E-4)$
Backward LTSA	$2.5015E-2(2.50E-4)$	$1.9490E-2(2.23E-4)$	$1.4808E-2(1.94E-4)$
Benchmark	$2.4871E-2$	$1.9441E-2$	$1.4789E-2$
$\sigma = 15\%$			
Euler Scheme	$3.4044E-2(4.60E-4)$	$2.8765E-2(4,27E-4)$	$2.4047E-2(3.93E-4)$
Euler Scheme C.V.	$3.4938E-2(2.93E-4)$	$2.9617E-2(2,83E-4)$	$2.4848E-2(2.70E-4)$
Forward RMQA	$3.4071E-2(4.57E-4)$	$2.8818E-2(4,23E-4)$	$2.4113E-2(3.89E-4)$
Forward LTSA	$3.4378E-2(4.62E-4)$	$2.9075E-2(4,29E-4)$	$2.4334E-2(3.95E-4)$
Backward RMQA	$3.3832E-2(2.78E-4)$	$2.8548E-2(2.57E-4)$	$2.3820E-2(2.36E-4)$
Backward LTSA	$3.4027E-2(3.51E-4)$	$2.8680E-2(3.24E-4)$	$2.3921E-2(2.95E-4)$
Benchmark	$3.411E-2$	$2.883E-2$	$2.4130E-2$
$\sigma = 20\%$			
Euler Scheme	$4.3345E-2(6.05E-4)$	$3.8166E-2(5.72E-4)$	$3.3414E-2(5.38E-4)$
Euler Scheme C.V.	$4.4514E-2(3.88E-4)$	$3.9306E-2(3.77E-4)$	$3.4506E-2(3.58E-4)$
Forward RMQA	$4.3372E-2(5.97E-4)$	$3.8191E-2(5.63E-4)$	$3.3428E-2(5.29E-4)$
Forward LTSA	$4.3422E-2(6.01E-4)$	$3.8246E-2(5.67E-4)$	$3.3479E-2(5.34E-4)$
Backward RMQA	$4.3013E-2(3.62E-4)$	$3.7836E-2(3.41E-4)$	$3.3086E-2(3.20E-4)$
Backward LTSA	$4.3505E-2(4.55E-4)$	$3.8291E-2(4.28E-4)$	$3.3502E-2(3.99E-4)$
Benchmark	$4.3426E-2$	$3.8265E-2$	$3.3524E-2$

NOTES:Numerical values for *Euler Scheme*, *Euler Scheme C.V.*, *Forward RMQA*, *Forward LTSA*, *Backward RMQA*, *Backward LTSA*, *Benchmark* prices for an Asian call option and CEV dynamics. *Errors* (in brackets) correspond to one standard deviation. The initial spot price is $\bar{X}_0 = 1.36$, whereas the strike is 1.35, 1.36 and 1.37 for ITM, ATM and OTM options respectively. The maturity T is 6 months. The Euler time step is $\Delta = 0.01$ ($n = 51$). σ varies from 5% to 20% with step of 5%.

Figure 5.5: Plot of the *Error ratio* as a function of the parameter σ for CEV model when pricing up-and-out barrier call option. The initial spot price is $\bar{X}_0 = 1.36$, whereas the pair strike-barrier is set to $(1.35, 1.39)$, $(1.36, 1.39)$, $(1.37, 1.39)$ for ITM, ATM and OTM options respectively.



and the transition probability matrices are recovered by means of the LTSA. In order to compare the two MC methodologies we fix a set of call-dates $\{t_1^c, \ldots, t_4^c\}$ and a set of pre-fixed coupon $\{Q_1, \ldots, Q_4\}$ and we vary the value of the barrier b. Precisely, $\{t_1, \ldots, t_4\} = \{1, 3, 6, 12\}$ months, $\{Q_1, \ldots, Q_4\} = \{5\%, 10\%, 15\%, 20\%\}$ with unitary notional, and $\mathrm{b} \in \{\bar{X}_0, 1.05\bar{X}_0, 1.1\bar{X}_0\}$. As usual, at pricing date the EUR/USD exchange rate is $\bar{X}_0 = 1.36$. Results in Table 5.5 show that the improvement on pricing accuracy due to the backward simulation of the price paths widens with the increase of the value of the barrier b (in real market usually one has $\mathrm{b} > \bar{X}_0$). The ratio between the pricing error of the Forward MC and that of the Backward MC is ≈ 0.8, ≈ 2 and ≈ 5.5 for $\mathrm{b} = \bar{X}_0$, $\mathrm{b} = 1.05\bar{X}_0$ and $\mathrm{b} = 1.1\bar{X}_0$ respectively. Intuitively, this happens because an increase in the value of the barrier b makes the early exercise of the option less probable, and a larger number of paths will reach the final domain of integration. Then, our methodology allows to efficiently sample paths drawing from regions that would be infrequently explored by the forward Euler algorithm.

Figure 5.6: Plot of the *Error ratio* as a function of the maturity T for LV model when pricing Asian call option. The initial spot price is $\bar{X}_0 = 1.36$, whereas the strike is set to 1.35, 1.36 and 1.37 for ITM, ATM and OTM options respectively.

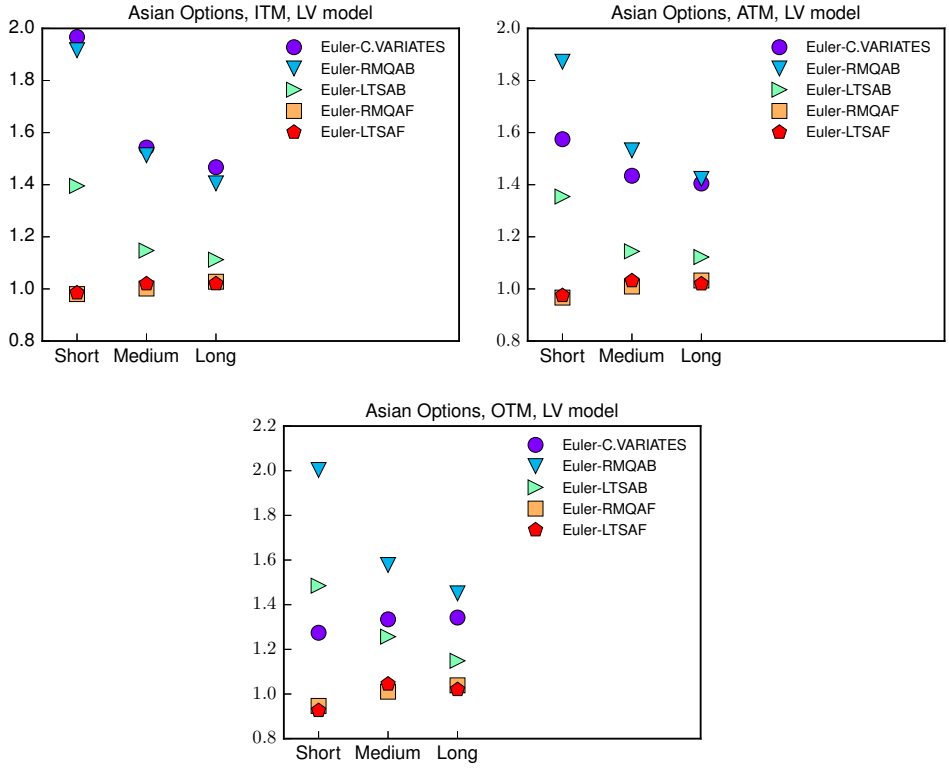
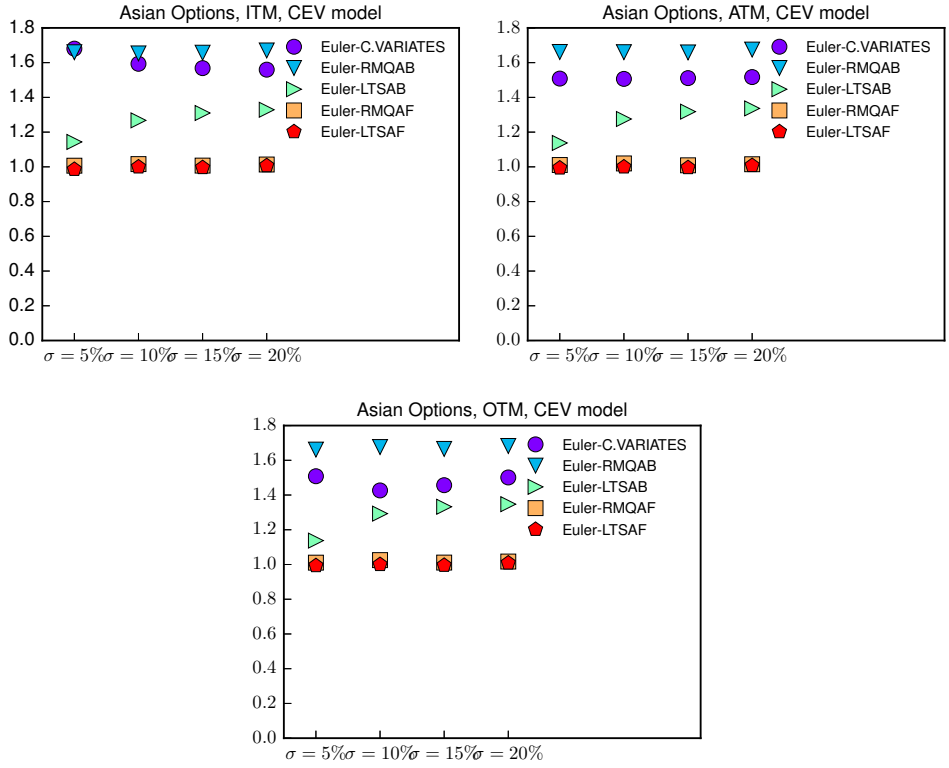


Table 5.5: Summary output of the results when pricing an auto-callable option in a LV dynamics

Auto-callable option			
Algorithm	**ITM**	**ATM**	**OTM**
Barrier b	$\mathbf{b} = \bar{X}_0$	$\mathbf{b} = 1.05\bar{X}_0$	$\mathbf{b} = 1.1\bar{X}_0$
	Local Volatility model		
Forward LTSA	0.04107(0.00056)	0.01902(0.00074)	0.00447(0.00058)
Backward LTSA	0.04099(0.00072)	0.01856(0.00039)	0.00377(0.00011)
Benchmark	0.04099	0.01820	0.00357

NOTES:Numerical values for *Forward*, *Backward* and *Benchmark* prices for an auto-callable option in the LV. *Errors* (in brackets) correspond to one standard deviation. The initial spot price is $\bar{X}_0 = 1.36$. The value of the barrier b is equal to $\bar{X}_0$, $1.05\bar{X}_0$ and $1.1\bar{X}_0$ for the three experiments considered. The call-dates are fixed to $\{1, 3, 6, 12\}$ months and the pre-fixed coupons are fixed to $\{5\%, 10\%, 15\%, 20\%\}$ of a unitary notional.

Figure 5.7: Plot of the *Error ratio* as a function of the parameter σ for CEV model when pricing Asian call options. The initial spot price is $\bar{X}_0 = 1.36$, whereas the strike is set to 1.35, 1.36 and 1.37 for ITM, ATM and OTM options respectively.



5.5 Conclusions

In this chapter, we presented a novel approach – termed backward Monte Carlo – to the Monte Carlo simulation of continuous time diffusion processes. We exploit recent advances in the quantization of diffusion processes to approximate the continuous process with a discrete-time Markov Chain defined on a finite grid of points. Specifically, we consider the Recursive Marginal Quantization Algorithm and as a first contribution we investigate a fixed-point scheme – termed Lloyd I method with Anderson acceleration – to compute the optimal grid in a robust way. As a complementary approach, we consider the grid associated with the explicit scheme approximation of the Markov generator of a piecewise constant volatility process. The latter approach – termed Large Time Step Algorithm – turns out to be competitive in pricing pay-off specifications which require the observation of the price process over a finite number of pre-specified dates. Both methods – quantization and the explicit scheme – provide us with the marginal and transition probabilities associated with the points of the approximating grid. Sampling from the discrete grid backward – from the terminal point to the spot value of the process – we design a simple but effective mechanism to draw Monte Carlo paths and achieve a sizeable reduction of the variance associated with Monte Carlo estimators. Our conclusion has been extensively supported by the numerical results presented in the final section.

B Appendix

B.1 The distortion function and companion parameters

We suppose to have access to the quantizer Γ_k of $\widetilde{X}_{t_k}$ and to the related Voronoi tessellations $\{C_i(\Gamma_k)\}_{i=1,\dots,N}$. We derive an explicit expression for the distortion function $\widetilde{D}(\Gamma_{k+1})$ as follows:

$$
\begin{aligned}
&\widetilde{D}(\Gamma_{k+1}) = \mathbb{E}\left[d(\mathcal{E}_k(\widehat{\widetilde{X}}_{t_k}, \Delta t; Z_{t_{k+1}}), \Gamma_{k+1})^2\right] \\
&= \sum_{i=1}^{N} \mathbb{E}\left[d(\mathcal{E}_k(\gamma_i^k, \Delta t; Z_{t_{k+1}}), \Gamma_{k+1})^2\right] \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k)) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} (m_k(\gamma_i^k) - \gamma_j^{k+1})^2 (\Phi(\gamma_{k+1,j^+}(\gamma_i^k)) - \Phi(\gamma_{k+1,j^-}(\gamma_i^k))) \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k)) \\
&- 2\sum_{i=1}^{N}\sum_{j=1}^{N} (m_k(\gamma_i^k) - \gamma_j^{k+1}) v_k(\gamma_i^k) (\varphi(\gamma_{k+1,j^+}(\gamma_i^k)) - \varphi(\gamma_{k+1,j^-}(\gamma_i^k))) \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k)) \\
&+ \sum_{i=1}^{N}\sum_{j=1}^{N} v_k(\gamma_i^k)^2 (\gamma_{k+1,j^-}(\gamma_i^k)\varphi(\gamma_{k+1,j^-}(\gamma_i^k)) - \gamma_{k+1,j^+}(\gamma_i^k)\varphi(\gamma_{k+1,j^+}(\gamma_i^k))) \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k)) \\
&+ \sum_{i=1}^{N}\sum_{j=1}^{N} v_k(\gamma_i^k)^2 (\Phi(\gamma_{k+1,j^+}(\gamma_i^k)) - \Phi(\gamma_{k+1,j^-}(\gamma_i^k))) \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k)) \,,
\end{aligned}
\tag{B.1}
$$

where Φ and φ indicate the cumulative distribution function and the probability density function of a standard Normal random variable, respectively. To simplify notation, in

Equation (B.1), we set for all $k \in \{0, \dots, n-1\}$ and for all $j \in \{1, \dots, N\}$

$$\gamma_{k+1,j^+}(\gamma) \doteq \frac{\gamma_{j+1/2}^{k+1} - m_k(\gamma)}{v_k(\gamma)} \quad \text{and} \quad \gamma_{k+1,j^-}(\gamma) \doteq \frac{\gamma_{j-1/2}^{k+1} - m_k(\gamma)}{v_k(\gamma)} \quad \text{where}$$
$$\gamma_{j-1/2}^{k+1} \equiv \frac{\gamma_j^{k+1} + \gamma_{j-1}^{k+1}}{2}, \quad \gamma_{j+1/2}^{k+1} \equiv \frac{\gamma_j^{k+1} + \gamma_{j+1}^{k+1}}{2}, \quad \gamma_{1/2}^{k+1} \doteq -\infty, \quad \text{and} \quad \gamma_{N+1/2}^{k+1} \doteq +\infty.$$

The so-called companion parameters $\{\mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k))\}_{i=1,\dots,N}$ and $\{\mathbb{P}(\widetilde{X}_{t_k} \in C_j(\Gamma_k) | \widetilde{X}_{t_{k-1}} \in C_i(\Gamma_k))\}_{j=1,\dots,N}$ are computed in a recursive way as follows:

$$\mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k)) = \sum_{j=1}^{N} (\Phi(\gamma_{k,i^+}(\gamma_j^{k-1})) - \Phi(\gamma_{k,i^-}(\gamma_j^{k-1})))\mathbb{P}(\widetilde{X}_{t_{k-1}} \in C_j(\Gamma_{k-1})),$$
$$\mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k) | \widetilde{X}_{t_{k-1}} \in C_j(\Gamma_{k-1})) = \Phi(\gamma_{k,i^+}(\gamma_j^{k-1})) - \Phi(\gamma_{k,i^-}(\gamma_j^{k-1})).$$

B.2 Lloyd I method within the RMQA

We present a brief review of the Lloyd I method within the Recursive Marginal Quantization framework. Let us fix $t_k \in \{t_1, \dots, t_n\}$ and suppose we have access to the quantizer Γ_k of $\widetilde{X}_{t_k}$ and to the associated Voronoi tessellations $\{C_i(\Gamma_k)\}_{i=1,\dots,N}$. We want to quantize $\widetilde{X}_{t_{k+1}} = \mathcal{E}_k(\widehat{\widetilde{X}}_{t_k}, \Delta t; Z_{t_{k+1}})$ by means of a quantizer $\Gamma_{k+1} \equiv \{\gamma_1^{k+1}, \dots, \gamma_N^{k+1}\}$ of cardinality N. One starts with an initial guess Γ_{k+1}^0 and then one sets recursively a sequence $(\Gamma_{k+1}^l)_{l \in \mathbb{N}}$ such that

$$\gamma_j^{k+1,l+1} = \mathbb{E}\left[\widetilde{X}_{t_{k+1}} | \widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}^l)\right], \tag{B.2}$$

where l indicates the running iteration number. One can easily check that previous equation implies that

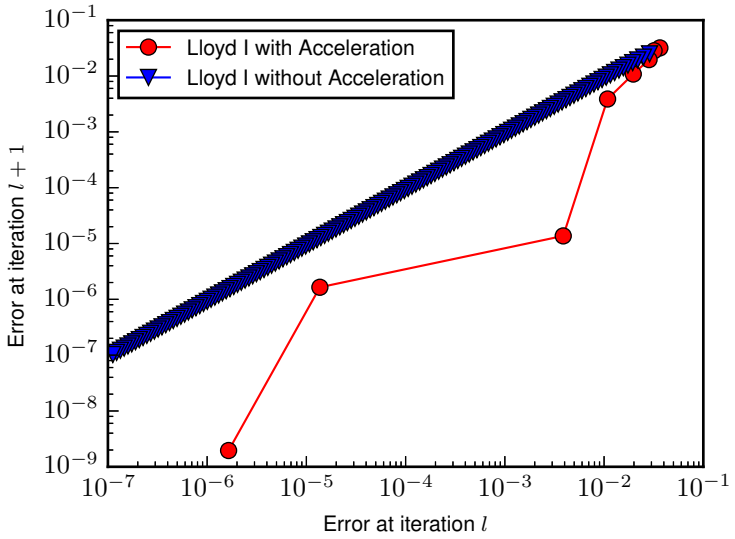
$$q_N^{l+1}(\widetilde{X}_{t_{k+1}}) = \mathbb{E}\left[\widetilde{X}_{t_{k+1}} | q_N^l(\widetilde{X}_{t_{k+1}})\right] \doteq \left(\mathbb{E}\left[\widetilde{X}_{t_{k+1}} | \widetilde{X}_{t_{k+1}} \in C_i(\Gamma_{k+1}^l)\right]\right)_{1 \le i \le N},$$

where q_N^l is the quantization associated with Γ_{k+1}^l. It has been proven (see Pagès et al., 2004; Carmona et al., 2012) that $\{\|\widetilde{X}_{t_{k+1}} - q_N^l(\widetilde{X}_{t_{k+1}})\|_2, l \in \mathbb{N}^+\}$ is a non-increasing sequence and that $q_N^l(\widetilde{X}_{t_{k+1}})$ converges towards some random variable taking N values as l tends to infinity. From Equation (B.2) and exploiting the idea of RMQA we have

$$\begin{aligned}
\gamma_j^{k+1,l+1} &= \mathbb{E}\left[\widetilde{X}_{t_{k+1}} | \widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}^l)\right] \\
&= \frac{\mathbb{E}\left[\widetilde{X}_{t_{k+1}} \mathbb{1}_{\{\widetilde{X} \in C_j(\Gamma_{k+1}^l)\}}\right]}{\mathbb{P}(\widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}^l))} \\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[\widetilde{X}_{t_{k+1}} \mathbb{1}_{\{\widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}^l)\}} | \widetilde{X}_{t_k}\right]\right]}{\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{\widetilde{X}_{t_{k+1}} \in C_j(\Gamma_{k+1}^l)\}} | \widetilde{X}_{t_k}\right]\right]} \\
&= \frac{\sum_{i=1}^N \mathbb{E}\left[\mathcal{E}_k(\gamma_i^k, \Delta t; Z_{t_{k+1}}) \mathbb{1}_{\{\mathcal{E}_k(\gamma_i^k, \Delta t; Z_{t_{k+1}}) \in C_j(\Gamma_{k+1}^l)\}}\right] \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k))}{\sum_{i=1}^N \mathbb{P}(\mathcal{E}_k(\gamma_i^k, \Delta t; Z_{t_{k+1}}) \in C_j(\Gamma_{k+1}^l)) \mathbb{P}(\widetilde{X}_{t_k} \in C_i(\Gamma_k))}.
\end{aligned}$$

The last term in previous equation is equivalent to the stationary condition in Equation (5.9) for the quantization $q_N(\widetilde{X}_{t_{k+1}})$. Then, the stationary condition is equivalent to a fixed point relation for the quantizer.

Figure B.1: Quantization of a standard Normal random variable: Comparison of the convergence of Lloyd I method with and without Anderson acceleration.



B.3 Robustness checks

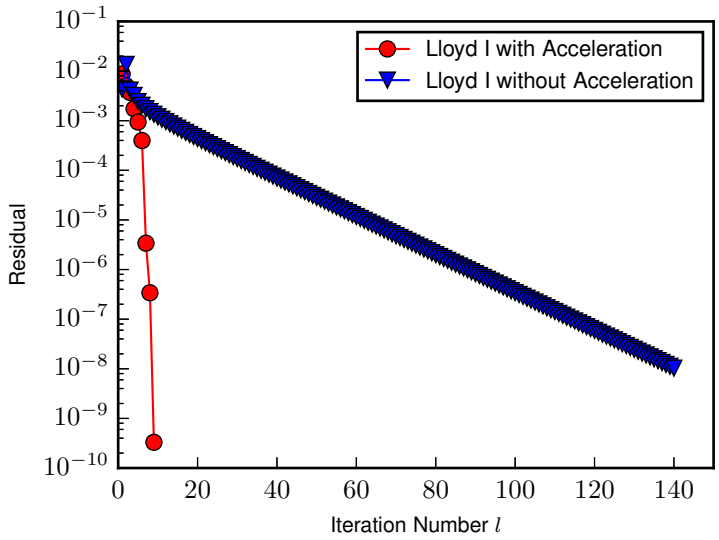
We test the convergence of Lloyd I method with and without Anderson acceleration on the quantization of a standard Normal random variable initialised from a distorted quantizer (at `www.quantize.maths-fi.com` a database providing quadratic optimal quantizers of the standard univariate Gaussian distribution from level $N = 1$ to $N = 1000$ is available). We indicate by $\Gamma^*_{\mathcal{N}(0,1)}$ the optimal quantizer of a standard Normal random variable and we distort it through the multiplication by a constant c, that is to say $c \times \Gamma^*_{\mathcal{N}(0,1)}$. Then, we monitor the convergence of both algorithms to $\Gamma^*_{\mathcal{N}(0,1)}$ starting from $c \times \Gamma^*_{\mathcal{N}(0,1)}$. The error at iteration l is defined as $\|\Gamma^*_{\mathcal{N}(0,1)} - \Gamma^l_{\mathcal{N}(0,1)}\|_2$, with $\Gamma^l_{\mathcal{N}(0,1)}$ the quantizer found by the algorithms at the l-th iteration and $\|\cdot\|_2$ the Euclidean norm in $\mathbb{R}^N$. The stopping criteria is set to $\|\Gamma^{l+1}_{\mathcal{N}(0,1)} - \Gamma^l_{\mathcal{N}(0,1)}\|_2 \leq 10^{-7}$, the level of the quantizer to $N = 10$, and the constant c to 1.01. The results of our investigation are summarized in Figure B.1. We can graphically assess the rate of convergence for both algorithms. We recall that a sequence $(\Gamma^l)_{l\in\mathbb{N}}$ converging to a $\Gamma^* \neq \Gamma^l$ for all l is said to converge to Γ^* with order α and asymptotic error constant λ if there exist positive constants α and λ such that

$$\lim_{l\to\infty} \frac{\|\Gamma^{l+1} - \Gamma^*\|_2}{\|\Gamma^l - \Gamma^*\|_2^\alpha} = \lambda\,.$$

In case of Lloyd I method without acceleration the convergence is, as expected, linear. For Lloyd I method with acceleration the rate is not well defined, but Figure B.1 shows the impressive improvement in the convergence towards the known optimal quantizer. Figure B.2 supports the same conclusion in terms of the number of iterations necessary to reach the stopping criterion.

Then, we investigate numerically the sensitivity of Lloyd I method with Anderson acceleration and Newton-Raphson algorithm to the initial guess as a function of the distortion c applied to the optimal quantizer $\Gamma^*_{\mathcal{N}(0,1)}$. The results of our investigation are summarized in Figure B.3. The four panels correspond to different levels of distortion $c = \{1.10, 1.20, 1.25, 1.35\}$. As before, we set $N = 10$ whereas on the y axis we report the residual at iteration l, $\|\Gamma^{l+1} - \Gamma^l\|_2$. For low levels of distortion Newton-Raphson method

Figure B.2: Quantization of a standard Normal random variable: Comparison of the number of iterations necessary to reach the stopping criterion between Lloyd I method with and without Anderson acceleration.



converges to the optimal solution more quickly than Lloyd I method. This result confirms the theoretical behavior due to the quadratic rate of convergence of the Newton-Raphson algorithm. However, when the initial guess is quite far from the solution – as it is for the cases of 25% and 35% distortion – the algorithm may spend many cycles far away from the optimal grid.

Remark 6. *Notice that it is also possible to design a hybrid algorithm which starts as Lloyd with acceleration and after some iterations, when it is convenient, it switches to Newton-Raphson.*

Finally, we examine the convergence of Lloyd I and Newton-Raphson algorithms when considering the following Euler-Maruyama discrete scheme

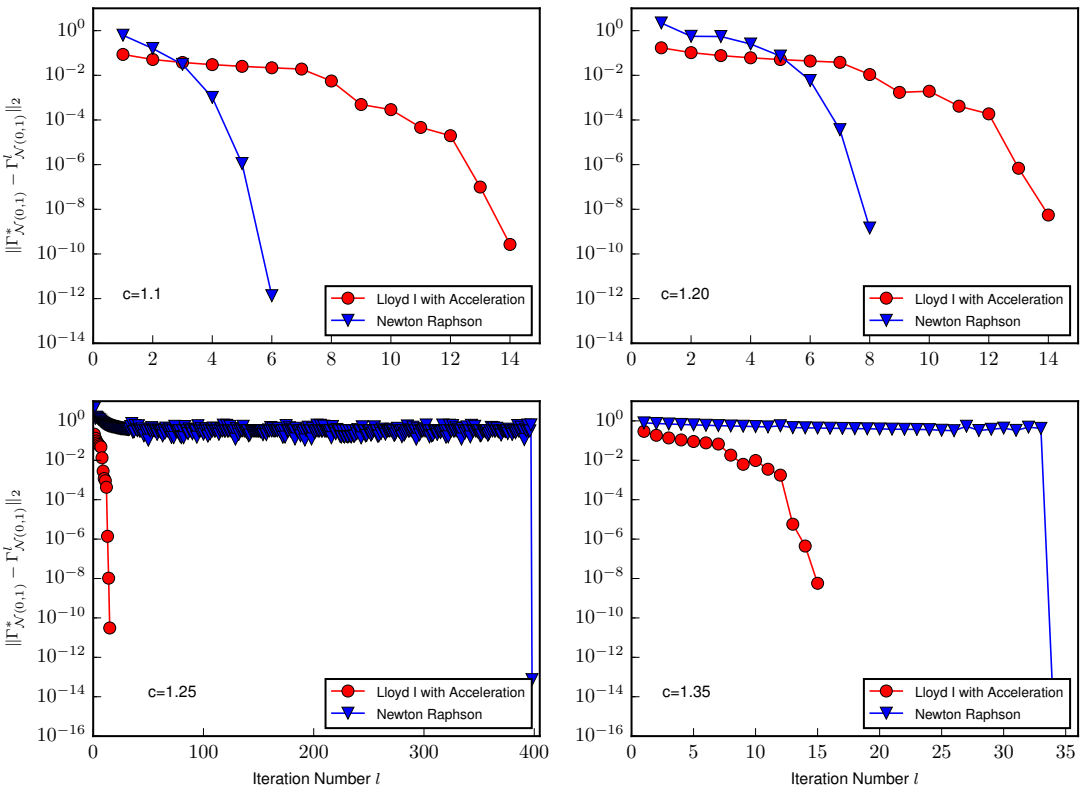
$$\begin{cases} \bar{X}_{t_{k+1}} = \bar{X}_{t_k} + r\bar{X}_{t_k}\Delta t + \sigma\bar{X}_{t_k}\sqrt{\Delta t}\ Z_{t_k}\,, \\ X_0 = x_0\,, \end{cases}$$

with r, σ, and x_0 strictly positive real constants, and $\Delta t = t_{k+1} - t_k$ for all $k = 1, \ldots, n-1$. To enlighten the greater sensitivity of the Newton-Raphson method to the grid initialisation in comparison with the Lloyd I with Anderson acceleration it is sufficient to stop at $n = 2$ with $\Delta t = 0.01$. We set the level of the quantizers Γ_1 and Γ_2 equal to $N = 30$ and $x_0 = 1$. By definition the random variable $\bar{X}_{t_1} \sim \mathcal{N}(m_0(x_0), v_0(x_0))$ where $m_0(x_0) = x_0 + rx_0\Delta t$ and $v_0(x_0) = \sigma x_0$. In order to compute the quantizer for $\bar{X}_{t_1}$ we initialise the algorithms at time t_1 to $m_0(x_0) + v_0(x_0)\Gamma^*_{\mathcal{N}(0,1)}$, with $\Gamma^*_{\mathcal{N}(0,1)}$ the optimal quantizer of a standard Normal random variable. Once we have obtain the optimal quantizer $\Gamma_1^* = \{\gamma_1^{*1}, \cdots, \gamma_1^{*30}\}$ we set the initialisation of the quantizer $\Gamma_2^{Init} = \{\gamma_2^1, \cdots, \gamma_2^{30}\}$ at time t_2 using one of the following alternatives

i. the optimal quantizer at the previous step

$$\Gamma_2^{Init} = \Gamma_1^*;$$

Figure B.3: Quantization of a standard Normal random variable: Comparison between Lloyd I method with Anderson acceleration and Newton-Raphson algorithm.



ii. the Euler operator

$$\gamma_2^i = m_1(\gamma_1^i) + v_1(\gamma_1^i)\Gamma^{*,i}_{\mathcal{N}(0,1)}\,,$$

for $i = 1, \ldots, 30$;

iii. the mid point between Euler operator and the optimal quantizer at the previous step

$$\gamma_2^i = 0.5\gamma_1^i + 0.5(m_1(\gamma_1^i) + v_1(\gamma_1^i)\Gamma^{*,i}_{\mathcal{N}(0,1)})\,,$$

for $i = 1, \ldots, 30$;

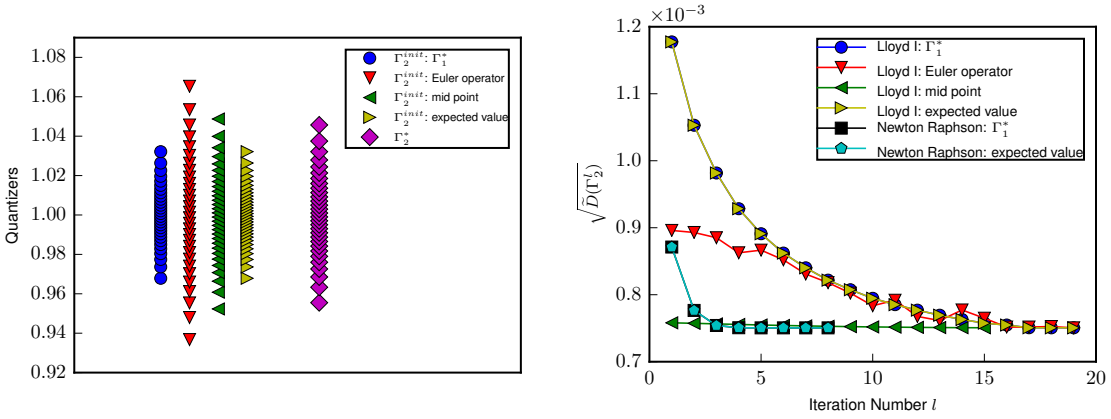
iv. the expected value

$$\gamma_2^i = m_1(\gamma_1^i)\,,$$

for $i = 1, \ldots, 30$.

The left panel of Figure B.4 shows the four different initial grids which correspond to above specifications. In the same panel, on the right side, we also plot the optimal quantizer Γ_2^* to which both Lloyd I with Anderson acceleration and Newton-Raphson methods should converge. The right panel report the quantization error – defined as $\sqrt{\widetilde{D}_2(\Gamma_2^l)}$ – as a function of the iteration number l. We stop the algorithm when the residual falls below 10^{-5}. The numerical investigation shows that the Newton-Raphson method converges to the optimal grid faster than the Lloyd I method, with the only exception represented by the case Γ_2^{init} equal to the mid point. However, when initialized with the Euler operator or the mid point Newton-Raphson algorithm fails to converge due to the bad condition number of the Hessian matrix (corresponding lines are not reported on the Figure). This

Figure B.4: Quantization of the Euler-Maruyama scheme associated to a Geometric Brownian motion. Left panel: four different initial grids Γ_2^{init} and optimal grid Γ_2^* on the left and right sides, respectively. Right panel: quantization error as a function of the iteration number.



result is in line with the findings in Pagès and Printems (2003) where the authors stress that the Newton-Raphson method may fail even for symmetric initial vectors since the anomalous behaviour of some components of the Voronoi tessellation.

In light of above explorations, we finally conclude that the approach based on a fixed-point algorithm such as the Lloyd I method with Anderson acceleration is much more robust than a Newton-Raphson approach – which in the present application relies on the computation of the Hessian of the matrix.

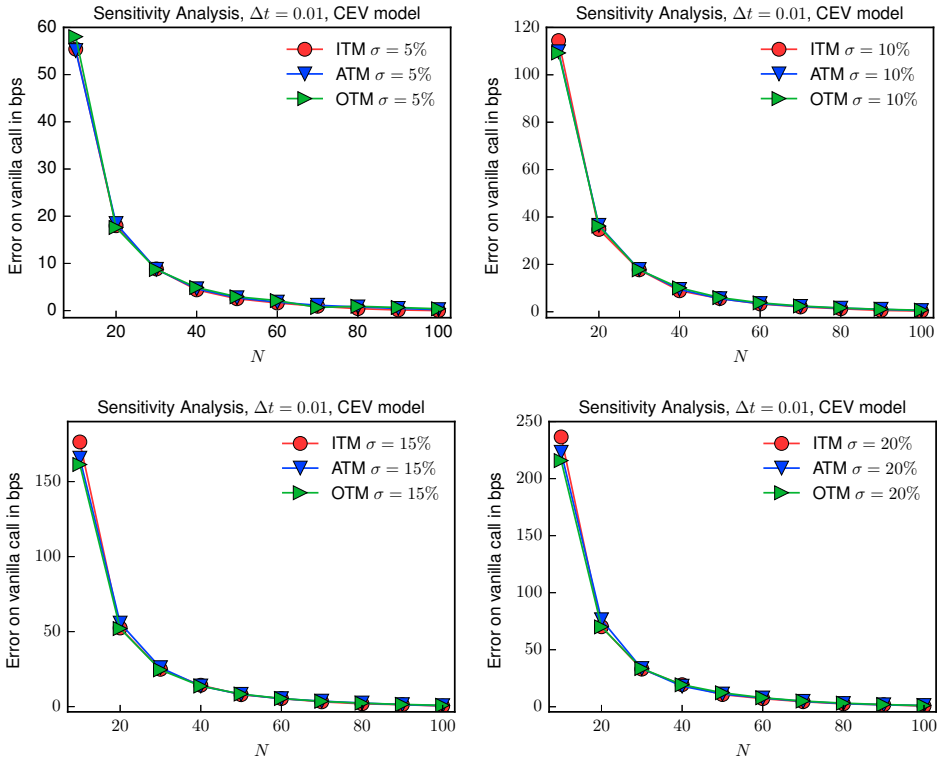
Finally, we investigate the impact of the state space finite granularity on option prices. In the numerical experiments, we employ the CEV model – see Equation (5.20) – with $r_d - r_f = 0.32\%$, $\alpha = 0.5$, and $X_0 = \bar{X}_0 = 1.36$. Besides $\Delta t = 0.01$. The error on vanilla call option prices has been computed according to Equation (5.22) as a function of N for different values of σ and different exercise conditions (ITM, ATM, OTM). The four panels of Figure B.5 report the numerical results. When the level of the volatility parameter σ grows the error increases consistently. As expected the mispricing is more noticeable for ITM than for OTM options. N equal to 100 guarantees – at least empirically – the reduction of the absolute error below 5 bps. Accordingly, in numerical applications in the main text we fix $N = 100$.

B.4 Construction of the Markov generator $\mathcal{L}_\Gamma$ and sensitivity analysis

We define $\Delta\gamma \doteq \gamma_{i+1} - \gamma_i$, $1 \le i \le N-1$. The finite difference approximation of the first and second partial derivative in Equation (5.16) is defined as

$$\frac{\partial u}{\partial \gamma}(\gamma, t) \approx \frac{u(\gamma_i + \Delta\gamma, t) - u(\gamma_i - \Delta\gamma, t)}{2\Delta\gamma},$$
$$\frac{\partial^2 u}{\partial \gamma^2}(\gamma, t) \approx \frac{u(\gamma_i + \Delta\gamma, t) - 2u(\gamma_i, t) + u(\gamma_i - \Delta\gamma, t)}{(\Delta\gamma)^2},$$

Figure B.5: Study of the influence of the granularity N of the state space on the pricing of call vanilla options when employing the RMQA, CEV model. *Top left*: $\sigma = 5\%$, *Top right*: $\sigma = 10\%$, *Bottom left*: $\sigma = 15\%$, *Bottom right*: $\sigma = 20\%$.



for all $t \in [0, T]$. The Markov generator $\mathcal{L}_\Gamma$ is the $N \times N$ matrix defined as

$$\mathcal{L}_\Gamma \doteq \begin{pmatrix} d_1 & u_1 & 0 & 0 & \cdots & 0 & 0 \\ l_2 & d_2 & u_2 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \cdots & 0 & 0 \\ 0 & 0 & l_i & d_i & u_i & 0 & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & l_{N-1} & d_{N-1} & u_{N-1} \\ 0 & 0 & 0 & 0 & 0 & l_N & d_N \end{pmatrix},$$

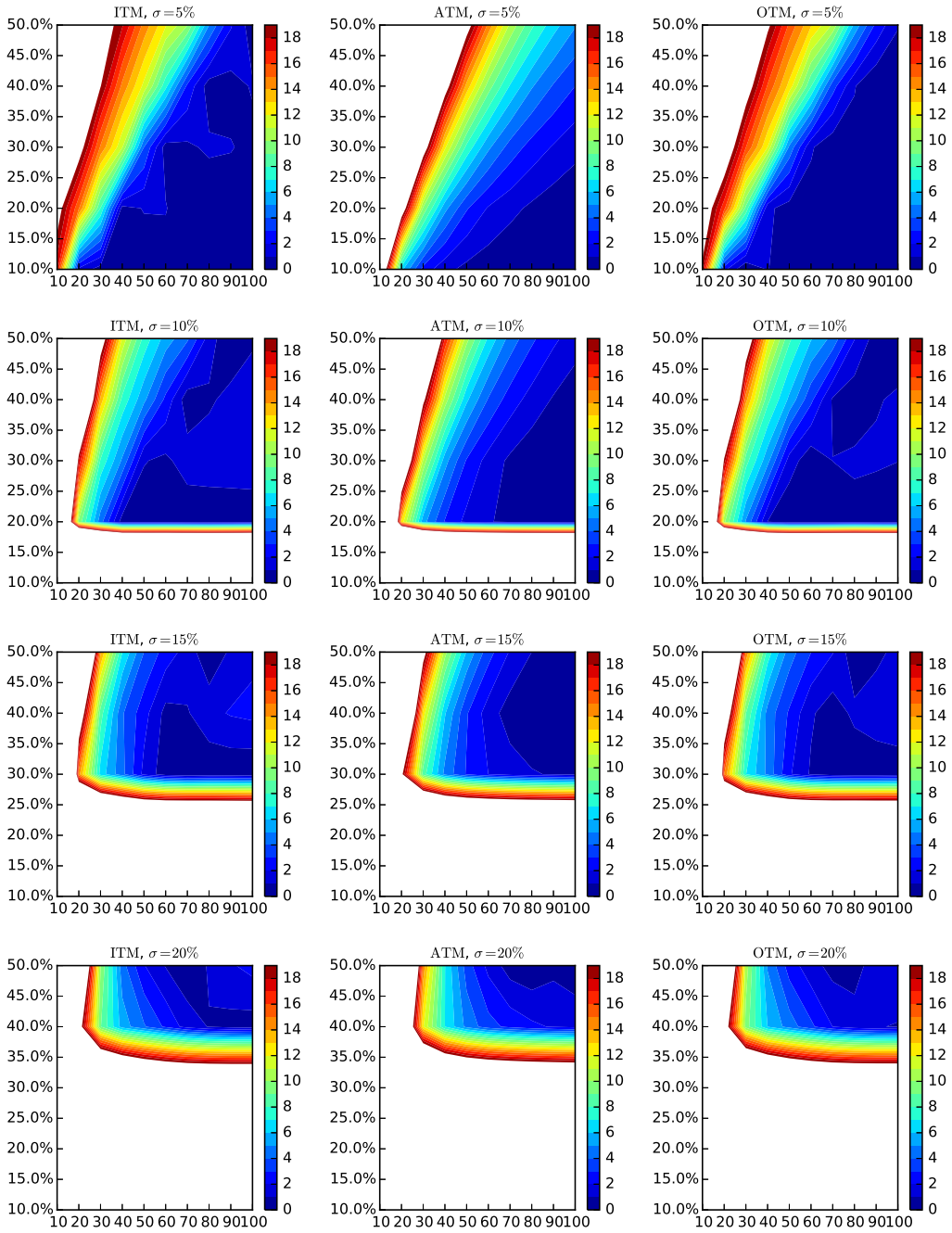
where the coefficients l_i, d_i and u_i are given by

$$\begin{aligned} l_i &= -\frac{b(\gamma_i)}{2\Delta\gamma} + \frac{1}{2}\frac{\sigma(\gamma_i)^2}{(\Delta\gamma)^2}, \\ d_i &= -\frac{\sigma(\gamma_i)^2}{(\Delta\gamma)^2}, \\ u_i &= +\frac{b(\gamma_i)}{2\Delta\gamma} + \frac{1}{2}\frac{\sigma(\gamma_i)^2}{(\Delta\gamma)^2}, \end{aligned}$$

for all $1 \leq i \leq N$. The coefficients of the first and last row are chosen so that the Markov chain is reflected at the boundaries of the state domain. The choice of the boundary conditions should have a negligible effect provided that the range of the state domain is sufficiently large. To quantify what sufficiently large means in practical situation, we consider again the CEV model corresponding to Equation (5.20) with $r_d - r_f = 0.32\%$,

$\alpha = 0.5$ and $X_0 = \bar{X}_0 = 1.36$. Besides $\Delta t = 0.01$. The contribution of the drift to the dynamics is negligible, and to choose the grid boundary values we proceed as follows. For a given value of σ, we first take symmetric intervals around the starting point X_0, i.e. $X_0 \pm \beta X_0$, for different values of the coefficient β. Then, for a fixed β, we vary the granularity N of the state space and we compute the error on vanilla prices according to Equation (5.22). We display a contour plot CONTOUR(X, Y, Z), where X is the vector of N values, Y contains the values of the coefficient β, and Z is the matrix of errors in basis points. The panels of Figure B.6 report the sensitivity analysis for different values of σ, and different exercise conditions (ITM, ATM, OTM). Accordingly, in numerical applications we fix the cardinality of the quantizer to 100 and β to 0.15, 0.25, 0.35, and 0.45 for $\sigma = 5\%$, $\sigma = 10\%$, $\sigma = 15\%$, and $\sigma = 20\%$, respectively.

Figure B.6: Sensitivity analysis of the influence of the boundary conditions on the pricing of Call vanilla options for the LTSA and CEV dynamics. *From left to right, from top to bottom*: Contour plot, CONTOUR(X, Y, Z) where X is the vector containing the variable N, Y is the vector containing the coefficient β, and Z is the matrix of errors in basis points. The white regions correspond to errors larger than 19 basis points. *First row*: $\sigma = 5\%$, *Second row*: $\sigma = 10\%$, *Third row*: $\sigma = 15\%$, *Fourth row*: $\sigma = 20\%$.



Conclusions

The objective of the thesis is manifold.

First, we motivate the use of rough volatility models for applications, notably for option pricing and hedging. We show that rough volatility models seem to be compatible with observed option prices. Thus, they can be used consistently. We investigate the nature of the financial data and what they are telling us about the behaviour of the volatility process.

Second, we motivate and introduce the first fully analytical discrete-time option pricing stochastic volatility model – SV-LHARG(p) – combining stochastic volatility with realized measures. Our model is characterized by two measurement equations relating both the observed return and realized measures to the latent conditional variance, which features a multiple components structure in both volatility and leverage. Following Majewski et al. (2015), we derive analytically the recursive formulae for the MGF under $\mathbb{P}$ and $\mathbb{Q}$ measures, the formal change of measure, as well as the no-arbitrage condition. The change of measure is obtained using an exponential affine SDF with two risk premia.

Third, we extend the work of Creal (2015) to the case of two-measurement equations. Thus, contribute to the non-linear and non-Gaussian filtering literature by providing new analytical filtering and smoothing relations for the basic version of the SV-LHARG(p). We demonstrate the importance of such relations and we develop an effective Metropolis-Hasting algorithm for the general SV-LHARG(p). We confirm the efficacy of the MCMC algorithm through an extensive simulation exercise.

Fourth, we empirically assess the importance of our modelling in an option pricing exercise. The proposed model suggest significant improvement compared to existing models in the literature in tracking dynamics of the short-end of the implied volatility surface and in pricing medium-long maturity options.

Fifth, we motivate and introduce an algorithm – the backward Monte Carlo – based on Monte Carlo simulation for the effective pricing of path-dependent options where sampled path runs in a backward way on top of a multinomial tree. We provide two ways to construct the tree and we fix some flaws highlighted in the existing literature for one of these two methodologies. We assess the robustness of the proposed solution in some representative cases.

Finally, we demonstrate the importance of sampling from the multinomial tree in a backward way, from the terminal point to the initial value of the underlying asset price in a option pricing exercise. We investigate the performance of the backward Monte Carlo, especially in combination with different variance reduction techniques, by pricing Asian, barrier, and auto-callable options.

Bibliography

Adrian, T. and J. Rosenberg (2008). Stock returns and volatility: Pricing the short-run and long-run components of market risk. *The Journal of Finance 63*(6), 2997–3030.

Albanese, C. (2007). Operator methods, abelian processes and dynamic conditioning. *Available at* SSRN 1018490.

Albanese, C., H. Lo, and A. Mijatović (2009). Spectral methods for volatility derivatives. *Quantitative Finance 9*(6), 663–692.

Albanese, C. and A. Mijatovic (2007). Convergence rates for diffusions on continuous-time lattices. *Available at* SSRN 1018609.

Alitab, D., G. Bormetti, F. Corsi, and A. A. Majewski (2015). A jump and smile ride: Continuous and jump variance risk premia in option pricing. *Available at* SSRN 2631155.

Alziary, B., J.-P. Décamps, and P.-F. Koehl (1997). A PDE approach to Asian options: analytical and numerical evidence. *Journal of Banking & Finance 21*(5), 613–640.

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

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*(4), 885–905.

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

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

Andersen, T. G., D. Dobrev, and E. Schaumburg (2008). Duration based volatility estimation. *Manuscript, Northwestern University*.

Anderson, D. G. (1965). Iterative procedures for nonlinear integral equations. *Journal of the ACM (JACM) 12*(4), 547–560.

Andreasen, J. and B. Huge (2011). Volatility interpolation. *Risk Magazine 24*(3), 76.

Ané, T. and H. Geman (2000). Order flow, transaction clock, and normality of asset returns. *The Journal of Finance 55*(5), 2259–2284.

Asai, M. and M. McAleer (2009). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics 150*(2), 182–192.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: A review. *Econometric Reviews 25*(2-3), 145–175.

Bachelier, L. (1900). *Théorie de la spéculation*. Gauthier-Villars.

Baker, G. A. and P. R. Graves-Morris (1996). *Padé approximants*, Volume 59. Cambridge University Press.

Bakshi, G. and N. Kapadia (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies 16*(2), 527–566.

Bally, V., G. Pages, et al. (2003). A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli 9*(6), 1003–1049.

Bandi, F. M. and J. R. Russell (2006). Separating microstructure noise from volatility. *Journal of Financial Economics 79*(3), 655–692.

Bandi, F. M. and J. R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies 75*(2), 339–369.

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

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

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(2), 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics 17*(5), 457–477.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*(1), 1–37.

Barndorff-Nielsen, O. E. and N. Shephard (2005). How accurate is the asymptotic approximation to the distribution of realised variance? In *Identification and Inference for Econometric Models*. Cambridge University Press.

Barone-Adesi, G., R. F. Engle, and L. Mancini (2008). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies 21*(3), 1223–1258.

Barraquand, J. and D. Martineau (1995). Numerical valuation of high dimensional multivariate american securities. *The Journal of Financial and Quantitative Analysis 30*(3), 383–405.

Bates, D. S. (1996a). 20 testing option pricing models. *Handbook of statistics 14*, 567–611.

Bates, D. S. (1996b). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *Review of Financial Studies 9*(1), 69–107.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics 94*(1), 181–238.

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

Becker, R., A. E. Clements, and S. I. White (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance 31*(8), 2535–2549.

Bekierman, J. and B. Gribisch (2016). Estimating stochastic volatility models using realized measures. *Studies in Nonlinear Dynamics & Econometrics 20*(3), 279–300.

Bennedsen, M., A. Lunde, and M. S. Pakkanen (2015). Hybrid scheme for Brownian semistationary processes. *Available at* SSRN 2636829.

Bennedsen, M., A. Lunde, and M. S. Pakkanen (2016). Decoupling the short-and long-term behavior of stochastic volatility. *Available at* SSRN 2846756.

Beran, J. (1994). *Statistics for long-memory processes*, Volume 61. CRC press.

Berkaoui, A., M. Bossy, and A. Diop (2008). Euler scheme for SDEs with non-lipschitz diffusion coefficient: strong convergence. *ESAIM: Probability and Statistics 12*, 1–11.

Bernardo, J., M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West (2003). Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, pp. 307. Oxford University Press, USA.

Bertholon, H., A. Monfort, and F. Pegoraro (2008). Econometric asset pricing modelling. *Journal of Financial Econometrics 6*(4), 407–458.

Billio, M., R. Casarin, and A. Osuntuyi (2016). Efficient Gibbs sampling for markov switching garch models. *Computational Statistics & Data Analysis 100*(C), 37–57.

Björk, T. (2009). *Arbitrage theory in continuous time*. Oxford university press.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The journal of Political Economy 81*(3), 637–654.

Blair, B. J., S.-H. Poon, and S. J. Taylor (2010). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. In *Handbook of Quantitative Finance and Risk Management*, pp. 1333–1344. Springer.

Blanes, S., F. Casas, J. Oteo, and J. Ros (2009). The Magnus expansion and some of its applications. *Physics Reports 470*(5), 151–238.

Bloomberg (2008). Introduction into the new Bloomberg implied volatility calculations.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006). Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics 4*(3), 353–384.

Bollerslev, T., G. Tauchen, and H. Zhou (2009). Expected stock returns and variance risk premia. *Review of Financial studies 22*(11), 4463–4492.

Bollerslev, T. and J. H. Wright (2001). High-frequency data, frequency domain inference, and volatility forecasting. *Review of Economics and Statistics 83*(4), 596–602.

Bormetti, G., G. Callegaro, G. Livieri, and A. Pallavicini (2017). A backward monte carlo approach to exotic option pricing. *European Journal of Applied Mathematics* DOI*: https://doi.org/10.1017/S0956792517000079*.

Bormetti, G., R. Casarin, F. Corsi, and G. Livieri (2016). Smile at errors: A discrete-time stochastic volatility framework for pricing options with realized measures. *Available at* SSRN 28114304.

Bormetti, G., F. Corsi, and A. A. Majewski (2015). Term structure of variance risk premium in multi-component GARCH models. *Available at* SSRN 2619278.

Bormetti, G., G. Montagna, N. Moreni, and O. Nicrosini (2006). Pricing exotic options in a path integral approach. *Quantitative Finance 6*(1), 55–66.

Bossy, M. and A. Diop (2007). *An efficient discretisation scheme for one dimensional SDEs with a diffusion coefficient function of the form* $|x|^a$*, a in* $[1/2, 1)$. Ph. D. thesis.

Bouchaud, J.-P. and M. Potters (2003). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge University press.

Bühlmann, H., F. Delbaen, P. Embrechts, and A. N. Shiryaev (1996). No-arbitrage, change of measure and conditional Esscher transforms. *CWI quarterly 9*(4), 291–317.

Callegaro, G., L. Fiorin, and M. Grasselli (2015). Pricing and calibration in local volatility models via fast quantization. *Risk Magazine 28*(4), 62–67.

Calvet, L. E. and A. J. Fisher (2004). How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics 2*(1), 49–83.

Carmona, R., P. Del Moral, P. Hu, and N. Oudjane (2012). *Numerical Methods in Finance: Bordeaux, June 2010*. Springer Science & Business Media.

Casarin, R., R. V. Craiu, and F. Leisen (2011). Interacting multiple try algorithms with different proposal distributions. *Statistics and Computing 23*(2), 185–200.

Casarin, R., L. Dalla Valle, F. Leisen, et al. (2012). Bayesian model selection for beta autoregressive processes. *Bayesian Analysis 7*(2), 385–410.

Casarin, R., J.-M. Marin, et al. (2009). Online data processing: Comparison of Bayesian regularized particle filters. *Electronic Journal of Statistics 3*, 239–258.

Casarin, R., D. Sartore, and M. Tronzano (2016). A Bayesian Markov-switching correlation model for contagion analysis on exchange rate markets. *Journal of Business & Economic Statistics* (just-accepted), 1–46.

Casella, G. and C. P. Robert (2004). *Monte Carlo Statistical Methods*. New York: Springer Verlag.

Chen, Z., R. T. Daigler, and A. M. Parhizgari (2006). Persistence of volatility in futures markets. *Journal of Futures Markets 26*(6), 571–594.

Cheridito, P., H. Kawaguchi, M. Maejima, et al. (2003). Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab 8*(3), 1–14.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics 10*(4), 407–432.

Christoffersen, P., B. Feunou, K. Jacobs, and N. Meddahi (2014). The economic value of realized volatility: Using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis 49*(03), 663–697.

Christoffersen, P., B. Feunou, and Y. Jeon (2015). Option valuation with observable volatility and jump dynamics. *Journal of Banking & Finance 61*, S101–S120.

Christoffersen, P., S. Heston, and K. Jacobs (2013). Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies*, 1963–2006.

Christoffersen, P., K. Jacobs, C. Ornthanalai, and Y. Wang (2008). Option valuation with long-run and short-run volatility components. *Journal of Financial Economics 90*(3), 272–297.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, 135–155.

Clewlow, L. and C. Strickland (1996). Implementing derivative models (Wiley Series in Financial Engineering).

Coleman, T. F., Y. Li, and A. Verma (1999). Reconstructing the unknown local volatility function. *Journal of Computational Finance 2*(3), 77–100.

Comte, F., L. Coutin, and É. Renault (2012). Affine fractional stochastic volatility models. *Annals of Finance 8*(2-3), 337–378.

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

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

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, nbp001.

Corsi, F., N. Fusari, and D. La Vecchia (2013). Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics 107*(2), 284–304.

Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008). The volatility of realized volatility. *Econometric Reviews 27*(1-3), 46–78.

Corsi, F., D. Pirino, and R. Renó (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics 159*(2), 276–288.

Corsi, F. and R. Renò (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics 30*(3), 368–380.

Cox, J. (1975). Notes on option pricing i: Constant elasticity of variance diffusions. *Unpublished note, Stanford University, Graduate School of Business*.

Creal, D. D. (2015). A class of non-Gaussian state space models with exact likelihood inference. *Journal of Business & Economic Statistics* (just-accepted).

Darolles, S., C. Gourieroux, and J. Jasiak (2006). Structural Laplace transform and compound autoregressive models. *Journal of Time Series Analysis 27*(4), 477–503.

de Pinho, F. M., G. C. Franco, and R. S. Silva (2016). Modeling volatility using state space models with heavy tailed distributions. *Mathematics and Computers in Simulation 119*, 108 – 127.

Deng, G., J. Mallett, and C. McCann (2011). Modeling autocallable structured products. *Journal of Derivatives & Hedge Funds 17*(4), 326–340.

Derman, E. and Z. J. Kani, I. (1996). The local volatility surface: Unlocking the information in index options prices. *Financial Analysts Journal 52*(4), 25–36.

Deschamps, P. (2011). Bayesian estimation of an extended local scale stochastic volatility model. *Journal of Applied Econometrics 162*, 369–382.

Dewynne, J. and W. Shaw (2008). Differential equations and asymptotic solutions for arithmetic Asian options: ′Black–Scholes′ formula for Asian rate calls. *European Journal of Applied Mathematics 19*(4), 353–391.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Dobrev, D. and P. Szerszen (2010). The information content of high-frequency data for estimating equity return models and forecasting risk. *Available at* SSRN 1895533.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag.

Duan, J.-C. (1995). The garch option pricing model. *Mathematical finance 5*(1), 13–32.

Duffie, D., D. Filipović, and W. Schachermayer (2003). Affine processes and applications in finance. *Annals of Applied Probability*, 984–1053.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 1343–1376.

Dupire, B. (1994). Pricing with a smile. *Risk 7*(1), 18–20.

El Euch, O., M. Fukasawa, and M. Rosenbaum (2016). The microstructural foundations of leverage effect and rough volatility. *arXiv preprint arXiv:1609.05177*.

El Euch, O. and M. Rosenbaum (2016). The characteristic function of rough Heston models. *arXiv preprint arXiv:1609.02108*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1), 3–27.

Engle, R. F., T. Ito, and W.-L. Lin (1990). Meteor showers or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica 58*(3), 525–542.

Fang, F. and C. W. Oosterlee (2008). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing 31*(2), 826–848.

Ferrante, M. and P. Vidoni (1998). Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises. *Stochastic Process. Appl. 77*, 69–81.

Fiorentini, G., C. Planas, and A. Rossi (2014). Efficient MCMC sampling in dynamic mixture models. *Statistics and Computing 24*(1), 77–89.

Forde, M. and H. Zhang (2015). Asymptotics for rough stochastic volatility and Lévy models. *arXiv preprint arXiv:1610.08878*.

Fouque, J.-P. and M. J. Lorig (2011). A fast mean-reverting correction to Heston's stochastic volatility model. *SIAM Journal on Financial Mathematics 2*(1), 221–254.

Fritsch, F. N. and R. E. Carlson (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis 17*(2), 238–246.

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation for state space models based on reparameterizations. *State Space and Unobserved Component Models: Theory and Applications*, 123–151.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Fukasawa, M. (2011). Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics 15*(4), 635–654.

Fukasawa, M. (2016). Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance, to appear*.

Funahashi, H. and M. Kijima (2015). Does the Hurst index matter for option prices under fractional volatility? *Annals of Finance*, 1–20.

Gagliardini, P., C. Gourieroux, and E. Renault (2011). Efficient derivative pricing by the extended method of moments. *Econometrica*, 1181–1232.

Gallo, G. M. and E. Otranto (2015). Forecasting realized volatility with changing average levels. *International Journal of Forecasting 31*(3), 620–634.

Gatheral, J. (2011). *The volatility surface: a practitioner's guide*, Volume 357. John Wiley & Sons.

Gatheral, J., T. Jaisson, and M. Rosenbaum (2014). Volatility is rough. *Available at* SSRN 2509457.

Geman, H., D. B. Madan, and M. Yor (2001). Time changes for Lévy processes. *Mathematical Finance 11*(1), 79–96.

Gerber, H. U., E. S. Shiu, et al. (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries 46*(99), 140.

Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, Volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Glasserman, P. (2004). *Monte Carlo methods in Financial Engineering. Applications of Mathematics (New York). Stochastic Modeling and Applied Probability.*, Volume 53. Springer-Verlag, New York.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

Gobet, E. (2000). Weak approximation of killed diffusion using Euler schemes. *Stochastic processes and their applications 87*(2), 167–197.

Golub, G. H. and C. F. Van Loan (2012). *Matrix computations*, Volume 3. JHU Press.

Gouriéroux, C. and J. Jasiak (2006). Autoregressive gamma processes. *Journal of Forecasting 25*(2), 129–152.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009). The Wishart Autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*(2), 167–181.

Gourieroux, C. and A. Monfort (2007). Econometric specification of stochastic discount factor models. *Journal of Econometrics 136*(2), 509–530.

Gourieroux, C., A. Monfort, and V. Polimenis (2006). Affine models for credit risk analysis. *Journal of Financial Econometrics 4*(3), 494–530.

Graf, S. and H. Luschgy (2000). *Foundations of quantization for probability distributions*. Springer.

Guennoun, H., A. Jacquier, and P. Roome (2014). Asymptotic behaviour of the fractional Heston model. *Available at* SSRN 2531468.

Hagan, P. S., D. Kumar, A. S. Lesniewski, and D. E. Woodward (2002). Managing smile risk. *Wilmott Magazine*, 84–108.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hansen, P. R. and G. Horel (2009). Quadratic variation by Markov chains. *Univ. of Aarhus Dept. of Economics Research Paper* (2009-13).

Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics 20*(7), 873–889.

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

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *The Review of Economic Studies 61*(2), 247–264.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*(2), 327–343.

Heston, S. L. and S. Nandi (2000). A closed-form GARCH option valuation model. *Review of Financial Studies 13*(3), 585–625.

Higgins, M. L. and A. K. Bera (1992). A class of nonlinear ARCH models. *International Economic Review*, 137–158.

Higham, N. J. (2005). The scaling and squaring method for the matrix exponential revisited. *SIAM Journal on Matrix Analysis and Applications 26*(4), 1179–1193.

Hui, C. H., C.-F. Lo, and P. Yuen (2000). Comment on 'Pricing double barrier options using Laplace transforms' by Antoon Pelsser. *Finance and Stochastics 4*(1), 105–107.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance 42*(2), 281–300.

Hull, J. C. (2006). *Options, futures, and other derivatives*. Pearson Education India.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

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

Jeanblanc, M., M. Yor, and M. Chesney (2009). *Mathematical methods for financial markets*. Springer Science & Business Media.

Jin, X. and J. M. Maheu (2012). Modeling realized covariances and returns. *Journal of Financial Econometrics 11*(2), 335–369.

Joshi, M. S. and T. S. Leung (2011). Using Monte Carlo simulation and importance sampling to rapidly obtain jump-diffusion prices of continuous barrier options. *Journal of Computational Finance 10*(4), 93–105.

Kahalé, N. (2004). An arbitrage-free interpolation of volatilities. *Risk Magazine 17*(5), 102–106.

Karatzas, I. and S. Shreve (2012). *Brownian Motion and Stochastic Calculus*, Volume 113. Graduate Texts in Mathematics, Springer-Verlag, New York.

Kieffer, J. B. (1982). Exponential rate of convergence for Lloyd's method i. *IEEE Transactions on Information Theory 28*(2), 205–210.

Kijima, M. (1997). *Markov Processes for Stochastic Modelling*, Volume 6. Stochastic Modeling Series. Chapman & Hall, London.

Koopman, S. J., B. Jungbacker, and E. Hol (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance 12*(3), 445–475.

Koopman, S. J. and M. Scharth (2013). The analysis of stochastic volatility in the presence of daily realized measures. *Journal of Financial Econometrics 11*(1), 76–115.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management science 48*(8), 1086–1101.

Kronmal, R. A. and A. V. Peterson Jr (1979). On the Alias method for generating random variables from a discrete distribution. *The American Statistician 33*(4), 214–218.

Kushner, H. and P. G. Dupuis (2001). *Numerical methods for stochastic control problems in continuous time*, Volume 24. Second Edition. Applications of Mathematics (New York), 24. Stochastic Modeling and Applied Probability. Springer-Verlag, New York.

Labbé, C., B. Rémillard, and J.-F. Renaud (2011). A simple discretization scheme for nonnegative diffusion processes with applications to option pricing. *The Journal of Computational Finance 15*(2), 3.

Li, G. and C. Zhang (2010). On the number of state variables in options pricing. *Management Science 56*(11), 2058–2075.

Lipton, A. and A. Sepp (2011). Filling the gaps. *Risk 24*(10), 78.

Livieri, G., S. Mouti, A. Pallavicini, and M. Rosenbaum (2017). Rough volatility: evidence from option prices. *Available at* SSRN 2914086.

Maheu, J. M. and T. H. McCurdy (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics 160*(1), 69–76.

Majewski, A. A., G. Bormetti, and F. Corsi (2015). Smile from the past: A general option pricing framework with multiple volatility and leverage components. *Journal of Econometrics 187*(2), 521–531.

Mandelbrot, B. B. (1997). The variation of certain speculative prices. In *Fractals and Scaling in Finance*, pp. 371–418. Springer.

Mandelbrot, B. B. and J. W. Van Ness (1968). Fractional brownian motions, fractional noises and applications. *SIAM review 10*(4), 422–437.

Martens, M. (2002). Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets 22*(6), 497–518.

Medvedev, A. and O. Scaillet (2007). Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Review of Financial Studies 20*(2), 427–459.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science 4*(1), 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3*(1-2), 125–144.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics 8*(4), 323–361.

Mitchell, A. R. and D. F. Griffiths (1980). *The finite difference method in partial differential equations*. John Wiley.

Mühle-Karbe, J. and M. Nutz (2011). Small-time asymptotics of option prices and first absolute moments. *Journal of Applied Probability 48*(4), 1003–1020.

Müller, U. A., M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, and J. E. von Weizsäcker (1997). Volatilities of different time resolutionsâ€"analyzing the dynamics of market components. *Journal of Empirical Finance 4*(2), 213–239.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Neuenkirch, A. and T. Shalaiko (2016). The order barrier for strong approximation of rough volatility models. *arXiv preprint arXiv:1606.03854*.

Pagès, G. (2014). Introduction to optimal vector quantization and its applications for numerics. *Available at http://hal.archives-ouvertes.fr/INSMI/hal-01034196*.

Pagès, G., H. Pham, and J. Printems (2004). Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, pp. 253–297. Springer.

Pagès, G. and J. Printems (2003). Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods and Applications 9*(2), 135–165.

Pagés, G. and A. Sagna (2015). Recursive marginal quantization of the Euler scheme of a diffusion process. *Applied Mathematical Finance 22*(5), 463–498.

Pascucci, A. (2011). *PDE and martingale methods in option pricing*. Springer Science & Business Media.

Philipov, A. and M. E. Glickman (2006). Multivariate stochastic volatility via Wishart processes. *Journal of Business & Economic Statistics 24*(3), 313–328.

Pooter, M. d., M. Martens, and D. v. Dijk (2008). Predicting the daily covariance matrix for s&p 100 stocks using intraday data—but which frequency to use? *Econometric Reviews 27*(1-3), 199–229.

Predota, M. (2005). On european and asian option pricing in the generalized hyperbolic model. *European Journal of Applied Mathematics 16*(1), 111–144.

Printems, J. et al. (2005). Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Applications mcma 11*(4), 407–446.

Reghai, A., G. Boya, and G. Vong (2012). Local volatility: Smooth calibration and fast usage. *Available at* SSRN 2008215.

Reiswich, D. and W. Uwe (2012). Fx volatility smile construction. *Wilmott 2012*(60), 58–69.

Ren, Y., D. Madan, and M. Q. Qian (2007). Calibrating and pricing with embedded local volatility models. *Risk Magazine 20*(9), 138.

Renault, E. (1997). Econometric models of option pricing errors. *Econometric Society Monographs 28*, 223–278.

Renault, E. and N. Touzi (1996). Option hedging and implied volatilities in a stochastic volatility model1. *Mathematical Finance 6*(3), 279–302.

Roberts, G. O., A. Gelman, W. R. Gilks, et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability 7*(1), 110–120.

Roberts, G. O., O. Papaspiliopoulos, and P. Dellaportas (2004). Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(2), 369–393.

Rogers, L. C. G. and Z. Shi (1995). The value of an asian option. *Journal of Applied Probability 32*(04), 1077–1088.

Rosenbaum, M. (2008). Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Processes and their Applications 118*(8), 1434–1462.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics 63*(1), 289–306.

Samuelson, P. A. (1965). Rational theory of warrant pricing. *Industrial management review 6*(2), 13–32.

Scharth, M. and M. C. Medeiros (2009). Asymmetric effects and long memory in the volatility of Dow Jones stocks. *International Journal of Forecasting 25*(2), 304–327.

Schwert, G. W. (2011). Stock volatility during the recent financial crisis. *European Financial Management 17*(5), 789–805.

Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative analysis 22*(04), 419–438.

Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies 62*(4), 639–661.

Shephard, N. (1994). Local scale model: state space alternative to integrated garch processes. *Journal of Applied Econometrics 60*, 181–202.

Shephard, N. and M. Pitt (1997). Likeklihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–667.

Shephard, N. and K. Sheppard (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*(2), 197–231.

Shirota, S., T. Hizu, and Y. Omori (2014). Realized stochastic volatility with leverage and long memory. *Computational Statistics & Data Analysis 76*, 618–641.

Sidje, R. B. (1998). Expokit: a software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS) 24*(1), 130–156.

Siyanko, S. (2012). Essentially exact asymptotic solutions for asian derivatives. *European Journal of Applied Mathematics 23*(3), 395–415.

Smith, R. L. and J. E. Miller (1986). A non-gaussian state space model and application to the prediction of records. *Journal of the Royal Statistical Society. Series B 48*, 79–88.

So, M. E. P., K. Lam, and W. K. Li (1998). A stochastic volatility model with Markov switching. *Journal of Business & Economic Statistics 16*(2), 244–253.

So, M. K. (2006). Bayesian analysis of nonlinear and non-Gaussian state space models via multiple-try sampling methods. *Statistics and Computing 16*(2), 125–141.

So, M. K., W. Li, and K. Lam (2002). A threshold stochastic volatility model. *Journal of Forecasting 21*(7), 473–500.

Stentoft, L. (2008a). Option pricing using realized volatility. Working Paper at CREATES, University of Copenhagen.

Stentoft, L. P. (2008b). Option pricing using realized volatility.

Takahashi, M., Y. Omori, and T. Watanabe (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis 53*(6), 2404–2426.

Tanizaki, H. (1996). *Nonlinear Filters*. Spring.

Tankov, P. (2003). *Financial modelling with jump processes*, Volume 2. CRC press.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*(398), 528–540.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance 4*(2), 183–204.

Taylor, S. J. (2007). Modelling financial time series (second edition).

Vecer, J. and M. Xu (2004). Pricing Asian options in a semimartingale model. *Quantitative Finance 4*(2), 170–175.

Vermaak, J., C. Andrieu, A. Doucet, and S. Godsill (2004). Reversible jump Markov Chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Journal of Time Series Analysis 25*(6), 785–809.

Vidoni, P. (1999). Exponential family state space models based on a conjugate latent process. *Journal of the Royal Statistical Society. Series B 61*(1), 213–221.

Walker, H. F. and P. Ni (2011). Anderson acceleration for fixed-point iterations. *SIAM Journal on Numerical Analysis 49*(4), 1715–1735.

Ward, R. C. (1977). Numerical computation of the matrix exponential with accuracy estimate. *SIAM Journal on Numerical Analysis 14*(4), 600–610.

Wilmott, P., J. Dewynne, and S. Howison (1993). *Option pricing: mathematical models and computation*. Oxford financial press.

Wu, L. (2011). Variance dynamics: Joint evidence from options and high-frequency returns. *Journal of Econometrics 160*(1), 280–287.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control 18*(5), 931–955.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales. *Journal of the American Statistical Association 100*(472).

Zvan, R., K. R. Vetzal, and P. A. Forsyth (2000). PDE methods for pricing barrier options. *Journal of Economic Dynamics and Control 24*(11), 1563–1590.