Hedging of time discrete auto-regressive stochastic volatility options

Alexandru Badescu[1], Joan del Castillo[2], and Juan-Pablo Ortega[3]

Abstract

Numerous empirical proofs indicate the adequacy of the time discrete auto-regressive stochastic volatility models introduced by Taylor [Tayl 82, Tayl 86, Tayl 05] in the dynamical description of the log-returns of financial assets. The pricing and hedging of contingent products that use these models for their underlying assets is a complicated task due to the incomplete nature of the corresponding market and the non-observability of the associated volatility process. In this paper we introduce new pricing kernels for this setup and apply two existing volatility filtering techniques available in the literature for these models, namely Kalman filtering and the hierarchical-likelihood approach, in order to implement various pricing and dynamical hedging strategies. An extensive empirical analysis using both historical returns and options data illustrates the advantages of this model when compared with more standard approaches, namely Black-Scholes and GARCH.

Keywords: Stochastic volatility models, ARSV models, hedging techniques, incomplete markets, local risk minimization, Kalman filter, hierarchical-likelihood.

1 Introduction

Ever since Black, Merton, and Scholes introduced their celebrated option valuation formula [Blac 72, Mert 76], much effort has been dedicated to reproduce that result using more realistic stochastic models for the underlying asset. In the discrete time modeling setup, the GARCH parametric family [Engl 82, Boll 86] has been a very popular and successful choice for which the option pricing and hedging problem has been profusely studied; see for example [Duan 95, Hest 00, Bade 11, Chor 12, Orte 12], and references therein. Even though the GARCH specification can accommodate most stylized features of financial returns like leptokurticity, volatility clustering, and autocorrelation of squared returns, there are mathematical relations between some of their moments that impose undesirable constraints on some of the parameter values. For example, a well-known phenomenon [Carn 04] has to do with the relation that links the kurtosis of the process with the parameter that accounts for the persistence of the volatility shocks (the sum of the ARCH and the GARCH parameters, bound to be smaller than one in order to ensure stationarity). This relation implies that for highly volatility persistent time series like the ones observed in practice, the ARCH coefficient is automatically forced to be very small, which is in turn incompatible with having sizeable autocorrelation (ACF) for the squares (the ACF of the squares at lag 1 is linear in this coefficient). This situation aggravates when innovations with fat tails are used in order to better reproduce leptokurticity.

The structural rigidities associated to the finiteness of the fourth moment are of particular importance when using quadratic hedging methods for there is an imperative need to remain in the category of square

---

[1]Department of Mathematics and Statistics, University of Calgary, Calgary, Canada. `abadescu@math.ucalgary.ca`

[2]Universitat Autònoma de Barcelona. E-08192 Barcelona, Spain. `castillo@mat.uab.cat`

[3]Corresponding author. Faculty of Mathematics and Statistics, Universität Sankt Gallen, CH-9000 Sankt Gallen, Switzerland, and Centre National de la Recherche Scientifique (CNRS), France. `Juan-Pablo.Ortega@unisg.ch`

summable objects and finite kurtosis is a convenient way to ensure that (see, for example [Orte 12, Theorem 3.1 (iv)]).

The auto-regressive stochastic volatility (ARSV) models [Tayl 82, Tayl 86, Tayl 05] that are described later on in Section 2 are a parametric family designed in part to overcome the difficulties that we just explained. The defining feature of these models is the fact that the volatility (or a function of it) is determined by an autoregressive process that, unlike the GARCH situation, is exclusively driven by past volatility values and by innovations that may or may not be independent from those driving the returns. The use of additional innovations introduces in the model a huge structural leeway. For example, this time around, the same constraint that ensures stationarity guarantees the existence of finite moments of arbitrary order which is of much value in the context of option pricing and quadratic hedging. Additionally, having two sources of randomness, they resemble more than GARCH the continuous time stochastic volatility models that have been successfully implemented in the literature (see for example [Hull 87, Scot 87, Wigg 87, Stei 91, Hest 93] and references therein, as well as [Flem 03] for a comparison between GARCH and ARSV).

The features in the ARSV prescription that enrich the dynamics come with a price that has to do with the fact that conditional volatilities are not determined by the price process, as it was the case in the GARCH situation. This causes added difficulties at the time of volatility and parameter estimation; in particular, a conditional likelihood cannot be written down in this context (see later on in Section 2.2). This also has a serious impact when these models are used in derivatives pricing and hedging because the existence of additional noise sources driving the dynamics accentuates the incomplete character of the corresponding market and makes more complex the pricing and hedging of contingent products.

The main goal of this paper is showing that appropriate implementations of the local risk minimization pricing and hedging scheme [Foll 86, Foll 91, Schw 01], tailored to the ARSV setup provide a competitive tool in the handling of European style contingent products. As we will see later on, this approach presents several advantages that explain its good performance in a discrete time modeling context. For example, it can be adapted to prescribed changes in the hedging frequency, which regularly happens in real life applications. This feature provides a competitive advantage to our approach when compared with most sensitivity based (delta) hedging methods that are constructed by discretizing a continuous time hedging argument and that lose pertinence as the hedging frequency diminishes.

The implementation of this objective requires two ingredients that constitute the main contributions of the paper. First, we introduce various numerically viable pricing kernels in this context and, second, we combine them with two existing volatility filtering techniques, namely Kalman filtering and the hierarchical-likelihood approach, in order to implement the associated pricing and dynamical hedging strategies.

Concerning the pricing kernels, we recall that this term refers to a probability measure equivalent to the physical one with respect to which the discounted prices are martingales. The expectation of the discounted payoff of a derivative product with respect to any of these equivalent measures yields an arbitrage free price for it. There are two equivalent martingale measures that we will be using. The first one is inspired by the so called Extended Girsanov Principle introduced in [Elli 98]. This principle allows the construction of a martingale measure in a general setup under which the process behaves as its "martingale component" used to do under the physical probability; this is the reason why this measure is sometimes referred to as the mean-correcting martingale measure, a denomination that we will adopt. This construction has been widely used in the GARCH context (see [Bade 11, Orte 12] and references therein) where it admits a particularly simple and numerically efficient expression in terms of the probability density function of the model innovations. In the ARSV case, this feature is not anymore available (see Theorem 2.6 and Remark 2.7). We will hence work with a measure inspired in the predictable situation that, even though does not satisfy the Extended Girsanov Principle, it is still a martingale measure (see Theorem 2.8). In Theorem 2.4 we construct the so called minimal martingale measure $Q_{\min}$ in the ARSV setup. The importance in our context of this measure is given

by the fact that the value process of the local risk-minimizing strategy with respect to the physical measure for a derivative product coincides with its arbitrage free price when using $Q_{\min}$ as a pricing kernel. A concern with $Q_{\min}$ in the ARSV setup is that this measure is in general signed; fortunately, the occurrence of negative Radon-Nikodym derivatives is extremely rare for the usual parameter values that one encounters in financial time series. Consequently, the bias introduced by censoring paths that yield negative Radon-Nikodym derivatives and using $Q_{\min}$ as a well-defined positive measure is hardly noticeable. A point that is worth emphasizing is that even though the value processes obtained when carrying out local risk minimization with respect to the physical and the minimal martingale measures are identical, the hedges are in general *not* the same and consequently so are the hedging errors; this difference is studied in Proposition 3.6.

Regarding the volatility estimation, we already mentioned that for these models this dynamical feature is not determined by the price process and hence becomes a genuinely hidden variable that needs to be estimated separately. In this work we use two different techniques to tackle this problem: first, the Kalman filtering approach advocated in [Harv 94, Harv 96] and second, the so called hierarchical likelihood (h-likelihood) strategy [Lee 96, Lee 06, Cast 08, Lim 11] which, roughly speaking, consists of carrying out a likelihood estimation while considering the volatilities as unobserved parameters that are part of the optimization problem. These two methods should be seen as complementary since, even though they are both adequate and numerically feasible volatility estimation techniques, they are subjected to dissimilar hypotheses. On the one hand, the Kalman based approach requires that the returns dynamics can be formulated as a state-space model but, in exchange, the Kalman filter always yields a minimum variance linear unbiased estimator of the volatilities, no matter how the model innovations are distributed (see [Durb 12, Section 2.2.4]). On the other hand, the h-likelihood technique is not subjected to the rigidities of the state space representation necessary for Kalman and hence can be used for stochastic volatility models with complex link functions, however, volatility estimation in the h-likelihood approach is carried out by finding the extrema of the relevant (conditional) probability functions which requires that the distribution of the model innovations is symmetric and unimodal. References regarding other estimation techniques that we do not use in this paper are [Frid 98, Shep 98, Sand 98, Lies 03, Meye 03, Shim 05].

The paper is organized in four sections. Section 2 contains a brief introduction to the ARSV models and their dynamical features of interest to our study. In particular, it includes subsections that explain the volatility filtering techniques that we will be using and the martingale measures that we mentioned above. The details about the implementation of the local risk minimization strategy are contained in Section 3. Section 4 contains an extensive numerical study that uses a dataset of almost three thousand S&P500 European put option contracts quoted every Wednesday of the years 2012 and 2013. The aim of that section is twofold: first, we document that the presence of a non-zero leverage effect in the ARSV modeling improves the overall hedging performance by around 21%, and secondly, we show that the ARSV models with correlated driving innovations outperforms an asymmetric GARCH model with comparable dynamical features and the Black-Scholes delta hedging scheme.

Conventions and notations: The proofs of all the results in the paper are contained in the appendices in Section 6. Given a filtered probability space $(\Omega, P, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{N}})$ and X, Y two random variables, we will denote by $E_t[X] := E[X|\mathcal{F}_t]$ the conditional expectation, $\operatorname{cov}_t(X, Y) := \operatorname{cov}(X, Y|\mathcal{F}_t) := E_t[XY] - E_t[X]E_t[Y]$ the conditional covariance, and by $\operatorname{var}_t(X) := E_t[X^2] - E_t[X]^2$ the conditional variance. A discrete-time stochastic process $\{X_t\}_{t \in \mathbb{N}}$ is called predictable when X_t is $\mathcal{F}_{t-1}$-measurable, for any $t \in \mathbb{N}$. Given two time steps t_1, t_2 and the stochastic process $\{X_t\}_{t \in \mathbb{N}}$, we define $X_{t_1|t_2} := \mathrm{E}\left[X_{t_1}|\mathcal{F}_{t_2}\right]$.

2 Risk-neutral measures for auto-regressive stochastic volatility (ARSV) processes and their evaluation

In this section we introduce two risk-neutral probability measures for the auto-regressive stochastic volatility processes (ARSV) described in the following paragraphs. These models are used later on to describe the price behavior dynamics of the underlying assets of European style contingent products that we will show how to price and hedge in Section 3 using quadratic methods.

2.1 The ARSV model

The auto-regressive stochastic volatility (ARSV) model was introduced in [Tayl 82, Tayl 86] with the objective of capturing some of the most common stylized features observed in the excess returns of financial time series: volatility clustering, excess kurtosis, and autodecorrelation in the presence of dependence; this last feature can be visualized by noticing that financial log-returns time series exhibit autocorrelation close to zero at, say lag 1, while the autocorrelation of the squared returns is significantly not null. This model was generalized later on in [Harv 96] (see also [Yu 05]) in order to accommodate a conditionally asymmetric behavior that is often found in financial time series and that can be described by saying that falls in stock prices tend to be associated with increases in volatility. This last point is of particular importance in derivative pricing/hedging context since it is connected to the empirically observed asymmetry of option smiles [Rubi 94, Ait 98, Jack 00, Fore 05].

The model. Let S_t be the price at time t of the asset under consideration and $y_t := \log(S_t/S_{t-1})$ the associated log-return. The ARSV model [Tayl 82, Tayl 86, Harv 96] is given by the prescription:

$$\left\{ \begin{array}{rcl} y_t & = & m + \sigma_t \epsilon_t, \\ b_t & = & \gamma + \phi b_{t-1} + w_t, \end{array} \right. \quad \text{with} \quad \left(\begin{array}{c} \epsilon_t \\ w_{t+1} \end{array} \right) \sim \text{IID} \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & \rho \sigma_w \\ \rho \sigma_w & \sigma_w^2 \end{array} \right) \right) \quad (2.1)$$

and where $b_t := \log(\sigma_t^2)$, m and γ are real parameters, and $\phi \in (-1, 1)$. The correlation ρ between $\{\epsilon_t\}$ and $\{w_{t+1}\}$ can be exploited to model the conditionally asymmetric behavior that we discussed in the previous paragraph. Notice that in the ARSV model (2.1), the volatility process $\{\sigma_t\}$ is a non-traded stochastic latent variable that, unlike the situation in GARCH-like models [Engl 82, Boll 86] is not a predictable process that can be written as a function of previous returns and volatilities.

It is easy to prove that the unique stationary returns process induced by (2.1) available in the presence of the constraint $\phi \in (-1, 1)$ is a white noise (the returns have no autocorrelation) with finite moments of arbitrary order. In particular, the unconditional variance σ_b^2 of the stationary process $\{b_t\}$ is given by

$$\sigma_b^2 = \frac{\sigma_w^2}{1 - \phi^2},$$

and if the innovations process $\{(\epsilon_t, w_t)'\}$ is Gaussian then the unconditional variance and kurtosis of the process $\{y_t\}$ are given by

$$\text{var}(y_t) = \text{E}[\sigma_t^2] = \exp \left[\frac{\gamma}{1 - \phi} + \frac{1}{2} \sigma_b^2 \right], \quad \text{and} \quad \text{kurtosis}(y_t) = 3 \exp\left(\sigma_b^2 \right). \quad (2.2)$$

Moreover, it can be shown [Tayl 86] that whenever σ_b^2 is small and/or ϕ is close to one then the autocorrelation $\gamma(h)$ of the squared returns at lag h can be approximated by

$$\gamma(h) \simeq \frac{\exp(\sigma_b^2) - 1}{3 \exp(\sigma_b^2) - 1} \phi^h.$$

The existence of finite moments is particularly important in the context of the quadratic hedging methods that we will use later on in the paper. For example, in Theorem 3.1 of [Orte 12] it is shown that the finiteness of the kurtosis is a sufficient condition for the availability of adequate integrability conditions necessary to carry out pricing and hedging via local risk minimization.

Filtrations and conditional cumulant functions. Let (Ω, P) be the probability space where the model (2.1) has been formulated and let $\mathcal{F}_t$ be the information set generated by the observables $\{S_0, S_1, \ldots, S_t\}$. This statement can be mathematically coded by setting $\mathcal{F}_t := \sigma(S_0, S_1, \ldots, S_t)$, where $\sigma(S_0, S_1, \ldots, S_t)$ is the sigma algebra generated by the prices $\{S_0, S_1, \ldots, S_t\}$ up to time t. As several equivalent probability measures will appear in our discussion, we will refer to P as the physical or historical probability measure.

In the rest of the discussion we will not assume, unless we indicate otherwise, that the innovations process $\{(\epsilon_t, w_{t+1})'\}$ is Gaussian. We define the cumulant and the conditional cumulant functions of $\{\epsilon_t\}$ with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \in \{0, \ldots, T\}}$ as:

$$
\begin{aligned}
L_{\epsilon_t}^P(z) &= \log E^P\left[e^{z\epsilon_t}\right], \qquad z \in \mathbb{R}, \\
K_{\epsilon_t}^P(U) &= \log E^P\left[e^{U\epsilon_t} | \mathcal{F}_{t-1}\right], \qquad \text{with } U \text{ a random variable,}
\end{aligned}
$$

respectively. If the innovations $\{\epsilon_t\}$ are Gaussian, we obviously have $L_{\epsilon_t}^P(z) = z^2/2$. When the random variables U and ϵ_t are P-independent, the following relation between $L_{\epsilon_t}^P$ and $K_{\epsilon_t}^P$, proved in Appendix 6.1, holds.

Lemma 2.1 *Let $U : (\Omega, P) \to \mathbb{R}$ be a random variable independent of ϵ_t and $\mathcal{F}_{t-1} = \sigma(S_0, S_1, \ldots, S_{t-1}) \in \mathcal{F}$. Then*

$$
K_{\epsilon_t}^P(U) = \log E^P\left[e^{L_{\epsilon_t}^P(U)} \mid \mathcal{F}_{t-1}\right]. \tag{2.3}
$$

A quick observation of the dynamical prescription (2.1) shows that the volatility process σ_t of the ARSV model is independent, at any given time step t, of the innovation ϵ_t and of $\mathcal{F}_{t-1} = \sigma(S_0, S_1, \ldots, S_{t-1})$. The Lemma 2.1 hence proves the following corollary.

Corollary 2.2 *Let $\{\sigma_t\}$ be the volatility process associated to the ARSV model defined in (2.1) and let $\mathcal{F}_{t-1} = \sigma(S_0, S_1, \ldots, S_{t-1})$ be the information set determined by the observed prices up to time $t-1$. Then,*

$$
K_{\epsilon_t}^P(\sigma_t) = \log E^P\left[e^{L_{\epsilon_t}^P(\sigma_t)} \mid \mathcal{F}_{t-1}\right]. \tag{2.4}
$$

The cumulant functions that we just introduced are very useful at the time of writing down the conditional means and variances of price changes with respect to the physical probability; these quantities will show up frequently in our developments later on. Before we provide these expressions we introduce the following notation for discounted prices:

$$
\widetilde{S}_t := S_t e^{-rt},
$$

where r denotes the risk-free interest rate annualized to the sampling frequency of the model. The tilde will be used in general for discounted processes. Using this notation and the relation (2.4), it is easy to show that,

$$
\begin{aligned}
E^P\left[\widetilde{S}_t - \widetilde{S}_{t-1} \mid \mathcal{F}_{t-1}\right] &= \widetilde{S}_{t-1}\left(e^{K_{\epsilon_t}^P(\sigma_t)} - 1\right) = \widetilde{S}_{t-1} E^P\left[e^{L_{\epsilon_t}^P(\sigma_t)} \mid \mathcal{F}_{t-1}\right], \tag{2.5} \\
\mathrm{var}\left[\widetilde{S}_t - \widetilde{S}_{t-1} \mid \mathcal{F}_{t-1}\right] &= \widetilde{S}_{t-1}^2\left[e^{K_{\epsilon_t}^P(2\sigma_t)} - e^{2K_{\epsilon_t}^P(\sigma_t)}\right]. \tag{2.6}
\end{aligned}
$$

2.2 Volatility and model estimation

The main complication in the estimation of ARSV models is due to the fact that $\mathcal{F}_t$ does not determine the random variable σ_t and hence makes impossible the writing of a likelihood function in a traditional sense. Many procedures have been developed over the years to go around this difficulty based on different techniques: moment matching [Tayl 86], generalized method of moments [Meli 90, Jacq 94, Ande 96], combinations of quasi-maximum likelihood with the Kalman filter [Harv 94], simulated maximum likelihood [Dani 94], MCMC [Jacq 94] and, more recently, hierarchical-likelihood [Lee 96, Lee 06, Cast 08, Lim 11] (abbreviated in what follows as h-likelihood). An excellent overview of some of these methods is provided in Chapter 11 of the monograph [Tayl 05].

In this paper we will focus on the Kalman and h-likelihood approaches since both are based on numerically efficient estimations and forecasts of the volatility, a point of much importance in our developments. Additionally, as we already mentioned in the introduction, these two methods should be seen as complementary since, even though they are both adequate volatility estimation techniques, they are subjected to dissimilar hypotheses. On the one hand, the Kalman based approach requires that the returns dynamics can be formulated as a state-space model but, in exchange, the Kalman filter always yields a minimum variance linear unbiased estimator of the volatilities, no matter how the model innovations are distributed (see [Durb 12], Section 2.2.4]). On the other hand, the h-likelihood technique is not subjected to the rigidities of the state space representation necessary for Kalman and hence can be used for stochastic volatility models with complex link functions, however, volatility estimation in the h-likelihood approach is carried out by finding the extrema of the relevant (conditional) probability functions which requires that the distribution of the model innovations is symmetric and unimodal.

State space representation and the Kalman approach. This method was introduced in [Harv 94] for models with independent returns and volatility innovations and was generalized later on in [Harv 96] to accommodate the general case in (2.1). The first step in this approach consists of considering the logarithm of the squared returns equation in (2.1), that is,

$$\log(y_t - m)^2 = b_t + \log \epsilon_t^2,$$

or, equivalently,

$$\log(y_t - m)^2 = \omega + b_t + \xi_t, \quad \text{with} \quad \omega := \mathrm{E}\left[\log \epsilon_t^2\right] \quad \text{and} \quad \xi_t := \log \epsilon_t^2 - \mathrm{E}\left[\log \epsilon_t^2\right], \tag{2.7}$$

and to think of this relation as the observation equation in a state space model in which the defining equation for b_t in (2.1), that is, $b_t = \gamma + \phi b_{t-1} + w_t$, is the transition or state equation. Finally, the correlation ρ between the innovations can be made apparent in this representation by reformulating the state-space so that the information about the sign s_t of the observations $y_t - m$ is taken into account. More specifically, let s_t be the random variable that takes the value 1 when $y_t - m$ is positive and -1 otherwise. The state-space form of (2.1) can then be rephrased as:

$$\begin{cases} \log(y_t - m)^2 &= \omega + b_t + \xi_t, \\ b_t &= \gamma + \left(\phi - \frac{\gamma^* s_{t-1}}{\sigma_\xi^2}\right) b_{t-1} + s_{t-1}\left(\mu^* + \frac{\gamma^*}{\sigma_\xi^2}\left(\log(y_{t-1} - m)^2 - \omega\right)\right) w_t^+, \end{cases} \quad \text{where} \tag{2.8}$$

$$\left.\left(\begin{array}{c} \xi_t \\ w_{t+1}^+ \end{array}\right)\right|_{s_t} \sim \mathrm{IID}\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma_\xi^2 & 0 \\ 0 & \sigma_w^2 - \mu^{*\,2} - \frac{\gamma^{*\,2}}{\sigma_\xi^2} \end{array}\right)\right), \tag{2.9}$$

and where σ_ξ^2 is the variance of $\{\xi_t\}$, $\mu^* := \mathrm{E}\left[w_{t+1} \mid s_t = 1\right]$, and $\gamma^* := \mathrm{cov}\left(w_{t+1}\xi_t \mid s_t = 1\right)$. These values, together with $\omega = \mathrm{E}\left[\log \epsilon_t^2\right]$, depend on the specific distribution that is chosen for the innovations. For example, when $\{(\epsilon_t, w_{t+1})'\}$ are bivariate normal, then $\omega = -1.270$, $\sigma_\xi^2 = 4.934$, $\mu^* = .7979\rho\sigma_w$, and $\gamma^* = 1.1061\rho\sigma_w$.

These relations can be used, together with the Kalman filter and the associated quasi-likelihood, to produce estimates for the model parameters γ, ϕ, ρ, and σ_w once a returns sample has been fixed. On the other hand, if a sample and the model parameters are available, the Kalman filter yields an estimate of the forecast $b_{t|t-1} := \mathrm{E}\left[b_t \mid \mathcal{F}_{t-1}\right]$ that will be used later on in the implementation of the proposed option hedging schemes.

The h-likelihood approach. This method [Lee 96, Lee 06, Cast 08, Lim 11] consists of carrying out a likelihood estimation in which the volatilities are considered as unobserved parameters that are part of the optimization problem. The reference [Lim 11] makes explicit the implementation of this technique for ARSV models with independent innovations for the returns and volatility equations and it also considers a generalization to the simultaneously correlated case, that is, for each time step t, there is a dependence between ϵ_t and w_t. This model differs from the one introduced in [Harv 96] that we consider in this paper, in which there is a correlation between ϵ_t and the shifted volatility innovation w_{t+1}. Apart from various empirical reasons for this shift that can be found in [Harv 96, Tayl 05], its presence is of much importance in our setup since it guarantees the independence for any fixed t between the volatility process σ_t and the innovations ϵ_t; this fact is needed for the conclusion of Corollary 2.2 that is used in what follows.

The following paragraphs adapt the h-likelihood approach in [Lim 11] to the case with correlation between the shifted innovations. Let $\boldsymbol{\theta} := (\gamma, \phi, \sigma_w, \rho)$ be the parameters vector that we are interested in estimating, $y := (y_1, \ldots, y_T)$ a vector containing T observed returns, and $b = (b_1, \ldots, b_T)$ the corresponding unobserved log-variances $b_t := \log(\sigma_t^2)$. If $f(y_1, \ldots, y_T, b_1, \ldots, b_T)$ is the joint probability function of y and b, then the associated h-(log)likelihood is defined as

$$h(y; b, \boldsymbol{\theta}) = f(y_1, \ldots, y_T, b_1, \ldots, b_T) = \sum_{t=2}^{T-1} \log f(y_t, b_{t+1}|y_{t-1}, b_t) + \log f(y_T|b_T, y_{T-1}, \ldots, b_2, y_1). \quad (2.10)$$

In order to provide an explicit expression for the h-likelihood, suppose now that the model innovations are Gaussian, that is,

$$\begin{pmatrix} \epsilon_t \\ w_{t+1} \end{pmatrix} \sim \mathrm{IN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_w \\ \rho\sigma_w & \sigma_w^2 \end{pmatrix} \right), \text{ in which case } \begin{cases} \epsilon_t|w_{t+1} & \sim & \mathrm{N}\left(\rho w_{t+1}/\sigma_w, 1-\rho^2\right) \\ w_{t+1}|\epsilon_t & \sim & \mathrm{N}\left(\rho\sigma_w\epsilon_t, \sigma_w^2(1-\rho^2)\right). \end{cases} \quad (2.11)$$

This assumption has two immediate consequences:

$$\begin{aligned}
y_t|b_{t+1}, b_t, w_{t+1} &= (m + \exp(b_t/2)\epsilon_t)\,|w_{t+1}, b_t \sim \mathrm{N}\left(m + \exp(b_t/2)\frac{\rho\left(b_{t+1} - \gamma - \phi b_t\right)}{\sigma_w}, \exp(b_t)(1-\rho^2) \right) \\
b_{t+1}|b_t &= (\gamma + \phi b_t + w_{t+1})\,|b_t \sim \mathrm{N}\left(\gamma + \phi b_t, \sigma_w^2\right).
\end{aligned}$$

Therefore, for any $t \in \{1, \ldots, T\}$, we have

$$\begin{aligned}
\log f(y_t, b_{t+1}|y_{t-1}, b_t) &= \log f(y_t|y_{t-1}, b_{t+1}, b_t) + \log f(b_{t+1}|b_t, y_{t-1}) \\
&= \log f(y_t|b_{t+1}, b_t, w_{t+1}) + \log f(b_{t+1}|b_t) \\
&= -\frac{1}{2}\log\left(2\pi\exp(b_t)(1-\rho^2)\right) - \frac{1}{2\exp(b_t)(1-\rho^2)}\left(y_t - m - \exp(b_t/2)\frac{\rho\left(b_{t+1} - \gamma - \phi b_t\right)}{\sigma_w} \right)^2 \\
&\quad - \log\left(2\pi\sigma_w^2\right) - \frac{1}{2\sigma_w^2}\left(b_{t+1} - \gamma - \phi b_t\right)^2,
\end{aligned} \quad (2.12)$$

and given that $y_t|\,(b_t, y_{t-1}, \ldots, b_2, y_1) = (m + \exp(b_t/2)\epsilon_t)\,|b_t \sim \mathrm{N}\left(m, \exp(b_t)\right)$, we conclude that

$$\log f(y_t|b_t, y_{t-1}, \ldots, b_2, y_1) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}b_t - \frac{(y_t - m)^2}{2\exp(b_t)}. \quad (2.13)$$

Expressions (2.12) and (2.13) substituted in (2.10) yield an explicit expression of the h-likelihood in the Gaussian case.

The smoothing, filtering (updating), and forecasting of the conditional log-variances in the h-likelihood approach to the ARSV model is carried out by assuming that the relevant (conditional) probability functions are symmetric and unimodal and that hence conditional expectations can be computed by finding their extrema. Indeed, in the presence of those hypotheses, given a sample of observed returns $y = (y_1, \ldots, y_T)$ of length T, an estimate of the smoothed conditional log-variances $(b_{1|T}, \ldots, b_{T|T})$, with $b_{t|T} := \mathrm{E}\,[b_t | \mathcal{F}_T]$, is obtained as:

$$(b_{1|T}, \ldots, b_{T|T}) = \underset{b \in \mathbb{R}^T}{\arg\max}\, h(y; b, \boldsymbol{\theta}).$$

The solution of this optimization problem is usually obtained via the score equation $\nabla_b h(y; b, \boldsymbol{\theta}) = 0$ (see [Lim 11] for a numerically efficient technique to solve this relation using sparsity techniques). This equation determines, for the given sample y, a function $b(\boldsymbol{\theta})$ whose values are the smoothed conditional log-variances and that are used to construct the adjusted profile h-likelihood. The maximum of the adjusted profile h-likelihood determines the h-likelihood estimator $\widehat{\boldsymbol{\theta}}$ of the parameters $\boldsymbol{\theta}$ [Lee 96].

Regarding the forecasted and the filtered (updated) values of the conditional log-variances, that is, $b_{t|t-1} := \mathrm{E}\,[b_t | \mathcal{F}_{t-1}]$ and $b_{t|t} := \mathrm{E}\,[b_t | \mathcal{F}_t]$, respectively, these are obtained by maximizing recursively the conditional densities $f(b_t | b_{t-1}, y_{t-1})$ and $f(b_t | y_t, b_{t-1})$. More explicitly, at each time step t, we set

$$
\begin{aligned}
b_{t|t-1} &= \underset{b_t \in \mathbb{R}}{\arg\max}\, f(b_t | b_{t-1|t-1}, y_{t-1}), & (2.14)
\end{aligned}
$$

$$
b_{t|t} = \underset{b_t \in \mathbb{R}}{\arg\max}\, f(b_t | y_t, y_{t-1}, b_{t-1|t-1}) = \underset{b_t \in \mathbb{R}}{\arg\max}\, f(y_t | b_t, y_{t-1}, b_{t-1}) f(y_{t-1}, b_t | b_{t-1}). \quad (2.15)
$$

The last equality is justified by the identity

$$f(y_t, b_t, y_{t-1}, b_{t-1}) = f(y_t | b_t, y_{t-1}, b_{t-1}) f(y_{t-1}, b_t | b_{t-1}) = f(b_t | y_t, y_{t-1}, b_{t-1}) f(y_t, y_{t-1}, b_{t-1}),$$

that shows that $f(b_t | y_t, y_{t-1}, b_{t-1|t-1})$ and $f(y_t | b_t, y_{t-1}, b_{t-1}) f(y_{t-1}, b_t | b_{t-1})$ are proportional with a proportionality constant that does not depend on the variable with respect to which we carry out the optimization, that is b_t. In practice, it is usually the conditional log-densities instead of the densities that are maximized.

We now provide explicit expressions for the optimization problems (2.14) and (2.15) in the Gaussian case. Notice first that in that situation,

$$b_t | b_{t-1}, y_{t-1} = (\gamma + \phi b_{t-1} + w_t) | \epsilon_{t-1}, \quad \text{with} \quad \epsilon_{t-1} = (y_{t-1} - m)/\exp(b_{t-1}/2).$$

Consequently, by (2.11), $b_t | b_{t-1}, y_{t-1} \sim \mathrm{N}\left(\gamma + \phi b_{t-1} + \rho\sigma_w\left((y_{t-1} - m)/\exp(b_{t-1}/2)\right), \sigma_w^2(1 - \rho^2)\right)$ and hence the optimization problem (2.14) is equivalent to:

$$b_{t|t-1} = \underset{b_t \in \mathbb{R}}{\arg\min}\left(b_t - \gamma - \phi b_{t-1|t-1} - \frac{\rho\sigma_w(y_{t-1} - m)}{\exp(b_{t-1|t-1}/2)}\right)^2 = \gamma + \phi b_{t-1|t-1} + \frac{\rho\sigma_w(y_{t-1} - m)}{\exp(b_{t-1|t-1}/2)}.$$

Finally, by the second equality in (2.15) together with (2.12) and (2.13) we conclude that:

$$
\begin{aligned}
b_{t|t} = \underset{b_t \in \mathbb{R}}{\arg\min}\Bigg(& \frac{1}{\exp(b_{t-1|t-1})(1 - \rho^2)}\left(y_{t-1} - m - \exp(b_{t-1|t-1}/2)\frac{\rho\left(b_t - \gamma - \phi b_{t-1|t-1}\right)}{\sigma_w}\right)^2 \\
& + b_t + \frac{(y_t - m)^2}{\exp(b_t)} + \frac{1}{\sigma_w^2}\left(b_t - \gamma - \phi b_{t-1|t-1}\right)^2\Bigg).
\end{aligned}
$$

Remark 2.3 There are other volatility forecasting and estimation techniques available in the literature for models of the form (2.1). A particularly well-known approach that yields excellent performances is presented in [Lies 03]; the technique presented in that paper is representative of a collection of methods based on the use of importance sampling and/or MCMC (Monte Carlo Markov Chains). Their implementation in the pricing/hedging context that we introduce later on in the paper is unfortunately not viable due to the numerical effort it demands. As we show in the next section, prices and hedges are computed via Monte Carlo, and volatilities need to be forecasted for each time step up to maturity for each path generated in the simulation; this circumstance makes preferable Kalman or h-likelihood-based one-step-ahead forecasting techniques that are faster and cheaper when compared to importance sampling related approaches like the one that we just quoted.

2.3   Equivalent martingale measures for stochastic volatility models

Any technique for the pricing and hedging of options based on non-arbitrage theory requires the use of an equivalent measure for the probability space used for the modeling of the underlying asset under which its discounted prices are martingales with respect to the filtration generated by the observables. Measures with this property are usually referred to as risk-neutral or simply martingale measures and the Radon-Nikodym derivative that links this measure with the historical or physical one is called a pricing kernel.

A number of martingale measures have been formulated in the literature in the context of GARCH-like time discrete processes with predictable conditional volatility; see [Bade 11] for a good comparative account of many of them. Those constructions do not generalize to the SV context mainly due to the fact that the volatility process $\{\sigma_t\}$ is not uniquely determined by the price process $\{S_t\}$ and hence it is not predictable with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t\}$ generated by $\{S_t\}$.

In this piece of work we will explore two solutions to this problem that, when applied to option pricing and hedging, yields a good combination of theoretical computability and numerical efficiency. The first one has to do with the so called minimal martingale measure and the second approach is based on an approximation inspired in the Extended Girsanov Principle [Elli 98].

The minimal martingale measure. As we will see in the next section, this measure is particularly convenient when using local risk minimization with respect to the physical measure [Foll 86, Foll 91, Foll 02] as a pricing/hedging technique, for it provides the necessary tools to interpret the associated value process as an arbitrage free price for the contingent product under consideration.

The minimal martingale measure $Q_{\min}$ is a measure equivalent to P defined by the following property: every P-martingale $M \in L^2(\Omega, P)$ that is strongly orthogonal to the discounted price process $\widetilde{S}$, is also a $Q_{\min}$-martingale. The following result spells out the specific form that $Q_{\min}$ takes when it comes to the ARSV models.

Theorem 2.4 *Consider the price process $S = \{S_0, S_1, \ldots, S_T\}$ associated to the ARSV model given by the expression (2.1). In this setup, the minimal martingale measure is determined by the Radon-Nikodym derivative $dQ_{\min}/dP$ that is obtained by evaluating at time T the P-martingale $\{Z_t\}_{t \in \{1, \ldots, T\}}$ defined by*

$$Z_t := \prod_{k=1}^{t} \left(1 + \frac{\left(e^{K_{\epsilon_t}^P(\sigma_k)} - 1\right)\left(e^{\sigma_k \epsilon_k} - e^{K_{\epsilon_t}^P(\sigma_k)}\right)}{e^{2K_{\epsilon_t}^P(\sigma_k)} - e^{K_{\epsilon_t}^P(2\sigma_k)}} \right). \tag{2.16}$$

Proof.   By Corollaries 10.28 and 10.29 and Theorem 10.30 in [Foll 02], the minimal martingale measure, when it exists, is unique and is determined by the Radon-Nikodym derivative obtained by

evaluating at T the P-martingale

$$Z_t := \prod_{k=1}^{t} 1 + \lambda_k \left(\widetilde{Y}_k - \widetilde{Y}_{k-1} \right), \quad \text{where} \quad \lambda_k := -\frac{E_{k-1}^P \left[\widetilde{S}_k - \widetilde{S}_{k-1} \right]}{\mathrm{var}_{k-1} \left(\widetilde{S}_k - \widetilde{S}_{k-1} \right)}, \tag{2.17}$$

and $\widetilde{Y}_k$ is the martingale part in the Doob decomposition of $\widetilde{S}_k$ with respect to P. Hence we have

$$\widetilde{Y}_k - \widetilde{Y}_{k-1} = \widetilde{S}_k - E_{k-1}^P \left[\widetilde{S}_k \right] = \widetilde{S}_{k-1} \left(e^{\sigma_k \epsilon_k} - e^{K_{\epsilon_k}^P(\sigma_k)} \right). \tag{2.18}$$

The equality (2.16) follows from (2.17), (2.18), (2.5), and (2.6). ∎

Remark 2.5 The measure $Q_{\min}$ obtained by using (2.16) is in general signed. Indeed, as the random variable $\sigma_k \epsilon_k$ is $\mathcal{F}_k$-adapted and $K_{\epsilon_k}^P(\sigma_k)$ is $\mathcal{F}_k$-predictable, the term $\left(e^{\sigma_k \epsilon_k} - e^{K_{\epsilon_k}^P(\sigma_k)} \right)$ can take arbitrarily negative values that can force Z_t to become negative. We will see in our numerical experiments that even though this is in general the case, negative occurrences are extremely unlikely for the usual parameter values that one encounters in financial time series. Consequently, the bias introduced by censoring paths that yield negative Radon-Nikodym derivatives and using $Q_{\min}$ as a well-defined positive measure is not noticeable.

The Extended Girsanov Principle and the mean correcting martingale measure. This construction has been introduced in [Elli 98] as an extension in discrete time and for multivariate processes of the classical Girsanov Theorem. This measure is designed so that when the process is considered with respect to it, its dynamical behavior coincides with that of its martingale component under the original historical probability measure. These martingale measures are widely used in the GARCH context (see for example [Bade 11, Orte 12] and references therein).

In this paragraph we will work with models that are slightly more general than (2.1) in the sense that we will allow for predictable trend terms and generalized innovations, that is, our ARSV model will take the form

$$\begin{cases} y_t &= m_t + \sigma_t \epsilon_t, \\ b_t &= \gamma + \phi b_{t-1} + w_t, \end{cases} \quad \text{with} \quad \begin{pmatrix} \epsilon_t \\ w_{t+1} \end{pmatrix} \sim \mathrm{IID} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \sigma_w \\ \rho \sigma_w & \sigma_w^2 \end{pmatrix} \right), \tag{2.19}$$

and where m_t an $\mathcal{F}_{t-1}$-measurable random variable.

Theorem 2.6 *Given the model specified by (2.19), let $f_{\sigma_t \epsilon_t}^P$ be the conditional probability density function under P of the random variable $\sigma_t \epsilon_t$ given $\mathcal{F}_{t-1}$ and let $\{N_t\}$ be the stochastic discount factors defined by:*

$$N_t := \frac{f_{\sigma_t \epsilon_t}^P (\sigma_t \epsilon_t + m_t - r + \log E_{t-1}^P [e^{\sigma_t \epsilon_t}])}{f_{\sigma_t \epsilon_t}^P (\sigma_t \epsilon_t)}. \tag{2.20}$$

The process $Z_t := \prod_{k=1}^{t} N_k$, is a $(\mathcal{F}_t, P)$-martingale such that $E^P[Z_t] = 1$ and Z_T defines in the ARSV setup the martingale measure Q_{EGP} associated to Extended Girsanov Principle via the identity $Z_T = dQ_{\mathrm{EGP}}/dP$.

Remark 2.7 As σ_t and ϵ_t are independent processes, we can write (see [Roha 76]):

$$f_{\sigma_t \epsilon_t}^P(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{\sigma_t}^P(x) f_{\epsilon_t}^P \left(\frac{z}{x} \right) dx.$$

Unfortunately, there is no closed form expression for this integral even in the Gaussian setup. Though *ad hoc* numerical have been developed to treat it (see for instance [Glen 04]), the computation of the stochastic discount factors N_t for each time step t may prove to be a computationally heavy task.

The point that we just raised in the remark leads us to introduce yet another martingale measure inspired by the analytical expression that takes the martingale measure associated to the Extended Girsanov Principle for GARCH models. Indeed, in that situation, that measure is determined by the stochastic discount factor (see for example expression (2.7) in [Bade 11]):

$$N_t(\epsilon_t, \rho_t) = \frac{f_{\epsilon_t}^P(\epsilon_t + \rho_t)}{f_{\epsilon_t}^P(\epsilon_t)}, \qquad (2.21)$$

where $f_{\epsilon_t}^P$ is the conditional probability density function of the innovations $\{\epsilon_t\}$ with respect to the measure P given $\mathcal{F}_{t-1}$ and ρ_t is the **market price of risk** process defined by:

$$\rho_t := \frac{m_t + K_{\epsilon_t}^P(\sigma_t) - r}{\sigma_t}. \qquad (2.22)$$

The stochastic discount factor (2.21) is different from the one in (2.20) and consequently the measure that it defines is obviously not the martingale measure associated to the Extended Girsanov Principle for discounted ARSV based prices. Nevertheless, as we show in the next theorem whose proof can be found in the appendix, this discount factor still defines an equivalent martingale measure. We emphasize that the economic foundations for the use of the Extended Girsanov Principle do not apply to the one that we just introduced, in particular this change of measure is not consistent with the minimization of the conditional expectation of squared discounted risk adjusted hedging costs (see [Elli 98, Theorem 4.2]).

Theorem 2.8 *Using the notation that we just introduced, we have that*

(i) *The process $Z_t := \prod_{k=1}^t N_k(\epsilon_k, \rho_k)$, is a $(\mathcal{F}_t, P)$-martingale such that $E^P[Z_t] = 1$.*

(ii) *Z_T defines an equivalent measure Q_{mc} such that $Z_T = dQ_{\text{mc}}/dP$ under which the discounted price process $\{S_0, S_1, \ldots, S_T\}$ determined by (2.19) is a martingale. By an abuse of language we will refer to Q_{mc} as the ARSV* **mean correcting martingale measure**.

Remark 2.9 The mean correcting martingale measure takes a particularly simple form in the Gaussian case. Indeed, in that situation $L_{\epsilon_t}^P(z) = z^2/2$, and hence by Lemma 2.1 we have that

$$K_{\epsilon_t}^P(\sigma_t) = \log E_{t-1}^P\left[e^{\frac{\sigma_t^2}{2}}\right].$$

Moreover, it is easy to see that the stochastic discount factor (2.21) equals in this case

$$N_t(\epsilon_t, \rho_t) = \exp\left[-\frac{1}{2}\rho_t\left(\rho_t + 2\epsilon_t\right)\right].$$

2.4 Numerical implementation of the martingale measures

The computation of the martingale measures that we just introduced requires one step ahead forecasts $\sigma_{t|t-1}$ of the volatilities σ_t as well as evaluations of the conditional cumulant function $K_{\epsilon_t}^P(\sigma_t) = \log E^P\left[e^{\sigma_t \epsilon_t}|\mathcal{F}_{t-1}\right]$ introduced in (2.16) at those volatility values. In Section 2.2 we have introduced the Kalman and h-likelihood techniques for the filtering and the forecasting of the conditional log-variances b_t. Both approaches yield estimates for the random variables $b_{t|t}$ and $b_{t|t-1}$.

As to the conditional cumulant function $K_{\epsilon_t}^P(\sigma_t) = \log E^P\left[e^{\sigma_t \epsilon_t}|\mathcal{F}_{t-1}\right]$, its exact evaluation is difficult and there are no closed form expressions for it, even in the Gaussian case. This explains why, in the empirical study that we will carry out later on in Section 4, we will use the plug-in estimator that we now

describe. First, we approximate $\sigma_{t|t-1}$ by setting $\sigma_{t|t-1} := \exp(b_{t|t-1}/2)$. Second, the Lemma 2.1 and the Corollary 2.2 allow us to rewrite $K^P_{\epsilon_t}(\sigma_t)$ as the conditional expectation of the function $\exp(L^P_{\epsilon_t}(\sigma_t))$ that depends only on σ_t. Hence, if $\sigma_{t|t-1}$ is the estimation of the forecasted volatility coming either from Kalman or h-likelihood, we will approximate $K^P_{\epsilon_t}(\sigma_t)$ by $L^P_{\epsilon_t}(\sigma_{t|t-1})$. In particular, if the innovations $\{\epsilon_t\}$ are Gaussian and we hence have $L^P_{\epsilon_t}(z) = z^2/2$, we will set

$$K^P_{\epsilon_t}(\sigma_t) \simeq (\sigma_{t|t-1})^2/2. \tag{2.23}$$

In order to assess the quality of these volatility forecasts and of the approximations of the conditional cumulant function that we just described, we have conducted some simulations a price generating Gaussian ARSV process with the following parameters:

$$m = 0.1/252, \qquad \gamma = -0.821, \qquad \phi = 0.9, \qquad \sigma_w = 0.675, \qquad \rho = 0, \tag{2.24}$$

which produces extremely volatile and leptokurtic series. We have used the corresponding model to generate 1200 paths containing 1000 time steps each, with which we have computed the mean absolute percentage error committed in the volatility forecasts and in the conditional cumulant function evaluations when using the Kalman and the h-likelihood techniques. In the case of the volatility forecast, the error is computed by comparing with the actual volatility associated to the model. Regarding the conditional cumulant function evaluation, we compare for each path and each of its time steps the approximations in (2.23) with a Monte Carlo evaluation of

$$K^P_{\epsilon_t}(\sigma_t) = \log E^P_{t-1}\left[e^{\frac{\sigma_t^2}{2}}\right].$$

using 10^5 paths. The results of this exercise are reported in Table 2.1, which shows that the Kalman based technique performs slightly better than h-likelihood and that the errors committed in the evaluation of $K^P_{\epsilon_t}(\sigma_t)$ by using the plug-in estimators (2.23) are of the same magnitude as the volatility forecasting error. Additionally, we show the time needed to estimate the volatilities associated to path with 2500 time steps via both techniques using the same computer and comparable code for both. Again, Kalman performs slightly better.

The graphs in Figure 2.1 show the real and the forecasted daily volatilities associated to a given price path using the Kalman filter and the h-likelihood method. The bottom panel compares the evaluation of $K^P_{\epsilon_t}(\sigma_t)$ at each time step using a Monte Carlo estimation that uses the real unobserved volatility values with those obtained via the plug-in estimators in (2.23).

Method	Comparison of performances between Kalman and h-likelihood		
	Volatility forecasting error **(%/100)**	$K^P_{\epsilon_t}(\sigma_t)$ **evaluation error** **(%/100)**	**Numerical effort** **(seconds)**
Kalman	0.4332 (0.0284)	0.5964 (0.0294)	9.52
h-likelihood	0.5054 (0.0462)	0.7033 (0.0791)	11.69

Table 2.1: Average percentage errors committed by the Kalman and the h-likelihood based techniques in the estimation of volatility forecasts and in the evaluation of the conditional cumulant function. The figures reported correspond to Monte Carlo estimates using paths of the ARSV model with the parameters in (2.24). Both the volatility forecasts and the conditional cumulant function are produced using the plug-in estimator described in the text.
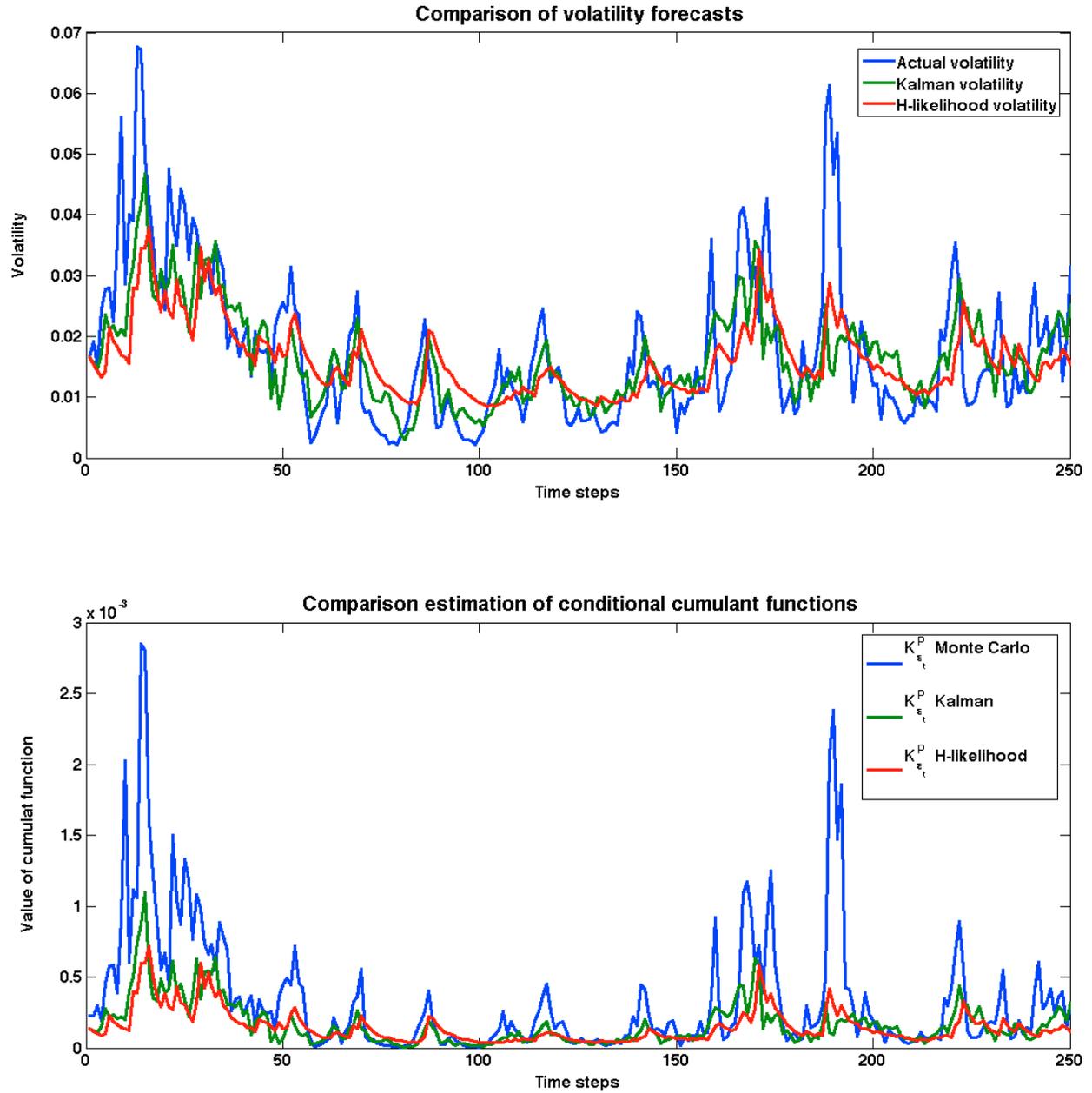
Figure 2.1: Illustration of volatility and conditional cumulant function evaluation paths produced by the ARSV model with the parameters in (2.24) and their estimates using the plug-in Kalman and h-likelihood based estimators described in the text. The top figure shows the real and forecasted daily volatilities associated to a given price path using the Kalman filter and the h-likelihood method. The one at the bottom compares the evaluation of $K^P_{\epsilon_t}(\sigma_t)$ at each time step using a Monte Carlo estimation that uses the real unobserved volatility values with those obtained via the plug-in estimators in (2.23).

3 Local risk minimization for ARSV options

In this section we use the local risk minimization pricing/hedging technique developed by Föllmer, Schweizer, and Sondermann (see [Foll 86, Foll 91, Schw 01], and references therein) as well as the different measures and volatility estimation techniques introduced in the previous section to come up with prices and hedging ratios for European options that have an ARSV process as model for the underlying asset.

As we will see, this technique provides simultaneous expressions for prices and hedging ratios that, even though require Monte Carlo computations in most cases, admit convenient interpretations based on the notion of hedging error minimization that can be adapted to various models for the underlying asset. This hedging approach has been studied in [Orte 12] in the context of GARCH models with Gaussian innovations and in [Bade 14] for more general innovations. Additionally, the technique can be tuned in order to accommodate different prescribed hedging frequencies and hence adapts very well to realistic situations encountered by practitioners. In the negative side, like any other quadratic method, local risk minimization penalizes equally shortfall and windfall hedging errors, which may sometimes lead to inadequate hedging decisions.

3.1 Generalized trading strategies and local risk minimization

We now briefly review the necessary concepts on pricing by local risk minimization that are needed in the sequel. The reader is encouraged to check with Chapter 10 of the monograph [Foll 02] for a self-contained and comprehensive presentation of the subject.

As we have done so far, we will denote by S_t the price of the underlying asset at time t. The symbol r_t denotes the continuously composed risk-free interest rate paid on the currency of the underlying in the period that goes from time $t-1$ to t; we will assume that $\{r_t\}$ is a predictable process. Denote by

$$R_t := \sum_{j=0}^{t} r_j.$$

The price at time t of the riskless asset S^0 such that $S_0^0 = 1$, is given by $S_t^0 = \mathrm{e}^{R_t}$.

Let now $H(S_T)$ be a European contingent claim that depends on the terminal value of the risky asset S_t. In the context of an incomplete market, it will be in general impossible to replicate the payoff H by using a self-financing portfolio. Therefore, we introduce the notion of generalized trading strategy, in which the possibility of additional investment in the riskless asset throughout the trading periods up to expiry time T is allowed. All the following statements are made with respect to a fixed filtered probability space $(\Omega, P, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\{0,\ldots,T\}})$.

A generalized trading strategy is a pair of stochastic processes (ξ^0, ξ) such that $\{\xi_t^0\}_{t\in\{0,\ldots,T\}}$ is adapted and $\{\xi_t\}_{t\in\{1,\ldots,T\}}$ is predictable. The associated value process V of (ξ^0, ξ) is defined as

$$V_0 := \xi_0^0, \quad \text{and} \quad V_t := \xi_t^0 \cdot S_t^0 + \xi_t \cdot S_t, \quad t \geq 1.$$

The gains process G of the generalized trading strategy (ξ^0, ξ) is given by

$$G_0 := 0 \quad \text{and} \quad G_t := \sum_{k=1}^{t} \xi_k \cdot (S_k - S_{k-1}), \quad t = 0, \ldots, T.$$

The cost process C is defined by the difference

$$C_t := V_t - G_t, \quad t = 0, \ldots, T.$$

All these processes have discounted versions $\widetilde{V}_t$, $\widetilde{G}_t$, and $\widetilde{C}_t$ defined as:

$$\widetilde{V}_t := V_t e^{-R_t}, \qquad \widetilde{G}_t := \sum_{k=1}^{t} \xi_k \cdot (\widetilde{S}_k - \widetilde{S}_{k-1}), \quad \text{and} \quad \widetilde{C}_t := \widetilde{V}_t - \widetilde{G}_t.$$

Assume now that both H and the $\{S_n\}_{n \in \{0,\ldots,T\}}$ are in $L^2(\Omega, P)$. A generalized trading strategy is called admissible for H whenever it is in $L^2(\Omega, P)$ and its associated value process is such that

$$V_T = H, \quad \text{P } a.s., \qquad V_t, G_t \in L^2(\Omega, P), \quad \text{for each } t.$$

The hedging technique via local risk minimization consists of finding the strategies $(\widehat{\xi}^0, \widehat{\xi})$ that minimize the local risk process

$$R_t(\xi^0, \xi) := E^P \left[(\widetilde{C}_{t+1} - \widetilde{C}_t)^2 \mid \mathcal{F}_t \right], \quad t = 0, \ldots, T - 1, \tag{3.1}$$

within the set of admissible strategies (ξ^0, ξ). More specifically, the admissible strategy $(\widehat{\xi}^0, \widehat{\xi})$ is called local risk-minimizing if

$$R_t(\widehat{\xi}^0, \widehat{\xi}) \leq R_t(\xi^0, \xi), \quad \text{P } a.s.$$

for all t and each admissible strategy (ξ^0, ξ). It can be shown that [Foll 02, Theorem 10.9] an admissible strategy is local risk-minimizing if and only if the discounted cost process is a P-martingale and it is strongly orthogonal to $\widetilde{S}$, in the sense that $\text{cov}_t(\widetilde{S}_{t+1} - \widetilde{S}_t, \widetilde{C}_{t+1} - \widetilde{C}_t) = 0$, P-a.s., for any $t = 0, \ldots, T - 1$. More explicitly, whenever a local risk-minimizing technique exists, it is fully determined by the backwards recursions:

$$V_T = H, \tag{3.2}$$

$$\xi_{t+1} = \frac{\text{cov}^P(\widetilde{V}_{t+1}, \widetilde{S}_{t+1} - \widetilde{S}_t \mid \mathcal{F}_t)}{\text{var}^P(\widetilde{S}_{t+1} - \widetilde{S}_t \mid \mathcal{F}_t)}, \tag{3.3}$$

$$\widetilde{V}_t = E^P \left[\widetilde{V}_{t+1} \mid \mathcal{F}_t \right] - \xi_{t+1} E^P \left[\left(\widetilde{S}_{t+1} - \widetilde{S}_t \right) \mid \mathcal{F}_t \right]. \tag{3.4}$$

The initial investment V_0 determined by these relations will be referred to as the local risk minimization option price associated to the measure P.

3.2 Local risk minimization with respect to a martingale measure

As we have explicitly indicated in expressions (3.2)–(3.4), the option values and the hedging ratios that they provide depend on a previously chosen probability measure and a filtration. In the absence of external sources of information, the natural filtration $\{\mathcal{F}_t\}$ to be used is the one generated by the price process. Regarding the choice of the probability measure, the first candidate that should be considered is the physical or historical probability associated to the price process since from a risk management perspective this is the natural measure that should be used in order to construct the local risk (3.1).

In practice, the use of the physical probability encounters two major difficulties: on the one hand, when (3.2)–(3.4) are written down with respect to this measure, the resulting expressions are convoluted and numerically difficult to estimate due to the high variance of the associated Monte Carlo estimators; a hint of this difficulty in the GARCH situation can be seen in Proposition 2.5 of [Orte 12]. On the other hand, unless the discounted prices are martingales with respect to the physical probability measure or there is a minimal martingale measure available, the option prices that result from this technique cannot be interpreted as arbitrage free prices.

These reasons lead us to explore the local risk minimization strategy for martingale measures and, more specifically, for the martingale measures introduced in Section 2.3. Given that the trend terms that separate the physical measure from being a martingale measure are usually very small when dealing with daily or weekly financial returns, it is expected that the inaccuracy committed by using a conveniently chosen martingale measure will be smaller than the numerical error that we would face in the Monte Carlo evaluation of the expressions corresponding to the physical measure; see [Orte 12, Proposition 3.2] for an argument in this direction in the GARCH context. The next section is dedicated to an empirical comparison between the hedging performances obtained using the different measures in Section 2.3, as well as the various volatility estimation techniques that we described and that are necessary to implement them.

The following proposition is proved using a straightforward recursive argument in expressions (3.2)–(3.4) combined with the martingale hypothesis in its statement.

Proposition 3.1 *Let Q be an equivalent martingale measure for the price process $\{S_t\}$ and $H(S_T)$ be a European contingent claim that depends on the terminal value of the risky asset S_t. The local risk minimizing strategy with respect to the measure Q is determined by the recursions:*

$$V_T = H, \tag{3.5}$$

$$\xi_{t+1} = \frac{1}{\Sigma_{t+1}^2} E_t^Q \left[e^{-(R_T+R_t)} H(S_T) \left(S_{t+1} e^{-r_{t+1}} - S_t \right) \right], \tag{3.6}$$

$$V_t = E_t^Q \left[e^{-(R_T-R_t)} H(S_T) \right], \tag{3.7}$$

where

$$\Sigma_{t+1}^2 := \mathrm{var}^Q(\widetilde{S}_{t+1} - \widetilde{S}_t \mid \mathcal{F}_t) = e^{-2R_t} E_t^Q \left[S_{t+1}^2 e^{-2r_{t+1}} - S_t^2 \right] = e^{-2R_t} \mathrm{var}^Q(S_{t+1} e^{-r_{t+1}} \mid \mathcal{F}_t). \tag{3.8}$$

Remark 3.2 When expression (3.7) is evaluated at $t = 0$ it yields the initial investment V_0 necessary to setup the generalized local risk minimizing trading strategy that fully replicates the derivative H and coincides with the arbitrage free price for H that results from using Q as a pricing measure. Obviously, this connection only holds when local risk minimization is carried out with respect to a martingale measure.

Remark 3.3 Local risk-minimizing trading strategies computed with respect to a martingale measure Q also minimize [Foll 02, Proposition 10.34] the so called **remaining conditional risk**, defined as the process $R_t^Q(\xi^0, \xi) := E_t[(\widetilde{C}_T - \widetilde{C}_t)^2]$, $t = 0, \ldots, T$; this is in general not true outside the martingale framework (see [Schw 01, Proposition 3.1] for a counterexample). Analogously, local risk minimizing strategies are also **variance-optimal**, that is, they minimize $E^Q\left[\left(\widetilde{H} - V_0 - \widetilde{G}_T\right)^2\right]$ (see [Foll 02, Proposition 10.37]). This is particularly relevant in the ARSV context in which a standard sufficient condition (see [Foll 02, Theorem 10.40]) that guarantees that local risk minimization with respect to the physical measure implies variance optimality does not hold; we recall that this condition demands for the deterministic character of the quotient

$$\beta_t := \frac{\left(E_{t-1}^P [S_t - S_{t-1}]\right)^2}{\mathrm{var}_{t-1}^P [S_t - S_{t-1}]}.$$

Indeed, expressions (2.5) and (2.6), together with Lemma 2.3 imply that in our situation

$$\beta_t := \frac{e^{2K_{\epsilon_t}^P(\sigma_t)}}{\left[e^{K_{\epsilon_t}^P(2\sigma_t)} - e^{2K_{\epsilon_t}^P(\sigma_t)}\right]},$$

which is obviously not deterministic.

Remark 3.4 The last equality in (3.8) is of much importance when using Monte Carlo simulations to evaluate (3.5)–(3.7) and suggests the correct estimator that must be used in practice: if we generate under the measure Q a set of N price paths all of which start at S_t at time t and take values $S_{t+1}^1, \ldots, S_{t+1}^N$ at time $t+1$, then in view of the last equality in (3.8) we write

$$\Sigma_{t+1}^2 = e^{-2R_t} \mathrm{var}^Q(S_{t+1}e^{-r_{t+1}} \mid \mathcal{F}_t) \simeq \frac{e^{-2R_t}}{N} \sum_{i=1}^N \left(S_{t+1}^i e^{-r_{t+1}} - S_t\right)^2.$$

We warn the reader that an estimator for Σ_{t+1}^2 based on the Monte Carlo evaluation of

$$e^{-2R_t} E_t^Q \left[S_{t+1}^2 e^{-2r_{t+1}} - S_t^2\right]$$

would have much more variance and once inserted in the denominator of (3.6) would produce unacceptable results.

Remark 3.5 The Monte Carlo estimation of the expressions (3.5)–(3.7) requires the generation of price paths with respect to the martingale measure Q. In the GARCH context this can be easily carried out for a variety of pricing kernels by rewriting the process in connection with the new equivalent measure in terms of new innovations (see for example [Duan 95, Orte 12, Chor 12, Bade 11]). This method can be combined with modified Monte Carlo estimators that enforce the martingale condition and that have a beneficial variance reduction effect like, for example, the Empirical Martingale Simulation technique introduced in [Duan 98, Duan 01]. Unfortunately, this is difficult to carry out in our context. Given that all the pricing measures introduced in the previous section are constructed out of the physical one, the best option in our setup consists of carrying out the path simulation with respect to the physical measure and computing the expectations in (3.5)–(3.7) using the corresponding Radon-Nikodym derivative. More specifically, let Q be an equivalent martingale measure that is obtained out of the physical measure P by constructing a Radon-Nikodym derivative of the form:

$$\frac{dQ}{dP} = \prod_{k=1}^T N_k,$$

such that the process $\{Z_t\}$ defined by $Z_t := \prod_{k=1}^t N_k$ is a P-martingale that satisfies $E^P[Z_t] = 1$; notice that all the measures introduced in the previous section are of that form. In that setup, we can easily rewrite Q-expectations in terms of P-expectations since for any $\mathcal{F}_T$ measurable function f and any $t \in \{0, \ldots, T\}$:

$$E^Q[f \mid \mathcal{F}_t] = E^P[N_t f \mid \mathcal{F}_t]. \tag{3.9}$$

The proof of this equality is a straightforward consequence of [Foll 02, Proposition A.11] and of the specific way in which the measure Q is constructed.

3.3 Local risk minimization and changes in the hedging frequency

As we already pointed out one of the major advantages of the local risk minimization hedging scheme when compared to other sensitivity based methods, is its adaptability to prescribed changes in the hedging frequency. Indeed, suppose that the life of the option H with maturity in T time steps is partitioned into identical time intervals of duration j; this assumption implies the existence of an integer k such that $kj = T$.

We now want to set up a local risk minimizing replication strategy for H in which hedging is carried out once every j time steps. We will denote by ξ_{t+j} the hedging ratio at time t that presupposes that

the next hedging will take place at time $t + j$. The value of such ratios will be obtained by minimizing the j-spaced local risk process:

$$R_t^j(\xi^0, \xi) := E^P\left[(\widetilde{C}_{t+j}^j - \widetilde{C}_t^j)^2 \mid \mathcal{F}_t\right], \quad t = 0, j, 2j, \ldots, (k-1)j = T - j,$$

where $\{\widetilde{C}_t^j\}$ is a cost process constructed out of value and gains processes, $\{V_t^j\}$ and $\left\{G_t^j\right\}$ that only take into account the prices of the underlying assets at time steps $t = 0, j, 2j, \ldots, kj = T$, in particular, given an integer l such that $t = lj$

$$\widetilde{G}_t^j := \sum_{r=1}^{l} \xi_{rj} \cdot (\widetilde{S}_{rj} - \widetilde{S}_{(r-1)j}).$$

A straightforward modification of the argument in [Foll 02, Theorem 10.9] proves that the solution of this local risk minimization problem with modified hedging frequency with respect to a martingale measure is given by the expressions:

$$V_T^j \quad = \quad H, \tag{3.10}$$

$$\xi_{t+j} \quad = \quad e^{-(R_T - R_t)} \frac{E_t^Q\left[H(S_T)\left(S_{t+j}e^{-(R_{t+j} - R_t)} - S_t\right)\right]}{E_t^Q\left[S_{t+j}^2 e^{-2(R_{t+j} - R_t)} - S_t^2\right]}, \tag{3.11}$$

$$V_t^j \quad = \quad E_t^Q\left[e^{-(R_T - R_t)} H(S_T)\right], \tag{3.12}$$

for any $t = 0, j, 2j, \ldots, (k-1)j = T - j$. As a follow up to what we pointed out in the remark 3.4, the denominator in (3.11) should be computed by first noticing that

$$E_t^Q\left[S_{t+j}^2 e^{-2(R_{t+j} - R_t)} - S_t^2\right] = \text{var}_t^Q\left[S_{t+j}e^{-(R_{t+j} - R_t)}\right],$$

and then using the appropriate Monte Carlo estimator for the variance, that is,

$$\text{var}^Q(S_{t+1}e^{-r_{t+1}} \mid \mathcal{F}_t) \simeq \frac{1}{N} \sum_{i=1}^{N} \left(S_{t+j}^i e^{-(R_{t+j}^i - R_t)} - S_t\right)^2, \tag{3.13}$$

where $S_{t+j}^1, \ldots, S_{t+j}^N$ are N realizations of the price process under Q at time $t + j$, using paths that have all the same origin S_t at time t. If the interest rate process $\{r_t\}$ is constant and equal to r, then (3.13) obviously reduces to

$$\text{var}^Q\left[S_{t+j}e^{-r_{t+1}} \mid \mathcal{F}_t\right] \simeq \frac{1}{N} \sum_{i=1}^{N} \left(S_{t+j}^i e^{-jr} - S_t\right)^2.$$

3.4 Local risk minimization and the minimal martingale measure

Local risk-minimization requires picking a particular probability measure in the problem. We also saw that it is only when discounted prices are martingales with respect to that measure that the obtained value process for the option in question coincides with the arbitrage free price that one would obtain by using that martingale measure as a pricing kernel. This observation particularly concerns the physical probability which one would take as the first candidate to use this technique since it is the natural measure to be used when quantifying the local risk.

The solution of this problem is one of the motivations for the introduction of the minimal martingale measure $Q_{\min}$ that we defined in Section 2.3. Indeed, it can be proved that (see Theorem 10.22

in [Foll 02]) the value process of the local risk-minimizing strategy *with respect to the physical measure* coincides with arbitrage free price for H obtained by using $Q_{\min}$ as a pricing kernel. More specifically, if V_t^P is the value process for H obtained out of formulas (3.2)–(3.4) using the physical measure then

$$V_t^P = E_t^{Q_{\min}}\left[e^{-(R_T - R_t)}H(S_T)\right]. \tag{3.14}$$

We emphasize that, as we pointed out in Remark 2.5, the minimal martingale measure in the ARSV setup is in general signed. Nevertheless, the occurrences of negative Radon-Nikodym derivatives are extremely unlikely, at least when dealing with Gaussian innovations, which justifies its use for local risk minimization.

We also underline that even though the value processes obtained when carrying out local risk minimization with respect to the physical and the minimal martingale measures are identical, the hedges are in general *not* the same and consequently so are the hedging errors. In the next proposition we provide a relation that links both hedges and identify situations in which they coincide. Before we proceed we need to introduce the notion of global (hedging) risk process: it can be proved that once a probability measure P has been fixed, if there exists a local risk-minimizing strategy (ξ^0, ξ) with respect to it, then it is unique (see [Foll 02, Proposition 10.9]) and the discounted payoff $\widetilde{H}$ can be decomposed as (see [Foll 02, Corollary 10.14])

$$\widetilde{H} = V_0 + \widetilde{G}_T + \widetilde{L}_T, \tag{3.15}$$

with $\widetilde{G}_t$ the discounted gains process associated to (ξ^0, ξ) and $\widetilde{L}_t := \widetilde{C}_t - C_0$, $t = 0, \dots, T$ a sequence $\{\widetilde{L}_t\}_{t \in \{0,\dots,T\}}$ that we will call the (discounted) global (hedging) risk process. $\{\widetilde{L}_t\}_{t \in \{0,\dots,T\}}$ is a square integrable P-martingale that satisfies $L_0 = 0$ and that is strongly orthogonal to $\widetilde{S}$ in the sense that

$$\operatorname{cov}^P((\widetilde{L}_{t+1} - \widetilde{L}_t)(\widetilde{S}_{t+1} - \widetilde{S}_t) \mid \mathcal{F}_t) = 0 \quad \text{for any} \quad t = 0, \dots, T - 1.$$

The decomposition (3.15) shows that $\widetilde{L}_T$ measures how far $\widetilde{H}$ is from the terminal value of the self-financing portfolio uniquely determined by the initial investment V_0 and the trading strategy given by $\{\xi_t\}$ (see [Lamb 08, Proposition 1.1.3]).

Proposition 3.6 *Let $H(S_T)$ be a European contingent claim that depends on the terminal value of the risky asset S_t and let $\{\xi_t^{Q_{\min}}\}_{t=1}^T$ and $\{\xi_t^P\}_{t=1}^T$ be the local risk minimizing hedges associated to the minimal martingale measure $Q_{\min}$ and the physical measure P, respectively. Let $\left\{\widetilde{L}_t^P\right\}_{t=0}^T$ be the associated P-global risk process. Then, for any $t \in \{1, \dots, T\}$:*

$$\xi_t^{Q_{\min}} = \xi_t^P + \frac{E_{t-1}^{Q_{\min}}\left[\widetilde{L}_t^P(\widetilde{S}_t - \widetilde{S}_{t-1})\right]}{\operatorname{var}_{t-1}^{Q_{\min}}\left[\widetilde{S}_t - \widetilde{S}_{t-1}\right]}. \tag{3.16}$$

If the processes $\left\{\widetilde{L}_t^P\right\}_{t=0}^T$ and $\left\{\widetilde{S}_t\right\}_{t=0}^T$ are either $Q_{\min}$–strongly orthogonal or $Q_{\min}$–independent, then $\{\xi_t^{Q_{\min}}\}_{t=1}^T = \{\xi_t^P\}_{t=1}^T$.

4 Empirical analysis

The goal of this section is comparing the hedging performances obtained by implementing the local risk minimization scheme using the different volatility estimation techniques and martingale measures introduced in Section 2. The exercise consists of hedging a dataset of almost three thousand European

puts on the S&P500 index, quoted every Wednesday of the years 2012 and 2013 and with different maturities and moneyness. The empirical assessment is carried out by comparing the results with those obtained using a GARCH model with comparable dynamical features and the Black-Scholes deltas. The necessary model parameters and volatility estimates are obtained using historical information coming exclusively from the asset returns. We leave for a forthcoming paper the implementation of ad hoc techniques along the lines presented in [Bade 14] that improve the performance of the hedging setup by using information on option prices.

The options dataset. The empirical pricing performance is tested using a dataset of 2,728 S&P500 European put options obtained from OptionMetrics, whose prices were quoted all the Wednesdays during the period spanning January 1st, 2012–December 31st, 2013. The dataset comprises contracts with maturities between 20 and 250 days and moneyness between 0.95 and 1.1. In order to only use significant contracts, we applied various filters similar to those introduced in Bakshi *et al.* [Baks 97]. The basic features of these two datasets, including the number of contracts, average prices, and implied volatilities are reported in Table 4.2 for the same array of maturities and moneyness intervals that is used later on to present the hedging performance results. The global average price and implied volatility for the dataset are $ 53.093 and 16.8%, respectively.

Models and estimation. We use two ARSV models of the type in (2.1) with Gaussian innovations and constant drift structure. In the first one, we do not contemplate the possibility of a conditionally asymmetric behavior in the returns and we set the correlation parameter ρ equal to zero. For the second model, this parameter is allowed to be non-zero and the determination of its numerical value is part of the estimation process. Model parameters are estimated using the Kalman and the h-likelihood based methods introduced in Section 2.2. Estimation is carried out only twice for each model at the beginning of each of the two years 2012 and 2013 by using the historical returns of the underlying asset (the S&P500 index) quoted in the preceding ten years. The parameters are then kept fixed when hedging the contracts of that year.

As a benchmark we have chosen an asymmetric NGARCH model with constant drift structure that prescribes a returns dynamics of the form:

$$y_t = m + \sigma_t \epsilon_t, \quad \epsilon_t | \mathcal{F}_{t-1} \sim N(0,1), \tag{4.1}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 (\epsilon_{t-1} - \gamma)^2 + \beta_1 \sigma_{t-1}^2. \tag{4.2}$$

The parameter γ allows the modeling of the asymmetric behavior that we obtain in the ARSV context by considering non-zero correlations ρ between the (shifted) innovations for the returns and volatility processes. The scheme for model parameters estimation and updating is identical to the one followed for ARSV but this time using quasi-maximum likelihood.

In order to give the reader an indication of the parameter values that are obtained for each of these models, we report the values obtained for the year 2012 based on historical returns recorded during the preceding ten years. For the ARSV model with decorrelated innovations, we obtain the following values using Kalman and h-likelihood (in parenthesis):

$$m = -5.366 \cdot 10^{-5}, \quad \gamma = -0.084 \ (-0.094), \quad \phi : 0.991 \ (0.989), \quad \sigma_w = 0.113 \ (0.139).$$

In the correlated case we have:

$$m = -5.366 \cdot 10^{-5}, \ \gamma = -0.189 \ (-0.150), \ \phi : 0.9789 \ (0.983), \ \sigma_w = 0.177 \ (0.169), \ \rho = -0.8929 \ (-0.814).$$

Finally, the estimated NGARCH parameters obtained via maximum likelihood are:

$$m = 7.858 \cdot 10^{-11}, \ \alpha_0 = 1.923 \cdot 10^{-6}, \ \alpha_1 = 0.058, \ \beta_1 = 0.824, \ \gamma = 1.354.$$

The hedging exercise. Each of the put option contracts in the dataset under consideration is hedged via local risk minimization with a weekly frequency using the formulas in Section 2.2. More specifically, given a contract whose price is quoted for a given Wednesday, we construct a replicating self-financing portfolio that is rebalanced every week until the Wednesday preceding its maturity; for that particular contract, let $\{t_0, \ldots, t_l\}$ be the set of dates such that t_0 is the day for which the option price V_0 is quoted and for which the first hedging ratio ξ_0 is computed, the hedging portfolio is rebalanced at dates $t_1, \ldots, t_{l-1}$ using the hedges $\xi_1, \ldots, \xi_{l-1}$, and at $t_l := T$ the contract reaches its maturity. We will measure the performance of each of the models under consideration by comparing their associated normalized hedging square errors that we define, for each contract, as:

$$\text{NHSE} = \frac{\left(H(S_T) - V_0 - \sum_{i=0}^{l-1} \xi_i \left(S_{t_{i+1}} - S_{t_i}\right)\right)^2}{V_0^2}. \tag{4.3}$$

In this empirical exercise we compare the performances obtained when computing the hedges using (3.11) with ARSV models with independent and correlated innovations for the returns and the log-variance equations, and both estimated with either the Kalman or the h-likelihood approaches. We also study the sensitivity of the obtained results to the choice of the martingale measure Q used in the computation of the hedges; more specifically, we implement (3.11) using the minimal martingale measure $Q_{\min}$ and the mean correcting martingale measure Q_{mc}.

The hedges are computed via Monte Carlo by simulating 10^6 price paths under the physical measure P. In order to estimate the value of the conditional expectation under the martingale measures needed in (3.11), we use (3.9) by computing the values of the Radon-Nikodym derivatives associated to each path using the numerical implementation scheme described in Section 2.4, that uses estimates of the forecasted volatility using Kalman or h-likelihood. In the construction of the different discounted expressions, we use the corresponding period T-Bill rates interpolated to match the option maturity.

A crucial step in the path generation process is the choice of the initial volatility. In this exercise we choose to use exclusively information about the underlying asset and not on option prices. That is why we take as the initial volatility, the filtered value for the hedging date and corresponding to the method that is being tested (Kalman or h-likelihood).

Finally, the performance of the benchmark GARCH model (4.1)–(4.2) is evaluated by carrying out local risk minimization with respect to the martingale measure coming from the Extended Girsanov Principle (see [Bade 11]) with hedges evaluated via Monte Carlo using 10^6 paths initialized with the GARCH volatility. The GARCH volatility is also used in the construction of the Black-Scholes hedging scheme that completes the study.

Results. The empirical results are illustrated in Table 4.3. First, we notice that the ARSV model with correlated innovations provides the overall smallest hedging errors under the MCMM and for both estimation methods; the corresponding NHSE values are 0.337, for the Kalman filtering, and 0.338 for the h-likelihood approach, both slightly lower than the NGARCH benchmark of 0.357. The Black-Scholes model has the highest hedging error. The correlation between the returns innovations and the log-variance process has a significant effect on the hedging performance. For example, when the models are estimated using Kalman filtering, the overall NSHE is reduced by around 21% for both MCMM and MMM pricing kernels. Finally, we observe that the hedging errors are not very sensitive to the choice of risk-neutral measure used.

A closer look at Table 4.3 suggests that the ARSV models with correlated innovations outperforms the NGARCH counterpart for most maturity and moneyness groups considered. The largest differences are observed for long maturity and for in-the-money options. Moreover, the two estimation methodologies have a different impact on the hedging errors; the Kalman filtering approach slightly outperforms the h-likelihood for in-the-money and at-the-money options, while the latter methodology leads to bet-

ter results for short maturity contracts. The superior hedging performance of the ARSV models with a non-zero leverage effect is consistently documented for all groups of options.

BASIC FEATURES OF THE 2012-2013 OPTION DATASET (WEDNESDAYS)						
		Moneyness S_0/K					Across
	Maturities	[0.950, 0.975]	[0.975, 1.000]	[1.000, 1.025]	[1.025, 1.050]	[1.050, 1.100]	Moneyness
Number of Contracts	$T < 30$	18	67	225	210	346	866
	$30 \leq T < 80$	17	114	422	298	495	1346
	$80 \leq T < 180$	3	51	130	80	149	413
	$180 \leq T \leq 250$	0	12	41	21	29	103
Across Maturities		38	244	818	609	1019	2728
Average Prices	$T < 30$	58.933	33.557	19.404	10.318	5.034	25.449
	$30 \leq T < 80$	73.138	46.383	34.163	23.088	14.194	38.193
	$80 \leq T < 180$	98.767	79.584	65.802	50.420	38.948	66.704
	$180 \leq T \leq 250$	——	108.304	97.965	81.412	69.350	89.258
Across Maturities		76.946	66.957	54.333	41.309	31.881	53.093
Average Implied Volatilities	$T < 30$	0.133	0.145	0.148	0.168	0.198	0.158
	$30 \leq T < 80$	0.129	0.149	0.155	0.172	0.190	0.159
	$80 \leq T < 180$	0.154	0.167	0.168	0.180	0.198	0.173
	$180 \leq T \leq 250$	——	0.175	0.176	0.178	0.205	0.184
Across Maturities		0.139	0.159	0.162	0.175	0.198	0.168

Table 4.2: Basic features of the option dataset containing contract prices quoted every Wednesday during the period January 1st, 2012–December 31st, 2013.

5 Conclusions

This work introduces several explicit implementations of the local risk minimization hedging technique for ARSV options. Carrying this out required the design of various numerically viable pricing kernels combined with two existing volatility filtering techniques, namely Kalman filtering and the hierarchical-likelihood approach, in order to evaluate them and to implement pricing and dynamical hedging strategies. Two equivalent martingale measures were introduced. The first one is inspired by a modification in the so called Extended Girsanov Principle introduced in [Elli 98] and the second one is the so called minimal martingale measure.

The pertinence of the proposed approach was tested in an empirical study that shows that the ARSV model provides a competitive tool in the hedging of European style contingent products when compared to GARCH models with similar dynamical features, or the standard Black-Scholes delta hedging scheme. The study was based on a dataset of almost three thousand S&P500 European put option contracts quoted every Wednesday of the years 2012 and 2013. Our results indicate that the leverage effect plays a crucial role in the hedging performance of the ARSV models. Although the two proposed estimation schemes lead to similar overall hedging errors, the h-likelihood outperforms the Kalman filtering for short maturity and out-of-money contracts, while the latter provides a better alternative for long maturity and in-the-money options.

6 Appendix

6.1 Proof of Lemma 2.1

Let Z be an arbitrary $\mathcal{F}_{t-1}$ measurable random variable. As ϵ_t is independent of $\mathcal{F}_{t-1}$ and by hypothesis it is also independent of U, the joint law $P_{\epsilon_t,U,Z}$ of the random variable $(\epsilon_t, U, Z) : (\Omega, P) \to \mathbb{R}^3$ can be written as the product $P_{\epsilon_t,U,Z}(x,y,z) = P_{\epsilon_t}(x)P_{U,Z}(y,z)$. Hence, using Fubini's Theorem we have

$$
\begin{aligned}
E^P\left[e^{U\epsilon_t}Z\right] &= \int\int e^{yx}z\,dP_{\epsilon_t}(x)dP_{U,Z}(y,z) = \int\left[\int e^{yx}dP_{\epsilon_t}(x)\right]z\,dP_{U,Z}(y,z) \\
&= \int e^{L_{\epsilon_t}(y)}z\,dP_{U,Z}(y,z) = E^P\left[e^{L_{\epsilon_t}(U)}Z\right].
\end{aligned}
\tag{6.1}
$$

At the same time, by the definition of $K^P_{\epsilon_t}$, for any $\mathcal{F}_{t-1}$ measurable random variable Z, the equality (6.1) implies that:

$$
E^P\left[e^{K^P_{\epsilon_t}(U)}Z\right] = E^P\left[e^{U\epsilon_t}Z\right] = E^P\left[e^{L_{\epsilon_t}(U)}Z\right].
$$

As $e^{K^P_{\epsilon_t}(U)}$ is $\mathcal{F}_{t-1}$ measurable and Z arbitrary, the almost sure uniqueness of the conditional expectation implies that

$$
e^{K^P_{\epsilon_t}(U)} = E^P\left[e^{L_{\epsilon_t}(U)} \mid \mathcal{F}_{t-1}\right],
$$

as required. ■

The result that we just proved is a particular case of the following Lemma that we state for reference. The proof follows the same pattern as in the previous paragraphs.

Lemma 6.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{B} \subset \mathcal{F}$ a sub-σ-algebra. Let X and Y be two random variables such that X is simultaneously independent of Y and $\mathcal{B}$. For any measurable function $\Phi : \mathbb{R}^2 \to \mathbb{R}$ define the mapping $F : \mathbb{R} \to \mathbb{R}$ by $F(y) = E[\Phi(X, y)]$. Then,*

$$
E\left[\Phi(X, Y) \mid \mathcal{B}\right] = E[F(Y) \mid \mathcal{B}].
$$

		Moneyness S_0/K					Across
	Maturities	$[0.950, 0.975]$	$[0.975, 1.000]$	$[1.000, 1.025]$	$[1.025, 1.050]$	$[1.050, 1.100]$	Moneyness
				RESULTS FOR THE 2012-2013 HEDGING EXERCISE			
NHSE	$T < 30$	0.179	0.246	0.441	0.700	0.881	0.489
Black-Scholes	$\mathbf{30 \leq T < 80}$	0.141	0.281	0.342	0.498	0.654	0.383
	$\mathbf{80 \leq T < 180}$	0.349	0.368	0.381	0.458	0.586	0.429
	$\mathbf{180 \leq T \leq 250}$	——	0.407	0.384	0.408	0.539	0.435
Across Maturities		0.223	0.325	0.387	0.516	0.665	0.434
NHSE	$T < 30$	0.140	0.237	0.416	0.662	0.900	0.471
NGARCH	$\mathbf{30 \leq T < 80}$	0.115	0.234	0.288	0.419	0.540	0.319
	$\mathbf{80 \leq T < 180}$	0.270	0.282	0.280	0.334	0.387	0.311
	$\mathbf{180 \leq T \leq 250}$	——	0.310	0.254	0.256	0.360	0.295
Across Maturities		0.175	0.266	0.309	0.418	0.547	0.352
NHSE	$T < 30$	0.126	0.237	0.451	0.722	0.937	0.495
ARSV ($\rho = 0$)	$\mathbf{30 \leq T < 80}$	0.119	0.254	0.333	0.502	0.681	0.378
under MCMM and	$\mathbf{80 \leq T < 180}$	0.279	0.336	0.361	0.455	0.601	0.406
Kalman filtering	$\mathbf{180 \leq T \leq 250}$	——	0.356	0.357	0.416	0.535	0.416
Across Maturities		0.174	0.296	0.375	0.524	0.689	0.424
NHSE	$T < 30$	0.126	0.237	0.451	0.723	0.937	0.495
ARSV ($\rho = 0$)	$\mathbf{30 \leq T < 80}$	0.119	0.255	0.334	0.503	0.681	0.378
under MMM and	$\mathbf{80 \leq T < 180}$	0.279	0.337	0.362	0.455	0.601	0.407
Kalman filtering	$\mathbf{180 \leq T \leq 250}$	——	0.356	0.356	0.415	0.535	0.416
Across Maturities		0.175	0.296	0.376	0.524	0.689	0.424
NHSE	$T < 30$	0.113	0.230	0.408	0.680	0.890	0.464
ARSV (ρ arbitrary)	$\mathbf{30 \leq T < 80}$	0.121	0.226	0.282	0.425	0.568	0.324
under MCMM and	$\mathbf{80 \leq T < 180}$	0.237	0.248	0.250	0.313	0.369	**0.284**
Kalman filtering	$\mathbf{180 \leq T \leq 250}$	——	0.261	0.213	0.231	0.327	**0.258**
Across Maturities		0.157	**0.241**	**0.288**	0.412	0.538	**0.337**
NHSE	$T < 30$	0.113	0.229	0.408	0.680	0.890	0.464
ARSV (ρ arbitrary)	$\mathbf{30 \leq T < 80}$	0.118	0.226	0.283	0.426	0.570	0.325
under MMM and	$\mathbf{80 \leq T < 180}$	0.235	0.251	0.255	0.319	0.372	0.286
Kalman filtering	$\mathbf{180 \leq T \leq 250}$	——	0.280	0.223	0.240	0.343	0.272
Across Maturities		**0.155**	0.246	0.292	0.416	0.544	0.340
NHSE	$T < 30$	0.139	0.235	0.393	0.610	0.853	**0.446**
ARSV (ρ arbitrary)	$\mathbf{30 \leq T < 80}$	0.128	0.242	0.285	0.402	0.515	**0.315**
under MCMM and	$\mathbf{80 \leq T < 180}$	0.304	0.265	0.259	0.313	0.364	0.301
h-lklhd filtering	$\mathbf{180 \leq T \leq 250}$	——	0.291	0.234	0.249	0.340	0.279
Across Maturities		0.190	0.258	0.293	**0.394**	**0.518**	0.338
NHSE	$T < 30$	0.139	0.234	0.394	0.610	0.852	**0.446**
ARSV (ρ arbitrary)	$\mathbf{30 \leq T < 80}$	0.127	0.242	0.286	0.404	0.516	**0.315**
under MMM and	$\mathbf{80 \leq T < 180}$	0.303	0.267	0.263	0.318	0.366	0.304
h-lklhd filtering	$\mathbf{180 \leq T \leq 250}$	——	0.308	0.243	0.257	0.355	0.291
Across Maturities		0.190	0.263	0.297	0.397	0.522	0.341

Table 4.3: Results of the hedging exercise presented in Section 4. The table reports the averaged normalized hedging square errors (NHSE, see definition in (4.3)) committed in the hedging of the options in each of the different moneyness/maturities bin (see Table 4.2). Three main families of hedging schemes are considered: Black-Scholes delta hedging with GARCH filtered volatility and local risk minimization using NGARCH and ARSV models, with both correlated and uncorrelated innovations for the returns and the log-variances. In the case of the ARSV models, we study the sensitivity of the results to the method used to filter and forecast the volatility (Kalman or h-likelihood) and to the martingale measure (Mean correcting martingale measure (MCMM) or miminal martingale measure (MMM)) used to implement the local risk minimization approach.

6.2 Proof of Theorem 2.6

The proof is a straightforward consequence of Theorem 3.1 in [Elli 98]. Indeed, it suffices to identify in the ARSV case the two main ingredients necessary to carry out the Extended Girsanov Principle, namely the one period excess discounted return μ_t defined by:

$$e^{\mu_t} := E_{t-1}^P \left[\frac{\widetilde{S}_t}{\widetilde{S}_{t-1}} \right] = e^{m_t - r} E_{t-1}^P \left[e^{\sigma_t \epsilon_t} \right] = \frac{e^{m_t - r}}{K},$$

with $K := 1/E_{t-1}^P \left[e^{\sigma_t \epsilon_t} \right]$, and the process $\{W_t\}$ uniquely determined by the relation:

$$W_t := \frac{\widetilde{S}_t}{\widetilde{S}_{t-1}} e^{-\mu_t} = K e^{\sigma_t \epsilon_t}.$$

The Theorem follows from expression (3.8) in [Elli 98]. ∎

6.3 Proof of Theorem 2.8

(i) We start by noticing that since ϵ_t is independent of both ρ_t and $\mathcal{F}_{t-1}$, by Lemma 6.1,

$$E_{t-1}^P \left[N_t(\epsilon_t, \rho_t) \right] = E_{t-1}^P \left[F(\rho_t) \right], \tag{6.2}$$

where the real function $F : \mathbb{R} \to \mathbb{R}$ is defined by

$$F(y) = E^P \left[N_t(\epsilon_t, y) \right] = \int \frac{f_{\epsilon_t}^P(x+y)}{f_{\epsilon_t}^P(x)} f^P(x) dx = \int f_{\epsilon_t}^P(x+y) dx = \int f_{\epsilon_t}^P(x) dx = 1,$$

which substituted in (6.2) yields

$$E_{t-1}^P \left[N_t(\epsilon_t, \rho_t) \right] = 1. \tag{6.3}$$

This equality proves immediately the martingale property for $\{Z_t\}$. Indeed,

$$E_{t-1}^P \left[Z_t \right] = E_{t-1}^P \left[N_t(\epsilon_t, \rho_t) \right] Z_{t-1} = Z_{t-1}.$$

Finally, as $\{Z_t\}$ is a P-martingale

$$E^P \left[Z_t \right] = E^P \left[Z_1 \right] = E^P \left[N_1(\epsilon_1, \rho_1) \right] = 1.$$

(ii) Z_T is by construction non-negative and $E[Z_T] = \mathbb{P}(Z_T > 0) = 1$. This guarantees (see, for example, Remarks after Theorem 4.2.1 in [Lamb 08]) that Q_{mc} is a probability measure equivalent to P. In order to show that Q_{mc} is a martingale measure we note that a straightforward computation proves that:

$$E_{t-1}^{Q_{\mathrm{mc}}^1} \left[\frac{S_t}{S_{t-1}} \right] = E_{t-1}^P \left[e^{m_t + \sigma_t \epsilon_t} N_t(\epsilon_t, \rho_t) \right]. \tag{6.4}$$

Now, as ϵ_t is independent of both σ_t and $\mathcal{F}_{t-1}$, we can use Lemma 6.1 to prove that

$$E_{t-1}^P \left[e^{m_t + \sigma_t \epsilon_t} N_t(\epsilon_t, \rho_t) \right] = E_{t-1}^P \left[e^{m_t - \sigma_t \rho_t + L_{\epsilon_t}(\sigma_t)} \right]. \tag{6.5}$$

Indeed, we can write

$$E_{t-1}^P \left[e^{m_t + \sigma_t \epsilon_t} N_t(\epsilon_t, \rho_t) \right] = e^{m_t} E_{t-1}^P \left[e^{\sigma_t \epsilon_t} N_t(\epsilon_t, \rho_t) \right] = e^{m_t} E_{t-1}^P \left[F(\sigma_t, \rho_t) \right], \tag{6.6}$$

where the function $F : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$
\begin{aligned}
F(y, z) &= E^P \left[e^{y \epsilon_t} N_t(\epsilon_t, z) \right] = \int e^{yx} f^P_{\epsilon_t}(x) \frac{f^P_{\epsilon_t}(x+z)}{f^P_{\epsilon_t}(x)} dx = \int e^{y(s-z)} f^P_{\epsilon_t}(s) ds \\
&= E^P \left[e^{y(\epsilon_t - z)} \right] = e^{-yz} e^{L^P_{\epsilon_t}(y)},
\end{aligned}
$$

which substituted in (6.6) yields (6.5). Finally, if we insert in (6.5) the explicit expression (2.22) that defines the market price of risk, we obtain that

$$
E^{Q^1_{\mathrm{mc}}}_{t-1} \left[\frac{S_t}{S_{t-1}} \right] = E^P_{t-1} \left[e^{m_t - \sigma_t \rho_t + L_{\epsilon_t}(\sigma_t)} \right] = e^{r - K^P_{\epsilon_t}(\sigma_t)} E^P_{t-1} \left[e^{L^P_{\epsilon_t}(\sigma_t)} \right] = e^r,
$$

as required. The last equality in the previous expression follows from Lemma 2.1. ∎

6.4 Proof of Proposition 3.6

Consider the unique decomposition (3.15) of the discounted payoff $\widetilde{H}$ as a sum of the discounted P-gains and P-risk processes:

$$
\widetilde{H} = V_0 + \sum_{k=1}^T \xi^P_k \cdot \left(\widetilde{S}_k - \widetilde{S}_{k-1} \right) + \widetilde{L}^P_T. \tag{6.7}
$$

We recall that the minimal martingale measure is characterized by the property that every P-martingale that is strongly orthogonal to the discounted price process $\left\{ \widetilde{S}_t \right\}_{t=0}^T$, is also a $Q_{\min}$-martingale. Hence, as the P-martingale $\left\{ \widetilde{L}^P_t \right\}_{t=0}^T$ is P-strongly orthogonal to $\left\{ \widetilde{S}_t \right\}_{t=0}^T$, it is therefore a $Q_{\min}$-martingale. Having this in mind, as well as (3.14), we take conditional expectations $E^{Q_{\min}}_{t-1}$ on both sides of the decomposition (6.7) and we obtain:

$$
\widetilde{V}^P_t = V_0 + \sum_{k=1}^t \xi^P_k \cdot \left(\widetilde{S}_k - \widetilde{S}_{k-1} \right) + \widetilde{L}^P_t.
$$

As $\widetilde{V}^P_t$ is the local risk minimizing value process for both the P and the $Q_{\min}$ measures, that is $\widetilde{V}^P_t = \widetilde{V}^{Q_{\min}}_t$, if we multiply both sides of by $\left(\widetilde{S}_t - \widetilde{S}_{t-1} \right)$ and take conditional expectations $E^{Q_{\min}}_{t-1}$ we have:

$$
\begin{aligned}
E^{Q_{\min}}_{t-1} \left[\widetilde{V}^{Q_{\min}}_t \left(\widetilde{S}_t - \widetilde{S}_{t-1} \right) \right] &= E^{Q_{\min}}_{t-1} \left[V_0 \left(\widetilde{S}_t - \widetilde{S}_{t-1} \right) \right] + \sum_{k=1}^t E^{Q_{\min}}_{t-1} \left[\xi^P_k \cdot \left(\widetilde{S}_k - \widetilde{S}_{k-1} \right) \left(\widetilde{S}_t - \widetilde{S}_{t-1} \right) \right] \\
&\quad + E^{Q_{\min}}_{t-1} \left[\widetilde{L}^P_t \left(\widetilde{S}_t - \widetilde{S}_{t-1} \right) \right] \\
&= \xi^P_t E^{Q_{\min}}_{t-1} \left[\left(\widetilde{S}_t - \widetilde{S}_{t-1} \right)^2 \right] + E^{Q_{\min}}_{t-1} \left[\widetilde{L}^P_t \left(\widetilde{S}_t - \widetilde{S}_{t-1} \right) \right].
\end{aligned}
$$

If we divide both sides of this equality by $E^{Q_{\min}}_{t-1} \left[\left(\widetilde{S}_t - \widetilde{S}_{t-1} \right)^2 \right]$ and we use the relation (3.3) we obtain that

$$
\xi^{Q_{\min}}_t = \xi^P_t + \frac{E^{Q_{\min}}_{t-1} \left[\widetilde{L}^P_t (\widetilde{S}_t - \widetilde{S}_{t-1}) \right]}{\mathrm{var}^{Q_{\min}}_{t-1} \left[\widetilde{S}_t - \widetilde{S}_{t-1} \right]},
$$

as required. Finally, we notice that

$$E_{t-1}^{Q_{\min}}\left[\left(\widetilde{L}_t^P - \widetilde{L}_{t-1}^P\right)(\widetilde{S}_t - \widetilde{S}_{t-1})\right] = E_{t-1}^{Q_{\min}}\left[\widetilde{L}_t^P(\widetilde{S}_t - \widetilde{S}_{t-1})\right] - E_{t-1}^{Q_{\min}}\left[\widetilde{L}_{t-1}^P(\widetilde{S}_t - \widetilde{S}_{t-1})\right] = E_{t-1}^{Q_{\min}}\left[\widetilde{L}_t^P(\widetilde{S}_t - \widetilde{S}_{t-1})\right],$$

hence, either the $Q_{\min}$-independence or the orthogonality of $\left\{\widetilde{L}_t^P\right\}_{t=0}^T$ and $\left\{\widetilde{S}_t\right\}_{t=0}^T$ imply that $E_{t-1}^{Q_{\min}}\left[\widetilde{L}_t^P(\widetilde{S}_t - \widetilde{S}_{t-1})\right] = 0$ and hence $\{\xi_t^{Q_{\min}}\}_{t=1}^T = \{\xi_t^P\}_{t=1}^T$. ∎

References

[Ait 98] Y. Aït-Sahalia and A. W. Lo. "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices". *The Journal of Finance*, Vol. 53, No. 2, pp. 499–547, apr 1998.

[Ande 96] T. G. Andersen and B. E. Sørensen. "GMM estimation of a stochastic volatility model: A Monte Carlo study". *Journal of Business and Economic Statistics*, Vol. 14, pp. 328–352, 1996.

[Bade 11] A. Badescu, R. J. Elliott, R. Kulperger, M. Jarkko, and T. K. Siu. "A comparison of pricing kernels for GARCH option pricing with generalized hyperbolic distributions". *International Journal of Theoretical and Applied Finance*, Vol. 14, No. 5, pp. 669–708, 2011.

[Bade 14] A. Badescu, R. J. Elliott, and J.-P. Ortega. "Quadratic hedging schemes for non-Gaussian GARCH models". *Journal of Economic Dynamics and Control*, Vol. 42, pp. 13–32, 2014.

[Baks 97] G. Bakshi, C. Cao, and Z. Chen. "Empirical performance of alternative option pricing models". *The Journal of Finance*, Vol. 52, No. 5, pp. 2003–2049, dec 1997.

[Blac 72] F. Black and M. Scholes. "The valuation of option contracts and a test of market efficiency". *J. Finance*, Vol. 27, pp. 399–417, 1972.

[Boll 86] T. Bollerslev. "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics*, Vol. 31, No. 3, pp. 307–327, 1986.

[Carn 04] M. A. Carnero, D. Peña, and E. Ruiz. "Persistence and kurtosis in GARCH and stochastic volatility models". *Journal of Financial Econometrics*, Vol. 2, No. 2, pp. 319–342, 2004.

[Cast 08] J. del Castillo and Y. Lee. "GLM-methods for volatility models". *Stat. Model.*, Vol. 8, No. 3, pp. 263–283, 2008.

[Chor 12] C. Chorro, D. Guegan, and F. Ielpo. "Option pricing for GARCH-type models with generalized hyperbolic innovations". *Quantitative Finance*, Vol. 12, No. 7, pp. 1079–1094, 2012.

[Dani 94] J. Daníelsson. "Stochastic volatility in asset prices: estimation with simulated maximum likelihood". *Journal of Econometrics*, Vol. 64, pp. 375–400, 1994.

[Duan 01] J.-C. Duan, G. Gauthier, and J.-G. Simonato. "Asymptotic distribution of the EMS option price estimator". *Mgmt. Sci.*, Vol. 47, No. 8, pp. 1122–1132, 2001.

[Duan 95] J.-C. Duan. "The GARCH option pricing model". *Mathematical Finance*, Vol. 5, No. 1, pp. 13–32, 1995.

[Duan 98] J.-C. Duan and J.-G. Simonato. "Empirical martingale simulation of asset prices". *Management Science*, Vol. 44, No. 9, pp. 1218–1233, 1998.

[Durb 12]   J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods.* Oxford University Press, 2012.

[Elli 98]   R. J. Elliott and D. B. Madan. "A discrete time equivalent martingale measure". *Mathematical Finance*, Vol. 8, No. 2, pp. 127–152, 1998.

[Engl 82]   R. F. Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica*, Vol. 50, No. 4, pp. 987–1007, 1982.

[Flem 03]   J. Fleming. "A closer look at the relation between GARCH and stochastic autoregressive volatility". *Journal of Financial Econometrics*, Vol. 1, No. 3, pp. 365–419, sep 2003.

[Foll 02]   H. Föllmer and A. Schied. *Stochastic Finance.* Vol. 27 of *de Gruyter Studies in Mathematics*, Walter de Gruyter and Co., Berlin, 2002.

[Foll 86]   H. Föllmer and D. Sondermann. "Hedging of nonredundant contingent claims". In: *Contributions to mathematical economics*, pp. 205–223, North-Holland, Amsterdam, 1986.

[Foll 91]   H. Föllmer and M. Schweizer. "Hedging of contingent claims under incomplete information". In: *Applied stochastic analysis (London, 1989)*, pp. 389–414, Gordon and Breach, New York, 1991.

[Fore 05]   S. Foresi and L. Wu. "CrashOPhobia". *The Journal of Derivatives*, Vol. 13, No. 2, pp. 8–21, jan 2005.

[Frid 98]   M. Fridman and L. Harris. "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models". *Journal of Business & Economic Statistics*, Vol. 16, No. 3, pp. 284–291, 1998.

[Glen 04]   A. G. Glen, L. M. Leemis, and J. H. Drew. "Computing the distribution of the product of two continuous random variables". *Comput. Statist. Data Anal.*, Vol. 44, No. 3, pp. 451–464, 2004.

[Harv 94]   A. C. Harvey, E. Ruiz, and N. Shephard. "Multivariate stochastic variance models". *Review of Economic Studies*, Vol. 61, pp. 247–264, 1994.

[Harv 96]   A. C. Harvey and N. Shephard. "Estimation of an asymmetric stochastic volatility model for asset returns". *Journal of Business and Economic Statistics*, Vol. 14, No. 4, pp. 429–434, jul 1996.

[Hest 00]   S. L. Heston and S. Nandi. "A closed-form GARCH option valuation model". *Review of Financial Studies*, Vol. 13, No. 3, pp. 585–625, 2000.

[Hest 93]   S. L. Heston. "A closed-form solution for options with stochastic volatility with applications to bond and currency options". *Review of Financial Studies*, Vol. 6, No. 2, pp. 327–343, apr 1993.

[Hull 87]   J. Hull and A. White. "The pricing of options on assets with stochastic volatilities". *The Journal of Finance*, Vol. 42, No. 2, pp. 281–300, jun 1987.

[Jack 00]   J. C. Jackwerth. "Recovering risk aversion from option prices and realized returns". *Review of Financial Studies*, Vol. 13, No. 2, pp. 433–451, apr 2000.

[Jacq 94]   E. Jacquier, N. G. Polson, and P. E. Rossi. "Bayesian analysis of stochastic volatility models". *Journal of Business and Economic Statistics*, Vol. 12, pp. 371–417, 1994.

[Lamb 08]   D. Lamberton and B. Lapeyre. *Introduction to stochastic calculus applied to finance. Chapman & Hall/CRC Financial Mathematics Series*, Chapman and Hall/CRC, second Ed., 2008.

[Lee 06]   Y. Lee and J. A. Nelder. "Double hierarchical generalized linear models". *J. Roy. Statist. Soc. Ser. C*, Vol. 55, No. 2, pp. 139–185, 2006.

[Lee 96]   Y. Lee and J. A. Nelder. "Hierarchical generalized linear models". *J. Roy. Statist. Soc. Ser. B*, Vol. 58, No. 4, pp. 619–678, 1996.

[Lies 03]   R. Liesenfeld and J.-F. Richard. "Univariate and multivariate stochastic volatility models: estimation and diagnostics". *Journal of Empirical Finance*, Vol. 10, No. 4, pp. 505–531, 2003.

[Lim 11]   J. Lim, L. Woojoo, Y. Lee, and J. del Castillo. "The hierarchical-likelihood approach to autoregressive stochastic volatility models". *Computational Statistics and Data Analysis*, Vol. 55, No. 55, pp. 248–260, 2011.

[Meli 90]   A. Melino and S. M. Turnbull. "Pricing foreign currency options with stochastic volatility". *Journal of Econometrics*, Vol. 45, pp. 239–265, 1990.

[Mert 76]   R. Merton. "Option pricing when underlying stock returns are discontinuous". *J. Finan. Econ.*, Vol. 3, pp. 125–144, 1976.

[Meye 03]   R. Meyer, D. A. Fournier, and A. Berg. "Stochastic volatility: Bayesian computation using automatic differentiation and the extended Kalman filter". *Econometrics Journal*, Vol. 6, No. 2, pp. 408–420, dec 2003.

[Orte 12]   J.-P. Ortega. "GARCH options via local risk minimization". *Quantitative Finance*, Vol. 12, No. 7, pp. 1095–1110, oct 2012.

[Roha 76]   V. K. Rohatgi and A. K. M. E. Saleh. *An Introduction to Probability and Statistics*. John Wiley & Sons, 1976.

[Rubi 94]   M. Rubinstein. "Implied Binomial Trees". *The Journal of Finance*, Vol. 49, No. 3, pp. 771–818, jul 1994.

[Sand 98]   G. Sandmann and S. J. Koopman. "Estimation of stochastic volatility models via Monte Carlo maximum likelihood". *Journal of Econometrics*, Vol. 87, No. 2, pp. 271–301, dec 1998.

[Schw 01]   M. Schweizer. "A guided tour through quadratic hedging approaches". In: *Option pricing, interest rates and risk management, Handbook of Mathematical Finance*, pp. 538–574, Cambridge University Press, Cambridge, 2001.

[Scot 87]   L. O. Scott. "Option pricing when the variance changes randomly: theory, estimation, and an application". *Journal of Financial and Quantitative Analysis*, Vol. 22, No. 04, pp. 419–438, dec 1987.

[Shep 98]   K. S. Shephard and S. Chib. "Stochastic volatility: likelihood inference and comparison with ARCH models". *Review of Economic Studies*, Vol. 65, pp. 361–393, 1998.

[Shim 05]   J. Shimada and Y. Tsukuda. "Estimation of Stochastic Volatility Models: An Approximation to the Nonlinear State Space Representation". *Communications in Statistics - Simulation and Computation*, Vol. 34, No. 2, pp. 429–450, apr 2005.

[Stei 91]   E. M. Stein. "Stock price distributions with stochastic volatility: an analytic approach".
Review of Financial Studies, Vol. 4, No. 4, pp. 727–752, oct 1991.

[Tayl 05]   S. J. Taylor. *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press,
Princeton, 2005.

[Tayl 82]   S. J. Taylor. "Financial returns modelled by the product of two stochastic processes, a study
of daily sugar prices". In: B. D. Anderson, Ed., *Time series analysis: theory and practice I*,
pp. 1961–1979, 1982.

[Tayl 86]   S. J. Taylor. *Modelling Financial Time Series.* John Wiley & Sons, Chichester, 1986.

[Wigg 87]   J. B. Wiggins. "Option values under stochastic volatility: Theory and empirical estimates".
Journal of Financial Economics, Vol. 19, No. 2, pp. 351–372, dec 1987.

[Yu 05]   J. Yu. "On leverage in a stochastic volatility model". *Journal of Econometrics*, Vol. 127,
No. 2, pp. 165–178, aug 2005.