Regime-switching Stochastic Volatility Model : Estimation and Calibration to VIX options

Stéphane Goutte, Amine Ismail, Huyên Pham

Regime-switching Stochastic Volatility Model : Estimation and Calibration to VIX options

Stéphane GOUTTE[1,2,*], Amine ISMAIL[3,5] and Huyên PHAM[3,4]

[1]Université Paris 8, LED , 2 rue de la Liberté, 93526 Saint-Denis Cedex, France.
[2] PSB, Paris School of Business, 59 rue Nationale 75013 Paris, France.
[3] Laboratoire de Probabilités et Modèles Aléatoires, CNRS, UMR 7599, Universités Paris 7 Diderot.
[4]CREST-ENSAE
[5]Natixis, 47 Quai d'Austerlitz, 75013 Paris.

Abstract

We develop and implement a method for maximum likelihood estimation of a regime-switching stochastic volatility model. Our model uses a continuous time stochastic process for the stock dynamics with the instantaneous variance driven by a Cox-Ingersoll-Ross (CIR) process and each parameter modulated by a hidden Markov chain. We propose an extension of the EM algorithm through the Baum-Welch implementation to estimate our model and filter the hidden state of the Markov chain while using the VIX index to invert the latent volatility state. Using Monte Carlo simulations, we test the convergence of our algorithm and compare it with an approximate likelihood procedure where the volatility state is replaced by the VIX index. We found that our method is more accurate than the approximate procedure. Then, we apply Fourier methods to derive a semi-analytical expression of S&P 500 and VIX option prices, which we calibrate to market data. We show that the model is sufficiently rich to encapsulate important features of the joint dynamics of the stock and the volatility and to consistently fit option market prices.

Keywords: Regime-switching model; Stochastic volatility; Implied volatility; EM algorithm; VIX index; Options; Baum-Welch algorithm.

JEL classification: G12, C58, C51.

MSC classification: 91G70, 91G60, 60J05.

1 Introduction

A central research question in option pricing for academics and market practitioners is the specification of asset returns dynamics. Under the pricing measure, this question reduces to specify the dynamics of the volatility of the asset. Since the Black-Scholes constant volatility model does not provide a good fit to historic price data of the underlying and does not, by definition, have the ability to explain vanilla option prices, it cannot be reliable for pricing and hedging most of the recent exotic derivatives. Indeed, some of these structures cannot be hedged statically and require frequent rebalancing of vanilla options in order to hedge the vega risk. This leads to a penalizing hedging cost in the case of a mismatch between the market and model prices of the vanilla options used for the hedge (see Bergomi [9]). Over the years, several approaches, including local volatility and stochastic volatility models, have been proposed to overcome the issue of fitting market-implied volatilities across strikes and maturities.

Stochastic volatility models are increasingly important because they capture a richer set of empirical and theoretical characteristics than other volatility models. First, stochastic volatility models generate return distributions similar to what is empirically observed: the return distribution has a fatter left tail and kurtosis compared to normal distributions, with tail asymmetry controlled by the correlation between the stock and the volatility process (see Musiela and Rutkowski [35]). Second, stochastic volatility models allow reproduction of the main features of the volatility behaviour: mean reversion and volatility clustering (see Durham [22]). Third, stochastic volatility with a zero correlation always produces implied volatilities with a smile (see Renault and Touzi [37]). Fourth, Trolle and Schwartz [40] developed a tractable and flexible stochastic volatility multifactor model of the interest rates term structure. This model allows them to match the implied cap skews and the dynamics of implied volatilities. Moreover Christoffersen et al. [15] investigated alternatives to the affine square root stochastic volatility model, by comparing its empirical performance with different but equally parsimonious stochastic volatility models. They provide empirical evidence from three different sources: realized volatilities, S&P 500 returns, and an extensive panel of option data. Finally, historic volatility shows significantly higher variability than would be expected from local or time-dependent volatility, which could be better explained by a stochastic process. Among stochastic volatility models, the Heston model (see Heston [31]) is an industry standard. Its parameters are known to exert clear and specific control over the implied volatility skew/smile, and it can mimic the implied volatilities of around-the-money options with a fair degree of accuracy.

The recent introduction of derivatives on the VIX index contributed new and valuable information regarding S&P 500 index returns. Introduced by the CBOE in 1993, the VIX index non-parametrically approximates the expected future realized volatility of the S&P 500 returns over the next 30 days. VIX options started trading in 2006 and, as of today, represent almost a quarter of the vega notional of all the derivatives written on the S&P 500. By definition, the VIX index, VIX options, and S&P 500 options are directly linked to the S&P 500 index; therefore, the VIX option smile encodes information about the dynamics of volatility in a stochastic volatility model. If we want to value exotic options that are sensitive to the precise dynamics of S&P 500 implied volatilities or volatility derivatives, it seems natural to choose a model that consistently prices both VIX and S&P 500 options. However, because it has been widely studied in Duan and Yeh [19], Gatheral [25], Mencìa and Sentana [34], Wong and Lo [41], the Heston model is not able to fit the information of the VIX option market because the volatility process generated by this model remains too close to the mean level and because insufficient mass is spread in the tail of the probability distribution of the process.

Recently, several approaches have been proposed by academics and practitioners to overcome this shortcoming, which can be classified basically into two groups. The philosophy of the the first group consists in enriching the volatility process with features that describe a more realistic behaviour. In particular, a square-root process with jumps in both the underlying and the volatility was used in Sepp

[38] to price VIX derivatives, reinforced by an increasing body of statistical evidence for jumps in price and volatility dynamics. We can refer to Ait-Sahalia and Jacod [1], Cont and Mancini [14]) or Todorov [39] who have demonstrated that the variance risk is manifest in two salient features of financial returns: stochastic volatility and jumps. In Baldeaux and Badran [3], Drimus [18] a so-called 3/2 model is used as a candidate, allowing the volatility of volatility changes to be highly sensitive to the actual level of volatility. Another important approach that was derived recently uses additional factors to model stochastic volatility: see Bardgett et al. [6], in which a double Heston model with jumps is applied, and Bayer et al. [8], in which a three-factor double-mean-reverting model is calibrated to option market data. These models offer a good flexibility to consistently model VIX and S&P 500 options while providing a realistic representation of the path of volatility, see Gatheral [25], Mencìa and Sentana [34]. The second main alternative is to model the forward variance swap dynamics instead of the instantaneous variance for a discrete tenor of maturities, as originally suggested by Dupire [21] and subsequently formalized by Bergomi [10] and Buehler [12]. This approach is quite flexible and is able to reproduce the VIX option skew (Bao et al. [5]) and some important features of the variance swap dynamics and the term structure of the volatilities of volatility, while allowing better control over the desired level of forward skew of the underlying and its dependence on the level of volatility. In a recent work, Cont and Kokholm [13] combine both approaches incorporating jumps features in the forward variance process and the underlying process while offering more tractability than the Bergomi [11] model.

In this paper, we incorporate a regime switching feature in the stochastic volatility process in line with the first group in the literature as described above. The original introduction of regime switching in the econometric literature can be attributed to Hamilton [30], the relevance of regime shifts in return GARCH models is highlighted in Dueker [20], Hillebrand [32] and a multivariate Markov-modulated Gaussian model is calibrated to stock returns in Date and Mitra [16]. The first extension of the Heston model with regime switching to price VIX options was recently reported by Papanicolau and Sircar [36], in which an observed Markov chain modeling the state of volatility modulates two components of the stock process: the intensity of jumps and an additional multiplicative factor for volatility. Although this model overcomes the shortcoming of VIX skew, regime shifts drive the stock returns, and thus, it is not clear how the dynamics of volatility itself can be monitored using this model. Moreover, as shown by Gatheral [25], there is a very good consistency between forward variance swap rates estimates from S&P 500 and VIX options, invalidating the existence of a jump premium priced in the market, since the unique underlying assumption behind the computation of the VIX index is the continuity of the stock price process.

The first incorporation of regime switching in the volatility process itself was achieved by Elliot et al. [23, 24], who propose an extension of the Heston model, in which the mean-reverting level of volatility is modulated by an observable Markov chain, and use it to derive the price of volatility derivatives, such as variance swaps and volatility swaps.

In this paper, we generalize this approach by considering that a hidden Markov chain modulates the speed of mean reversion, the mean-reversion level, the volatility of volatility, and the correlation with the stock index. This model was originally studied by Goutte [27] for the pricing and hedging of derivatives. To the best of our knowledge, this is the first time that a combination of a Cox-Ingersoll-Ross framework with a Markov regime switching model has been used for pricing S&P 500 and VIX options. We also provide a method to estimate this model by deriving a complete maximum likelihood procedure to estimate the parameters of the model and the filtering of the hidden Markov chain. This model makes two main contributions: It facilitates better capturing important features of the (joint) dynamics of the stock and volatility and is able to consistently match the S&P 500 and the VIX option-implied volatilities.

The article is structured as follows. Section 2 provides an introductory analysis of VIX and S&P 500 time series, and we discuss the sources of motivation for the model. Section 3 focuses on describing the model and the calculations of the VIX index. Section 4 presents our estimation and filtering method in

detail and then tests the accuracy of the method by performing Monte Carlo simulations and comparing it with an approximate likelihood procedure in which the latent volatility state is replaced by a proxy. In Section 5, we derive closed-form expressions for the price of S&P 500 and VIX options using this model, calibrate the model to market prices by assuming a small rate of regime-change, and show through a joint calibration that the model offers sufficient flexibility to fit both smiles.

2 Motivations & Stylized Facts

We know that some important features of the behaviour of the S&P 500 and its volatility as well as their joint dynamics can be reproduced by the Heston model. These main features are the excess skewness and kurtosis of the distribution of stock returns, the mean reversion of the volatility process, and the so-called leverage effect for the stock-vol joint dynamics through the negative correlation between the two processes (see Figure 1). In this section, we will explore the main limitations of the Heston model and, therefore, the foundation of our work according to three aspects: the dynamics of the volatility process, the dynamics of stock returns, and their joint dynamics. We finally conclude this section by applying the regime-switching feature to the risk management of an equity portfolio using VIX futures.

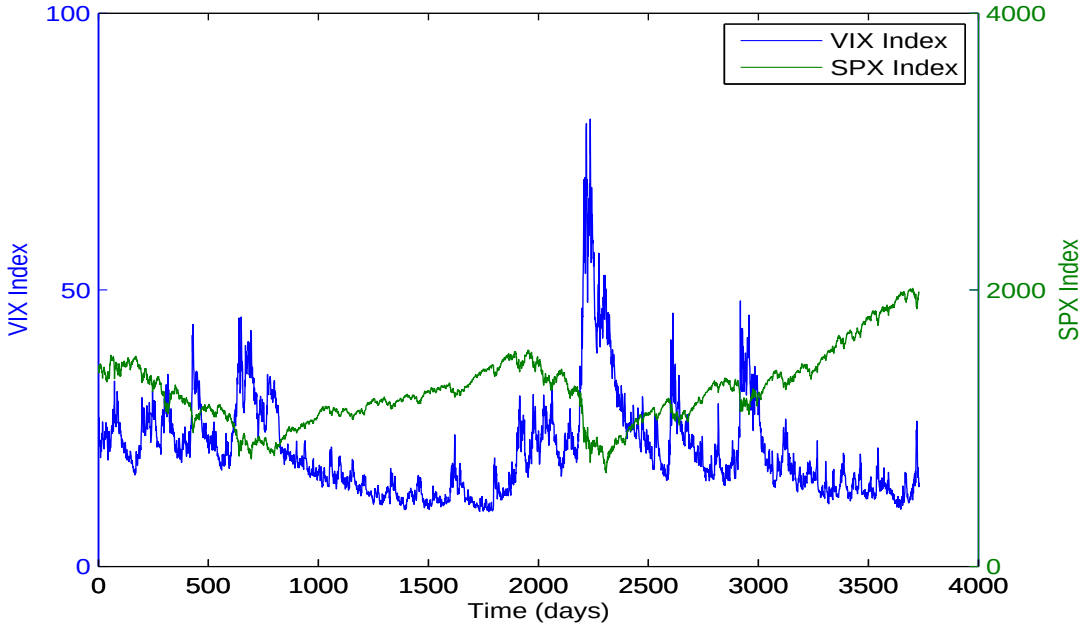

Figure 1: The VIX and S&P 500 indices from January 2000 to January 2015

2.1 Dynamics of volatility

To test how the Heston parameters shift over time, we start by examining a maximum likelihood estimation of the volatility process in the Heston model over a rolling window using the squared VIX index as a proxy for the instantaneous variance. We also estimate the correlation parameter between the instantaneous variance and the S&P 500 over the same rolling window. The results reported in Figure 2 not only demonstrate that these parameters are not constant over time but also highlight sharp level changes, such

as regime shifts, especially for the θ mean-reversion level, the volatility of volatility ξ and the correlation ρ. Moreover, by superposing in Figure 4 the graphs of instantaneous volatility (through its proxy) and the estimated volatility of volatility, these results reveal an important correlation between the two quantities, which is a crucial feature of the statistical behaviour of volatility. This feature cannot be captured by the Heston model, for which the instantaneous variance is modeled by a CIR process:

$$dV_t = \kappa(\theta - V_t)dt + \xi\sqrt{V_t}dW_t \tag{2.1}$$

Assuming that $2\kappa\theta > \xi^2$ and applying the Itô's lemma to the log-volatility process $\ln(\sigma_t)$, where $\sigma_t = \sqrt{V_t}$, we can see that the volatility of volatility is equal to $\frac{\xi}{2\sigma_t}$ in the Heston model:

$$d\ln(\sigma_t) = \frac{1}{2\sigma_t^2}\left(\kappa(\theta - \sigma_t^2) - \frac{\xi^2}{2}\right)dt + \frac{\xi}{2\sigma_t}dW_t \tag{2.2}$$

Thus, the Heston model implies that the volatility and the volatility of volatility will move in opposite directions. This suggests the use of an exogenous factor in the volatility process to drive both the mean level of volatility and the volatility of volatility (which has never been reported in the literature and motivates extending the work of Elliot et al. [23, 24]). The same motivation was previously inspired by the work of Bergomi [9], in which, after calibrating the Heston parameters to the option market using a large sample of time, he found that the daily variations of the calibrated instantaneous variance and the volatility of volatility showed an impressive correlation of almost 60 %. Thus, the market was already pricing a feature of the behaviour of volatility that Heston's model misprices by construction, thereby leading to the mispricing of derivatives that are highly sensitive to the dynamics of implied volatility. However, what was previously hidden became more transparent with the development of volatility products, allowing more insight regarding how the market prices the behaviour of volatility and, therefore, how it should be modeled. The correlation between volatility and the volatility of volatility can indeed be illustrated using the VVIX index as a proxy for the vol-of-vol. The VVIX index, which is also called the VIX of VIX, is calculated from the price of a portfolio of liquid VIX options and gives a model-free measure of the 1-month volatility of the VIX as implied by the market. The daily log-variations of the two series in in Figure 4 exhibit a high correlation level of 60 %.

The information encapsulated in the positive skew of VIX (see Bao et al. [5]) implied volatilities demands the same improvement: the market implies that a higher VIX will be concomitant with a high level of volatility of the VIX, which is in good agreement with the historical behaviour but cannot be fulfilled by the Heston model (see Figure 3). The positive slope of the VIX implied volatilities also provides the market's perspectives regarding the distribution of $\log(VIX_t)$ conditional upon V_0, which is closely linked to the conditional distribution of $\log(\sqrt{V_t})$ (here again, we proxy $\sqrt{V_t}$ with VIX_t). It is clear that the conditional moments and higher moments of a log-chi-squared distribution (as given for $\log(V_t)$ by the Heston model) are not compatible with the market-implied positive skewness and fat tails of the distribution of $\log(VIX_t)$. We would like to mention that recently, Bayer et al. [7], resolved this problematic of Heston model (i.e. cover typical short term maturity features of observed data (e.g. volatility smiles)) via a completely different approach.

To show how the regime-switching feature can fill this gap, we compare model-implied conditional skewness and kurtosis for $\log VIX_t$ generated by the Heston model and its regime-switching extension. The results are presented in Table 1 and show how the fits are improved for the conditional higher moments because of regime switching. Table 1 also provides the model-implied unconditional higher moments for the VIX index compared with their sample data values, resulting in a better fit and the

model's ability to generate more realistic implied volatility paths for the pricing of forward implied volatility surface-dependent derivatives.

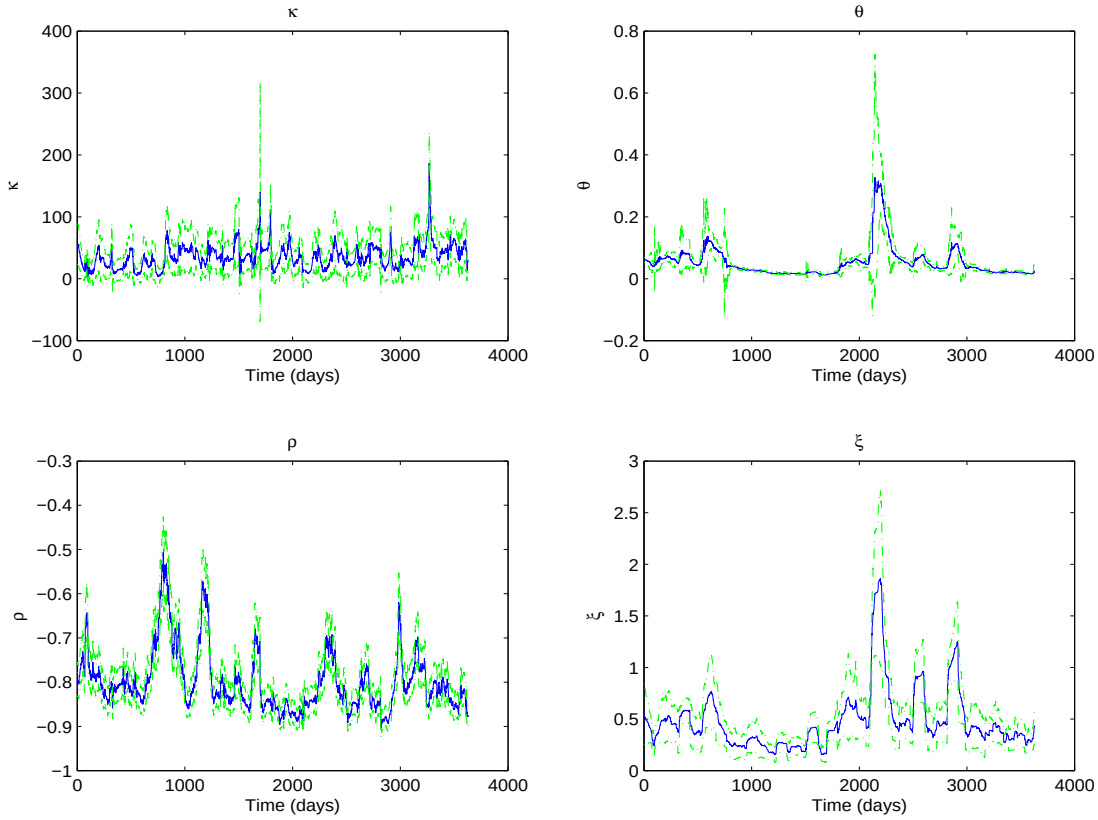

Figure 2: Outputs of the 100 days rolling maximum likelihood for κ, θ and ξ. The correlation parameter is estimated over the same period. Estimates are displayed with the 95% confidence bounds

2.2 Conditional moments of stock returns

It is well known that the Heston model has the ability to generate realistic higher moments in the unconditional distribution of stock returns. However, the model-implied conditional higher moments may be too close to normality, especially for short-term horizons, which explains why the model falls short of explaining the implied volatility smile for short-dated options. In fact, an analysis of conditional moments of stock returns shows near-Gaussian behaviour for all the stochastic variance models, particularly at short horizons, as argued by Jones [33]. Here, we show how adding regime switching to the Heston stochastic volatility model allows overcoming this shortcoming. Considering 5- and 21-day time horizons, we use a Monte Carlo simulation to compute the skewness and kurtosis of the distribution of stock returns conditional upon the level of the instantaneous variance $(V_0 = 0.02)$ and the level of the volatility regime $(Z_0 = 1)$. The results are presented in Table 2 and show how the regime-switching feature allows the non-Gaussianity of the conditional stock returns to significantly increase at short time horizons. As a benchmarking model, we use the CEV class model studied by Jones [33] (which generalizes the '3/2' model) and manages, similar to our model, to address some important features, such as the correlation of

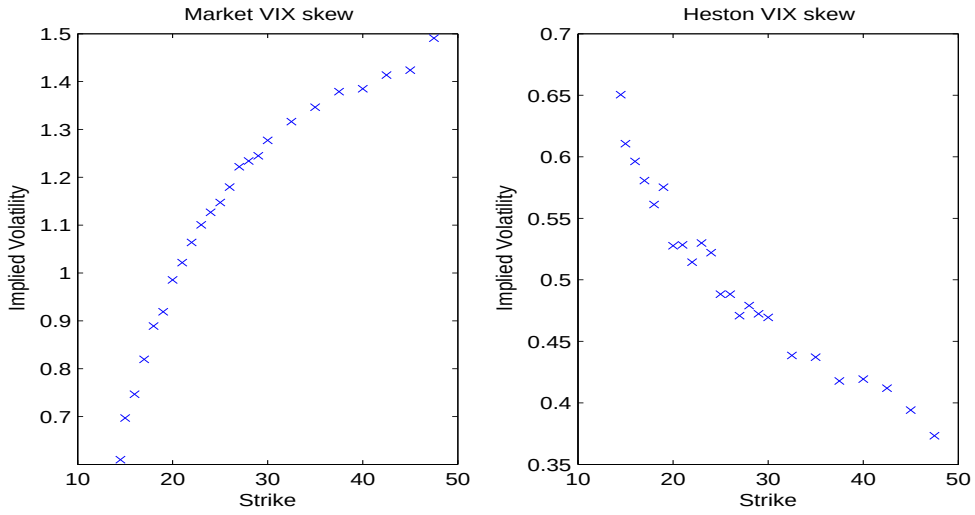

Figure 3: The VIX option-implied volatilities versus the skew implied by the Heston Model

	Heston model		RSSV	
	skewness	kurtosis	skewness	kurtosis
Conditional moments (t=21 days, $V_0 = 0.02$)	-0.11	3.31	0.68	4.28
Unconditional moments - Model implied	-0.46	3.25	0.78	8.23
Unconditional moments - Sample data	0.63	7.07	0.63	7.07

Table 1: Conditional and unconditional higher moments for the VIX index log returns

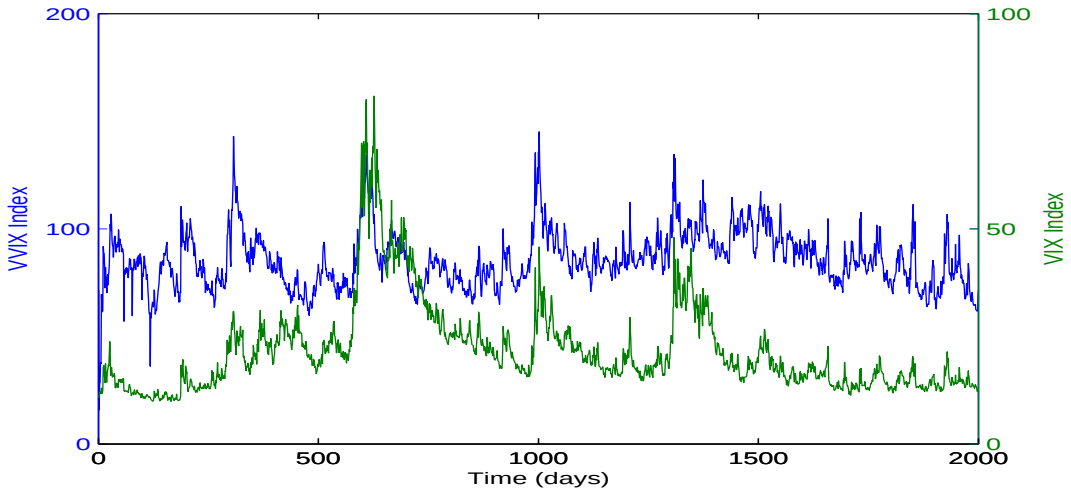

Figure 4: Daily closes of the VIX index and the VVIX index from March 2006 to January 2015

	Heston model		CEV		RSSV	
	skewness	kurtosis	skewness	kurtosis	skewness	kurtosis
Conditional moments (t=5 days, $V_0 = 0.02$)	-0.09	3.03	-0.095	3.05	-0.15	3.13
Conditional moments (t=21 days, $V_0 = 0.02$)	-0.20	3.12	-0.22	3.13	-0.38	3.71

Table 2: Conditional higher moments for stock returns

volatility and the vol-of-vol, time-varying leverage effect and unconditional skewness and kurtosis of the VIX and stock distributions. However, it fails to reproduce the conditional higher moments. Although the typically suggested way to tackle this issue is to add a jump component to the stock process, here, we show that this stylized fact can be embedded in a regime-switching stochastic volatility process.

2.3 Joint dynamics of the stock and the volatility process

The empirical study shown in Figure 2 reveals sharp level changes in the correlation between the volatility and the stock process. Most importantly, this correlation tends to increase when volatility increases, indicating that the leverage effect becomes substantially stronger, as noted by Bandi and Reno [4]. Adding a regime-switching feature to the stock-vol correlation allows capturing this stylized fact and produces a far more realistic distribution of stock returns by allowing more extreme events at short and long horizons. From a pricing perspective, this feature prices the fact that when the spot moves down, the implied volatility and the implied volatility skew increase even more, which is particularly crucial for barrier option valuation.

To summarize, we have demonstrated that the following features are better reproduced by the regime-switching Heston model:

- the correlation between the volatility and the volatility of volatility;
- the conditional and unconditional higher moments of the volatility distribution;
- the short-term conditional higher moments of stock returns; and
- the time-varying leverage effect

All of these points justify the use of a new extension of the standard Heston model with regime-switching parameters in both all the parameters of the volatility process and the correlation factor between the stock and the volatility processes.

3 The stochastic model

3.1 Regime-switching Heston model

Below, we work in a filtered probability space $(\Omega, \mathbb{F} := (\mathcal{F}_t)_{[0,T]}, \mathbb{P})$, where $\mathbb{P}$ denotes a risk neutral or equivalent martingale measure. We introduce (Z_t), a homogeneous continuous time Markov chain on a finite space $E := \{1, ..., S\}$, with an initial value of μ. This Markov chain represents the regime-state of volatility. We denote by Π the generator matrix of Z, which is given for all $i, j \in E$ by $\Pi_{ij} \geq 0$ if $i \neq j$ and $\Pi_{ii} = -\sum_{i \neq j} \Pi_{ij}$. We assume that the transition probabilities from a state $i \in E$ at time t towards a

state $j \in E$ at time $t+h$ are stationary. This leads to an infinitesimal generator Π independent of time. Let us denote now by $T(h)$ the matrix of transition probabilities defined by $T_{ij}(h) = \mathbb{P}(Z_h = j \mid Z_0 = i)$, for all $i, j \in E$. Then, the following relation holds:

$$\frac{dT_{ij}(h)}{dh} = \Pi_{ij} \; for \; all \; i, \; j \; \in E$$

Note that this relation implies that $T_{ij}(h) = (e^{\Pi h})_{ij}$. Finally, let us denote by P the matrix of transition probabilities of the embedded DTMC of Z (discrete time Markov chain). P is given by:

$$P_{ij} = \begin{cases} \frac{\Pi_{ij}}{\sum_{i \neq j} \Pi_{ij}} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

Now, we can introduce the model that we will use to model the $S\&P500$ index. Let $S = (S_t)$ be a stochastic process in our probability space that models the spot value of the $S\&P500$ index. Let $V = (V_t)$ be another stochastic process that models the instantaneous variance of S. We assume that the dynamic of our model is given by the following system of stochastic differential equations:

$$\begin{cases} dS_t = S_t(rdt + \sqrt{V_t} dW_t^1), \quad S_0 = s \\ dV_t = \kappa(Z_t)(\theta(Z_t) - V_t)dt + \xi(Z_t)\sqrt{V_t} dW_t^2, \quad V_0 = v_0. \end{cases} \tag{3.3}$$

where $r \geq 0$ is the risk-free rate of return and W^1 and W^2 are two $\mathbb{P}$-Brownian motions such that $\mathbb{E}^{\mathbb{P}}[dW_t^1 dW_t^2] = \rho(Z_t)dt$ with $\rho \in [-1, 1]$.

Assumption 3.1. *We assume that Z and the pair of processes W^1, W^2 are independent. Therefore, the Markov process Z is considered to be an exogenous factor of the market information.*

Remark 3.1. *It is important to note that all the parameters of the volatility process V and that the correlation factor between the $S\&P500$ index and its instantaneous variance V depend on the homogeneous continuous time Markov chain Z.*

Now, we define some filtrations. Let $\mathcal{F}_t^W$ be the filtration generated by the two Brownian motions W^1 and W^2. Thus, $\mathcal{F}_t^W := \sigma((W^1, W^2)_s, s \leqslant t)$. Then, let $\mathcal{F}_t^Z$ be the one generated by the process Z, $\mathcal{F}_t^Z := \sigma(Z_s, s \leqslant t)$. Finally, we denote by $\mathcal{F}$ the global filtration that is given by $\mathcal{F}_t := \mathcal{F}_t^W \vee \mathcal{F}_t^Z$.

3.2 The VIX Index

Definition 3.1. *The VIX index is calculated as the strike of the one-month variance swap contract on the S&P500 index. Let τ_{vix} be the duration corresponding to one month. Because S_t has no jumps, the following relationship holds:*

$$VIX_t^2 = \mathbb{E}^{\mathbb{P}}\left[\frac{1}{\tau_{vix}} \int_t^{t+\tau_{vix}} V_s ds \mid \mathcal{F}_t\right] \tag{3.4}$$

We can now give two decomposition results based on the VIX process.

Lemma 3.1. *Let M be a discrete random variable in $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$, such that $M-1$ is the random number of jumps of the Markov chain Z between 0 and τ_{vix}. Let $(\tau_1, \ldots, \tau_{M-1})$ be a sequence of $M-1$ random jump times, and let us denote by $\Delta_{t_k} := \tau_{k+1} - \tau_k$ for each $k \in \{0, \ldots, M-1\}$ with $\tau_M = \tau_{vix}$ and $\tau_0 = 0$. For each $k \in \{0, \ldots, M-1\}$, there are two families of functions $\{f_k\}_{0 \leq k \leq M-1}$ and $\{g_k\}_{0 \leq k \leq M-1}$ defined from E^k to $\mathbb{R}$, such that:*

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[\int_0^{\tau_{vix}} V_s ds \mid V_0 = v_0, \mathcal{F}^Z_{\tau_{vix}}\right] &= \left(\sum_{k=0}^{M-1} a(Z_{\tau_k}) f_k(Z_0, \ldots, Z_{\tau_{k-1}})\right) v_0 \\ &+ \sum_{k=0}^{M-1} \left(b(Z_{\tau_k}) + a(Z_{\tau_k}) g_k(Z_0, \ldots, Z_{\tau_{k-1}})\right) \end{aligned}$$

where the functions $\{g_k\}_{0 \leq k \leq M-1}$ and $\{f_k\}_{0 \leq k \leq M-1}$ are determined as follows:

$$g_k(Z_{\tau_0}, \ldots, Z_{\tau_{k-1}}) = \begin{cases} g_{k-1}(Z_{\tau_0}, \ldots, Z_{\tau_{k-2}})(1 - \kappa(Z_{\tau_{k-2}})a(Z_{\tau_{k-2}})) & \text{if } k \in \{2, \ldots, M-1\} \\ +\kappa(Z_{\tau_{k-2}})(\theta(Z_{\tau_{k-2}})\Delta_{t_{k-2}} - b(Z_{\tau_{k-2}})) & \\ \kappa(Z_{\tau_0})(\theta(Z_{\tau_0})\Delta_{t_0} - b(Z_{\tau_0})) & \text{if } k = 1 \\ 0 & \text{if } k = 0 \end{cases}$$

$$f_k(Z_{\tau_0}, \ldots, Z_{\tau_{k-1}}) = \begin{cases} \prod_{j=0}^{k-1} (1 - \kappa(Z_{\tau_j}))a(Z_{\tau_j}) & \text{if } k \in \{1, \ldots, M-1\} \\ 1 & \text{if } k = 0 \end{cases}$$

Additionally, $a(Z_{\tau_k})$ and $b(Z_{\tau_k})$ are given by:

$$\begin{aligned} a(Z_{\tau_k}) &= \frac{1 - \exp(-\kappa(Z_{\tau_k})\Delta_{t_k})}{\kappa(Z_{\tau_k})} \\ b(Z_{\tau_k}) &= \theta(Z_{\tau_k})(\Delta_{t_k} - a(Z_{\tau_k})) \end{aligned}$$

Proof The proof of this lemma can be found in the appendix.

Now, we can report an important result that describes the relationship between the process VIX_t and the couple (V_t, Z_t). This result will subsequently be crucial.

Proposition 3.1. *For each integer $M \in \mathbb{N}^*$ and each $z_0 \in E$, let us denote by $S^{\tau_{vix}}_{z_0,M}$ the set of all*

possible sequences of states through a path of the continuous time Markov Chain Z containing $M-1$ jumps between $t=0$ and τ_{vix} and starting at the state z_0,i.e.

$$S_{z_0,M}^{\tau_{vix}} = \{(s_i)_{0\leq i\leq M} | s_i \in E, s_0 = z_0, s_{i+1} \neq s_i \text{ if } i \in \{0,\ldots,M-2\}, s_M = s_{M-1}\}$$

Now, with each set of M dwell times $(\Delta_{t_0},\ldots,\Delta_{t_{M-1}})$ of the continuous time Markov chain Z and each sequence $s \in S_{z_0,M}^{\tau_{vix}}$, we can associate a unique path of Z containing $M-1$ jumps between $t=0$ and $t=\tau_{vix}$. For each of these paths, we define the following functions:

$$\begin{aligned}
\mathcal{A}_M(s,(\Delta_{t_k})_{0\leq k\leq M-1}) &= \sum_{k=0}^{M-1} a(Z_{\tau_k}) f_k(Z_0,\ldots,Z_{\tau_{k-1}}) \\
\mathcal{B}_M(s,(\Delta_{t_k})_{0\leq k\leq M-1}) &= \sum_{k=0}^{M-1} \left(b(Z_{\tau_k}) + a(Z_{\tau_k}) g_k(Z_0,\ldots,Z_{\tau_{k-1}})\right)
\end{aligned}$$

Thus, the probability of this path, which is given by the function:

$$\mathcal{L}_M^{\tau_{vix}}(s,(\Delta_{t_k})_{0\leq k\leq M-1}) = \begin{cases} \left(\prod_{i=0}^{M-1} q_{s_i} e^{-q_{s_i}\Delta_{t_i}} P_{s_i s_{i+1}}\right) e^{-q_{s_M}(\tau_{vix}-\sum_{k=0}^{M-1}\Delta_{t_k})} & \text{if } \sum_{k=0}^{M-1}\Delta_{t_k} < \tau_{vix} \\ 0 & \text{otherwise} \end{cases}$$

where $q_{s_i} = \Pi(s_i,s_i)$.
If the variance process V_t follows the described dynamic, then there is a linear relationship between the squared VIX_0^2 and v_0 with switching coefficients depending on the state of the Markov chain z_0. This relationship is given by:

$$VIX_t^2 = \alpha(z_0)V_t + \beta(z_0)$$

where α and β are two functions from the set E to $\mathbb{R}$ given by:

$$\begin{aligned}
\alpha(z_0) &= \frac{1}{\tau_{vix}} \sum_{M=1}^{\infty} \sum_{s\in S_{z_0,M}^{\tau_{vix}}} \underbrace{\int\cdots\int}_{\sum_{k=0}^{M-1}\Delta_{t_k}<\tau_{vix}} \mathcal{A}_M(s,(\Delta_{t_k})_{0\leq k\leq M-1}) \mathcal{L}_M^{\tau_{vix}}(s,(\Delta_{t_k})_{0\leq k\leq M-1})\, d\Delta_{t_0}\ldots d\Delta_{t_{M-1}} \\
\beta(z_0) &= \frac{1}{\tau_{vix}} \sum_{M=1}^{\infty} \sum_{s\in S_{z_0,M}^{\tau_{vix}}} \underbrace{\int\cdots\int}_{\sum_{k=0}^{M-1}\Delta_{t_k}<\tau_{vix}} \mathcal{B}_M(s,(\Delta_{t_k})_{0\leq k\leq M-1}) \mathcal{L}_M^{\tau_{vix}}(s,(\Delta_{t_k})_{0\leq k\leq M-1})\, d\Delta_{t_0}\ldots d\Delta_{t_{M-1}}
\end{aligned}$$

Proof

We can write

$$\begin{aligned} VIX_0^2 &= \mathbb{E}^{\mathbb{P}}\left[\frac{1}{\tau_{vix}}\int_0^{\tau_{vix}} V_s ds \mid \mathcal{F}_0\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[\frac{1}{\tau_{vix}}\mathbb{E}^{\mathbb{P}}\left[\int_0^{\tau_{vix}} V_s ds \mid \mathcal{F}^Z_{\tau_{vix}}, V_0 = v\right] \mid \mathcal{F}_0\right] \end{aligned}$$

Finally, by denoting:

$$\begin{aligned} \alpha(z_0) &= \frac{1}{\tau_{vix}}\mathbb{E}^{\mathbb{P}}[\sum_{k=0}^{M-1} a(Z_{\tau_k}) f_k(Z_0, \ldots, Z_{\tau_{k-1}}) \mid Z_0 = z_0] \\ \beta(z_0) &= \frac{1}{\tau_{vix}}\mathbb{E}^{\mathbb{P}}[\sum_{k=0}^{M-1} b(Z_{\tau_k}) + a(Z_{\tau_k}) g_k(Z_0, \ldots, Z_{\tau_{k-1}}) \mid Z_0 = z_0] \end{aligned}$$

we obtain the expected result.

The expression of the functions α and β can be derived explicitly because each multidimensional integral in the expression of α and β is a linear combination of integrals of the form $\int_\Omega \exp(x.a)dx$ or $\int_\Omega (x.b)\exp(x.a)dx$ where a and b are constant vectors and Ω is a triangular domain. However, practically, these expressions are derived up to a fixed dimension $\bar{M}$, which represents the maximum number of jumps considered for the Markov chain during a month. In the following simulations, we will choose $\bar{M} = 5$, which is coherent with the assumption that the jump intensity of the Markov chain is moderate.

4 Maximum Likelihood Estimation

In this section, we explore a maximum likelihood procedure to estimate the set of parameters appearing in the dynamic of V, $\Theta := \{(\theta_i)_{i\in E}, (\kappa_i)_{i\in E}, (\xi_i)_{i\in E}, \Pi\}$. The main goal is to use the Expectation-Maximization (EM) algorithm initiated in Dempster, Laird and Rubin [17], which consists of completing the vector of the observed data with the unobserved data and iteratively maximizing a pseudo-likelihood function of the complete data. When the unobserved data is a hidden Markov chain driving the parameters of a continuous time-observed process, a well-known application of the EM principle is the Baum-Welch algorithm. This algorithm is used by Goutte and Zou [29] to estimate the parameters of a continuous time regime-switching model, in which the observed data are the foreign exchange rates and the unobserved data consist of a hidden Markov chain. The main difference and main difficulty in this work relate to the fact that, in this case, the process $V = (V_t)$ is not observable, unlike the spot foreign exchange rate. An interesting discussion can be found in Papanicolau and Sircar [36] on risk-neutral filtering and smile effects caused by volatility uncertainty. Thus, we are confronted with the problem of the the maximum likelihood estimation for a set of latent data driven by a hidden Markov chain. To solve the problem of the latent variance, Ait-Sahalia and Kimmel [2] proposed two different approaches: a full likelihood procedure, in which a quoted security (an option, for example) is inverted into the unobservable volatility state, and an approximate likelihood procedure, in which the volatility process is replaced by a good proxy, such as the implied volatility of a short dated at the money option. In our paper, we use the Baum-Welch algorithm to address the problem of the unobservable Markov chain in the model, and we adapt it by proposing using the relationship established between the observable VIX Index and the couple (V_t, Z_t) to overcome the issue of the unobservable variable V_t.

4.1 Algorithm estimation

Let us suppose that we observe the value of the VIX Index daily. Let the size of the historical data be $H+1$ and Γ be the corresponding increasing sequence of time in which these data are collected:

$$\Gamma = \{t_j; 0 = t_0 \leq t_1 \leq ...t_{H-1} \leq t_H = H * \delta\},\ with\, \delta = t_j - t_{j-1}.$$

Here, the time step δ is constant. Let us denote $Y_t = VIX_t^2$, and by (y_k, z_k), realize the random variables (Y_{t_k}, Z_{t_k}) for every $k \in \{0, \ldots, H\}$. Finally, we denote a realization of the full history $\mathbb{Y} := (Y_{t_k})_{0 \leq k \leq H}$ and $\mathbb{Z} := (Z_{t_k})_{0 \leq k \leq H}$ by y and z. Using this algorithm, we deduce the transition density function of the observed variable $VIX_{t_k}^2$ from the transition density function of the unobserved variable V_{t_k}, conditional upon the state of the Markov chain at time t_k and t_{k-1}. This approach is used in Ait-Sahalia and Kimmel's work, as mentioned above, where the transition density function of the observed asset prices (the stock and one or more options written on that stock) is derived from the transition function of the state vector (the stock and the variance processes). However, our work is based on two aspects. First, we assume that the information embedded in the VIX is sufficient to reflect all of the empirical features in the variance behaviour and that we do not need the information contained in the stock process S_t, thereby simplifying the computation of the state vector transition function. Second, we must take into account the unobservable state of the Markov chain in our maximum likelihood procedure because the hidden Markov chain is needed to complete the observed data.
Let $\{\epsilon_{t_k}\}_{0 \leq k \leq H}$ be a sequence of i.i.d random variables such that $\epsilon_{t_k} \sim \mathcal{N}(0,1)$. We can write thus the following discretization of the process V:

$$\begin{aligned} V_{t_k} - V_{t_{k-1}} &= \kappa(Z_{t_k})(\theta(Z_{t_k}) - V_{t_{k-1}})\delta + \xi(Z_{t_k})\sqrt{\delta}\sqrt{V_{t_{k-1}}}\epsilon_{t_k} \\ V_{t_k} &= \kappa(Z_{t_k})\theta(Z_{t_k})\delta + (1 - \kappa(Z_{t_k})\delta)V_{t_{k-1}} + \xi(Z_{t_k})\sqrt{\delta}\sqrt{V_{t_{k-1}}}\epsilon_{t_k} \end{aligned}$$

Using the relationship established in Proposition 1.1, we get:

$$\frac{Y_{t_k} - \beta(Z_{t_k})}{\alpha(Z_{t_k})} = \kappa(Z_{t_k})\theta(Z_{t_k}))\delta + (1 - \kappa(Z_{t_k})\delta)\frac{Y_{t_{k-1}} - \beta(Z_{t_{k-1}})}{\alpha(Z_{t_{k-1}})} + \xi(Z_{t_k})\sqrt{\delta}\sqrt{\frac{Y_{t_{k-1}} - \beta(Z_{t_{k-1}})}{\alpha(Z_{t_{k-1}})}}\epsilon_{t_k}$$

In what follows, we well denote $\alpha(Z_t, \Theta) := \alpha(Z_t)$ and $\beta(Z_t, \Theta) := \beta(Z_t)$. Thus, for each $(i,j) \in E^2$, we have the following expression for the density function of Y_{t_k} evaluated in y_k conditional upon $(Y_{t_{k-1}} = y_{k-1}, Z_{t_k} = j, Z_{t_{k-1}} = i)$:

$$\begin{aligned} f_Y(y_k \mid Z_{t_k} = j, Z_{t_{k-1}} = i, y_{k-1}, \Theta) &= \frac{1}{\mid \alpha(j,\Theta) \mid \sqrt{2\pi\delta\frac{y_{k-1} - \beta(i,\Theta)}{\alpha(i,\Theta)}\xi_j}} \\ &\times \exp\left\{ -\frac{\left(\frac{y_k - \beta(j,\Theta)}{\alpha(j,\Theta)} - \theta_j\kappa_j\delta - (1 - \kappa_j\delta)\frac{y_{k-1} - \beta(i,\Theta)}{\alpha(i,\Theta)}\right)^2}{2\delta\xi_j^2\frac{y_{k-1} - \beta(i,\Theta)}{\alpha(i,\Theta)}} \right\} \end{aligned}$$

Our objective is to maximize the likelihood of the observed data. Thus, we aim to find the values of

Θ that maximize the probability:

$$J := \mathcal{L}(\Theta \mid Y) := \mathbb{P}(Y_0 = y_0, ..., Y_{t_H} = y_H \mid \Theta)$$

Clearly, it is not easy to compute the function $\mathcal{L}(\Theta \mid Y)$. Therefore, it is useful to complete the observed data Y with the hidden data Z and then compute the probability of the joint events:

$$\mathbb{P}(Y_0 = y_0, ..., Y_{t_H} = y_H, Z_0 = z_0, ..., Z_{t_H} = z_H \mid \Theta)$$

By successively applying the Bayes formula to this collection of events, we can write:

$$\begin{aligned} J &= \mathbb{P}(Z_0 = z_0) \prod_{k=1}^{H} \mathbb{P}(Z_{t_k} = z_k \mid Z_0 = z_0, \ldots, Z_{t_{k-1}} = z_{k-1}, \Theta) \\ &* \prod_{k=1}^{H} \mathbb{P}(Y_{t_k} = y_k \mid Z_0 = z_0, \ldots, Z_{t_H} = z_H, Y_0 = y_0, \ldots, Y_{t_{k-1}} = y_{k-1}, \Theta) \end{aligned}$$

We can simplify the first product term because of the Markovity of Z and the second product term because the conditional probability of Y_{t_k} depends only on $(Y_{t_{k-1}}, Z_{t_k}, Z_{t_{k-1}})$. Then, the expression of J reduces to

$$\begin{aligned} J &= \mathbb{P}(Z_0 = z_0) \prod_{k=1}^{H} \mathbb{P}(Z_{t_k} = z_k \mid Z_{t_{k-1}} = z_{k-1}, \Theta) \\ &\times \prod_{k=1}^{H} \mathbb{P}(Y_{t_k} = y_k \mid Z_{t_k} = z_k, Z_{t_{k-1}} = z_{k-1}, Y_{t_{k-1}} = y_{k-1}, \Theta) \end{aligned}$$

Finally:

$$J = \mathbb{P}(Z_0 = z_0) \prod_{k=1}^{H} T_{z_{k-1}, z_k}(\delta) * f_Y(y_k \mid Z_{t_k} = z_k, Z_{t_{k-1}} = z_{k-1}, Y_{t_{k-1}} = y_{k-1}, \Theta)$$

We recall that the aim of the Expectation-Maximization algorithm is to iteratively maximize the following function over Θ, assuming that we know the set of parameters $\Theta^{(n)}$ after step n of the iteration:

$$\begin{aligned} \mathcal{Q}(\Theta, \Theta^{(n)}) &:= \mathbb{E}^{\mathbb{P}}[\ln \mathbb{P}(\mathbb{Y} = y, \mathbb{Z} \mid \Theta) \mid \mathbb{Y} = y, \Theta^{(n)}] \\ &= \sum_{z \in \mathcal{Z}} \ln \left(\mathbb{P}(\mathbb{Y} = y, \mathbb{Z} = z \mid \Theta)\right) \mathbb{P}(\mathbb{Z} = z \mid \mathbb{Y} = y, \Theta^{(n)}) \end{aligned}$$

where $\mathcal{Z}$ is the set of possible values that can be taken by $\mathbb{Z}$.

Proposition 4.2. *The pseudo-likelihood function $\mathcal{Q}(\Theta, \Theta^{(n)})$ is given by*

$$\mathcal{Q}(\Theta, \Theta^{(n)}) = Q_1(\Theta^{(n)}) + Q_2(\Theta, \Theta^{(n)}) + Q_3(\Theta, \Theta^{(n)}), \tag{4.5}$$

where

$$
\begin{aligned}
Q_1(\Theta^{(n)}) &:= \sum_{j\in E} \ln\left(\mathbb{P}(Z_0 = j)\right) \mathbb{P}(\mathbb{Z}_0 = j \mid \mathbb{Y} = y, \Theta^{(n)}), \\
Q_2(\Theta, \Theta^{(n)}) &:= \sum_{(i,j)\in E^2} \sum_{k=1}^{H} \ln\left(T_{ij}(\delta)\right) \mathbb{P}(Z_{t_{k-1}} = i, Z_{t_k} = j \mid \mathbb{Y} = y, \Theta^{(n)}), \\
Q_3(\Theta, \Theta^{(n)}) &:= \sum_{(i,j)\in E^2} \sum_{k=1}^{H} \ln\left(f_Y(y_k \mid Z_{t_k} = j, Z_{t_{k-1}} = i, y_{k-1}, \Theta)\right) \mathbb{P}(Z_{t_{k-1}} = i, Z_{t_k} = j \mid \mathbb{Y} = y, \Theta^{(n)}).
\end{aligned}
$$

Moreover, we can obtain more specific formulas for

$$
\mathbb{P}(Z_{t_{k-1}} = i, Z_{t_k} = j \mid \mathbb{Y} = y, \Theta^{(n)}) = \frac{w_i(t_{k-1}) v_j(t_k) T_{ij}^{(n)}(\delta) f_Y(y_k \mid Z_{t_k} = j, Z_{t_{k-1}} = i, y_{k-1}, \Theta^{(n)})}{\sum_{k=1}^{H} w_i(t_k)},
$$

Additionally, we can recursively evaluate the quantities w and v for $k \in \{0, \ldots, H\}$:

$$
\begin{aligned}
w_i(t_k) &= \begin{cases} \sum_{j\in E} w_j(t_{k-1}) * f_Y(y_k \mid Z_{t_k} = i, Z_{t_{k-1}} = j, y_{k-1}, \Theta^{(n)}) T_{ji}^{(n)}(\delta) & \text{if } k \in \{1, \ldots, H\} \\ \mathbb{P}(Z_0 = j) & \text{if } k = 0 \end{cases} \\
v_i(t_k) &= \begin{cases} \sum_{j\in E} v_j(t_{k+1}) * f_Y(y_{k+1} \mid Z_{t_{k+1}} = j, Z_{t_k} = i, y_k, \Theta^{(n)}) T_{ij}^{(n)}(\delta) & \text{if } k \in \{0, \ldots, H-1\} \\ 1 & \text{if } k = H \end{cases}
\end{aligned}
$$

Proof
We have

$$
\begin{aligned}
\mathcal{Q}(\Theta, \Theta^{(n)}) &= \mathbb{E}[\ln(\mathbb{P}(\mathbb{Y} = y, \mathbb{Z} \mid \Theta)) \mid \mathbb{Y} = y, \Theta^{(n)}] \\
&= \sum_{z\in\mathcal{Z}} \ln(\mathbb{P}(\mathbb{Y} = y, \mathbb{Z} = z \mid \Theta)) \mathbb{P}(\mathbb{Z} = z \mid \mathbb{Y} = y, \Theta^{(n)}) \\
&= \underbrace{\sum_{z\in\mathcal{Z}} \ln(\mathbb{P}(Z_0 = z_0)) \mathbb{P}(\mathbb{Z} = z \mid \mathbb{Y} = y, \Theta^{(n)})}_{Q_1(\Theta^{(n)})} \\
&+ \underbrace{\sum_{z\in\mathcal{Z}} \sum_{k=1}^{H} \ln(T_{z_{k-1}, z_k}(\delta)) \mathbb{P}(\mathbb{Z} = z \mid \mathbb{Y} = y, \Theta^{(n)})}_{Q_2(\Theta, \Theta^{(n)})} \\
&+ \underbrace{\sum_{z\in\mathcal{Z}} \sum_{k=1}^{H} \ln\left(f_Y(y_k \mid z_k, z_{k-1}, y_{k-1}, \Theta)\right) \mathbb{P}(\mathbb{Z} = z \mid \mathbb{Y} = y, \Theta^{(n)})}_{Q_3(\Theta, \Theta^{(n)})}
\end{aligned}
$$

Because Z_t is a continuous time Markov chain, we can obtain the expression of Q_1, Q_2 and Q_3 shown in the proposition. Regarding $w_i(t_k)$ and $v_i(t_k)$, the recursive relations are derived according what follows:

$$\begin{aligned}
w_i(t_k) &= \mathbb{P}(Z_{t_k} = i, Y_0 = y_0, \ldots, Y_{t_k} = y_k) \\
&= \sum_{j \in E} \mathbb{P}(Z_{t_k} = i, Z_{t_{k-1}} = j, Y_0 = y_0, \ldots, Y_{t_k} = y_k) \\
&= \sum_{j \in E} \mathbb{P}(Z_{t_{k-1}} = j, Y_0 = y_0, \ldots, Y_{t_{k-1}} = y_{k-1}) \\
&\quad \times \mathbb{P}(Z_{t_k} = i, Y_{t_k} = y_k \mid Z_{t_{k-1}} = j, Y_0 = y_0, \ldots, Y_{t_{k-1}} = y_{k-1}) \\
&= \sum_{j \in E} w_j(t_{k-1}) \mathbb{P}(Z_{t_k} = i, Y_{t_k} = y_k \mid Z_{t_{k-1}} = j, Y_0 = y_0, \ldots, Y_{t_{k-1}} = y_{k-1}) \\
&= \sum_{j \in E} w_j(t_{k-1}) f_Y(y_k \mid Z_{t_k} = i, Z_{t_{k-1}} = j, y_{k-1}, \Theta^{(n)}) \times T_{ji}^{(n)}(\delta)
\end{aligned}$$

for $k \in [1, H]$ and with $w_i(0) = \mu_i$.

$$\begin{aligned}
v_i(t_k) &= \mathbb{P}(Y_{t_{k+1}} = y_{k+1}, \ldots, Y_T = y_H \mid Z_{t_k} = i, Y_{t_k} = y_k) \\
&= \sum_{j \in E} \mathbb{P}(Z_{t_{k+1}} = j, Y_{t_{k+1}} = y_{k+1}, \ldots, Y_T = y_H \mid Z_{t_k} = i, Y_{t_k} = y_k) \\
&= \sum_{j \in E} \mathbb{P}(Y_{t_{k+2}}, \ldots, Y_T = y_H \mid Z_{t_{k+1}} = j, Z_{t_k} = i, Y_{t_{k+1}} = y_{k+1}, Y_{t_k} = y_k) \\
&\quad \times \mathbb{P}(Y_{t_{k+1}} = y_{k+1}, Z_{t_{k+1}} = j \mid Z_{t_k} = i, Y_{t_k} = y_k) \\
&= \sum_{j \in E} \mathbb{P}(Y_{t_{k+2}}, \ldots, Y_T = y_H \mid Z_{t_{k+1}} = j, Z_{t_k} = i, Y_{t_{k+1}} = y_{k+1}) \\
&\quad \times \mathbb{P}(Y_{t_{k+1}} = y_{k+1} \mid Z_{t_{k+1}} = j, Z_{t_k} = i, Y_{t_k} = y_k) \mathbb{P}(Z_{t_{k+1}} = j \mid Z_{t_k} = i) \\
&= \sum_{j \in E} v_j(t_{k+1}) f_Y(y_{k+1} \mid Z_{t_{k+1}} = j, Z_{t_k} = i, y_k, \Theta^{(n)}) \times T_{ij}^{(n)}(\delta)
\end{aligned}$$

for $k \in [0, H-1]$ and with $v_i(t_H) = 1$.

Algorithm

1. Start with an initial vector set $\Theta^{(0)} := \{(\theta_i^{(0)}, \kappa_i^{(0)}, \xi_i^{(0)})_{i \in E}, \Pi^{(0)}\}$. Let us fix N the maximum number of iterations authorized of the algorithm and a positive constant ϵ for the estimated pseudo log-likelihood function.

2. Assume that we are at the $n+1 < N$ step and that the n step yielded to the vector set $\Theta^{(n)} := \{(\theta_i^{(n)}, \kappa_i^{(n)}, \xi_i^{(n)})_{i \in E}, \Pi^{(n)}\}$.

   – **E-Step:**
   This step consists of computing the function $\mathcal{Q}(\Theta, \Theta^{(n)})$ following Proposition 4.2.

   – **M-Step:**
   The computation of $\theta_i^{(n+1)}$, $\kappa_i^{(n+1)}$, $\xi_i^{(n+1)}$ and $\Pi_{ij}^{(n+1)}$ is achieved by maximizing the function

$\mathcal{Q}(\Theta, \Theta^{(n)})$:

$$\Theta^{(n+1)} = \arg\max_{\Theta}(\mathcal{Q}(\Theta, \Theta^{(n)}))$$

4.2 Numerical results and interpretations

To estimate our model, we use daily data ($\delta = 1/252$) provided by the CBOE for the VIX index between January 2000 and January 2015. The maximum number of iterations of the EM algorithm is set at $N = 100$, and we use the Matlab function *fmincon* to maximize the pseudo likelihood function $\mathcal{Q}(\Theta, \Theta^{(n)})$ at each iteration n. The results of the maximum likelihood estimation at the last iteration of the algorithm are given below:

Regimes	κ^{ML}	θ^{ML}	ξ^{ML}
state 1	9.44	0.0172	0.18
state 2	13.72	0.0525	0.48
state 3	14.04	0.24	1.49

$$Q^{ML} = \begin{pmatrix} -13.6762 & 13.5095 & 0.1667 \\ 15.1125 & -18.9127 & 3.8002 \\ 0.4061 & 35.5938 & -35.9999 \end{pmatrix}$$

This generator indicates that from one day to the next, the state of the Markov chain may switch (or not) to another state according to the transition probability matrix for one day given by $T^{ML} = e^{Q^{ML}\delta}$ in 4.6, which reveals an important inertia for each regime, giving a true economic sense to each: state 1 corresponds to a low-volatility regime, state 2 is a transition regime, and state 3 represents a high-volatility regime. The transition regime discriminates quite effectively between short-term spikes, which are quite frequent in the market (but also revert frequently), and highly stressed situations (see 2001, 2008, 2010 and 2011 in Figure 5).

Therefore, the model is sufficiently flexible to reproduce both short-term and long-term features of volatility.

$$T^{ML} = \begin{pmatrix} 0.9487 & 0.0503 & 0.0010 \\ 0.0563 & 0.9302 & 0.0136 \\ 0.0053 & 0.1268 & 0.8678 \end{pmatrix} \tag{4.6}$$

The results of the maximum likelihood estimation clearly exhibit three different levels, especially for the mean level of volatility and the volatility of volatility. The filtering function for the hidden Markov chain is deduced from the estimation procedure and is given by:

$$\hat{Z}_t = \arg\max_{i} \mathbb{P}(Z_t = i \mid \mathbb{Y}, \Theta^{ML})$$

where Θ^{ML} is the set of parameters given by the maximum likelihood. The result of the filtering procedure is shown in Figure 5.

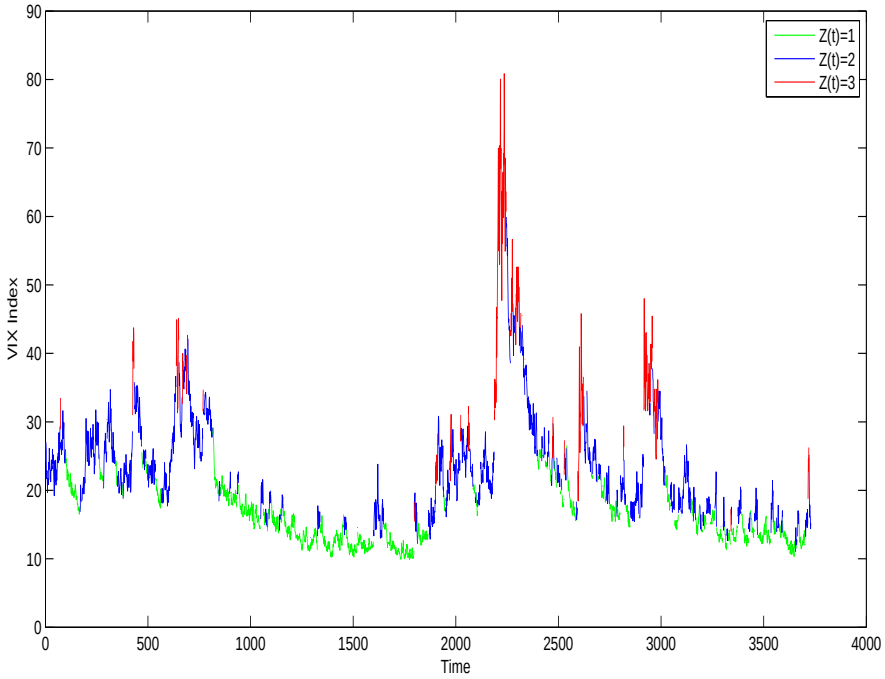

Figure 5: The VIX index and the filtered regimes between January 2000 *and January* 2015

4.3 Convergence of the algorithm

In this section, we show that the convergence of the developed algorithm can be attributed to the general theory of the EM method. The paper of Fort and Moulines [26] deals with the convergence of the EM algorithm using Monte Carlo method. This is s a specific case of the EM algorithm when we use Monte Carlo method at the step Expectation of this algorithm. However, in our case, we derive at the E step, an analytical procedure based on the Baum-Welch Algorithm instead of using Monte Carlo method. This allows us to have better results both in term of estimations error and in term of speed of the algorithm calculus.

4.3.1 Theoretical convergence

Let n and L be positive integers and Θ be a subset of $\mathbb{R}^n$. Let Z_t be a continuous time Markov Chain with a transition probability matrix $A = (a_{ij})$ and initial probability distribution $a = (a_i)$, and let $\mu(z)$ be the probability distribution of $(Z_{t_0}, \ldots, Z_{t_L})$. Let V_t be a Markovian stochastic process with a transition probability function $f_i(t, v, s, u) := \mathbb{P}(V_t = v \mid V_s = u, Z_s = i)$. Let M be a function defined from $\mathbb{R} \times E$ to $\mathbb{R}$ such that for each $i \in E$, the function $m_i(v, \theta) := M(v, i, \theta)$ is a diffeomorphism of $\mathbb{R}$. Let $Y_t := M(V_t, Z_t, \theta)$ be a stochastic process, and let $\Psi_{ij}(t, x, s, y)$ be a family of transition probability functions defined as follows:

$$\begin{aligned} \Psi_{ij}(t, x, s, y) : &= \mathbb{P}(Y_t = x \mid Y_s = y, Z_s = i, Z_t = j) \\ &= \frac{1}{m_j'(m_j^{-1}(x))} f_j(t, m_j^{-1}(x), s, m_i^{-1}(y)) \end{aligned}$$

For the triple $A, a, \Psi = \{\Psi_{ij}\}$, we define the density of the stochastic process $(Y_{t_0}, \ldots, Y_{t_L})$ for which we have the set of observations $y := (y_0, y_1, \ldots, y_L)$:

$$
\begin{aligned}
&P(A, a, \Psi)\{Y_{t_0} = y_0, Y_{t_1} = y_1, \ldots, Y_{t_L} = y_L\} \\
&= \sum_{z_{t_0}, \ldots, z_L} a_{z_{t_0}} a_{z_{t_0} z_{t_1}} \Psi_{z_{t_0} z_{t_1}}(t_1, y_1, t_0, y_0) a_{z_{t_1} z_{t_2}} \Psi_{z_{t_1} z_{t_2}}(t_2, y_2, t_1, y_1) \ldots a_{z_{T-1} z_T} \Psi_{z_{t_{L-1}} z_{t_L}}(t_L, y_L, t_{L-1}, y_{L-1})
\end{aligned}
$$

Assuming that for each $\theta \in \Theta$ there is a smooth assignment $\theta \to (A(\theta), a(\theta), \Psi(\theta))$, we will denote this expression $P(\theta)$. We observe that $P(\theta)$ is of the form $P(\theta) = \sum_{z \in E^{L+1}} p(z, y, \theta)$ with:

$$
\begin{aligned}
p(z, y, \theta) &= a_{z_{t_0}}(\theta) \prod_{k=1}^{L} a_{z_{t_{k-1}} z_{t_k}}(\theta) \Psi_{z_{t_{k-1}} z_{t_k}}(t_k, y_k, t_{k-1}, y_{k-1}, \theta) \\
&= a_{z_{t_0}}(\theta) \prod_{k=1}^{L} a_{z_{t_{k-1}} z_{t_k}}(\theta) \frac{1}{m'_{z_{t_k}}(m^{-1}_{z_{t_k}}(y_k, \theta), \theta)} f_{z_{t_k}}(t_k, m^{-1}_{z_{t_k}}(y_k, \theta), t_{k-1}, m^{-1}_{z_{t_{k-1}}}(y_{k-1}, \theta), \theta)
\end{aligned}
$$

Now, for each couple $(\theta, \theta') \in \Theta$, we define the following function:

$$
\begin{aligned}
Q(\theta, \theta') : &= E_z \left[\log\left(p(z, y, \theta)\right) \mid Y = y, \theta' \right] \\
&= \sum_z \nu_y(z, \theta') \log\left(p(z, y, \theta)\right)
\end{aligned}
$$

where $\nu_y(z, \theta') := \frac{p(z, y, \theta')}{P(\theta')}$ is the probability distribution of Z conditional upon Y. Finally, let $p_z(y, \theta)$ be the probability distribution of Y conditional upon Z. We thus have

$$
\begin{aligned}
\Delta(\theta, \theta') : &= \log\left(P(\theta)\right) - \log\left(P(\theta')\right) \\
&= \log\left(\sum_z p_z(y, \theta) \mu(z, \theta)\right) - \sum_z \nu_y(z, \theta') \log\left(P(\theta')\right) \\
&= \log\left(\sum_z \frac{p_z(y, \theta) \mu(z, \theta)}{\nu_y(z, \theta')} \nu_y(z, \theta')\right) - \sum_z \nu_y(z, \theta') \log\left(P(\theta')\right)
\end{aligned}
$$

Given that $\sum_z \nu_y(z, \theta') = 1$, we apply Jensen's inequality and obtain

$$
\begin{aligned}
\Delta(\theta, \theta') &\geq \sum_z \nu_y(z, \theta') \log\left(\frac{p_z(y, \theta) \mu(z, \theta)}{\nu_y(z, \theta')}\right) - \sum_z \nu_y(z, \theta') \log\left(P(\theta')\right) \\
&= \sum_z \nu_y(z, \theta') \log\left(\frac{p_z(y, \theta) \mu(z, \theta)}{\nu_y(z, \theta') P(\theta')}\right) \\
&= \sum_z \nu_y(z, \theta') \log\left(\frac{p(z, y, \theta)}{p(z, y, \theta')}\right) \\
&=: \delta(\theta, \theta')
\end{aligned}
$$

Finally, denoting:

$$\theta^{''} := \arg\max_{\Theta} \delta(\theta, \theta^{'}) = \arg\max_{\Theta} Q(\theta, \theta^{'})$$

We have:

$$\begin{aligned} \log\left(P(\theta^{''})\right) - \log\left(P(\theta^{'})\right) &= \Delta(\theta^{''}, \theta^{'}) \\ &\geq \delta(\theta^{''}, \theta^{'}) \\ &\geq \delta(\theta^{'}, \theta^{'}) \\ &\geq 0 \end{aligned}$$

Now, we have shown the monotone increase in the likelihood at each step of the proposed algorithm. The convergence toward a local maximum is a classical result related to the EM theory.

4.3.2 Empirical test of convergence

Now, we aim to empirically test the convergence of the algorithm. To do so, we perform several batches of Monte Carlo simulations, increasing the number of simulations at each batch from $N = 100$ to $N = 10000$ simulations. We use sample lengths of 1000 daily observations. Following the approach of Ait-Sahalia and Kimmel [2], we compare the accuracy of the developed method with an approximate likelihood procedure in which the unobservable volatility state is replaced by a proxy: the VIX index. As a result, we gain in tractability because the update of Θ^n at each step of the EM algorithm is obtained explicitly.

As shown in Figures 6 and 7, we find that both of the procedures converge to a biased value because of the sampling error. Comparing the bias of the estimators for a sufficiently large number of N to annihilate the variance resulting from the Monte Carlo simulations, we deduce that the exact likelihood procedure gives a more accurate result (a smaller bias) and that this additional bias is small compared to the bias resulting from the sampling error. Indeed, the resulting error is augmented by a factor of 10% to 15% for each parameter in our observations, except for the parameters ξ, for which the approximation procedure does not produce an additional error because the functions α and β do not depend on ξ.

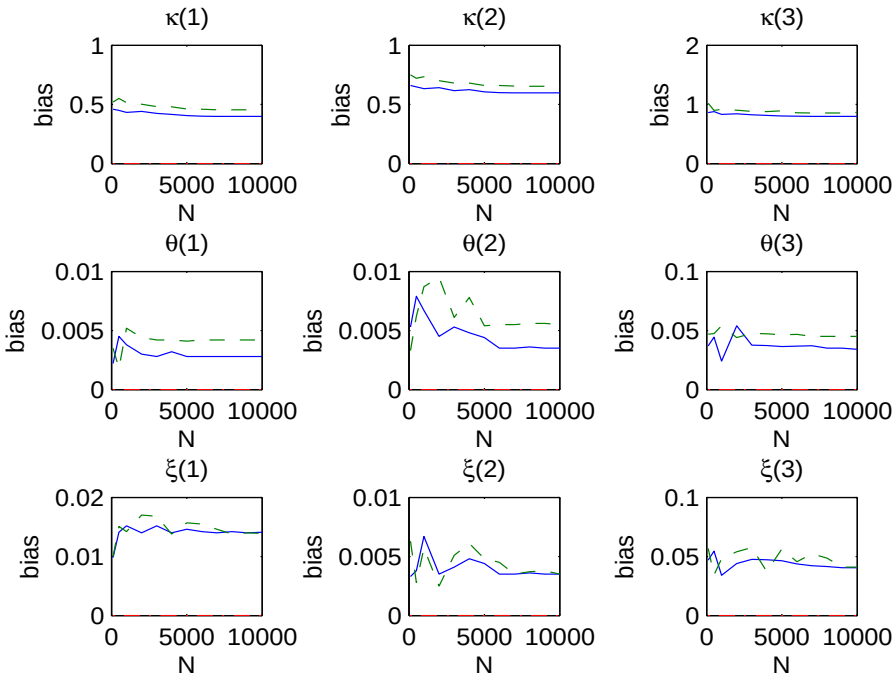

Figure 6: Outputs of the Monte Carlo simulations for κ,θ and ξ for the complete maximum likelihood (solid line) versus the approximate maximum likelihood (dashed line). For each batch of simulations, we compute the estimated absolute bias of the estimator.

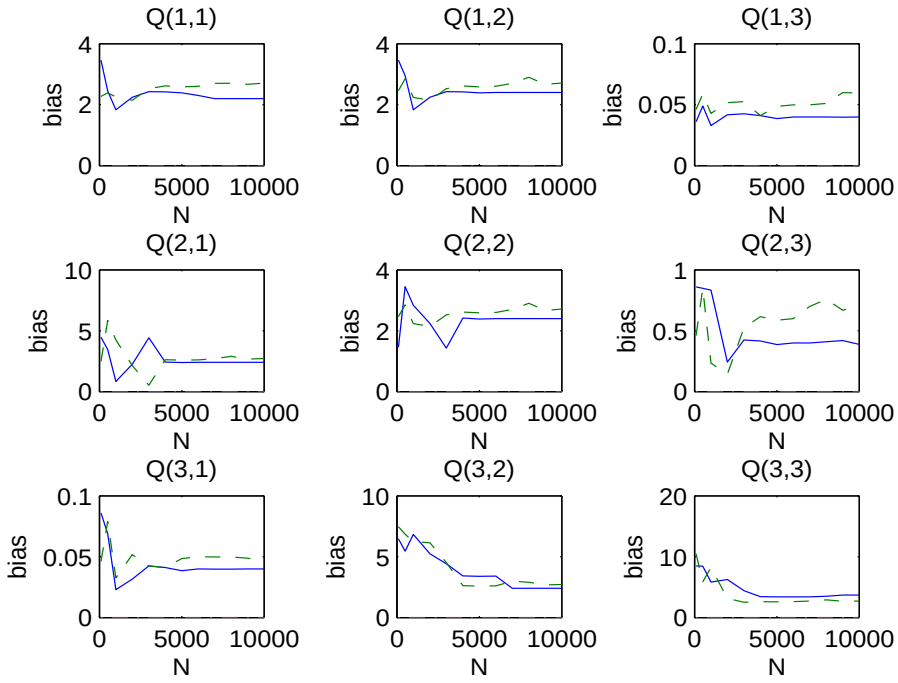

Figure 7: Outputs of the Monte Carlo simulations for Q for the complete maximum likelihood (solid line) versus the approximate maximum likelihood (dashed line). For each batch of simulations, we compute the estimated absolute bias of the estimator.

5 Consistent Pricing of VIX and S&P500 European Options

Given the ability of the proposed model to capture some important features of volatility dynamics for the variance process of the S&P500, we aim to explore whether this model can be used as a pricing model. It is widely known that the commonly used and industry-standard Heston stochastic volatility model is not able to fit VIX option prices because the positive smile implies a much wider support for the probability distribution of the variance process than implied by the square-root process. This issue has recently and successfully been addressed in Papanicolau and Sircar [36], where a Heston-type volatility process is multiplied by a regime-switching factor. The main difference here is that a joint calibration of both S&P500 and VIX options, enabled by the fact that our regime-switching stochastic volatility model offers more degrees of freedom, is proposed. Below, we take advantage of the piecewise affine property of the volatility process to derive analytical formulas for S&P500 and VIX options up to the numerical integration of multidimensional integrals, which are computed using a Monte Carlo method. The small regime shift rate exhibited in the maximum likelihood estimation through 15 years of sample data makes these numerical integrations feasible.

5.1 Pricing VIX European options

In this section, we derive a semi-analytical pricing formula for a European option written on the VIX Index with a maturity T and a strike K. For this purpose, we use a semi-analytical expression of the characteristic function of V_T conditional upon the state of the Markov chain at time T to obtain the pricing formula by Fourier inversion. First, let us introduce a preliminary lemma that provides a useful function to compute the price of the VIX option. This lemma is closely related to derivations of the characteristic function of processes whose volatility is driven by a Cox-Ingersoll-Ross process with piecewise constant parameters, as considered in Wu and Zhang [42] for the pricing of interest rate derivatives. Our approach derives the characteristic function of such a CIR process relying on previous results in Goutte and Ngoupeyou [28], where is derived the characteristic function of the integral of a regime switching CIR process.

Lemma 5.2. *Let M be a discrete random variable in $\mathbb{N}^*$, such that $M-1$ is the random number of jumps of the Markov chain Z between 0 and T. Let $(\tau_1, \ldots, \tau_{M-1})$ be a sequence of $M-1$ random jump times, and let us denote by $\Delta_{t_k} := \tau_{k+1} - \tau_k$ for each $k \in \{0, \ldots, M-1\}$ with $\tau_M = T$ and $\tau_0 = 0$. For every $x \in \mathbb{R}$, there exist two families of functions: $\{B_j\}_{0 \leq j \leq M-1}$ and $\{A_j\}_{0 \leq j \leq M-1}$, such that:*

$$\mathbb{E}^{\mathbb{P}}\left[e^{ixV_T} \mid V_0 = v_0, \mathcal{F}_T^Z\right] = \exp\left((ix - A_0(\Delta_{t_0}, Z_{\tau_0})v_0 + \sum_{j=0}^{M-1}(-B_j(\Delta_{t_j}, Z_{\tau_j}) + ix\Delta_{t_j}\theta(Z_{\tau_j})\kappa(Z_{\tau_j}))\right)$$

The functions $\{B_j\}_{0 \leq j \leq M-1}$ and $\{A_j\}_{0 \leq j \leq M-1}$ are determined recursively following the resolution of

successive systems of Ricatti equations, such that for each $\Delta_t \in [0,T]$ and $z \in E$:

$$\begin{aligned} A_j(\Delta_t, z) &= a_{1,z}\frac{1-\frac{a_{2,z}}{a_{1,z}}C_{j,z}e^{-\sqrt{\kappa(z)}\Delta_t}}{1-C_{j,z}e^{-\sqrt{\kappa(z)}\Delta_t}} \\ B_j(\Delta_t, z) &= a_{1,z}\tilde{\kappa}(z)\tilde{\theta}(z)(\Delta_t + \frac{1}{\sqrt{\kappa(z)}}(1-\frac{a_{2,z}}{a_{1,z}})\ln\frac{C_{j,z}e^{-\sqrt{\kappa(z)}\Delta_t}-1}{C_{j,z}-1}) \end{aligned}$$

where for each $j \in \{0,\dots,M-1\}$:

$$a_{1,z} = -\frac{\tilde{\kappa}(z)-\sqrt{\kappa}}{2\xi(z)^2} \qquad \gamma(z) = -\kappa(z)ix - \frac{1}{2}x^2\xi(z)^2$$

$$a_{2,z} = -\frac{\tilde{\kappa}(z)+\sqrt{\kappa}}{2\xi(z)^2} \qquad \tilde{\theta}(z) = \frac{\kappa(z)\theta(z)}{\tilde{\kappa}(z)}$$

$$\tilde{\kappa}(z) = \kappa(z) - ix\xi(z)^2 \qquad C_{j,z} = \begin{cases} \frac{a_{1,z}-A_{j+1}(\Delta_{t_{j+1}},Z_{\tau_{j+1}})}{a_{2,z}-A_{j+1}(\Delta_{t_{j+1}},Z_{\tau_{j+1}})} & \text{if } j \in \{0,\dots,M-2\} \\ 0 & \text{if } j = M-1 \end{cases}$$

Proof
The proof of this lemma can be found in the appendices. □

Thus, we can give the expression of the characteristic function of V_T conditional upon the state of the Markov chain at time T.

Proposition 5.3. *Let M be an integer in $\mathbb{N}^*$ and z_0 an element of E. With each set of M dwell times $(\Delta_{t_0},\dots,\Delta_{t_{M-1}})$ of the continuous time Markov chain Z and each sequence $s \in S^T_{z_0,M}$, we can associate a unique path of Z containing $M-1$ jumps. For each of these paths and for each $j \in E$ and $x \in \mathbb{R}$, we define the following functions:*

$$\Xi^j_M(v_0,x,s,(\Delta_{t_k})_{0\le k\le M-1}) =$$
$$\mathbb{I}_{\{s_M=j\}} \times \exp\left((ix - A_0(\Delta_{t_0},s_0)v_0 + \sum_{k=0}^{M-1}(-B_k(\Delta_{t_k},s_k) + ix\Delta_{t_k}\theta(s_k)\kappa(s_k))\right)$$

Finally, for each $j \in E$, let us denote by Φ_j the characteristic function of V_T conditional upon $Z_T = j$ and $(V_0 = v_0, Z_0 = z_0)$. For every $x \in \mathbb{R}$, $\phi_j(x,v_0,z_0)$ is given by:

$$\Phi_j(x,v_0,z_0) := \mathbb{E}^{\mathbb{P}}_{v_0,z_0}[e^{ixV_T} \mid Z_T = j]$$

$$= \frac{1}{T_{z_0,j}(T)}\sum_{M=1}^{\infty}\sum_{s\in S_{z_0,M}}\underbrace{\int\cdots\int}_{\sum_{k=0}^{M-1}\Delta_{t_k}<T}\Xi^j_M(v,x,s,(\Delta_{t_k})_{0\le k\le M-1})\mathcal{L}^T_M(s,(\Delta_{t_k})_{0\le k\le M-1})\,d\Delta_{t_0}\dots d\Delta_{t_{M-1}}$$

Proof
According to Bayes, the following holds:

$$\begin{aligned}
\Phi_j(x, v_0, z_0) &= \mathbb{E}^{\mathbb{P}}\left[e^{ixV_T} \mid Z_T = j, V_0 = v_0, Z_0 = z_0\right] \\
&= \frac{\mathbb{E}^{\mathbb{P}}\left[e^{ixV_T}\mathbb{I}_{\{Z_T=j\}} \mid V_0 = v_0, Z_0 = z_0\right]}{\mathbb{P}(Z_T = j \mid V_0 = v_0, Z_0 = z_0)} \\
&= \frac{\mathbb{E}^{\mathbb{P}}_{v_0,z_0}\left[\mathbb{I}_{\{Z_T=j\}}\mathbb{E}^{\mathbb{P}}[e^{ixV_T} \mid \mathcal{F}_T^Z, V_0 = v_0]\right]}{\mathbb{P}(Z_T = j \mid Z_0 = z_0)} \\
&= \frac{\mathbb{E}^{\mathbb{P}}_{v_0,z_0}\left[\mathbb{I}_{\{Z_T=j\}}\mathbb{E}^{\mathbb{P}}[e^{ixV_T} \mid \mathcal{F}_T^Z, V_0 = v_0]\right]}{T_{z_0,j}(T)}
\end{aligned}$$

By using the lemma, we have the expression of $\mathbb{E}^{\mathbb{P}}[e^{ixV_T} \mid \mathbb{F}_T^Z, V_0 = v]$, which we multiply by $\mathbb{I}_{\{s_M=j\}}$ to obtain the expression of the function Ξ_M^j. Finally, the function $\mathcal{L}_M$ for the likelihood of a given path of a continuous time Markov Chain with $M-1$ jumps is well known. □

With an expression of the characteristic function of V_T conditional upon $(Z_T = j, V_0 = v_0, Z_0 = z_0)$, we can obtain the density function of V_T conditional upon $(Z_T = j, V_0 = v_0, Z_0 = z_0)$ by Fourier inversion:

$$\mathcal{P}_{V_T|(Z_T=j,V_0=v_0,Z_0=z_0)}(v) = \frac{1}{2\pi}\int_{\mathbb{R}} \Phi_j(x, v_0, z_0)e^{-ixv}dx$$

Proposition 5.4. *The price at $t = 0$ of a call option written on the VIX Index with maturity T and strike K is given by*

$$Call_0(v_0, z_0, K, T) = e^{-rT}\sum_{j\in E} T_{z_0,j}(T)\int_{\mathbb{R}^+} h(v, j)\mathcal{P}_{V_T|(Z_T=j,V_0=v_0,Z_0=z_0)}(v)dv$$

where h is a function defined by:

$$h(v, j) = \left(\sqrt{\alpha(j)v + \beta(j)} - K\right)_+$$

Proof

$$\begin{aligned}
Call_0(v_0, z_0, K, T) &= e^{-rT}\mathbb{E}^{\mathbb{P}}[(VIX_T - K)_+ \mid V_0 = v_0, Z_0 = z_0] \\
&= e^{-rT}\mathbb{E}^{\mathbb{P}}[(\sqrt{\alpha(Z_T)V_T + \beta(Z_T)} - K)_+ \mid V_0 = v_0, Z_0 = z_0] \\
&= e^{-rT}\sum_{j\in E}\mathbb{P}(Z_T = j \mid Z_0 = z)\int_{\mathbb{R}^+} h(v, j)\mathcal{P}_{V_T|(Z_T=j,V_0=v_0,Z_0=z_0)}(v)dv \\
&= e^{-rT}\sum_{j\in E} T_{z_0,j}(T)\int_{\mathbb{R}^+} h(v, j)\mathcal{P}_{V_T|(Z_T=j,V_0=v_0,Z_0=z_0)}(v)dv
\end{aligned}$$

where h is a function defined by:

$$h(v,j) = \left(\sqrt{\alpha(j)v + \beta(j)} - K\right)_+$$

□

5.2 Pricing S&P500 European options

In this section, we derive a semi-analytical pricing formula for a European option written on the S&P500 Index with a maturity T and a strike K. For this purpose, we use a semi-analytical expression of the characteristic function of $X_T := \ln(\frac{S_t}{S_0}) - rT$ to obtain the pricing formula by Fourier inversion. The key element in the subsequent lemma is defining a new probability measure retaining the affine structure of the variance diffusion.

Lemma 5.3. *Let $\Phi(u,v,z_0,x)$ be the characteristic function of X_T. The following expression holds:*

$$\Phi(u,v,z,x) := \mathbb{E}^{\mathbb{P}}\left[e^{iuX_T} \mid X_0 = x, V_0 = v, Z_0 = z\right]$$

$$= e^{iux}\sum_{M=1}^{\infty}\sum_{s\in S_{z_0,M}}\underbrace{\int\cdots\int}_{\sum_{k=0}^{M-1}\Delta_{t_k}<T}\Xi_M(v,u,s,(\Delta_{t_k})_{0\le k\le M-1})\mathcal{L}_M^T(s,(\Delta_{t_k})_{0\le k\le M-1})\,d\Delta_{t_0}\ldots d\Delta_{t_{M-1}}$$

where $\Xi_M(v,u,s,(\Delta_{t_k})_{0\le k\le M-1})$ is a function defined by:

$$\Xi_M(v,u,s,(\Delta_{t_k})_{0\le k\le M-1}) := \exp\left(-A_0^s(\Delta_{t_0},s_0)v_0 + \sum_{k=0}^{M-1}(-B_k^s(\Delta_{t_k},s_k)\right)$$

where the functions A^s and B^s have the same expression as A and B in Lemma 5.2 with $\gamma = -\frac{u^2+iu}{2}$ and $\tilde{\kappa} = \kappa - iu\rho\xi$.

Proposition 5.5. *Let $P(s_0,z_0,v_0,K,T)$ be the price of a call option on the $S\&P500$ Index with a maturity T and a strike K at time $t=0$.*

$$\begin{aligned} P(s_0,z_0,v_0,K,T) &= e^{-rT}\mathbb{E}^{\mathbb{P}}[(S_T-K)_+ \mid V_0=v_0, Z_0=z_0, S_0=s_0] \\ &= s_0\int_{\mathbb{R}^+} g(x)\mathcal{P}_{X_T\mid(V_0=v_0,Z_0=z_0,X_0=x_0)}(x)dx \end{aligned}$$

where g and $\mathcal{P}_{X_T|(V_0=v_0,Z_0=z_0,X_0=x_0)}$ are functions defined by:

$$\begin{aligned} g(x) &= \left(e^x - \frac{Ke^{-rT}}{S_0}\right)_+ \\ \mathcal{P}_{X_T|(V_0=v_0,Z_0=z_0,X_0=x_0)}(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} \Phi(u, v_0, z_0, x_0)e^{-iux}du \end{aligned}$$

5.3 Numerical results & calibration

To numerically compute the analytical expression for VIX option prices, we must derive an approximation for the infinite sum in the expression of Φ_j. This approximation, denoted by $\hat{\Phi}_j$, is obtained by truncating the infinite sum up to a positive integer $\bar{M}$, which represents the maximum number of jumps considered when we integrate over all the possible paths of the continuous time Markov Chain between 0 and T. For each $x \in \mathbb{R}$, $v_0 \in \mathbb{R}^+$ and $z_0 \in E$, $\hat{\Phi}_j(x, v_0, z_0)$ is given by:

$$\begin{aligned} &\hat{\Phi}_j(x, v_0, z_0) \\ &:= \frac{1}{T_{z_0,j}(T)} \sum_{M=1}^{\hat{M}} \sum_{s\in S_{z_0,M}} \underbrace{\int\cdots\int}_{\sum_{k=0}^{M-1}\Delta_{t_k}<T} \Xi_M^j(v, x, s, (\Delta_{t_k})_{0\le k\le M-1})\mathcal{L}_M^T(s, (\Delta_{t_k})_{0\le k\le M-1})\, d\Delta_{t_0}\dots d\Delta_{t_{M-1}} \end{aligned}$$

The following technical lemma will allow us to derive a bound for the error incurred by truncating the infinite sum.

Lemma 5.4. *Let $M^{\mathcal{Z}}$ be the random number of jumps of the continuous time Markov Chain Z between 0 and T, and let $\bar{M}$ be a positive integer. Let $\Lambda := \max(Q)$ be the highest transition rate of Z; then, the following inequality holds:*

$$\mathbb{P}(M^{\mathcal{Z}} \ge \bar{M} + 1) \le e^{-\Lambda T} \sum_{M=\bar{M}+1}^{\infty} \frac{(\Lambda T)^M}{M!}$$

Now, we can derive a bound for the error between Φ_j and $\hat{\Phi}_j$:

Lemma 5.5. *For each $x \in \mathbb{R}$, $v_0 \in \mathbb{R}^+$ and $(z_0, j) \in E^2$, the following inequality holds:*

$$|\hat{\Phi}_j(x, v_0, z_0) - \Phi_j(x, v_0, z_0)| \le \frac{1}{T_{z_0,j}(T)} e^{-\Lambda T} \sum_{M=\bar{M}+1}^{\infty} \frac{(\Lambda T)^M}{M!}$$

Proof We have for each $x \in \mathbb{R}$, $v_0 \in \mathbb{R}^+$ and $(z_0, j) \in E^2$:

$$
\begin{aligned}
&|\hat{\Phi}_j(x, v_0, z_0) - \Phi_j(x, v_0, z_0)| \\
&= |\frac{1}{T_{z_0,j}(T)} \sum_{M=\bar{M}+1}^{\infty} \sum_{s \in S_{z_0,M}} \underbrace{\int \cdots \int}_{\sum_{k=0}^{M-1} \Delta_{t_k} < T} \Xi_M^j(v_0, x, s, (\Delta_{t_k})_{0 \le k \le M-1}) \mathcal{L}_M^T(s, (\Delta_{t_k})_{0 \le k \le M-1}) \, d\Delta_{t_0} \dots d\Delta_{t_{M-1}}| \\
&\le \frac{1}{T_{z_0,j}(T)} \sum_{M=\bar{M}+1}^{\infty} \sum_{s \in S_{z_0,M}} \underbrace{\int \cdots \int}_{\sum_{k=0}^{M-1} \Delta_{t_k} < T} |\Xi_M^j(v_0, x, s, (\Delta_{t_k})_{0 \le k \le M-1})| \mathcal{L}_M^T(s, (\Delta_{t_k})_{0 \le k \le M-1}) \, d\Delta_{t_0} \dots d\Delta_{t_{M-1}} \\
&\le \frac{1}{T_{z_0,j}(T)} \sum_{M=\bar{M}+1}^{\infty} \sum_{s \in S_{z_0,M}} \underbrace{\int \cdots \int}_{\sum_{k=0}^{M-1} \Delta_{t_k} < T} \mathcal{L}_M^T(s, (\Delta_{t_k})_{0 \le k \le M-1}) \, d\Delta_{t_0} \dots d\Delta_{t_{M-1}} \\
&\le \frac{1}{T_{z_0,j}(T)} e^{-\Lambda T} \sum_{M=\bar{M}+1}^{\infty} \mathbb{P}(M^{\mathcal{Z}} = M) \\
&\le \frac{1}{T_{z_0,j}(T)} e^{-\Lambda T} \mathbb{P}(M^{\mathcal{Z}} \ge \bar{M} + 1)
\end{aligned}
$$

Thus, using Lemma 5.4, we obtained the announced result.
Now, let us define the approximation for the VIX option price that we will use in the calibration procedure. This approximation also requires truncating the infinite integrals and is given by:

$$
\hat{Call}_0(v_0, z_0, K, T) := e^{-rT} \sum_{j \in E} T_{z_0,j}(T) \int_0^{v_m} h(v, j) \hat{\mathcal{P}}_{V_T|(Z_T=j, V_0=v_0, Z_0=z_0)}(v) dv
$$

where $\hat{\mathcal{P}}_{V_T|(Z_T=j, V_0=v_0, Z_0=z_0)}$ is defined by:

$$
\hat{\mathcal{P}}_{V_T|(Z_T=j, V_0=v_0, Z_0=z_0)}(v) = \frac{1}{2\pi} \int_{-x_m}^{x_m} \hat{\Phi}_j(x, v_0, z_0) e^{-ixv} dx
$$

We can thus write the following inequalities:

$$
\begin{aligned}
&|\hat{Call}_0(v_0, z_0, K, T) - Call_0(v_0, z_0, K, T)| \\
&\le e^{-rT} \sum_{j \in E} T_{z_0,j}(T) \int_0^{v_m} h(v, j) |\mathcal{P}_{V_T|(Z_T=j, V_0=v_0, Z_0=z_0)} - \hat{\mathcal{P}}_{V_T|(Z_T=j, V_0=v_0, Z_0=z_0)}|(v) dv \\
&\le e^{-rT} \sum_{j \in E} T_{z_0,j}(T) \int_0^{v_m} h(v, j) \frac{1}{2\pi} \int_{-x_m}^{x_m} |\hat{\Phi}_j(x, v_0, z_0) - \Phi_j(x, v_0, z_0)| dx dv \\
&\le C e^{-\Lambda T} \sum_{M=\bar{M}+1}^{\infty} \frac{(\Lambda T)^M}{M!}
\end{aligned}
$$

where C is a constant defined by:

$$C = \frac{x_m}{\pi} \sum_{j \in E} \int_0^{v_m} h(v, j) dv$$

Using similar arguments, we define an approximation for the expression of S&P 500 option prices by truncating up to $\bar{M}$ the characteristic function of X_T. Now, we can use these analytical pricing formulas in the calibration procedure using options data to determine whether the proposed model has the potential to capture market dynamics. To conduct a joint calibration of both markets and following the approach of Gatheral [25], we proceed as follows:

1. Initially, the parameters $\{Q, \kappa_i, \theta_i\}$ and the instantaneous variance V_t are taken from the historical estimation detailed below;

2. Second, the parameters $\{\xi_i\}$ are calibrated to VIX options across a range of strikes for four maturities.

3. Then, the parameters $\{\rho_i\}$ are calibrated to S&P 500 options across a range of strikes for four maturities.

4. Finally, the parameters $\{\kappa_i, \theta_i\}$ and the instantaneous variance V_t are refined through a joint calibration of S&P 500 and VIX options.

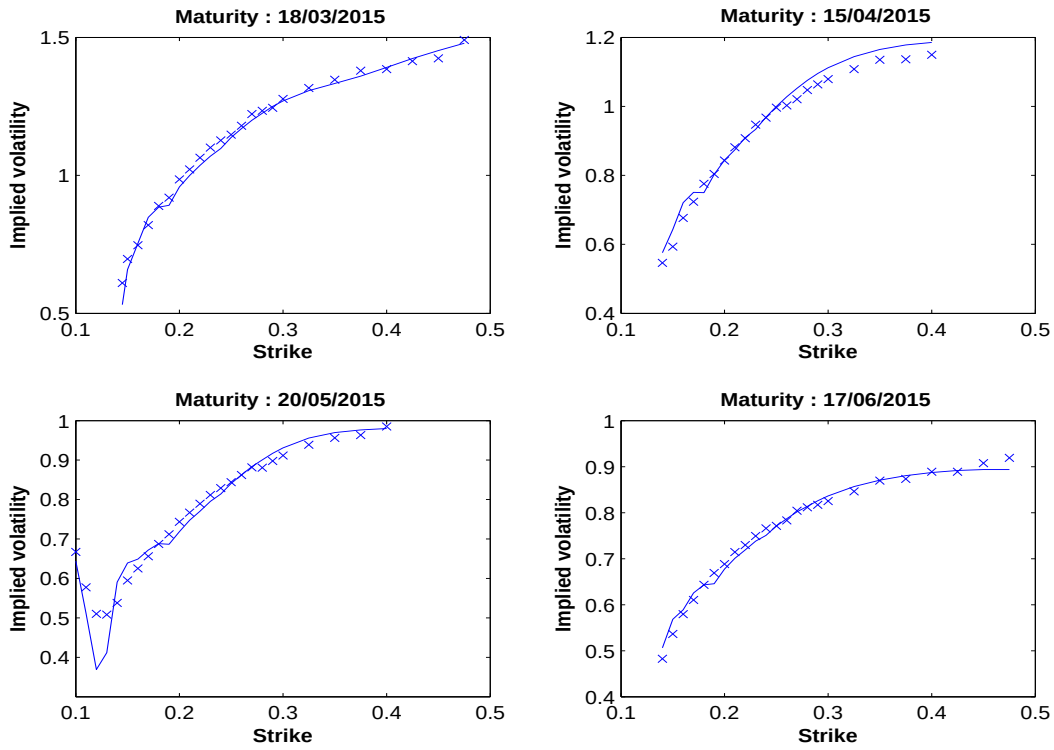

Figure 8: Implied volatilities of February 13, 2015, for VIX call options and the calibrated smile.

The figures of fitted smiles 8 and 9 show good fits for both option markets because of the flexibility offered by our regime-switching stochastic volatility model and the joint calibration procedure.

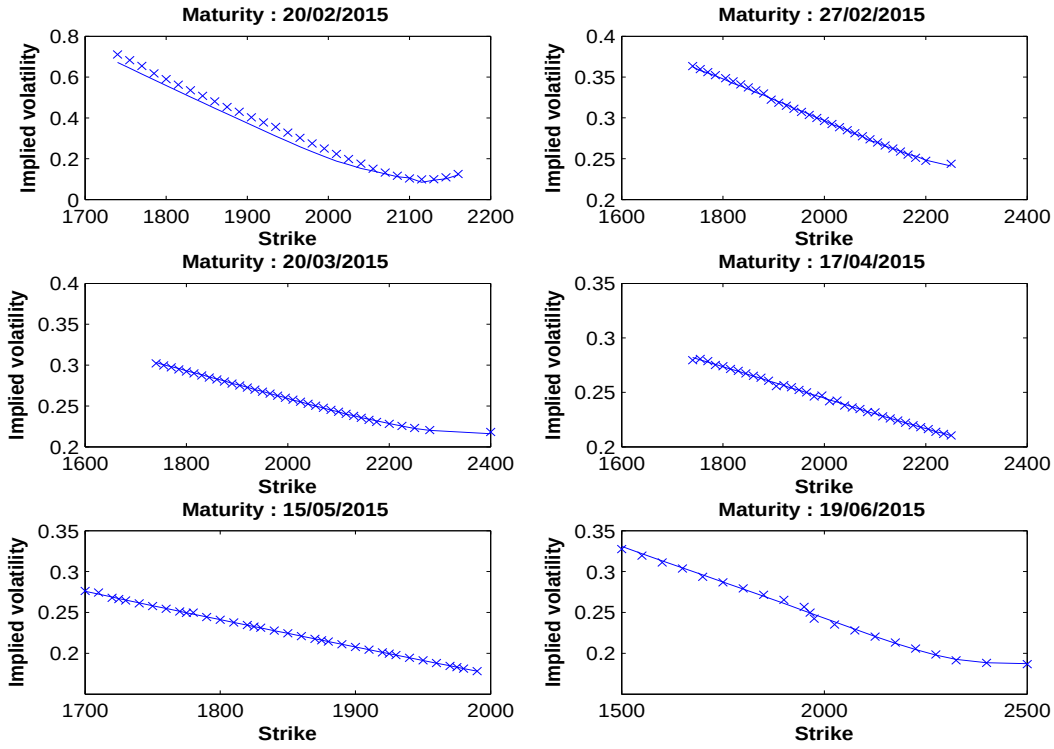

Figure 9: Implied volatilities of February 13, 2015, for S&P 500 call options and the calibrated smile.

This result is in contrast with the conclusions of Papanicolau and Sircar [36], who state that there is a systematic discrepancy between the two markets. Although they do not test a joint calibration procedure in their work, here, we can argue that our extension of the Heston model is able to consistently match both option markets with the same model, showing that rather than being inconsistent, both markets are actually providing complementary information: the VIX option market allows calibration of the volatility of volatility, while the S&P 500 implied skew captures the stock-vol correlation. This improved fit of the option markets must be directly linked to the improved matching of various stylized facts that the model is able to reproduce. As a remark, a good simultaneous fit has also been achieved during crisis time as shown in Gatheral [25], which goes in contradiction with the recent results of Bardgett et al. [6] that VIX and SPX markets have different information during times of crisis. As a conclusion, the seek of a model for pricing equity and volatility derivatives should be driven by the imperative of insuring consistency between VIX and SPX markets in any conditions.

This achieved consistency has two practical implications. First, we see an increasing demand for hybrid equity-volatility products, especially from hedge funds. These products are replicated using combinations of S&P 500 and VIX options, and thus, they require models that consistently jointly price both markets. Second, by more realistically modeling the implied volatility dynamics, this model provides more robust pricing of equity exotic products, which are highly sensitive to the dynamics of implied volatility.

6 Conclusion

In this paper, we propose an extension of the Heston model to consistently price both S&P 500 and VIX options while reproducing some important features of the joint dynamics of the stock and its volatility.

We derive analytical expressions for the prices of both options, allowing for efficient calibration and taking advantage of the piecewise affine structure of the volatility process and the good inertia of each volatility regime (or, equivalently, the small regime shift rate). We also provide an extension of the Baum-Welch algorithm to filter the hidden state of the Markov Chain driving the instantaneous variance process, although it is unobservable, and show how this problem falls into the EM setting.

There are several possible extensions that could be pursued in future research. First, we can explore the implications of the model regarding the pricing of volatility derivatives, such as variance options and VIX exotic derivatives; hybrid equity-volatility products; and forward skew and forward volatility-dependent equity exotic derivatives, such as cliquet or ratchet-related options. For all of these products, the choice of the volatility model has critical implications regarding valuation, and evidence shows that our model is reliable. However, it remains challenging to explore whether this model can be used efficiently as a hedging model because incompleteness matters. It could be also interesting to explore applications of the filtered state of the Markov Chain as a trading signal for several investment and hedging opportunities: the hedging of the downside risk of an equity portfolio or the harvesting of volatility risk premia by trading S&P 500 and VIX options .

References

[1] Y. Ait-Sahalia and J. Jacod. Estimating the Degree of Activity of Jumps in High Frequency Data. *Annals of Statistics*, 19:556–584, 2009.

[2] Y. Ait-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83:413–452, 2007.

[3] J. Baldeaux and A. Badran. Consistent Modeling of VIX and Equity Derivatives Using a 3/2 plus Jumps Model. *Applied Mathematical Finance*, 21(4):299–312, 2014.

[4] F.M. Bandi and R. Reno. Time-varying leverage effects. *Journal of Econometrics*, 169:94–113, 2012.

[5] Q. Bao, S. Li, and D. Gong. Pricing vxx option with default risk and positive volatility skew. *European Journal of Operational Research*, 223(1):246–255, 2012.

[6] C. Bardgett, M. Leippold, and E. Gourier. Inferring volatility dynamics and risk-premia from the S&P 500 and VIX markets. Working Paper, 2015.

[7] C. Bayer, P. Friz and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[8] C. Bayer, J. Gatheral, and M. Karlsmark. Fast Ninomiya-Victoir calibration of the double-mean-reverting model. *Quantitative Finance*, 13(11):1813–1829, 2013.

[9] L. Bergomi. Smile Dynamics. *Risk*, 17:117–123, 2004.

[10] L. Bergomi. Smile Dynamics III. *Risk*, 18:67–73, 2008.

[11] L. Bergomi. Smile Dynamics IV. *Risk*, pages 94–100, 2009.

[12] H. Buehler. Consistent variance curve models. *Finance and Stochastics*, 10:178–203, 2006.

[13] R. Cont and Th. Kokholm. A Consistent Pricing Model for Index Options and Volatility Derivatives. *Mathematical Finance*, 12(8):1199–1218, 2012.

[14] R. Cont and C. Mancini. Nonparametric Tests for Analyzing the Fine Structure of Price Fluctuations. Technical report, Columbia University, 2007. Financial Engineering Report 2007-13.

[15] P. Christoffersen, K. Jacobs and K. Mimouni. Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns, and Option Prices. *The Review of Financial Studies*, 23 (8), 2010.

[16] P. Date and S. Mitra. Regime switching volatility calibration by the Baum-Welch method. *Journal of Computational and Applied Mathematics*, 234(12):3243–3260, 2010.

[17] A.P. Dempster, N.M. Laird and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Discussion. *J.R. Stat.Soc*, 39:1–38, 1977.

[18] G. Drimus. Options on realized variance by transform methods : a non-affine stochastic volatility model. *Quantitative Finance*, 12(11):1679–1694, 2012.

[19] J. Duan and C. Yeh. Jump and volatility risk premiums implied by vix. *Journal of Economic Dynamics and Control*, 34(11):2232–2244, 2010.

[20] M. Dueker. Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business and Economic Statistics*, 15:26–34, 1997.

[21] B. Dupire. A Unified Theory of Volatility. In P. Carr, editor, *Derivatives Pricing : The Classic Collection*. Risk Books, 2004.

[22] G.B. Durham. SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics*, 85(3):822–856, 2007.

[23] R.J. Elliot, L. Chan, J.W. Lau, and T.K. Siu. Pricing options under a generalized markov-modulated jump-diffusion model. *Stochastic Analysis and Applications*, 25(4):821–843, 2007.

[24] R.J. Elliot, L. Chan, and T.K. Siu. Pricing volatility swaps under heston's stochastic volatility model with regime change. *Applied Mathematical Finance*, 14(1):41–62, 2007.

[25] J. Gatheral. Consistent Modeling of SPX and VIX options. Presented at the Fifth World Congress of the Bachelier Finance Society in London, 2008.

[26] G. Fort and E. Moulines. Convergence of the Monte Carlo expecta- tion maximization for curved exponential families. *Ann. Stat*, 31, 4, 1220-1259.

[27] S. Goutte. Pricing and hedging in stochastic volatility regime switching models. *Journal of Mathematical Finance*, 3:70–80, 2013.

[28] S. Goutte. and A.Ngoupeyou. Defaultable bond pricing using regime switching intensity model. *Journal of Applied Mathematics and Informatics*, 31:(3) 711–732, 2013.

[29] S. Goutte and B. Zou. Continuous time regime switching model applied to foreign exchange rate. *Mathematical Finance letters*, 8:1–37, 2013.

[30] J. Hamilton. Regime switching models. In S. Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.

[31] S.L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[32] E. Hillebrand. Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129 (1-2):121–138, 2005.

[33] C.S. Jones. The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics*, 116:181–224, 2003.

[34] J. Mencìa and E. Sentana. Valuation of VIX derivatives. *Journal of Financial Economics*, 108(2): 367–391, 2013.

[35] M. Musiela and N. Rutkowski. *Martingale Methods in Financial Modelling*, Springer, 2005.

[36] A. Papanicolau and R. Sircar. A Regime-Switching Heston Model for VIX and S&P 500 Implied Volatilities. *Quantitative Finance*, 14(10):1811–1827, 2014.

[37] E. Renault and N. Touzi. Option Hedging and Implied Volatilities in a Stochastic Volatility Model. *Mathematical Finance*, 6(3):279–302, 1996.

[38] A. Sepp. Pricing options on realized variance in the Heston's model with jumps in returns and volatility. *Journal of Computational Finance*, 11(4):33–70, 2008.

[39] V. Todorov Variance Risk-Premium Dynamics: The Role of Jumps. *The Review of Financial Studies*, 23 (1), 2010.

[40] A. B. Trolle and E. S. Schwartz. A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives. *The Review of Financial Studies*, 22 (5), 2009.

[41] H.Y. Wong and Y.W. Lo. Option pricing with mean reversion and stochastic volatility. *European Journal of Operational Research*, 197(1):179–187, 2009.

[42] L. Wu and F. Zhang. Fast swaption pricing under the market model with a square-root process. *Quantitative Finance*, 8(2):163–180, 2008.

Appendices

Proof of Lemma 3.1

The following relationship holds:

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}_{v,Z}\left[\int_0^{\tau_{vix}} V_s ds\right] &= \int_0^{\tau_{vix}} \mathbb{E}_{v,Z}[V_s] ds \\ &= \sum_{k=0}^{M-1} \mathbb{E}^{\mathbb{P}}_{v,Z}\left[\int_{\tau_k}^{\tau_{k+1}} V_s ds\right] \\ &= \sum_{k=0}^{M-1} \mathbb{E}^{\mathbb{P}}_{v,Z}\left[\mathbb{E}^{\mathbb{P}}\left[\int_{\tau_k}^{\tau_{k+1}} V_s ds \mid \mathcal{F}^Z_{\tau_k}\right]\right]\end{aligned}$$

It is well known that we have:

$$\mathbb{E}^{\mathbb{P}}[\int_{\tau_k}^{\tau_{k+1}} V_s ds \mid \mathcal{F}^Z_{\tau k}] = a(Z_{\tau_k})V_{\tau_k} + b(Z_{\tau_k})$$

with

$$\begin{aligned}a(Z_{\tau_k}) &= \frac{1-\exp(-\kappa(Z_{\tau_k})\Delta_{t_k})}{\kappa(Z_{\tau_k})\Delta_{t_k}} \\ b(Z_{\tau_k}) &= \theta(Z_{\tau_k})(1-a(Z_{\tau_k}))\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}_{v,Z}\left[\int_0^{\tau_{vix}} V_s ds\right] &= \sum_{k=0}^{M-1} \mathbb{E}^{\mathbb{P}}_{v,Z}[a(Z_{\tau_k})V_{\tau_k} + b(Z_{\tau_k})] \\ &= \sum_{k=0}^{M-1} \Big(a(Z_{\tau_k})\mathbb{E}^{\mathbb{P}}_{v,Z}[V_{\tau_k}] + b(Z_{\tau_k})\Big)\end{aligned}$$

Now, we need to calculate $\mathbb{E}^{\mathbb{P}}_{v,Z}[V_{\tau_k}]$. If we integrate the SDE of the process V_t between τ_{k-1} and τ_k, we get

$$\mathbb{E}^{\mathbb{P}}_{v,Z}[V_{\tau_k}] = \mathbb{E}^{\mathbb{P}}_{v,Z}[V_{\tau_{k-1}}] + \Delta_{t_{k-1}}\theta(Z_{\tau_{k-1}})\kappa(Z_{\tau_{k-1}}) - \kappa(Z_{\tau_{k-1}})(a(Z_{\tau_{k-1}})V_{\tau_{k-1}} + b(Z_{\tau_{k-1}}))$$

Hence, we can recursively calculate the quantity $\mathbb{E}^{\mathbb{P}}_{v,Z}[V_{\tau_k}]$ starting with $\mathbb{E}_{v,\mathbb{Z}}[V_{\tau_0}] = v_0$, and we can obtain a family of functions $(f_k(Z_0,\ldots,Z_{\tau_{k-1}}), g_k(Z_0,\ldots,Z_{\tau_{k-1}}))$, such that:

$$\mathbb{E}^{\mathbb{P}}_{v,\mathbb{Z}}[V_{\tau_k}] = f_k(Z_0,\ldots,Z_{\tau_{k-1}})v_0 + g_k(Z_0,\ldots,Z_{\tau_{k-1}})$$

Thus, we have:

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}_{v,Z}\left[\int_0^{\tau_{vix}} V_s ds\right] &= \sum_{k=0}^{M-1}\left(a(Z_{\tau_k})(f_k(Z_0,\ldots,Z_{\tau_{k-1}})v_0 + g_k(Z_0,\ldots,Z_{\tau_{k-1}})) + b(Z_{\tau_k})\right)\\ &= \left(\sum_{k=0}^{M-1} a(Z_{\tau_k}) f_k(Z_0,\ldots,Z_{\tau_{k-1}})\right) v_0 + \sum_{k=0}^{M-1} b(Z_{\tau_k}) + a(Z_{\tau_k}) g_k(Z_0,\ldots,Z_{\tau_{k-1}})\end{aligned}$$

Proof of Lemma 5.2

By integrating the SDE of the process (V_t) between 0 and T, we have:

$$\begin{aligned}ixV_T &= ixv + ix\int_0^T \kappa(Z_s)(\theta(Z_s) - V_s)ds + \int_0^T ix\xi(Z_s)\sqrt{V_s}ds\\ e^{ixV_T} &= \exp(ixv)\exp\left(ix\int_0^T \kappa(Z_s)(\theta(Z_s) - V_s)ds\right)\exp\left(\int_0^T ix\xi(Z_s)\sqrt{V_s}ds\right)\end{aligned}$$

Let us define $\mathbb{P}^*$ as a measure of probability in the space Ω by its Radon-Nikodym derivative with respect to $\mathbb{P}$:

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = \exp\left(\int_0^T ix\xi(Z_s)\sqrt{(}V_s)dW_s - \int_0^T \frac{1}{2}(ix\xi(Z_s)\sqrt{(}V_s))^2 ds\right)$$

Thus, we have:

$$\mathbb{E}^{\mathbb{P}}[e^{ixV_T} \mid V_0 = v, \mathcal{F}_T^Z] = \exp(ixv)\exp\left(ix\int_0^T \kappa(Z_s)\theta(Z_s)ds\right)\mathbb{E}^{\mathbb{P}^*}\left[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right) \mid V_0 = v, \mathcal{F}_T^Z\right]$$

where $\gamma(Z_s) := -\kappa(Z_s)ix - \frac{1}{2}x^2\xi(Z_s)^2$.
The next part of the proof will be dedicated to the calculation of $\mathbb{E}^{\mathbb{P}^*}[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right) \mid V_0 = v, \mathcal{F}_T^Z] := E^*_{v,Z}[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right)]$
Recall that we have a sequence of increasing times $0 = \tau_0 < \tau_1 < \cdots < \tau_{M-1} < \tau_M = T$ in which the Markov chain Z changes its value. Hence,

$$\begin{aligned}\mathbb{E}^*_{v,Z}\left[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right)\right] &= \mathbb{E}^*_{v,Z}\left[\exp\left(\sum_{k=0}^{M-1}\int_{\tau_k}^{\tau_{k+1}} \gamma(Z_s)V_s ds\right)\right]\\ &= \mathbb{E}^*_{v,Z}\left[\prod_{k=0}^{M-1}\exp\left(\int_{\tau_k}^{\tau_{k+1}} \gamma(Z_s)V_s ds\right)\right]\end{aligned}$$

Hypothetically, Z is independent of W; then, by conditioning with respect to $\mathcal{F}_{M-1} := \mathcal{F}^W_{\tau_{M-1}} \vee \mathcal{F}^Z_{\tau_{M-1}}$, we obtain:

$$\begin{aligned}\mathbb{E}^*_{v,Z}\left[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right)\right] &= \mathbb{E}^*_{v,Z}\left[\mathbb{E}^*\left[\prod_{k=0}^{M-1}\exp\left(\int_{\tau_k}^{\tau_{k+1}}\gamma(Z_s)V_s ds\right) \mid \mathcal{F}_{M-1}\right]\right] \\ &= \mathbb{E}^*_{v,Z}\left[\prod_{k=0}^{M-2}\exp\left(\int_{\tau_k}^{\tau_{k+1}}\gamma(Z_s)V_s ds\right)\mathbb{E}^*\left[\exp\left(\int_{\tau_{M-1}}^{\tau_M}\gamma(Z_s)V_s ds\right) \mid \mathcal{F}_{M-1}\right]\right]\end{aligned}$$

Now, we define two families of functions $\{B_j\}_{0\le j\le M-1}$ and $\{A_j\}_{0\le j\le M-1}$ that are determined recursively following the resolution of successive systems of Ricatti equations, such that for each $\Delta_t \in [0,T]$ and $z \in E$:

$$\begin{aligned}A_j(\Delta_t, z) &= a_{1,z}\frac{1-\frac{a_{2,z}}{a_{1,z}}C_{j,z}e^{-\sqrt{\kappa(z)}\Delta_t}}{1-C_{j,z}e^{-\sqrt{\kappa(z)}\Delta_t}} \\ B_j(\Delta_t, z) &= a_{1,z}\tilde{\kappa}(z)\tilde{\theta}(z)(\Delta_t + \frac{1}{\sqrt{\kappa(z)}}(1-\frac{a_{2,z}}{a_{1,z}})\ln\frac{C_{j,z}e^{-\sqrt{\kappa(z)}\Delta_t}-1}{C_{j,z}-1})\end{aligned}$$

where for each $j \in \{0,\dots,M-1\}$:

$$a_{1,z} = -\frac{\tilde{\kappa}(z)-\sqrt{\kappa}}{2\xi(z)^2} \qquad \gamma(z) = -\kappa(z)ix - \frac{1}{2}x^2\xi(z)^2$$

$$a_{2,z} = -\frac{\tilde{\kappa}(z)+\sqrt{\kappa}}{2\xi(z)^2} \qquad \tilde{\theta}(z) = \frac{\kappa(z)\theta(z)}{\tilde{\kappa}(z)}$$

$$\tilde{\kappa}(z) = \kappa(z) - ix\xi(z)^2 \qquad C_{j,z} = \begin{cases}\frac{a_{1,z}-A_{j+1}(\Delta_{t_{j+1}},Z_{\tau_{j+1}})}{a_{2,z}-A_{j+1}(\Delta_{t_{j+1}},Z_{\tau_{j+1}})} & \text{if } j\in\{0,\dots,M-2\}\\ 0 & \text{if } j = M-1\end{cases}$$

We have the following proposition:

Proposition 6.6. *For each integer $k \in \{1,\dots,M-1\}$, let us define:*

$$\phi_{\tau_{M-k},\Delta_{t_{M-k}}} := \mathbb{E}^*\left[\exp\left(\int_{\tau_{M-k}}^{\tau_{M-k+1}}\gamma(Z_s)V_s ds - A_{M-k+1}(\Delta_{t_{M-k+1}}, Z_{\tau_{M-k+1}})V_{\tau_{M-k+1}}\right) \mid \mathcal{F}_{M-k}\right]$$

We have the following equality:

$$\phi_{\tau_{M-k},\Delta_{t_{M-k}}} = \exp\left(-A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})V_{\tau_{M-k}} - B_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})\right)$$

Proof

We will use the following notation:

$$\phi_{\tau_{M-k},\Delta_{t_{M-k}}} := \mathbb{E}^*_{M-k}\left[\exp\left(\int_{\tau_{M-k}}^{\tau_{M-k+1}}\gamma(Z_s)V_s ds - A_{M-k+1}(\Delta_{t_{M-k+1}}, Z_{\tau_{M-k+1}})V_{\tau_{M-k}+\Delta_{t_{M-k}}}\right)\right]$$

Taking a small time interval $dt \ll \Delta_{t_{M-k}}$, we obtain:

$$
\begin{aligned}
&\mathbb{E}^*_{M-k}\left[\exp\left(\int_{\tau_{M-k}}^{\tau_{M-k+1}}\gamma(Z_s)V_s ds\right) - A_{M-k+1}(\Delta_{t_{M-k+1}}, Z_{\tau_{M-k+1}})V_{\tau_{M-k}+\Delta_{t_{M-k}}}\right]\\
=&\mathbb{E}^*_{M-k}\left[\mathbb{E}^*_{M-k+dt}\left[\exp\left(\int_{\tau_{M-k}}^{\tau_{M-k+1}}\gamma(Z_s)V_s ds\right) - A_{M-k+1}(\Delta_{t_{M-k+1}}, Z_{\tau_{M-k+1}})V_{\tau_{M-k}+\Delta_{t_{M-k}}}\right]\right]\\
=&\mathbb{E}^*_{M-k}\left[\mathbb{E}^{\left(\int_{\tau_{M-k}}^{\tau_{M-k+dt}}\gamma(Z_s)V_s ds\right)}\right.\\
&\left.\times\,\mathbb{E}^*_{M-k+dt}\left[\exp\left(\int_{\tau_{M-k+dt}}^{\tau_{M-k+1}}\gamma(Z_s)V_s ds\right) - A_{M-k+1}(\Delta_{t_{M-k+1}}, Z_{\tau_{M-k+1}})V_{\tau_{M-k}+\Delta_{t_{M-k}}}\right]\right]\\
=&\mathbb{E}^*_{M-k}\left[\phi_{\tau_{M-k}+dt,\Delta_{t_{M-k}}-dt}\exp\left(\int_{\tau_{M-k}}^{\tau_{M-k+dt}}\gamma(Z_s)V_s ds\right)\right]
\end{aligned}
$$

We now use the hypothesis is the form of ϕ to obtain:

$$
\begin{aligned}
= \quad \mathbb{E}^*_{M-k}\Big[&\exp\left(-A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})V_{\tau_{M-k}+dt} - B_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})\right)\\
&\times\exp\left(\textstyle\int_{\tau_{M-k}}^{\tau_{M-k+dt}}\gamma(Z_s)V_s ds\right)\Big]
\end{aligned}
$$

Because dt is small, we can use the stochastic differential equation of V to obtain:

$$
\begin{aligned}
&= \mathbb{E}^*_{M-k}\Big[\exp\left(-A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})(V_{\tau_{M-k}} + \tilde{\kappa}_{M-k}(\tilde{\theta}_{M-k} - V_{\tau_{M-k}})dt + \xi_{M-k}\sqrt{V_{\tau_{M-k}}}dW_t)\right)\\
&\times\exp\left(-B_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}}) - \gamma_{M-k}V_{\tau_{M-k}}dt\right)\Big]\\
&= \exp\left(-A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})V_{\tau_{M-k}} - A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})\tilde{\kappa}_{M-k}(\tilde{\theta}_{M-k} - V_{\tau_{M-k}})dt\right)\\
&\times\exp\left(-B_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}}) - \gamma_{M-k}V_{\tau_{M-k}}dt\right)\mathbb{E}^*_{M-k}\left[\exp\left(-A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})\xi_{M-k}\sqrt{V_{\tau_{M-k}}}dW_t\right)\right]\\
&= \exp\left(-A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})V_{\tau_{M-k}} - A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})\tilde{\kappa}_{M-k}(\tilde{\theta}_{M-k} - V_{\tau_{M-k}})dt\right)\\
&\times\exp\left(-B_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}}) - \gamma_{M-k}V_{\tau_{M-k}}dt\right)\exp\left(\frac{1}{2}A^2_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})(\xi_{M-k})^2V_{\tau_{M-k}}dt\right)
\end{aligned}
$$

where $\tilde{\kappa}_{M-k} = \tilde{\kappa}(Z_{\tau_{M-k}}), \tilde{\theta}_{M-k} = \tilde{\theta}(Z_{\tau_{M-k}}), \xi_{M-k} = \xi(Z_{\tau_{M-k}})$, and $\gamma_{M-k} = \gamma(Z_{\tau_{M-k}})$. By identifying the assumed expression of ϕ, we obtain:

$$
\begin{aligned}
A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}}) &= A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}}) - A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})\tilde{\kappa}_{M-k}dt\\
&\quad -\frac{1}{2}A^2_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})(\xi_{M-k})^2dt - \gamma_{M-k}dt\\
B_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}}) &= B_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}}) + A_{M-k}(\Delta_{t_{M-k}}-dt, Z_{\tau_{M-k}})\tilde{\kappa}_{M-k}\tilde{\theta}_{M-k}dt
\end{aligned}
$$

Taking dt close to zero, we obtain:

$$\begin{cases} \frac{\partial A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})}{\partial \Delta_{t_{M-k}}} & = -A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})\kappa^{M-k} - \frac{1}{2}A^2_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})(\xi_{M-k})^2 - \gamma_{M-k} \\ \frac{\partial B_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})}{\partial \Delta_{t_{M-k}}} & = A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})\tilde{\kappa}_{M-k}\tilde{\theta}_{M-k} \end{cases}$$

with the initial conditions set to:

$$\begin{cases} A_{M-k}(0, Z_{\tau_{M-k}}) & = \begin{cases} A_{M-k+1}(\Delta_{t_{M-k+1}}, Z_{\tau_{M-k+1}}) & \text{if } k \in \{0, \dots, M-2\} \\ 0 & \text{if } k = M-1 \end{cases} \\ B_{M-k}(0, Z_{\tau_{M-k}}) & = 0 \end{cases}$$

Thus, it is possible to obtain an explicit solution for the family of functions $\{B_k\}_{0\leq k\leq M-1}$ and $\{A_k\}_{0\leq k\leq M-1}$ by solving successive systems of Ricatti equations. □

We can now continue the proof of Lemma 5.2.
In what remains of the proof, we will use the simplified notation $A_{M-k}(\Delta_{t_{M-k}}) := A_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})$ and $B_{M-k}(\Delta_{t_{M-k}}) := B_{M-k}(\Delta_{t_{M-k}}, Z_{\tau_{M-k}})$. Hence:

$$\mathbb{E}^*\left[\exp\left(\int_{\tau_{M-1}}^{\tau_M} \gamma(Z_s)V_s ds\right) \mid \mathcal{F}_{M-1}\right] = \exp\left(-A_{M-1}(\Delta_{t_{M-1}})V_{\tau_{M-1}} - B_{M-1}(\Delta_{t_{M-1}})\right)$$

By reconditioning with respect to $\mathcal{F}_{M-2}$, we obtain:

$$\begin{aligned} &\mathbb{E}^*_{v,Z}\left[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right)\right] \\ =&e^{-B_{M-1}(\Delta_{t_{M-1}})}\mathbb{E}^*_{v,Z}\left[\mathbb{E}^*\left[\prod_{k=0}^{M-2}\exp\int_{\tau_k}^{\tau_{k+1}}\gamma(Z_s)V_s ds - A_{M-1}(\Delta_{t_{M-1}})V_{\tau_{M-1}} \mid \mathcal{F}_{M-2}\right]\right] \\ =&e^{-B_{M-1}(\Delta_{t_{M-1}})}\mathbb{E}^*_{v,Z}\left[\prod_{k=0}^{M-3}\exp\left(\int_{\tau_k}^{\tau_{k+1}}\gamma(Z_s)V_s ds\right)\times\right. \\ &\left.\mathbb{E}^*\left[\exp\left(\int_{\tau_{M-2}}^{\tau_{M-1}}\gamma(Z_s)V_s ds - A_{M-1}(\Delta_{t_{M-1}})V_{\tau_{M-1}}\right) \mid \mathcal{F}_{M-2}\right]\right] \\ =&e^{-B_{M-1}(\Delta_{t_{M-1}})}\mathbb{E}^*_{v,Z}\left[\prod_{k=0}^{M-3}\exp\left(\int_{\tau_k}^{\tau_{k+1}}\gamma(Z_s)V_s ds\right)\exp\left(-A_{M-2}(\Delta_{t_{M-2}})V_{\tau_{M-2}} - B_{M-2}(\Delta_{t_{M-2}})\right)\right] \\ =&\exp\left(-B_{M-1}(\Delta_{t_{M-1}}) - B_{M-2}(\Delta_{t_{M-2}})\right)\mathbb{E}^*_{v,Z}\left[\prod_{k=0}^{M-3}\exp\left(\int_{\tau_k}^{\tau_{k+1}}\gamma(Z_s)V_s ds - A_{M-2}(\Delta_{t_{M-2}})V_{\tau_{M-2}}\right)\right] \end{aligned}$$

By iterating the conditioning with respect to $\mathcal{F}_{M-k}$, where k ranges from 3 to M, we finally obtain:

$$\mathbb{E}^*_{v,Z}\left[\exp\left(\int_0^T \gamma(Z_s)V_s ds\right)\right] = \exp\left(-A_0(\Delta_{t_0})v - \sum_{k=1}^{M} B_{M-k}(\Delta_{t_{M-k}})\right)$$

□