

The Term Structure of Variance Risk Premia*

Dante Amengual

*Department of Economics, Princeton University,
Fisher Hall, Princeton, NJ 08544, USA
<amengual@princeton.edu>*

November 2008

Abstract

I study volatility dynamics and volatility risk premia under alternative stochastic volatility models. Using daily data on S&P 500 returns and variance swap rates I find that two-factor models, in which the additional factor represents the level to which the spot variance reverts, significantly improve the fit of the term structure of variance derivatives. Market prices of variance shocks are negative and economically large under all the specifications, leading to significant gains from taking short positions in variance swaps. One-factor models may lead to risk premia term structures that understate the impact of the current state of the economy at longer horizons. Moreover, term structures from two-factor models generate different patterns during quiet and turbulent market periods.

Keywords: Stochastic Volatility, Variance Risk, Bayesian Inference.

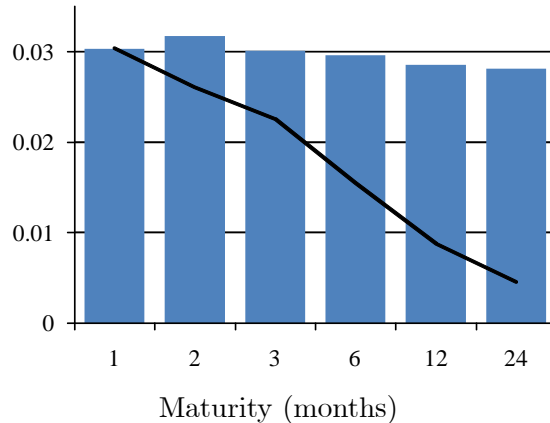
JEL: G12, C32, C52

*Preliminary. I am thankful to Yacine Aït-Sahalia for stimulating conversations, and also for providing the data. I would like to thank Harrison Hong, Jakub Jurek, Enrique Sentana, Mark Watson and Motohiro Yogo for useful discussions. I am also grateful to Gara Afonso, Pooyan Amir, Gorazd Brumen, Markus Brunnermeier, Germán Cubas, Konstantin Milbradt, Hyun Shin and Wei Xiong for helpful comments. Finally, I am deeply indebted to Chris Sims for his help, support and encouragement. All errors are my own.

1 Introduction

Empirical evidence on variance derivatives typically yields a paradoxical relationship between the mean reversion speed of volatility for short maturity derivatives and the term structure of derivative prices. For instance, the columns in Figure 1 show a similar dispersion of variance swap rates irrespective of the maturity of the contract.¹ This would imply low levels of mean reversion which are inconsistent with short maturity price dynamics. Put differently, one-factor stochastic volatility models predict a much lower price variation at longer maturities for reasonable values of mean reversion speed (solid line in Figure 1). Moreover, observed changes in the level and slope of the term structure of variance derivatives present some puzzling features *if* we think of variance as a one-factor mean reverting process. Even though these facts may suggest the existence of predictable investor’s misreaction to information –e.g. Stein (1989) or Poteshman (2001)–, these claims are always conditional on model specification.

Figure 1. Standard deviation of variance swap rates



Note: Columns denote standard deviation of variance swap rates over the period January 2, 1996 – August 29, 2008. The solid line represents implied standard deviations derived from a one-factor linear mean reverting stochastic volatility model with speed of mean reversion taken from Pan (2002) and variance calibrated from 1 month-variance swap rate.

In this paper, I propose an alternative explanation to agents’ misreaction based on popular pricing models being misspecified. I study variance dynamics, variance risk premia and their

¹A variance swap is a forward contract on the future quadratic variation that pays, at maturity, the difference between the quadratic variation over the horizon of the contract and a fixed rate called the variance swap rate determined at the inception of the contract. Variance swaps are popular volatility derivatives —frequently used by market practitioners as a hedge for volatility risk— that have been actively traded in over-the-counter markets since the collapse of LTCM in late 1998. See <http://math.uchicago.edu/~sbossu/VarSwaps.pdf> for additional information on variance swaps.

term structures under alternative stochastic volatility models using S&P 500 returns and variance swap rates on different maturities over the period January 4, 1996 to August 29, 2008. I find that two-factor extensions of the double correlated jump-diffusion model of Duffie et al. (2000), in which the additional factor represents the *level* to which the spot variance reverts, significantly improve the fit of the term structure of variance derivatives. In particular, the model in which the second factor evolves as a square-root process yields the best in-sample and predictive performances.

One drawback of single-factor models is that all points in the implied variance term structure become perfectly correlated. Another limitation of these models is that they only allow –at each point in time– a monotonically upward or downward sloping term structure of implicit variances, depending on whether the current spot variance is below or above its long-run level. These tight cross-sectional restrictions on derivative prices, which are frequently refuted by the data, yield misreaction to the arrival of new information as a plausible argument to justify the empirical findings. The results in this paper suggest that part of this misreaction in the early literature could be explained by introducing a second factor. Moreover, the near martingale behavior of this second factor explains changes in the level and slope of the variance curves that otherwise may seem puzzling.

I find that market prices of variance risks are always negative and economically large under alternative specifications implying a more persistent dynamics of variance factors under the pricing measure, and leading to significant gains from taking short positions in variance swaps. The magnitude of the (annualized) unconditional risk premium is larger at longer horizons suggesting that the long-run factor is the main component in explaining negative premia. By decomposing the risk premium into its components, I find that one-factor models lead to risk premia term structures that understate the impact of the current state of the economy at longer horizons. Furthermore, for the best-fit model empirical results suggest that the spot variance contribution to the risk premia is decreasing for maturities longer than six months while the second factor has a significant contribution at horizons as long as ten years or more.

Motivated by its ability to characterize the main cross-sectional properties of option prices, I consider the double correlated jump-diffusion model of Duffie et al. (2000) as the benchmark case. I then analyze two simple two-factor extensions of this model where the level to which the spot variance reverts is stochastic. Following Bates (2000) –who uses options data– and Chernov et al. (2003) –based on returns data only–, I consider a second model in which the second factor is a diffusion. However, instead of assuming the stock diffusion being composed

by two independent stochastic processes, I follow Egloff et al. (2007) in the parameterization of the square-root process as a parameter drift of the spot variance. Nevertheless, a potential limitation of such continuous space models is that they do not allow for sudden and correlated changes in prices. While adding jumps would help in this regard, jump-diffusions may not generate persistent periods of turbulent and quiet markets. For that reason, I also examine a third model in which the long-run variance is modulated by a hidden Markov chain. The latter constitutes a flexible class of models to capture changes in the underlying parameters that drive the spot volatility. While both two-factor extensions are similar in relaxing the cross sectional properties of derivatives with respect to one-factor models, they generate different dynamics for the spot volatility.

I rely on observations on the joint time-series of the underlying asset price and several variance swap rates with different maturities to estimate the models. A primary benefit of this approach is that risk premia related to the different volatility components can be estimated. For instance, risk premia parameters could not be identified if the data consisted only of stock returns –as in Chernov et al. (2003) and Eraker et al. (2003)–, and hence the dynamics of the volatility under the risk-neutral measure (parameters of interest for pricing) could not be recovered. Secondly, the one-to-one correspondence of derivative prices to the conditional returns distribution allows to better identify the spot variance and its dynamics under both the pricing and the physical measure. Finally, an advantage of variance swaps over other derivative prices is that they induce prices that are linear in the latent variables –as long as the drift functions of the underlying stochastic processes that govern the volatility dynamics are affine. This fact facilitates the interpretation of different parameterizations, and more importantly, it improves the computational efficiency of the estimation procedures.

I consider a likelihood-based Bayesian approach for estimating multivariate Markov processes using Markov Chain Monte Carlo (MCMC) methods. There are at least three reasons that make the use of Bayesian methods convenient when estimating stochastic volatility models. First, their treatment of latent variables does not only provide estimates of the stochastic volatility components and jump times and sizes, but it also gives a quantification of the uncertainty of those estimates by accounting for the parameters’ estimation risk. Secondly, Bayesian methods are well-suited to conduct formal model comparison. Third, priors can be utilized to disentangle jumps from diffusions in an intuitive manner when dealing with discretely observed samples of continuous time processes. This can be done by using informative priors on jump parameters i.e. beliefs that reflect the nature of jumps as large and infrequent moves in returns and

volatility.

From a methodological perspective this paper fills the blank in Table 1 as it presents an integrated analysis of stochastic volatility models based on both stock returns and variance swap rates using Bayesian methods.²

Table 1. Literature overview

	Stock returns	Returns and/or option prices	Returns and/or variance swaps
Frequentist	Chernov et al. (2003) Bates (2008)	Bates (2000), Coval and Shumway (2001), Pan (2002) Bondarenko (2003) Bakshi and Kapadia (2003) Aït-Sahalia and Kimmel (2007)	Egloff et al. (2007) Todorov (2007) Carr and Wu (2008)
Bayesian	Eraker et al. (2003)	Eraker (2004)	*

The outline of the paper is as follows. In the next section, I specify the benchmark model, its extensions and the pricing kernel, and I then discuss the different term structures of variance swap rates under alternative models. I introduce the econometric methodology in Section 3. The description of the data and a summary of the empirical findings are presented in Section 4. Then, Section 5 discusses implications of the empirical results for the term structure of variance risk premia. Finally, Section 6 concludes. Some derivations and additional results are gathered in the Appendices.

2 The models

2.1 Data-generating processes

As is well known, Heston (1993)’s model requires implausibly large shocks to generate the returns behavior observed in the data. To overcome this issue, recent works in empirical option pricing mostly focus on alternative extensions of the square-root stochastic volatility model that retain its one-factor structure. Jones (2003) shows that moving away from the affine diffusion specification can lead to more realistic model predictions (i.e. the constant

²To the best of my knowledge this paper is the only one to fit in the lower right corner of Table 1. In the other cells I report representative papers that may not necessarily exhaust the corresponding list of papers.

elasticity of volatility model; hereinafter CEV). Pan (2002) studies the volatility and jump risk premia when stock returns can jump; she also finds that adding jumps leads to a more accurate representation of stock returns while keeping the affine structure of the model, which is convenient for pricing purposes. Advantages of allowing jumps to affect both stock returns and volatility are documented in Eraker et al. (2003) and Eraker (2004).

I consider three extensions of Heston’s (1993) stochastic volatility model. In order to obtain analytical tractability for option pricing I take the correlated-jump affine diffusion model introduced by Duffie et al. (2000) as the benchmark case. For the extended models, the dynamics of the stochastic long-run variance are also restricted accordingly in order to get option prices in semi-closed form. In what follows, fix a probability space (Ω, \mathcal{F}, P) and an information filtration $(\mathcal{F}_t)_{t \geq 0}$ satisfying the usual conditions (see e.g. Protter (1990)).

2.1.1 Benchmark: the affine double-jump diffusion model

Under the objective probability measure P , we assume the following data generating process for the log-stock price y and its spot variance V

$$\begin{aligned} dy_t &= \mu^P(V_t)dt + \sqrt{V_t} \left[\sqrt{1 - \rho^2} dW_{0t}^P + \rho dW_{1t}^P \right] + Z_t^y dJ_t \\ dV_t &= \kappa_V^P(\bar{\alpha}^P - V_t)dt + \gamma_V \sqrt{V_t} dW_{1t}^P + Z_t^V dJ_t, \end{aligned} \quad (1)$$

where κ_V^P is the mean reversion parameter, $\bar{\alpha}^P$ is the mean level of volatility, γ_V is commonly referred as “volatility-of-volatility”, $(W_0, W_1)'$ is an adapted standard Brownian motion, and $\mu^P(V_t)$ is a parametric function that depends on the specification of the market prices of risk. The Brownian shocks associated to the volatility are allowed to be correlated with those driving the stock price dynamics through the parameter ρ , hence capturing the so-called “leverage effect” of Black (1976): A stylized fact that stock returns are typically negatively correlated with changes in volatility, generating negative skewness in the transition density of the stock price.

Jumps in volatility and prices are driven by the same Poisson process J , allowing for correlated jump sizes. In particular,

$$Z_t^V \sim \exp(\beta_1 V_t), \text{ and } Z_t^y | Z_t^V \sim N(\beta_2 + \beta_3 Z_t^V, \beta_4)$$

where $\beta_i, i = 1, \dots, 4$ are constant parameters, and J has a constant intensity parameter λ . Eraker et al. (2003) find that jumps in volatility are an important feature in stock returns. Jumps allow for fast increments in volatility which helps to close the gap of the square-root

diffusion specification when compared to more general models e.g. the constant elasticity of variance specification of Jones (2003). In essence, the specification considered here implies that jumps will simultaneously impact both prices and volatility and that jumps occur at a constant rate.³

2.1.2 Extensions: stochastic volatility level

While multi-factor models have been extensively used in the bond pricing literature, their use in stochastic volatility models is more recent and scarce. Bates (2000) emphasizes the potential importance of more than one volatility factors for explaining the term structure of return volatilities; he includes two independent volatility factors in his model. Using a long time-series of the Dow Jones index, Chernov et al. (2003) find strong evidence for a second factor within the affine family of models they study in that paper. Eraker (2004) argues that the option pricing errors from his empirical analysis based on one-factor jump-diffusion models are consistent with two-factor volatility models where the second factor determines the long-run volatility. More recently, Egloff et al. (2007) find evidence for a second factor using data on variance swap rates only.

I consider two simple two-factor stochastic volatility models where the second factor has a clear economic interpretation: the level to which the spot volatility reverts. As is described below, both specifications relax the cross sectional dynamics of derivative prices in a similar way –allowing for a better fit of the variance term structure–, but they imply different dynamics for the spot volatility.

Square-root process Bates (2000) and Chernov et al. (2003) consider models where the stock diffusion term takes the form $\sqrt{V_{1t}}dW_{1t} + \sqrt{V_{2t}}dW_{2t}$. Given sufficiently generous correlation structure, factor rotations can make such specification observationally equivalent to the stochastic volatility model considered here where α evolves according to

$$d\alpha_t = \kappa_\alpha^P(\bar{\alpha}^P - \alpha_t)dt + \gamma_\alpha\sqrt{\alpha_t}dW_{2t}^P \quad (2)$$

where κ_α^P , $\bar{\alpha}^P$ and γ_α are the mean reversion parameter, the long-run level of the volatility level, and the volatility-of-volatility (vol-vol) parameter that characterize the dynamics of α under the physical measure, respectively. W_2 is a standard Brownian motion independent of

³Alternative specifications have been considered in Bates (2000), Eraker (2004), Eraker et al. (2003), and Pan (2002) among others.

$(W_0, W_1)'$: I assume that shocks associated to α are not correlated to those that drive the dynamics of y and V ; the reason is that the additional factor W_2 is aimed to characterize the low-frequency volatility dynamics while (W_0, W_1) are mainly related to the high-frequency.

I do not explicitly consider jumps in the volatility level α to base the subsequent analysis on three essentially different dynamic patterns for α , say, (i) a constant, (ii) a stochastic process with continuous paths, and (iii) a stochastic process which only take values on a discrete space, which is next.

Regime-switching process A potential limitation of a diffusion model for α is that it does not allow for sudden and correlated changes in prices. While adding jumps would help in this regard, jump-diffusions may not generate persistent periods of “turbulent” and “quiet” markets. This can be accomplished with regime-switching models. Since the seminal work of Hamilton (1989), many authors have built econometric models whose parameters are driven by a Markov state variable. Again, its use is more popular in bond pricing; for example Hamilton (1988), Lewis (1991) and Ang and Bekaert (2002) have examined discrete-time models of regime switches in interest rates, while Dai and Singleton (2000), Landen (2000) among others do the same in a continuous-time framework.

Most of the work on regime-switching stochastic volatility models (e.g. Yao et al (2003)), however, assumes that the spot volatility itself evolves according to a hidden Markov chain, which often requires a large number of states in order to get a reasonable representation of the data. In this paper, I assume that the volatility level α is modulated by a Markov chain $s_t : \Omega \mapsto \{1, \dots, d\}$ with a $d \times d$ generator matrix $\mathbf{P} = (p_{ij})$ in which all rows sum to zero and whose elements are assumed to be constant.⁴ In this sense, regime-switching models constitute an attractive and flexible class of models to capture changes on the underlying parameters that drive the spot volatility: Compared to the benchmark case, this model includes $d^2 - 1$ additional parameters to describe the dynamics of the underlying level of volatility.

2.2 Pricing variance swaps

2.2.1 Pricing measure and Q dynamics

In contrast to the complete market setting of Black and Scholes (1973), the random jump sizes make the market incomplete with respect to the riskfree bank account, the underlying

⁴The assumption of constant intensity could be easily relaxed as soon as the intensities remain constant under the martingale measure in order to analytically price derivatives.

stock, and a finite number of variance swap rates. Consequently, the state-price density (or pricing kernel) is not unique. Hence, for pricing purposes I impose standard risk premium assumptions in the literature and define Q as the equivalent martingale measure associated to that pricing kernel. In particular, I assume that jump risk is not priced due to: (i) the short period of time considered for which variance swap data on different time-to-maturities are available makes difficult to obtain precise estimates of those parameters, and (ii) the purpose of including jumps in this work is to improve the models' fit under the objective measure –e.g. to obtain better estimates of the diffusion parameters. Accordingly, Z^y , Z^V and J have the same distribution under the Q and P measures, and hence the candidate pricing kernel prices the following sources of risk: (i) diffusive price shocks, (ii) spot variance shocks, and for the second and third models, (iii) variance level shocks.

Specifically, the pricing kernel M will satisfy

$$\frac{dM_t}{M_t} = -r dt - [\eta_S \sqrt{(1-\rho^2)V_t}, \eta_V \sqrt{V_t}] \cdot \begin{bmatrix} dW_{0t}^P \\ dW_{1t}^P \end{bmatrix} - \vartheta_t^j \quad (3)$$

where ϑ_t^j is model specific. $\vartheta_t^I = 0$ in the benchmark case (model *I*); when α follows a “square-root” process, then $\vartheta_t^{II} = \eta_\alpha \sqrt{\alpha_t} dW_{2t}^P$ (model *II*). As for the regime-switching model, denoting by $\varpi^j(s_t)$ the market price of a shift from the current regime s_t to regime j , we have $\vartheta_t^{III} = \sum_{j=1}^d \varpi^j(s_t) [\mathbf{1}\{s_t = j\} - P_j dt]$ (model *III*) where $\mathbf{1}\{A\}$ is the indicator function of event A , and $P_j = \sum_{i=1}^d \mathbf{1}\{s_t = i\} p_{ij}$, with η 's and $\varpi^j(s_t)$'s assumed to be constant.

If we state the dynamics of the log-stock price y under the Q measure as

$$dy_t = \left(r - \delta - \frac{1}{2} V_t \right) dt + \sqrt{V_t} \left[\sqrt{1-\rho^2} dW_{0t}^Q + \rho dW_{1t}^Q \right] + Z_t^y dJ_t$$

we obtain that its drift in (1) is of the form

$$\mu^P(V) = r - \delta + \left(\eta_S (1 - \rho^2) + \eta_V \rho - \frac{1}{2} \right) V$$

where r is the risk-free rate and δ is the dividend yield (both taken to be constant for simplicity only). As for the spot volatility, we have

$$\begin{aligned} dV_t &= [\kappa_V^P (\alpha^P - V_t) - \eta_V \gamma_V V_t] dt + \gamma_V \sqrt{V_t} dW_{1t}^Q + Z_t^V dJ_t \\ &= \kappa_V^Q (\alpha^Q - V_t) dt + \gamma_V \sqrt{V_t} dW_{1t}^Q + Z_t^V dJ_t \end{aligned}$$

from where we can redefine the P -parameters in terms of Q -parameters and market prices of risk: $\kappa_V^P = \kappa_V^Q - \eta_V \gamma_V$ and $\alpha^P = \alpha^Q \kappa_V^Q / (\kappa_V^Q - \eta_V \gamma_V)$. Therefore, under the physical measure

the spot variance is reverting at the rate $\kappa_V^Q - \gamma_V \eta_V$, where η_V is the risk premium process associated with shocks to V . A negative market price of variance risk γ_V is related to a slower mean reversion under the pricing measure than under the physical measure.

Similarly, when the underlying level of variance evolves as a square root process, ϑ_t^{II} and (2) imply that under the Q measure α satisfies

$$\begin{aligned} d\alpha_t &= [\kappa_\alpha^P(\bar{\alpha}^P - \alpha_t) - \eta_\alpha \gamma_\alpha \alpha_t] dt + \gamma_\alpha \sqrt{\alpha_t} dW_{2t}^Q \\ &= \kappa_\alpha^Q(\bar{\alpha}^Q - \alpha_t) dt + \gamma_\alpha \sqrt{\alpha_t} dW_{2t}^Q, \end{aligned}$$

so that $\kappa_\alpha^P = \kappa_\alpha^Q - \eta_\alpha \gamma_\alpha$ and $\bar{\alpha}^P = \bar{\alpha}^Q \kappa_\alpha^Q / (\kappa_\alpha^Q - \eta_\alpha \gamma_\alpha)$; where again, the difference in the speed of mean reversion under both measures reflects the market price of shocks to the level of variance.

As for the hidden Markov chain model for α , we can characterize its dynamics under the risk-neutral measure by a $d \times d$ generator matrix \mathbf{Q} with elements defined by

$$q_{ij} = [1 - \varpi_j(s_t = i)] p_{ij} \text{ for } i \neq j \text{ and } q_{ii} = - \sum_{j \neq i} q_{ij}.$$

I also assume that (i) the pricing kernel can be written as $M_t := \sum_{j=1}^d \mathbf{1}\{s_t = j\} M(s_t, V_t)$, and (ii) V follows a continuous path even across regime shifts. These additional conditions require $\varpi_i(s_t = i) = 0$ for $i = 1, \dots, d$ and $(1 - \varpi_j(s_t = i))(1 - \varpi_i(s_t = j)) = 1$ for $i, j = 1, \dots, d$. Therefore, for the two-regime model considered in the empirical section we would only have one price of risk associated to α (see e.g. Dai and Singleton (2003) for details).

2.2.2 Prices

Since the variance swap contract involves no initial payment, the price of the contract is equal to the expected value under the risk-neutral measure of the future quadratic variation over the contract's horizon. For a contract with a length of τ years, the standardized variance swap rate is

$$Y_t^\tau = \frac{1}{\tau} E_t^Q \left[\int_t^{t+\tau} V_s ds \right]. \quad (4)$$

Given the affine structure of the first two models, variance swap rates can be easily derived by using the transform analysis for affine jump-diffusion processes of Duffie et al. (2000) (see Egloff et al. (2007) for a detailed derivation). In fact, it turns out that the same formulas

apply to the case of non-affine instantaneous variance models such as the CEV model. The only requirement is linearity of the drift functions (see Aït-Sahalia and Amengual (2008)).

The variance swap rate in the one factor model is a linear combination of the spot variance V and its long-run level; specifically,

$$Y_t^\tau = \phi^I(\tau)V_t + [1 - \phi^I(\tau)]\bar{\alpha}^Q \quad \text{with} \quad \phi^I(\tau) = \frac{1 - e^{-\kappa_V^Q \tau}}{\tau \kappa_V^Q}.$$

It converges to the current level of variance as time-to-maturity goes to zero and to the long-run level when τ goes to infinity. For a given τ , the weight of V is decreasing with the speed of mean-reversion κ^Q . Interestingly, we can note that one-factor models impose strong restrictions on the shape the term structure of volatility may take: it is decreasing if the spot volatility is above its long-run level and vice versa.

Not surprisingly, a second factor introduces more flexibility in the cross-section of variance swaps. A similar linear relationship holds when the underlying level of variance evolves as a linearly mean reverting process, Y_t^τ is a linear combination of the current spot variance, the level α to which V is reverting, and the long-run level $\bar{\alpha}$, that is,

$$Y_t^\tau = \phi^I(\tau)V_t + \phi^{II}(\tau)\alpha_t + [1 - \phi^I(\tau) - \phi^{II}(\tau)]\bar{\alpha}^Q \quad (5)$$

with

$$\phi^{II}(\tau) = \frac{1}{\tau \kappa_\alpha^Q} \left[1 + \frac{\kappa_\alpha^Q}{\kappa_V^Q - \kappa_\alpha^Q} e^{-\kappa_V^Q \tau} - \frac{\kappa_V^Q}{\kappa_V^Q - \kappa_\alpha^Q} e^{-\kappa_\alpha^Q \tau} \right].$$

It shares the same properties of the one-factor case when time-to-maturity goes either to zero or infinity. However, the extra term $\phi^{II}(\tau)[\alpha_t - \bar{\alpha}^Q]$ which is due to the second Brownian factor, leads to more flexible term structures of volatilities that become more relevant at intermediate maturities: For instance, it allows for hump shaped term structure of variance. Notice also that under this model, variance swap rates with different maturities can present different degrees of persistence: Assuming α captures the low-frequency dynamics of the volatility, i.e. when $\kappa_\alpha^Q < \kappa_V^Q$, we would observe more persistence in variance swap rates at longer maturities.

Finally, consider the case in which α_t is a Markov chain with state space $\{1, \dots, d\}$, and intensity matrix $\mathbf{Q} = (q_{ij})$. Given the simple structure of the model (in which only the constant term of the drift function is allowed to change across regimes), it is shown in Appendix A that the price P of a derivative security can be decomposed as

$$P(t, \alpha = \alpha_i, S, V) = \bar{P}(t, \bar{\alpha}, S, V) + P_i(t, \alpha, S, V)$$

where $\bar{P}(t, \bar{\alpha}, S, V)$ refers to the corresponding price in the benchmark model where $\bar{\alpha}^Q$ denotes the expected level of the invariant distribution of α under Q plus a constant $P_i(t, \alpha, S, V)$ that reflect the state of the chain. The set of d constants should satisfy

$$\mathbf{P}(\tau; \alpha, S, V) = -\kappa \int_0^\tau \exp[\mathbf{Q}(\tau - u)](\bar{\alpha}^Q \ell - \boldsymbol{\alpha}) \frac{\partial \bar{P}(u, S, V)}{\partial V} du \quad (6)$$

where $\mathbf{P}(\tau; \alpha, S, V) = (P_1(\tau, \alpha, S, V), \dots, P_d(\tau, \alpha, S, V))'$, ℓ is a vector of ones, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)'$ and τ denotes time-to-maturity. Note that this integral equation can be easily evaluated given that $\partial \bar{P}(u, S, V)/\partial V$ is available in semi-closed form.

In particular, when pricing variance swaps $\partial \bar{P}(u, S, V)/\partial V$ takes the simple form $\phi^I(\tau)$ and prices –conditional on the regime– become linear in V and $\bar{\alpha}^Q$. Specifically, in the case $d = 2$ with $s_t \in \{L, H\}$ we have

$$\begin{aligned} Y_t^\tau &= \frac{q_{HL}^Q \alpha_L + q_{LH}^Q \alpha_H}{q} [1 - \phi^I(\tau)] + \phi^I(\tau) V_t \\ &+ \left[\mathbf{1}\{\alpha = \alpha_L\} q_{LH}^Q \frac{\alpha_L - \alpha_H}{q} + \mathbf{1}\{\alpha = \alpha_H\} q_{HL}^Q \frac{\alpha_H - \alpha_L}{q} \right] \phi^{III}(\tau) \end{aligned} \quad (7)$$

where q_{ij}^Q is the rate of moving from regime i to j under the Q measure, $q = q_{HL} + q_{LH}$ and

$$\phi^{III}(\tau) = \frac{1 - e^{-q\tau}}{q\tau} - \frac{1 - e^{-(\kappa^Q + q)\tau}}{(\kappa^Q + q)\tau}.$$

The variance swap curves that this model can generate are quite similar to the the ones in the square-root level of variance. However its dynamics allow for correlated discontinuous changes in prices of magnitude $(\alpha_H - \alpha_L)\phi^{III}(\tau)$, while in the second model changes derived from changes in α are arbitrarily small as the sampling interval approaches zero. Hence, this type of models may accommodate sudden changes in market participants' beliefs about the mean level of the market volatility.

3 Likelihood inference

In order to get the notation as simple as possible, assume that variance swap rates are determined by a subset of state variables included in the set of all latent variables X (typically, X will contain V , α in the two-factor extensions, and the remaining latent variables as jump sizes and jump times) and a parameter vector θ_M that includes the parameters of the theoretical model –excluding the jump parameters– through a known function $Y_t^\tau = Y(X_t, \theta_M)$; whose specific functional form was described in Section 2 (recall that in spite of not being explicit in

this notation, theoretical variance swap rates do not depend on the jump times or jump sizes). Let θ_J denote the vector of the jump process parameters.

Assume there are observations available on n different time-to-maturity variance swap rates, over the period $[0, T]$. Let also assume that observations are recorded daily, so that the total number of time periods under consideration is $N = T/\Delta$ with T being the length of the sample in years and $\Delta = 1/252$. Given that there are more variance swap rates than sources of uncertainty, to avoid stochastic singularity we typically assume there are pricing errors. Specifically, I assume a multiplicative possibly autocorrelated pricing error $\varepsilon_{i\Delta}^{\tau_j}$ so that the observed variance swap rates satisfy

$$Y_{i\Delta}^{\tau_j} = Y(X_{i\Delta}, \theta) \times e^{\varepsilon_{i\Delta}^{\tau_j}}$$

with $\varepsilon_{i\Delta}^{\tau_j} \sim N(\rho_j \varepsilon_{(i-1)\Delta}^{\tau_j}, s_j^2)$ and $\varepsilon_{i\Delta}^{\tau_j}$ is independent of $\varepsilon_{i\Delta}^{\tau_h}$ for $h \neq j$. ρ_j 's and s_j 's are stacked in θ_E .

I use a likelihood-based Bayesian approach for estimating multivariate Markov processes using Markov Chain Monte Carlo (MCMC) methods. The purpose of MCMC sampling in this context is to obtain a sample of parameters $\Theta = (\theta_M, \theta_J, \theta_E)$ and latent variables X from their joint posterior density $p(X, \Theta | \mathbf{Y})$ where $\mathbf{Y} = (y, Y)$ contains data on the stock price and variance swap rates. For model A_i , the posterior distribution summarizes the sample information regarding Θ and X ,

$$p(X, \Theta | \mathbf{Y}, A_i) \propto \mathcal{L}(\mathbf{Y} | X, \Theta, A_i) \cdot \mathcal{H}(X | \Theta, A_i) \cdot p(\Theta | A_i) \quad (8)$$

where $\mathcal{L}(\mathbf{Y} | X, \Theta, A_i)$ denotes the likelihood function, $\mathcal{H}(X | \Theta, A_i)$ is the density of the latent variables and $p(\Theta | A_i)$ is the prior density over the parameter vector Θ . Notice that we can estimate the marginal posterior density of the parameters, $p(\Theta | Y)$, by designing a sampling scheme that produces random draws, $\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(G)}$, whose density is $p(X, \Theta | Y)$ since, by a simple application of Bayes theorem, $p(\Theta | Y) \propto p(X, \Theta | Y)$.

3.1 The posterior simulator

The model is estimated using a Gibbs sampling procedure. In essence, this amounts to reducing a complex problem –sampling from the joint posterior distribution– into a sequence of tractable ones for which the literature already provides a solution –sampling from conditional distributions for a subset of the parameters conditional on all the other parameters. The Gibbs

sampling procedure contains several blocks: two of them involving Metropolis-within-Gibbs steps (see e.g. Geweke (2005) for an advanced textbook treatment). A brief description of them is next.

3.1.1 Spot and long-run volatilities

The conditional posterior for the spot volatility is not known in closed form. To sample from it, I use a Gaussian approximation to the transition density $p(V_{i\Delta}|V_{(i-1)\Delta}, \dots, \Theta)$ corresponding to the Euler discretization of the process (Eraker et al. (2003) show that discretization performs well with daily data; the simulation exercise in Appendix B confirms this fact). At the g th iteration of the sampler, I then draw from

$$p(V_{i\Delta}|\dots) \propto p(Y_{i\Delta}|V_{i\Delta}, \Theta^{(g-1)}) \cdot p(V_{i\Delta}|V_{(i-1)\Delta}^{(g)}, Z_{i\Delta}^{V(g-1)}, J_{i\Delta}^{(g-1)}, \theta_M^{(g-1)}) \quad (9)$$

$$\cdot p(y_{(i+1)\Delta}, V_{(i+1)\Delta}^{(g-1)}|V_{i\Delta}, Z_{(i+1)\Delta}^{y(g-1)}, Z_{(i+1)\Delta}^{V(g-1)}, J_{(i+1)\Delta}^{(g-1)}, \theta_M^{(g-1)})$$

using a random-walk Metropolis algorithm with normal proposal density with mean and variance computed as in Proposition 2 of Eraker (2001) but taking into account the presence of jumps. The acceptance rate of this step is in the 30-50% range.

The same idea is applied when the long-run variance follows a square-root process. As $p(Y_{i\Delta}|V_{i\Delta}, \alpha_{i\Delta}, \Theta)$ contains most of the information about those latent variables, the way they enter in variance swap rates may lead them to be highly correlated, and hence I sample $V_{i\Delta}$ and $\alpha_{i\Delta}$ jointly from the conditional posterior $p(V_{i\Delta}, \alpha_{i\Delta}|\dots)$, whose expression is a straightforward modification of (9) including two additional terms: $p(\alpha_{i\Delta}|\alpha_{(i-1)\Delta}^{(g)}, \theta_M^{(g-1)})$ and $p(\alpha_{(i+1)\Delta}^{(g-1)}|\alpha_{i\Delta}, \theta_M^{(g-1)})$.

As for the regime switching model, a time discretization of the model yields a discrete Markov chain with transition matrix $[1 - p_{LH}\Delta, p_{HL}\Delta; p_{LH}\Delta, 1 - p_{HL}\Delta]$. Following Albert and Chib (1993), I use the forward-filtering backward-sampling algorithm to sample s from its joint conditional distribution, which accelerates the convergence of the simulator when compared to single-move steps e.g. when sampling (V, s) jointly. Hence, I sample V from (9) conditional on the state of the chain at the iteration $g - 1$, and use Hamilton (1989)'s filter and Kim (1993)'s smoother to sample $\{s_{i\Delta}\}_{i=1}^N$ in a block.

3.1.2 Parameter vector θ_M

Conditional on jump sizes, jump times, spot variance (variance level in two-factor models) and (θ_J, θ_E) , the posterior of θ_M is proportional to (8). Since this conditional distribution is nonstandard, it is sampled using a Metropolis step using a normal source density centered at the current draw and with covariance matrix proportional to the Hessian of $\mathcal{L}(\mathbf{Y}|X, \Theta, A_i) \cdot \mathcal{H}(X|\Theta, A_i)$ at the peak of θ_M where the latent variables and remaining parameters are concentrated out using their posterior means from a preliminary run of the algorithm. The acceptance rate of this step is around 40% for the three models. The priors are relatively uninformative but still imposing the relevant constraints (e.g. κ 's and γ 's being non-negative). Specifically, they are $N(0, 5)$ for the drift parameters of the stock price, $\kappa_j^i \sim \text{Gam}(1, 5)$ and $\gamma_j \sim \text{Gam}(.2, 2)$ for $i = Q, P$ and $j = V, \alpha$, $\rho \sim U[-1, 1]$, $p_{LH}, p_{HL} \sim \text{Gam}(1, 1)$ and $1 - \varpi \sim \text{Gam}(1, 1)$. Table 2 presents mean, standard deviation and interquartile range for these priors.

3.1.3 Jump parameters, sizes and times

Conditioning on the parameter vector $(\theta'_{M_j}, \theta'_E)'$ and stochastic volatilities (either V or both V and α , depending on the model in hand), we can use standard methods to sample: (i) θ_J conditional on jump sizes and times, (ii) jump times conditional on jump sizes and θ_J , and (iii) jump sizes conditional on jump times and θ_J . Provided the priors are conjugate, the conditional posteriors of the elements of θ_J are standard: β_1 is a Gamma random variate, $(\beta_2, \beta_3)'$ is normal, β_4 is Inverse Gamma and β_5 is Beta (for details, see Johannes and Polson (2003)). The assumed data generating process (DGP, henceforth) for Z^V and Z^y leads to truncated normal and normal conditional posteriors, respectively. Expressions for the posterior means and variances for the jump sizes can be found in Appendix A. Sampling the jump times conditional on jump sizes is also straightforward: its conditional posterior is Bernoulli (for details about the computation of the Bernoulli probability see Appendix A in Eraker, Johannes and Polson (2003)). To disentangle Brownian increments from jumps i.e. to reflect the nature of jumps as large and infrequent moves in returns, I set priors that place small probability in small jump sizes: for Z^V , Exponential with expected value 0.1 and $Z^S \sim N(-0.5Z^V, \beta_4)$ with $\beta_4^{-1} \sim \text{Gam}(10, 10)$; and $\lambda\Delta \sim \text{Beta}(2, 40)$, i.e. about one jump is expected per month.

3.1.4 Pricing error parameters

For each j , I sample ρ_j and s_j using standard procedures. Conditional on $Y_{i\Delta}^{\tau_j}$, $V_{i\Delta}$ and θ_M , $\varepsilon_{i\Delta}^{\tau_j} = \log Y_{i\Delta}^{\tau_j} - \log Y(V_{i\Delta}, \theta) \sim N(\rho_j \varepsilon_{(i-1)\Delta}^{\tau_j}; s_j^2)$. Hence conditional on s_j , and with a Gaussian prior for $\rho_j | s_j$, $N(\delta_{\rho 1_j}, s_j^2 \delta_{\rho 2_j}^{-1})$, the full conditional posterior of ρ_j is normal with posterior mean and variance that are easy to compute. Assuming that $s_j^{-2} \sim \text{Gam}(\delta_{s_{j1}}, \delta_{s_{j2}})$ we have that its conditional posterior is $\text{Gam}(\delta_{s_{j1}}^*, \delta_{s_{j2}}^*)$ where again I omit their derivation since it can be found in standard textbooks. The prior parameters are $\delta_{\rho 1_j} = 0$, $\delta_{\rho 2_j}^{-1} = 100$, $\delta_{s_{j1}} = 10$ and $\delta_{s_{j2}} = 20$. The choice of the priors is such that gives high probability to small pricing errors and low probability to near unit-root behavior of the pricing errors.⁵ The value of $\delta_{\rho 1_j}$ and $\delta_{\rho 2_j}$ can be understood as increasing the sample size by 5% with dummy observations with uncorrelated pricing errors.

3.2 Formal model comparison

The marginal likelihood of model A_i , $p(\mathbf{Y}|A_i)$, is the measure of how well model A_j predicted the data \mathbf{Y} that is relevant for model comparison. Approximating the integral that comprises $p(\mathbf{Y}|A_i)$ using simulation methods is considerably more difficult than approximating the posterior distribution of the parameter vector Θ or latent variables X . For this reason I consider two different approaches: To obtain (i) the marginal likelihood and (ii) the predictive Bayes factor, which allow to determine how the different models described in Section 2 fit the data.

3.2.1 Modified harmonic mean method

Following Gelfand and Dey (1994), we can obtain an estimate of the marginal likelihood $p(\mathbf{Y}|A_i)$ by using draws from the posterior simulator to numerically evaluate

$$E \left[\frac{f(\Theta, X)}{\mathcal{L}(\mathbf{Y}|X, \Theta, A_i) \cdot \mathcal{H}(X|\Theta, A_i) \cdot p(\Theta|A_i)} \right] = \frac{1}{p(\mathbf{Y}|A_i)} \quad (10)$$

where f is a probability density function whose support is contained in the parameter space of (X, Θ) , i.e. it is required that f to have “thin tails” relative to the likelihood function. To implement it, a proper choice of f is in order. I find convenient to deal with parameters and latent variables separately, for instance in model A_I , we can write $f(\Theta, X) = f_{\Theta}(\Theta) \cdot f_V(V) \cdots \cdots f_{Z^V}(Z^V)$ and then find good candidates for the f_j ’s. To do so, I follow Geweke (1999) and choose truncated normal pdfs for Θ and the continuous-state elements of X . In

⁵The latter becomes particularly important when estimating the benchmark model: trials with priors that impose very little information on ρ lead to posterior means close to 0.99 for longer maturities.

the empirical section I set the truncating parameter p equal to 0.2. The proposed f guarantees the boundedness condition of the ratio in the right hand side of (10) when the support of the parameter space is not bounded. For that reason, I also include an additional truncating term to take into account the non-negativity constraints of V , Z^V and some elements of Θ .

3.2.2 Predictive Bayes factor

Predictive likelihoods provide an alternative method of model comparison that is directly related to the posterior model probabilities and is less sensitive to the prior distribution than direct computation of marginal likelihoods (see Geweke (2005), Section 2.6.2). We can use the posterior simulator to approximate the predictive likelihood of $\mathbf{y}_{(N+1)\Delta}$ conditional on \mathbf{Y}_N and model A_j , $p(\mathbf{y}_{(N+1)\Delta}|\mathbf{Y}_N, A_j)$ by

$$p(\mathbf{y}_{(N+1)\Delta}|\mathbf{Y}_N, A_j) \simeq G^{-1} \sum_{g=1}^G p(\mathbf{y}_{(N+1)\Delta}|\mathbf{Y}_N, X^{(g)}, \Theta^{(g)}, A_j)$$

where \mathbf{y}_t is the vector of data corresponding to date t , and $\mathbf{Y}_N = (\mathbf{y}_\Delta, \dots, \mathbf{y}_{N\Delta})'$. In the notation is implicit the use of Building samples, i.e. the use of the entire sample beginning with the start of the returns and variance swap series, as opposed to Rolling samples, which keep fixed a number of observations $N - N_0$ in \mathbf{Y} , say $\mathbf{Y}_N^0 = (\mathbf{y}_{\Delta(N-N_0)}, \dots, \mathbf{y}_{N\Delta})'$. As in Geweke and Amisano (2007), the basis for model comparison is $\sum_{i=N_1}^{N_2} \log p(\mathbf{y}_{(N+1)\Delta}|\mathbf{Y}_N, A_j)$, which allows to compute the geometric mean of the proportionate increase in the conditional probability density of the next day's realization evaluated using model A_i relative to the conditional probability density of the next day's realization evaluated using model A_j by

$$\exp \left[\frac{1}{N_2 - N_1 + 1} \sum_{n=N_1}^{N_2} (\log p(\mathbf{y}_{n\Delta}|\mathbf{Y}_{n-1}, A_i) - \log p(\mathbf{y}_{n\Delta}|\mathbf{Y}_{n-1}, A_j)) \right] - 1.$$

In the empirical Section $N_2 - N_1$ is relatively large, which makes the computation of the predictive likelihood quite time consuming. In order to obtain the previous quantities in reasonable time, I only update the most recent dates of the latent variables and Θ daily by using the so-called practical filter approach to sequential estimation.

3.3 Parameter and latent variables updating

Recent research on sequential estimation focusses on alternative schemes for implementing Bayesian sequential estimation. Here, I use the so-called practical filter of Johannes et al.

(2003) and Stroud et al. (2003) which uses a fixed-lag filtering. The practical filter relies on the following decomposition of the joint distribution of parameters and latent variables

$$p(\Theta, X_{(i+1)\Delta} | \mathbf{Y}_{i+1}) = \int p(\Theta, X_{(i-k+1)\Delta}^{(i+1)\Delta} | X_{(i-k)\Delta}, \mathbf{Y}_{i+1}) p(X_{(i-k)\Delta} | \mathbf{Y}_{i+1}) dX_{(i-k)\Delta}^{i\Delta}.$$

The fixed-lag filtering is based on the idea that today’s observation provides little information about states in the distant past, beyond the information embedded in past data. In the filtering recursion then $p(X_{(i-k)\Delta} | \mathbf{Y}_{i+1}) \approx p(X_{(i-k)\Delta} | \mathbf{Y}_i)$ i.e. $Y_{(i+1)\Delta}$ as an additional observation has little impact on $X_{(i-k)\Delta}$. Then, the posterior simulator described above can be used to generate samples from $p(\Theta, X_{(i-k+1)\Delta}^{(i+1)\Delta} | X_{(i-k)\Delta}, \mathbf{Y}_{i+1})$, which provides samples from $p(\Theta, X_{(i+1)\Delta} | \mathbf{Y}_{i+1})$. This way, we reduce the computational cost by sampling k realizations of the latent variables instead of $N + 1$. For a detailed description of its implementation see Johannes et al. (2003).

The choice of the number of draws G and the value of k have a central effect in the efficiency and accuracy of the algorithm. While a large value of k improves the approximation, the computational costs increase linearly. Following Johannes et al. (2003) I set $k = 21$ so that latent variables updating occurs to realizations of the last month. As for G I keep 500 draws using a thinning parameter of 25. Starting from the stationary distribution of the “smoothed” estimators from over 2,000 observations, adding one observation may not have a drastic impact in the posteriors so that we can expect convergence of the chain after 10,000 iterations.

Even though the practical filter is useful to alleviate the computational cost of computing the predictive Bayes factor using over a thousand observations, its output is of interest itself. It provides information that reflect changes in volatility and the agents’ view of the parameters. Sequential parameter estimation by itself becomes a relevant issue when dealing with stochastic volatility models as is discussed in the next Section.

4 Empirical results

4.1 The data

I estimate the models using daily data on the S&P 500 index returns and variance swap rates on six different maturities (1, 2, 3, 6, 12 and 24 months) over the period January 4, 1996 to December 31, 2003. The number of daily observations is 2013, excluding weekends and holidays.⁶ The remaining sampling period, January 2, 2004 to August 29, 2008 is kept for sequential parameter and state estimation, and to compute the predictive Bayes factor (see Figure 2). The sample is constructed as follows: it contains data on variance swaps quotes from Bank of America over the period January 4, 1996 to March 30, 2007, while the data for the remaining period is obtained from Bloomberg.

Descriptive statistics for the two subsamples are in Table 3. The first subsample is characterized by a significantly higher market volatility which is due in part to the Asian, Russian and LTCM crises. Nevertheless, the last year of the second subsample presents a moderate high level of volatility that does not seem to be represented in the summary statistics due to the “quiet” period that characterized the market between the years 2004 and 2007. For both subsamples we observe an upward sloping mean term structure that, according to the specifications of Section 2, suggests negative risk premia for volatilities (column 2 of Tables 3.A and 3.B). Finally, the eigenvalue decomposition of the covariance matrix of variance swap rates in Table 4 shows that the main principal component explains around 90% of the total variation in the data, while the first two explain more than 99%. The corresponding eigenvectors suggest that the first eigenvalue is related to level shifts in the curve while the second one captures changes in the slope of the variance curve. Interestingly, when compared with the values of Litterman and Scheinkman (1991) we find a simpler structure in the term structure of variances than the corresponding to bond returns: The convexity effect becomes negligible for variance swap data. Notice also that the contribution of the second factor is more substantial in the first subperiod i.e. the one that includes more “turbulent” periods.

4.2 Parameter estimates

In Table 5 I report the posterior means and 90% high probability densities of the parameter vector θ_M for the three models. Parameters are defined in annual terms following the convention

⁶Different from Egloff et al. (2007), I use all the data and not the every-Wednesday weekly sample strategy they use due to two reasons. First, the Euler discretization used for estimation: Its impact in parameter estimates is negligible at the daily sampling frequency (see e.g. Johannes et al. (2003) or Appendix B). Second, I do not find significant weekday’s effect through a preliminar examination of the data.

in the empirical option pricing literature (e.g. Bates (2000) or Aït-Sahalia and Kimmel (2007)). Convergence diagnostics are in the Appendix C.

Estimates from model *I* present some interesting features. The speed of mean reversion of the stochastic variance is significantly lower than the values found in recent works based on data from stock returns and/or option prices (see e.g. Pan (2002) or Aït-Sahalia and Kimmel (2007)). Still they are consistent with the empirical results of Egloff et al. (2007), who also use data on variance swap rates. This mean reversion is remarkably lower under the pricing measure Q , which implies a substantial negative market price of the risk associate to the spot variance. This feature, which is also present in models *II* and *III*, is well documented in the related literature including different data and estimation methodologies (e.g. Aït-Sahalia and Kimmel (2007), Bondarenko (2007) and Pan (2002) among others). The values of mean reversion parameters under the Q measure are somewhat lower than the ones found in previous works: the daily autocorrelation they imply, which can be computed as $e^{-\kappa\Delta}$, is very close to 1: 0.994 and 0.997 under both the pricing and physical measure, respectively. In the same line, the long-run level of volatility is slightly higher than the values suggested in the option pricing literature: The posterior mean of $\bar{\alpha}^Q$ suggests a mean level of volatility around 25%. With respect to the volatility of volatility γ_V , the results are similar to previous works that use data on option prices, and as is well documented, its mean value is higher than the values reported in empirical work based on returns data only (e.g. Eraker et al. (2003)). As for correlation parameters, we find that over the sampling period the negative correlation between shocks to returns and volatility shocks is remarkably strong.

I now turn to discuss the parameter estimates of model *II*, in which α_t follows a square-root process. In this model, the mean reversion parameter of V is closer to the values found in the literature under both probability measures. As noted in the previous paragraph, the risk premium associated to V is again negative and significative. Interestingly, the estimates of the mean reversion parameter of α under both the pricing and physical measure are very low, implying that shocks to α have a half life of several years. These parameter estimates are consistent with the interpretation of the second factor, α , as capturing the low-frequency dynamics of the market volatility. The market price of risk associated to shocks from W_2 is negative; even though its 90% HPD includes a small region in the positive real line, only 20% of the draws lie in that region. The fact that the posterior distribution of the risk premium associated to α contains large uncertainty about its value, however, is not surprising given that the sample period is 8 years long, and that risk premia are typically poorly estimated even

in longer samples. In fact, since $\eta = (\kappa^Q - \kappa^P)/\gamma$ we should note that while the likelihood function provides rich information about about κ_α^Q –because it enters in the price function– and the daily sampling frequency does the same with γ_α ; κ_α^P only appears in the drift function under the objective measure, and the quality of the estimates of drift parameters typically depends only on the length of the sample, and not the sampling frequency. Of course the same argument applies to the drift parameters of V , as can be seen from the respective 90% HPDs. Finally, the estimated correlation between W_0 and W_1 is in line with the results from model *I*.

Next, we take a look at the parameter estimates of model *III*: the two-state regime-switching specification for α . The mean reversion speed of the spot variance under the objective measure increases with respect to the benchmark model, i.e. a change in the same direction as in model *II*, but smaller in magnitude. As for the remaining parameters characterizing the dynamics of the spot variance, the parameter estimates are quite similar to the ones for the benchmark model.

The transition parameters estimates of the hidden Markov chain suggest that quiet periods of the market –with lower mean level of volatility α – (denoted with the subscript L) have an average duration of two years, and a duration of only one quarter for turbulent periods (denoted with the subscript H). Estimates of ϖ show a negative market price of the risk associated to changes in regime, which are congruent with results in the previous models. As for the levels of α , the results suggest that the spot volatility fluctuates around 22% in quiet periods, but this value increases to 50% in high volatility regimes.

Table 6 contains summary statistics of the posterior distribution of the jump parameters θ_J . First note that the posterior means for the different parameters are quite similar across the models. The short sampling period makes jump-size parameters difficult to identify: less than 20 jumps are expected in the 8 years period analyzed according to the estimates of λ . Still, parameter estimates are consistent with previous works (see e.g. Eraker et al. (2003), Eraker (2004)): Jumps are very infrequent, as they occur around twice per year in average, and jumps in the spot and volatility are negatively correlated.

4.3 Latent variables

The posterior simulator generates samples of the latent variables drawn from the joint posterior distribution. Hence, the MCMC approach provides a straightforward method to estimate the volatilities, jump times and jump sizes by computing the posterior expectation

of these variables. Furthermore, these estimates take into account parameter uncertainty. For instance, from the samples of $V_{i\Delta}$, $\{V_{i\Delta}^{(g)}\}_{g=1}^G$, we compute an estimate of $E(V_{i\Delta}|\mathbf{Y})$ by calculating $G^{-1}\sum_{g=1}^G V_{i\Delta}^{(g)}$ as opposed to $E(V_{i\Delta}|\mathbf{Y}, \hat{\Theta})$, with $\hat{\Theta}$ denoting an estimator of the parameter vector Θ , which treats parameter estimates as known, ignoring the fact that they are random variables.

Figures 3 to 5 report the posterior mean of the latent variables for the three models. The posterior distribution of jump sizes are similar across the models, with a high concentration of jumps during the LTCM crisis. Even though parameter estimates suggest a couple of jumps per year during the period, most of the jumps occur in the two-months period right after the Summer of 1998. Other jumps of significant size are mainly associated to the Asian crisis and to September 11, 2001.

As expected, the variance level α in model *II* increases above 0.10 in September of 1998 and stays at high levels for more than one year with a very slow reversion –that takes a couple of years– to the magnitudes previous to LTCM crisis, which is consistent with the estimates of the mean reversion parameter under the physical measure for this process. Interestingly, a projection of the posterior mean of V from model *I* onto the posterior means of V and α from model *II* yields statistically significant coefficients of 0.76 and 0.25 respectively. That is, if we believe that model *I* is misspecified and the data was generated under model *II*, this fact suggests that the one-factor specification estimates of the spot variance are in fact estimates of a combination of the true spot variance and its tendency. In fact, parameter estimates of the volatility of volatility for the two processes are consistent with the previous interpretation of the factors: γ_V in model *I* lies between the estimates of γ_V and γ_α in model *II*; the same happens with the mean reversion parameters.

Finally, the smoothed estimates of the hidden Markov chain also present some interesting features. The second panel of Figure 5 presents estimates of the probability of high volatility level. Interestingly, the plot suggests that the market anticipates the turbulent period that starts in August 1998 by three months. Note that this fact is in contrast with the estimates of the volatility level of model *II*, which starts increasing by late August of that year. The discrete space nature of α in model *III* hence suggest some ability to accommodate sudden changes in market’s volatility, at least in-sample.

4.4 Model specification analysis

By analyzing the properties of pricing errors we can learn about the statistical performance of the different models in an informal but intuitive way. Ideally, pricing errors should be uncorrelated and with small variance. I present the correlation and standard deviation parameters that characterize the pricing error parameters θ_E in Table 7. Looking at the first column, we see that the one-factor model performs poorly, as suggested by the high values of the autocorrelation coefficients at longer maturities. While extending the model in the direction of model *III* slightly reduces these values, still the reduction in magnitude of the standard deviation parameters is small. Clearly, the parameter estimates of model *II* suggest that it outperforms the other models. Except for the correlation parameter of the 1 month time-to maturity variance swap pricing error, the values of ρ_j are much smaller when compared to either models *I* or *III*. Similarly, there is a significant reduction in the standard deviation of the pricing errors' innovations which become less than a half of the corresponding ones in the competing models for longer maturities.

Now we take a look at formal model comparison procedures. Panel A of Table 8 provides the log marginal data densities for the three models i.e. the quantities that –together with the prior odd ratio– allow for computation of the posterior odd ratios. Model *II* clearly outperforms its competitors. Also model *III* significantly improves the fit of the data when compared to the benchmark case (recall that the results are in logs). It is important to note that posterior odd ratios do not necessarily favor more complex models; it is a method that automatically penalizes models with unnecessary free parameters.

The predictive Bayes factor provides additional evidence in favor of the two-factor model with square-root level of variance. As can be seen from Panel B of Table 8, the geometric mean of the proportionate increase in the conditional probability density of the next day's realization evaluated using model *II* relative to the conditional probability density of the next day's realization evaluated using model *I* is over 300%. Even though this is a very large improvement in the statistical fit of the data, it should not be surprising since at each point in time model *II* has two degrees of freedom to fit the cross-section of variance swaps. As for model *III*, it underperforms the benchmark by 1%.

From the previous results, it is clear that a second factor is needed. At daily sampling frequency, a two-state regime-switching model for the level of variance allows for identification of turbulent and quiet market periods, however it is not as flexible as model *II* in which α takes

values on the positive real line. While we could conjecture that the distance in explanatory power between models *II* and *III* may be smaller if the data were sampled on a monthly basis, extensions of model *III* that include more regimes could also close the gap between the alternative formulations of the two-factor models.

Finally, we can use the output from the practical filter to check for model specification through analyzing parameters stability. It is frequently argued that structural models are often misspecified; in this sense, the subsample choice provides a tough test for model misspecification since the average volatility during the second period is significantly lower than over the period used for smoothed inference. Filtered estimates for the best-fit model are in Figure 6. As expected, the low volatility of years 2004 to 2006 drives the vol-vol parameters down suggesting that an improved fit should require a non-affine diffusion specification even if we allow for jumps in both returns and volatility. However, we can see that speed of mean reversion parameters, which are the main estimates to analyze the term structure of variance risk premia, are quite stable. Similar results are obtained for the remaining models, with the exception that $\bar{\alpha}^Q$ in model *I* and α_L in model *III* present a small downward trend.

5 The term structure of variance risk premia

5.1 τ -maturity risk premia

While several works emphasize an instantaneous negative variance risk premia (Aït-Sahalia and Kimmel (2007), Bates (2000), Bondarenko (2004) and Pan (2002) among others), it is interesting to analyze implications of parameter estimates under the alternative models for the risk premia associated to a given time-to-maturity τ . In this sense, the parametric assumptions of Section 2 give us a simple framework to study the term structure of the different components of the variance risk premium. Furthermore, posterior draws allow us to incorporate into the analysis both the uncertainty about parameters and latent variables for each model in a straightforward manner.

At maturity the payoff of the long side of the swap is equal to the difference between the realized variance over the life of the contract and the variance swap rate, $(RV_{t,t+\tau} - Y_t^\tau) \times L$, where $RV_{t,T}$ denotes the realized annualized return variance between t and $t + \tau$ and L denotes the notional dollar amount that converts the variance difference into a dollar payoff. As discussed in Carr and Wu (2008), this quantity represents the dollar capital gain from going

long a L par variance swap contract and holding it to maturity. Here we focus on the quantity $E_t^P[RV_{t,t+\tau}] - Y_t^\tau$.

Using (3) to write (4) as

$$\begin{aligned} Y_t^\tau &= \frac{1}{\tau} \frac{1}{E_t^P[M_{t,t+\tau}]} E_t^P \left[M_{t,t+\tau} \int_t^{t+\tau} V_s ds \right] \\ &= \frac{1}{\tau} E_t^P \left[\int_t^{t+\tau} V_s ds \right] + \frac{1}{\tau} Cov_t^P \left(\frac{M_{t,t+\tau}}{E_t^P[M_{t,t+\tau}]}, \int_t^{t+\tau} V_s ds \right) \end{aligned}$$

where $M_{t,t+\tau}$ denotes the stochastic exponential of the pricing kernel defined in (3), we observe that the variance swap rate can be decomposed into the time-series conditional mean of the realized variance and the conditional covariance between the normalized pricing kernel and the realized variance. The negative of this second term is defined by the variance risk premia. The negative estimates market prices of variance risk, then, give an estimate of the terminal profit and loss from going long a variance swap contract and holding it to maturity.

The pricing formulas of Section 2 can be used to obtain explicit expressions for the variance risk premia. In particular, they take the simple linear structure

$$RP_t^\tau = -\frac{1}{\tau} Cov_t^P \left(\frac{M_{t,t+\tau}}{E_t^P[M_{t,t+\tau}]}, \int_t^{t+\tau} V_s ds \right) \quad (11)$$

$$= A_j(\tau) \bar{\alpha}^Q + B_j(\tau) V_t + C_j(\tau) \alpha_t, \quad (12)$$

for $j = I, II, III$ and where the expressions for A_j , B_j and C_j are easy to compute. In order to analyze the risk premia implied by each model I present plots that characterize the posterior distribution of A_j , B_j and C_j in Figures 7, 8 and 9 as a function of time-to-maturity as well as the marginal risk premia i.e. the unconditional expectation of (11).

The term structure for model I is reported in Figure 7. First, we can see that the mean of the posterior distribution of the loadings associated to both $\bar{\alpha}^Q$ and V_t are negative; which is an obvious consequence of $\eta_V < 0$. This result may appear contrary to the fact that risk averse agents require a positive premium in order to accept an increase in the unconditional market volatility; however, the loading A_I is capturing the forward variance risk at longer maturities –once the forecast of V_t is close to its unconditional level– which is also negative. In this sense, we should also note that as τ goes to infinity A_I converges to $\kappa_V^Q / (\kappa_V^Q - \eta_V \gamma_V) - 1$. The low mean reversion of V_t leads to a significant contribution of this factor to the total risk premia even for maturities longer than one year. As a natural consequence of A_I and B_I being negative, the marginal term structure of risk premium (i.e. once we integrate V out) is also

negative. In fact, this is also the case for models *II* and *III* since all market prices of variance risk are found to be negative.

The term structure for the best-fit model in Figure 8 presents some remarkable facts. First the loading for $\bar{\alpha}^Q$ is positive for short maturities while the loadings associated to the stochastic components of volatility are negative, which suggests that both V and α serve as a hedge against return specific risks. According to parameter estimates and our interpretation of volatility factors, B_{II} becomes dominant at short maturities –mainly around six months– while the near martingale behavior of α_t makes the loading C_{II} increasing in time-to-maturity –even at 10-year horizons. Note the difference with the loadings from the one-factor structure: At a 10 year horizon, the impact of the current state of the stochastic variance on total risk premia is much smaller. The requirement for this unique factor to fit both the short-run and long-run dynamics of variance does not allow to capture the sensitivity of risk premia to the current state of the economy at long horizons.

The corresponding curves for model *III*, which are reported in Figure 9, offer ambiguous results. As in the best-fit model, the loading associated to the unconditional level of variance is positive for short maturities but then it becomes negative at longer horizons. As for the loading of the spot variance, it is increasing in time-to-maturity and similar to the corresponding to the one-factor model. Interestingly, C_{III} is positive for short horizons and it becomes negative for maturities larger than 6 months. The fact that this loading is positive for small τ can be explained by the discrete and sudden nature of a change in regime: When there is a change in regime V_t is driven towards a new level which according to parameter estimates is far away from the previous mean level, the low mean of reversion for V suggest that this convergence may take a few months.

5.2 Risk premia dynamics

Parametric assumptions implicit in the specification of the Q -measure dynamics and market prices of risk restrict the risk premia to linear combinations of risk factors with weights being functions of time-to-maturity. In other words, Itô's lemma imply that risk premia inherits the dynamic properties of the variance factors. Still, it is interesting to study the variance risk premia dynamics over the sample, and in particular to see what volatility factor is dominant and how this contribution changes over time, particularly during crises or periods of high volatility. To do so, in this Subsection I restrict attention to the best-fit model.

In what follows I consider a couple of examples to analyze the risk premia dynamics. From the posterior draws of both parameters and latent variables we can compute the τ -maturity variance swap risk premia for each day over the entire sample. In particular, I choose 3-months and 1-year maturities since they are the ones that model *II* fits better. The dynamics for those risk premia dynamics as well as its difference, i.e. shorting one unit of 1-year variance swap while being long in the other, are presented in Figure 10.

By looking at Figure 10 we find a larger risk premia associated to the longer maturity rate; moreover, the difference between the two variance swap risk premiums $RP_t^{1/4} - RP_t^1$ increases during turbulent periods. It is also noteworthy that this difference is positive during the whole sample. This is not a trivial result: A large increase in V_t that is not corresponded by a similar increase of α_t would yield negative values of $RP_t^{1/4} - RP_t^1$. In fact, as can be seen in Figure 11 –which presents a decomposition of the 1 year time-to-maturity variance risk premia–, we can see that during turbulent periods the short-run factor is the dominant. Note, however, that for most of the sample its contribution is below the one corresponding to the long-run factor. Of course, at shorter horizons we expect a larger contribution of the spot variance risk.

5.3 Two P -factors, one Q -factor

For the best-fit model, estimates of the mean reversion parameter of α under the physical measure but particularly under the pricing measure are very low, implying a near unit-root behavior of the process under both measures. A low value of κ_α^Q implies term structures of variance swap rates monotonically increasing or decreasing; a feature that is supported by the principal components decomposition of variance swap rates: over 99% of the prices variation comes from changes in level and slope leaving little space for changes in convexity as a significant common factor driving the term structure of variance risk premia.

Given the small value of κ_α^Q that can be inferred from its posterior distribution, which implies that shocks to α have a half life of more than a decade on average, it is interesting to study the differences in prices that arise from imposing the restriction $\kappa_\alpha^Q = 0$. This restriction can be understood as market participants' perception of the level of market volatility as being unpredictable; so that the best forecast they can use for today's valuation is yesterday level of volatility. Not surprisingly, when this is the case we have that

$$\lim_{\kappa_\alpha^Q \rightarrow 0} \phi^{II}(\tau) = 1 - \phi^I(\tau) \quad \text{and} \quad \lim_{q \rightarrow 0} \phi^{III}(\tau) = 1 - \phi^I(\tau)$$

so that (5) become simply a weighted average of the current values of V and α ,

$$Y_t^\tau = \phi^I(\tau)V_t + [1 - \phi^I(\tau)]\alpha_t,$$

that is, a *local* one-factor structure for the spot variance. The same result holds for model *III*; which immediately follows by rewriting (7) as

$$Y_t^\tau = \bar{\alpha}^Q [1 - \phi^I(\tau) - \phi^{III}(\tau)] + \phi^I(\tau)V_t + \phi^{III}(\tau)\mathbf{1}\{\alpha = \alpha_j\}\alpha_j$$

where $\bar{\alpha}^Q = (q_{HL}^Q\alpha_L + q_{LH}^Q\alpha_H)/q$ and $j = L, H$.

Provided the current value of the variance level α_t is treated as a parameter, we could borrow standard formulae from one-factor models for pricing purposes. Moreover, this result suggest that we could parameterize α as subject to more complex dynamics under the P measure, while keeping the asset pricing in closed form. The only requirement is that α be a martingale under the Q measure. Of course, for practical purposes the value of second factor under the P -measure has to be filtered in order to get the value of the “parameter” α . Notice also that this implication, which is associated to the martingale dynamics of α under the pricing measure, is not specific to variance swap rates; it does apply to other derivatives.

Draws from the posterior simulator can be directly used to quantify the accuracy of prices obtained from this *local* one-factor model by computing the posterior distribution of the loadings entering in the pricing formulas. Posterior means and 90% HPD for the loadings of models *II* and *III* are in Table 9. For both models, the loading associated to $\bar{\alpha}^Q$ takes values smaller than 0.025 for maturities up to six months. For the best-fit model, in particular, it is remarkable that the loading corresponding to $\bar{\alpha}^Q$ remains below 0.05 even for the 2 year time-to-maturity variance swap rate. If we compare the contribution of $\bar{\alpha}^Q$ to Y_t^τ with the magnitude of the pricing errors, we find that biases that may arise from ignoring the third loading can be accommodated within the pricing error, even taking as a reference the pricing error parameters of the three-months rate in model *II*. This is particularly the case for maturities shorter than one year. It is interesting to note also the different pattern of the loadings of V_t and α_t as time-to-maturity increases; α_t becomes the main factor for longer maturities in model *II* while for model *III* the loading of V_t remains above 0.60 even for two-years time-to-maturity variance swaps.

6 Concluding remarks

Fitting one-factor stochastic volatility models to the term structure of variance derivatives may lead to the following misleading results: (i) low mean reversion of the spot variance which is not consistent with the observed mean reversion from the analysis of short time-to-maturity derivatives; and (ii) understatement of the long-run risk, the dominant component for explaining the variance risk premia. This long-run risk, associated with a near martingale behavior of the corresponding factor, suggests agents face large uncertainty about the future level of volatility, and it has an important impact in prices of variance derivatives and risk premia at long horizons.

While in the context of stochastic volatility models, Bates (2000) was probably the first in analyzing models with two variance factors almost a decade ago, its relevance to fit variance dynamics, and particularly term structure properties of both risk premia and derivative prices has been understated. This paper shows that a second factor is necessary. Whether it is sufficient or not to also fit the cross section of option prices for a given time-to-maturity is left for future research.

Appendix

A Auxiliary calculations

A.1 Derivation of equation (6)

First, let's introduce some additional notation for the Markov chain s_t with state space $\{1, \dots, d\}$ and intensity matrix $\mathbf{Q} = (q_{ij})$. Let the mark space E be given by $E = \{(i, j) : i \in \{1, \dots, d\}, j \in \{1, \dots, d\}, i \neq j\}$ with σ -algebra $\mathcal{E} = 2^E$, and let m be the marked point process on this mark space with compensator

$$\xi(t, s_{t-}(\omega), dz)dt = \left[\sum_{i \neq j} q_{ij} \mathbf{1}\{s_{t-}(\omega) = i\} \epsilon_{(i,j)}(dz) \right] dt,$$

where $\mathbf{1}\{A\}$ denotes the indicator function of the set A , and ϵ_p denotes the Dirac measure at point p . The relevant filtration for our purposes is $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$, the filtration generated by both W and m , i.e.

$$\mathcal{F}_t = \sigma\{W_s, m([0, s] \times A), B; 0 \leq s \leq t, A \in \mathcal{E}, B \in \mathcal{N}\}$$

where \mathcal{N} is the collection of Q -null sets from \mathcal{F} . For simplicity in what follows I assume that S and V cannot jump; still the same result holds. Then, we can represent the dynamics of the regime-switching variance level model under the Q measure as

$$\begin{aligned} dS_t &= (r - \delta) dt + \sqrt{V} S_t \left[\sqrt{1 - \rho^2} dW_{0t}^Q + \rho dW_{1t}^Q \right] \\ dV_t &= \kappa_V^Q(\alpha_t(s_t) - V_t) dt + \gamma_V \sqrt{V} dW_{1t}^Q \\ ds_t &= \int_E \delta(z) m(dt, dz) \end{aligned}$$

where $\delta(z) = \delta((i, j)) = j - i$.

Standard arbitrage arguments imply that the price P of a derivative security must satisfy the following PDE

$$\begin{aligned} 0 &= \frac{\partial P(t, \alpha, S, V)}{\partial t} + \frac{\partial P(t, \alpha, S, V)}{\partial S} (r - \delta) + \frac{\partial P(t, \alpha, S, V)}{\partial V} \kappa_V^Q(\alpha_t - V) \\ &+ \frac{1}{2} \frac{\partial^2 P(t, \alpha, S, V)}{\partial S^2} S^2 V + \frac{1}{2} \frac{\partial^2 P(t, \alpha, S, V)}{\partial V^2} \gamma_V^2 V + \frac{\partial^2 P(t, \alpha, S, V)}{\partial S^2 \partial V} \gamma_V \rho S V \\ &+ \int_E [P(t, \alpha(s_t + \delta(t, z)), S, V) - P(t, \alpha(s_t), S, V)] \xi(t, dz). \end{aligned}$$

Next, write $P(t, \alpha = \alpha_i, S, V) = \bar{P}(t, \bar{\alpha}, S, V) + P_i(t, \alpha, S, V)$ where $\bar{P}(t, \bar{\alpha}, S, V)$ satisfies

$$\begin{aligned} 0 &= \frac{\partial \bar{P}(t, S, V)}{\partial t} + \frac{\partial \bar{P}(t, S, V)}{\partial S} (r - \delta) + \frac{\partial \bar{P}(t, S, V)}{\partial V} \kappa(\bar{\alpha} - V) \\ &+ \frac{1}{2} \frac{\partial^2 \bar{P}(t, S, V)}{\partial S^2} V S^2 + \frac{1}{2} \frac{\partial^2 \bar{P}(t, S, V)}{\partial V^2} \gamma_V^2 V + \frac{\partial^2 \bar{P}(t, S, V)}{\partial S^2 \partial V} \gamma_V \rho V S, \end{aligned}$$

and $\bar{\alpha}$ is the expected value of the invariant distribution of the long-run level of volatility, that is $\bar{\alpha} = \boldsymbol{\alpha} \cdot \bar{\mathbf{q}}$ where $\bar{\mathbf{q}} : \bar{\mathbf{q}}' \mathbf{Q} = 0$. Notice that under our assumptions the affine structure of the model allows us to compute \bar{P} using the transform-based approach analysis [see e.g. Heston (1993), Bates (2000) and Duffie, Pan and Singleton (2000)]. The final step then consists in solving for the P_i 's. Denoting by \mathbf{P} the vector of stacked P_i 's, $(P_1(t, \alpha, S, V), \dots, P_d(t, \alpha, S, V))'$, the first PDE implies

$$\frac{\partial \mathbf{P}(\tau; S, V)}{\partial \tau} = \mathbf{Q} \cdot \mathbf{P}(\tau; S, V) - \kappa_V^Q (\bar{\alpha} \ell - \boldsymbol{\alpha}) \frac{\partial \bar{P}(\tau, S, V)}{\partial V}$$

where $\ell = (1, \dots, 1)'$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ and $\tau = T - t$. Given that $\mathbf{P}(0; S, V) = \mathbf{0}$, we have that the solution to the system of differential equations above is given by (6).

A.2 Derivation of equation (7)

Write $Y_t^\tau = \mathbf{A}(\tau) + B(\tau)V_t$, where \mathbf{A} is the vector that captures the impact of the regimes into the price. In the two-state case \mathbf{Q} has a decomposition $\mathbf{J} = \mathbf{U}^{-1} \mathbf{Q} \mathbf{U}$ where

$$\mathbf{J} = \begin{bmatrix} 0 & 0 \\ 0 & -q \end{bmatrix}, \quad \mathbf{U} = \frac{1}{q} \begin{bmatrix} q_{HL} & q_{LH} \\ q_{LH} & -q_{HL} \end{bmatrix}$$

with $q = q_{LH} + q_{HL}$. Then, to obtain the integrand of $\mathbf{A}(\tau)$ we first compute

$$\mathbf{U} \begin{bmatrix} 1 & 0 \\ 0 & e^{-q\xi} \end{bmatrix} \mathbf{U}^{-1} \boldsymbol{\alpha} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{q_{HL}\alpha_L + q_{LH}\alpha_H}{q} + \begin{bmatrix} q_{LH} \\ -q_{HL} \end{bmatrix} \frac{(\alpha_L - \alpha_H)e^{-q\xi}}{q}.$$

Next, as $B(s) = [1 - e^{-\kappa s}]/\kappa$, $\int_0^\tau e^{-\kappa s} ds = (1 - e^{-\kappa\tau})/\kappa$, $\int_0^\tau e^{-qs} ds = (1 - e^{-q\tau})/q$ and $\int_0^\tau e^{-(\kappa+q)s} ds = (1 - e^{-(\kappa+q)\tau})/(\kappa + q)$, we have that

$$\begin{aligned} \mathbf{A}(\tau) &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{q_{HL}\alpha_L + q_{LH}\alpha_H}{q} \left[\tau - \frac{1 - e^{-\kappa\tau}}{\kappa} \right] \\ &\quad + \begin{bmatrix} q_{LH} \\ -q_{HL} \end{bmatrix} \frac{\alpha_L - \alpha_H}{q} \left[\frac{1 - e^{-q\tau}}{q} - \frac{1 - e^{-(\kappa+q)\tau}}{\kappa + q} \right] \end{aligned}$$

from where equation (7) follows.

A.3 Posterior of jump sizes

The conditional posterior of the volatility jumps is $p(z_{t\Delta}^V | N_{t\Delta} = 1, z_{t\Delta}^S, X, \Theta)$ which, by Bayes rule is

$$p(y_{t\Delta}, V_{t\Delta} | V_{(t-1)\Delta}, J_{t\Delta} = 1, z_{t\Delta}^S, z_{t\Delta}^V, \Theta) \cdot p(z_{t\Delta}^V | J_{t\Delta} = 1, z_{t\Delta}^S, \Theta)$$

where the first term is a bivariate normal distribution

$$p(y_{(i+1)\Delta}, V_{(i+1)\Delta} | V_{i\Delta}, \dots) \propto \frac{1}{(1 - \rho^2)\gamma_V^2 V_{i\Delta}\Delta} \exp\left(-\frac{1}{2V_{(i-1)\Delta}\Delta} v' \begin{bmatrix} 1 & \rho\gamma_V \\ \rho\gamma_V & \gamma_V^2 \end{bmatrix}^{-1} v\right).$$

where $v = (x_{1,(i+1)\Delta}, x_{2,(i+1)\Delta})'$, $x_{1,(i+1)\Delta} = y_{(i+1)\Delta} - z_{(i+1)\Delta}^S j_{(i+1)\Delta} - b_0 - b_1 V_{i\Delta}$ and $x_{2,(i+1)\Delta} = V_{(i+1)\Delta} - z_{(i+1)\Delta}^V j_{(i+1)\Delta} - c_0 - c_1 V_{i\Delta}$ with

$$\begin{aligned} b_0 &= (r - \delta)\Delta, \\ b_1 &= \left(\eta_S(1 - \rho^2) + \eta_V\rho - \frac{1}{2}\right)\Delta, \\ c_0 &= \frac{\kappa_V^Q}{\kappa_V^Q - \eta_V\gamma_V} \alpha^Q \Delta, \text{ and} \\ c_1 &= 1 - (\kappa_V^Q - \eta_V\gamma_V)\Delta, \end{aligned}$$

As for the second term, by Bayes rule, it is proportional to the product of $p(z_{t\Delta}^S | z_{t\Delta}^V, J_{t\Delta} = 1, \Theta)$ and $p(z_{t\Delta}^V | J_{t\Delta} = 1, \Theta)$, i.e. the product of a normal, $z_{t\Delta}^S | z_{t\Delta}^V \sim N(\beta_2 + \beta_3 z_{t\Delta}^V, \beta_4)$, and an exponential, $z_{t\Delta}^V \sim \exp(\beta_1)$. To find the conditional posterior, completing the square for all three terms as a function of $z_{t\Delta}^V$ leads to a truncated normal.

By completing the squares, the conditional posterior of $z_{t\Delta}^V$ is $TN(\mu_{t\Delta}^*, \sigma_{t\Delta}^{*2}; z_{t\Delta}^V > 0)$ with

$$\begin{aligned} \mu_{t\Delta}^{V*} &= \frac{\beta_3^2 V_{(t-1)\Delta} \Delta \gamma^2 (1 - \rho^2)}{\beta_4 + \beta_3^2 V_{(t-1)\Delta} \Delta \gamma^2 (1 - \rho^2)} \frac{z_{t\Delta}^S - \beta_2}{\beta_3} + \frac{\beta_4}{\beta_4 + \beta_3^2 V_{(t-1)\Delta} \Delta \gamma^2 (1 - \rho^2)} \\ &\quad \times [V_{t\Delta} - \gamma_V \rho (y_{t\Delta} - z_{t\Delta}^S) - c_0 - c_1 V_{(t-1)\Delta} + \gamma_V \rho (b_0 + b_1 V_{(t-1)\Delta}) - \gamma_V^2 \beta_1 (1 - \rho^2) \Delta V_{(t-1)\Delta}] \end{aligned}$$

and

$$\sigma_{t\Delta}^{V*2} = \frac{V_{(t-1)\Delta} \Delta \gamma^2 (1 - \rho^2) \beta_4}{\beta_4 + \beta_3^2 V_{(t-1)\Delta} \Delta \gamma^2 (1 - \rho^2)}.$$

When $N_{t\Delta} = 0$, the conditional posterior is $\exp(\beta_1)$, as the data provides no information about the jump sizes.

The posterior for the jumps in returns is similarly derived; Bayes rule implies that $p(z_{t\Delta}^S | J_{t\Delta} = 1, z_{t\Delta}^V, X, \Theta)$ is proportional to

$$p(y_{t\Delta} | V_{(t-1)\Delta}, J_{t\Delta} = 1, z_{t\Delta}^S, z_{t\Delta}^V, \Theta) \cdot p(z_{t\Delta}^S | J_{t\Delta} = 1, z_{t\Delta}^V, \Theta)$$

where the first term is a normal distribution with mean $b_0 + b_1 V_{(t-1)\Delta} + z_{t\Delta}^S$ and variance $V_{(t-1)\Delta} \Delta$, while the second term is also normal with mean $\beta_2 + \beta_3 z_{t\Delta}^V$ and variance β_4 , so that

the conditional posterior of $z_{t\Delta}^S$ is $N(\mu_{t\Delta}^{S*}, \sigma_{t\Delta}^{S*2})$ with

$$\begin{aligned}\mu_{t\Delta}^{S*} &= \frac{\Delta V_{(t-1)\Delta}}{\Delta V_{(t-1)\Delta} + \beta_4} (\beta_2 + \beta_3 z_{t\Delta}^V) + \frac{\beta_4}{\Delta V_{(t-1)\Delta} + \beta_4} (y_{t\Delta} - b_0 - b_1 V_{(t-1)\Delta}) \\ \sigma_{t\Delta}^{S*2} &= \frac{\beta_4 \Delta V_{(t-1)\Delta}}{\Delta V_{(t-1)\Delta} + \beta_4}.\end{aligned}$$

B Simulation exercise

I use a sample length of 2,000, which is in line with the sample size used for smoothed inference. The parameter values used in simulations are similar to the values obtained from the empirical application. For each model, I generate 100 sample paths using an Euler discretization of the process, using 20 sub-intervals per sampling interval which is one day: $\Delta = 1/252$ and $\Delta_{sim} = \Delta/20$ so that for instance for model I ,

$$\begin{aligned}y_{i\Delta_{sim}} &= b_0 + b_1 V_{(i-1)\Delta_{sim}} + \sqrt{V_{(i-1)\Delta_{sim}} \Delta_{sim}} \epsilon_i^S + z_i^S j_{i\Delta_{sim}} \\ V_{i\Delta_{sim}} &= c_0 + c_1 V_{(i-1)\Delta_{sim}} + \gamma_V \sqrt{V_{(i-1)\Delta_{sim}} \Delta_{sim}} \epsilon_i^V + z_i^V j_{i\Delta_{sim}}\end{aligned}$$

where ϵ_i^S and ϵ_i^V are standard normal variates with correlation ρ and $j_{i\Delta_{sim}}$ is Bernoulli with constant intensity $\lambda \Delta_{sim}$. Each simulated data is initialized with the volatility states at its unconditional mean. Initial 1,000 observations are generated and then discarded; the last of these observations is then taken as the starting point for the simulated series.

Tables A1 and A2 report mean, standard deviation and 90% high probability regions of the parameter posteriors. Diffusion and Q measure drift parameters are precisely estimated. The posterior distribution of P measure drift parameters contains large uncertainty about their values given that the sample period is only 8 years long. As for the jump parameters, given the realistic setup in which two jumps occurs in average per year, we see that they have low precision. This is due to the fact that in practice the effective sample size for jump parameters inference is about 20 observations only. Finally, we note an upward bias in the standard deviation parameters of the pricing error that reflects the uncertainty of the latent variables estimates.

C Convergence diagnostics

Figures A1–A4 report histograms, means and standard deviations for the first and last thirds of the posterior draws for the most parsimonious model (model *I*). We should expect to obtain the same marginal distribution for different subsamples from the posterior draws once the posterior simulator had converged. In this sense, the marginal distributions in the tables do not present evidence of failure of convergence. Similar results are obtained for the remaining specifications, and hence they are omitted. They are available upon request.

References

- Aït-Sahalia, Y. and B. Kimmel (2007). “Maximum likelihood estimation of stochastic volatility models”, *Journal of Financial Economics* 83, 413–452.
- Aït-Sahalia, Y. and D. Amengual (2008). “Market-based estimation of stochastic volatility models”, mimeo, Princeton University.
- Albert, J. and S. Chib (1993). “Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts”; *Journal of Business and Economic Statistics*, 11(1),1–15.
- Ang, A. and G. Bekaert (2002). “Regime switches in interest rates”, *Journal of Business and Economic Statistics* 20, 163–182.
- Bakshi, G. and N. Kapadia (2003). “Delta-hedged gains and the negative market volatility risk premium”, *Review of Financial Studies* 16, 527–566.
- Bates, D. (2000). “Post-’87 Crash fears in S&P 500 futures options”, *Journal of Econometrics* 94, 181–238.
- Black, F. (1976) “Studies in stock price volatility changes”, in *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, American Statistical Association, 177–181.
- Black, F. and M. Scholes (1973). “The pricing of options and corporate liabilities”, *Journal of Political Economy* 81, 637–654.
- Bondarenko, O. (2004). “Market price of variance risk and performance of hedge funds”, mimeo, University of Illinois at Chicago.
- Carr, P. and L. Wu (2008). “Variance risk premiums”, *Review of Financial Studies* , forthcoming.
- Chernov, M. A. R. Gallant, E. Ghysels, and G. Tauchen (2003). “Alternative models for stock price dynamics”, *Journal of Econometrics* 116, 225–257.
- Coval, J.D. and T. Shumway (2001). “Expected option returns”, *Journal of Finance* 56, 983–1009.

Dai, Q. and K. J. Singleton (2003). “Term structure modeling in theory and reality”, *Review of Financial Studies* 16, 631–678.

Duffie, D., J. Pan, and K. J. Singleton (2000). “Transform analysis and asset pricing for affine jump-diffusions”, *Econometrica* 68, 1343–1376.

Egloff, D. M. Leippold and L. Wu (2007). “Variance risk dynamics, variance risk premia and optimal variance swap investments”, mimeo, Baruch college.

Eraker, B. (2001). “MCMC analysis of diffusion models with application to finance”, *Journal of Business and Economic Statistics* 19, 177–191.

Eraker, B. (2004). “Do stock prices and volatility jump? Reconciling evidence from spot and option prices”, *Journal of Finance* 59, 1269–1300.

Eraker, B., M. S. Johannes, and N. G. Polson (2003). “The impact of jumps in returns and volatility”, *Journal of Finance* 58, 1269–1300.

Gelfand, A. E., and Dey, D. K. (1994). “Bayesian model choice: asymptotic and exact calculations”, *Journal of the Royal Statistical Society, Series B* 56, 501–514.

Geweke, J. (1999). “Using simulation methods for Bayesian econometric models: inference, development and communication”, *Econometric Reviews* 18, 1–126.

Geweke, J. (1999). “Simulation methods for model criticism and robustness analysis” in J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.), *Bayesian Statistics* 6, 275–299. Oxford University Press.

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*, John Wiley and Sons.

Geweke, J. and G. Amisano (2007). “Hierarchical Markov normal mixture models with applications to financial asset returns”, mimeo, University of Iowa.

Hamilton, J. D. (1988). “Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates”, *Journal of Economic Dynamics and Control* 12, 385–423.

Hamilton, J. D. (1989). “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica* 57(2), 357–384.

Heston, S. (1993). “Closed-form solution of options with stochastic volatility with application to bond and currency options”, *Review of Financial Studies* 6, 327–343.

Johannes, M. and N. Polson (2002). “MCMC methods for financial applications”, forthcoming *Handbook of Financial Econometrics*, edited by Yacine Aït-Sahalia and Lars Hansen (Elsevier).

Jones, C.S. (2003). “The dynamics of stochastic volatility: evidence from underlying and options markets”, *Journal of Econometrics* 116, 181–224.

Landén, C. (2000). “Bond pricing in a hidden Markov model of the short rate”, *Finance and Stochastics* 4, 371–389.

Pan, Jun (2002). “The jump-risk premia implicit in options: evidence from an integrated time-series study”, *Journal of Financial Economics* 63, 3–50.

Poteshman, A. (2001). “Underreaction, overreaction, and increasing misreaction to information in the options market”, *Journal of Finance* 56, 851–876.

Protter, P. (2005). *Stochastic integration and differential equations*, Second Edition, Springer-Verlag, Heidelberg.

Stein, J. (1989). “Overreactions in the options market”, *Journal of Finance* 44, 1011–1023.

Todorov, V (2007). “Variance risk premium dynamics”, mimeo, Northwestern University.

Yao, D., Q. Zhang and X. Y. Zhou (2003). “A regime-switching model for European option pricing”, mimeo, Columbia University.

Table 2. Priors

	Mean	Std.Dev.	IQR
κ_V^P, κ_V^Q	5.0	5.0	[1.4, 6.9]
η_V	0.0	7.1	[-3.5, 3.5]
$\kappa_\alpha^P, \kappa_\alpha^Q$	5.0	5.0	[1.4, 6.9]
η_α	0.0	7.1	[-3.5, 3.5]
γ_V, γ_α	0.4	0.9	[0.0, 0.4]
ρ	0.0	0.6	[-0.5, 0.5]
$\bar{\alpha}, \alpha_L, \alpha_H$	0.4	0.9	[0.0, 0.4]
p_{LH}, p_{HL}	1.0	1.0	[0.3, 1.4]
ϖ	0.0	1.0	[-0.4, 0.7]
β_1	0.05	0.03	[0.03, 0.06]
λ	12.0	8.2	[5.9, 16.2]

Notes: For each parameter I report its prior mean and standard deviation as well as its interquantile range. The models and parameterizations are given in Section 2: κ_i^j 's denote mean reversion speed parameter for process i under measure j ; γ_i 's denote volatility of volatility parameter for process i ; η_i 's denote market price of risk for process i (ϖ does the same for hidden Markov chain); ρ denotes the correlation parameter between Brownian processes driving returns and spot variance; $\bar{\alpha}$ denotes the long-run level of variance; finally p 's are the transition rates characterizing the hidden Markov chain dynamics (model *III*). $\beta_2, \beta_3, \beta_4, \rho_j$'s and s_j 's are omitted since they are standard conjugated priors.

Table 3. Summary statistics for S&P 500 returns and variance swap rates

Panel A. Sampling period: January 4, 1996 – December 31, 2003

Descriptive statistics for univariate time-series

	Mean	Std. Dev.	Skewness	Kurtosis
Variance swap rates				
1 month	0.058	0.031	1.709	6.584
2 months	0.059	0.033	2.180	10.434
3 months	0.059	0.030	2.059	9.866
6 months	0.061	0.030	2.055	10.570
1 year	0.064	0.028	1.458	7.217
2 years	0.067	0.028	1.034	4.832
S&P 500				
	0.000	0.013	-0.091	5.416

Panel B. Sampling period: January 2, 2004 – August 29, 2008

Descriptive statistics for univariate time-series

	Mean	Std. Dev.	Skewness	Kurtosis
Variance swap rates				
1 month	0.027	0.017	1.542	4.846
2 months	0.028	0.017	1.384	3.973
3 months	0.029	0.016	1.259	3.567
6 months	0.031	0.016	1.156	3.185
1 year	0.033	0.014	0.965	2.682
2 years	0.035	0.014	0.914	2.555
S&P 500				
	0.000	0.009	-0.192	5.088

Table 4. Principal component decomposition of variance swap rates

Panel A. Sampling period: January 4, 1996 – December 31, 2003						
	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>4th</i>	<i>5th</i>	<i>6th</i>
Eigenvalues	89.92%	9.05%	0.82%	0.10%	0.07%	0.04%
Eigenvectors	0.386	-0.554	0.639	-0.349	0.117	-0.001
	0.404	-0.400	-0.214	0.762	0.217	0.056
	0.400	-0.204	-0.394	-0.213	-0.772	0.044
	0.410	0.113	-0.464	-0.402	0.507	-0.431
	0.419	0.416	0.021	-0.104	0.170	0.782
	0.428	0.553	0.418	0.282	-0.241	-0.445
Panel B. Sampling period: January 2, 2004 – August 29, 2008						
	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>4th</i>	<i>5th</i>	<i>6th</i>
Eigenvalues	96.73%	2.89%	0.26%	0.06%	0.04%	0.02%
Eigenvectors	0.374	-0.539	0.685	-0.315	-0.027	-0.034
	0.386	-0.398	-0.179	0.766	0.131	-0.237
	0.394	-0.237	-0.454	-0.189	-0.066	0.736
	0.412	0.049	-0.414	-0.398	-0.378	-0.596
	0.428	0.380	0.014	-0.175	0.798	-0.072
	0.451	0.590	0.348	0.299	-0.445	0.200

Table 5. Estimates for drift and diffusion parameters

Parameter	Model <i>I</i>		Model <i>II</i>		Model <i>III</i>	
	Mean	90%HPD	Mean	90%HPD	Mean	90%HPD
κ_V^P	1.484	[0.636, 2.319]	4.641	[3.466, 5.819]	1.768	[1.293, 2.280]
η_V	-2.536	[-4.866, -0.159]	-3.577	[-5.994, -1.139]	-3.526	[-4.985, -2.248]
κ_V^Q	0.577	[0.499, 0.668]	2.939	[2.749, 3.080]	0.505	[0.385, 0.551]
$\bar{\alpha}^Q$	0.066	[0.061, 0.070]	0.111	[0.092, 0.131]		
γ_V	0.358	[0.348, 0.368]	0.476	[0.462, 0.490]	0.358	[0.347, 0.374]
ρ	-0.861	[-0.871, -0.851]	-0.854	[-0.865, -0.843]	-0.853	[-0.864, -0.841]
κ_α^P			0.155	[0.016, 0.357]		
η_α			-0.875	[-2.702, 0.371]		
κ_α^Q			0.057	[0.043, 0.076]		
γ_α			0.112	[0.106, 0.120]		
α_L					0.047	[0.043, 0.055]
α_H					0.248	[0.230, 0.277]
p_{LH}					0.472	[0.291, 0.570]
p_{HL}					3.838	[3.091, 4.244]
ϖ					-0.824	[-0.846, -0.804]

Notes: Sampling period: January 4, 1996 – December 31, 2003. For each parameter I report its posterior mean and the 90% high probability density interval. The models and parameterizations are given in Section 2: κ_i^j 's denote mean reversion speed parameter for process i under measure j ; γ_i 's denote volatility of volatility parameter for process i ; η_i 's denote market price of risk for process i (ϖ does the same for hidden Markov chain); ρ denotes the correlation parameter between Brownian processes driving returns and spot variance; $\bar{\alpha}$ denotes the long-run level of variance; finally p 's are the transition rates characterizing the hidden Markov chain dynamics (model *III*).

Table 6. Estimates for jump parameters

	Model <i>I</i>		Model <i>II</i>		Model <i>III</i>	
	Mean	90%HPD	Mean	90%HPD	Mean	90%HPD
β_1	28.031	[16.418, 41.136]	22.803	[13.562,33.678]	32.489	[20.221,46.897]
β_2	-0.011	[-0.059, 0.035]	-0.005	[-0.054, 0.045]	-0.007	[-0.049, 0.033]
β_3	-0.508	[-0.951,-0.070]	-0.497	[-0.928,-0.066]	-0.513	[-0.927,-0.089]
β_4	0.007	[0.005, 0.011]	0.007	[0.005, 0.011]	0.007	[0.004, 0.010]
λ	1.565	[0.805, 2.495]	1.584	[0.833, 2.488]	1.878	[0.989, 2.977]

Notes: Sampling period: January 4, 1996 – December 31, 2003. For each parameter I report its posterior mean and the 90% high probability density interval. The models and parameterizations are given in Section 2: β_1 is the exponential parameter of variance jump size; $\beta_2, \beta_3, \beta_4$ characterize jumps in returns while λ is the jump intensity of the Poisson process J .

Table 7. Estimates for pricing error parameters

Parameter	Model <i>I</i>		Model <i>II</i>		Model <i>III</i>	
	Mean	90%HPD	Mean	90%HPD	Mean	90%HPD
ρ_{1M}	0.684	[0.665, 0.704]	0.437	[0.416, 0.458]	0.586	[0.564, 0.609]
ρ_{2M}	0.322	[0.301, 0.344]	0.159	[0.141, 0.177]	0.272	[0.251, 0.293]
ρ_{3M}	0.004	[-0.005, 0.014]	0.006	[-0.001, 0.014]	0.011	[0.001, 0.021]
ρ_{6M}	0.537	[0.516, 0.558]	0.059	[0.048, 0.070]	0.325	[0.304, 0.346]
ρ_{1Y}	0.798	[0.781, 0.816]	0.009	[0.002, 0.015]	0.585	[0.564, 0.607]
ρ_{2Y}	0.872	[0.857, 0.886]	0.072	[0.059, 0.086]	0.715	[0.692, 0.737]
s_{1M}	0.089	[0.087, 0.091]	0.087	[0.084, 0.089]	0.085	[0.082, 0.087]
s_{2M}	0.062	[0.060, 0.064]	0.055	[0.054, 0.057]	0.059	[0.057, 0.061]
s_{3M}	0.022	[0.020, 0.023]	0.022	[0.021, 0.023]	0.023	[0.022, 0.025]
s_{6M}	0.069	[0.067, 0.071]	0.033	[0.032, 0.034]	0.056	[0.055, 0.058]
s_{1Y}	0.083	[0.081, 0.086]	0.020	[0.019, 0.021]	0.076	[0.074, 0.078]
s_{2Y}	0.086	[0.083, 0.088]	0.036	[0.034, 0.037]	0.083	[0.081, 0.085]

Notes: Sampling period: January 4, 1996 – December 31, 2003. For each parameter I report its posterior mean and the 90% high probability density interval. The models and parameterizations are given in Section 2: s_j denotes standard deviation of pricing error of maturity j while ρ_j is the corresponding autocorrelation coefficient.

Table 8. Model comparison results

Panel A. Log marginal data densities			
	Model <i>I</i>	Model <i>II</i>	Model <i>III</i>
Mean	40,337	56,104	42,216
Std.Dev.	91.42	81.48	89.13

Panel B. Predictive Bayes factor		
	Model <i>II</i> vs. Model <i>I</i>	Model <i>III</i> vs. Model <i>I</i>
Geometric mean	3.08	-0.01

Notes: Sampling period: Log marginal data densities are computed using the output of the posterior simulator over the period January 4, 1996 – December 31, 2003. To compute the predictive Bayes factor I uses the output from the practical filter over the period January 2, 2004 – August 29, 2008.

Table 9. Posterior distribution of variance swap rates loadingsPanel A. Loadings for model *II*

Maturity	$V_t(\phi^I)$		$\alpha_t(\phi^{II})$		$\bar{\alpha}^Q(1 - \phi^I - \phi^{II})$	
	Mean	90%HPD	Mean	90%HPD	Mean	90%HPD
1 month	0.887	[0.882, 0.894]	0.113	[0.106, 0.118]	0.000	[0.000, 0.000]
2 months	0.791	[0.782, 0.802]	0.209	[0.197, 0.217]	0.001	[0.001, 0.001]
3 months	0.708	[0.697, 0.723]	0.290	[0.275, 0.302]	0.002	[0.001, 0.002]
6 months	0.524	[0.510, 0.544]	0.471	[0.450, 0.486]	0.005	[0.004, 0.006]
1 year	0.322	[0.310, 0.341]	0.663	[0.640, 0.678]	0.015	[0.012, 0.020]
2 years	0.170	[0.162, 0.181]	0.791	[0.768, 0.807]	0.040	[0.031, 0.051]

Panel B. Loadings for model *III*

Maturity	$V_t(\phi^I)$		$\alpha_t(\phi^{III})$		$\bar{\alpha}^Q(1 - \phi^I - \phi^{III})$	
	Mean	90%HPD	Mean	90%HPD	Mean	90%HPD
1 month	0.979	[0.977, 0.984]	0.020	[0.015, 0.022]	0.001	[0.001, 0.001]
2 months	0.959	[0.956, 0.969]	0.038	[0.029, 0.041]	0.003	[0.002, 0.004]
3 months	0.940	[0.934, 0.953]	0.053	[0.042, 0.058]	0.007	[0.005, 0.008]
6 months	0.884	[0.874, 0.910]	0.091	[0.073, 0.099]	0.025	[0.017, 0.029]
1 year	0.786	[0.769, 0.830]	0.133	[0.113, 0.146]	0.0813	[0.057, 0.091]
2 years	0.630	[0.606, 0.697]	0.151	[0.135, 0.167]	0.219	[0.164, 0.239]

Notes: Sampling period: January 4, 1996 – December 31, 2003. For each parameter I report its posterior mean and the 90% high probability density interval. The models and parameterizations are given in Section 2.

Table A1. Monte Carlo results for model *II*

	True value	Posterior mean	Posterior std.dev	90% credible interval
Theoretical model parameters				
a	0.000	-0.048	0.389	[-0.694, 0.582]
b	1.000	0.582	3.121	[-4.410, 5.853]
κ_V^P	5.000	4.911	1.062	[3.283, 6.708]
κ_V^Q	3.000	2.929	0.149	[2.707, 3.193]
κ_α^P	2.000	2.491	1.194	[0.723, 4.500]
κ_α^Q	1.000	1.039	0.113	[0.810, 1.207]
α^Q	0.100	0.100	0.001	[0.099, 0.101]
γ_V	0.300	0.297	0.009	[0.285, 0.313]
γ_α	0.100	0.116	0.018	[0.088, 0.149]
ρ	-0.500	-0.484	0.027	[-0.524,-0.439]
β_1	20.000	18.093	4.190	[11.869,25.462]
β_2	-0.100	-0.100	0.023	[-0.139,-0.063]
β_3	-0.500	-0.519	0.215	[-0.871,-0.167]
β_4	0.003	0.006	0.001	[0.004, 0.009]
λ	4.000	4.079	1.076	[2.479, 6.000]
Pricing error parameters				
ρ_{1M}	0	0.001	0.003	[-0.004, 0.005]
ρ_{2M}	0	0.000	0.003	[-0.004, 0.005]
ρ_{3M}	0	0.000	0.003	[-0.004, 0.004]
ρ_{6M}	0	0.000	0.003	[-0.004, 0.005]
ρ_{1Y}	0	0.001	0.003	[-0.004, 0.005]
ρ_{2Y}	0	0.001	0.004	[-0.003, 0.006]
s_{1M}	0.002	0.009	0.000	[0.009, 0.010]
s_{2M}	0.002	0.008	0.000	[0.008, 0.009]
s_{3M}	0.002	0.008	0.000	[0.008, 0.009]
s_{6M}	0.002	0.008	0.000	[0.008, 0.009]
s_{1Y}	0.002	0.009	0.000	[0.008, 0.009]
s_{2Y}	0.002	0.009	0.001	[0.008, 0.009]

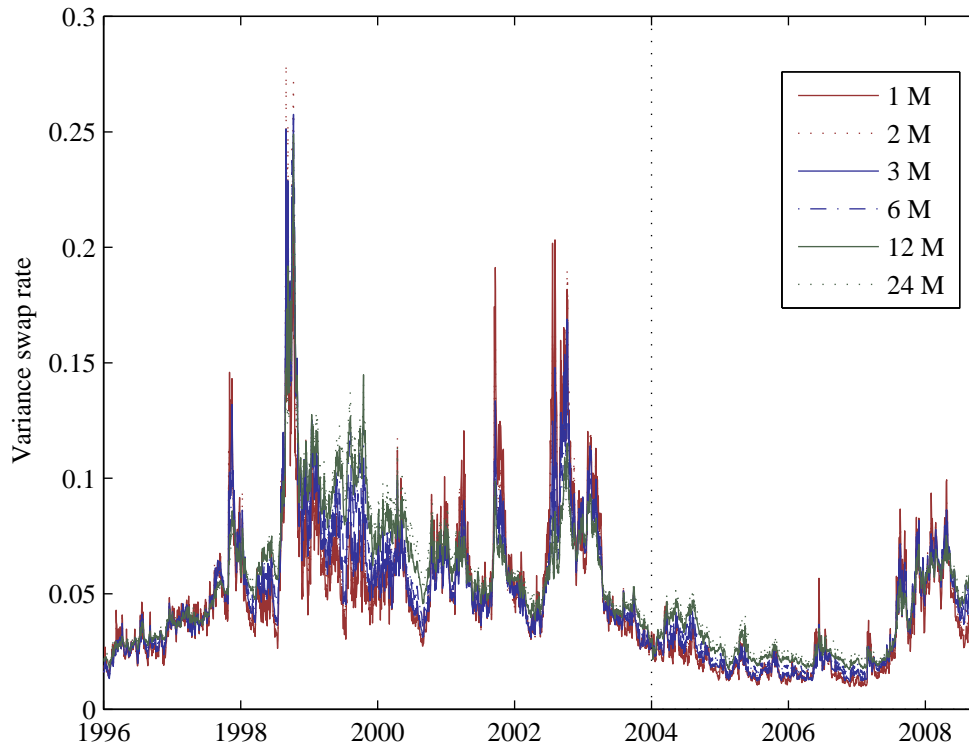
Notes: Parameter definitions are given in Section 2. a and b is a short-hand notation for the drift coefficients of y , that is $\mu_y(v) = a + bv$.

Table A2. Monte Carlo results for model *III*

	True value	Posterior mean	Posterior std.dev	90% credible interval
Theoretical model parameters				
a	0.000	-0.066	0.460	[-0.818, 0.673]
b	1.000	0.871	2.858	[-3.747, 5.738]
κ_V^P	5.000	4.899	0.916	[3.463, 6.456]
κ_V^Q	3.000	3.000	0.005	[2.991, 3.009]
α_L^Q	0.100	0.100	0.000	[0.100, 0.101]
α_H^Q	0.200	0.200	0.000	[0.200, 0.201]
p_{LH}	1.000	1.012	0.095	[0.861, 1.170]
p_{HL}	1.000	1.011	0.095	[0.861, 1.169]
ϖ	-0.500	-0.501	0.047	[-0.572,-0.420]
γ_V	0.300	0.293	0.007	[0.283, 0.304]
ρ	-0.500	-0.498	0.022	[-0.533,-0.460]
β_1	20.000	17.835	4.154	[11.622,25.193]
β_2	-0.100	-0.096	0.024	[-0.136,-0.057]
β_3	-0.500	-0.541	0.223	[-0.907,-0.172]
β_4	0.003	0.006	0.002	[0.004, 0.009]
λ	4.000	4.036	1.090	[2.435, 6.004]
Pricing error parameters				
ρ_{1M}	0	0.000	0.003	[-0.004, 0.004]
ρ_{2M}	0	-0.000	0.003	[-0.004, 0.004]
ρ_{3M}	0	-0.000	0.003	[-0.004, 0.004]
ρ_{6M}	0	0.000	0.002	[-0.004, 0.004]
ρ_{1Y}	0	0.000	0.002	[-0.004, 0.004]
ρ_{2Y}	0	0.000	0.002	[-0.004, 0.004]
s_{1M}	0.002	0.009	0.000	[0.008, 0.009]
s_{2M}	0.002	0.008	0.000	[0.008, 0.009]
s_{3M}	0.002	0.008	0.000	[0.008, 0.008]
s_{6M}	0.002	0.008	0.000	[0.008, 0.008]
s_{1Y}	0.002	0.008	0.000	[0.007, 0.008]
s_{2Y}	0.002	0.007	0.000	[0.007, 0.008]

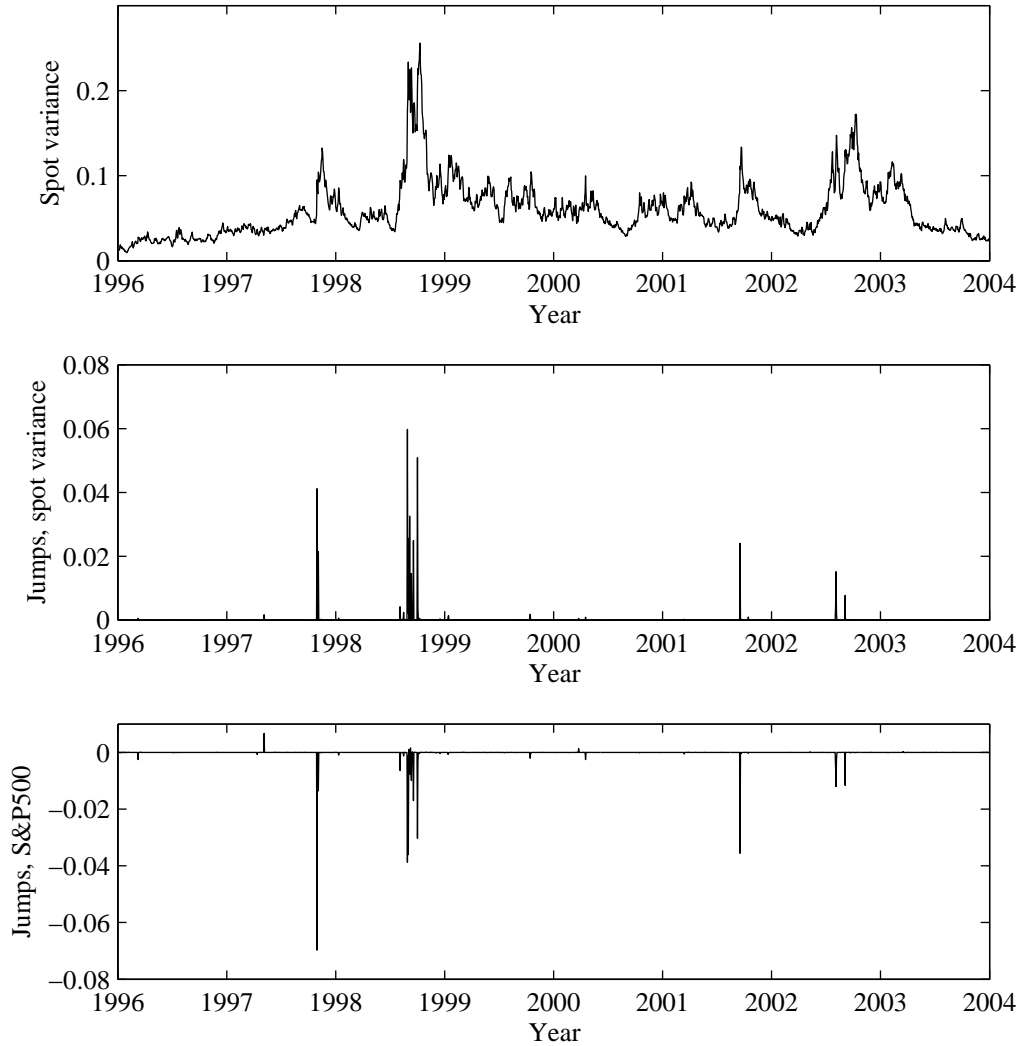
Notes: Parameter definitions are given in Section 2. a and b is a short-hand notation for the drift coefficients of y , that is $\mu_y(v) = a + bv$.

Figure 2. Variance swap rates



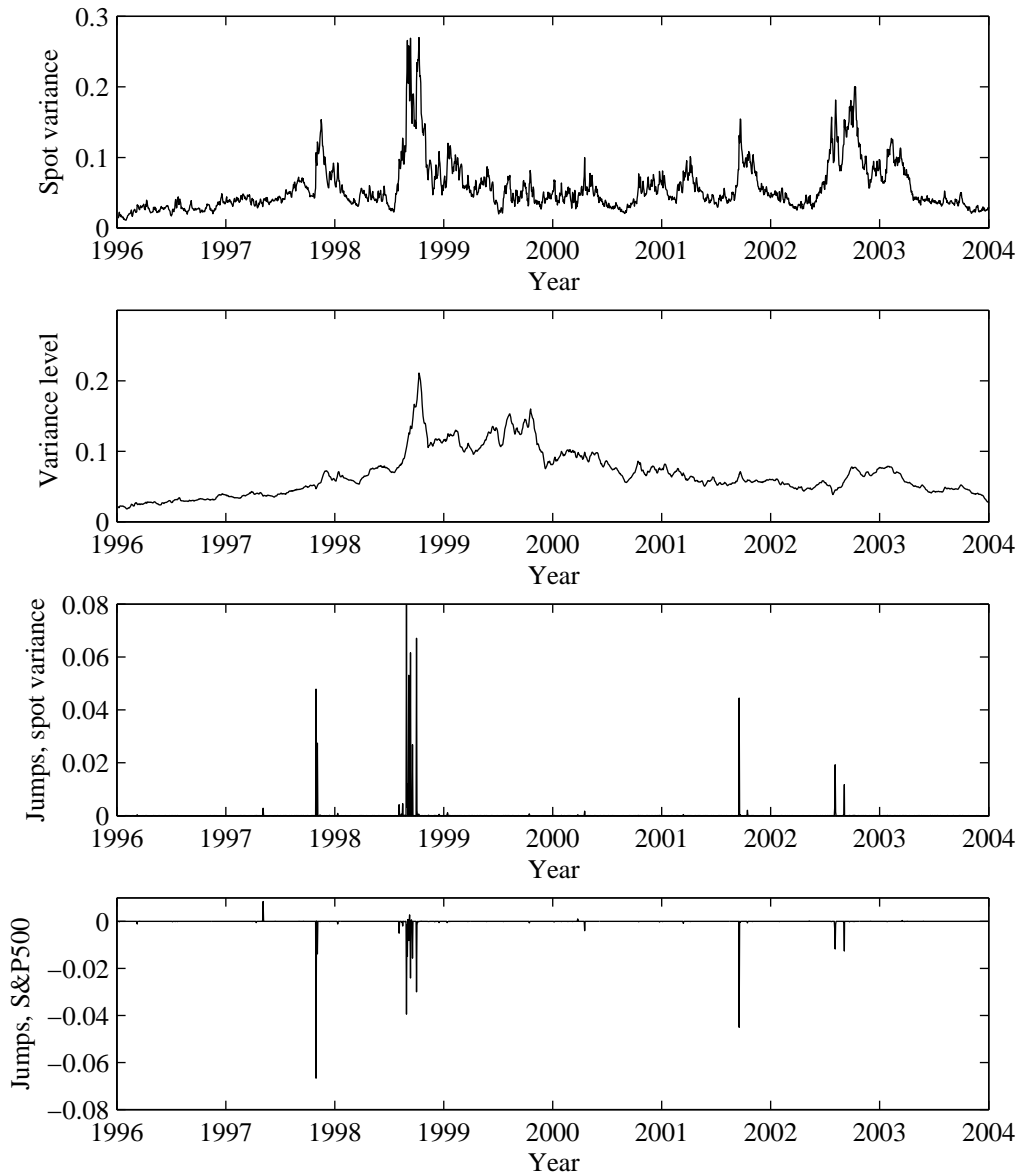
Notes: The vertical line indicates the sampling period used for smoothed inference, January 4, 1996 – December 31, 2003, and the corresponding to sequential inference, January 2, 2004 – August 29, 2008.

Figure 3. Posterior mean of latent variables, Model *I*
Constant level of volatility



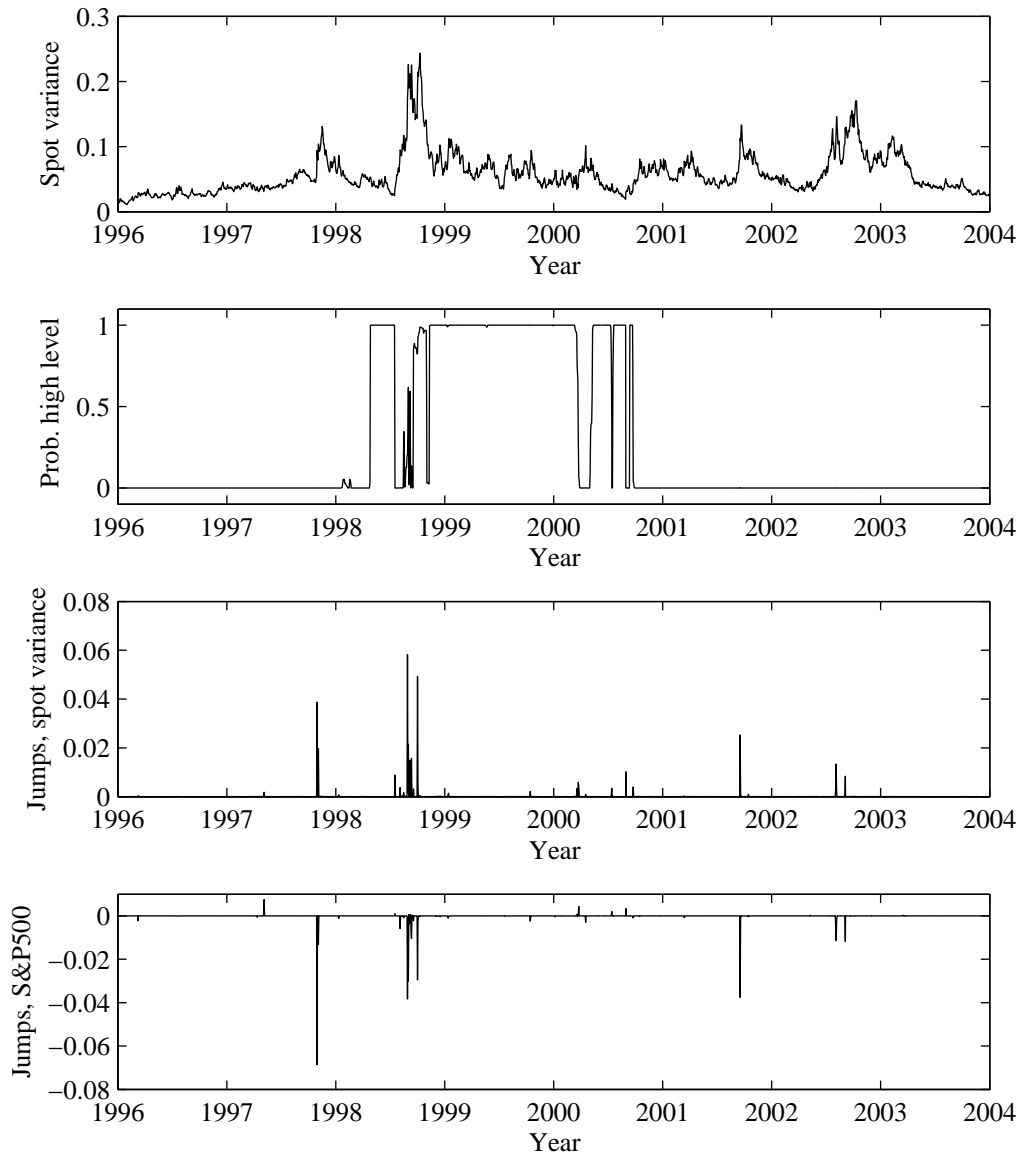
Notes: Sampling period: January 4, 1996 – December 31, 2003. Labels: “Spot variance” stands for V , “Jumps, spot variance” stands for Z^V and “Jumps, S&P 500” stands for Z^S . Details about the parameterization are given in Section 2.

Figure 4. Posterior mean of latent variables, Model *II*
 Square-root process level of volatility



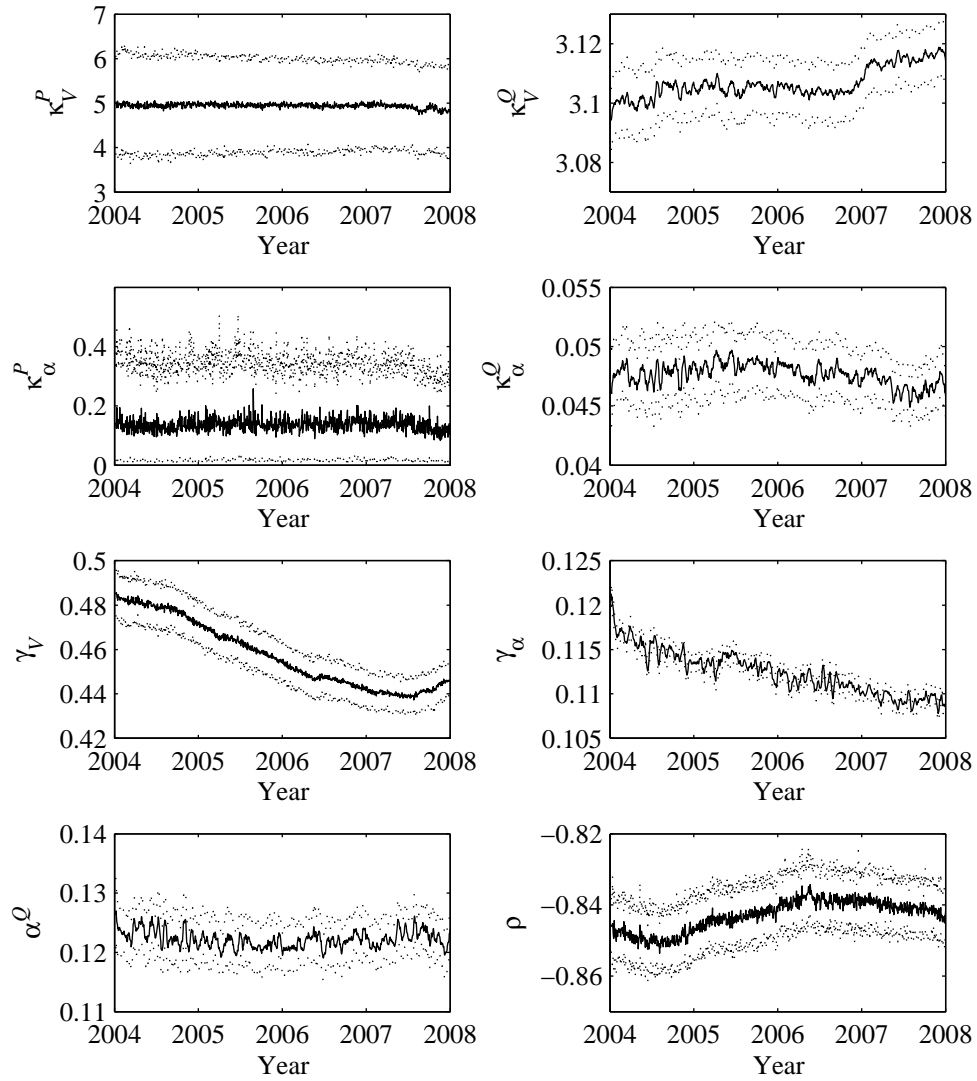
Notes: Sampling period: January 4, 1996 – December 31, 2003. Labels: “Spot variance” stands for V , “Variance level” stands for α , “Jumps, spot variance” stands for Z^V and “Jumps, S&P 500” stands for Z^S . Details about the parameterization are given in Section 2.

Figure 5. Posterior mean of latent variables, Model III
Regime-switching level of volatility



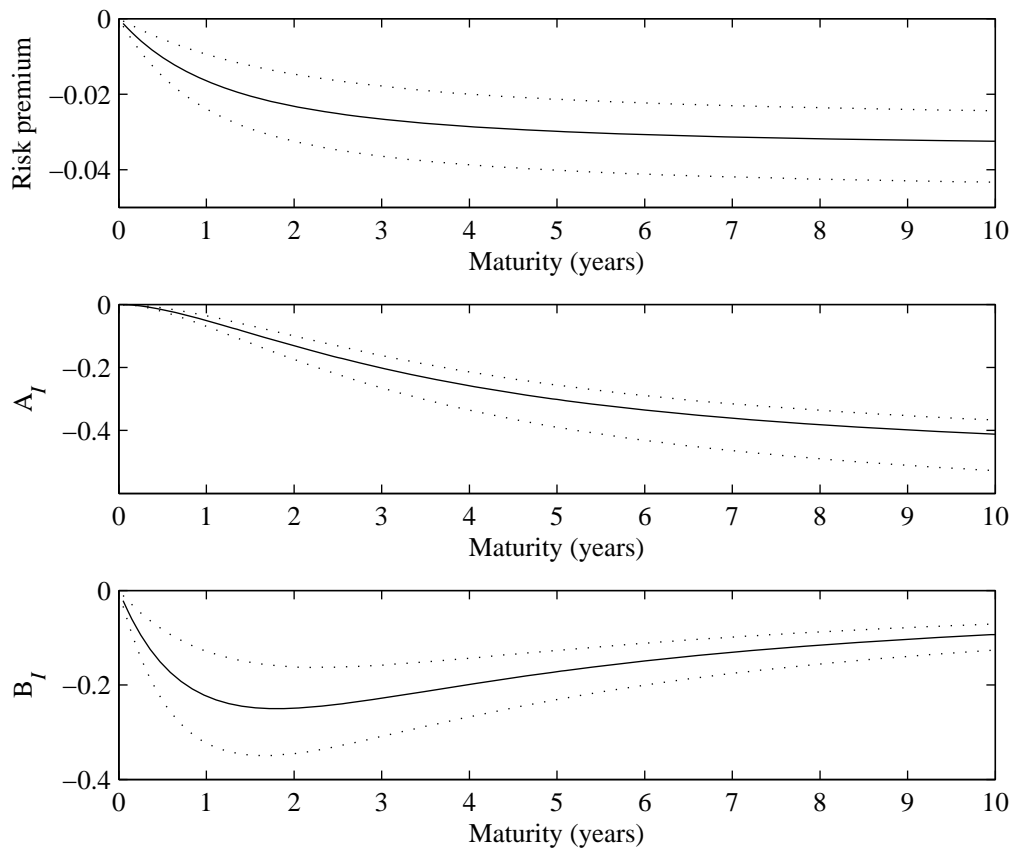
Notes: Sampling period: January 4, 1996 – December 31, 2003. Labels: “Spot variance” stands for V , “Prob high level” denotes the probability of the state $\alpha = \alpha_H$, “Jumps, spot variance” stands for Z^V and “Jumps, S&P 500” stands for Z^S . Details about the parameterization are given in Section 2.

Figure 6. Filtered posterior distribution for parameters, Model II
Square-root process level of volatility



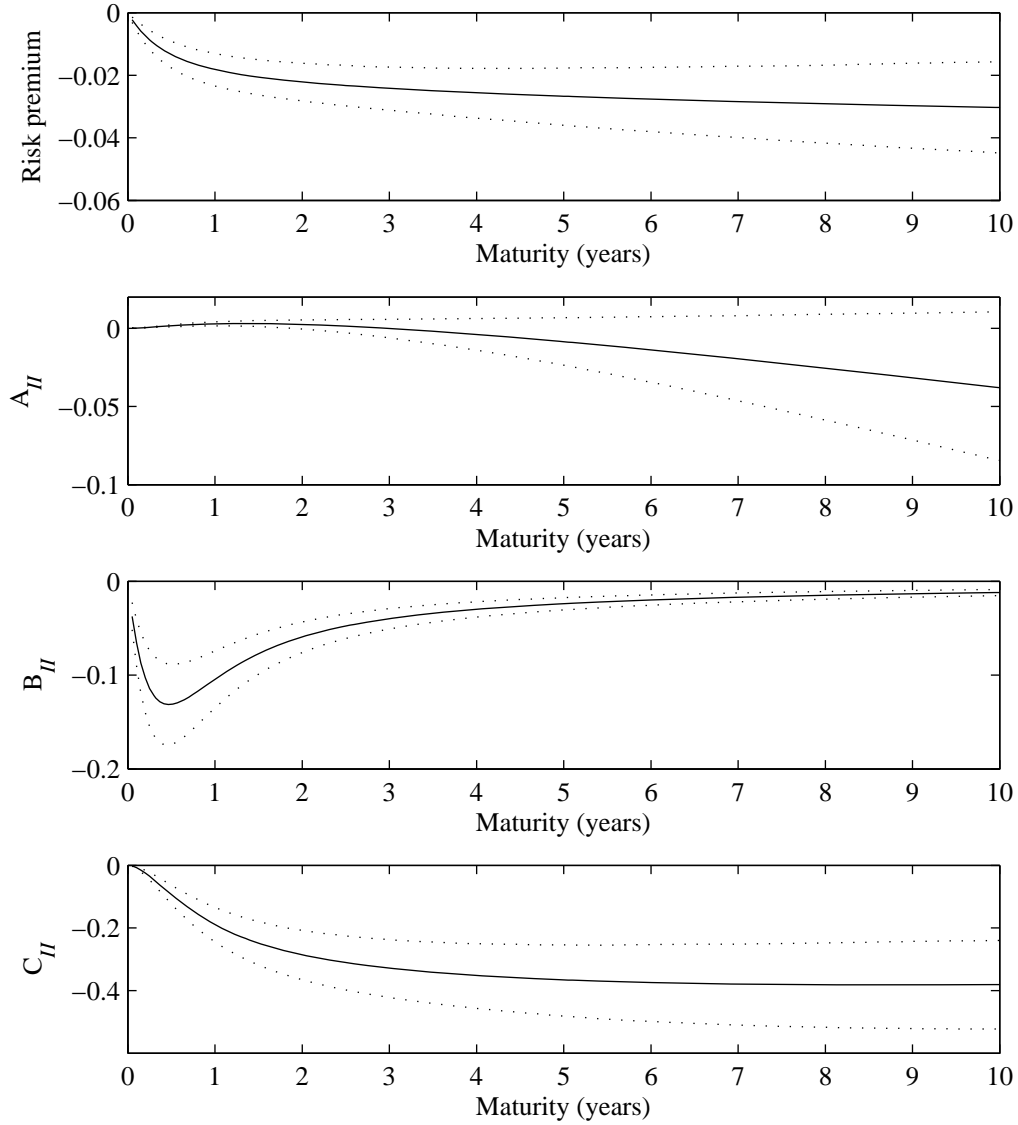
Notes: Solid lines denote the posterior mean while dotted lines represent the 90% probability intervals. κ_i^j 's denote mean reversion speed parameter for process i under measure j ; γ_i 's denote volatility of volatility parameter for process i ; η_i 's denote market price of risk for process i ; ρ denotes the correlation parameter between Brownian processes driving returns and spot variance.

Figure 7. Term structure of variance risk premia, Model *I*
Constant level of volatility



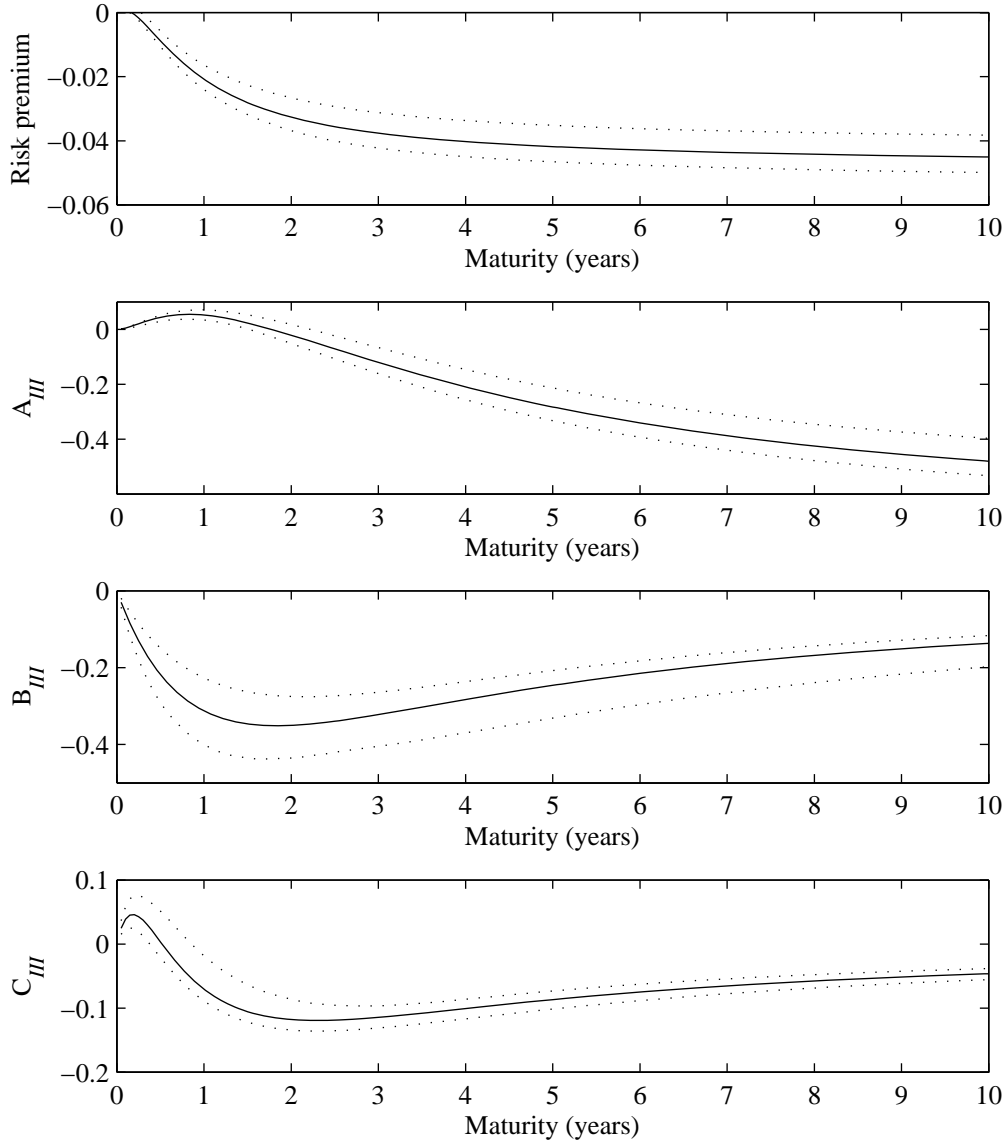
Notes: Solid lines denote the posterior mean while dotted lines represent the 66% probability intervals. Risk premium in the first plot denotes the difference of unconditional expectations of integrated variance under the pricing and objective measures. The subsequent plots show the posterior distribution of the loadings entering in the risk premia: $A_I(\tau)$ is the loading for $\bar{\alpha}^Q$ and $B_I(\tau)$ is the loading for V_t .

Figure 8. Term structure of variance risk premia, Model *II*
 Square-root process level of volatility



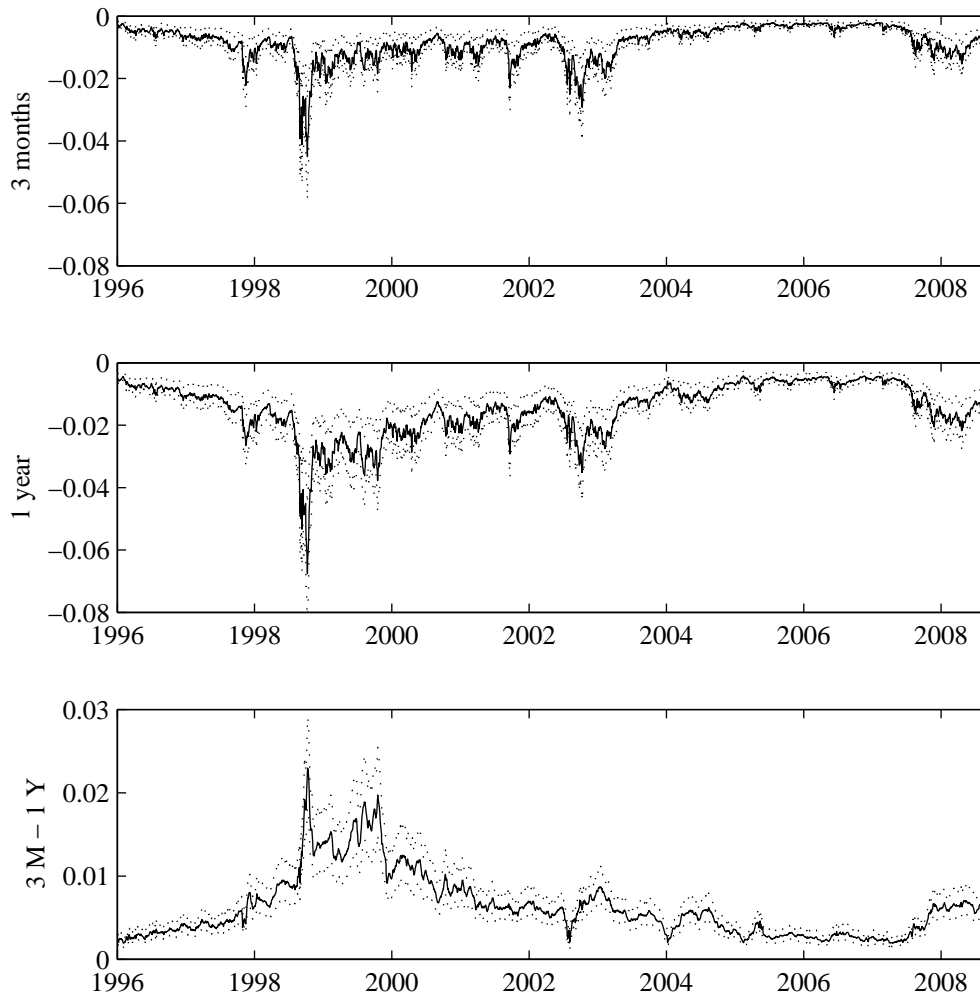
Notes: Solid lines denote the posterior mean while dotted lines represent the 66% probability intervals. Risk premium in the first plot denotes the difference of unconditional expectations of integrated variance under the pricing and objective measures. The subsequent plots show the posterior distribution of the loadings entering in the risk premia: $A_{II}(\tau)$ is the loading for $\bar{\alpha}^Q$, $B_{II}(\tau)$ is the loading for V_t and $C_{II}(\tau)$ is the loading for α_t .

Figure 9. Term structure of variance risk premia, Model *III*
Regime-switching level of volatility



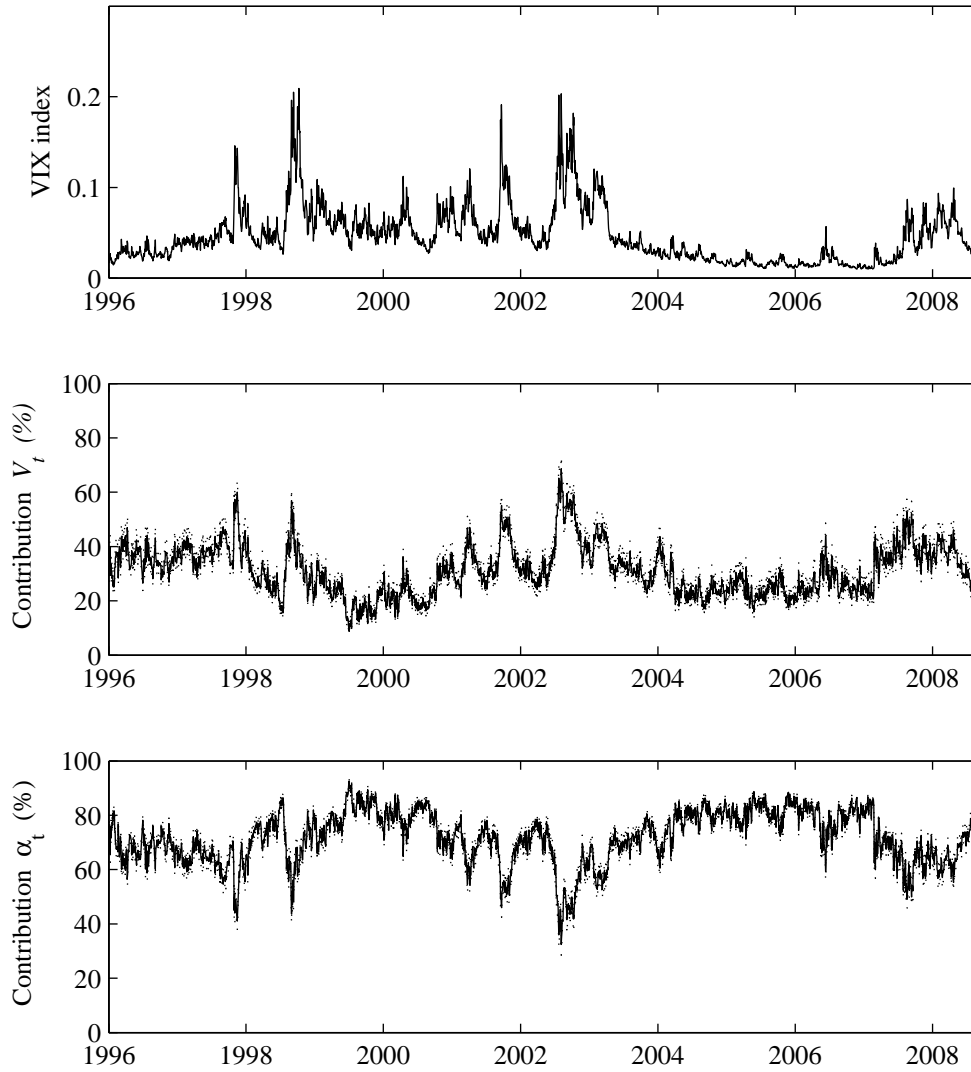
Notes: Solid lines denote the posterior mean while dotted lines represent the 66% probability intervals. Risk premium in the first plot denotes the difference of unconditional expectations of integrated variance under the pricing and objective measures. The subsequent plots show the posterior distribution of the loadings entering in the risk premia: $A_{III}(\tau)$ is the loading for $\bar{\alpha}^Q$, $B_{III}(\tau)$ is the loading for V_t and $C_{III}(\tau)$ is the loading for α_t .

Figure 10. Variance swap risk premia, Model *II*
Square-root process level of volatility



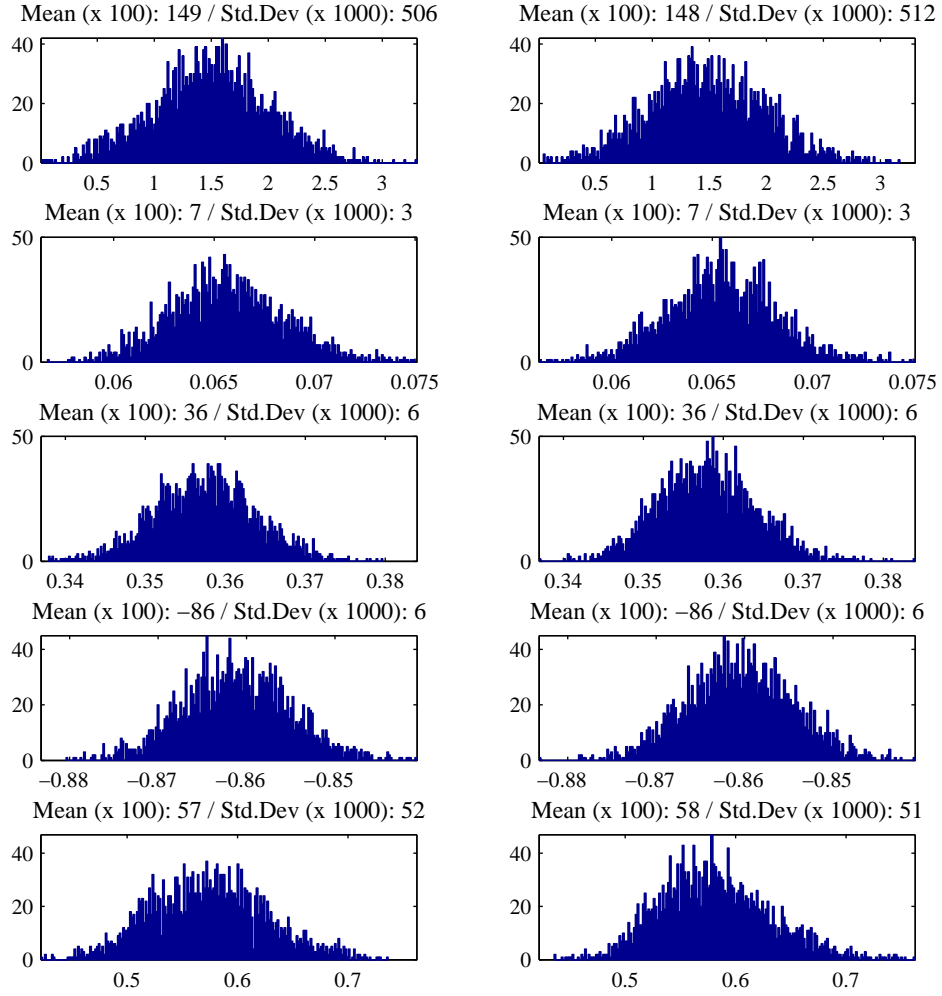
Notes: Solid lines denote the posterior mean while dotted lines represent the 66% probability intervals.

Figure 11. Decomposition of variance risk premia for 1-year variance swap rate
Model *II*, Square-root process level of volatility



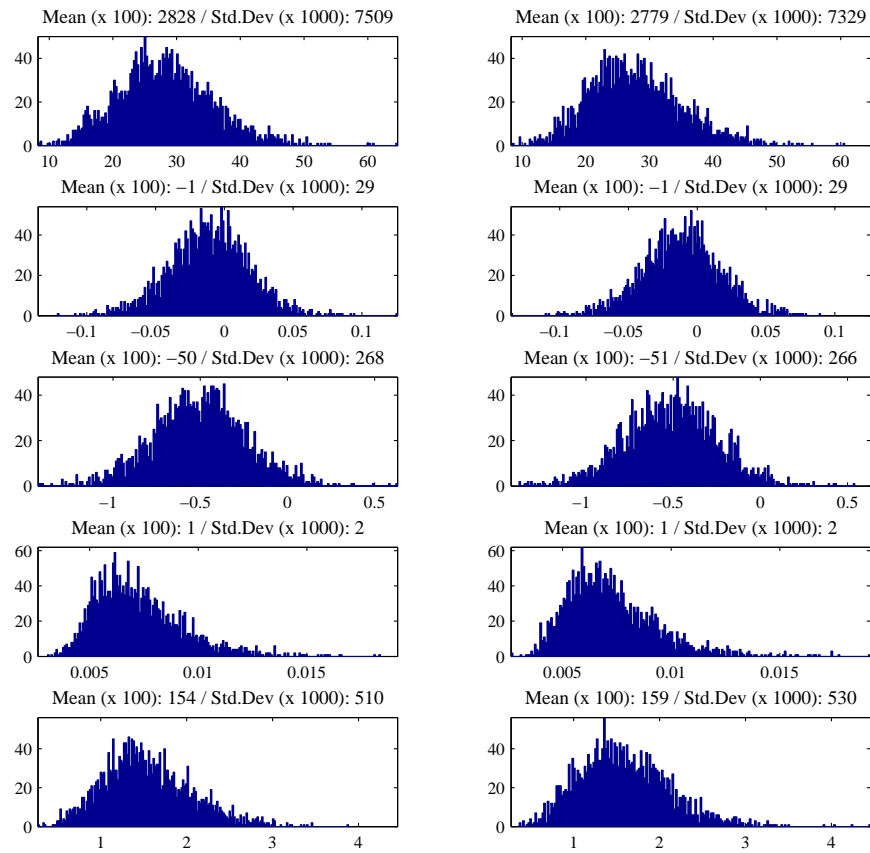
Notes: Solid lines denote the posterior mean while dotted lines represent the 66% probability intervals.

Figure A.1. Convergence diagnostics for model *I*. Histograms



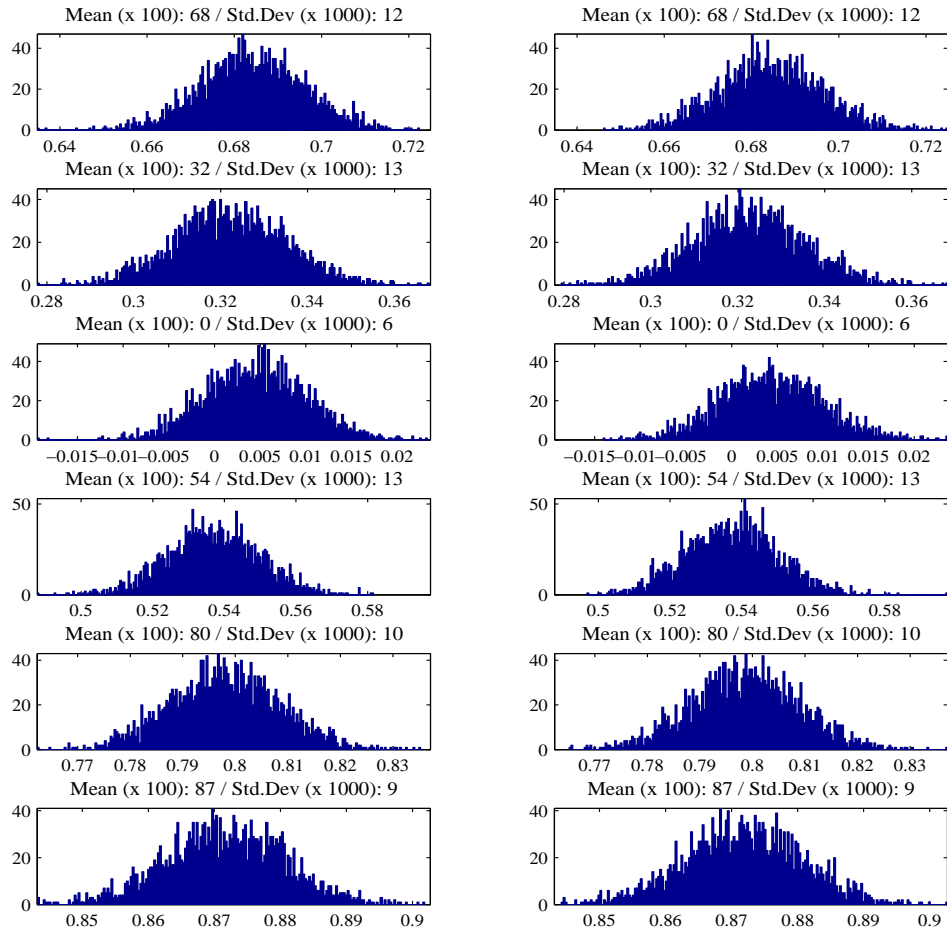
Notes: Histograms of the first third (left) and last third (right) of the Markov chain. First row is κ_V^P , second row is $\bar{\alpha}^Q$, third row is γ_V , fourth row is ρ and fifth row is κ_V^Q .

Figure A.2. Convergence diagnostics for model *I*. Histograms



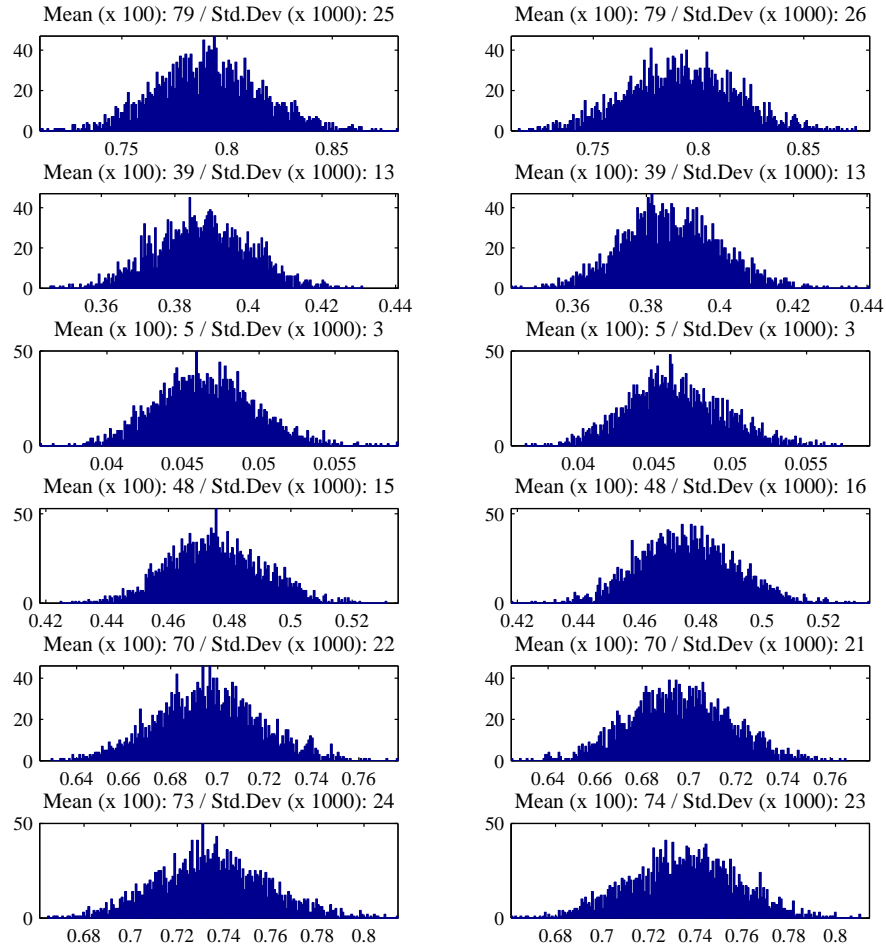
Notes: Histograms of the first third (left) and last third (right) of the Markov chain. First row is β_1 , second row is β_2 , third row is β_3 , fourth row is β_4 and fifth row is λ .

Figure A.3. Convergence diagnostics for model *I*. Histograms



Notes: Histograms of the first third (left) and last third (right) of the Markov chain. Pricing errors autocorrelation coefficients: each row correspond to a different maturity: first row is one month, second row is two months, third row is three months, fourth row is six months, fifth row is one year and sixth row is two years.

Figure A.4. Convergence diagnostics for model *I*. Histograms



Notes: Histograms of the first third (left) and last third (right) of the Markov chain. Pricing errors standard deviation coefficients ($\times 100$): each row correspond to a different maturity: first row is one month, second row is two months, third row is three months, fourth row is six months, fifth row is one year and sixth row is two years.