2016/42

A New Approach to Volatility Modeling : The High-Dimensional Markov Model

MACIEJ AUGUSTYNIAK, LUC BAUWENS AND ARNAUD DUFAYS

50 YEARS OF
CORE
DISCUSSION PAPERS

CORE

Voie du Roman Pays 34, L1.03.01

Tel (32 10) 47 43 04

Fax (32 10) 47 43 01

Email: immaq-library@uclouvain.be

<http://www.uclouvain.be/en-44508.html>

A new approach to volatility modeling: the High-Dimensional Markov model

Maciej Augustyniak^{a,b}, Luc Bauwens^{c,d} and Arnaud Dufays^{*,e,f}

^a*Département de mathématiques et de statistique, Université de Montréal, PO Box 6128, Station Centre-ville, Montreal, Quebec, Canada H3C 3J7*

^b*Quantact Actuarial and Financial Mathematics Laboratory, Centre de recherches mathématiques, Université de Montréal, PO Box 6128, Station Centre-ville, Montreal, Quebec, Canada H3C 3J7*

^c*Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve, Belgium,*

^d*Université Côte d’Azur – SKEMA, France*

^e*Département d’économie, Université Laval, 1025 avenue des Sciences-Humaines, Quebec City, Quebec, Canada, G1V 0A6*

^f*CRREP research associate*

Version: December 13, 2016

Abstract

A new model – the high-dimensional Markov (HDM) model – is proposed for financial returns and their latent variances. It is also applicable to model directly realized variances. Volatility is modeled as a product of three components: a Markov chain driving volatility persistence, an independent discrete process capable of generating jumps in the volatility, and a predictable (data-driven) process capturing the leverage effect. The Markov chain and jump components allow volatility to switch abruptly between thousands of states. The transition probability matrix of the Markov chain is structured in such a way that the multiplicity of the second largest eigenvalue can be greater than one. This distinctive feature generates a high degree of volatility persistence. The statistical properties of the HDM model are derived and an economic interpretation is attached to each component. In-sample results on six financial time series highlight that the HDM model compares favorably to the main

Maciej Augustyniak gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Luc Bauwens acknowledges support of the “Communauté française de Belgique” through contract “Projet d’Actions de Recherche Concertées 12/17-045”, granted by the “Académie universitaire Louvain”.

Arnaud Dufays gratefully acknowledges financial support from the Fonds de Recherche du Québec - Société et Culture.

* Corresponding author.

Email addresses: `augusty@dms.umontreal.ca` (Maciej Augustyniak), `luc.bauwens@uclouvain.be` (Luc Bauwens), `arnaud.dufays@ecn.ulaval.ca` (Arnaud Dufays).

existing volatility processes. A forecasting experiment shows that the HDM model significantly outperforms its competitors when predicting volatility over time horizons longer than five days.

JEL classification: C22, C51, C58

Keywords: Volatility, Markov-switching, Persistence, Leverage effect.

1 Introduction

Building on the seminal contribution of Goldfeld and Quandt (1973), Hamilton (1989) has popularized the use of regime-switching models in economics and finance. These models allow us to model sharp changes in the dynamics of economic or financial time series by introducing a finite-valued latent stochastic process that governs the evolution of the parameters of the time series model. In most applications, this latent process is a Markov chain, and, consequently, Markov-switching and hidden Markov models are sometimes used interchangeably with regime-switching models. In the past twenty-five years, the emphasis in the literature has been on models with a relatively low number of states – between two and four (e.g., Ang and Bekaert, 2002; Bauwens et al., 2014; Dai et al., 2007). On one hand, this choice is motivated by parsimony because the number of parameters in the transition matrix of the Markov chain increases quadratically with the number of states. On the other hand, it is generally easier to attach an economic interpretation to a low-dimensional state space (e.g., a Markov chain with two states can be used to represent bull and bear market regimes).

Rydén et al. (1998) showed that hidden Markov models can reproduce reasonably well most of the stylized facts of financial return series. However, they also argue that the model seems to be “doomed from the start” for replicating the high degree of persistence in volatility that is empirically observed. This is because, similarly to traditional stationary autoregressive moving-average models, regime-switching models based on a Markovian switching process have a short memory, that is, they can only generate an autocorrelation function that eventually decays exponentially. However, at finite lags the decay in this autocorrelation function can still potentially be quite slow. For instance, much research has shown that a time series generated with a short memory process contaminated by occasional breaks can exhibit statistical properties that are akin to those that would be obtained from a genuine long memory process (e.g., Diebold and Inoue, 2001; Granger and Hyung, 2004; Mikosch and Starica, 2004; Perron and Qu, 2010; Starica and Granger, 2005). This

observation explains why many studies in financial econometrics consider models in which a low-dimensional regime-switching process is used as a way to govern time-variation in the parameters of an existing econometric model. An example of such a combination is the regime-switching generalized autoregressive conditional heteroskedasticity (GARCH) model (Dueker, 1997; Gray, 1996; Haas et al., 2004; Klaassen, 2002).

An alternative to these types of models is to consider regime-switching processes with a finite, but high-dimensional, state space, such as the Markov switching multifractal (MSM) model proposed by Calvet and Fisher (2004). These authors demonstrate that this process has the ability to generate a high degree of volatility persistence and show that it outperforms GARCH, fractionally integrated GARCH, as well as regime-switching GARCH models, when modeling exchange rate volatility. Although these empirical results offer a motivation for considering pure regime-switching specifications with a large number of states, very few models of this type have since been proposed in the literature.

Building on the MSM approach, the objective of this article is to propose a new parsimonious regime-switching volatility model with a high-dimensional finite state space: the high-dimensional Markov (HDM) model. The volatility dynamics in this model originate from the product of three components: a high-dimensional finite state space Markov chain driving volatility persistence, a jump process capable of generating non-persistent changes in volatility, and a data-driven component capturing a leverage effect. The structure of the Markov chain component shares some similarities with the structure of the MSM model, because it is constructed by multiplying a large number of independent two-state Markov chains. However, the specific formulation that we adopt leads to four important differences. First, all of our two-state Markov chains are not constrained to take identical values as in the MSM model. As a consequence, the support of the volatility distribution in the HDM model comprises thousands of points, whereas the MSM models implemented by Calvet and Fisher (2004) only allowed the volatility process to switch between at most eleven different values. Second, the transition probability matrix of our Markov chain component is structured in

such a way that the multiplicity of the second largest eigenvalue can be greater than one. This distinctive characteristic enables us to generate a high degree of volatility persistence, which translates into a very slow decay of the autocorrelation function at finite lags. A further novelty of our approach versus the MSM model is that we allow for non-persistent jumps and integrate a leverage effect. As a final advantage, the HDM model is specified in such a way that only one estimation with a high number of chains is needed while several model estimations are required to select the number of chains of the MSM process.

We perform an empirical analysis of fit and forecasting performance on return and realized volatility data from the S&P 500 index, the NASDAQ Composite index, and the USD/EUR exchange rate over the period 2000–2016. When modeling returns, we find that in terms of information criteria, the fit of the HDM model is superior to that of the MSM model, and comparable to those of GARCH and regime-switching GARCH models. When modeling realized variances, the HDM model is considerably superior to heterogeneous autoregressions (HAR) (Corsi, 2009) and multiplicative error models (MEM) (Engle, 2002). The forecasting comparison reveals that at short-term horizons (five days or less), the root mean squared forecast errors generated by the HDM models are as good as those of popular volatility processes, and that at longer-term horizons (up to one hundred), they are considerably smaller than those of its competitors. Hence, the HDM models are much superior at long forecasting horizons, due to their superior ability to capture the high degree of volatility persistence of squared returns and realized variances.

The paper is structured as follows. Section 2 introduces the HDM model and exposes its statistical properties. Section 3 presents the results of the in-sample fit and out-of-sample forecasting performance for different time series of financial returns and realized variances. Section 4 concludes.

2 Model definition and properties

The HDM model is designed to fit a time series of financial returns, taking into account their time-varying volatility. It is also suitable to model directly a series of realized variances. Its central component is a discrete-time positive latent stochastic process, denoted by $\{V_t\}$. This process corresponds to the latent variance of returns in the first case, and to the expected value of the realized variance in the second case. Before defining this component in detail, we introduce the modeling framework that enables us to link it either to financial returns or to realized variances.

2.1 Basic model framework

2.1.1 Returns

Let r_t , $t = 1, \dots, T$, denoted by $\{r_t\}$, represent a time series of demeaned daily financial returns. As is typical in the financial econometrics literature, we model r_t as

$$r_t = \sqrt{V_t} \epsilon_t, \tag{1}$$

where $\{\epsilon_t\}$ is an independent and identically distributed (i.i.d.) innovation process with mean 0 and variance 1, which is assumed to be independent of $\{V_t\}$.

2.1.2 Realized variances

Let $\{RV_t\}$ represent a time series of daily realized variances, computed as the sum of intraday squared returns.¹ Because the realized variance is a positive process, we choose to model it

¹ Other measures of realized variances can be used. Assuming that the logarithmic asset price is represented by a continuous-time jump diffusion process, Andersen et al. (2003) show that, under suitable conditions, the intraday squared returns converges in probability to the increment of the quadratic variation process as the sampling frequency of the intraday returns increases.

with a multiplicative error structure (Engle, 2002) as,

$$RV_t = V_t \eta_t, \quad (2)$$

where $\{\eta_t\}$ is a positive i.i.d. innovation process with mean 1, which is assumed to be independent of $\{V_t\}$.² MEM have been shown to perform well on realized volatility data by Engle and Gallo (2006), Gallo and Otranto (2015) and Lanne (2006), among others.

Remark 1. The return model considered in equation (1) implies a MEM for squared returns, as

$$r_t^2 = V_t \eta_t, \quad (3)$$

where, in this specific context, $\eta_t = \epsilon_t^2$.³

2.2 Latent variance model

We first define the latent variance process, $\{V_t\}$, without a leverage component as this allows us to highlight the properties of the model without the added complication of the leverage component. We model V_t as

$$V_t = \sigma^2 C_t M_t, \quad (4)$$

where $\{C_t\}$ is a Markov chain, with a discrete state space, satisfying $\mathbb{E}(C_t) = 1$ and $\{M_t\}$ is a sequence of i.i.d. discrete random variables, assumed independent of $\{C_t\}$ and to satisfy

² As argued by Engle (2002), the main advantage of the MEM approach is that the variable of interest is modeled without any transformation by a process that ensures its positivity. However, note that if it is assumed that η_t follows a lognormal distribution, then $\log RV_t$ can be equivalently expressed as the sum of $\log V_t$ and a normally distributed innovation term, which is a popular specification used to model positive processes in financial econometrics. Therefore, if no specific structure is imposed on V_t , the MEM approach nests many existing specifications based on a logarithmic transformation (e.g., the logarithmic version of the HAR model considered by Andersen et al., 2007; Corsi et al., 2008).

³ Note that although r_t^2 and RV_t share the same model structure, they do not necessarily serve as a proxy for the same latent return variance V_t . This is simply because the realized variance measures return variance inside the trading hours of a day, whereas the squared return quantifies return variance over the return period, which can extend outside of trading hours (e.g., close-to-close returns). Nevertheless, if open-to-close returns are modeled in equation (1), then it could be assumed that the processes $\{V_t\}$ in equations (2) and (3) refer to the same return variance process; in this case, RV_t is known to be a more accurate proxy of V_t than r_t^2 .

$\mathbb{E}(M_t) = 1$. As a consequence, the parameter σ^2 denotes the unconditional expectation of the return variance process, that is, $\mathbb{E}(V_t) = \sigma^2$.

The economic interpretation that we attach to the model is one where volatility is impacted by the arrival of news in the financial market, with varying degrees of importance from day to day. The processes $\{C_t\}$ and $\{M_t\}$ are both used to capture the impacts of these news. The C_t component takes the effect of news whose effect persists over time, whereas M_t represents the impact of non-persistent news and can be interpreted as a jump component. These interpretations become more apparent in Sections 2.2.1 and 2.2.2, where we define C_t and M_t , respectively.

2.2.1 Structure and interpretation of C_t

The process $\{C_t\}$ is constructed as a product of N independent two-state Markov chains, denoted by $\{C_t^{(i)}\}$, $i = 1, \dots, N$:

$$C_t = c_0 \prod_{i=1}^N C_t^{(i)}, \quad (5)$$

where $c_0 = 1/\mathbb{E}[\prod_{i=1}^N C_t^{(i)}]$ is a normalizing constant ensuring that $\mathbb{E}(C_t) = 1$. These Markov chains are assumed to share the same 2×2 transition probability matrix (t.p.m.),

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}, \quad (6)$$

where $p \in (0, 1)$. However, they do not share the same state space as we assume that $C_t^{(i)} \in \{c_i, 1\}$, where $c_1 > 1$ and

$$\begin{aligned} c_i &= (1 - \theta_c) + \theta_c c_{i-1} \\ &= 1 + \theta_c^{i-1}(c_1 - 1), \quad \text{for } i = 2, \dots, N \text{ and } \theta_c \in [0, 1]. \end{aligned}$$

The normalizing constant in equation (5) is therefore easily obtained as

$$c_0 = \left[\prod_{i=1}^N \left(1 + \theta_c^{i-1} (c_1 - 1)/2 \right) \right]^{-1}.$$

Note that $c_1 \geq c_2 \geq \dots \geq c_N \geq 1$, which implies a hierarchical structure in the components of C_t . For instance, if we say that the component $C_t^{(i)}$ is turned ON at time t when $C_t^{(i)} = c_i$ and turned OFF when $C_t^{(i)} = 1$, then $C_t^{(1)}$ and $C_t^{(N)}$ have, respectively, the greatest and weakest impact on volatility when turned ON.⁴

The two-state Markov chains $\{C_t^{(i)}\}$, $i = 1, \dots, N$, are used to model the impact of news arriving in the financial market, so that when any one of these chains is turned ON, volatility increases proportionally to the news importance, measured by the value of c_i . The impact of news on volatility then persists for a number of time periods that follows a geometric distribution with parameter p ; in the applications reported in Section 3, the estimated value of p is very close to 1.

Our construction of $\{C_t\}$ implies that this process is a Markov chain on a state space $\mathcal{X}_C$ with 2^N elements, generated by the Kronecker product of the state spaces of $\{C_t^{(i)}\}$, $i = 1, \dots, N$, that is,

$$\mathcal{X}_C = c_0 \cdot \{c_1, 1\} \otimes \{c_2, 1\} \otimes \dots \otimes \{c_N, 1\}.$$

Its $2^N \times 2^N$ t.p.m., denoted by $\mathbf{P}_C$, is simply

$$\mathbf{P}_C = \mathbf{P}^{\otimes N},$$

where $\mathbf{P}^{\otimes N}$ is the N th Kronecker power of $\mathbf{P}$.⁵

⁴ Note that if $\theta_c = 0$, then, $\forall i, c_i = 1$, and, $\forall t, C_t = 1$, which implies that C_t has no impact on volatility. On the other hand, if $\theta_c = 1$, then, $\forall i, c_i = c_1$, which implies that $\{C_t\}$ corresponds to the product of independent Markov chains with identical state spaces.

⁵ The k th Kronecker power of $\mathbf{P}$, denoted by $\mathbf{P}^{\otimes k}$, is defined inductively for $k \in \mathbb{N}$ by

$$\mathbf{P}^{\otimes 1} = \mathbf{P} \quad \text{and} \quad \mathbf{P}^{\otimes k} = \mathbf{P} \otimes \mathbf{P}^{\otimes (k-1)}, \quad k = 2, 3, \dots$$

Remark 2. Because we assumed that $p \in (0, 1)$, $\mathbf{P}_C$ is a positive matrix (i.e., all elements of $\mathbf{P}_C$ are positive), which implies that all states of the Markov chain $\{C_t\}$ communicate with each other in a single time step. Therefore, $\{C_t\}$ is an ergodic Markov chain (i.e., it is irreducible and aperiodic) with a unique stationary distribution, which we denote by $\boldsymbol{\pi}_C$. Lemma 5 implies that⁶

$$\boldsymbol{\pi}_C = 2^{-N} \mathbf{1}_{2^N},$$

where $\mathbf{1}_n$ is used to denote the n -dimensional column vector of ones, for $n = 1, 2, \dots$

2.2.2 Structure and interpretation of M_t

The process $\{M_t\}$ is defined to be a sequence of i.i.d. discrete random variables with probability mass function

$$\Pr(M_t = m_0 \cdot m_i) = \begin{cases} q(N-1)^{-1}, & \text{if } i = 1, \dots, N-1, \\ 1-q, & \text{if } i = N, \end{cases}$$

where $q \in (0, 1)$, $m_1 > 1$,

$$\begin{aligned} m_i &= (1 - \theta_m) + \theta_m m_{i-1} \\ &= 1 + \theta_m^{i-1} (m_1 - 1), \quad \text{for } i = 2, \dots, N-1, \end{aligned}$$

and $m_N = 1$. We assume that $\theta_m \in [0, 1]$, which implies that $m_1 \geq m_2 \geq \dots \geq m_N = 1$, and use m_0 as a normalizing constant to ensure $\mathbb{E}(M_t) = 1$, which leads to

$$m_0 = \left[1 + q \frac{(m_1 - 1)(1 - \theta_m^{N-1})}{(N-1)(1 - \theta_m)} \right]^{-1}.$$

We interpret $\{M_t\}$ as a process capturing the non-persistent impact on volatility of the

⁶ The stationary distribution of an ergodic Markov chain is given by the left eigenvector of its t.p.m. associated with an eigenvalue of 1, appropriately normalized such that its elements sum to 1.

arrival of news in the financial market. The parameter q corresponds to the probability of this kind of news arriving in a given time period. This news has a multiplicative impact on volatility, given by one of the values $m_1, m_2, \dots, m_{N-1}$, chosen with equal probabilities (ON states), with m_1 representing the greatest impact and m_{N-1} the weakest impact. The probability of no news arriving is $1 - q$, which is associated with $m_N = 1$ (OFF state). In contrast to $\{C_t\}$, the impact of news generated by the $\{M_t\}$ process does not persist over time since it is an independent process. Consequently, this component of the model serves to generate non-persistent jumps of different magnitudes on volatility,.

For further developments, it is convenient to express $\{M_t\}$ in the form of a Markov chain. To this end, let $\boldsymbol{\pi}_M$ be the column vector of the N component probabilities,

$$\boldsymbol{\pi}_M = \left(\underbrace{\frac{q}{N-1}, \dots, \frac{q}{N-1}}_{(N-1) \text{ terms}}, 1 - q \right)'. \quad (7)$$

Then, $\{M_t\}$ can be expressed as a Markov chain with $N \times N$ t.p.m.,

$$\boldsymbol{P}_M = \mathbf{1}_N \boldsymbol{\pi}_M',$$

on the state space $\mathcal{X}_M$ with N elements,

$$\mathcal{X}_M = m_0 \cdot \{m_1, m_2, \dots, m_N\}.$$

Remark 3. Because $q \in (0, 1)$, $\boldsymbol{P}_M$ is a positive matrix, and $\{M_t\}$ is an ergodic Markov chain with stationary distribution $\boldsymbol{\pi}_M$ (see Remark 2 and Lemma 6).

2.2.3 Markov chain structure of V_t

The latent return variance at time t , V_t , is the product of C_t and M_t , as specified in equation (4), hence it combines the effects on volatility of the arrival of persistent and non-

persistent news in the financial market. Since $\{V_t\}$ is a product of two Markov chains, it is itself a Markov chain with $(N \cdot 2^N) \times (N \cdot 2^N)$ t.p.m.,

$$\mathbf{P}_V = \mathbf{P}_C \otimes \mathbf{P}_M,$$

on the state space $\mathcal{X}_V$ with $N \cdot 2^N$ elements, where

$$\mathcal{X}_V = \sigma^2 \cdot \mathcal{X}_C \otimes \mathcal{X}_M.$$

Remark 4. $\{V_t\}$ is potentially a high-dimensional Markov chain (e.g., for $N = 10$, the number of states is 10240), but it is parsimoniously indexed by the seven parameters $\{\sigma^2, p, q, c_1, m_1, \theta_c, \theta_m\}$.

Remark 5. Because $\mathbf{P}_C$ and $\mathbf{P}_M$ are positive matrices (see Remarks 2 and 3), $\mathbf{P}_V$ is itself a positive matrix, and $\{V_t\}$ is an ergodic Markov chain with stationary distribution given by (see Lemma 7)

$$\boldsymbol{\pi}_V = \boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M.$$

Remark 6. In constructing model, we assumed that the number of components used as building blocks of $\{M_t\}$ and $\{C_t\}$ is the same and equal to N . Although nothing prevents us from considering a different number of components in each component, in our view it makes sense to specify N as large as possible up to computational constraints in both components because the effect of additional components on volatility, measured by the variables c_i and m_i , is structured to decrease geometrically. Therefore, when N is large, the model has the ability to adjust itself, through the parameters c_1 , m_1 , θ_c and θ_m , and assign very little importance to superfluous components.

2.2.4 Volatility persistence

It is a well-known empirical fact that the volatility of returns on financial assets exhibits a high degree of persistence (see e.g., Mandelbrot, 1963; Bollerslev, 1986). In the HDM model, volatility persistence can be characterized by the speed at which $\text{Cov}(V_t, V_{t+k})$ approaches zero as k increases. Let $\mathbf{v}$ denote the $N \cdot 2^N$ column vector of the elements of $\mathcal{X}_V$, and let $\mathbf{\Upsilon}$ denote the $(N \cdot 2^N) \times (N \cdot 2^N)$ diagonal matrix with the elements of $\mathbf{v}$ on its diagonal (i.e., $\mathbf{v} = \mathbf{\Upsilon} \mathbf{1}_{N \cdot 2^N}$). Then, based on standard Markov chain theory (see Hamilton, 1994, Chapter 22), we have

$$\begin{aligned} \text{Cov}(V_t, V_{t+k}) &= \boldsymbol{\pi}'_V \mathbf{\Upsilon} \mathbf{P}_V^k \mathbf{v} - (\boldsymbol{\pi}'_V \mathbf{v})^2 \\ &= \boldsymbol{\pi}'_V \mathbf{\Upsilon} (\mathbf{P}_V^k - \mathbf{1}_{N \cdot 2^N} \boldsymbol{\pi}'_V) \mathbf{v}, \quad k = 0, 1, \dots, \end{aligned} \tag{8}$$

and $\text{Cov}(V_t, V_{t+k}) \rightarrow 0$ as $k \rightarrow \infty$.

Clearly, the rate at which volatility tends to persist in time is directly related to the rate of convergence of the matrix $\mathbf{P}_V^k$ as k tends to infinity. It is well known that if γ denotes the second largest eigenvalue (in absolute value) of $\mathbf{P}_V$, then $|\gamma|^k$ is the dominating term in its asymptotic rate of convergence (see Poskitt and Chung, 1996). This observation led Rydén et al. (1998) to affirm that hidden Markov models “can only produce series with exponentially decaying autocorrelation functions,” and that these models are therefore “doomed from the start” for replicating the high degree of persistence in volatility which is empirically observed. Although this affirmation holds asymptotically, Theorem 1 shows that the particular structure that we introduce to construct the Markov chain $\{V_t\}$, specifically the multiplication of N two-state Markov chains with identical t.p.m., offers a way to slow down the convergence of $\mathbf{P}_V^k$ as $k = 1, 2, \dots$

Theorem 1 (Rate of convergence of $\mathbf{P}_V$). *Let $\gamma = 2p - 1$ and $\boldsymbol{\Pi}_V = \lim_{k \rightarrow \infty} \mathbf{P}_V^k$.*

(i) Asymptotic limit of $\mathbf{P}_V^k$ as $k \rightarrow \infty$:

$$\mathbf{\Pi}_V = \mathbf{1}_{N \cdot 2^N} \boldsymbol{\pi}_V'.$$

(ii) Nonasymptotic rate of convergence of $\mathbf{P}_V^k$ as $k = 1, 2, \dots$:

$$\|\mathbf{P}_V^k - \mathbf{\Pi}_V\|_\infty \leq (1 + |\gamma|^k)^N - 1, \quad (9)$$

where $\|\cdot\|_\infty$ is the maximum absolute row sum norm, and, for $\gamma \in [0, 1)$,⁷

$$\|\mathbf{P}_V^k - \mathbf{\Pi}_V\|_{\max} = \left((1 + \gamma^k)^N - 1\right) \|\boldsymbol{\pi}_V\|_\infty, \quad (10)$$

with $\|\boldsymbol{\pi}_V\|_\infty = 2^{-N} \max\{q/(N-1), 1-q\}$, where $\|\cdot\|_{\max}$ is the max norm, that is, the maximum absolute element of the given matrix.

(iii) Asymptotic rate of convergence of $\mathbf{P}_V^k$ as $k \rightarrow \infty$:

$$\mathbf{P}_V^k - \mathbf{\Pi}_V = O(k^{N-1}|\gamma|^k). \quad (11)$$

Proof. The proof is given in Appendix 5.3. □

Remark 7. From a linear algebra standpoint, N corresponds to the algebraic multiplicity of the eigenvalue γ of the matrix $\mathbf{P}_V$, which is its largest eigenvalue (in absolute value) that is smaller than 1. Note that the 2×2 matrix $\mathbf{P}$ also has an eigenvalue of $\gamma = 2p - 1$, but its algebraic multiplicity is 1. Since N corresponds to the number of components used in the construction of $\{C_t\}$, the algebraic multiplicity of the eigenvalue γ of the matrix $\mathbf{P}_V$ increases by one unit each time a component is added.

Theorem 1 shows that the number of latent components N impacts the rate of convergence of $\mathbf{P}_V^k$ as $k = 1, 2, \dots$. For instance, if $N = 1$, we have $\|\mathbf{P}_V^k - \mathbf{\Pi}_V\|_\infty \leq |\gamma|^k$ and $\mathbf{P}_V^k - \mathbf{\Pi}_V =$

⁷ When $\gamma \in (-1, 0)$, equation (10) also holds for even k .

$O(|\gamma|^k)$. Equations (9)–(11) indicate that higher values of N generally lead to a slower decay of $\mathbf{P}_V^k$ as $k = 1, 2, \dots$, and that the impact of a higher N is magnified the closer γ (or equivalently p) is to 1 (in the non-asymptotic case). The chosen structure of the volatility process therefore offers a flexible way to capture varying degrees of volatility persistence (by the choice of N), and this is an important motivation for this structure. In fact, as can be seen in the empirical applications of Section 3, the HDM model very well mimicks the autocorrelation structure of squared returns and realized variances.

2.3 Autocovariance structure and moments

Although the Markov chain process $\{V_t\}$ exhibits a particular structure and has a high-dimensional state space, it nevertheless remains a time-homogeneous Markov chain on a finite state space. Consequently, the models for financial returns and realized variances, presented in Section 2.1 and 2.2, are comprised in the class of hidden Markov models. Accordingly, the autocovariance structure, the conditional and unconditional moments, as well as the log-likelihood function can all be computed in closed-form based on standard techniques.

2.3.1 Autocovariance structure

First, let us consider the (unconditional) autocorrelation function of $\{r_t^2\}$ and $\{\text{RV}_t\}$. Since Section 2.1.2 showed that HDM models for r_t^2 and RV_t share the same multiplicative error structure, the derivation of the autocorrelation function for these two variables is treated at once in Proposition 1 by introducing a new variable x_t that represents either r_t^2 or RV_t .

Proposition 1 (Autocovariance structure). *Let*

$$x_t = V_t \eta_t,$$

where V_t is defined by (4) and $\{\eta_t\}$ is a positive i.i.d. random process with mean 1 and finite

variance, which is assumed independent of $\{V_t\}$, and let

$$\phi_i = \left(\frac{c_i - 1}{c_i + 1} \right)^2 = \left(\frac{\theta_c^{i-1}(c_1 - 1)}{\theta_c^{i-1}(c_1 - 1) + 2} \right)^2 \in [0, 1], \quad i = 1, \dots, N.$$

For $k = 1, 2, \dots$, we have:

(i)

$$\text{Cov}(x_t, x_{t+k}) = \text{Cov}(V_t, V_{t+k}), \quad (12)$$

(ii)

$$\text{Cov}(x_t, x_{t+k}) = \sigma^4 \left(\prod_{i=1}^N (1 + \phi_i \gamma^k) - 1 \right), \quad (13)$$

(iii)

$$\text{Var}(x_t) = \sigma^4 \left(\mathbb{E}[\eta_t^2] m_0^2 \left(\prod_{i=1}^N (1 + \phi_i) \right) \left(\frac{q}{N-1} \sum_{i=1}^{N-1} m_i^2 + (1-q) \right) - 1 \right), \quad (14)$$

(iv)

$$\text{Corr}(x_t, x_{t+k}) = \frac{\prod_{i=1}^N (1 + \phi_i \gamma^k) - 1}{\mathbb{E}[\eta_t^2] m_0^2 \left(\prod_{i=1}^N (1 + \phi_i) \right) \left(\frac{q}{N-1} \sum_{i=1}^{N-1} m_i^2 + (1-q) \right) - 1},$$

where $\gamma = 2p - 1$, p being the parameter of the t.p.m. defined in (6).

Proof. The proof is given in Appendix 5.4. □

Note that although equation (8) provides us with a generic expression to compute $\text{Cov}(V_t, V_{t+k})$, and hence $\text{Cov}(x_t, x_{t+k})$ by equation (12), Proposition 1 has the advantage of expressing this autocovariance function directly in terms of the parameters of the model. For instance, equation (13) shows that it is strongly influenced by the number of latent components N . In order to see how N impacts on the autocovariances, let us consider two HDM models differing by only one latent component. If both models share the same parameters, σ^2 , p and c_i , $i = 1, \dots, N-1$, then the autocovariances of the model with $N-1$ components, denoted by $\text{Cov}_{N-1}(x_t, x_{t+k})$, are always smaller than or equal to the autocovariances of the model

exhibiting one extra component, denoted by $\text{Cov}_N(x_t, x_{t+k})$, since we have

$$\begin{aligned}\text{Cov}_N(x_t, x_{t+k}) &= (1 + \phi_N \gamma^k) \sigma^4 \left(\prod_{i=1}^{N-1} (1 + \phi_i \gamma^k) - 1 \right) + \phi_N \gamma^k \sigma^4 \\ &= (1 + \phi_N \gamma^k) \text{Cov}_{N-1}(x_t, x_{t+k}) + \phi_N \gamma^k \sigma^4 \\ &\geq \text{Cov}_{N-1}(x_t, x_{t+k}).\end{aligned}$$

Another interesting feature of Proposition 1 comes from the variance equation (14), which shows that the excess kurtosis typically observed in financial returns can be captured either by the latent components C_t and M_t , or by $\mathbb{E}(\eta_t^2)$.⁸

2.3.2 Moments

Of particular interest is the conditional moment forecast of x_{t+h} , for $h = 1, 2, \dots$, based on the available information up to time t (as in Section 2.3.1, x_t represents either r_t^2 or RV_t). To compute this forecast, one must first obtain the vector of filtered probabilities, denoted by $\boldsymbol{\xi}_{t|t}$, using standard filtering techniques developed for hidden Markov models (e.g., Hamilton, 1994, Chapter 22). Let $v_1, v_2, \dots, v_{N \cdot 2^N}$ denote the elements of $\mathbf{v}$, and let $\boldsymbol{\xi}_{t+h|t}$ be the $N \cdot 2^N$ column vector with elements,

$$\xi_{i,t+h|t} = \Pr(V_{t+h} = v_i \mid x_t, x_{t-1}, \dots), \quad i = 1, \dots, N \cdot 2^N, \quad (15)$$

where $h = 0, 1, \dots$. These conditional forecast probabilities are directly obtained from the filtered probabilities, since $\boldsymbol{\xi}'_{t+h|t} = \boldsymbol{\xi}'_{t|t} \mathbf{P}_V^h$, $h = 1, 2, \dots$. It is then simple to compute the conditional moment forecast, $\mathbb{E}[g(x_{t+h}) \mid x_t, x_{t-1}, \dots]$, for any real-valued function $g(\cdot)$, from the following expression:

$$\mathbb{E}[g(x_{t+h}) \mid x_t, x_{t-1}, \dots] = \sum_{i=1}^{N \cdot 2^N} \xi_{i,t+h|t} \mathbb{E}[g(v_i \eta_{t+h})]. \quad (16)$$

⁸ In the case of returns, $\mathbb{E}(\eta_t^2)$ is the fourth moment of ϵ_t : like in GARCH models, the kurtosis of returns depends on the kurtosis of the innovation.

When $g(x) = x^r$, expression (16) simplifies to

$$\mathbb{E} \left[x_{t+h}^r \mid x_t, x_{t-1}, \dots \right] = \mathbb{E} \left[\eta_{t+h}^r \right] \sum_{i=1}^{N \cdot 2^N} \xi_{i,t+h|t} v_i^r.$$

Note that to compute unconditional moments, one must simply replace the probability vector $\boldsymbol{\xi}_{t+h|t}$ by the stationary distribution $\boldsymbol{\pi}_V$ (in fact, $\boldsymbol{\xi}_{t+h|t} \rightarrow \boldsymbol{\pi}_V$ as $h \rightarrow \infty$).

2.4 Relationship to the MSM model

Since we mentioned in the introduction that the construction of the HDM model is motivated by the success of the MSM approach of Calvet and Fisher (2004), it is instructive to relate our model to the MSM model. The MSM process was initially proposed as a model for financial returns, and it thus admits the general form given in equation (1). Its latent variance is specified as

$$\tilde{V}_t = \tilde{\sigma}^2 \prod_{i=1}^N \tilde{C}_t^{(i)},$$

where, for $i = 1, \dots, N$,

$$\tilde{C}_t^{(i)} = \begin{cases} \tilde{C}_{t-1}^{(i)}, & \text{with probability } \tilde{p}_i, \\ \tilde{c}, & \text{with probability } (1 - \tilde{p}_i)/2, \\ 2 - \tilde{c}, & \text{with probability } (1 - \tilde{p}_i)/2, \end{cases}$$

with $\tilde{c} \in (0, 1)$, and

$$\tilde{p}_i = (1 - \tilde{a})^{\tilde{b}^{i-1}}, \quad i = 1, \dots, N,$$

where $\tilde{a} \in (0, 1)$ and $\tilde{b} \in (1, \infty)$. Note that $\mathbb{E}(\tilde{V}_t) = \tilde{\sigma}^2$. The MSM model therefore includes four parameters, $\{\tilde{\sigma}^2, \tilde{a}, \tilde{b}, \tilde{c}\}$.

It is easily seen that $\{\tilde{C}_t^{(i)}\}$, $i = 1, \dots, N$, are independent two-state Markov chains defined on a common state space comprised of the values $\{\tilde{c}, 2 - \tilde{c}\}$. Consequently, $\tilde{V}_t$ can only take $N + 1$ distinct values in the set $\{\tilde{\sigma}^2 \tilde{c}^i (2 - \tilde{c})^{N-i}\}_{i=0}^N$. This represents a first

important difference with respect to the HDM model. Moreover, in contrast to our approach, the Markov chains, $\{\tilde{C}_t^{(i)}\}$, $i = 1, \dots, N$, do not all share the same t.p.m., which implies that the structure of the MSM model does not necessarily benefit from the results of Theorem 1 and Proposition 1.

A model that is related to the MSM model is the component-driven regime-switching (CDRS) model of Fleming and Kirby (2013). This model also represents the latent variance by way of two-state Markov chains with identical state spaces, but it allows some of these Markov chains to share the same t.p.m.. Our approach, however, has the advantage of practically eliminating the need for model selection, that is, the fitting procedure has the flexibility to assign little to no importance to superfluous components, which indicates that one can simply consider a value of N as large as is possible from a computational standpoint (see Remark 6).⁹ Moreover, the support of the volatility distribution of the HDM model can comprise thousands of points (10240 for $N = 10$), whereas the models considered by Calvet and Fisher (2004) and Fleming and Kirby (2013) are restricted practically to switch between at most eleven different values ($N = 10$).

2.5 Leverage effect

An additional novelty of the HDM model, that is not shared by the MSM and CDRS processes, is the possibility to include in it a time-varying leverage effect. The empirical analyses of Section 3 show that this component significantly enhances the in-sample fit and out-of-sample forecasting performance of the model on data from the S&P 500 index and the NASDAQ Composite index.

With a leverage effect, the latent variance specification introduced in (4) is extended to include an additional component:

$$V_t = \sigma^2 C_t M_t L_t,$$

⁹ Using the exact filter of Rabiner (1989), a number of $N = 10$ components can feasibly be integrated out. More components could be handled with a sequential Monte Carlo algorithm such as those of Gordon et al. (1993) and of Pitt and Shephard (1999).

where

$$L_t = \prod_{i=1}^{N_L} L_t^{(i)},$$

and

$$L_t^{(i)} = \begin{cases} 1, & \text{if } r_{t-i} \geq 0, \\ 1 + l_i \frac{|r_{t-i}|}{\sqrt{L_{t-i}}}, & \text{if } r_{t-i} < 0, \end{cases}$$

for $i = 1, \dots, N_L$, with $l_1 > 0$, and

$$l_i = \theta_l^{i-1} l_1, \quad \text{for } i = 2, \dots, N_L \text{ and } \theta_l \in [0, 1].$$

Observe that the leverage component, L_t , adds two parameters, $\{l_1, \theta_l\}$, and is specified as a predictable process, that is, its value at any given time index is fully determined by the observed returns up to the previous time step. This entails that the leverage component may be extracted and studied from an econometric perspective.¹⁰

Moreover, the process $\{L_t\}$ may be interpreted similarly to $\{C_t\}$ and $\{M_t\}$. For instance, if we let the returns $\{r_t\}$ represent a type of news, then we may say that the component $L_t^{(i)}$ is turned ON if $r_{t-i} < 0$ and turned OFF if $r_{t-i} \geq 0$. The impact of this component on volatility is then influenced by the importance of the news, which is a function of the magnitude of the negative return, $|r_{t-i}|$, and of a multiplicative factor, l_i , structured in such a way as to give less importance to more distant news. Since the specific value of N_L has a

¹⁰It must be noted that Cordis and Kirby (2014) and Rossi and Gallo (2006) consider regime-switching models with more than four states and discuss the integration of a leverage effect. However, their approach is very different from ours because they directly model the volatility states and the elements of the transition matrix with parametric functions. For instance, Cordis and Kirby (2014) construct the transition matrix by relating the volatility process to a discrete first- or second-order autoregressive process and use a binomial distribution to model off-diagonal elements. They propose to incorporate a leverage effect by making the probability parameter in the binomial distribution dependent upon past realizations of the process through a logistic function. On the other hand, Rossi and Gallo (2006) model the transition matrix with a tridiagonal structure and account for the leverage effect by allowing the elements of this matrix to be influenced by the sign of the return. Nonetheless, the impact of the leverage effect in these models cannot be extracted. Calvet et al. (2015) introduce a regime-switching stochastic volatility model with jumps and a leverage effect, where regime switching is specified as in the MSM model, regime switches affect the long-run level of the stochastic variance, and the leverage effect is specified by correlating the innovations of the variance process and the log-price process.

negligible impact on the computational burden, the number of components in the leverage effect may be chosen to be very large in applications.

3 Applications to daily returns and realized variances

We compare the HDM process to popular models on daily percentage log-returns and realized variances from the Standard and Poor’s 500 index (S&P 500), the Nasdaq Composite index (NASDAQ) and the exchange rate USD/EUR. Daily percentage log-returns span the period extending from January 3, 2000 to June 30, 2016, which yields 4150 observations (source: Federal Reserve Economic Data (FRED) database). Realized kernel variances cover the same period, except for the USD/EUR exchange rate, for which the data is available until March 3, 2009 (2337 observations) (source: Oxford-Man Institute of Quantitative Finance).

In the subsequent sections, we use the same number of components and innovations for the HDM model. The distribution of the innovation (ϵ_t) of the return process is a standard normal, whereas that of the innovation (η_t) of the realized variance process is Gamma($1/v, v$), with mean equal to 1.¹¹ Regarding the number of latent components, we set N to 10 and the number of leverage components N_L to 70 (slightly more than three financial months). With this number of components, the jump process $\{M_t\}$ and the Markov-chain $\{C_t\}$ generate altogether 10240 volatility states. As the leverage effect component depends on previous returns, the HDM model with leverage exhibits a continuous volatility process. The MATLAB program to implement the HDM model is available on the corresponding author’s website.

3.1 In-sample results for S&P 500 data

We first study the S&P 500 data displayed in Figure 1. Table 1 presents the log-likelihood function evaluated at the maximum likelihood estimate, the Akaike information criterion

¹¹ We use the following parameterization of the gamma density: if $X \sim \text{Gamma}(a, b)$, then $f_X(x | a, b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}$, and $E(X) = a b$.

(AIC) and the Bayesian information criterion (BIC) for seven models that are briefly reviewed in Appendix 6 : the MSM (Calvet and Fisher, 2004), the GARCH(1,1) (Bollerslev, 1986), the GJR-GARCH(1,1) (GJR) (Glosten et al., 1993), the two-state Markov-switching GARCH(1,1) (MS-GARCH) (Haas et al., 2004), the two-state Markov-switching GJR-GARCH(1,1) (MS-GJR), and the HDM model with (HDM-lev) and without (HDM) a leverage component. To compare with cutting-edge models, we consider Student- t distributions for the GARCH innovations, which is reminded by t in the model acronyms. Unsurprisingly, on both information criteria, the models that include a leverage effect strongly dominate the corresponding models without this effect. Focusing firstly on models without leverage components, we observe that the HDM process competes favorably with the most flexible alternative model, i.e. the MS-GARCH- t process. For the models with leverage, judging by the BIC, the HDM model with a leverage component highly dominates the GJR- t model, and is equivalent to the MS-GJR- t model. Regarding the AIC, the MS-GJR- t process slightly dominates the best HDM one, and both strongly dominate the GJR- t process. To summarize, the HDM models with and without leverage component deliver as good fits as their flexible competitors on this time series. Moreover, the HDM process could readily be improved by including breaks into its model parameters while the MS-GARCH and GJR models have already incorporated this extra feature.

With respect to realized variance data, Table 1 shows the statistics for the MEM (Engle, 2002), the asymmetric MEM (AMEM) (Gallo and Otranto, 2015), the two-state Markov-switching MEM (Gallo and Otranto, 2015), and the HAR model (Corsi, 2009). The specifications of these alternative processes are briefly presented in Appendix 7. For a fair comparison, the innovations of these models are driven by a gamma distribution as in the HDM models. The results show that the realized variance HDM model, with leverage, strongly dominates its competitors on both information criteria, and that the leverage component improves each model with respect to the corresponding model that does not include such a component.

Figure 2 displays the empirical autocorrelations of the squared percentage log-returns and

Figure 1 – S&P 500 data from January 3, 2000 to June 30, 2016.

Table 1 – S&P 500: Log-likelihood and information criteria.

Percentage log-returns							
	MSM	GARCH- t	GJR- t	MS-GARCH- t	MS-GJR- t	HDM	HDM-lev
Np	4	4	5	9	11	7	9
log-lik	-5874.1	-5870.4	-5782.0	-5861.3	-5770.5	-5862.9	-5770.1
AIC	-5878.1	-5874.4	-5787.0	-5870.3	-5781.5	-5869.9	-5779.1
BIC	-5890.8	-5887.1	-5802.8	-5898.8	-5816.3	-5892.1	-5807.5
Realized kernel variances							
	HAR	MEM	AMEM	MS-MEM	MS-AMEM	HDM	HDM-lev
Np	5	4	5	9	11	8	10
log-lik	-1473.5	-1465.0	-1366.4	-1234.0	-1120.7	-1148.4	-956.9
AIC	-1478.5	-1469.0	-1371.4	-1243.0	-1131.7	-1156.4	-966.9
BIC	-1494.3	-1481.7	-1387.2	-1271.5	-1166.5	-1181.7	-998.5

Np: number of parameters; log-lik: Maximum of the log-likelihood; AIC: Akaike Information criterion; BIC: Bayesian information criterion; The highest values are bolded.

of the realized variances with respect to the theoretical ones implied by the HDM models.¹² We observe a long-lasting volatility persistence that is reasonably well tracked by the implied autocorrelations of the HDM model, especially with leverage.

Focusing now only on the HDM model with leverage, Table 2 documents the parameter estimates. Regarding the returns, we observe that the independent component is weaker than the Markov-chain component (i.e., $m_0 m_1 < c_1 c_0^{\frac{1}{N}}$) meaning that non-persistent news

¹²The autocorrelations implied by the HDM model with a leverage effect component are computed by simulation.

(a) Percentage log-returns : without leverage effect

(b) Realized kernel variances : without leverage effect

(c) Percentage log-returns : with leverage effect

(d) Realized kernel variances : with leverage effect

Figure 2 – S&P 500 : Theoretical autocorrelations implied by the HDM models are displayed in solid lines with respect to the empirical ones in dashed lines.

have a smaller impact on volatility than persistent news.¹³ In contrast, when looking at the realized variances, non-persistent news have a much larger effect than enduring news. Additionally, note that for the two series, the persistence parameters $(\theta_m, \theta_c, \theta_l)$ are different from one, which indicates that the impact of the multiplicative components on volatility decays exponentially.

Figure 3 displays the smoothed probabilities of the most influential jump component in the HDM model for returns. Note that some of these probabilities are very high compared to

¹³ The jump state m_9 is far from unity, which could indicate that extra states are needed. We increased the number of states up to 70 but this did not significantly improve the fit.

Table 2 – S&P 500: Maximum likelihood estimates of the HDM model with leverage.

Percentage log-returns									
Constant component : $\sigma^2 = 0.22$									
Mixture component : $\theta_m = 0.87, m_1 = 23.55, q = 0.93$									
m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
23.55	20.71	18.23	16.06	14.16	12.51	11.06	9.79	8.68	1.00
$m_1 m_0$	$m_2 m_0$	$m_3 m_0$	$m_4 m_0$	$m_5 m_0$	$m_6 m_0$	$m_7 m_0$	$m_8 m_0$	$m_9 m_0$	m_0
1.69	1.48	1.31	1.15	1.02	0.90	0.79	0.70	0.62	0.07
Markov-chain component : $\theta_c = 0.51, c_1 = 1.99, p = 0.9986$									
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
1.99	1.50	1.26	1.13	1.07	1.03	1.02	1.01	1.00	1.00
$c_1 c_0^{\frac{1}{N}}$	$c_2 c_0^{\frac{1}{N}}$	$c_3 c_0^{\frac{1}{N}}$	$c_4 c_0^{\frac{1}{N}}$	$c_5 c_0^{\frac{1}{N}}$	$c_6 c_0^{\frac{1}{N}}$	$c_7 c_0^{\frac{1}{N}}$	$c_8 c_0^{\frac{1}{N}}$	$c_9 c_0^{\frac{1}{N}}$	$c_{10} c_0^{\frac{1}{N}}$
1.82	1.38	1.15	1.04	0.98	0.95	0.93	0.92	0.92	0.92
Leverage component : $\theta_l = 0.92, l_1 = 1.00$									
Realized kernel variances									
Constant component : $\sigma^2 = 0.51$									
Mixture component : $\theta_m = 0.75, m_1 = 3.45, q = 0.12$									
m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
3.29	2.65	2.19	1.86	1.62	1.45	1.32	1.23	1.17	1.00
$m_1 m_0$	$m_2 m_0$	$m_3 m_0$	$m_4 m_0$	$m_5 m_0$	$m_6 m_0$	$m_7 m_0$	$m_8 m_0$	$m_9 m_0$	m_0
3.08	2.53	2.12	1.81	1.58	1.40	1.27	1.18	1.11	0.89
Markov-chain component : $\theta_c = 0.81, c_1 = 2.19, p = 0.9897$									
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
2.34	2.21	2.09	1.99	1.89	1.80	1.72	1.65	1.59	1.53
$c_1 c_0^{\frac{1}{N}}$	$c_2 c_0^{\frac{1}{N}}$	$c_3 c_0^{\frac{1}{N}}$	$c_4 c_0^{\frac{1}{N}}$	$c_5 c_0^{\frac{1}{N}}$	$c_6 c_0^{\frac{1}{N}}$	$c_7 c_0^{\frac{1}{N}}$	$c_8 c_0^{\frac{1}{N}}$	$c_9 c_0^{\frac{1}{N}}$	$c_{10} c_0^{\frac{1}{N}}$
1.73	1.55	1.41	1.29	1.20	1.12	1.06	1.01	0.97	0.93
Leverage component : $\theta_l = 0.81, l_1 = 0.40$									

the unconditional probability of 10%, which allows us to capture aberrant values and identify large jumps.

Figure 4 provides the smoothed probabilities over time of each one of the first three $C_t^{(i)}$ components being turned ON. The solid line is associated with the component having the strongest impact on volatility. For the two series, this component is likely to be active during the dot-com crisis as well as during the subprime mortgage crisis. It mimics a long-run volatility dynamic and could be interpreted as a bull and bear effect. We also see a correlation between the different probabilities that could be accounted for.

Figure 5 shows the values taken by the leverage effect component L_t . For both series,

Figure 3 – S&P 500: Smoothed probabilities of the strongest mixture component.

(a) Percentage log-returns

(b) Realized kernel variances

Figure 4 – S&P 500 : Smoothed probabilities of the three most influential $C_t^{(i)}$ components. The component with the largest impact on volatility is displayed in solid line. The dash-dotted line denotes the second component, the dotted line corresponds to the third component.

we observe that its effect is very strong during the subprime mortgage crisis and (to a lesser extent) during the dot-com crisis. Since the leverage effect component is specified in a non-traditional way, it can be questioned whether it really corresponds to the so-called leverage effect captured by standard volatility models, such as the GJR model. To investigate this

issue, note first that the GJR volatility process can be decomposed as follows:¹⁴

$$\begin{aligned}\sigma_t^2 &= \omega + (\alpha + \delta \mathbb{1}_{\{y_{t-1} < 0\}}) y_{t-1}^2 + \beta \sigma_{t-1}^2, \\ &= \omega \left(\sum_{i=0}^{t-1} \beta^i \right) + \alpha \left(\sum_{i=0}^{t-1} \beta^i y_{t-1-i}^2 \right) + \underbrace{\delta \left(\sum_{i=0}^{t-1} \beta^i \mathbb{1}_{\{y_{t-1-i} < 0\}} y_{t-1-i}^2 \right)}_{L_t^{GJR}} + \beta^t \sigma_0^2.\end{aligned}\quad (17)$$

The decomposition (17) makes explicit the contribution of the leverage component of the GJR model to the variance dynamics. In fact, the leverage effect at time t is given by

$$L_t^{GJR} = \delta \left(\sum_{i=0}^{t-1} \beta^i \mathbb{1}_{\{y_{t-1-i} < 0\}} y_{t-1-i}^2 \right).$$

Obviously, this effect depends on several previous squared returns. More precisely, the impact of the i th lagged squared return is scaled by the coefficient $l_i^{GJR} = \delta \beta^{i-1}$, which geometrically decays over time. Although there is a clear connection between L_t and L_t^{GJR} , as well as between l_i and l_i^{GJR} , the impacts of these components on volatility cannot be directly compared. This is due to the fact that the leverage component of the HDM model is specified to have a multiplicative impact on volatility, whereas the impact of GJR leverage component is additive. Nevertheless, the correlation between L_t and L_t^{GJR} can be computed and is in fact equal to 0.92, which confirms that both components capture a similar effect. Furthermore, we illustrate the coefficients l_i for $i = 1, \dots, 70$ in Figure 6. Interestingly, more than 60 lagged returns are relevant to build the leverage component. We can interpret this long-lasting impact as the time needed for the financial market to completely react to a negative return. The leverage effect coefficients l_i^{GJR} of the GJR model also corroborate this long-term effect as the impact of one negative return generally persists for approximately two financial months.

When modeling returns, the HDM process without leverage effect specifies a discrete

¹⁴For this leverage analysis, we consider a GJR-GARCH model with Normal innovations. So, the random variable σ_t^2 coincides with the variance of the process when we condition to the information up to time $t - 1$.

Figure 5 – S&P 500 : Leverage effect over time.

process for the latent volatility. This feature may raise some concerns about the flexibility of the dynamics. Figure 7 documents the median of the HDM conditional volatility with and without leverage effect.¹⁵ These dynamics are compared with the MSM conditional volatility. While the MSM conditional volatility oscillates between 6 different values (i.e. the best MSM models exhibits 5 latent components), the HDM conditional volatility dynamics are way richer. Without leverage, the HDM conditional volatility evolve over 10240 different states which is already larger than the sample size. From Figure 7, we observe that the HDM conditional volatility smoothly tracks the continuous conditional volatility dynamic of the HDM process with leverage effect.

3.2 In-sample results for NASDAQ and USD/EUR data

Table 3 provides in-sample results for the NASDAQ and USD/EUR time series. Like for the S&P 500 data, the HDM process compares favorably well to competing models and its performance particularly stands out on the realized variances series. Two remarks are worth mentioning. First, in accordance with the financial econometrics literature, the inclusion of a leverage effect does not have a significant impact on the fit of the exchange rate series.

¹⁵The median is computed as $Q_{\sqrt{V_t}|y_{1:T}}(0.5)$ where $Q_X(p)$ denotes the quantile function of the random variable X .

(a) Percentage log-returns

(b) Realized kernel variances

(c) Percentage log-returns - GJR-GARCH

Figure 6 – S&P 500 : Leverage coefficients l_i for $i = 1, \dots, 70$.

Second, although the MSM process has been originally proposed for exchange rate series, the HDM model decidedly outperforms it in terms of information criteria.

3.3 Out-of-sample results

In this section, we carry out a forecasting exercise over the last three years of our data sample periods in order to compare the predictions of the HDM models and of some competitors on short- and long-run forecasting periods. Each time we move forward by one day in the in-sample period, the models are re-estimated, and cumulative variance forecasts, $\sum_{i=1}^h \hat{V}_{t+i}$, are computed for time horizons of $h = 1, 5, 10, 25, 50, 75$ and 100. These forecasts are then

(a) HDM and MSM conditional volatilities

(b) HDM conditional volatilities with/out leverage

Figure 7 – S&P 500 : The left panel shows the MSM conditional volatility in relation to the HDM conditional volatility with leverage effect. The former is displayed in black while the grey dynamic comes from the HDM model. The right panel shows the HDM conditional volatilities. The HDM process without leverage is displayed in black and the HDM model with leverage in grey.

compared to their associated observed values $\sum_{i=1}^h x_{t+i}$, where x_t denotes either the realized kernel variance or the squared percentage log-return.¹⁶

Table 4 presents the root mean squared errors (RMSE) divided by $\sqrt{h}$ (to document the loss functions at a similar scale) between the predicted and observed values for the different forecast horizons considered. Focusing on log-returns time series, the HDM model dominates all the competitors at all horizons for the S&P 500 index and at long horizons (starting from $h = 10$) for the NASDAQ series. Regarding the exchange rate, the HDM process dominates the MSM and the GARCH models but does not produce as good predictions as MS-GARCH models. Looking next at realized variances, with the exception of the HAR model with gamma-distributed innovations on the USD/EUR realized variances, the HDM model dominates all the alternatives. The outperformance is in fact very large for S&P 500 and NASDAQ data at long horizons. This outlines that the HDM model is particularly well suited to capture the long-range dependence in volatility.

¹⁶ As in many papers such as in Calvet and Fisher (2004), we use cumulative forecasts to measure the aggregate error over multiple period predictions.

Table 3 – Log-likelihoods and information criteria.

Percentage log-returns							
Models	MSM	GARCH- t	GJR- t	MS-GARCH- t	MS-GJR- t	HDM	HDM-lev
Np	4	4	5	9	11	7	9
NASDAQ							
log-lik	−7261.3	−7259.5	−7197.5	−7248.4	−7175.5	−7253.9	−7180.2
AIC	−7265.3	−7263.5	−7202.5	−7257.4	−7186.5	−7260.9	−7189.2
BIC	−7278.0	−7276.2	−7218.3	−7285.9	−7221.3	−7283.0	−7217.7
USD/EUR							
log-lik	−3762.2	−3747.5	−3745.6	−3740.7	−3740.5	−3738.6	−3737.7
AIC	−3766.2	−3751.5	−3750.6	−3749.7	−3751.5	−3745.6	−3746.7
BIC	−3778.9	−3764.2	−3766.4	−3778.2	−3786.3	−3767.8	−3775.1
Realized kernel variances							
Models	HAR	MEM	AMEM	MS-MEM	MS-AMEM	HDM	HDM-lev
Np	5	4	5	9	11	8	10
NASDAQ							
log-lik	−1640.6	−1669.0	−1596.7	−1464.6	−1389.8	−1430.3	−1225.5
AIC	−1645.6	−1673.0	−1601.7	−1473.6	−1400.8	−1438.3	−1235.5
BIC	−1661.4	−1685.6	−1617.5	−1502.1	−1435.6	−1463.6	−1267.2
USD/EUR							
log-lik	1174.1	1150.0	1151.8	1266.7	1269.5	1325.3	1335.5
AIC	1169.1	1146.0	1146.8	1257.7	1258.5	1317.3	1325.5
BIC	1154.7	1134.5	1132.4	1231.8	1226.8	1294.3	1296.7

log-lik: Maximum of the log-likelihood; AIC: Akaike Information criterion; BIC: Bayesian information criterion; The highest values are bolded.

With respect to the realized variance series, the striking performance of the HDM model is largely attributable to the inclusion of the leverage effect component.¹⁷ However, note that the HDM model without leverage still compares favorably well at shorter horizons (less than 25), which implies that the contribution of the other components is also important.

4 Conclusion

We propose the high dimensional Markov (HDM) model, a new volatility process that is suited for financial returns and realized variances. We specify the variance as a high dimen-

¹⁷ To forecast the realized kernel variances with a leverage effect component, we assume that the percentage log-returns $r_t \sim \text{i.i.d. } N(\hat{\mu}, \hat{\sigma}^2)$ where $\hat{\mu}$ denotes the empirical mean and $\hat{\sigma}^2$ stands for the empirical variance over the last three financial years. The benefit of the leverage effect could be further improved by considering a bivariate model for returns and realized variances, but we leave this extension to further research.

sional Markov-chain that is decomposed into a product of three hidden components that can be economically interpreted. In particular, the jump process captures the reaction of the financial market to non-persistent news whereas the Markov chain component reflects news with a long-lasting impact. The last component controls for the leverage effect. The specification of the latter process differs from what is typically found in the literature and exhibits the attractive property of being extractable. These three processes are parsimoniously and coherently specified and create a continuum of volatility states altogether. In addition, we derive the moments of the process and show that the autocovariance function can exhibit a slower decay than traditional hidden Markov model thanks to the multiplicity of the second largest eigenvalue of the transition probability matrix. This property seems beneficial empirically as we show that the HDM model dominates the MSM process as well as the GARCH- t and GJR- t models while competing favorably well with the MS-GARCH- t and the MS-GJR- t processes in terms of in-sample criteria such as the AIC and BIC on three financial data sets. Moreover, the HDM process also outperforms standard realized variances models (i.e., HAR, MEM, MS-MEM and AMEM) according to these criteria on three realized kernel variance series. Regarding the predictive performance, the HDM process competes very well with several alternatives in short forecasting horizons (less than 5 days). In long-run horizons, it significantly improves over the other models especially when the leverage component is active.

We view this variance modeling attempt with a high dimensional hidden Markov chain as very promising since many extensions can be entertained. We can for instance add a fourth component which would take into account the trading volume or we could introduce correlated components since the diverse news seem to be related. Additionally, a multivariate extension in the spirit of Calvet et al. (2006) could be undertaken.

SUPPLEMENTARY MATERIAL

5 Proofs

5.1 Properties of the Kronecker product

We first review some properties of the Kronecker product.

Lemma 1. *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{p \times q}$, $\mathbf{C} \in \mathbb{R}^{n \times k}$ and $\mathbf{D} \in \mathbb{R}^{q \times r}$. Then,*

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}).$$

Proof. See Broxson (2006, Theorem 7). □

Lemma 2. *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$. If λ_A is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{v}_A \in \mathbb{R}^n$ and if λ_B is an eigenvalue of $\mathbf{B}$ with corresponding eigenvector $\mathbf{v}_B \in \mathbb{R}^m$, then $\lambda_A \lambda_B$ is an eigenvalue of $\mathbf{A} \otimes \mathbf{B}$ with corresponding eigenvector $\mathbf{v}_A \otimes \mathbf{v}_B \in \mathbb{R}^{mn}$.*

Proof. See Broxson (2006, Theorem 15). □

Lemma 3. *Let $\mathbf{A}$ and $\mathbf{B}$ be non-negative matrices (i.e., all elements of $\mathbf{A}$ and $\mathbf{B}$ are non-negative). Then, $\mathbf{A} \otimes \mathbf{B}$ is diagonalizable if and only if both $\mathbf{A}$ and $\mathbf{B}$ are diagonalizable.*

Proof. See Broxson (2006, Theorem 57). □

5.2 Eigenvalues and eigenvectors

Lemma 4 (Eigenvalues and eigenvectors of the 2×2 matrix $\mathbf{P}$).

- (i) *The eigenvalues of $\mathbf{P}$ are 1 and $\gamma = 2p - 1$, where $p \in (0, 1)$.*
- (ii) *The corresponding right eigenvectors are $\mathbf{v} = (1, 1)'$ and $\mathbf{v}_\gamma = (-1, 1)'$.*
- (iii) *The corresponding left eigenvectors are $\mathbf{w} = (1/2, 1/2)' = 2^{-1}\mathbf{v}$ and $\mathbf{w}_\gamma = (-1/2, 1/2)' = 2^{-1}\mathbf{v}_\gamma$.*
- (iv) *The right and left eigenvectors defined in items (ii) and (iii) satisfy $\mathbf{w}'\mathbf{v} = \mathbf{w}'_\gamma\mathbf{v}_\gamma = 1$ and $\mathbf{w}'\mathbf{v}_\gamma = \mathbf{w}'_\gamma\mathbf{v} = 0$ (i.e., both sets of eigenvectors are orthogonal).*

(v) $\mathbf{P}$ is diagonalizable, and we may write $\mathbf{P} = \mathbf{v}\mathbf{w}' + \gamma\mathbf{v}_\gamma\mathbf{w}'_\gamma$.

Proof. Items (i)–(v) follow directly from basic linear algebra theory. $\square$

Lemma 5 (Eigenvalues and eigenvectors of the matrix $\mathbf{P}_C = \mathbf{P}^{\otimes N}$). *Let $\{\lambda_i\}_{i=1}^{2^N}$ represent the set of eigenvalues of the matrix $\mathbf{P}_C$, such that $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_{2^N}|$, and let $\{\mathbf{v}_i\}_{i=1}^{2^N}$ and $\{\mathbf{w}_i\}_{i=1}^{2^N}$ denote, respectively, the associated sets of right and left eigenvectors.*

Case $p \neq 1/2$ (and therefore $\gamma \neq 0$).

- (i) The set, $\{\lambda_i\}_{i=1}^{2^N}$, contains $N + 1$ distinct non-zero eigenvalues, namely $\{\gamma^j\}_{j=0}^N$, each with an algebraic multiplicity of $\binom{N}{j}$. In particular, we have $\lambda_1 = 1$ and N second largest eigenvalues (in absolute terms): $\lambda_i = \gamma$, $i = 2, \dots, N + 1$.
- (ii) The right eigenvectors, $\{\mathbf{v}_i\}_{i=1}^{2^N}$, can be obtained by way of a Kronecker product of N vectors among $\mathbf{v} = (1, 1)'$ and $\mathbf{v}_\gamma = (-1, 1)'$. Therefore, these vectors can only contain values in the set $\{-1, 1\}$. In particular, we have $\mathbf{v}_1 = \mathbf{v}^{\otimes N} = \mathbf{1}_{2^N}$.
- (iii) The left eigenvectors, $\{\mathbf{w}_i\}_{i=1}^{2^N}$, satisfy the relationship,

$$\mathbf{w}_i = 2^{-N} \mathbf{v}_i, \quad i = 1, \dots, 2^N.$$

In particular, we have $\mathbf{w}_1 = 2^{-N} \mathbf{1}_{2^N}$.

- (iv) Let the matrices $\mathbf{V}_C$ and $\mathbf{W}_C$ contain, respectively, the right and left eigenvectors of $\mathbf{P}_C$ (as defined in items (ii) and (iii)) along their columns. Then, $\mathbf{W}'_C \mathbf{V}_C = \mathbf{I}_{2^N}$, where $\mathbf{I}_{2^N}$ is the 2^N -dimensional identity matrix (note that this implies $\mathbf{W}'_C = \mathbf{V}_C^{-1}$). Specifically, the right and left eigenvectors satisfy:

$$\mathbf{w}'_i \mathbf{v}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

These two sets of eigenvectors are therefore orthogonal.

- (v) $\mathbf{P}_C$ is diagonalizable and we may write $\mathbf{P}_C = \mathbf{V}_C \mathbf{\Lambda} \mathbf{W}'_C$, where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues $\{\lambda_i\}_{i=1}^{2^N}$ on its diagonal.

Case $p = 1/2$ (and therefore $\gamma = 0$).

(vi) $\lambda_1 = 1$ and $\lambda_i = 0$, $i = 2, \dots, 2^N$, that is, the algebraic multiplicity of the eigenvalue 0 is $2^N - 1$. The associated sets of right and left eigenvectors are identical to the ones defined in items (ii) and (iii), respectively.

(vii) $\mathbf{P}_C = \mathbf{1}_{2^N} \boldsymbol{\pi}'_C$, where $\boldsymbol{\pi}_C = 2^{-N} \mathbf{1}_{2^N}$.

Proof. Items (i)–(iii) follow from Lemma 2. Regarding item (iv), we have,

$$\mathbf{w}'_i \mathbf{v}_i = 2^{-N} \mathbf{v}'_i \mathbf{v}_i = 2^{-N} 2^N = 1, \quad i = 1, \dots, 2^N,$$

because $\mathbf{v}_i$ only contains values in the set $\{-1, 1\}$. To see why $\mathbf{w}'_i \mathbf{v}_j = 0$, if $i \neq j$, note that Lemma 1 implies that

$$(\mathbf{w} \otimes \mathbf{w})'(\mathbf{v} \otimes \mathbf{v}_\gamma) = (\mathbf{w}'\mathbf{v}) \otimes (\mathbf{w}'\mathbf{v}_\gamma) = 0, \quad \text{and}$$

$$(\mathbf{w} \otimes \mathbf{w}_\gamma)'(\mathbf{v} \otimes \mathbf{v}) = (\mathbf{w}'\mathbf{v}) \otimes (\mathbf{w}'_\gamma \mathbf{v}) = 0.$$

When $i \neq j$, the term $\mathbf{w}'\mathbf{v}_\gamma = 0$ or $\mathbf{w}'_\gamma \mathbf{v} = 0$ appears at least once when using the above property of the Kronecker product on $\mathbf{w}'_i \mathbf{v}_j$, which explains why $\mathbf{w}'_i \mathbf{v}_j = 0$, if $i \neq j$. Moreover, item (v) follows directly from Lemma 3 and from the spectral decomposition.

Finally, item (vi) follows directly from Lemma 2, whereas item (vii) is easily obtained by noting that $\mathbf{P}_C = \mathbf{P}^{\otimes N} = (2^{-1} \mathbf{1}_2 \mathbf{1}'_2)^{\otimes N} = 2^{-N} \mathbf{1}_{2^N} \mathbf{1}'_{2^N}$.

□

Lemma 6 (Eigenvalues and eigenvectors of the matrix $\mathbf{P}_M = \mathbf{1}_N \boldsymbol{\pi}'_M$).

(i) 1 is an eigenvalue of $\mathbf{P}_M$ with an algebraic multiplicity of 1. $\mathbf{1}_N$ and $\boldsymbol{\pi}_M$ are, respectively, right and left eigenvectors corresponding to this eigenvalue.

(ii) 0 is an eigenvalue of $\mathbf{P}_M$ with an algebraic multiplicity of $N - 1$.

(iii) $\mathbf{P}_M$ is diagonalizable.

Proof. Because $\boldsymbol{\pi}_M$ is a vector of probabilities, we have $\boldsymbol{\pi}'_M \mathbf{1}_N = 1$, and hence

$$\mathbf{P}_M \mathbf{1}_N = \mathbf{1}_N \boldsymbol{\pi}'_M \mathbf{1}_N = \mathbf{1}_N, \quad \text{and}$$

$$\boldsymbol{\pi}'_M \mathbf{P}_M = \boldsymbol{\pi}'_M \mathbf{1}_N \boldsymbol{\pi}'_M = \boldsymbol{\pi}'_M.$$

Therefore, $\mathbf{1}_N$ and $\boldsymbol{\pi}_M$ are, respectively, right and left eigenvectors associated with the eigenvalue 1.

Since $\mathbf{P}_M$ is an $N \times N$ matrix with N repeated rows, it follows that $\mathbf{P}_M$ has rank 1 and nullity of $N - 1$. The singularity of $\mathbf{P}_M$ implies that this matrix has an eigenvalue of 0. The geometric multiplicity of this eigenvalue corresponds to the nullity of $\mathbf{P}_M$, which is $N - 1$. It follows that its algebraic multiplicity is also $N - 1$. To see why, note that on one hand, the algebraic multiplicity of an eigenvalue cannot be smaller than its geometric multiplicity, and on the other hand, the algebraic multiplicity of the eigenvalue 0 is at most $N - 1$ because we know that 1 is an eigenvalue.

Finally, the fact that the algebraic and geometric multiplicities of each of the eigenvalues are equal implies that the matrix $\mathbf{P}_M$ is diagonalizable. $\square$

Lemma 7 (Eigenvalues and eigenvectors of the matrix $\mathbf{P}_V = \mathbf{P}_C \otimes \mathbf{P}_M$).

Case $p \neq 1/2$ (and therefore $\gamma \neq 0$).

- (i) The set of non-zero eigenvalues of $\mathbf{P}_V$ corresponds to the set of all eigenvalues of $\mathbf{P}_C$, denoted by $\{\lambda_i\}_{i=1}^{2^N}$ (see item(i) of Lemma 5).
- (ii) The sets of right and left eigenvectors of $\mathbf{P}_V$ associated with these eigenvalues correspond to $\{\mathbf{v}_i \otimes \mathbf{1}_N\}_{i=1}^{2^N}$, and $\{\mathbf{w}_i \otimes \boldsymbol{\pi}_M\}_{i=1}^{2^N}$, respectively, where $\{\mathbf{v}_i\}_{i=1}^{2^N}$, and $\{\mathbf{w}_i\}_{i=1}^{2^N}$ denote the right and left eigenvectors of $\mathbf{P}_C$, respectively.
- (iii) Let the matrices $\mathbf{V}_V$ and $\mathbf{W}_V$ contain, respectively, the right and left eigenvectors of $\mathbf{P}_V$ associated with non-zero eigenvalues (as defined in item (ii)) along their columns. Accordingly, $\mathbf{V}_V$ and $\mathbf{W}_V$ are matrices of dimension $(N \cdot 2^N) \times 2^N$. Then, $\mathbf{W}'_V \mathbf{V}_V = \mathbf{I}_{2^N}$.

Specifically, the right and left eigenvectors in item (ii) satisfy:

$$(\mathbf{w}_i \otimes \boldsymbol{\pi}_M)'(\mathbf{v}_j \otimes \mathbf{1}_N) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

These two sets of eigenvectors are therefore orthogonal.

(iv) $\mathbf{P}_V$ is diagonalizable and we may write $\mathbf{P}_V = \mathbf{V}_V \boldsymbol{\Lambda} \mathbf{W}_V'$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues $\{\lambda_i\}_{i=1}^{2^N}$ on its diagonal.

Case $p = 1/2$ (and therefore $\gamma = 0$).

(v) $\mathbf{P}_V$ has a single non-zero eigenvalue that is equal to 1. The right and left eigenvectors associated with this eigenvalue correspond to $(\mathbf{v}_1 \otimes \mathbf{1}_N)$ and $(\mathbf{w}_1 \otimes \boldsymbol{\pi}_M)$, respectively (as defined in item (ii)).

(vi) $\mathbf{P}_V = \mathbf{1}_{N \cdot 2^N} \boldsymbol{\pi}_V'$, where $\boldsymbol{\pi}_V = \boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M$.

Proof. First, note that if $p \neq 1/2$ then all of the eigenvalues of $\mathbf{P}_C$, $\{\lambda_i\}_{i=1}^{2^N}$, are non-zero (see item (i) of Lemma 5), and observe that $\mathbf{P}_M$ possesses a single non-zero eigenvalue which is equal to 1 (see item (i) of Lemma 6). It then follows from Lemma 2 that the set of non-zero eigenvalues of $\mathbf{P}_V = \mathbf{P}_C \otimes \mathbf{P}_M$ must correspond to $\{\lambda_i\}_{i=1}^{2^N}$. This shows item (i).

Because the vectors $\mathbf{1}_N$ and $\boldsymbol{\pi}_M$ are, respectively, the right and left eigenvectors of $\mathbf{P}_M$ associated with the eigenvalue 1 (see item (i) of Lemma 6), it then follows, once again from Lemma 2, that $\{\mathbf{v}_i \otimes \mathbf{1}_N\}_{i=1}^{2^N}$, and $\{\mathbf{w}_i \otimes \boldsymbol{\pi}_M\}_{i=1}^{2^N}$ are, respectively, the right and left eigenvectors of $\mathbf{P}_V$ associated with the eigenvalues $\{\lambda_i\}_{i=1}^{2^N}$. This shows item (ii).

Using Lemma 1, we can write,

$$(\mathbf{w}_i \otimes \boldsymbol{\pi}_M)'(\mathbf{v}_j \otimes \mathbf{1}_N) = (\mathbf{w}_i' \mathbf{v}_j) \otimes (\boldsymbol{\pi}_M' \mathbf{1}_N) = \mathbf{w}_i' \mathbf{v}_j, \quad i, j = 1, \dots, 2^N.$$

Item (iii) then follows directly from item (iv) of Lemma 5.

Item (iv) follows from Lemma 3, because $\mathbf{P}_C$ and $\mathbf{P}_M$ are both diagonalizable (see item (v) of Lemma 5 and item (iii) of Lemma 6), and from the spectral decomposition. Note that

only non-zero eigenvalues and its associated eigenvectors are relevant when performing the spectral decomposition.

Finally, item (v) is a direct application of Lemma 2 on the results of Lemmas 5 and 6, whereas item (vi) is obtained from Lemma 1 by noting that $\mathbf{P}_V = \mathbf{P}_C \otimes \mathbf{P}_M = (\mathbf{1}_{2^N} \boldsymbol{\pi}'_C) \otimes (\mathbf{1}_N \boldsymbol{\pi}'_M) = \mathbf{1}_{N \cdot 2^N} (\boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M)'$. $\square$

5.3 Proof of Theorem 1

Before proving Theorem 1, we show the following two lemmas.

Lemma 8. *Let $\boldsymbol{\pi}$ be a n -dimensional column vector of probabilities such that $\boldsymbol{\pi}' \mathbf{1}_n = 1$, and let $\mathbf{Q} = \mathbf{1}_n \boldsymbol{\pi}'$. Then, $\mathbf{Q}^k = \mathbf{Q}$, for $k = 1, 2, \dots$*

Proof. The result is trivially satisfied for $k = 1$. We have $\mathbf{Q}^2 = (\mathbf{1}_n \boldsymbol{\pi}') (\mathbf{1}_n \boldsymbol{\pi}') = \mathbf{1}_n (\boldsymbol{\pi}' \mathbf{1}_n) \boldsymbol{\pi}' = \mathbf{Q}$. The result then follows by induction. $\square$

Lemma 9 (Properties of the matrices $\mathbf{v}_i \mathbf{w}'_i$, $i = 1, \dots, 2^N$). *Let $\{\mathbf{v}_i\}_{i=1}^{2^N}$ and $\{\mathbf{w}_i\}_{i=1}^{2^N}$ be two sets of eigenvectors as defined in Lemma 5.*

(i) *We have,*

$$\mathbf{v}_i \mathbf{w}'_i = 2^{-N} \mathbf{v}_i \mathbf{v}'_i, \quad i = 1, \dots, 2^N.$$

(ii) *Let $|\mathbf{v}_i \mathbf{w}'_i|$ denote the matrix of absolute elements of $\mathbf{v}_i \mathbf{w}'_i$. Then,*

$$|\mathbf{v}_i \mathbf{w}'_i| = 2^{-N} \mathbf{1}_{2^N} \mathbf{1}'_{2^N}, \quad i = 1, \dots, N.$$

(iii) *Let $\text{diag}\{\mathbf{v}_i \mathbf{w}'_i\}$ denote the column vector of diagonal elements of the matrix $\mathbf{v}_i \mathbf{w}'_i$. Then,*

$$\text{diag}\{\mathbf{v}_i \mathbf{w}'_i\} = 2^{-N} \mathbf{1}_{2^N}, \quad i = 1, \dots, 2^N.$$

Proof. Lemma 5 shows that $\mathbf{w}_i = 2^{-N} \mathbf{v}_i$, $i = 1, \dots, 2^N$, and that the vectors $\{\mathbf{v}_i\}_{i=1}^{2^N}$ only contain values in the set $\{-1, 1\}$. Items (i)–(iii) follow directly from these facts. $\square$

We now prove Theorem 1.

Proof. First, consider the case $p = 1/2$ (and therefore $\gamma = 0$). In this specific case we have $\mathbf{P}_V = \mathbf{\Pi}_V$ (see item (vi) of Lemma 7), and therefore, by Lemma 8, $\mathbf{P}_V^k = \mathbf{\Pi}_V$, $k = 1, 2, \dots$. The results of the theorem then follow trivially from this fact.

Now, consider the case $p \neq 1/2$ (and therefore $\gamma \neq 0$). Since we assumed that $p \in (0, 1)$ and $q \in (0, 1)$, then $\mathbf{P}_V$ is a positive matrix. The Perron-Frobenius theorem (Seneta, 2006) then implies that

$$\lim_{k \rightarrow \infty} \mathbf{P}_V^k = (\mathbf{v}_1 \otimes \mathbf{1}_N) (\mathbf{w}_1 \otimes \boldsymbol{\pi}_M)',$$

where $(\mathbf{v}_1 \otimes \mathbf{1}_N)$ and $(\mathbf{w}_1 \otimes \boldsymbol{\pi}_M)$, are, respectively, the right and left eigenvectors of $\mathbf{P}_V$ associated with the eigenvalue 1, normalized such that $(\mathbf{w}_1 \otimes \boldsymbol{\pi}_M)' (\mathbf{v}_1 \otimes \mathbf{1}_N) = 1$ (see Lemma 7). From Lemma 5, we obtain that $\mathbf{v}_1 = \mathbf{1}_{2^N}$ and $\mathbf{w}_1 = 2^{-N} \mathbf{1}_{2^N} = \boldsymbol{\pi}_C$, which implies that the left eigenvector, $(\mathbf{w}_1 \otimes \boldsymbol{\pi}_M)$, corresponds to the vector $(\boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M) = \boldsymbol{\pi}_V$. It then immediately follows that

$$\lim_{k \rightarrow \infty} \mathbf{P}_V^k = \mathbf{1}_{N \cdot 2^N} \boldsymbol{\pi}_V'.$$

This completes the proof of part (i) of the theorem.

The spectral decomposition of $\mathbf{P}_V$ presented in Lemma 7 implies the following relationship:

$$\begin{aligned} |\mathbf{P}_V^k - \mathbf{\Pi}_V| &= \left| \sum_{i=2}^{2^N} \lambda_i^k (\mathbf{v}_i \otimes \mathbf{1}_N) (\mathbf{w}_i \otimes \boldsymbol{\pi}_M)' \right| \\ &= \left| \sum_{i=2}^{2^N} \lambda_i^k (\mathbf{v}_i \mathbf{w}_i') \otimes (\mathbf{1}_N \boldsymbol{\pi}_M') \right| && \text{[Lemma 1]} \\ &\leq \sum_{i=2}^{2^N} |\lambda_i^k| |\mathbf{v}_i \mathbf{w}_i'| \otimes (\mathbf{1}_N \boldsymbol{\pi}_M') \\ &= \sum_{i=2}^{2^N} |\lambda_i^k| (2^{-N} \mathbf{1}_{2^N} \mathbf{1}_{2^N}') \otimes (\mathbf{1}_N \boldsymbol{\pi}_M') && \text{[Lemma 9]} \\ &= \left(2^{-N} \sum_{i=2}^{2^N} |\lambda_i^k| \right) \cdot (\mathbf{1}_{2^N} \mathbf{1}_{2^N}') \otimes (\mathbf{1}_N \boldsymbol{\pi}_M') \\ &= 2^{-N} \left(\sum_{j=1}^N \binom{N}{j} (|\gamma|^k)^j \right) \cdot (\mathbf{1}_{2^N} \mathbf{1}_{2^N}') \otimes (\mathbf{1}_N \boldsymbol{\pi}_M') && \text{[Lemma 5]} \\ &= 2^{-N} \left((1 + |\gamma|^k)^N - 1 \right) \cdot (\mathbf{1}_{2^N} \mathbf{1}_{2^N}') \otimes (\mathbf{1}_N \boldsymbol{\pi}_M') && \text{[Binomial theorem]} \end{aligned}$$

$$\begin{aligned}
&= 2^{-N} \left((1 + |\gamma|^k)^N - 1 \right) \cdot (\mathbf{1}_{2^N} \otimes \mathbf{1}_N) (\mathbf{1}'_{2^N} \otimes \boldsymbol{\pi}'_M) && [\text{Lemma 1}] \\
&= 2^{-N} \left((1 + |\gamma|^k)^N - 1 \right) \cdot \mathbf{1}_{N \cdot 2^N} (\mathbf{1}'_{2^N} \otimes \boldsymbol{\pi}'_M) \\
&= \left((1 + |\gamma|^k)^N - 1 \right) \cdot \mathbf{1}_{N \cdot 2^N} \left(2^{-N} \mathbf{1}'_{2^N} \otimes \boldsymbol{\pi}'_M \right) \\
&= \left((1 + |\gamma|^k)^N - 1 \right) \cdot \mathbf{1}_{N \cdot 2^N} (\boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M)' \\
&= \left((1 + |\gamma|^k)^N - 1 \right) \cdot \boldsymbol{\Pi}_V. && (18)
\end{aligned}$$

Equation (9) of the theorem then follows directly by noting that all row sums of $\boldsymbol{\Pi}_V$ are equal to 1, whereas equation (10) of the theorem is due to the fact that the upper bound in equation (18) is actually attained on the diagonal of $(\mathbf{P}_V^k - \boldsymbol{\Pi}_V)$ when $\gamma > 0$. To see why the latter is true, consider

$$\begin{aligned}
\text{diag} \{ \mathbf{P}_V^k - \boldsymbol{\Pi}_V \} &= \text{diag} \left\{ \sum_{i=2}^{2^N} \lambda_i^k (\mathbf{v}_i \mathbf{w}'_i) \otimes (\mathbf{1}_N \boldsymbol{\pi}'_M) \right\} \\
&= \sum_{i=2}^{2^N} \lambda_i^k \text{diag} \{ \mathbf{v}_i \mathbf{w}'_i \} \otimes \text{diag} \{ \mathbf{1}_N \boldsymbol{\pi}'_M \} \\
&= \sum_{i=2}^{2^N} \lambda_i^k \left(2^{-N} \mathbf{1}_{2^N} \right) \otimes \boldsymbol{\pi}_M && [\text{Lemma 9}] \\
&= \left(\sum_{j=1}^N \binom{N}{j} (\gamma^k)^j \right) \cdot (\boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M) \\
&= \left((1 + \gamma^k)^N - 1 \right) \cdot \boldsymbol{\pi}_V. && [\text{Binomial theorem}]
\end{aligned}$$

Consequently, for $\gamma > 0$, we must have

$$\| \mathbf{P}_V^k - \boldsymbol{\Pi}_V \|_{\max} = \left((1 + \gamma^k)^N - 1 \right) \| \boldsymbol{\pi}_V \|_{\infty}.$$

Because $\boldsymbol{\pi}_V = \boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M$, where $\boldsymbol{\pi}_C = 2^{-N} \mathbf{1}_{2^N}$ and $\boldsymbol{\pi}_M$ only contains elements in the set $\{q/(N-1), 1-q\}$ (see equation (7)), then it follows directly that $\| \boldsymbol{\pi}_V \|_{\infty} = 2^{-N} \max\{q/(N-1), 1-q\}$. This completes the proof of part (ii) of the theorem.

Finally, Lemma 7 shows that γ is the largest eigenvalue of $\mathbf{P}_V$ (in absolute terms) that is smaller than 1, and that it has an algebraic multiplicity of N . Part (iii) of the theorem is then a direct consequence of Seneta (2006, Theorem 1.2). $\square$

5.4 Proof of Proposition 1

We prove Proposition 1.

Proof. Item (i) is easily obtained by noting that

$$\begin{aligned}
\text{Cov}(x_t, x_{t+k}) &= \mathbb{E}[x_t x_{t+k}] - \mathbb{E}[x_t] \mathbb{E}[x_{t+k}] \\
&= \mathbb{E}[V_t V_{t+k}] \mathbb{E}[\eta_t] \mathbb{E}[\eta_{t+k}] - \mathbb{E}[V_t] \mathbb{E}[\eta_t] \mathbb{E}[V_{t+k}] \mathbb{E}[\eta_{t+k}] \\
&= \mathbb{E}[V_t V_{t+k}] - \mathbb{E}[V_t] \mathbb{E}[V_{t+k}] \\
&= \text{Cov}(V_t, V_{t+k}).
\end{aligned}$$

With respect to item (ii), first let $p_k = \Pr(C_{t+k}^{(i)} = 1 \mid C_t^{(i)} = 1)$ for $k = 1, 2, \dots$, and note that the symmetry of the ON and OFF states in the matrix $\mathbf{P}$ implies that p_k is also equal to $\Pr(C_{t+k}^{(i)} = c_i \mid C_t^{(i)} = c_i)$, and that $\Pr(C_t^{(i)} = 1) = \Pr(C_t^{(i)} = c_i) = 1/2$, for $i = 1, \dots, N$. Since item (v) of Lemma 4 entails that $p_k = 1/2 + \gamma^k/2$, where $\gamma = 2p - 1$, it then follows that, for $i = 1, \dots, N$ and $k = 1, 2, \dots$,

$$\begin{aligned}
\frac{\mathbb{E}[C_t^{(i)} C_{t+k}^{(i)}]}{\mathbb{E}[C_t^{(i)}] \mathbb{E}[C_{t+k}^{(i)}]} &= \frac{1 \cdot p_k/2 + c_i \cdot (1 - p_k)/2 + c_i \cdot (1 - p_k)/2 + c_i^2 \cdot p_k/2}{(c_i + 1)/2 \cdot (c_i + 1)/2} \\
&= \frac{(c_i + 1)^2/4 + \gamma^k(c_i - 1)^2/4}{(c_i + 1)^2/4} \\
&= 1 + \phi_i \gamma^k.
\end{aligned}$$

The autocovariance structure of $\{V_t\}$ can now be derived as follows:

$$\begin{aligned}
\text{Cov}(V_t, V_{t+k}) &= \sigma^4 (\mathbb{E}[(C_t M_t)(C_{t+k} M_{t+k})] - \mathbb{E}[C_t M_t] \mathbb{E}[C_{t+k} M_{t+k}]) \\
&= \sigma^4 (\mathbb{E}[C_t C_{t+k}] - 1) \\
&= \sigma^4 \left(\prod_{i=1}^N \left(\frac{\mathbb{E}[C_t^{(i)} C_{t+k}^{(i)}]}{\mathbb{E}[C_t^{(i)}] \mathbb{E}[C_{t+k}^{(i)}]} \right) - 1 \right) \\
&= \sigma^4 \left(\prod_{i=1}^N (1 + \phi_i \gamma^k) - 1 \right), \quad k = 1, 2, \dots
\end{aligned}$$

Now, let us consider item (iii). We have,

$$\begin{aligned}
\text{Var}(x_t) &= \mathbb{E}[x_t^2] - (\mathbb{E}[x_t])^2 \\
&= \mathbb{E}[V_t^2]\mathbb{E}[\eta_t^2] - (\mathbb{E}[V_t]\mathbb{E}[\eta_t])^2 \\
&= \mathbb{E}[V_t^2]\mathbb{E}[\eta_t^2] - (\mathbb{E}[V_t])^2 \\
&= \sigma^4 \left(\mathbb{E}[C_t^2]\mathbb{E}[M_t^2]\mathbb{E}[\eta_t^2] - 1 \right),
\end{aligned}$$

where,

$$\begin{aligned}
\mathbb{E}[C_t^2] &= \prod_{i=1}^N \left(\frac{\mathbb{E} \left[\left(C_t^{(i)} \right)^2 \right]}{\left(\mathbb{E}[C_t^{(i)}] \right)^2} \right) \\
&= \prod_{i=1}^N \left(\frac{1/2 + c_i^2/2}{(1/2 + c_i/2)^2} \right) \\
&= \prod_{i=1}^N (1 + \phi_i),
\end{aligned}$$

and,

$$\mathbb{E}[M_t^2] = m_0^2 \left(\frac{q}{N-1} \sum_{i=1}^{N-1} m_i^2 + (1-q) \right).$$

Finally, item (iv) follows directly from the definition of the correlation function. $\square$

6 Alternative models for daily log-returns

We briefly review the specifications of the processes used in the paper for modelling the daily log-returns. In this section, the time series $\{r_t\}$ denotes the demeaned daily financial log-returns.

6.1 MSM

The MSM model is fully specified in Section 2.4.

6.2 GARCH(1,1)

As mentioned in Section 6.3, the GARCH model is a particular case of the GJR process when $\delta = 0$.

6.3 GJR-GARCH(1,1)

We consider the standard GJR-GARCH(1,1) model with Student- t distribution specified as

$$\begin{aligned} r_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + (\alpha + \delta \mathbb{1}_{\{r_{t-1} < 0\}}) r_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

in which ϵ_t is driven by a standardized Student- t distribution with degree of freedom ν . The GJR-GARCH parameters are constrained to be positive in order to ensure the positivity of the variance. When $\delta = 0$, the model collapses into the standard GARCH(1,1) process also used in the paper.

Regarding the estimation, we impose the weak stationarity condition through the constraints $\alpha + 0.5\delta + \beta < 1$ and the initial variance is set to the theoretical unconditional one.

6.4 MS-GJR-GARCH(1,1)

The MS-GJR-GARCH(1,1) model that we consider is based on the specification introduced by Haas et al. (2004) and extended by Haas (2010). Let $\{X_t\}$ be a discrete time hidden Markov chain with finite state space $\{1, \dots, M\}$ and transition matrix $(p_{ij})_{i,j=1}^M$, where $p_{ij} = \Pr(X_t = j \mid X_{t-1} = i)$, for $i, j = 1, \dots, M$. The MS-GJR-GARCH(1,1) process is given by

$$\begin{aligned} r_t &= \sigma_{t,X_t} \epsilon_t, \\ \sigma_{t,j}^2 &= \omega_j + \alpha_j (|r_{t-1}| - \gamma_j r_{t-1})^2 + \beta_j \sigma_{t-1,j}^2, \end{aligned}$$

where ϵ_t is a Gaussian innovation with zero mean and unit variance and $\sigma_{t,j}^2$ represents the conditional variance process in regime j , $j = 1, \dots, M$, which evolves according to the GARCH structure.

The estimated MS-GJR-GARCH(1,1) process exhibits two regimes (i.e. $M = 2$) and the positivity constraint as well as the weakly stationary conditions are imposed to all the parameters. Lastly, the regime-specific initial variance is set to the corresponding theoretical unconditional one.

7 Alternative models for the realized volatility

In this section, we provide a brief review of the benchmark models used in the paper. The time series of daily realized variances is denoted as $\{RV_t\}$.

7.1 HAR

The standard HAR model is given by

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^{(5)} + \beta_3 RV_{t-1}^{(22)} + \epsilon_t,$$

with $RV_{t-1}^{(n)} = (RV_{t-1} + RV_{t-2} + \dots + RV_{t-n})/n$ and with $\epsilon_t \sim N(0, \sigma^2)$.

7.2 MEM

The MEM is a particular case of the AMEM when γ is set to 0 (see Section 7.3).

7.3 AMEM

The specification of the AMEM is given by

$$\begin{aligned} RV_t &= \mu_t \epsilon_t, \quad \text{with } \epsilon \sim G(a, 1/a) \\ \mu_t &= \omega + (\alpha + \gamma \mathbf{1}_{\{r_{t-1} < 0\}}) RV_{t-1} + \beta \mu_{t-1}, \end{aligned}$$

where ϵ_t is driven by a Gamma distribution with mean equal to one and variance equal to $\frac{1}{a}$. Positivity constraints (i.e. $\omega \geq 0, \alpha \geq 0, \gamma \geq 0$ and $\beta \geq 0$) as well as a stationarity condition (i.e. $\alpha + 0.5\gamma + \beta < 1$) are assumed. The initial volatility μ_1 is set to the theoretical unconditional volatility given by $\frac{\omega}{1-\alpha-0.5\gamma-\beta}$.

7.4 MS-MEM

The MS-MEM is a particular case of the MS-AMEM. See Section 7.5 for more information.

7.5 MS-AMEM

Let $\{X_t\}$ be a discrete time hidden Markov chain with finite state space $\{1, \dots, M\}$ and transition matrix $(p_{ij})_{i,j=1}^M$, where $p_{ij} = \Pr(X_t = j \mid X_{t-1} = i)$, for $i, j = 1, \dots, M$. The MS-AMEM introduced in Gallo and Otranto (2015) is specified as

$$\begin{aligned} \text{RV}_t &= \mu_{t,X_t} \epsilon_t, \\ \mu_{t,X_t} &= \omega + \sum_{i=1}^M \omega_i \mathbb{1}_{\{X_t \leq i\}} + \alpha_{X_t} (\text{RV}_{t-1} - \mu_{t-1,X_{t-1}}) \\ &\quad \dots + \beta_{X_t} \mu_{t-1,X_{t-1}} + \gamma_{X_t} \mathbb{1}_{\{r_{t-1} < 0\}} (\text{RV}_{t-1} - \mu_{t-1,X_{t-1}}), \end{aligned}$$

in which $\{r_t\}$ stands for the demeaned daily log-returns and $\epsilon_t | X_t \sim G(a_{X_t}, \frac{1}{a_{X_t}})$.¹⁸ Positivity constraints as well as the stationary condition for each regime (i.e. $\alpha_i + \beta_i + \gamma_i/2 < 1 \forall i \in [1, M]$) are imposed to the parameters. As the model exhibits a path dependence problem, it is estimated with the Kim's filter (see Kim (1994)). The estimated MS-AMEM has a number of regimes set to two. Note that when $\gamma_i = 0 \forall i \in [1, M]$, the process collapses into a standard MS-MEM.

¹⁸ i.e. a Gamma distribution with mean equal to one and variance equal to $\frac{1}{a_{X_t}}$.

References

- Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89(4):701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- Ang, A. and Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20(2):163–182.
- Bauwens, L., Dufays, A., and Rombouts, J. V. K. (2014). Marginal likelihood for Markov-switching and change-point GARCH models. *Journal of Econometrics*, 178(part 3):508–522.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Broxson, B. J. (2006). The Kronecker product. Master’s thesis, University of North Florida. Department of Mathematics and Statistics. Paper 25. <http://digitalcommons.unf.edu/etd/2>.
- Calvet, L., Fearnley, M., Fisher, A. J., and Leippold, M. (2015). What is beneath the surface? Option pricing with multifrequency latent states. *Journal of Econometrics*, 187(2):498–511.
- Calvet, L. E. and Fisher, A. J. (2004). How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2(1):49–83.
- Calvet, L. E., Fisher, A. J., and Thompson, S. B. (2006). Volatility comovement: a multi-frequency approach. *Journal of Econometrics*, 131(1-2):179–215.

- Cordis, A. S. and Kirby, C. (2014). Discrete stochastic autoregressive volatility. *Journal of Banking & Finance*, 43:160–178.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.
- Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78.
- Dai, Q., Singleton, K. J., and Yang, W. (2007). Regime shifts in a dynamic term structure model of U.S. Treasury bond yields. *The Review of Financial Studies*, 20(5):1669–1706.
- Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.
- Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics*, 15(1):26–34.
- Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446.
- Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1-2):3–27.
- Fleming, J. and Kirby, C. (2013). Component-driven regime-switching volatility. *Journal of Financial Econometrics*, 11(2):263–301.
- Gallo, G. M. and Otranto, E. (2015). Forecasting realized volatility with changing average levels. *International Journal of Forecasting*, 31(3):620–634.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

- Goldfeld, S. M. and Quandt, R. E. (1973). A Markov model for switching regressions. *Journal of Econometrics*, 1(1):3–16.
- Gordon, N., Salmond, D., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107–113.
- Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.
- Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.
- Haas, M. (2010). Skew-normal mixture and Markov-switching GARCH processes. *Studies in Nonlinear Dynamics & Econometrics*, 14(4):Article 1.
- Haas, M., Mittnik, S., and Paoletta, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press, New Jersey.
- Kim, C. (1994). Dynamic linear models with Markov switching. *Journal of Econometrics*, 60:1–22.
- Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27:363–394.
- Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics*, 4(4):594–616.

- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36:394–419.
- Mikosch, T. and Starica, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics*, 86(1):378–390.
- Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics*, 28(2):275–290.
- Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94:590–599.
- Poskitt, D. S. and Chung, S.-H. (1996). Markov chain models, time series analysis and extreme value theory. *Advances in Applied Probability*, 28(2):405–425.
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, pages 257–286.
- Rossi, A. and Gallo, G. M. (2006). Volatility estimation via hidden Markov models. *Journal of Empirical Finance*, 13(2):203–230.
- Rydén, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13:217–244.
- Seneta, E. (2006). *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544].
- Starica, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87:503–522.

Table 4 – Prediction RMSE (divided by $\sqrt{h}$) computed over the last three years of the data sample period.

Horizons	1	5	10	25	50	75	100
Percentage log-returns							
S&P 500							
HDM-lev	1.33	1.85	2.19	2.82	3.40	3.64	3.74
MS-GJR- t	1.35	1.91	2.30	2.96	3.54	3.87	4.01
GJR- t	1.33	1.94	2.44	3.53	4.71	5.42	5.66
MS-GARCH- t	1.38	2.00	2.45	3.39	4.50	5.22	5.54
GARCH- t	1.38	2.01	2.46	3.41	4.58	5.36	5.69
MSM	1.38	2.01	2.46	3.42	4.52	5.30	5.58
NASDAQ							
HDM-lev	2.00	2.59	3.05	3.87	4.61	4.83	4.71
MS-GJR- t	1.96	2.57	3.15	4.33	5.44	6.10	6.41
GJR- t	1.99	2.69	3.37	5.03	7.19	8.73	9.25
MS-GARCH- t	2.06	2.88	3.69	5.60	8.47	10.92	12.52
GARCH- t	2.05	2.79	3.42	4.73	6.41	7.49	7.68
MSM	2.04	2.78	3.41	4.71	6.46	7.80	8.51
USD/EUR							
HDM	0.66	0.67	0.68	0.76	0.94	1.16	1.45
MS-GJR- t	0.66	0.66	0.66	0.70	0.81	0.99	1.24
GJR- t	0.66	0.67	0.68	0.76	0.93	1.18	1.49
MS-GARCH- t	0.66	0.67	0.67	0.73	0.85	1.03	1.26
GARCH- t	0.66	0.68	0.69	0.79	1.00	1.25	1.57
MSM	0.66	0.68	0.70	0.80	1.00	1.22	1.50
Realized kernel variances							
S&P 500							
HDM-lev	0.75	1.17	1.46	1.94	2.35	2.44	2.41
MS-AMEM	0.83	1.49	2.01	3.00	4.14	4.95	5.51
AMEM	0.83	1.48	2.00	2.95	4.00	4.71	5.18
MS-MEM	0.82	1.44	1.96	3.09	4.65	6.06	7.21
MEM	0.82	1.44	1.96	3.03	4.48	5.72	6.74
HAR	0.82	1.36	1.77	2.66	3.90	4.88	5.53
HAR-norm	0.80	1.28	1.65	2.45	3.62	4.58	5.36
NASDAQ							
HDM-lev	0.56	0.95	1.24	1.72	2.21	2.33	2.30
MS-AMEM	0.59	1.10	1.55	2.53	4.06	5.45	6.69
AMEM	0.58	1.08	1.51	2.41	3.76	4.95	5.98
MS-MEM	0.62	1.19	1.69	2.82	4.62	6.34	7.86
MEM	0.62	1.18	1.68	2.78	4.53	6.25	7.83
HAR	0.61	1.09	1.48	2.32	3.61	4.67	5.38
HAR-norm	0.60	1.07	1.45	2.38	3.85	5.19	6.31
USD/EUR							
HDM	0.31	0.61	0.80	1.40	2.36	3.24	3.87
MS-AMEM	0.31	0.62	0.80	1.37	2.25	3.10	3.84
AMEM	0.31	0.62	0.80	1.33	2.15	2.97	3.67
MS-MEM	0.30	0.61	0.80	1.35	2.23	3.08	3.83
MEM	0.31	0.61	0.79	1.33	2.14	2.98	3.67
HAR	0.30	0.60	0.77	1.27	2.10	2.96	3.63
HAR-norm	0.30	0.63	0.83	1.46	2.44	3.37	4.01

Models with a leverage effect have been used to fit the percentage log-returns on the S&P 500 and NASDAQ indices. HDM-lev denotes the HDM model with gamma innovations and leverage. The smallest values are bolded.