Supplementary Appendix to “A new approach to volatility modeling: the factorial hidden Markov volatility model”

Abstract

Section 1 provides a discussion of hierarchical and factorial hidden Markov models in the context of volatility modeling, with some economic interpretations. Section 2 contains the proofs of Theorem 1 and Propositions 1 and 2 of the paper. Section 3 discusses some computational aspects associated with the estimation of the FHMV model. Sections 4 and 5 describe, respectively, the competing return and realized variance models used in the empirical study (Section 4 of the paper).

JEL classification: C22, C51, C58

Keywords: volatility, Markov-switching, hierarchical hidden Markov model, persistence, leverage effect

1 Hierarchical and factorial hidden Markov models

This section provides a deeper discussion on hierarchical and factorial hidden Markov models (briefly mentioned in Remark 3 of the paper) and on some of their economic interpretations in the context of volatility modeling.

1.1 Definitions

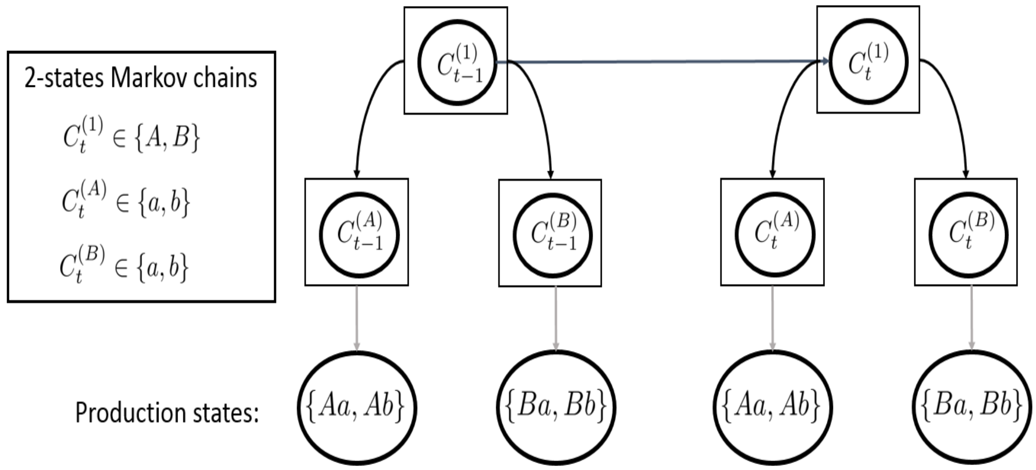
Hierarchical hidden Markov (HHM) models (see Fine et al., 1998) are structured in layers of hidden Markov chains. The state generated by the hidden Markov chain in the top layer can influence the dynamics of the hidden Markov chain in the second layer, and so on. The hidden Markov chain in the bottom layer serves as a "production state" because it generates the relevant state that will influence the observed process.

Factorial hidden Markov (FHM) models (see Ghahramani and Jordan, 1997) include multiple hidden Markov chains that evolve independently of each other and that are combined to produce the final state. The Markov-switching multifractal (MSM) model and the proposed model both fit into the FHM framework.

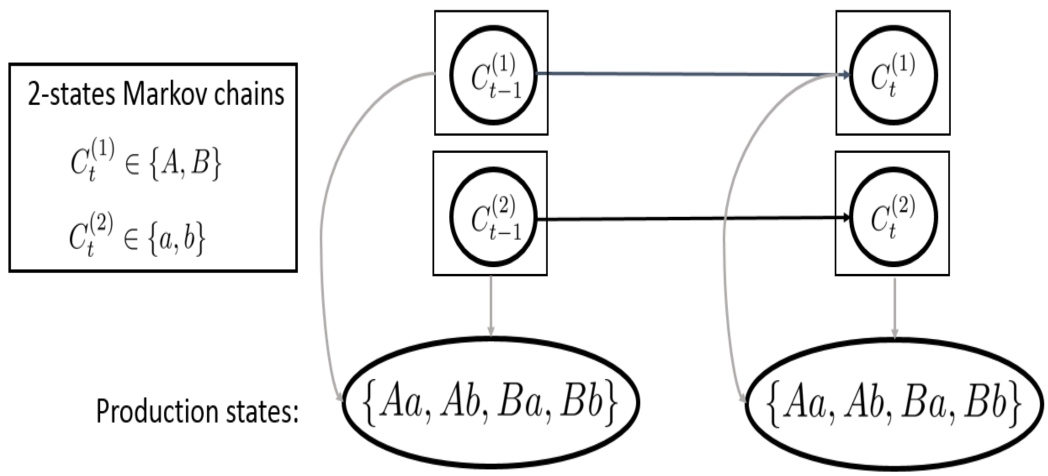
Figure 1 compares the structure of HHM and FHM models. Although the factorial structure can be seen as a particular case of the hierarchical structure (in which the different layers are independent of each other), both the HHM and FHM models can be formulated as a standard hidden Markov (HM) model. This follows from the fact that a combination of low-dimensional Markov chains can be reproduced by a single high-dimensional Markov chain. However, HHM and FHM models remain practical representations of a HM process because they allow us to consider a large number of states more parsimoniously.

1.2 Economic interpretations

Hamilton (1989) motivated his two regime Markov-switching (MS) model with a simple economic interpretation of the states. In the context of volatility modeling, one regime corresponds to a volatile market while the other mimics a tranquil period (i.e., bull versus bear periods). Nevertheless, empirical data typically invalidate this two-regime model as processes with a higher number



(a) Hierarchical hidden Markov framework



(b) Factorial hidden Markov framework

Figure 1: Comparison of hierarchical and factorial hidden Markov frameworks

of states are often preferred. Unfortunately, models with a high number of regimes are difficult to interpret and lack parsimony. HHM and FHM models offer a way to address these issues.

Let us reconsider the two-state MS model of Hamilton (1989) but let us introduce some breaks (or granularity) into the quiet and volatile periods. We are effectively building a HHM model, more specifically, the HHM model with K hidden layers can be expressed as follows:

$$r_t = \sigma(s_t^1, s_t^2, \ldots, s_t^K)\epsilon_t, \tag{1}$$

$$\sigma^2(s_t^1, s_t^2, \ldots, s_t^K) = \delta(s_t^1)' \begin{pmatrix} \sigma_L^2(s_t^2, \ldots, s_t^K) \\ \sigma_H^2(s_t^2, \ldots, s_t^K) \end{pmatrix}, \tag{2}$$

$$\sigma^2_L(s_t^2,\ldots,s_t^K) \;=\; \delta(s_t^2)' \begin{pmatrix} \sigma^2_{L,L}(s_t^3,\ldots,s_t^K) \\ \sigma^2_{L,H}(s_t^3,\ldots,s_t^K) \end{pmatrix}, \tag{3}$$

$$\sigma^2_H(s_t^2,\ldots,s_t^K) \;=\; \delta(s_t^2)' \begin{pmatrix} \sigma^2_{H,L}(s_t^3,\ldots,s_t^K) \\ \sigma^2_{H,H}(s_t^3,\ldots,s_t^K) \end{pmatrix}, \tag{4}$$

$$\ldots \;, \tag{5}$$

where r_t stands for the log-return at time t, $s_t^i \in \{1,2\}$ is the state variable in layer i ($i = 1,\ldots,K$), and $\delta(s_t^i)$ is a 2-dimensional vector with value equal to unity at entry s_t^i and zero at the other entry. The conditional variance $\sigma^2(s_t^1, s_t^2,\ldots,s_t^K)$ is therefore constructed hierarchically. The first layer determines if volatility is in a quiet or turbulent regime. The second layer breaks each one of these regimes in two to add granularity. For example, $\sigma^2_{H,L}$ means that the model lies in the lower volatility state of the most turbulent regime (i.e., $s_t^1 = 2$ and $s_t^2 = 1$). The interpretation of the other layers is analogous.

The dependence structure between the hidden Markov chains $\{s_t^1,\ldots,s_t^K\}$ determines if the process is a HHM model or a FHM model. The HHM model allows for a fairly general structure, whereas the FHM model imposes independence between Markov chains. In this case, the HHM model given in Equations (1)–(5) translates into a FHM model if we assume that each hidden layer is an independent two-state Markov chain with transition matrix given by

$$P^i \;=\; \begin{pmatrix} p_1^i & 1-p_1^i \\ 1-p_2^i & p_2^i \end{pmatrix}. \tag{6}$$

If we specify the conditional variance $\sigma^2(s_t^1, s_t^2,\ldots,s_t^K)$ as the product of the output of each hidden Markov chain, the FHM model is equivalent to a HM model with 2^K regimes and transition matrix $P = P^1 \otimes P^2 \otimes \ldots \otimes P^K$ (where $\otimes$ denotes the Kronecker product). In particular, the model has two equivalent representations given by

$$r_t \;=\; \sigma(s_t^1, s_t^2,\ldots,s_t^K)\epsilon_t, \tag{7}$$

$$\text{1st representation:}\quad \sigma^2(s_t^1, s_t^2,\ldots,s_t^K) \;=\; \prod_{i=1}^{K} \sigma^2_{s_t^i} \tag{8}$$

$$\text{2nd representation:}\quad \sigma^2(s_t^1, s_t^2,\ldots,s_t^K) \;=\; [\delta(s_t^1)\otimes\delta(s_t^2)\ldots\delta(s_t^k)]' \begin{pmatrix} \sigma_1^2 & \sigma_2^2 & \ldots & \sigma_{2^K}^2 \end{pmatrix}'. \tag{9}$$

Model (7) addresses two drawbacks of the standard MS model with a large number of regimes. On one hand, the number of parameters in the transition matrix increases linearly with K instead of quadratically. On the other hand, by adding constraints on the output of each Markov chain, it is possible to generate a bull and bear interpretation that would share some similarity with the general HHM model presented in Equation (5). The FHM model exposed in Equations (7)-(8) can in fact allow for different economic interpretations, as exemplified by the following three cases:

(i) If we assume that for each Markov chain i ($i = 1, \ldots, K$) the output of the first state is smaller than the output of the second state (i.e., $\sigma^2_{s^i_t=1} < \sigma^2_{s^i_t=2}$) and impose the constraint $\sigma_{s^1_t=1} \prod_{i=1}^K \sigma^2_{s^i_t=2} < \sigma_{s^1_t=2} \prod_{i=1}^K \sigma^2_{s^i_t=1}$, then we can interpret the first layer as a bull and bear regime. This is simply because under these conditions, the first layer has the strongest impact on volatility in the sense that the final volatility state generated when $s^1_t = 1$ is always smaller than when $s^1_t = 2$.

(ii) If we assume that for each Markov chain i ($i = 1, \ldots, K$) the transition probabilities $p^i_1 = p^i_2 = p^i$ are equal, but over different matrices these probabilities grow toward one (i.e., $p^1 < p^2 < \ldots < p^K \leq 1$), the FHM model becomes a generalization of the MSM process of Calvet and Fisher (2004). In such a framework, the Markov chains are interpreted as the impact on the financial market of agents selling and buying assets at different frequencies.

(iii) If we assume that all of the transition probabilities are equal (i.e., $p^i_1 = p^i_2 = p, \forall i = 1, \ldots, K$) and that the two possible states of each Markov chain i are $\{1, \sigma^2_i\}$ with $\sigma^2_{i+1} \leq \sigma^2_i$, the FHM model is then composed of the $\{C^{(i)}_t\}$ components of our FHMV model. In this case, the Markov chains can be interpreted as heterogeneous shocks that randomly hit the financial market.

2 Proofs of Theorem 1 and Propositions 1 and 2

Before proving Theorem 1 and Propositions 1 and 2, we review some properties of the Kronecker product and study the eigenvalues and eigenvectors of the transition probability matrix in the FHMV model.

2.1 Properties of the Kronecker product

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

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

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

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

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

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

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

2.2 Eigenvalues and eigenvectors

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

(i) The eigenvalues of $\boldsymbol{P}$ *are* 1 *and* $\gamma = 2p - 1$, *where* $p \in (0, 1)$.

(ii) The corresponding right eigenvectors are $\boldsymbol{v} = (1, 1)'$ *and* $\boldsymbol{v}_\gamma = (-1, 1)'$.

(iii) The corresponding left eigenvectors are $\boldsymbol{w} = (1/2, 1/2)' = 2^{-1}\boldsymbol{v}$ *and* $\boldsymbol{w}_\gamma = (-1/2, 1/2)' = 2^{-1}\boldsymbol{v}_\gamma$.

(iv) The right and left eigenvectors defined in items (ii) and (iii) satisfy $\boldsymbol{w}'\boldsymbol{v} = \boldsymbol{w}_\gamma'\boldsymbol{v}_\gamma = 1$ *and* $\boldsymbol{w}'\boldsymbol{v}_\gamma = \boldsymbol{w}_\gamma'\boldsymbol{v} = 0$ *(i.e., both sets of eigenvectors are orthogonal).*

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

Proof. Items *(i)–(v)* follow directly from basic linear algebra theory. □

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

<u>*Case* $p \neq 1/2$ *(and therefore* $\gamma \neq 0$*)*</u>.

(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):* $\lambda_i = \gamma$, $i = 2, \ldots, N+1$.

(ii) *The right eigenvectors,* $\{\boldsymbol{v}_i\}_{i=1}^{2^N}$, *can be obtained by way of a Kronecker product of* N *vectors among* $\boldsymbol{v} = (1,1)'$ *and* $\boldsymbol{v}_\gamma = (-1,1)'$. *Therefore, these vectors can only contain values in the set* $\{-1,1\}$. *In particular, we have* $\boldsymbol{v}_1 = \boldsymbol{v}^{\otimes N} = \mathbf{1}_{2^N}$.

(iii) *The left eigenvectors,* $\{\boldsymbol{w}_i\}_{i=1}^{2^N}$, *satisfy the relationship,*

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

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

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

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

These two sets of eigenvectors are therefore orthogonal.

(v) $\boldsymbol{P}_C$ *is diagonalizable and we may write* $\boldsymbol{P}_C = \boldsymbol{V}_C\boldsymbol{\Lambda}\boldsymbol{W}_C'$, *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* $\gamma = 0$*).*

(vi) $\lambda_1 = 1$ *and* $\lambda_i = 0$, $i = 2, \ldots, 2^N$, *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) $\boldsymbol{P}_C = \mathbf{1}_{2^N}\boldsymbol{\pi}_C'$, *where* $\boldsymbol{\pi}_C = 2^{-N}\mathbf{1}_{2^N}$.

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

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

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

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

When $i \neq j$, the term $\boldsymbol{w}'\boldsymbol{v}_\gamma = 0$ or $\boldsymbol{w}_\gamma'\boldsymbol{v} = 0$ appears at least once when using the above property of the Kronecker product on $\boldsymbol{w}_i'\boldsymbol{v}_j$, which explains why $\boldsymbol{w}_i'\boldsymbol{v}_j = 0$, 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 $\boldsymbol{P}_C = \boldsymbol{P}^{\otimes N} = (2^{-1}\mathbf{1}_2\mathbf{1}_2')^{\otimes N} = 2^{-N}\mathbf{1}_{2^N}\mathbf{1}_{2^N}'$. □

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

(i) *1 is an eigenvalue of $\boldsymbol{P}_M$ 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 $\boldsymbol{P}_M$ with an algebraic multiplicity of $N-1$.*

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

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

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

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

Since $\boldsymbol{P}_M$ is an $N \times N$ matrix with N repeated rows, it follows that $\boldsymbol{P}_M$ has rank 1 and nullity of $N-1$. The singularity of $\boldsymbol{P}_M$ implies that this matrix has an eigenvalue of 0. The geometric multiplicity of this eigenvalue corresponds to the nullity of $\boldsymbol{P}_M$, 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 $\boldsymbol{P}_M$ is diagonalizable. □

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

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

(i) The set of non-zero eigenvalues of $\boldsymbol{P}_V$ corresponds to the set of all eigenvalues of $\boldsymbol{P}_C$, denoted by $\{\lambda_i\}_{i=1}^{2^N}$ (see item(i) of Lemma 5).

(ii) The sets of right and left eigenvectors of $\boldsymbol{P}_V$ associated with these eigenvalues correspond to $\{\boldsymbol{v}_i \otimes \boldsymbol{1}_N\}_{i=1}^{2^N}$, and $\{\boldsymbol{w}_i \otimes \boldsymbol{\pi}_M\}_{i=1}^{2^N}$, respectively, where $\{\boldsymbol{v}_i\}_{i=1}^{2^N}$, and $\{\boldsymbol{w}_i\}_{i=1}^{2^N}$ denote the right and left eigenvectors of $\boldsymbol{P}_C$, respectively.

(iii) Let the matrices $\boldsymbol{V}_V$ and $\boldsymbol{W}_V$ contain, respectively, the right and left eigenvectors of $\boldsymbol{P}_V$ associated with non-zero eigenvalues (as defined in item (ii)) along their columns. Accordingly, $\boldsymbol{V}_V$ and $\boldsymbol{W}_V$ are matrices of dimension $(N \cdot 2^N) \times 2^N$. Then, $\boldsymbol{W}_V'\boldsymbol{V}_V = \boldsymbol{I}_{2^N}$. Specifically, the right and left eigenvectors in item (ii) satisfy:

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

These two sets of eigenvectors are therefore orthogonal.

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

(vi) $\boldsymbol{P}_V = \boldsymbol{1}_{N \cdot 2^N}\boldsymbol{\pi}_{\boldsymbol{V}}'$, 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 $\boldsymbol{P}_C$, $\{\lambda_i\}_{i=1}^{2^N}$, are non-zero (see item *(i)* of Lemma 5), and observe that $\boldsymbol{P}_M$ 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 $\boldsymbol{P}_V = \boldsymbol{P}_C \otimes \boldsymbol{P}_M$ must correspond to $\{\lambda_i\}_{i=1}^{2^N}$. This shows item *(i)*.

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

Using Lemma 1, we can write,

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

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

Item *(iv)* follows from Lemma 3, because $\boldsymbol{P}_C$ and $\boldsymbol{P}_M$ 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 $\boldsymbol{P}_V = \boldsymbol{P}_C \otimes \boldsymbol{P}_M = (\mathbf{1}_{2^N}\boldsymbol{\pi}_C') \otimes (\mathbf{1}_N\boldsymbol{\pi}_M') = \mathbf{1}_{N\cdot 2^N}(\boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M)'$. □

2.3 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* $\boldsymbol{Q} = \mathbf{1}_n\boldsymbol{\pi}'$. *Then,* $\boldsymbol{Q}^k = \boldsymbol{Q}$, *for* $k = 1, 2, \ldots$.

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

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

(i) We have,

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

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

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

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

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

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

We now prove Theorem 1.

Proof. First, consider the case $p = 1/2$ (and therefore $\gamma = 0$). In this specific case we have $\boldsymbol{P}_V = \boldsymbol{\Pi}_V$ (see item *(vi)* of Lemma 7), and therefore, by Lemma 8, $\boldsymbol{P}_V^k = \boldsymbol{\Pi}_V$, $k = 1, 2, \ldots$. 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 $\boldsymbol{P}_V$ is a positive matrix. The Perron-Frobenius theorem (Seneta, 2006) then implies that

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

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

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

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

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

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

$$
\begin{aligned}
&= \left(2^{-N}\sum_{i=2}^{2^N}|\lambda_i^k|\right)\cdot(\mathbf{1}_{2^N}\mathbf{1}_{2^N}')\otimes(\mathbf{1}_N\boldsymbol{\pi}_M') \\
&= 2^{-N}\left(\sum_{j=1}^{N}\binom{N}{j}(|\gamma|^k)^j\right)\cdot(\mathbf{1}_{2^N}\mathbf{1}_{2^N}')\otimes(\mathbf{1}_N\boldsymbol{\pi}_M') && \text{[Lemma 5]} \\
&= 2^{-N}\left((1+|\gamma|^k)^N-1\right)\cdot(\mathbf{1}_{2^N}\mathbf{1}_{2^N}')\otimes(\mathbf{1}_N\boldsymbol{\pi}_M') && \text{[Binomial theorem]} \\
&= 2^{-N}\left((1+|\gamma|^k)^N-1\right)\cdot(\mathbf{1}_{2^N}\otimes\mathbf{1}_N)(\mathbf{1}_{2^N}'\otimes\boldsymbol{\pi}_M') && \text{[Lemma 1]} \\
&= 2^{-N}\left((1+|\gamma|^k)^N-1\right)\cdot\mathbf{1}_{N\cdot 2^N}(\mathbf{1}_{2^N}'\otimes\boldsymbol{\pi}_M') \\
&= \left((1+|\gamma|^k)^N-1\right)\cdot\mathbf{1}_{N\cdot 2^N}\left(2^{-N}\mathbf{1}_{2^N}'\otimes\boldsymbol{\pi}_M'\right) \\
&= \left((1+|\gamma|^k)^N-1\right)\cdot\mathbf{1}_{N\cdot 2^N}(\boldsymbol{\pi}_C\otimes\boldsymbol{\pi}_M)' \\
&= \left((1+|\gamma|^k)^N-1\right)\cdot\mathbf{\Pi}_V. && (10)
\end{aligned}
$$

Equation (8) of the paper then follows directly by noting that all row sums of $\mathbf{\Pi}_V$ are equal to 1, whereas Equation (9) of the paper is due to the fact that the upper bound in Equation (10) above is actually attained on the diagonal of $(\boldsymbol{P}_V^k - \mathbf{\Pi}_V)$ when $\gamma > 0$. To see why the latter is true, consider

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

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

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

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 (6) in the paper), 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 $\boldsymbol{P}_V$ (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). $\square$

2.4 Proof of Proposition 1

We prove Proposition 1.

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

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

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

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

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

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

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

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

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. □

2.5 Proof of Proposition 2

We prove Proposition 2.

Proof. First, let $\boldsymbol{P}_i$ denote the t.p.m. of the Markov chain $\{\widetilde{C}_t^{(i)}\}$, for $i = 1, \ldots, N$. We have $\boldsymbol{P}_i = \begin{pmatrix} \frac{1}{2} + \frac{\tilde{p}_i}{2} & \frac{1}{2} - \frac{\tilde{p}_i}{2} \\ \frac{1}{2} - \frac{\tilde{p}_i}{2} & \frac{1}{2} + \frac{\tilde{p}_i}{2} \end{pmatrix}$, where $\boldsymbol{P}_i$ exhibits the same structure as $\boldsymbol{P}$ with $p = \frac{1}{2} + \frac{\tilde{p}_i}{2}$ (as defined in Equation (5) of our paper). Consequently, from Lemma 4, $\boldsymbol{P}_i = \boldsymbol{v}\boldsymbol{w}' + \tilde{p}_i \boldsymbol{v}_\gamma \boldsymbol{w}_\gamma'$ where $\tilde{p}_i$ stands for the second largest eigenvalue of $\boldsymbol{P}_i$. Additionally, $\boldsymbol{P}_i = \boldsymbol{V}\boldsymbol{\Lambda}_i\boldsymbol{W}$, where $\boldsymbol{\Lambda}_i = \mathrm{diag}(1, \tilde{p}_i)$, $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v} & \boldsymbol{v}_\gamma \end{pmatrix}$ and $\boldsymbol{W} = \boldsymbol{V}^{-1} = \begin{pmatrix} \boldsymbol{w} & \boldsymbol{w}_\gamma \end{pmatrix}'$.
By a property of the Kronecker product,

$$\boldsymbol{P}_{\mathrm{MSM}} = \boldsymbol{P}_1 \otimes \boldsymbol{P}_2 \otimes \ldots \otimes \boldsymbol{P}_N, \tag{11}$$

$$= \quad \boldsymbol{V}^{\otimes N} \underbrace{(\boldsymbol{\Lambda}_1 \otimes \boldsymbol{\Lambda}_2 \otimes \ldots \otimes \boldsymbol{\Lambda}_N)}_{\boldsymbol{\Lambda}_{\text{MSM}}} \boldsymbol{W}^{\otimes N}. \tag{12}$$

Regarding items (i) and (ii), from Lemma 5, the right eigenvector associated with the eigenvalue equal to one is given by $\boldsymbol{v}^{\otimes N} = \mathbf{1}_{2^N}$ while the left eigenvector is given by $\boldsymbol{w}^{\otimes N} = 2^{-N}\mathbf{1}_{2^N} = 2^{-N}\boldsymbol{v}^{\otimes N}$. Consequently, the asymptotic limit of $\boldsymbol{P}^k_{\text{MSM}}$ is

$$\lim_{k\to\infty} \boldsymbol{P}^k_{\text{MSM}} \quad = \quad \mathbf{1}_{2^N} \underbrace{2^{-N}\mathbf{1}'_{2^N}}_{\boldsymbol{\pi}'_{\text{MSM}}}.$$

To prove item (iii), Equation (12) implies that the second largest eigenvalue of $\boldsymbol{P}_{\text{MSM}}$ is given by the largest $\tilde{p}_i$ for $i = 1, \ldots, N$. Since $\tilde{p}_i = \tilde{a}^{\tilde{b}^{i-1}}$, where $\tilde{a} \in (0, 1)$ and $\tilde{b} \in (1, \infty)$, the largest value is obtained when $i = 1$, that is $\tilde{p}_1 = \tilde{a}$. The multiplicity of this eigenvalue is one since it appears only once in $\boldsymbol{\Lambda}_{\text{MSM}}$. □

3 Additional details on model estimation

This section discusses some computational aspects associated with the estimation of the FHMV model. Table 1 shows computing times required to evaluate the likelihood function when the number of components N increases (based on a Hamilton filter coded in MATLAB and running on a Intel(R) Core i7-4790 of 3.60GHz with 16 Gb of RAM). Note that since the jump component can be integrated out, the Hamilton filter only needs to iterate over at most 1,024 states (for $N = 10$). The computational burden is therefore similar to that of the MSM process (for a MSM model with N components, a Hamilton filter must be run on 2^N states). In general, the computing time scales exponentially with N, a computational curse that is well known in the factorial hidden Markov model literature. Elapsed times for finding maximum likelihood estimates depend on the optimization method, on starting values, and is proportional to the number of observations. To provide an example, if 1,000 likelihood evaluations were required to find the optimum (this number of evaluations is generally sufficient for a Newton-Raphson algorithm), maximum likelihood estimation would always be completed in less than 30 minutes based on the values given in Table 1.

To estimate the model with a number of components larger than $N = 10$, a Markov chain

Table 1: Computing times in seconds required to evaluate the likelihood function.

N	2	3	4	5	6	7	8	9	10
Percentage log-returns									
S&P 500 ($T = 4150$)									
FHMV	0.00	0.00	0.01	0.03	0.05	0.09	0.17	0.41	1.46
FHMV-lev	0.03	0.03	0.04	0.05	0.08	0.12	0.19	0.43	1.48
NASDAQ ($T = 4149$)									
FHMV	0.00	0.00	0.01	0.03	0.05	0.09	0.17	0.40	1.43
FHMV-lev	0.03	0.03	0.04	0.05	0.08	0.13	0.21	0.43	1.52
USD/EUR ($T = 4147$)									
FHMV	0.00	0.01	0.01	0.03	0.07	0.14	0.24	0.60	1.35
FHMV-lev	0.04	0.05	0.05	0.07	0.11	0.17	0.33	0.65	1.49
Realized variances									
S&P 500 ($T = 4120$)									
FHMV	0.00	0.01	0.02	0.05	0.10	0.17	0.31	0.58	1.64
FHMV-lev	0.05	0.05	0.06	0.10	0.14	0.21	0.36	0.63	1.60
NASDAQ ($T = 4124$)									
FHMV	0.00	0.01	0.02	0.06	0.09	0.17	0.21	0.49	1.60
FHMV-lev	0.04	0.06	0.07	0.10	0.13	0.20	0.24	0.52	1.68
USD/EUR ($T = 2328$)									
FHMV	0.00	0.00	0.01	0.02	0.04	0.07	0.13	0.29	0.93
FHMV-lev	0.02	0.02	0.02	0.04	0.05	0.09	0.15	0.31	0.92

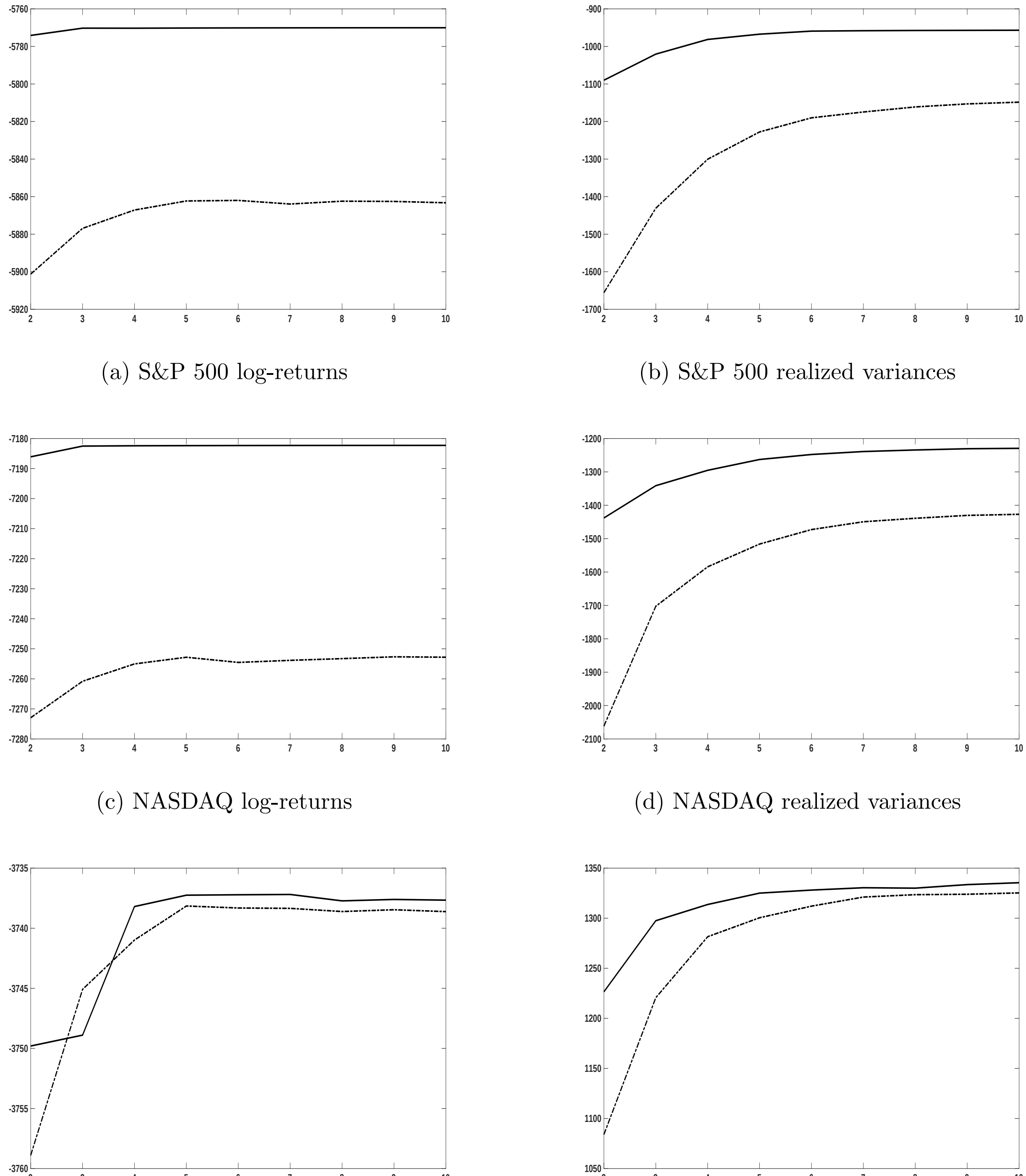
Monte Carlo algorithm, such as a simple Gibbs algorithm, could be considered. However, as empirically illustrated in Figure 2, $N = 10$ was enough in our applications. This figure shows the maximal value of the log-likelihood function that can be attained for different choices of N for the six data sets considered in the paper. We observe that additional components never significantly deteriorate the log-likelihood and that the choice $N = 10$ either yields the highest log-likelihood or is very close to it.

4 Alternative models for log-returns

We briefly review the specifications of the processes used in the paper for modeling returns. The time series $\{r_t\}$ refers to demeaned daily percentage log-returns. Whenever necessary, model parameters are constrained to ensure positivity of the variance.

4.1 MSM

The MSM model is fully specified in Section 2.4 of the paper.



(a) S&P 500 log-returns

(b) S&P 500 realized variances

(c) NASDAQ log-returns

(d) NASDAQ realized variances

(e) USD/EUR log-returns

(f) USD/EUR realized variances

Figure 2: Log-likelihood function evaluated at the MLE as a function of N. Results for the FHMV model with(out) leverage are displayed by a solid (dashed) line.

4.2 GARCH-t and GJR-t

The GJR-t model is defined as

$$r_t = \sigma_t \epsilon_t,$$

$$\sigma_t^2 = \omega + (\alpha + \gamma \mathbb{1}_{\{r_{t-1}<0\}}) r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where ϵ_t is driven by a standardized Student-t distribution with ν degrees of freedom. The GARCH-t model is a particular case of the GJR-t process with $\gamma = 0$.

4.3 MS-GARCH-t and MS-GJR-t

The MS-GJR-t 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 $\{1, \ldots, M\}$ and transition matrix $(p_{ij})_{i,j=1}^M$, where $p_{ij} = \Pr(X_t = j \mid X_{t-1} = i)$, for $i, j = 1, \ldots, M$. The MS-GJR-t process is given by

$$\begin{aligned} r_t &= \sigma_{t,X_t} \epsilon_t, \\ \sigma_{t,j}^2 &= \omega_j + (\alpha_j + \gamma_j \mathbb{1}_{\{r_{t-1}<0\}}) r_{t-1}^2 + \beta_j \sigma_{t-1,j}^2, \quad j = 1, \ldots, M, \end{aligned}$$

where ϵ_t is driven by a standardized Student-t distribution with ν degrees of freedom and $\sigma_{t,j}^2$ represents the conditional variance process in regime j ($j = 1, \ldots, M$). Our estimated MS-GJR-t model includes $M = 2$ regimes. The MS-GARCH-t model is a particular case of the MS-GJR-t process with $\gamma_j = 0$ for $j = 1, \ldots, M$.

5 Alternative models for realized variances

We briefly review the specifications of the processes used in the paper for modeling realized variances. The time series $\{\mathrm{RV}_t\}$ refers to daily realized kernel variances (scaled by a factor of 100^2). Whenever necessary, model parameters are constrained to ensure positivity of the variance.

5.1 log-HAR and log-HAR-lev

The log-HAR-lev process, developed by Corsi and Renò (2012) and Corsi et al. (2012), is specified as

$$\log \mathrm{RV}_t = \mu_t + \epsilon_t, \tag{13}$$

$$\mu_t = \beta_0 + \beta_1 \log \mathrm{RV}_t^{(1)} + \beta_2 \log \mathrm{RV}_t^{(5)} + \beta_3 \log \mathrm{RV}_t^{(22)} + \mathrm{lev}_t, \quad (14)$$

$$\mathrm{lev}_t = \beta_4 r_t^{(1)-} + \beta_5 r_t^{(5)-} + \beta_6 r_t^{(22)-}, \quad (15)$$

where ϵ_t is driven by a normal distribution with mean equal to one and variance equal to a, and the explanatory variables are defined as $\log \mathrm{RV}_t^{(h)} = \frac{1}{h}\sum_{i=1}^{h} \log \mathrm{RV}_{t-i}$ and $r_t^{(h)-} = \min(\frac{1}{h}\sum_{i=1}^{h} r_{t-i}, 0)$ for $h = 1, 5$ and 22, where r_t corresponds to the percentage log-return at time t. The model without a leverage effect (log-HAR) is obtained by setting the component lev_t to zero.

5.2 MEM and MEM-lev

The specification of the MEM-lev is given by

$$\begin{aligned} \mathrm{RV}_t &= \mu_t \eta_t, \\ \mu_t &= \omega + (\alpha + \gamma \mathbb{1}_{\{r_{t-1}<0\}})\mathrm{RV}_{t-1} + \beta \mu_{t-1}, \end{aligned}$$

where η_t is driven by a gamma distribution with mean equal to one and variance equal to a. The MEM is a particular case of the MEM-lev with $\gamma = 0$.

5.3 MS-MEM and MS-MEM-lev

The MS-MEM-lev model that we consider corresponds to a MEM version of the MS-GJR-t model introduced in Section 4.3. 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-MEM-lev is specified as

$$\begin{aligned} \mathrm{RV}_t &= \mu_{t,X_t} \eta_t, \\ \mu_{t,j} &= \omega_j + (\alpha_j + \gamma_j \mathbb{1}_{\{r_{t-1}<0\}})\mathrm{RV}_{t-1} + \beta_j \mu_{t-1,j}, \quad j = 1, \dots, M, \end{aligned}$$

where η_t is driven by a gamma distribution with mean equal to one and variance equal to a. Our estimated MS-MEM-lev model includes $M = 2$ regimes. The MS-MEM model is a particular case of the MS-MEM-lev with $\gamma_j = 0$ for $j = 1, \dots, M$.

References

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. 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.

Corsi, F., Audrino, F., and Renò, R. (2012). HAR modeling for realized volatility forecasting. In *Handbook of volatility models and their applications*, Wiley Handb. Finance Eng. Econom., pages 363–382. Wiley, Hoboken, NJ.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380.

Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. *Machine learning*, 32(1):41–62.

Ghahramani, Z. and Jordan, M. I. (1997). Factorial hidden Markov models. *Machine Learning*, 29(2):245–273.

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 Paolella, 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.

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].