

BAYESIAN NONPARAMETRIC PANEL MARKOV-SWITCHING GARCH MODELS*

BY ROBERTO CASARIN[†], MAURO COSTANTINI[‡] AND ANTHONY
OSUNTUYI[†]

University Ca' Foscari of Venice[†] and University of L'Aquila[‡]

Abstract This paper introduces a new model for panel data with Markov-switching GARCH effects. The model incorporates a series-specific hidden Markov chain process that drives the GARCH parameters. To cope with the high-dimensionality of the parameter space, the paper exploits the cross-sectional clustering of the series by first assuming a soft parameter pooling through a hierarchical prior distribution with two-step procedure, and then introducing clustering effects in the parameter space through a nonparametric prior distribution. The model and the proposed inference are evaluated through a simulation experiment. The results suggest that the inference is able to recover the true value of the parameters and the number of groups in each regime. An empirical application to 78 assets of the SP&100 index from 6th January 2000 to 3rd October 2020 is also carried out by using a two-regime Markov switching GARCH model. The findings shows the presence of 2 and 3 clusters among the constituents in the first and second regime, respectively.

1. Introduction. Over the last ten years, there has been an increasing interest in the study of volatility of large panels of asset returns, with a special focus on dynamic dependence and heterogeneity across assets (Pakel, Shephard and Sheppard, 2011; Barigozzi, Brownlees and Veredas, 2014; Ardia et al., 2018; Bollerslev, Patton and Quaedly, 2020). The empirical evidence has also shown the presence of regimes in the volatility of financial returns (see Ardia, 2008; Ang and Timmermann, 2012; Bauwens and Otranto, 2016; Haas and Liu, 2018, among others) and Markov switching (MS) GARCH models have been used to cope with regime changes and temporal clustering of the conditional volatility.

Several GARCH models have been proposed to account for dependence (for a review, see Virbickaite, Ausín and Galeano, 2015; Bauwens and Otranto, 2016, 2020), but the estimation of a large number of

*Corresponding author: Roberto Casarin E-mail: r.casarin@unive.it. The code is available upon request.

MSC 2010 subject classifications: Primary 62F15, 60K35; secondary 62P20, 91B64

Keywords and phrases: Bayesian nonparametrics, GARCH models, Gibbs sampling, Markov-switching, Time series

parameters with the available data dimension remains an open issue. In this respect, evidence of cluster-wise dependence in the distribution of financial asset returns (see Bauwens and Rombouts, 2007) has prompted researcher to adopt cross-sectional clustering of the time series as a building block for a dimensionality reduction step in large dimensional problems of the parameter space (see, for example, Hirano, 2002; Billio, Casarin and Rossini, 2019).

In this paper, we propose to model the cross-sectional clustering effects with a Bayesian nonparametric technique (Ferguson, 1973; Lo, 1984) where a hierarchical Pitman-Yor process prior (Pitman and Yor, 1997) for the MS-GARCH parameters is considered. Non-parametric Bayesian techniques have been largely and successfully used in different fields such as biostatistics (Do, Muller and Tang, 2005), biology (Arbel, Mengersen and Rousseau, 2016), medicine (Xu et al., 2016), and neuroimaging (Zhang et al., 2016). For an introduction to Bayesian non-parametrics see Hjort et al. (2010) and for a review of models and applications in different fields see Müller and Mitra (2013).

In our panel model, the first stage of the hierarchical prior allows for cross-unit heterogeneity, while shrinking all unit-specific parameters towards a common mean. The second stage of the hierarchy allows for mixed effects in the common mean. There are many advantages in using this hierarchical nonparametric prior. First, our approach allows for making inference on the number of mixture components in the cross-sectional clustering. Second, it adds flexibility to the model allowing for different shapes of the prior and posterior predictive distributions. Third, the predictive distribution incorporates uncertainty in the parameters and in the number of mixture components. Lastly, the Bayesian nonparametric combined with a data-augmentation strategy makes the inference more tractable for our high dimensional model.

The model and inference proposed in this paper are novel in some respects. As such, the paper contributes to the literature on Bayesian semiparametrics and nonparametrics for time series analysis (e.g., see Taddy and Kottas, 2009; Jensen and Maheu, 2010; Griffin and Steel, 2011; Di Lucca et al., 2013; Bassetti, Casarin and Leisen, 2014; Casarin, Molina and ter Horst, 2019; Billio, Casarin and Rossini, 2019; Nieto-Barajas and Quintana, 2016; Griffin and Kalli, 2018). The paper also innovates the Bayesian nonparametric dynamic panel model in Hirano (2002) by introducing Markov-switching and GARCH dynamics.

The paper also extends the nonparametric switching regression in Taddy and Kottas (2009) to a panel model with GARCH dynamics.

Our approach differs from those in Hirano (2002) and Taddy and Kottas (2009), and is in line with the strategies for large dimensional and over-parametrized models (e.g., see MacLehose and Dunson, 2010; Wang, 2010; Billio, Casarin and Rossini, 2019), where a multiple-stage hierarchical prior is used to combine partial pooling and clustering effects in the parameter space. Further, differently from Hirano (2002) and Taddy and Kottas (2009), the paper uses a MCMC algorithm for posterior approximation that relies on the efficient sampling method developed in Walker (2007); Kalli, Griffin and Walker (2011); Hatjispyros, Nicolieris and Walker (2011). Lastly, the paper makes a contribution to the literature on Bayesian Markov-switching panel models (e.g., see Kaufmann, 2010, 2015; Billio et al., 2016; Casarin et al., 2019) by introducing GARCH effects and allowing for a flexible nonparametric specification.

The estimation of MS-GARCH models is also a difficult task given the path dependence problem (Gray, 1996) and approximation methods have been considered (e.g., see Bauwens, Preminger and Rombouts, 2010; Henneke et al., 2011; Ardia, 2008; Haas, Mittnik and Paoletta, 2004; He and Maheu, 2010; Bauwens, Dufays and Rombouts, 2014; Elliott et al., 2012; Dufays, 2015; Wee, Chen and Dunsmuir, 2020). In this paper, we extend the univariate Gibbs sampler by Billio, Casarin and Osuntuyi (2016) to a multiple time series set-up and provide an efficient MCMC procedure for the hidden states of a panel MS-GARCH model. The proposed method relies on a combination of Gibbs and Metropolis samplers. The model and the proposed inference are evaluated through simulation experiments. The results show that the inference is able to recover the true value of the parameters and the number of groups in each regime.

Our model is applied to 78 assets of the SP&100 index from 6th January 2000 to 3rd October 2020. The analysis can be useful for portfolio making and style investing decisions. In particular, the analysis aims to identify the under- and over-performance regimes in expected returns. Then, clusters of assets within each regime are identified. Lastly, we use the sector classification and some fundamental financial ratios to study the composition of the clusters. The main empirical results are as follows. We find evidence of different clustering structure across regimes. Regime 1 (over-performance phase) and regime 2 (under-performance phase) comprise 2 and 3 clusters, respectively. While the composition of clusters varies across regimes, some common features are observed. In both regimes, medium size companies represent the largest majority and assets in cluster 2 seem to be overvalued by their Price-to-Earning ratio.

The paper is organized as follows. Section 2 introduces our MS-GARCH

panel model and the Bayesian nonparametric prior distribution. Section 3 presents the data augmentation strategy and the posterior approximation method. In Section 4, we present a simulation study and an empirical application to the financial returns data. Section 5 concludes.

2. A Bayesian nonparametric MS-GARCH model. We assume that the observable variable y_{it} for the i -th unit of the panel at time t satisfies

$$(1) \quad y_{it} = \mu_i(s_{it}) + \sigma_{it}\varepsilon_{it}, \quad \varepsilon_{it} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$$

for $t = 1, \dots, T$ and $i = 1, \dots, N$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with location μ and scale σ . The conditional variance is as follows:

$$(2) \quad \sigma_{it}^2 = \gamma_i(s_{it}) + \alpha_i(s_{it})\varepsilon_{it-1}^2 + \beta_i(s_{it})\sigma_{it-1}^2$$

which is the MS-GARCH model, and s_{it} , $t = 1 \dots, T$ is a hidden Markov chain process with transition probability

$$(3) \quad P(s_{it} = k | s_{it-1} = l) = p_{i,kl}$$

where $k, l = 1, \dots, K$ with K the number of states. The following functional form for the switching parameters is specified as follows:

$$(4) \quad \mu_i(s_{it}) = \sum_{k=1}^K \mu_{ik} \mathbb{I}(s_{it} = k), \quad \alpha_i(s_{it}) = \sum_{k=1}^K \alpha_{ik} \mathbb{I}(s_{it} = k)$$

$$(5) \quad \beta_i(s_{it}) = \sum_{k=1}^K \beta_{ik} \mathbb{I}(s_{it} = k), \quad \gamma_i(s_{it}) = \sum_{k=1}^K \gamma_{ik} \mathbb{I}(s_{it} = k)$$

We cope with the high-dimensionality of the parameter space due to the large cross-section dimension N , and related overfitting issues of the model, by exploiting cross-sectional clustering of the series. More specifically we propose to combine two modelling strategies. First, we assume soft parameter pooling through a hierarchical prior distribution with two stages, and second we introduce clustering effects in the parameter space through a nonparametric prior. The resulting joint prior distribution for the MS-GARCH parameters is given by the following.

In the first stage, the rows of the transition matrix are assumed to follow a Dirichlet distribution:

$$(6) \quad (p_{i,k1}, \dots, p_{i,kK}) \stackrel{iid}{\sim} \mathcal{D}(\phi r_{k1}, \dots, \phi r_{kK})$$

for all units $i = 1, \dots, N$ and regimes $k = 1, \dots, K$, where the precision parameter ϕ shrinks the unit-specific probabilities toward a common value (r_{k1}, \dots, r_{kK}) . For the second stage we assume

$$(7) \quad (r_{k1}, \dots, r_{kK}) \stackrel{iid}{\sim} \mathcal{D}(d, \dots, d)$$

with $d = 1/K$.

To cope with the high dimension of the parameter space, in the first stage of the hierarchical prior, we shrink the switching parameters toward some common values, and in the second stage we introduce a regime-specific process which is clustering the units in M_k groups $C_{1,k}, \dots, C_{M_k,k}$ such that $C_{h,k} \cap C_{l,k} = \emptyset$ for $h \neq l$ and $\cup_{h=1}^{M_k} C_{h,k} = \{1, \dots, N\}$. In the first stage, we assume the following

$$(8) \quad \mu_{ik} \sim \mathcal{N}(\tilde{\mu}_{ik}^*, s), \quad \gamma_{ik}/a \sim \mathcal{B}e(r\tilde{\gamma}_{ik}^*/a, r(1 - \tilde{\gamma}_{ik}^*/a)),$$

$$(9) \quad \alpha_{ik} \sim \mathcal{B}e(r\tilde{\alpha}_{ik}^*, r(1 - \tilde{\alpha}_{ik}^*)), \quad \beta_{ik} \sim \mathcal{B}e(r\tilde{\beta}_{ik}^*, r(1 - \tilde{\beta}_{ik}^*))$$

for $k = 1, \dots, K$ where $\mathcal{B}e(\alpha, \beta)$ denotes the beta distribution with mean $\alpha/(\alpha + \beta)$ and a is a real positive constant. The scale hyper-parameters s and r are shrinking $\boldsymbol{\theta}_{ik} = (\mu_{ik}, \gamma_{ik}, \alpha_{ik}, \beta_{ik}) \in \mathbb{R} \times [0, a] \times [0, 1]^2$ toward the parameter $\tilde{\boldsymbol{\theta}}_{ik}^* = (\tilde{\mu}_{ik}^*, \tilde{\gamma}_{ik}^*, \tilde{\alpha}_{ik}^*, \tilde{\beta}_{ik}^*) \in \mathbb{R} \times [0, a] \times [0, 1]^2$ which is assumed to be constant for all units in the same cluster, that is for all $i \in C_{hk}$ where $h = 1, \dots, M_k$ (for further details see Section 3 and Eq. 27). Since the parameters are non-identified due to the label switching problem, we follow a commonly used approach and impose a prior restriction on the intercepts $\mu_{i1} > \mu_{i2} > \dots > \mu_{iK}$ (e.g., see Celeux, 1998; Frühwirth-Schnatter, 2001, 2006).

The second stage of the hierarchy is generating the clusters of parameters. For each regime k we assume a Pitman-Yor process (PYP) prior

$$(10) \quad \tilde{\boldsymbol{\theta}}_{ik}^* | G_k \stackrel{iid}{\sim} G_k, \quad G_k \sim \text{PYP}(\nu, \psi, H_0)$$

with base measure H_0 and concentration and dispersion parameters $\nu \in [0, 1]$ and $\psi > -\nu$, respectively. We assume $H_0(\boldsymbol{\theta})$ is the product measure of the following independent normal and uniform distributions

$$(11) \quad \mathcal{N}(\mu; m^*, s^*), \quad \mathcal{U}(\gamma; 0, a), \quad \mathcal{U}(\alpha; 0, 1), \quad \mathcal{U}(\beta; 0, 1)$$

which are usually chosen as prior distributions in parametric Bayesian inference for MS-GARCH (e.g., see Billio, Casarin and Osuntuyi, 2016). The PYP introduced in Pitman and Yor (1997) is a generalization of the

Dirichlet process (DP) defined in Ferguson (1973) which can be obtained for $\nu = 0$.

Through the illustration of the Chinese Restaurant metaphor, the clustering structure of the PYP is defined by a Polya-Urn sampling scheme. The parameter θ_i^* of the i -th unit is either equal to one of the other units or a new one from the base distribution H_0 , i.e.:

$$(12) \quad \tilde{\theta}_{ik}^* | \tilde{\theta}_{1k}^*, \dots, \tilde{\theta}_{i-1k}^* = \frac{i}{\psi - \nu + i} \sum_{h=1}^{i-1} \delta_{\tilde{\theta}_{hk}^*}(\tilde{\theta}_{ik}^*) + \frac{\psi}{\psi - \nu + i} H_0(\tilde{\theta}_{ik}^*)$$

This sequential allocation procedure is generating clusters in the parameter space, where the number of clusters is random. The Pitman-Yor process induces the following prior distribution on the number of clusters M_k

$$P(M_k = h) = \frac{\nu^{h-1} \Gamma(\psi/\nu + h) \Gamma(\psi + 1)}{\Gamma(\psi/\nu + 1) \Gamma(\psi + N)} S_\nu(N, h)$$

with $h \in \mathbb{N}$, where $S_\nu(N, h)$ is a generalized Stirling number of the first kind, and $\Gamma(x)$ is the one-parameter gamma function (e.g., see Pitman (2006), Ch. 1 and 3). The following formula is used to evaluate the prior mean of the number of clusters:

$$\mathbb{E}(M_k) = \begin{cases} \sum_{h=1}^N \frac{\psi}{\psi+h-1}, & \text{if } \nu = 0, \\ \frac{\Gamma(\psi+\nu+N)\Gamma(\psi+1)}{\nu\Gamma(\psi+\nu)\Gamma(\psi+N)} - \frac{\psi}{\nu}, & \text{if } \nu \neq 0. \end{cases}$$

We summarize our Bayesian nonparametric model in the Directed Acyclic Graph representation of Fig. 1.

It is possible to show that the PYP clustering effects on the cross section of time series correspond to a probabilistic clustering of the parameters based on an infinite mixture distribution. The Pitman-Yor process prior can be written in a Sethuraman's like representation as a discrete random measure

$$(13) \quad G_k(\theta^*) = \sum_{h=1}^{\infty} W_{hk} \delta_{\theta_{hk}^*}(d\theta^*)$$

where the atoms θ_{hk}^* are i.i.d. random variables from the base measure H_0 and the random weights W_{hk} have the stick-breaking representation

$$(14) \quad W_{hk} = V_{hk} \prod_{l=1}^{h-1} (1 - V_{lk})$$

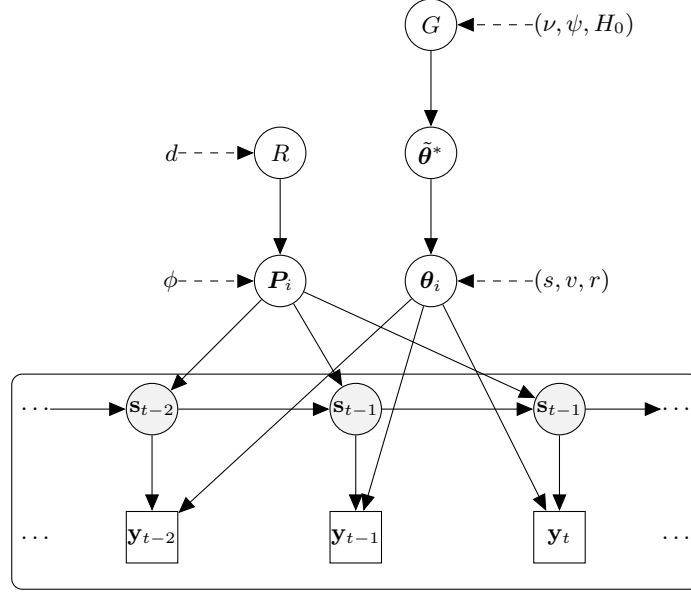


FIGURE 1. DAG of the Bayesian nonparametric MS-GARCH panel model. It exhibits the hierarchical structure of the observations $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})$ (boxes), the latent variables $\mathbf{s}_t = (s_{1t}, \dots, s_{Nt})$ (gray circles), the parameters $P_i = (p_{i,11}, \dots, p_{i,1K}, \dots, p_{i,K1}, \dots, p_{i,KK})$, $\boldsymbol{\theta}_i = (\mu_i, \gamma_i, \alpha_i, \beta_i)$, the hyperparameters of the first stage $R = (\mathbf{r}_1, \dots, \mathbf{r}_K)$, $\boldsymbol{\theta}_i^* = (\tilde{\mu}_i^*, \tilde{\gamma}_i^*, \tilde{\alpha}_i^*, \tilde{\beta}_i^*)$ and of the second stage G (white circles). The directed arrows show the causal dependence structure of the model.

with $V_{lk} \sim \mathcal{B}e(1 - \nu, \phi + \nu l)$ i.i.d. $l = 1, 2, \dots$ (see Arbel, Blasi and Prünster, 2019).

By integrating out the discrete part of the hierarchical prior one obtains the following infinite mixture representation of the prior distribution on $\boldsymbol{\theta}$

$$(15) \quad \boldsymbol{\theta}_{ik} | G_k \stackrel{ind}{\sim} \int \pi(\boldsymbol{\theta}_{ik} | \boldsymbol{\theta}^*) G_k(d\boldsymbol{\theta}^*) = \sum_{h=1}^{\infty} W_{hk} \pi(\boldsymbol{\theta}_{ik} | \boldsymbol{\theta}_{hk}^*)$$

where $\pi(\boldsymbol{\theta}_{ik} | \boldsymbol{\theta}^*)$ is the joint parameter distribution at the first stage of hierarchical prior (see Eqs. 8-9) and $G_k(\boldsymbol{\theta}^*)$ is the distribution at the second stage. In conclusion the PYP prior allows for probabilistic clustering in the parameter space.

The predictive density induced by our prior assumptions can be written as

$$(16) \quad y_{it} | G, s_{it} \stackrel{ind}{\sim} \sum_{h=1}^{\infty} W_{hs_{it}} \int f_t(y_{it} | s_{it}, \boldsymbol{\theta}) \pi(d\boldsymbol{\theta} | \boldsymbol{\theta}_{hs_{it}}^*)$$

where $f_t(y_{it}|s_{it} = k, \Theta) = f(y_{it}|\mu_{ik}, \sigma_{ik,t})$ is the transition kernel of the MS-GARCH with $\sigma_{ik,t}^2 = \gamma_{ik} + \alpha_{ik}\varepsilon_{it-1}^2 + \beta_{ik}\sigma_{it-1}^2$ for $k = 1, \dots, K$ and $\Theta = (\theta_1, \dots, \theta_K)$, $\theta_k = (\theta_{1k}, \dots, \theta_{Nk})$ and $\theta_{ik} = (\mu_{ik}, \gamma_{ik}, \alpha_{ik}, \beta_{ik})$. This prior predictive densities accounts for various forms of possible heterogeneity in the data such as asymmetry, excess of kurtosis and multimodality.

3. Posterior approximation. Let $\Theta = (\theta_1, \dots, \theta_K)$ be the collection of the unit- and regime-specific parameters $\theta_k = (\theta_{1k}, \dots, \theta_{Nk})$ and $\theta_{ik} = (\mu_{ik}, \gamma_{ik}, \alpha_{ik}, \beta_{ik})$, and $P = (P_1, \dots, P_N)$ the collection of transition probabilities. Let $Y = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ be the collection over time of the observation vector $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})$ and $S = (\mathbf{s}_1, \dots, \mathbf{s}_T)$ be the collection over time of the latent vectors $\mathbf{s}_t = (s_{1t}, \dots, s_{Nt})$. The likelihood function of the proposed MS-GARCH panel model is

$$(17) \quad L(Y|\Theta, P) = \sum_{s_1, \dots, s_T \in \{1, \dots, K\}} \prod_{t=1}^T \prod_{i=1}^N f(y_{it}|\theta_i, s_{it}) f(s_{it}|s_{it-1}, P_i)$$

where

$$(18) \quad f(s_{it}|s_{it-1}, P_i) = \prod_{k=1}^K \prod_{l=1}^k p_{i,kl}^{\mathbb{I}(s_{it}=l)\mathbb{I}(s_{it-1}=k)}$$

which is not tractable since it is written in integral form as usually in latent variable models. Nevertheless a data-augmentation principle can be applied (Tanner and Wong, 1987) in order to develop efficient posterior simulation methods. Following a common strategy in panel Markov-switching literature (e.g., see Billio et al., 2016; Casarin et al., 2019; Bianchi et al., 2019), we introduce the set of auxiliary allocation variables $\xi_{ikt} = \mathbb{I}(s_{it} = k)$ which allow us to write the complete-data likelihood function as follows

$$(19) \quad L(Y, \Xi|\Theta, P) = \prod_{t=1}^T \prod_{i=1}^N f(y_{it}|\theta_i, s_{it}) \prod_{k=1}^K \prod_{l=1}^k p_{i,kl}^{\xi_{ikt-1}\xi_{ilt}}$$

where $\Xi = (\Xi_1, \dots, \Xi_T)$ is the collection over time of the latent vectors $\Xi_t = (\xi_{1t}, \dots, \xi_{Nt})$ with $\xi_{it} = (\xi_{i1t}, \dots, \xi_{iKt})$.

The joint hierarchical prior distribution is

$$(20) \quad \pi(\Theta, G) = \prod_{k=1}^K \left(\prod_{i=1}^N \pi(\theta_{ik}|G_k) \prod_{l=1}^k p_{i,kl}^{r_l-1} \right) \pi(V_k) \pi(\Theta_k^*)$$

where

$$(21) \quad \pi(\boldsymbol{\theta}_{ik}|G_k) = \sum_{h=1}^{\infty} W_{hk} \pi(\boldsymbol{\theta}_{ik}|\boldsymbol{\theta}_{hk}^*)$$

is the infinite mixture prior where we recall $\pi(\boldsymbol{\theta}_{ik}|\boldsymbol{\theta}_{hk}^*) = \mathcal{N}(\mu_{ik}|\mu_{hk}^*, s)$ $\mathcal{B}e(\alpha_{ik}|\alpha_{hk}^*, r)$ $\mathcal{B}e(\beta_{ik}|\beta_{hk}^*, r)$ $\mathcal{B}e(\gamma_{ik}/a|\gamma_{hk}^*/a, r)$ is the first-stage joint prior distribution given in Eqs. 8-9 and

$$(22) \quad \pi(\Theta_k^*) = \prod_{h=1}^{\infty} \mathcal{N}(\mu_{hk}^*; m^*, s^*) \mathcal{U}(\alpha_{hk}^*; 0, 1) \mathcal{U}(\beta_{hk}^*; 0, 1) \mathcal{U}(\gamma_{hk}^*; 0, a)$$

$$(23) \quad \pi(V_k) = \prod_{l=1}^{\infty} \mathcal{B}e(V_{lk}; 1 - \nu, \psi + \nu l)$$

is joint distribution of the infinite collection of stick-breaking variables and atoms, $V_k = (V_{1k}, V_{2k}, \dots)$ and $\Theta_k^* = (\boldsymbol{\theta}_{1k}^*, \boldsymbol{\theta}_{2k}^*, \dots)$, respectively, which are involved in the definition of the random measures $G_k(\boldsymbol{\theta}_k^*)$ $k = 1, \dots, K$.

The joint prior distribution in a Bayesian nonparametric framework is usually not tractable since its support is the space of the discrete random measures which are infinite-dimensional objects (see Eqs. 21-23). Nevertheless, the data-augmentation principle can be applied in order to make the inference problem more tractable. Following the recent Bayesian nonparametrics literature (e.g., see Bassetti, Casarin and Leisen, 2014; Bassetti, Casarin and Ravazzolo, 2018; Billio, Casarin and Rossini, 2019), we introduce a set of slice variables $U_{ik} \sim \mathcal{U}(0, 1)$ and define the index set $\mathcal{A}_{ik} = \{h|U_{ik} < W_{hk}\}$. Then the infinite mixture can be demarginalized as follows

$$(24) \quad \begin{aligned} \pi(\boldsymbol{\theta}_{ik}|U_k, V_k, \boldsymbol{\theta}_k^*) &= \sum_{h=1}^{\infty} \mathbb{I}(U_{ik} < W_{hk}) \pi(\boldsymbol{\theta}_{ik}|\boldsymbol{\theta}_{hk}^*) \\ &= \sum_{h \in \mathcal{A}_{ik}} \mathbb{I}(U_{ik} < W_{hk}) \pi(\boldsymbol{\theta}_{ik}|\boldsymbol{\theta}_{hk}^*) \end{aligned}$$

which is a almost-surely finite mixture since $\text{Card}(\mathcal{A}_{ik}) < \infty$ a.s., where $U_k = (U_{1k}, \dots, U_{Nk})$ is the collection of slice variables.

Following the standard practice in finite mixture modelling we introduce the latent allocation variable $D_{ik} \in \mathcal{A}_{ik}$ and obtain

$$(25) \quad \pi(\boldsymbol{\theta}_{ik}|U_k, D_k, V_k, \boldsymbol{\theta}_k^*) = \mathbb{I}(U_{ik} < W_{D_{ik}k}) \pi(\boldsymbol{\theta}_{ik}|\boldsymbol{\theta}_{D_{ik}k}^*)$$

where $D_k = (D_{1k}, \dots, D_{Nk})$. Let us denote with $V = (V_1, \dots, V_K)$, $U = (U_1, \dots, U_K)$ and $\Theta^* = (\theta_1^*, \dots, \theta_K^*)$ the collections of regime-specific auxiliary variables and atoms. The joint posterior distribution $\pi(\Xi, \Theta, P, U, D, V, \Theta^* | Y)$ is proportional to

$$(26) \quad L(Y, \Xi | \Theta, P) = \prod_{k=1}^K \left(\prod_{i=1}^N \pi(\theta_{ik} | U_k, D_k, V_k, \Theta_k^*) \prod_{l=1}^k p_{i,kl}^{r_l-1} \right) \pi(V_k) \pi(\Theta_k^*).$$

Note that the allocation variables allows to reconcile the notations used in the hierarchical model of Eqs. 8-11 and the random measure representation in Eqs. 13-15 as follows:

$$(27) \quad \tilde{\theta}_{ik}^* = \theta_{D_{ik}k}^*$$

A Gibbs sampler is used to generate random samples from the joint posterior and to approximate the Bayesian estimator. The Gibbs sampler iterates the following steps

1. Sample slice and stick-breaking variables U and V given $\Xi, \Theta, P, D, \Theta^*, Y$
2. Sample the transition probabilities P given $\Xi, \Theta, U, D, V, \Theta^*, Y$
3. Sample the atoms Θ^* given $\Xi, \Theta, P, U, D, V, Y$
4. Sample the MS-GARCH parameters Θ given $\Xi, P, U, D, V, \Theta^*, Y$
5. Sample the switching allocation variables Ξ given $\Theta, P, U, D, V, \Theta^*, Y$
6. Sample the mixture allocation variables D given $\Xi, \Theta, P, U, V, \Theta^*, Y$

The derivation of the full conditional distributions is given in Appendix A.

4. Numerical illustrations.

4.1. *Simulation results.* For inference and model validation, we run a set of simulation experiments on synthetic datasets. In this section we report the results for one of the experiments, in which we examine the efficiency and effectiveness of our MCMC sampling scheme in estimating the number of clusters in each regime.

We generate a panel of 30 time series with length 300 each from the data generating process (DGP) corresponding to the model defined by Eqs. 1-9 for two regimes ($K = 2$), including their time-invariant transition probabilities and switching conditional mean and variance. The DGP is assumed to be as realistic as possible for illustrative purposes. In particular, the number of groups in the clusters across the two regimes is being kept relatively small.

In the first regime (i.e., $s_{it} = 1$), we assume that the units are clustered into two groups with equal probability. In formulas:

$$\begin{aligned} \mu_{i1} &= \begin{cases} 1 + 0.01\eta_{i1}, & \text{with probability } p_1 = 0.5, \\ 1.5 + 0.01\eta_{i1}, & \text{with probability } (1 - p_1) \end{cases} \\ \gamma_{i1} &= \begin{cases} 0.1 + 0.01\zeta_{i1}^2, & \text{with probability } p_1 = 0.5, \\ 0.2 + 0.01\zeta_{i2}^2, & \text{with probability } (1 - p_1), \end{cases} \\ (\alpha_{i1}, \beta_{i1}, x) &\sim \text{Dir}(1000(0.05, 0.8, 1 - 0.85)) \end{aligned}$$

In the second regime (i.e., $s_{it} = 2$), the units are clustered into three groups. In formulas:

$$\begin{aligned} \mu_{i2} &= \begin{cases} -1.1 + 0.01\eta_{i2}, & \text{with probability } p_1 = 0.3, \\ -1.5 + 0.01\eta_{i2}, & \text{with probability } p_2 = 0.3, \\ -1.0 + 0.01\eta_{i2}, & \text{with probability } (1 - p_1 - p_2), \end{cases} \\ \gamma_{i2} &= \begin{cases} 0.5 + 0.01\zeta_{i2}^2, & \text{with probability } p_1 = 0.3, \\ 0.8 + 0.01\zeta_{i2}^2, & \text{with probability } p_2 = 0.3, \\ 0.1 + 0.01\zeta_{i2}^2, & \text{with probability } (1 - p_1 - p_2), \end{cases} \\ (\alpha_{i2}, \beta_{i2}, x) &\sim \text{Dir}(1000(0.05, 0.8, 1 - 0.85)) \end{aligned}$$

where $\eta_{i1} \sim \mathcal{N}(0, 1)$, $\eta_{i2} \sim \mathcal{N}(0, 1)$, $\zeta_{i1} \sim \mathcal{N}(0, 1)$ and $\zeta_{i2} \sim \mathcal{N}(0, 1)$.

The transition probabilities are $p_{i,11} \sim \mathcal{Be}(1000p, 1000(1-p))$ and $p_{i,22} \sim \mathcal{Be}(1000p, 1000(1-p))$ i.i.d. for $i = 1, \dots, N$, where $p = 0.98$.

In Fig. 2, the true and the estimated values of the intercept (μ_i) of the measurement equation and GARCH parameter γ_i (see Eqs. 1 and 2) in regime 1 (red) and 2 (blue) are illustrated.¹ The findings seems to reveal that the inference is able to recover the true value of the parameters.

Figure 3 shows that data are informative about the number of clusters in each regimes, and there is a substantial revision of the prior distributions (red) and the posterior distributions (blue) concentrate about the true number of clusters in the two regimes. For our simulated dataset, the Maximum a Posteriori (MAP) estimation of the number of cluster is 1 for the first regime and 3 for the second regime.

4.2. Volatility clusters in the S&P 100. We consider 78 assets of the 101 constituents of the SP&100 index and collect the percentage log-returns at a weekly frequency. We do this to have a balanced panel of observations from 6th January to 3rd October 2020 (the sectorial classification of these assets is reported in Tab. 1 of Appendix C).

¹For other parameters and the trajectory of the Markov chain see Appendix 8.

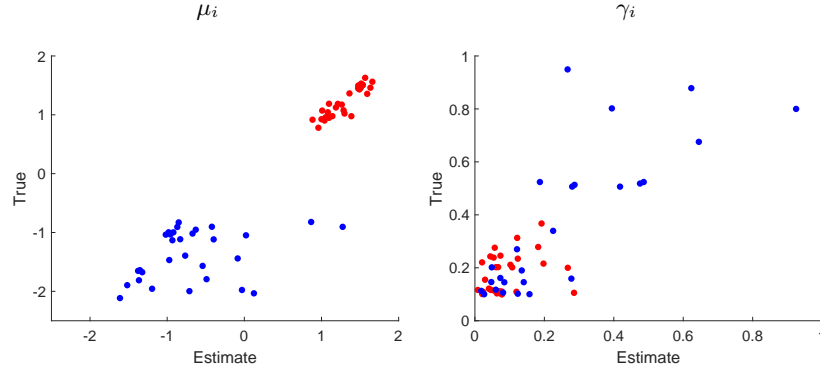


FIGURE 2. *Simulation results. True (vertical axis) and estimated (horizontal axis) values of the intercept (μ_i) of the measurement equation and GARCH parameter γ_i for each unit i in regime 1 (red) and 2 (blue).*

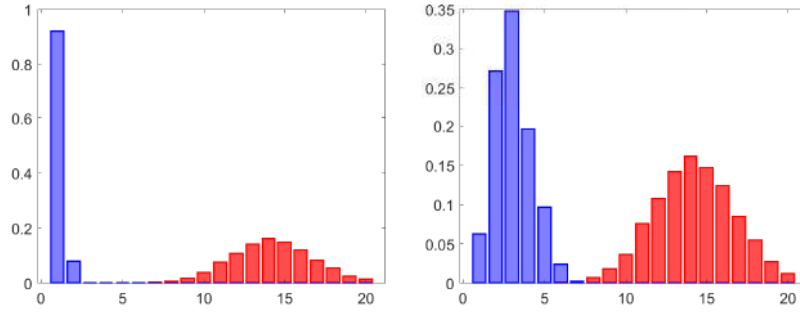


FIGURE 3. *Simulation results. Prior (red) and posterior (blue) distribution of the number of clusters in regime 1 (left) and 2 (right).*

The empirical analysis aims to identify regimes of under-performance and over-performance of expected returns. Moreover, we use the sector classification and some fundamental financial ratios to study the composition of the clusters.

As a preliminary analysis, we plot in Fig. 4 the estimates of the log-volatility and log-kurtosis of the 78 constituents considered in the analysis. This figure also shows the cross-sectional distribution of the log-volatility and log-kurtosis. The figure indicates that volatility and kurtosis change over

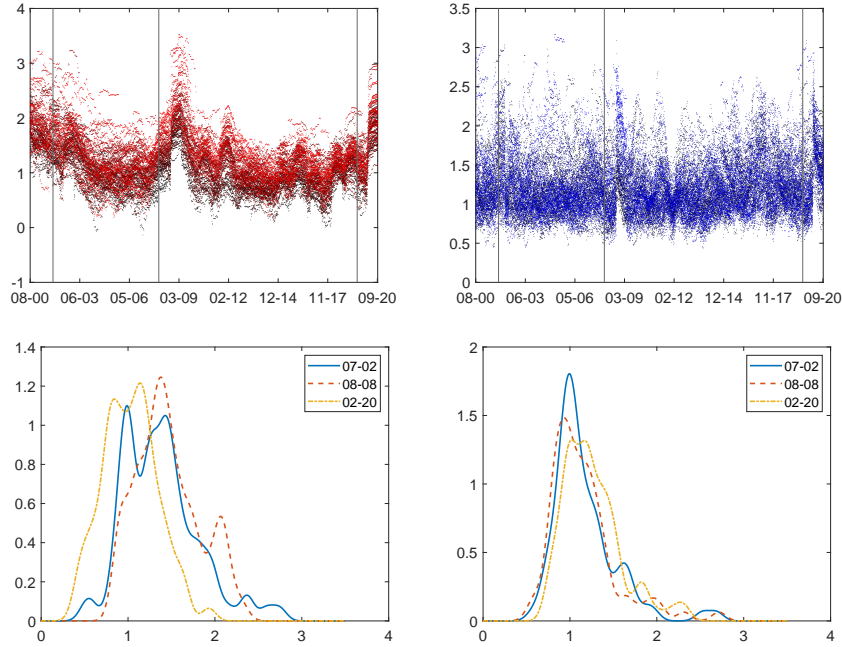


FIGURE 4. *Top: rolling window estimates of the log-volatility (left) and log-kurtosis (right) for the SP&100's constituents for the period 6th January 2000 to 3rd October 2020 (30 weeks window). Vertical bars indicate three reference dates: 6th July 2002, 23rd August 2008 and 22nd February 2020. Bottom: cross-sectional distribution of the log-volatility (left) and log-kurtosis (right) in three reference dates.*

time with time series clustering effects (see top plots of Fig. 4). This seems to suggest the use of GARCH and Markov-switching models. Furthermore, the cross-sectional distribution of the volatility and kurtosis exhibits multiple modes and long tails (see bottom plots of Fig. 4).² This fact seems to imply cross-section heterogeneity in the data with possible clustering effects in the parameters of the GARCH process. These effects cannot be captured only by a Markov-switching (MS) model, therefore there is a need of combining the MS-GARCH with a probabilistic clustering mechanism.

In our analysis, we first identify the two regimes, and then the clusters of assets within each regime. Lastly, we use the sector classification and some fundamental financial ratios to study the composition of the clusters.

²In Fig. 11 of Appendix C, we also report the estimates of the cross-sectional distribution of the log-volatility (left) and log-kurtosis (right) of the SP&100's constituents log-returns in the three dates (6th July 2002, 23rd August 2008 and 22nd February 2020) for three different sizes of the rolling window. The results show that the preliminary evidence on cross-sectional heterogeneity is robust with respect to the choice of the window size.

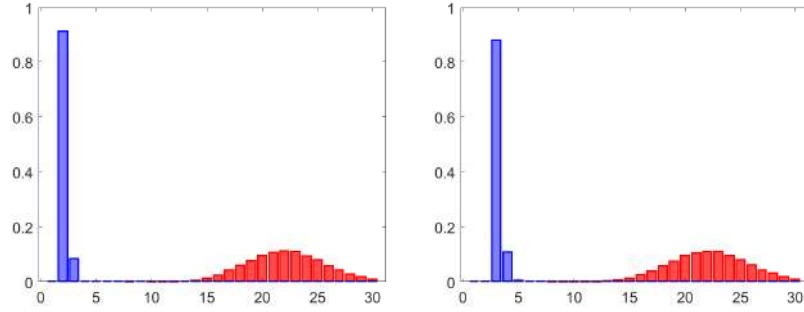


FIGURE 5. Prior (red) and posterior (blue) distribution of the number of clusters in the over-performing regime 1 (left) and under-performing regime 2 (right).

Regime identification is achieved by ordering the expected returns $\mu_{i1} > \mu_{i2}$, such that regime 1 corresponds to a relative over-performance state and regime 2 to an under-performance one. This identification constraint is strongly supported by the data and allows us to separate the assets returns in two performance regimes (see Fig. 12 in Appendix C).

Regarding the cluster identification, Fig. 5 reports the prior (red) and posterior (blue) distribution of the number of clusters in regime 1 (left) and 2 (right). We set $\nu = 0$ and $\phi = 10$ in the PYP prior in order to have quite diffuse prior distributions. The posterior distribution is concentrated suggesting a substantial revision of the prior information and the MAP estimates of the number of clusters is 2 and 3 for regime 1 and 2, respectively.

We also estimate the co-clustering probability matrix to delve into the composition of the clusters. In this way, the probability $\mathbb{P}(\{D_{ik} = D_{jk}\} | M_k, Y)$ that the parameters θ_{ik}^* and θ_{jk}^* are in the same cluster is given. This probability can be easily approximate by using the MCMC samples as follow

$$(28) \quad \frac{1}{\text{Card}(\mathcal{R}_k)} \sum_{r \in \mathcal{R}_k} \delta(D_{jk}^{(r)} - D_{ik}^{(r)})$$

where $D_{ik}^{(r)}$ is a sample of the allocation variable for the i -unit parameters in the regimes k and $\mathcal{R}_k = \{r = 1, \dots, R | N_k^{(r)} = M\}$ contains the values of MCMC iterations such that the parameters of the panel units have been allocated to exactly M mixture components. Note that a spectral clustering algorithm have been applied to re-order the series and to provide better graphical representation of the clusters.

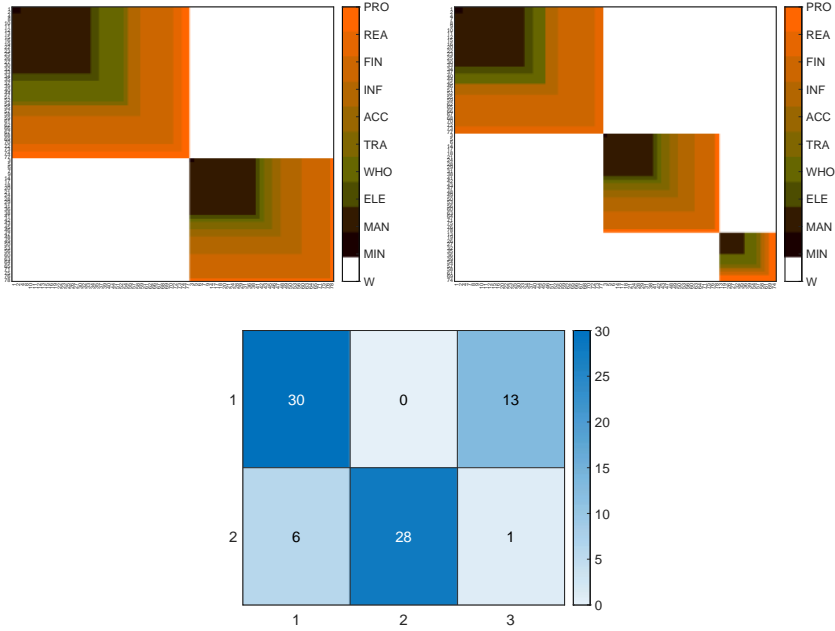


FIGURE 6. *Top: posterior co-clustering matrix in regime 1 (left) and 2 (right). In each block, colors represent the sector labels of the units. Bottom: the number of assets (cell entries) shared by two clustering structures in regime 1 (vertical axis) and regime 2 (horizontal axis).*

The top panel of Fig. 6 report the co-clustering matrices for the two regimes. In each block matrix, the algorithm identifies a constituent (asset) belonging to a cluster with the label 1 (color patch) and 0 (white patch) otherwise. The colors represent the sectors in the clusters. Following Wade and Ghahramani (2018), we use the variation of information (VI) metric proposed by Meilâ (2007) to compare the two regimes (in terms of clusters). This measure compares the information in the two regimes with the information shared between the two regimes. We compute the normalized value of VI (0.20), which suggests a substantial difference between the clustering and composition in the two regimes.³

The bottom panel of Fig. 6 shows the relationship between the clustering structures of the two regimes. We order the clusters following the numbers of constituents from the largest to the smallest. Most of the assets in cluster 1 in the first regime belong to cluster 1 in the second regime, whereas many

³VI lies in the interval $0 - \log_2(N)$ and a normalize value is obtained dividing VI by $\log_2(N)$.

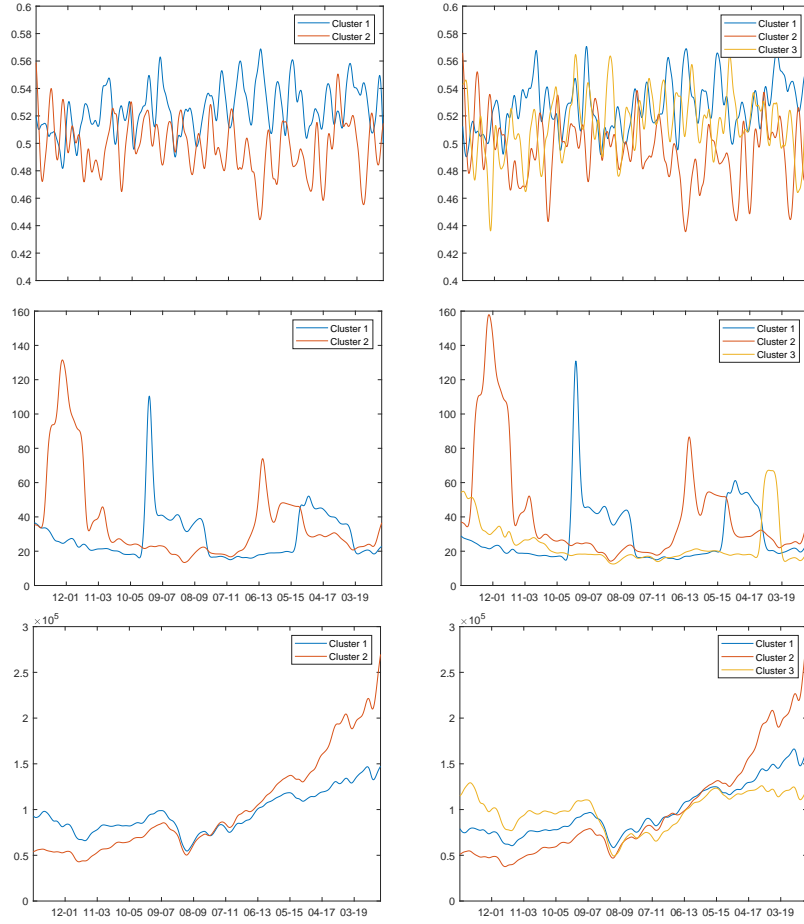


FIGURE 7. Average overperforming probability of the assets in the clusters (first row), Price-to-Earning (second row), and Market Capitalization (third row) for the assets in the clusters of regime 1 (left) and of regime 2 (right).

assets of the first group in regime 1 belong to the third group in regime 2. In particular, for cluster 1 in both regimes, we observe that the majority of the sectors representing the assets are: manufacturing (about 40% in both regimes), financial and insurance (19% in the first regime), and wholesale and retail (25% in the second regime). Similar results for the sectors are found for cluster 2 in the two regimes. More specifically, the manufacturing sector represents about 40% of the assets in the two regimes, while the financial and insurance sector is about 18% for regime 2, and information and communication is around 20% (for details on sectors see Tabs. 2 and 3

in Appendix C).

Further, in order to characterize the clusters in terms of the market size of the constituents, we first rank the companies by computing the average size of each of them using the last year of the sample period. Then, we classify the assets into three groups, namely small (bottom 30%), medium (middle 40%) and big (top 30%) companies (see Tab. 4). We also compute the percentage of companies belonging to the clusters in each regime in terms of size. In regime 1, companies with the medium size represent the largest majority about 40%, and a similar outcome is also observed for regime 2. More specifically, the following emerge.

In regime 1, we have:

1. Cluster 1 is characterized by 40% and 20% small and large size companies, respectively. This cluster also shows values on average of the market size, returns and their standard deviation equal to 1.15×10^5 , 0.21 and 4.72, respectively.⁴
2. Cluster 2 consists of 17% small and 42% medium size companies with an average market size of 1.50×10^5 , a return equal to 0.34 and standard deviation of 4.18.

The clusters composition in Regime 2 is as follows:

1. Cluster 1 comprises of small (39%) and medium size firms (36%) with an average market size of 1.26×10^5 , return of -0.02 and standard deviation of 4.27.
2. 17% and 42% of the assets in cluster 2 are small and medium size companies, respectively. While, on average the market capitalization, return and standard deviation are 1.48×10^5 , 0.15 and 4.50, respectively.
3. Cluster 3 is characterized by 29% small and 50% medium size companies. Moreover, the average market capitalization, returns and standard deviation are 1.09×10^5 , 0.08 and 4.78.

For the cluster composition, we also compute the value of the Price-to-Earnings ratio (PE) for all the clusters⁵. For regime 1, clusters 1 and 2 have values of PE equal to 26.97 and 32.38, respectively. For regime 2, these values are 27.08, 35.69 and 23.72 for clusters 1, 2 and 3, respectively. These results seem to indicate that in both regimes cluster 2 is overvalued.

⁴We computed the average return and standard deviation using the whole sample, while the average market size is calculated using the last 10 years of the sample period.

⁵Following the standard practice in style analysis the average PE is computed over the last 10 years.

To provide additional information on the composition of the clusters, we plot the dynamics over time of the average probabilities of the assets in each cluster, the market size and the PE in Fig. 7. Regarding the average probabilities, cluster 1 shows the highest probability in regime 1, while in regime 2 clusters 1 and 3 display similar probabilities. As for PE, the dynamics in clusters 1 and 2 in regime 1 resemble those in regime 2 (see second line of Fig. 7), and the market capitalization pattern indicates that relatively to cluster 1 in the over-performance state, assets in cluster 2 seems to show lowest values before the 2008/09 Global financial crisis and larger afterwards. The same dynamics for this two clusters is also observed in the under-performance state.

The following summarizes the results. There is evidence of time-varying clustering structures in our panel of time series. Regime 1 (over-performance phase) and regime 2 (under-performance phase) comprise 2 and 3 clusters, respectively. The composition of the clusters varies across regimes while some similarities in terms capitalization and financial ratios are observed. In regime 1 and 2 medium size companies represent the largest majority (about 40%) and assets in cluster 2 seem to be over valued by their PE ratio. The pattern of the market capitalization is similar across the two regimes as cluster 2 shows lower (larger) values compared to cluster 1 before (after) the 2008/09 Global financial crisis.

5. Conclusion. The increase of interest in the study of volatility of large panels of asset returns and the evidence of regimes in volatility of financial returns has suggested to adapt Markov switching models to GARCH effect. In this respect, this paper introduces a new model for panel data with Markov-switching GARCH effects.

In particular, we propose to model cross-sectional clustering effects with a Bayesian nonparametric technique that considers a hierarchical Pitman-Yor process prior for the Markov Switching GARCH parameters. The Bayesian nonparametric approach is a two-stage procedure. In the first stage, the hierarchical prior allows for cross-unit heterogeneity, while shrinking all unit-specific parameters towards a common mean. In the second stage, the hierarchical procedure allows for mixed effects in the common mean.

This paper makes a contribution in some respects. First, the new model allows us to make inference on the number of mixture components in the cross-sectional clustering. Second, the model is sufficiently flexible to embody different shapes of the prior and posterior predictive distributions. Third, uncertainty and the number of mixture components are incorporated in the predictive distribution. Lastly, through a data-augmentation strategy, this

paper makes the inference more tractable for our high dimensional model. A simulation exercise is carried out for inference and model validation.

We apply the new model to 78 assets of the SP&100 index from 6th January 2000 to 3rd October 2020. Our results may have some implications for portfolio making and for style investing decisions. The evidence shows that regime 1 (over-performance phase) and regime 2 (under-performance phase) differ in terms of clustering structures comprising 2 and 3 clusters, respectively. Within each regime the clusters differ substantially in terms of over-performance probability and in terms of style features, when considering capitalization and Price-to-Earnings. The heterogeneity of the clusters in terms of sectors and styles allows for portfolio diversification. Across regimes, the composition of the clusters changes, nevertheless some clusters share some similarities in terms of style features, allowing for the implementation of rotating style strategies.

Further research may consider the choice of the number of performance regimes, the sensitivity with respect to nonparametric prior specification and some forecasting comparisons with exogenous clustering models.

Acknowledgements. This research used the SCSCF multiprocessor cluster system provided by the Venice Centre for Risk Analytics (VERA) at University Ca' Foscari of Venice.

References.

- ANG, A. and TIMMERMANN, A. (2012). Regime Changes and Financial Markets. *Annual Review of Financial Economics* **4** 313-337.
- ARBEL, J., BLASI, P. D. and PRÜNSTER, I. (2019). Stochastic Approximations to the Pitman–Yor Process. *Bayesian Analysis* **14** 1201–1219.
- ARBEL, J., MENGENSEN, K. and ROUSSEAU, J. (2016). Bayesian nonparametric dependent model for partially replicated data: The influence of fuel spills on species diversity. *The Annals of Applied Statistics* **10** 1496–1516.
- ARDIA, D. (2008). *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications, volume 612 of Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, Germany.
- ARDIA, D., BLUTEAU, K., BOUDT, K. and CATANIA, L. (2018). Forecasting Risk with Markov-switching GARCH models: A large-scale performance study. *International Journal of Forecasting* **34** 733-747.
- BARIGOZZI, M., BROWNLEES, G. M. C. GALLO and VEREDAS, D. (2014). Disentangling Systematic and Idiosyncratic Dynamics in Panels of Volatility Measures. *Journal of Econometrics* **162** 364-384.
- BASSETTI, F., CASARIN, R. and LEISEN, F. (2014). Beta-product Dependent Pitman-Yor Processes for Bayesian Inference. *Journal of Econometrics* **180** 49-72.
- BASSETTI, F., CASARIN, R. and RAVAZZOLO, F. (2018). Bayesian Nonparametric Calibration and Combination of Predictive Distributions. *Journal of the American Statistical Association* **522** 675-685.

- BAUWENS, L., DUFAYS, A. and ROMBOUTS, J. (2014). Marginal Likelihood for Markov-switching and Change-Point GARCH. *Journal of Econometrics* **178** 508-522.
- BAUWENS, L. and OTRANTO, E. (2016). Modeling the Dependence of Conditional Correlations on Market Volatility. *Journal of Business and Economic Statistics* **34** 254-268.
- BAUWENS, L. and OTRANTO, E. (2020). Nonlinearities and Regimes in Conditional Correlations with Different Dynamics. *Journal of Econometrics* **217** 496-522.
- BAUWENS, L., PREMINGER, A. and ROMBOUTS, J. (2010). Theory and Inference for a Markov Switching GARCH Model. *Econometrics Journal* **13** 218-244.
- BAUWENS, L. and ROMBOUTS, J. (2007). Bayesian Clustering of Many GARCH Models. *Econometric Reviews* **26** 365-86.
- BIANCHI, D., BILLIO, M., CASARIN, R. and GUIDOLIN, M. (2019). Modeling Systemic Risk with Markov Switching Graphical SUR Models. *Journal of Econometrics* **210** 58-74.
- BILLIO, M., CASARIN, R. and OSUNTUYI, A. (2016). Efficient Gibbs Sampling for Markov Switching GARCH Models. *Computational Statistics and Data Analysis* **100** 37-57.
- BILLIO, M., CASARIN, R. and ROSSINI, L. (2019). Bayesian Nonparametric Sparse VAR Models. *Journal of Econometrics* **212** 97-115.
- BILLIO, M., CASARIN, R., RAVAZZOLO, F. and VAN DIJK, H. (2016). Interactions between Eurozone and US Booms and Busts: A Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics* **31** 1352-1370.
- BOLLERSLEV, T., PATTON, A. J. and QUAEDVLIIEG, R. (2020). Multivariate Leverage Effects and Realized Semicovariance GARCH Models. *Journal of Econometrics* **217** 411-430.
- CASARIN, R., MOLINA, G. and TER HORST, E. (2019). A Bayesian time varying approach to risk neutral density estimation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **182** 165-195.
- CASARIN, R., FORONI, C., MARCELLINO, M. and RAVAZZOLO, F. (2019). Uncertainty Through the Lenses of a Mixed-Frequency Bayesian Panel Markov Switching Model. *Annals of Applied Statistics* **12** 2559-2586.
- CELEUX, G. (1998). Bayesian Inference for Mixture: The Label Switching Problem. In *Compstat* (R. Payne and P. Green, eds.) Physica, Heidelberg.
- DI LUCCA, M., GUGLIELMI, A., MULLER, P. and QUINTANA, F. (2013). A Simple Class of Bayesian Nonparametric Autoregression Models. *Bayesian Analysis* **8** 63-88.
- DO, K.-A., MULLER, P. and TANG, F. (2005). A Bayesian mixture model for differential gene expression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54** 627-644.
- DUFAYS, A. (2015). Infinite-State Markov-Switching for Dynamic Volatility. *Journal of Financial Econometrics* **14** 418-460.
- ELLIOTT, R. J., LAU, J. W., MIAO, H. and SIU, T. K. (2012). Viterbi-Based Estimation for Markov Switching GARCH Models. *Applied Mathematical Finance* **19** 1-13.
- FERGUSON, T. S. (1973). A Bayesian Analysis of some Nonparametric Problems. *The Annals of Statistics* **1** 209-230.
- FRÜHWIRTH-SCHNATTER, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* **96** 194-209.
- FRÜHWIRTH-SCHNATTER, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.
- GRAY, S. F. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime-switching Process. *Journal of Financial Economics* **42** 27-62.
- GRIFFIN, J. and KALLI, M. (2018). Bayesian Nonparametric Vector Autoregressive

- Models. *Journal of Econometrics* **203** 267-282.
- GRIFFIN, J. E. and STEEL, M. F. J. (2011). Stick-breaking Autoregressive Processes. *Journal of Econometrics* **162** 383-396.
- HAAS, M. and LIU, J. C. (2018). A Multivariate Regime-switching GARCH Model with an Application to Global Stock Market and Real Estate Equity Returns. *Studies in Nonlinear Dynamics and Econometrics* **22**.
- HAAS, M., MITTNIK, S. and PAOLELLA, M. (2004). A new Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics* **2** 493-530.
- HATJISPYROS, S. J., NICOLERIS, T. N. and WALKER, S. G. (2011). Dependent Mixtures of Dirichlet Processes. *Computational Statistics & Data Analysis* **55** 2011-2025.
- HE, Z. and MAHEU, J. M. (2010). Real Time Detection of Structural Breaks in GARCH Models. *Computational Statistics and Data Analysis* **54** 2628-2640.
- HENNEKE, J. S., RACHEV, S. T., FABOZZI, F. J. and METODI, N. (2011). MCMC-based Estimation of Markov Switching ARMA-GARCH Models. *Applied Economics* **43** 259-271.
- HIRANO, K. (2002). Semiparametric Bayesian Inference in Autoregressive Panel Data Models. *Econometrica* **70** 781-799.
- HJORT, N. L., HOMES, C., MÜLLER, P. and WALKER, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.
- JENSEN, J. M. and MAHEU, M. J. (2010). Bayesian Semiparametric Stochastic Volatility Modeling. *Journal of Econometrics* **157** 306-316.
- KALLI, M., GRIFFIN, J. E. and WALKER, S. G. (2011). Slice Sampling Mixture Models. *Statistics and Computing* **21** 93-105.
- KAUFMANN, S. (2010). Dating and Forecasting Turning Points by Bayesian Clustering with Dynamic Structure: A Suggestion with an Application to Austrian Data. *Journal of Applied Econometrics* **25** 309-344.
- KAUFMANN, S. (2015). K-state Switching Models with Time-varying Transition Distributions: Does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics* **187** 82-94.
- KLAASSEN, F. (2002). Improving GARCH Volatility Forecasts with Regime Switching GARCH. *Empirical Economics* **27** 363-394.
- LO, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics* **12** 351-357.
- MACLEHOSE, R. and DUNSON, D. (2010). Bayesian Semiparametric Multiple Shrinkage. *Biometrics* **66** 455-462.
- MEILÂ, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis* **98** 873 - 895.
- MÜLLER, P. and MITRA, R. (2013). Bayesian Nonparametric Inference – Why and How. *Bayesian Analysis* **8** 269-302.
- NAKATSUMA, T. (1998). A Markov-chain Sampling Algorithm for GARCH Models. *Studies in Nonlinear Dynamics and Econometrics* **3** 107-117.
- NIETO-BARAJAS, L. E. and QUINTANA, F. A. (2016). A Bayesian Non-Parametric Dynamic AR Model for Multiple Time Series Analysis. *Journal of Time Series Analysis* **37** 675-689.
- PAKEL, C., SHEPHARD, N. and SHEPPARD, K. (2011). Nuisance Parameters, Composite Likelihood, and a Panel of GARCH Models. *Statistica Sinica* **21** 307-329.
- PITMAN, J. (2006). *Combinatorial Stochastic Processes, volume 1875 of Lecture Notes in Mathematics*. Springer-Verlag, Berlin.
- PITMAN, J. and YOR, M. (1997). The two Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator. *Annals of probability* **25** 855-900.

- TADDY, M. A. and KOTTAS, A. (2009). Markov Switching Dirichlet Process Mixture Regression. *Bayesian Analysis* **4** 793-816.
- TANNER, M. and WONG, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* **82** 528-550.
- VIRBICKAITE, A., AUSÍN, M. C. and GALEANO, P. (2015). Bayesian inference methods for univariate and multivariate GARCH models: A survey. *Journal of Economic Surveys* **29** 76-96.
- WADE, S. and GHAHRAMANI, Z. (2018). Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion). *Bayesian Analysis* **13** 559-626.
- WALKER, S. G. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation* **36** 45-54.
- WANG, H. (2010). Sparse Seemingly Unrelated Regression Modelling: Applications in Finance and Econometrics. *Computational Statistics & Data Analysis* **54** 2866-2877.
- WEE, D. C. H., CHEN, F. and DUNSMUIR, W. T. M. (2020). Likelihood Inference for Markov Switching GARCH(1,1) Models using Sequential Monte Carlo. *Econometrics and Statistics*.
- XU, Y., MÜLLER, P., WAHED, A. S. and THALL, P. F. (2016). Bayesian Nonparametric Estimation for Dynamic Treatment Regimes With Sequential Transition Times. *Journal of the American Statistical Association* **111** 921-950.
- ZHANG, L., GUINDANI, M., VERSACE, F., ENGELMANN, J. M., VANNUCCI, M. et al. (2016). A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. *The Annals of Applied Statistics* **10** 638-666.

APPENDIX A: PROOF OF THE RESULTS IN SECTION 3

We introduce for $h \geq 1$ the set of parameters allocated to the h -th mixture component in the regime k , $\mathcal{D}_{hk} = \{i = 1, \dots, N | D_{ik} = h\}$ and the set of the non-empty mixture components $\mathcal{D}_k^* = \{h | \mathcal{D}_{hk} \neq \emptyset\}$. The number of stick-breaking components needed for the finite mixture representation is $D_k^* = \max\{D_{ik}, i = 1, \dots, N\}$. When sampling from the full conditional distribution of Θ_k^* and V_k only N_k^* element are sampled where N_k^* is the smallest integer such that $\sum_{h=1}^{N_k^*} W_{hk} > 1 - U_k^*$ where $U_k^* = \min\{U_{ik}, i = 1, \dots, N\}$.

A.1. Full conditional distribution of V and U . Let us split V_k in three blocks: $V_k^* = \{V_{lk} : l \in \mathcal{D}_k^*\}$, $V_k^{**} = \{V_{kD_k^*+1}, \dots, V_{kD_k^*+N_k^*}\}$ and $V_k^{***} = \{V_{lk} : l > N_k^*\}$. The samples are generated from a collapsed Gibbs step

1. the full conditional of the elements in V_k^* given $\Xi, \Theta, P, D, \Theta^*, Y$

$$(29) \quad f(V_{lk} | \dots) \propto \mathcal{B}e \left(1 - \nu + \sum_{i=1}^N \mathbb{I}(D_{ik} = l), \psi + \nu l + \sum_{i=1}^N \mathbb{I}(D_{ik} > l) \right)$$

for $l \leq D_k^*$,

2. the full conditional of the elements elements of V_k^{**} and V_k^{***} given $\Xi, \Theta, P, D, V^*, \Theta^*, Y$, which coincide with the prior distributions $\mathcal{B}e(1 - \nu, \psi + \nu l)$ for $l > D_k^*$
3. the full conditional of the elements of U_k given V and $\Xi, \Theta, P, D, \Theta^*, Y$

$$(30) \quad f(U_{ik} | \dots) \propto \mathbb{I}(U_{ik} < W_{D_{ik}k})$$

which is uniform on the interval $(0, W_{D_{ik}k})$

A.2. Full conditional distribution of P and R . We apply a collapsed-Gibbs step and sample \mathbf{r}_k $k = 1, \dots, K$ given Ξ, Θ, D, V, Y and \mathbf{p}_k $k = 1, \dots, K$ from its conditional given R and $\Xi, \Theta_{-p}, D, V, Y$. As regards the transition probabilities, from standard calculations in Markov-switching regression models we obtain

$$(31) \quad f(\mathbf{p}_{i,k} | \dots) \propto \prod_{h=1}^K p_{i,kh}^{\phi r_{kh} + n_{i,kh} - 1} \propto \mathcal{D}(\phi r_{k1} + n_{i,k1}, \dots, \phi r_{kK} + n_{i,kK})$$

where

$$(32) \quad n_{i,kh} = \sum_{t=1}^T \xi_{ikt-1} \xi_{iht}$$

The marginal distribution is

$$(33) \quad f(\mathbf{r}_k | \dots) \propto \int_{\Delta_{[0,1]^K}^N} \prod_{i=1}^N \prod_{h=1}^K p_{i,kh}^{\phi r_{kh} + n_{i,kh} - 1} \frac{\Gamma(\phi)}{\Gamma(\phi r_{kh})} dp_{i,kh} \pi(\mathbf{r}_k) \\ \propto \left(\prod_{h=1}^K r_{kh}^{d-1} \right) \left(\prod_{i=1}^N \prod_{h=1}^K \frac{\Gamma(\phi r_{kh} + n_{i,kh})}{\Gamma(\phi + n_{i,k})} \frac{\Gamma(\phi)}{\Gamma(\phi r_{kh})} \right)$$

where $n_{i,k} = n_{i,k1} + \dots + n_{i,kK}$ and $\Delta_{[0,1]^K}^N = \{(p_1, \dots, p_K) \in \mathbb{R}^K | p_k > 0 \forall k, p_1 + \dots + p_K = 1\}$ is the K -dim standard simplex. From the properties of the gamma functions

$$\Gamma(\phi r_{kh} + n_{i,kh}) = \prod_{l=1}^{n_{i,kh}} (\phi r_{kh} + l - 1) \Gamma(\phi r_{kh}) \\ \Gamma(\phi + n_{i,k}) = \prod_{l=1}^{n_{i,k}} (\phi + l - 1) \Gamma(\phi)$$

we obtain

$$(34) \quad f(\mathbf{r}_k | \dots) \propto \mathcal{D}ir(d + m_{k1}, \dots, d + m_{kK}) g(\mathbf{r}_k)$$

where

$$g(\mathbf{r}_k) = \prod_{i=1}^N \left(\prod_{l=1}^{n_{i,k}} (\phi + l - 1) \right)^{-1} \prod_{h=1}^K \prod_{l=2}^{n_{i,kh}} (\phi r_{kh} + l - 1)$$

and $m_{kh} = \text{Card}(\mathcal{M}_{kh})$, $\mathcal{M}_{kh} = \{i = 1, \dots, N | n_{i,kh} > 0\}$. Samples from this full conditional distribution are obtain by a Metropolis-Hastings algorithm with independent proposal distribution $\text{Dir}(d + m_{k1}, \dots, d + m_{kK})$.

A.3. Full conditional distribution of Θ^* . The full conditional distribution of $\theta_{hk}^* = (\mu_{hk}^*, \gamma_{hk}^*, \alpha_{hk}^*, \beta_{hk}^*)$ can be sampled by simulating iteratively from the following conditional distributions. The full conditional of μ_{hk}^* :

$$(35) \quad \begin{aligned} f(\mu_{hk}^* | \dots) &\propto \mathcal{N}(\mu_{hk}^* | m^*, s^*) \prod_{i \in \mathcal{D}_{hk}} \mathcal{N}(\mu_{ik} | \mu_{hk}^*, s) \\ &\propto \mathcal{N}(\mu_{hk}^* | \bar{m}_{hk}, \bar{s}_{hk}) \end{aligned}$$

where,

$$\bar{m}_{hk} = \bar{s}_{hk}^2 \left(\frac{m^*}{s^{*2}} + \frac{\sum_{i \in \mathcal{D}_{hk}} \mu_{ik}}{s^2} \right), \quad \text{and} \quad \bar{s}_{hk} = \left(\frac{1}{s^{*2}} + \frac{\text{Card}(\mathcal{D}_{hk})}{s^2} \right)^{-1/2}$$

The full conditional distribution of γ_{hk}^*

$$(36) \quad \begin{aligned} f(\gamma_{hk}^* | \dots) &\propto \mathbb{I}_{[0,a]}(\gamma_{hk}^*) \prod_{i \in \mathcal{D}_{hk}} \text{Be}(\gamma_{ik}/a | r\gamma_{hk}^*/a, r(1 - \gamma_{hk}^*/a)) \\ &\propto \mathbb{I}_{[0,a]}(\gamma_{hk}^*) \prod_{i \in \mathcal{D}_{hk}} \frac{\exp\{(r\gamma_{hk}^*/a - 1) \log(\gamma_{ik}/a) + (r(1 - \gamma_{hk}^*/a) - 1) \log(1 - \gamma_{ik}/a)\}}{\Gamma(r\gamma_{hk}^*/a) \Gamma(r(1 - \gamma_{hk}^*/a))} \\ &\propto \exp\{-\kappa_{hk} \gamma_{hk}^*\} \left(\frac{1}{\Gamma(r\gamma_{hk}^*/a) \Gamma(r(1 - \gamma_{hk}^*/a))} \right)^{\text{Card}(\mathcal{D}_{hk})} \mathbb{I}_{[0,a]}(\gamma_{hk}^*) \end{aligned}$$

where

$$\kappa_{hk} = \frac{r}{a} \sum_{i \in \mathcal{D}_{hk}} \log((a - \gamma_{ik})/\gamma_{ik})$$

which can be simulated exactly by the inverse cdf method where the cdf is

$$(1 - \exp\{-\kappa_{hk} \gamma_{hk}^*\}) \frac{1}{1 - \exp\{-a\kappa_{hk}\}} \mathbb{I}_{[0,a]}(\gamma_{hk}^*).$$

The full conditional distribution of α_{hk}^*

$$\begin{aligned}
(37) \quad f(\alpha_{hk}^* | \dots) &\propto \mathbb{I}_{[0,1]}(\alpha_{hk}^*) \prod_{i \in \mathcal{D}_{hk}} \mathcal{B}e(\alpha_{ik} | r\alpha_{hk}^*, r(1 - \alpha_{hk}^*)) \\
&\propto \mathbb{I}_{[0,1]}(\alpha_{hk}^*) \prod_{i \in \mathcal{D}_{hk}} \frac{\exp\{(r\alpha_{hk}^* - 1) \log(\alpha_{ik}) + (r(1 - \alpha_{hk}^*) - 1) \log(1 - \alpha_{ik})\}}{\Gamma(r\alpha_{hk}^*) \Gamma(r(1 - \alpha_{hk}^*))} \\
&\propto \exp\{-\tau_{hk} \alpha_{hk}^*\} \left(\frac{1}{\Gamma(r\alpha_{hk}^*) \Gamma(r(1 - \alpha_{hk}^*))} \right)^{\text{Card}(\mathcal{D}_{hk})} \mathbb{I}_{[0,1]}(\alpha_{hk}^*)
\end{aligned}$$

where

$$\tau_{hk} = r \sum_{i \in \mathcal{D}_{hk}} \log((1 - \alpha_{ik})/\alpha_{ik})$$

which can be simulated exactly by the inverse cdf method where the cdf is

$$(1 - \exp\{-\tau_{hk} \alpha_{hk}^*\}) \frac{1}{1 - \exp\{-\tau_{hk}\}} \mathbb{I}_{[0,1]}(\alpha_{hk}^*).$$

Similar argument is applied to the full conditional distributions of β_{hk}^* .

A.4. Full conditional distribution of Θ . The full conditional distribution of the elements of θ_{ik} $k = 1, \dots, K$ are discussed. Let $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iK})$, its full conditional distribution

$$(38) \quad f(\boldsymbol{\mu}_i | \dots) \propto \left(\prod_{t=1}^T \mathcal{N}(y_{it} | \mu_i(s_{it}), \sigma_{it}^2) \right) \prod_{k=1}^K \mathcal{N}(\mu_{ik} | \tilde{\mu}_{ik}^*, s)$$

which is not tractable due to the recursive form of σ_{it}^2 . Thus we sample from the full conditional by Metropolis-Hastings with proposal distribution obtained through the approximation σ_{it}^{*2} of σ_{it}^2 . It can easily be shown, by the completing of the square argument, that the joint full conditional distribution of $\boldsymbol{\mu}_i$ can be approximated by a normal distribution with mean and covariance

$$(39) \quad \mathbf{m}_i = S_i \begin{pmatrix} m_{i1}/s_{i1}^2 \\ m_{i2}/s_{i2}^2 \\ \vdots \\ m_{iK}/s_{iK}^2 \end{pmatrix}, \quad S_i = \begin{pmatrix} s_{i1}^2 & 0 & \dots & 0 \\ 0 & s_{i2}^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & s_{iK}^2 \end{pmatrix}$$

where

$$m_{ik} = s_{ik}^2 \left(\frac{\tilde{\mu}_{ik}^*}{s^2} + \sum_{t \in \mathcal{T}_{y,ik}} \frac{y_{it}}{\sigma_{it}^{*2}} \right), \quad \text{and} \quad s_{ik}^2 = \left(\frac{1}{s^2} + \sum_{t \in \mathcal{T}_{y,ik}} \frac{1}{\sigma_{it}^{*2}} \right)^{-1}$$

with $\mathcal{T}_{y,ik} = \{t = 1, \dots, T | s_{it} = k\}$ and

$$\sigma_{it}^{*2} = \gamma_i(s_{it}) + \alpha_i(s_{it})(y_{t-1} - \mu_i(s_{it-1}))^2 + (\beta_i(s_{it}))\sigma_{t-1}^{*2}.$$

The mean and variance thus constructed are used in defining the parameters of the normal mixture proposal distribution for $\boldsymbol{\mu}_i$.

$$f(\boldsymbol{\mu}_i | \dots) = 0.05\mathcal{N}(\boldsymbol{\mu}_i; \mathbf{m}_i, S_i) + 0.95\mathcal{N}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_i^{(r-1)}, S_i)$$

As regards the parameters of the volatility process the full conditional probability distribution is

$$\text{let } \boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{iK}), \boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{iK}), \boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{iK}),$$

(40)

$$f(\boldsymbol{\gamma}_i, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_i | \dots) \propto \prod_{t=1}^T \mathcal{N}(y_{it} | \mu_i(s_{it}), \sigma_{it}) \prod_{k=1}^K \mathcal{B}e(\gamma_{ik}/a | r\tilde{\gamma}_{ik}^*/a, r(1 - \tilde{\gamma}_{ik}^*/a)) \\ \mathcal{B}e(\alpha_{ik} | r\tilde{\alpha}_{ik}^*, r(1 - \tilde{\alpha}_{ik}^*)) \mathcal{B}e(\beta_{ik} | r\tilde{\beta}_{ik}^*, r(1 - \tilde{\beta}_{ik}^*))$$

We follow the ARMA approximation of the MS-GARCH process, that is

$$(41) \quad \sigma_{it}^2 = \gamma_i(s_{it}) + \alpha_i(s_{it})\epsilon_{it-1}^2 + \beta_i(s_{it})\sigma_{it-1}^2$$

$$(42) \quad \epsilon_{it}^2 = \gamma_i(s_{it}) + (\alpha_i(s_{it}) + \beta_i(s_{it}))\epsilon_{it-1}^2 - \beta_i(s_{it})(\epsilon_{it-1}^2 - \sigma_{it-1}^2) + (\epsilon_{it}^2 - \sigma_{it}^2).$$

Let

$$w_{it} = \epsilon_{it}^2 - \sigma_{it}^2 = \left(\frac{\epsilon_{it}^2}{\sigma_{it}^2} - 1 \right) \sigma_{it}^2 = (\chi^2(1) - 1)\sigma_{it}^2$$

with

$$E_{t-1}[w_{it}] = 0; \quad \text{and} \quad \text{Var}_{t-1}[w_{it}] = 2\sigma_{it}^4.$$

Subject to the above and following Nakatsuma (1998) suggestion, we assume that $w_{it} \approx w_{it}^* \sim \mathcal{N}(0, 2\sigma_{it}^4)$. Then we have the following auxiliary ARMA model for the squared error term ϵ_{it}^2

$$(43) \quad \epsilon_{it}^2 = \gamma_i s_{it} + (\alpha_i(s_{it}) + \beta_i(s_{it}))\epsilon_{it-1}^2 - \beta_i(s_{it})w_{it-1}^* + w_{it}^*$$

with $w_{it}^* \sim \mathcal{N}(0, 2\sigma_{it}^4)$, which returns

$$(44) \quad w_{it}^* = \epsilon_{it}^2 - \gamma_i(s_{it}) - \alpha_i(s_{it})\epsilon_{it-1}^2 - \beta_i(s_{it})(\epsilon_{it-1}^2 - w_{it-1}^*).$$

Following Ardia (2008) we further express w_{it}^* as a linear function of the $(3K \times 1)$ vector $\boldsymbol{\theta}_{i\sigma} = (\gamma_{i1}, \dots, \gamma_{iK}, \alpha_{i1}, \dots, \alpha_{iK}, \beta_{i1}, \dots, \beta_{iK})'$. To do this, we approximate the function w_{it}^* by the first order Taylor's expansion about $\boldsymbol{\theta}_{i\sigma}^{(r-1)} = (\gamma_{i1}^{(r-1)}, \dots, \gamma_{iK}^{(r-1)}, \alpha_{i1}^{(r-1)}, \dots, \alpha_{iK}^{(r-1)}, \beta_{i1}^{(r-1)}, \dots, \beta_{iK}^{(r-1)})'$.

$$(45) \quad w_{it}^* \approx w_{it}^{**} = w_{it}^*(\boldsymbol{\theta}_{i\sigma}^{(r-1)}) + \nabla'_{it}(\boldsymbol{\theta}_{i\sigma} - \boldsymbol{\theta}_{i\sigma}^{(r-1)}),$$

where

$$(46) \quad \nabla_{it} = \text{vec} \begin{pmatrix} \nabla'_{it1} \\ \nabla'_{it2} \\ \vdots \\ \nabla'_{itK} \end{pmatrix}, \quad \nabla_{itk} = \begin{pmatrix} \frac{\partial w_{it}^*}{\partial \gamma_{ik}} \\ \frac{\partial w_{it}^*}{\partial \alpha_{ik}} \\ \frac{\partial w_{it}^*}{\partial \beta_{ik}} \end{pmatrix}$$

with

$$(47) \quad \begin{pmatrix} \nabla'_{it1} \\ \nabla'_{it2} \\ \vdots \\ \nabla'_{itK} \end{pmatrix} = \boldsymbol{\xi}'_{it} E_t + (\boldsymbol{\xi}_{it} \beta'_i) \begin{pmatrix} \nabla'_{it-1,1} \\ \nabla'_{it-1,2} \\ \vdots \\ \nabla'_{it-1,K} \end{pmatrix}$$

$E_t = (-1, -\epsilon_{it-1}^2, -(\epsilon_{it-1}^2 - w_{it-1}^*))$, $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{iK})$, $\nabla_{i0k} = \mathbf{0}$ and $\boldsymbol{\xi}_{it}$ is a row vector.

Upon defining $r_{it}^* = w_{it}^*(\boldsymbol{\theta}_{i\sigma}^{(r-1)}) - \nabla'_{it} \boldsymbol{\theta}_{i\sigma}^{(r-1)}$, it turns out that $w_{it}^{**} = r_{it}^* + \nabla'_{it} \boldsymbol{\theta}_{i\sigma}$. Furthermore, by defining $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{iK})$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK})$, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{iK})$, the $T \times 1$ vectors $\mathbf{w}_i = (w_{i1}^*, \dots, w_{iT}^*)'$, $\mathbf{r}_i^* = (r_{i1}^*, \dots, r_{iT}^*)'$, a $T \times 3K$ matrix $\nabla_i = (\nabla_{i1}, \nabla_{i2}, \dots, \nabla_{iT})'$ as well as a $T \times T$ matrix

$$(48) \quad \Upsilon_i = 2 \begin{pmatrix} \sigma_{i1}^{**4} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{iT}^{**4} \end{pmatrix},$$

with $\sigma_{it}^{**2} = (\boldsymbol{\xi}_{it} \boldsymbol{\gamma}_i^{(r-1)'}) + (\boldsymbol{\xi}_{it} \boldsymbol{\alpha}_i^{(r-1)'}) (y_{t-1} - \boldsymbol{\xi}_{t-1} \boldsymbol{\mu}_i^{(r)'})^2 + (\boldsymbol{\xi}_{it} \boldsymbol{\beta}_i^{(r-1)'}) \sigma_{it-1}^{**2}$, we end up with $\mathbf{w}_i = \mathbf{r}_i^* + \nabla_i \boldsymbol{\theta}_{i\sigma}$. Using this linear approximation, we can approximate the full conditional distribution of the volatility parameters as

$$(49) \quad f(\boldsymbol{\theta}_{i\sigma} | \boldsymbol{\xi}_{i,1:T}^{(r-1)}, \boldsymbol{\mu}_i^{(r)}, y_{1:T}) \propto \frac{1}{|\Upsilon_i|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{w}_i' \Upsilon_i^{-1} \mathbf{w}_i}{2}\right) \mathbb{I}_{\Theta}(\boldsymbol{\theta}_{i\sigma}) \\ \propto \mathcal{N}_{3K}(m_{i\sigma}, S_{i\sigma}) \mathbb{I}_{\Theta}(\boldsymbol{\theta}_{i\sigma}),$$

where $\Theta = \{\gamma_{i1} > 0, \dots, \gamma_{iK} > 0, 0 < \alpha_{i1} < 1, \dots, 0 < \alpha_{iK} < 1, 0 < \beta_{i1} < 1, \dots, 0 < \beta_{iK} < 1\}$ and

$$(50) \quad \begin{aligned} S_{i\sigma} &= (\nabla'_i \Upsilon_i^{-1} \nabla_i)^{-1} \\ m_{i\sigma} &= -S_{i\sigma} \nabla'_i \Upsilon_i^{-1} \mathbf{r}_i^*. \end{aligned}$$

The mean and variance defined above are used to characterize proposal distribution for $\theta_{i\sigma}$, that is a mixture of truncated normal distributions. In our MCMC exercise, we sample from the normal mixture and check that each sample satisfies the constraints.

$$f(\theta_{i\sigma} | \dots) = 0.05 \mathcal{N}(\theta_{i\sigma}; \mathbf{m}_{i\sigma}, S_{i\sigma}) + 0.95 \mathcal{N}(\theta_{i\sigma}; \theta_{i\sigma}^{(r-1)}, S_{i\sigma})$$

A.5. Full conditional distribution of Ξ . The full joint conditional distribution of the state variables, $\boldsymbol{\xi}_{i,1:T} = (\boldsymbol{\xi}_{i1}, \dots, \boldsymbol{\xi}_{iT})$ with $\boldsymbol{\xi}_{it} = (\xi_{i1,t}, \dots, \xi_{iK,t})$, given the parameter values and return series

$$(51) \quad p(\boldsymbol{\xi}_{i,1:T} | \dots) \propto \prod_{t=1}^T f(y_{it} | \theta_i, s_{it}) \prod_{k=1}^K \prod_{l=1}^k p_{i,kl}^{\xi_{ikt-1} \xi_{ilt}}$$

is a non-standard distribution. For this reason, following Billio, Casarin and Osuntuyi (2016), we propose a Metropolis-Hastings algorithm with proposal distribution given by an approximation of the smoothed probability $p(\boldsymbol{\xi}_{i,1:T} | \dots)$. Precisely, the algorithm involves running a Forward Filtering Backward Sampling (FFBS) on a auxiliary model to generate proposals at each iteration step. Among the several alternative MS-GARCH models based on collapsing procedure (see Billio, Casarin and Osuntuyi (2016)), we adopt the Klaassen (2002) MS-GARCH model as our auxiliary model because it accounts for the highest amount of information in its construction. We denote the proposal distribution by

$$(52) \quad q(\boldsymbol{\xi}_{i,1:T} | \theta_i, y_{i,1:T}) = q(\boldsymbol{\xi}_{iT} | \theta_i, y_{i,1:T}) \prod_{t=1}^{T-1} q(\boldsymbol{\xi}_{it} | \boldsymbol{\xi}_{it+1}, \theta_i, y_{i,1:t}),$$

where

$$q(\boldsymbol{\xi}_{it} | \boldsymbol{\xi}_{it+1}, \theta_i, y_{i,1:t}) = \frac{q(\boldsymbol{\xi}_{it} | y_{i,1:t}, \theta_i) q(\boldsymbol{\xi}_{it+1} | \boldsymbol{\xi}_{it}, \theta_i)}{q(\boldsymbol{\xi}_{it+1} | y_{i,1:t}, \theta_i)}$$

with $q(\boldsymbol{\xi}_{it} | y_{i,1:t}, \theta_i)$ representing filtered probability.

At time t , given θ_i and $y_{i,1:t}$ the prediction and filtering densities are respectively given by

$$(53) \quad q(\boldsymbol{\xi}_{it}|\theta_i, y_{i,1:t-1}) = \sum_{k=1}^K \left(\prod_{l=1}^K p_{i,lk}^{\xi_{il,t}} \right) q(\boldsymbol{\xi}_{it-1} = e_k|\theta_i, y_{i,1:t-1}),$$

and

$$(54) \quad q(\boldsymbol{\xi}_{it}|\theta_i, y_{i,1:t}) = \frac{g(y_{it}|\boldsymbol{\xi}_{it}, \theta_i, y_{i,1:t-1})q(\boldsymbol{\xi}_{it}|\theta_i, y_{i,1:t-1})}{\sum_{k=1}^K g(y_{it}|\boldsymbol{\xi}_{it} = e_k, \theta_i, y_{i,1:t-1})q(\boldsymbol{\xi}_{it} = e_k|\theta_i, y_{i,1:t-1})},$$

where e_k is the k -th row of a K -by- K identity matrix and $g(y_{it}|\boldsymbol{\xi}_{it}, \theta_i, y_{i,1:t-1})$ is the conditional density of unit i return process under the auxiliary model

$$(55) \quad g(y_{it}|\boldsymbol{\xi}_{it}, \theta_i, y_{i,1:t-1}) \propto \prod_{\tau=1}^t \frac{1}{h_{i\tau}} \exp\left(-\frac{(y_{i\tau} - \mu_i(s_{i\tau}))^2}{2h_{i\tau}^2}\right)$$

where

$$h_{it}^2 = \gamma_i(s_{it}) + \alpha_i(s_{it})\epsilon_{(y)it-1}^2 + \beta_i(s_{it})\sigma_{(y)i,kt-1}^2$$

with

$$(56) \quad \begin{aligned} \epsilon_{(y)it-1} &= y_{it-1} - \mu_{(y)i,kt-1} \\ \mu_{(y)ik,t-1} &= E[\mu_i(s_{it})|y_{i,1:t-1}, \boldsymbol{\xi}_{it} = e_k] \\ \sigma_{(y)i,kt-1}^2 &= E[\sigma_{it-1}^2(y_{i,1:t-2}, \boldsymbol{\xi}_{it-1}, \boldsymbol{\xi}_{it-2})|y_{i,1:t-1}, \boldsymbol{\xi}_{it} = e_k]. \end{aligned}$$

Using the output of the FF, we compute $q(\boldsymbol{\xi}_{iT}|\theta_i, y_{i,1:T})$ and

$$(57) \quad q(\boldsymbol{\xi}_{it}|\boldsymbol{\xi}_{it+1}, \theta_i, y_{i,1:t}) = \frac{\prod_{l=1}^K \left(\sum_{k=1}^K p_{i,lk} \xi_{i1,t} \right)^{\xi_{il,t+1}} q(\boldsymbol{\xi}_{it}|\theta_i, y_{i,1:t})}{q(\boldsymbol{\xi}_{it+1}|\theta_i, y_{i,1:t})},$$

for $t = T-1, T-2, \dots, 2, 1$. Then at each time step we sample $\boldsymbol{\xi}_T$ from $q(\boldsymbol{\xi}_T|\theta_i, y_{i,1:T})$ and $\boldsymbol{\xi}_{it}$ from $q(\boldsymbol{\xi}_{it}|\boldsymbol{\xi}_{it+1}, \theta_i, y_{i,1:t})$ iteratively for $t = T-1, T-2, \dots, 2, 1$. This is the BS step. The BS procedure is implemented by first noting that $\boldsymbol{\xi}_{it+1}$ is the most recent value sampled for the hidden Markov chain at $t+1$ and since $\boldsymbol{\xi}_{it}$ can take one of e_1, \dots, e_K , we compute the expression in equation (57) for each of these values. Sampling $\boldsymbol{\xi}_{it}$ from $q(\boldsymbol{\xi}_{it}|\boldsymbol{\xi}_{it+1}, \theta_i, y_{i,1:t})$ may be compared to multinomial sampling, provided that the probability of $\boldsymbol{\xi}_{ik} = e_k$, $k = 1, \dots, K$, are known.

A.6. Full conditional distribution of D . The full conditional of D_{ik} is $P(D_{ik} = h|\dots) = c_h/c$ for $h \in \mathcal{A}_{ki}$, with $c_h = \mathcal{N}(\mu_{ik}|\mu_{hk}^*, s) \mathcal{Be}(\alpha_{ik}|r\alpha_{hk}^*, r(1 - \alpha_{hk}^*)) \mathcal{Be}(\beta_{ik}|r\beta_{hk}^*, r(1 - \beta_{hk}^*)) \mathcal{Be}(\gamma_{ik}/a|r\gamma_{hk}^*/a, r(1 - \gamma_{hk}^*/a))/a$ where $c = \sum_{h \in \mathcal{A}_{ki}} c_h$ is the normalizing constant and a a real positive constant.

APPENDIX B: FURTHER DETAILS ON THE SIMULATION
EXERCISE

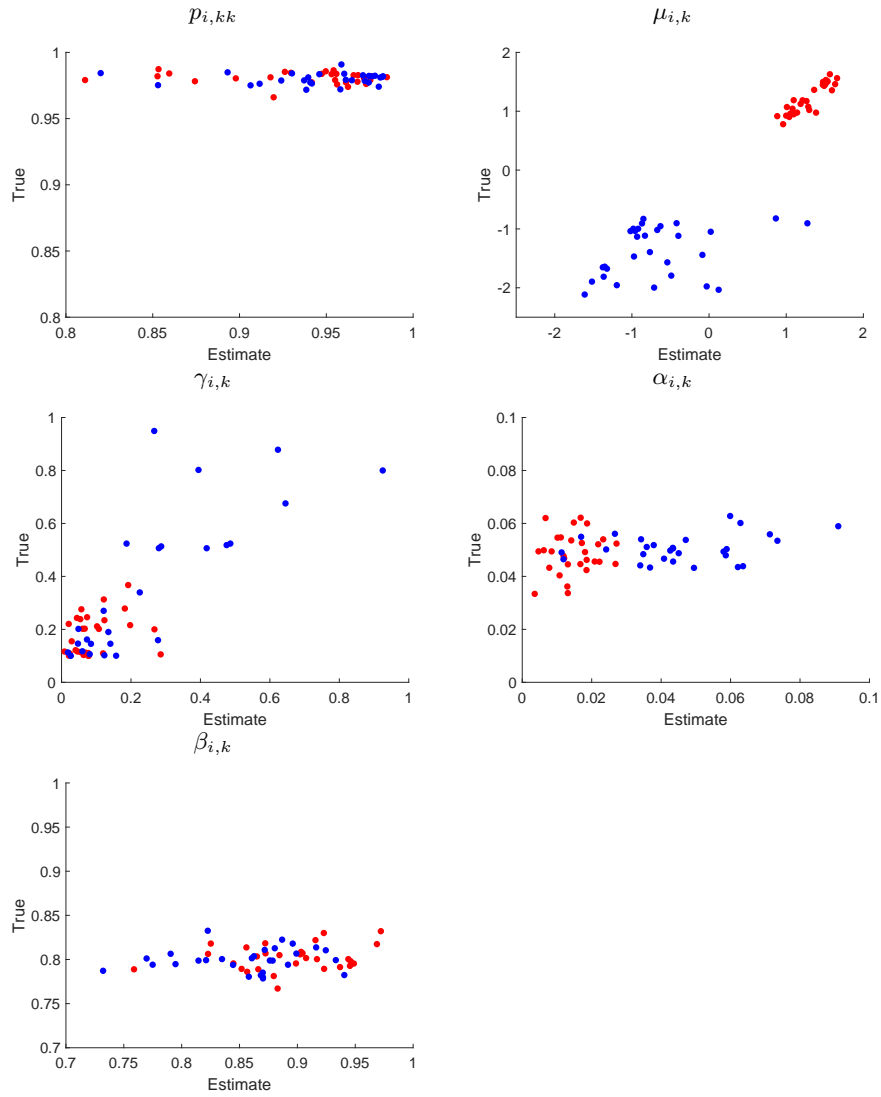


FIGURE 8. True (vertical axis) and estimated (horizontal axis) values of the parameters θ_{ik} for each unit i (dots) in regime $k = 1$ (red) and $k = 2$ (blue).

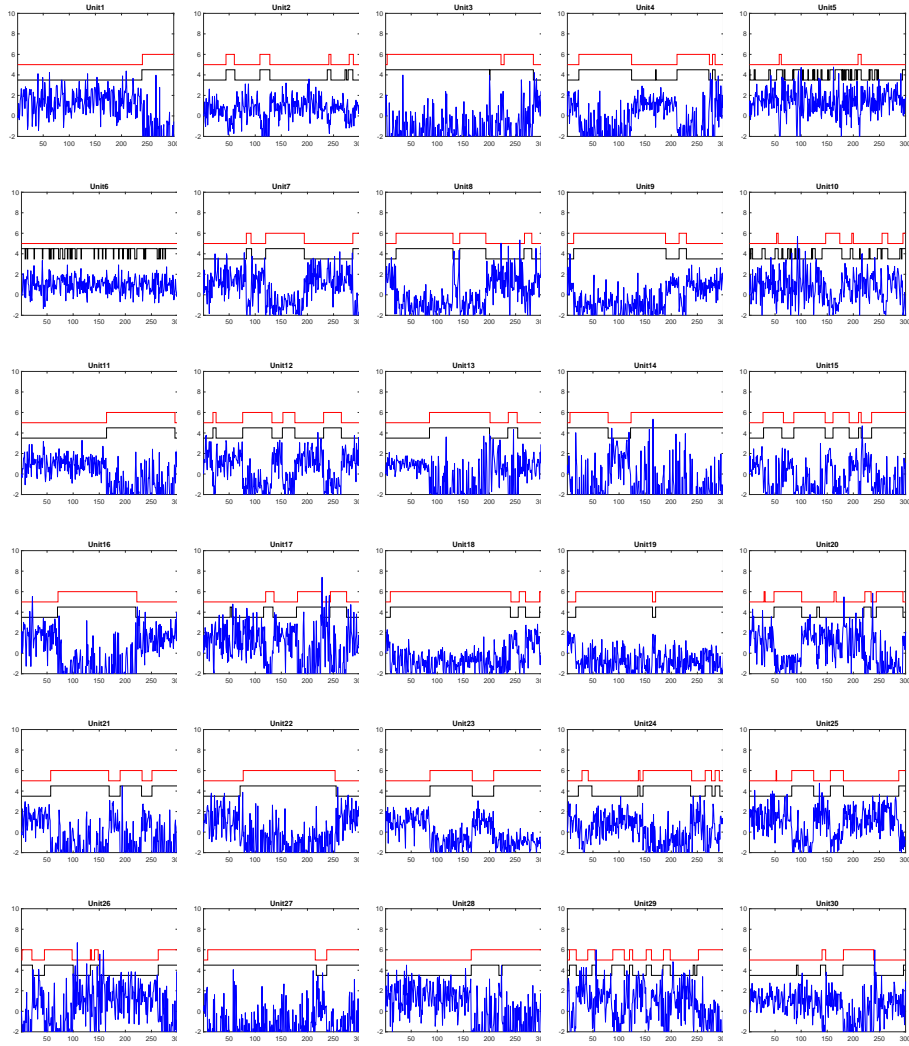


FIGURE 9. In each plot, the true (red) and estimated (black) trajectories of the hidden Markov chain process and the observed process y_{it} (blue line).

APPENDIX C: FURTHER DETAILS ON THE EMPIRICAL APPLICATION

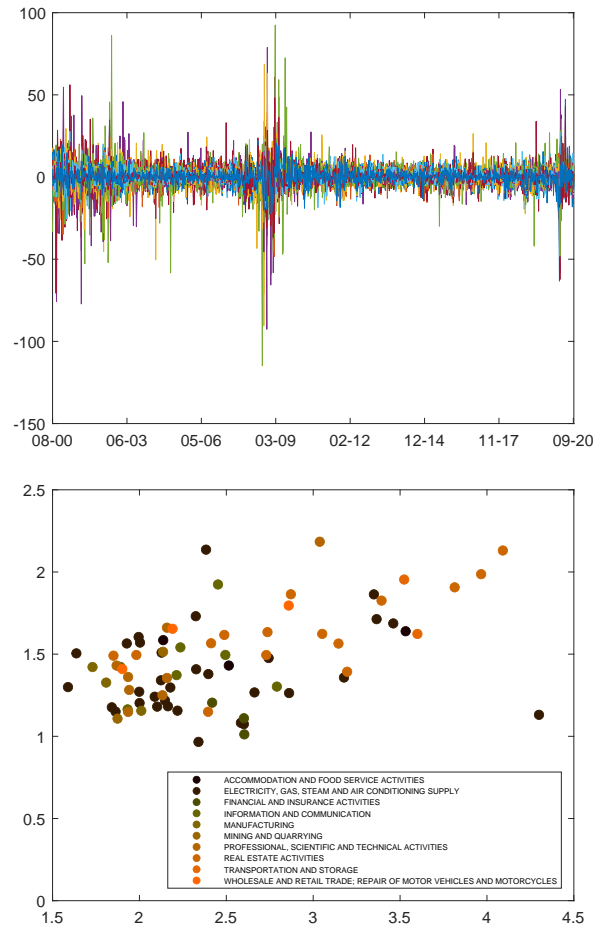


FIGURE 10. Top: percentage log-returns of the SP500's constituents for the period 3rd August 2000 to 3rd October 2020. Bottom: scatter plot of the log-variance and log-kurtosis of the financial returns. Colours indicate a different sector.

TABLE 1

Constituents of the S&P100 at the 1st October 2020. In the columns, the company label (Symbol), name (Name) and sector (S). Sector codes are: Mining and Quarrying (B), Financial and Insurance Activities (K), Information and Communication (J), Manufacturing (C), Real Estates Activities (L), Wholesale and Retail Trade; Repair of Motors (G), Accommodation and Food Service (I), Electricity Gas Steam and Air Cond. (D), Transp. and Storage (H), Professional Scientific and Technical Activities (M). The columns C indicates if a company is included in (1) in the analysis.

Symbol	Name	S	C	Symbol	Name	S	C
OXY	Occidental Petroleum Corp	B	1	COST	Costco Wholesale Corp	G	1
COP	ConocoPhillips	B	1	TGT	Target Corp	G	1
SLB	Schlumberger NV	B	1	LOW	Lowe's Cos Inc	G	1
MDLZ	Mondelez Int Inc	C	0	CVS	CVS Health Corp	G	1
BA	Boeing Co/The	C	1	UNP	Union Pacific Corp	H	1
CAT	Caterpillar Inc	C	1	KMI	Kinder Morgan Inc	H	0
CVX	Chevron Corp	C	1	FDX	FedEx Corp	H	1
KO	Coca-Cola Co/The	C	1	UPS	United Parcel Service Inc	H	1
XOM	Exxon Mobil Corp	C	1	MCD	McDonald's Corp	I	1
GE	General Electric Co	C	1	SBUX	Starbucks Corp	I	1
JNJ	Johnson & Johnson	C	1	VZ	Verizon Communications Inc	J	1
MRK	Merck & Co Inc	C	1	DIS	Walt Disney Co/The	J	1
MMM	3M Co	C	1	IBM	Int Business Machines Corp	J	1
PFE	Pfizer Inc	C	1	ACN	Accenture PLC	J	0
PG	Procter & Gamble Co/The	C	1	GOOG	Alphabet Inc	J	0
RTX	Raytheon Technologies Corp	C	1	T	AT&T Inc	J	1
CSCO	Cisco Systems Inc	C	1	CHTR	Charter Communications Inc	J	0
INTC	Intel Corp	C	1	MSFT	Microsoft Corp	J	1
NVDA	NVIDIA Corp	C	1	BKNG	Booking Holdings Inc	J	1
HON	Honeywell Int Inc	C	1	GOOGL	Alphabet Inc	J	0
MO	Altria Group Inc	C	1	NFLX	Netflix Inc	J	0
ABT	Abbott Laboratories	C	1	CRM	salesforce.com Inc	J	0
TXN	Texas Instruments Inc	C	0	ADBE	Adobe Inc	J	1
KHC	Kraft Heinz Co/The	C	0	CMCSA	Comcast Corp	J	1
TMO	Thermo Fisher Scientific Inc	C	1	ORCL	Oracle Corp	J	0
PM U	Philip Morris International Inc	C	0	FB	Facebook Inc	J	0
BMJ	Bristol Myers Squibb Co	C	1	AXP	American Express Co	K	1
AAPL	Apple Inc	C	1	JPM	JPMorgan Chase & Co	K	1
CL	Colgate-Palmolive Co	C	1	BAC	Bank of America Corp	K	1
ABBV	AbbVie Inc	C	0	C	Citigroup Inc	K	1
DHR	Danaher Corp	C	1	AIG	American International Group Inc	K	1
DOW	Dow Inc	C	0	GS	Goldman Sachs Group Inc/The	K	1
GM	General Motors Co	C	0	UNH	UnitedHealth Group Inc	K	1
EMR	Emerson Electric Co	C	1	BLK	BlackRock Inc	K	1
F	Ford Motor Co	C	1	BK	Bank of NY Mellon Corp/The	K	1
GD	General Dynamics Corp	C	1	MET	MetLife Inc	K	0
QCOM	QUALCOMM Inc	C	1	BRK/B	Berkshire Hathaway Inc	K	1
PEP	PepsiCo Inc	C	0	MA	Mastercard Inc	K	0
LLY	Eli Lilly and Co	C	1	V	Visa Inc	K	0
MDT	Medtronic PLC	C	1	PYPL	PayPal Holdings Inc	K	0
LMT	Lockheed Martin Corp	C	1	USB	US Bancorp	K	1
NKE	NIKE Inc	C	1	MS	Morgan Stanley	K	1
DD	DuPont de Nemours Inc	C	1	ALL	Allstate Corp/The	K	1
SO	Southern Co/The	D	1	COF	Capital One Financial Corp	K	1
DUK	Duke Energy Corp	D	1	WFC	Wells Fargo & Co	K	1
EXC	Exelon Corp	D	0	AMT	American Tower Corp	L	1
NEE	NextEra Energy Inc	D	1	SPG	Simon Property Group Inc	L	1
AMZN	Amazon.com Inc	G	1	AMGN	Amgen Inc	M	1
HD	Home Depot Inc/The	G	1	GILD U	Gilead Sciences Inc	M	1
WMT	Walmart Inc	G	1	BIIB	Biogen Inc	M	1
WBA	Walgreens Boots Alliance Inc	G	0				

date: December 21, 2020

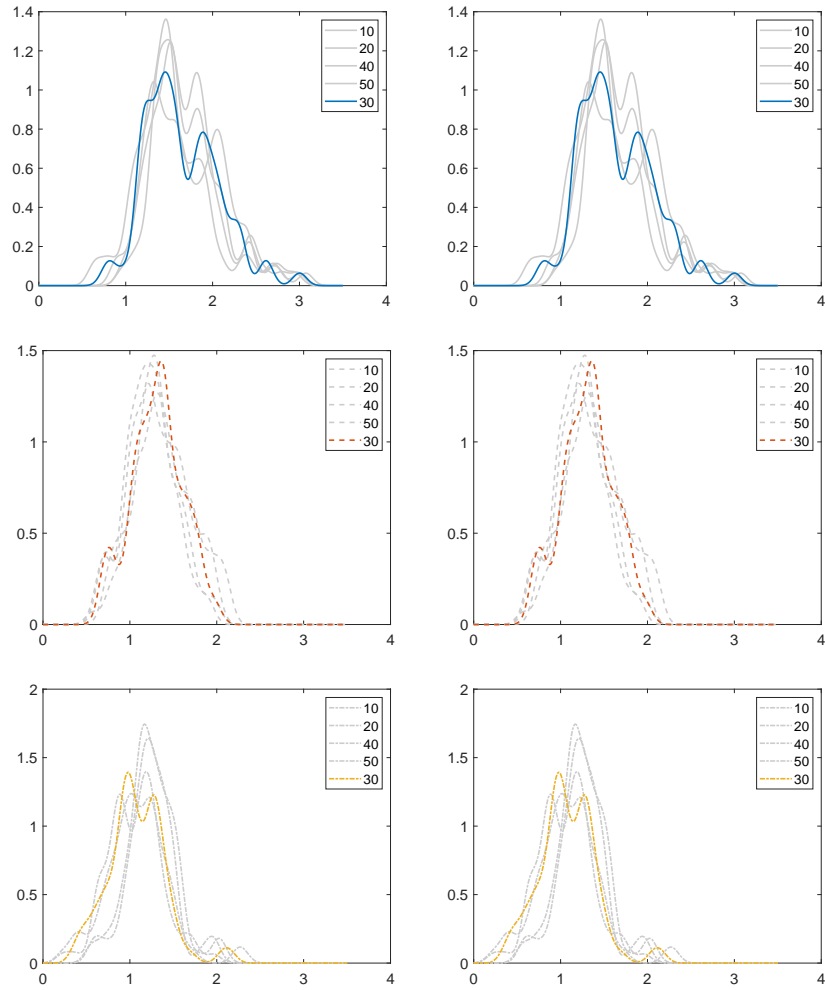


FIGURE 11. Cross-sectional distribution of the log-volatility (left) and log-kurtosis (right) of the SP500's constituents log-returns in the three dates: 6th July 2002, 23rd August 2008 and 22nd February 2020 (different rows). In each plot, the cross-section of statics is derived with different sizes of the rolling window (different lines).

TABLE 2
Cluster composition in Regime 1

Cluster 1					
Symbol	Name	S	Symbol	Name	S
GD	General Dynamics Corp	C	BK	Bank of New York Mellon Corp/The	K
MDT	Medtronic PLC	C	HD	American Express Co	K
BA	Boeing Co/The	C	C	Citigroup Inc	K
KO	Coca-Cola Co/The	C	MS	Morgan Stanley	K
F	Ford Motor Co	C	ALL	Allstate Corp/The	K
XOM	Exxon Mobil Corp	C	WFC	Wells Fargo & Co	K
LLY	Eli Lilly and Co	C	JPM	JPMorgan Chase & Co	K
GE	General Electric Co	C	AIG	American Int. Group Inc	K
CL	Colgate-Palmolive Co	C	GS	Goldman Sachs Group Inc/The	K
LMT	Lockheed Martin Corp	C	LOW	Lowe's Cos Inc	G
MRK	Merck & Co Inc	C	TGT	Target Corp	G
PFE	Pfizer Inc	C	AXP	Home Depot Inc/The	G
PG	Procter & Gamble Co/The	C	CVS	CVS Health Corp	G
RTX	Raytheon Technologies Corp	C	COST	Costco Wholesale Corp	G
HON	Honeywell Int. Inc	C	WMT	Walmart Inc	G
CSCO	Cisco Systems Inc	C	DUK	Duke Energy Corp	D
CAT	Caterpillar Inc	C	SO	Southern Co/The	D
TMO	Thermo Fisher Scientific Inc	C	OXY	Occidental Petroleum Corp	B
CVX	Chevron Corp	C	COP	ConocoPhillips	B
VZ	Verizon Communications Inc	J	AMT	American Tower Corp	L
MSFT	Microsoft Corp	J	SPG	Simon Property Group Inc	L
DIS	Walt Disney Co/The	J	GILD	Gilead Sciences Inc	M
BKNG	Booking Holdings Inc	J	BIIB	Biogen Inc	M
IBM	Int. Business Machines Corp	J	FDX	FedEx Corp	H
Cluster 2					
BMJ	Bristol Myers Squibb Co	C	UNH	UnitedHealth Group Inc	K
JNJ	Johnson & Johnson	C	BLK	BlackRock Inc	K
AAPL	Apple Inc	C	COF	Capital One Financial Corp	K
MMM	3M Co	C	BRK/B	Berkshire Hathaway Inc	K
EMR	Emerson Electric Co	C	BAC	Bank of America Corp	K
DHR	Danaher Corp	C	USB	US Bancorp	K
INTC	Intel Corp	C	CMCSA	Comcast Corp	J
QCOM	QUALCOMM Inc	C	ADBE	Adobe Inc	J
NVDA	NVIDIA Corp	C	T	AT&T Inc	J
MO	Altria Group Inc	C	SBUX	Starbucks Corp	I
ABT	Abbott Laboratories	C	MCD	McDonald's Corp	I
NKE	NIKE Inc	C	NEE	NextEra Energy Inc	D
DD	DuPont de Nemours Inc	C	AMGN	Amgen Inc	M
UPS	United Parcel Service Inc	H	SLB	Schlumberger NV	B
UNP	Union Pacific Corp	H	AMZN	Amazon.com Inc	G

date: December 21, 2020

TABLE 3
Cluster composition in Regime 2.

Cluster 1					
Symbol	Name	S	Symbol	Name	S
XOM	Exxon Mobil Corp	C	GS	Goldman Sachs Group Inc/The	K
MMM	3M Co	C	BK	Bank of New York Mellon Corp/The	K
BA	Boeing Co/The	C	AXP	American Express Co	K
CVX	Chevron Corp	C	BRK/B	Berkshire Hathaway Inc	K
KO	Coca-Cola Co/The	C	OXY	American Int. Group Inc	K
CL	Colgate-Palmolive Co	C	WFC	Wells Fargo & Co	K
F	Ford Motor Co	C	USB	US Bancorp	K
BMY	Bristol Myers Squibb Co	C	MS	Morgan Stanley	K
TMO	Thermo Fisher Scientific Inc	C	ALL	Allstate Corp/The	K
GD	General Dynamics Corp	C	HD	Home Depot Inc/The	G
LLY	Eli Lilly and Co	C	COST	Costco Wholesale Corp	G
MDT	Medtronic PLC	C	CVS	CVS Health Corp	G
PFE	Pfizer Inc	C	COP	ConocoPhillips	B
PG	Procter & Gamble Co/The	C	AIG	Occidental Petroleum Corp	B
RTX	Raytheon Technologies Corp	C	SPG	Simon Property Group Inc	L
MSFT	Microsoft Corp	J	AMT	American Tower Corp	L
IBM	Int. Business Machines Corp	J	DUK	Duke Energy Corp	D
DIS	Walt Disney Co/The	J	SO	Southern Co/The	D
Cluster 2					
DHR	Danaher Corp	C	JPM	JPMorgan Chase & Co	K
CAT	Caterpillar Inc	C	BAC	Bank of America Corp	K
AAPL	Apple Inc	C	T	AT&T Inc	J
DD	DuPont de Nemours Inc	C	ADBE	Adobe Inc	J
JNJ	Johnson & Johnson	C	CMCSA	Comcast Corp	J
QCOM	QUALCOMM Inc	C	VZ	Verizon Communications Inc	J
EMR	Emerson Electric Co	C	UNP	Union Pacific Corp	H
NKE	NIKE Inc	C	UPS	United Parcel Service Inc	H
INTC	Intel Corp	C	MCD	McDonald's Corp	I
NVDA	NVIDIA Corp	C	SBUX	Starbucks Corp	I
ABT	Abbott Laboratories	C	NEE	NextEra Energy Inc	D
COF	Capital One Financial Corp	K	SLB	Schlumberger NV	B
BLK	BlackRock Inc	K	AMGN	Amgen Inc	M
UNH	UnitedHealth Group Inc	K	AMZN	Amazon.com Inc	G
Cluster 3					
GE	General Electric Co	C	LOW	Lowe's Cos Inc	G
MRK	Merck & Co Inc	C	WMT	Walmart Inc	G
LMT	Lockheed Martin Corp	C	TGT	Target Corp	G
CSCO	Cisco Systems Inc	C	GILD	Gilead Sciences Inc	M
MO	Altria Group Inc	C	BIIB	Biogen Inc	M
HON	Honeywell International Inc	C	BKNG	Booking Holdings Inc	J
C UN	Citigroup Inc	K	FDX	FedEx Corp	H

TABLE 4

Cluster composition by market capitalization, small (bottom 30%), medium (middle 40%) and big (top 30%) companies, in the two regimes (panel (a) and (b)).

(a) Regime 1

Cluster 1			Cluster 2		
Small	Medium	Big	Small	Medium	Big
AIG	AMT	CSCO	CAT	ABT	AAPL
ALL	AXP	HD	COF	ADBE	AMZN
BIIB	BA	KO	DD	AMGN	BAC
BK	C	MRK	EMR	BLK	BRK/B
BKNG	COST	MSFT	SLB	BMY	CMCSA
CL	CVS	PFE	USB	DHR	CVX
COP	GE	PG		MCD	DIS
DUK	GILD	WMT		MMM	INTC
F	HON	XOM		MO	JNJ
FDX	IBM			NEE	JPM
GD	LLY			NKE	NVDA
GS	LMT			QCOM	T
MS	LOW			SBUX	UNH
OXY	MDT			UNP	VZ
SO	RTX			UPS	
SPG	TMO				
TGT	WFC				

(b) Regime 2

Cluster 1			Cluster 2			Cluster 3		
Small	Medium	Big	Small	Medium	Big	Small	Medium	Big
AIG	AMT	BRK/B	CAT	ABT	AAPL	BIIB	C	CSCO
ALL	AXP	CVX	COF	ADBE	AMZN	BKNG	GE	MRK
BK	BA	DIS	DD	AMGN	BAC	FDX	GILD	WMT
CL	BMY	HD	EMR	BLK	CMCSA	TGT	HON	
COP	COST	KO	SLB	DHR	INTC		LMT	
DUK	CVS	MSFT		MCD	JNJ		LOW	
F	IBM	PFE		NEE	JPM		MO	
GD	LLY	PG		NKE	NVDA			
GS	MDT	XOM		QCOM	T			
MS	MMM			SBUX	UNH			
OXY	RTX			UNP	VZ			
SO	TMO			UPS				
SPG	WFC							
USB								

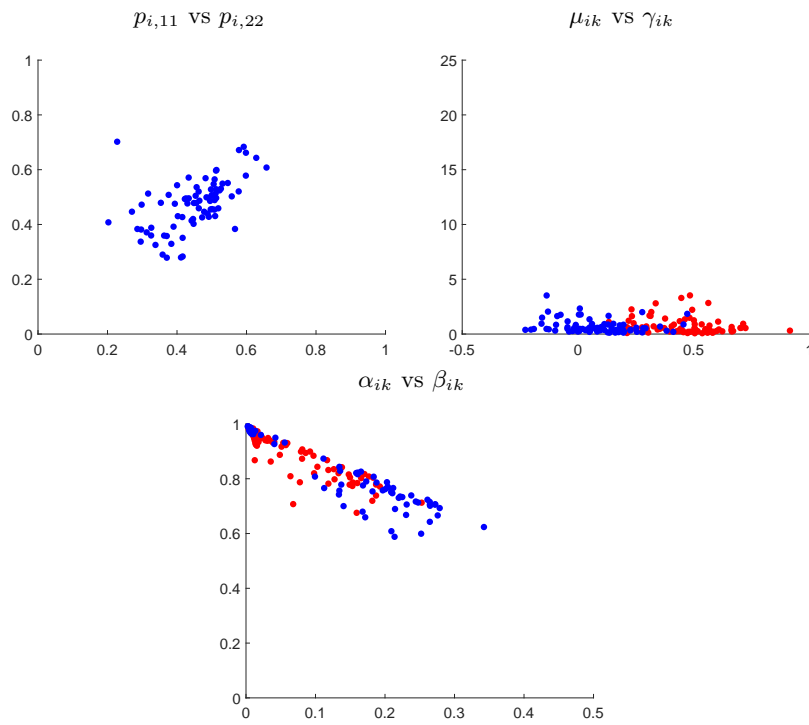


FIGURE 12. *Parameter estimates. Colors indicate regime-specific parameter values with under-performance regime in blue (regime $k = 1$) and over-performance regime in red (regime $k = 2$).*