Stochastics and Statistics

Clustering financial time series: New insights from an extended hidden Markov model

José G. Dias [a,*], Jeroen K. Vermunt [b], Sofia Ramos [c]

[a] *Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (UNIDE-IUL), Portugal*
[b] *Department of Methodology and Statistics, Tilburg University, The Netherlands*
[c] *NEOMA Business School, France & Business Research Unit (UNIDE-IUL), Portugal*

ARTICLE INFO

ABSTRACT

In recent years, large amounts of financial data have become available for analysis. We propose exploring returns from 21 European stock markets by model-based clustering of regime switching models. These econometric models identify clusters of time series with similar dynamic patterns and moreover allow relaxing assumptions of existing approaches, such as the assumption of conditional Gaussian returns. The proposed model handles simultaneously the heterogeneity across stock markets and over time, i.e., time-constant and time-varying discrete latent variables capture unobserved heterogeneity between and within stock markets, respectively. The results show a clear distinction between two groups of stock markets, each one characterized by different regime switching dynamics that correspond to different expected return-risk patterns. We identify three regimes: the so-called bull and bear regimes, as well as a stable regime with returns close to 0, which turns out to be the most frequently occurring regime. This is consistent with stylized facts in financial econometrics.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In recent years dealing with unobserved heterogeneity has become a predominant topic in many research areas. As Heckman emphasized in his Nobel lecture, one of the most important discoveries in microeconometrics is the pervasiveness of heterogeneity and diversity in economic life: "When a full analysis of heterogeneity in response was made, a variety of candidate averages emerged to describe the average person, and the long-standing edifice of the representative consumer was shown to lack empirical support. (Heckman, 2001, p. 674)". In finance research, heterogeneity has been mostly assumed observed (e.g., based on countries), where groups and their boundaries are delineated without regarding the intrinsic information on the observed data. However, there are plenty of examples in the academic and professional finance literature that show that heterogeneity exists among capital market participants, business managers, fund managers, among others.

The correct modeling of the dynamics of stock market returns has been an important challenge in modern financial research. Though the dominant approach followed by both academics and practitioners has been to assume that returns follow a normal distribution (see, e.g., Lundblad, 2007 and Fu, 2009), it has also been recognized that stock market returns and returns of financial assets contain skewness and excessive kurtosis. A common conclusion is that the normal distribution is inadequate for short period returns of financial assets (Fama, 1965; Mandelbrot, 1963; Praetz, 1972). Several alternative distributions have therefore been suggested for modeling returns, one of which is the Laplace distribution. These alternatives have in common that they try to accommodate for the excessive kurtosis in the empirical distribution of the returns. Whereas excess kurtosis of financial return distributions has been well addressed in the financial literature, the asymmetry of the distribution has not received much attention, and the few studies available tend to be inconclusive (Peiró, 1999; Simkowitz & Beedles, 1980; Singleton & Wingender, 1986).

Latent class or finite mixture modeling has proven to be a powerful tool for analyzing a wide range of social and behavioral science data (see, for example, Clogg, 1995 and Vermunt, 2003). The identification of distinct dynamics in time series data has been an important topic of research from a substantive point of view. We propose a latent class model for financial data analysis that takes into account unobserved heterogeneity by means of time-constant and time-varying discrete latent variables. A feature of latent class modeling is that it yields a model-based clustering of observational units that is especially attractive to the typical analysis in finance research, where it

* Corresponding author at: Av. das Forças Armadas, 1649-026 Lisboa, Portugal. Tel.: +351 217 903 208, 351 21 479 33 88; fax: +351 217 903 004.
E-mail addresses: jose.dias@iscte.pt (J. G. Dias), J.K.Vermunt@uvt.nl (J. K. Vermunt), sofia.ramos@neoma-bs.fr (S. Ramos).

is of interest to identify subpopulations of firms, investors, markets or countries that differ in their propensities to specific characteristics (regulation, company governance characteristics, etc.). The challenging task of clustering this type of observations is that one has to incorporate data dependency in the clustering process (Kakizawa, Shumway, & Taniguchi, 1998). Therefore, clustering of time series has attracted large attention in statistics and data mining literature. However, most of the proposals developed have been based on classic clustering algorithms in such a way that time series data can be handled. Interested readers may refer to Liao (2005) and Esling and Agon (2012) who provide a detailed discussion of time series clustering in the data mining literature. Contrary to alternative heuristic clustering techniques for financial time series analysis which operate directly on the correlations (e.g., Basalto et al., 2007; Mantegna, 1999) or other measures of similarity between time series being clustered (e.g., Bastos & Caiado, 2013), the approach proposed here is a model-based clustering technique that accommodates for serial dependencies and unobserved heterogeneity by assuming a regime-switching model (RSM), also known as hidden Markov model, underlying each cluster. There is a vast body of literature on RSMs in economics and empirical finance, including Hamilton (1989), Hamilton and Susmel (1994), and Gray (1996) to name just a few. Hidden Markov models and regime-switching models as discrete state models can also be connected to stochastic volatility models (see, e.g., Langrock, MacDonald, & Zucchini, 2012; Rossi & Gallo, 2006). The autoregressive conditional root (ACR) model (Bec, Rahbek, & Shephard, 2008) is another econometric model that connects to regime-switching models. Many extensions of the regime-switching models have been suggested, adding new possibilities and modeling additional stylized facts of financial time series. For instance, Lux and Morales-Arias (2010) proposes a model that takes long memory and heavy tails of return time series into account, whereas Guidolin and Timmermann (2007), Fu, Wei, and Yang (2014), and Bae, Kim, and Mulvey (2014) apply RSMs to asset allocation and portfolio optimization. RSMs have been applied to pension funds optimization (Hainaut, 2014) and weather derivatives (Elias, Wahab, & Fang, 2014). For recent surveys on the application of RSMs in empirical finance, we refer to Lange and Rahbek (2009) and Guidolin (2011). These models have broader fields of application, covering manpower systems, where both observable and latent sources of dynamic heterogeneity should be accounted for (Guerry, 2011), and reliability analysis (Zhou, Hu, Xu, Chen, & Zhou, 2010).

Here, RSMs are extended to take the clustering structure of the returns of 21 European stock market indexes into account. Stock markets are well-known for presenting cycles, however country idiosyncrasies are also likely to make them differ in their transition between boom and bust. As is illustrated below, the proposed approach is flexible in the sense that it can deal with the specific features of financial time series data, such as asymmetry, kurtosis, and unobserved heterogeneity, an aspect that tends to be neglected. Because we selected a rather large and heterogeneous sample of countries including both developed and emerging countries and both EMU (European Monetary Union) and non-EMU countries, we expect that heterogeneity in market returns due to country idiosyncrasies will show up in the results. For instance, emerging market return distributions tend to show larger deviations from normality, i.e., they are more skewed and have fatter tails (Harvey, 1995). In addition, stock markets are also known to contain asymmetry of volatility, i.e., volatility is higher in negative shocks than positive ones (see, e.g., Ang & Bekaert, 2002) and structural breaks in time, or in other words, regime switching.

The results show that stock markets are better described by three regimes: A high returns, a negative returns, and close to zero regimes that we interpret as a bull, a bear, and a stable regime. This, however, challenges the simplistic view that stock markets should be characterized by two regimes. The characterization of regimes is consistent with several stylized facts such as asymmetry in volatility, i.e., bear regimes are associated with larger volatility than bull regimes. Stock markets are clustered into two groups that are mainly distinguished by the propensity to switch to the bear regime, which includes countries that were more affected by crisis during the period of analysis as well as Eastern European emerging countries that were less integrated with the other European countries.

The paper is structured as follows. Section 2 describes the 21 country financial time series data set that is used throughout this paper. Section 3 presents the statistical framework for the analysis of heterogeneous financial time series. It also discusses parameter estimation by maximum likelihood and model selection issues. Section 4 reports the results obtained for the data set at hand. The paper concludes with a summary of the main findings and a description of possible implications.

2. Description of the data set

The data used in this article are daily closing prices from 27 January 1998 to 31 July 2013 for 21 European stock market indexes drawn from Datastream database.[1] The series are denominated in US dollars. In total, we have 4010 end-of-the-day observations per country. Let P_{it} be the observed daily closing price of market i on day t, $i = 1, \ldots, n$ and $t = 0, \ldots, T$. The daily rates of return are defined as the percentage log-return by $y_{it} = 100 \times \log(P_{it}/P_{i,t-1})$, $t = 1, \ldots, T$. This definition which is commonly used in the literature is justified by the fact that for expected small increases (decreases) of value, say r, $\log(1 + r) \simeq r$.

The 21 stock markets are listed in Table 1. Figs. 1 and 2 depict for six distinct countries the index and returns time series, respectively. As is well known, stock markets follow cycles. In the sample period there were two main periods of global stock market crises. The dot-com bubble bursting that started at the end of 1999 and went on until 2003, and the subprime crisis that had its first signs in the summer of 2007, and made stock markets plummet in September 2008 after the Lehman Brothers bankruptcy. From 2004 to 2007, stock markets registered a strong growth.

Fig. 2 depicts stock market returns. Russia and Turkey have the highest level of volatility, which is typical of emerging markets. It is worth to note that in August 1998 Russia defaulted a sovereign bond payment triggering the "ruble crisis" in financial markets. Market features like these seem well-suitable to test our econometric model.

Table 1 provides the relevant descriptive statistics for the 21 stock-return series. All markets show non-negative median returns. However, only 18 out of 21 had positive mean returns; that is, Greece, Italy, and Portugal showed negative mean returns. Emerging markets such as Czech Republic and Russia show larger positive mean return. These figures confirm that stock market distributions tend to be negatively skewed.

The 21 analyzed markets show very diverse patterns of dispersion, where the largest standard deviations are found for Russia (3.134) and Turkey (2.983) – both emerging markets – which are almost three times larger than for Switzerland, the stock market with the smallest dispersion with a standard deviation of 1.210. Moreover, the excess kurtosis shows values above 0, indicating heavier tails and more peakness than the normal distribution (which has a kurtosis of 0). The Jarque–Bera test rejects the null hypothesis of normality for each of the 21 stock markets' returns.

Figs. 3 and 4 depict rolling means and standard deviations (30-day window) for these markets. Although moving averages tend to smooth trends, the main booms and peaks in the stock markets' returns are still visible. The ruble crisis is visible in Russia and it propagates to neighbor markets such as Poland and Hungary (not shown). All stock markets show a volatility peak during the subprime crisis.

---

[1] Observations from different time zones can create problems of non-synchronization on the analysis, to eliminate such problems we focus on European markets.

Table 1
Summary statistics.

| Stock market	Mean	Median	Standard deviation	Skewness	Kurtosis	Jarque–Bera test	
						statistics	p-value
Austria (OE)	0.016	0.056	1.454	−0.348	9.694	7532.11	0.000
Belgium (BG)	0.015	0.079	1.387	−0.184	7.942	4083.96	0.000
Czech Rep. (CZ)	0.038	0.000	1.700	−0.120	15.293	25151.89	0.000
Denmark (DK)	0.028	0.077	1.424	−0.350	9.961	8139.56	0.000
Finland (FN)	0.019	0.050	2.118	−0.298	8.415	4934.37	0.000
France (FR)	0.017	0.066	1.533	−0.072	8.013	4182.10	0.000
Germany (BD)	0.013	0.083	1.488	0.087	9.917	7961.37	0.000
Greece (GR)	−0.015	0.045	1.963	0.001	6.473	2003.94	0.000
Hungary (HN)	0.006	0.000	2.109	−0.164	10.583	9581.98	0.000
Ireland (IR)	0.005	0.059	1.561	−0.554	9.321	6850.14	0.000
Italy (IT)	−0.001	0.051	1.610	−0.129	7.911	4020.05	0.000
Netherlands (NL)	0.004	0.067	1.515	−0.237	9.056	6136.97	0.000
Norway (NW)	0.025	0.095	1.871	−0.463	9.284	6710.94	0.000
Poland (PO)	0.022	0.056	1.940	−0.185	6.791	2411.65	0.000
Portugal (PT)	−0.005	0.039	1.375	−0.227	9.927	8015.17	0.000
Russia (RS)	0.040	0.079	3.134	−1.300	41.272	244910.66	0.000
Spain (ES)	0.009	0.058	1.592	−0.006	7.890	3975.55	0.000
Sweden (SD)	0.025	0.070	1.928	0.007	6.969	2617.88	0.000
Switzerland (SW)	0.019	0.045	1.210	−0.087	7.335	3128.88	0.000
Turkey (TK)	0.018	0.057	2.983	−0.166	10.118	8444.53	0.000
United Kingdom (UK)	0.007	0.047	1.389	−0.160	10.747	9998.91	0.000
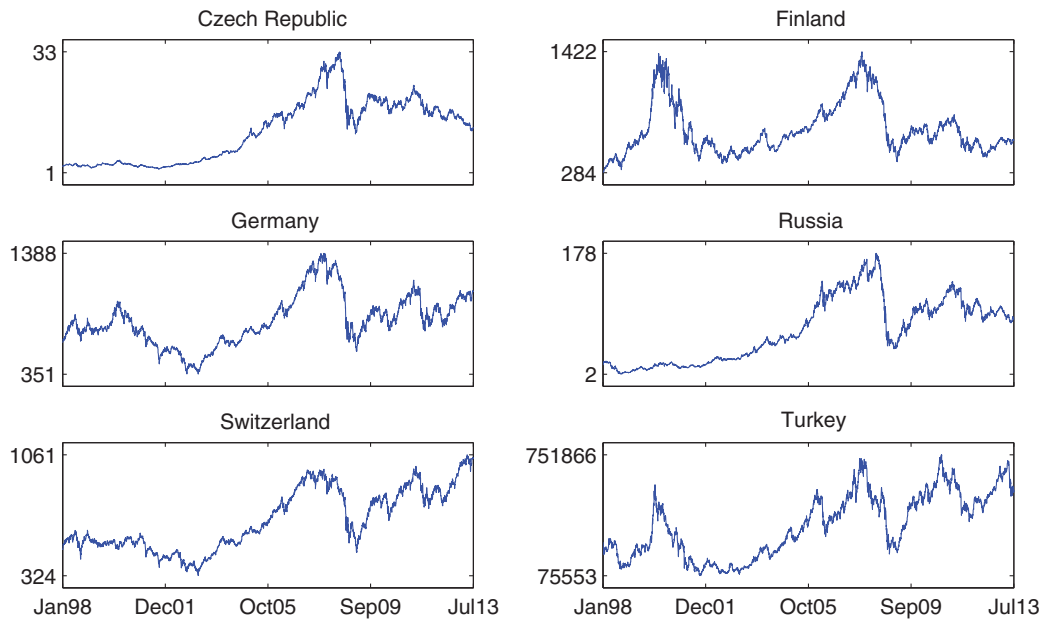
Fig. 1. Time series of price indexes for six European stock markets.

3. The econometric model

3.1. Model definition

The proposed model for statistical analysis of financial time series – the mixture Gaussian hidden Markov model (MGHMM) – contains three types of variables: a time-varying response variable, a time-constant discrete latent variable, and a time-varying discrete latent variable. Let y_{it} represent the metric response of observation i at time point t, where $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$. The time-constant and time-varying discrete latent variables are denoted by w_i and z_{it}, respectively, where $w_i \in \{1, \ldots, S\}$ and $z_{it} \in \{1, \ldots, K\}$. The latter implies that the number of categories of the two types of latent variables equals S and K, respectively. To make as clear as possible the distinction between the two types of latent variables, we will refer to w_i as a latent class and to z_{it} as a latent state or regime. The time-constant latent classes (w_i) can be seen as clusters for which the process under study differs. The time-varying latent variable which has a Markovian

transition structure (z_{it}) is used to flexibly model the distribution of the time-specific responses as well as to capture changes that occur across adjacent time points.

Fig. 5 provides the graph of the econometric model. The model is defined as:

$$f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{w_i=1}^{S} \sum_{z_{i1}=1}^{K} \sum_{z_{i2}=1}^{K} \cdots \sum_{z_{iT}=1}^{K} f(w_i, z_{i1}, \ldots, z_{iT}, y_{i1}, \ldots, y_{iT}) \quad (1)$$

where $f(w_i, z_{i1}, \ldots, z_{iT}, y_{i1}, \ldots, y_{iT})$ is the joint density of latent and observed variables and can be factorized into

$$f(w_i)f(z_{i1}|w_i)\prod_{t=2}^{T} f(z_{it}|z_{i,t-1}, w_i, y_{i,t-1}) \quad (2)$$

by

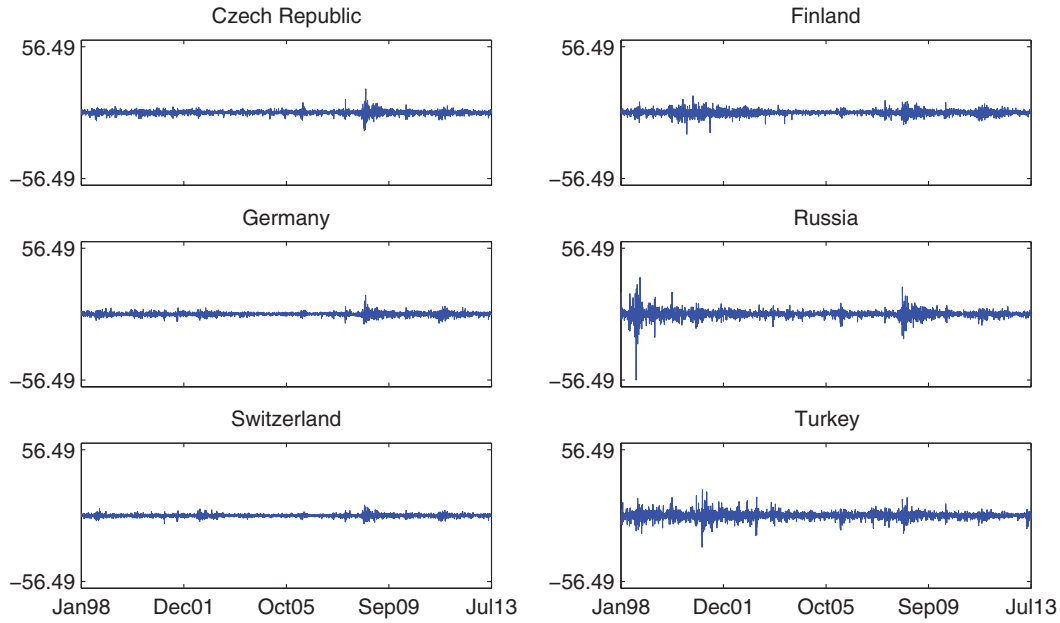$$\prod_{t=1}^{T} f(y_{it}|z_{it}). \quad (3)$$

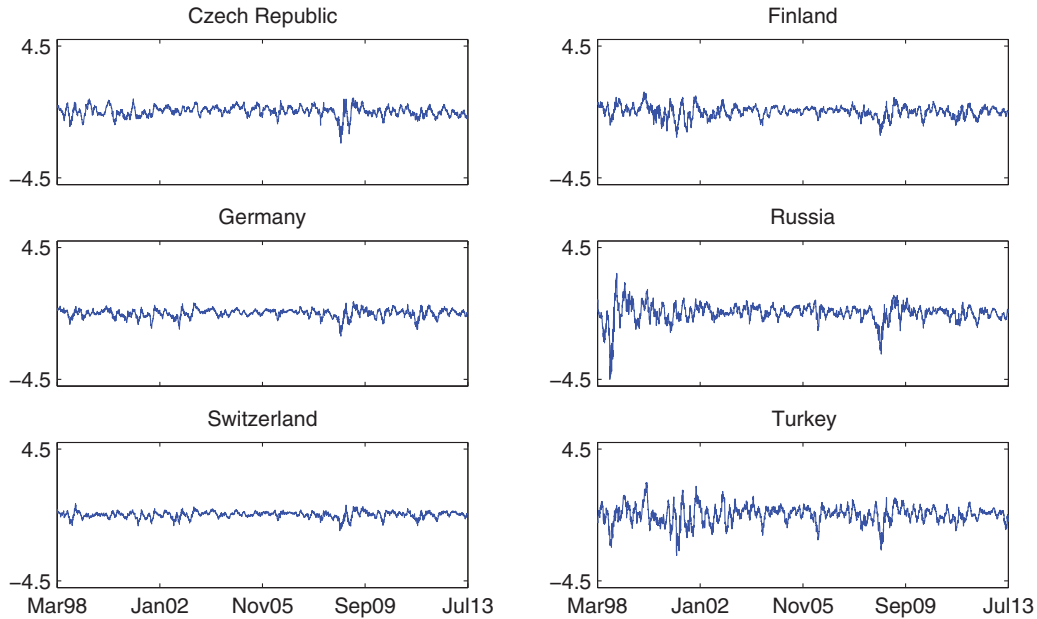**Fig. 2.** Time series of returns for six European stock markets.

Fig. 3. Time series of rolling means for six European stock markets (30-day window).

Eq. (1) describes $f(\mathbf{y}_i; \boldsymbol{\varphi})$, the (probability) density function associated with the time series of return rates for stock market index i. The right-hand side of this equation shows that we are dealing with a mixture model containing one time-constant latent variable and T time-varying latent variables. The total number of mixture components (or latent classes) equals $S \cdot K^T$ which is the product of the number of categories of w_i and z_{it} for $t = 1, 2, \ldots, T$. As in any mixture model, $f(\mathbf{y}_i; \boldsymbol{\varphi})$ results from the marginalization over the latent variables, which yields a weighted summation of class-specific probability densities with the (prior) class membership probabilities or mixture proportions as weights (McLachlan & Peel, 2000). The model described in Eqs. (1)–(3) is strongly related to the mixed latent Markov model proposed in the social sciences literature (van de Pol & Langeheine, 1990; Vermunt, Langeheine, & Bockenholt, 1999). Their model was, however, for categorical instead of continuous

variables and for (short) panel data rather than long time series like ours.

Fig. 5 and Eqs. (2) and (3) show the conditional independence assumptions implied by the MGHMM, and the resulting factorizations which simplify the model structure. More specifically, conditional on latent classes w_i and $y_{i,t-1}$, z_{it} is associated only with $z_{i,t-1}$ and $z_{i,t+1}$ and thus not with the latent states occupied at the other time points – the well-known first-order Markov assumption. Eq. (3) shows that conditionally on z_{it}, the response at occasion $t(y_{it})$ is independent of responses at other time points – usually referred to as the local independence assumption – and also independent of the latent classes and the latent states at the other time points.

Two remarks should be made about the first-order Markov assumption for the latent states. First, after marginalizing over w_i, the
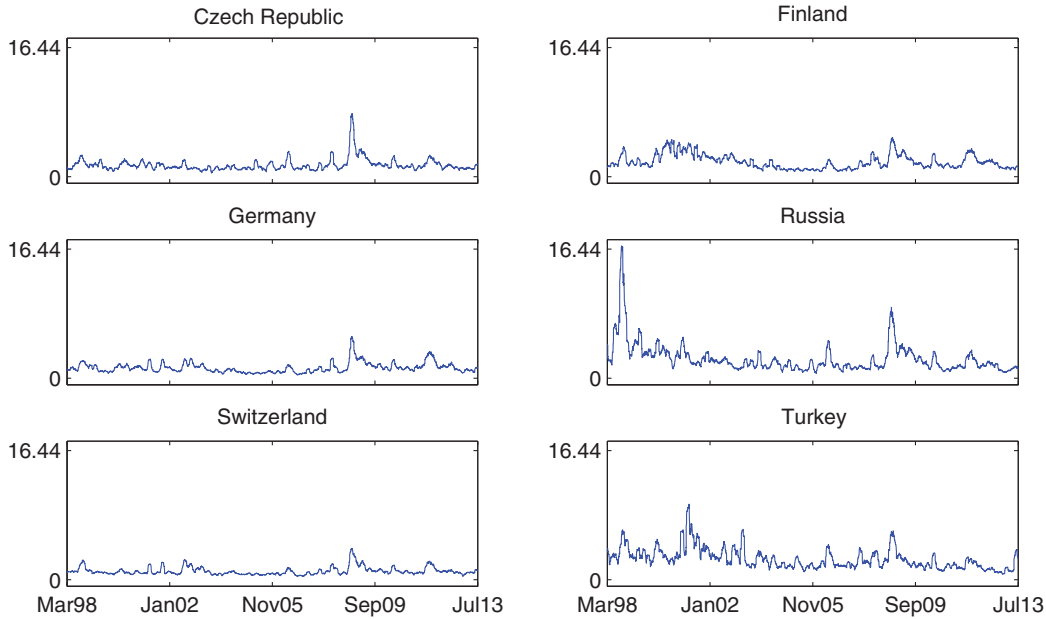
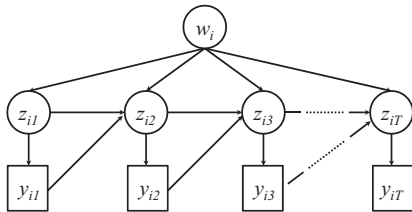**Fig. 4.** Time series of rolling standard deviations for six European stock markets (30-day window).

Fig. 5. The graph of the proposed model.

process for the sequence z_{it} is no longer Markovian.[2] Second, the Markov assumption for z_{it} conditionally on w_i allows a nonlinear structure for the responses y_{it}. This shows that the first-order Markov assumption is not as restrictive as one may initially think (see Fig. 5). Additionally, within each latent class the transition probabilities between regimes depend of the previous return, and consequently the Markov process is inhomogeneous.

The econometric model described in Eqs. (1)–(3) has three key relevant elements for the analysis of heterogeneous time series and takes into account: (1) time-invariant unobserved heterogeneity in the process under study, (2) serial dependency, and (3) flexible distributions that deviate in terms of skewness and kurtosis from normality (see, e.g., Dias & Wedel, 2004). Unobserved heterogeneity is captured by the time-constant latent variable w_i, serial dependencies are captured by the first-order Markov transition process in which the state at time point t depends on the state at time point $t - 1$ and the inhomogeneous Markov process that depends on the prior observed return, and, finally, flexible distributions of the returns are possible because of the time-specific mixture distribution for the response variable.

As can be seen from Eqs. (2) and (3), the model of interest is characterized by four sets of probability functions:

- $f(w_i)$ is the probability of belonging to a particular latent class w and $\pi_w = P(W_i = w)$;

---

[2] For example, for a given it the value z_{it} results from a mixture of S components with conditional multinomial distribution. As a mixture of Gaussian distributions is no longer (generally speaking) a Gaussian distribution, the same happens with mixtures of other distributions.

- $f(z_{i1}|w_i)$ is the initial regime probability; that is, the probability of having a particular latent initial state conditional on belonging to latent class w: $\lambda_{kw} = P(Z_{i1} = k|W_i = w)$, and $\boldsymbol{\lambda}_w$ is the vector containing initial probabilities in latent class w;
- $f(z_{it}|z_{i,t-1}, w_i, y_{i,t-1})$ is a latent transition probability; that is, the probability of being in a particular regime at time point t conditional on the latent state at time point $t - 1$, class membership, and observed return at time point $t - 1$; assuming an inhomogeneous transition process, we regress the transition probability on the prior stock return using a logit-link function

$$p_{jkw} = P(Z_{it} = k|Z_{i,t-1} = j, W_i = w, y_{i,t-1})$$
$$= \frac{e^{\gamma_{jkw}^{(0)} + \gamma_{jkw}^{(1)} y_{i,t-1}}}{\sum_{k'=1}^{K} e^{\gamma_{jk'w}^{(0)} + \gamma_{jk'w}^{(1)} y_{i,t-1}}},$$

where $\gamma_{jkw}^{(0)}$ and $\gamma_{jkw}^{(1)}$ are the intercept and slope coefficients of $y_{i,t-1}$ in the multinomial logit model for the transition from regime j to regime k within latent class w. As usual, identifying constraints on the gamma parameters are needed. The matrix with transition probabilities given w and $y_{i,t-1}$ is denoted by $\mathbf{P}_w(y_{i,t-1})$;
- $f(y_{it}|z_{it})$ is the Gaussian density function for the observed response, which is the probability density of having a particular observed stock return in index i at time point t conditional on the regime occupied at time point t. This distribution is characterized by the vector $\boldsymbol{\theta}_k = (\mu_k, \sigma_k^2)$ containing the means (μ_k) and variances (σ_k^2) for regime k, $k = 1, \ldots, K$ (and invariant across latent classes). The vector $\boldsymbol{\mu}$ contains all μ_k. Since the marginal distribution is a mixture of densities it defines a flexible model that takes skewness and kurtosis into account.

The $(SK + 1)(2K - 1)$ free parameters of the MGHMM ($\boldsymbol{\varphi}$) include the $S - 1$ class sizes, the $S(K - 1)$ initial state parameters and $2SK(K - 1)$ transition probability parameters, and the $2K$ conditional means and variances of the observed variables.

3.2. Restricted special cases of the model

Various special cases of the MGHMM defined in Eqs. (1)–(3) can be obtained by eliminating one or more of its three main elements, the

serial dependency structure, the time-varying latent variables, or the time-constant latent variable. For example, if we assume that there is no serial dependency between the time-varying discrete latent variables – that $p(z_{it}|z_{i,t-1}, w_i, y_{i,t-1}) \equiv p(z_{it}|w_i)$ – we obtain a model that is called a multilevel or hierarchical mixture model (Vermunt, 2003, 2007). This shows that the MGHMM can be seen as a hierarchical Gaussian mixture model that is expanded with a serial dependency structure.

The hidden Markov or Markov switching model (Baum, Petrie, Soules, & Weiss, 1970; Hamilton, 1989) is the special case of the MGHMM that is obtained by eliminating the time-constant latent variable w_i from the model, that is, by assuming that there is no unobserved heterogeneity at the upper level of analysis. This model can be obtained without modifying the formulae, but by simply assuming that $S = 1$; that is, that all stock markets belong to the same group.

The mixture Gaussian model can be seen as a restricted variant of the MGHMM that is obtained by removing the time-varying latent variables z_{it}, resulting in $f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{w_i=1}^{S} f(w_i) \prod_{t=1}^{T} f(y_{it}|w_i)$. Note that this model is equivalent to a latent class model for T response variables which are assumed to be conditionally independent within latent classes w. For $T = 1$ it yields the mixture of (univariate) Gaussian distributions (Dias & Wedel, 2004).

3.3. Parameter estimation by maximum likelihood

Maximum likelihood (ML) estimation of the parameters of the MGHMM involves maximizing the log-likelihood function: $\ell(\boldsymbol{\varphi}; \mathbf{y}) = \sum_{i=1}^{n} \log f(\mathbf{y}_i; \boldsymbol{\varphi})$, a problem that can be solved by means of the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977).

In the E step, we compute the expected value of the $T + 1$ latent variables given the observed data and the current provisional estimates of the model parameters. In the M step, standard complete data ML methods are used to update the unknown model parameters using an expanded data matrix with previous expectation as weights. Since the EM algorithm needs to compute and store the $S \cdot K^T$ entries of expectations over the latent space for each subject or for each unique data pattern, computation time and computer storage increases exponentially with the number of time points, which makes this algorithm impractical or even impossible to apply with more than a few time points. This explains why models like the one proposed here have been used in social science applications only with very short time series (van de Pol & Langeheine, 1990; Vermunt et al., 1999) and has hampered its application for longer time series (Schmittmann, Visser, & Raijmakers, 2006).

However, for hidden Markov models, a special variant of the EM algorithm has been proposed that is usually referred to as the forward–backward or Baum–Welch algorithm (Baum et al., 1970; Hamilton, 1989). This special algorithm is needed because the model for our data set contains a huge number of entries in the joint posterior latent distribution generated by the $T + 1$ latent variables. Recall that in our application T is quite large. This means that even for $S = 2$ and $K = 2$, the number of entries in the joint posterior distribution is too large to process and store for all n stock markets as has to be done within the standard EM algorithm. The Baum–Welch algorithm circumvents the computation of this joint posterior distribution making use of the conditional independencies implied by the model. Whereas this algorithm was originally proposed for hidden Markov models, here we expand it to deal with mixture (Gaussian) hidden Markov models. This algorithm exploits the conditional independence assumptions implied by the model in order to circumvent the computational issue which affects the traditional EM procedure when the number of time points is high. A detailed description of the extended Baum–Welch algorithm is provided in Appendix A.

3.4. Decision on the number of latent classes and regimes

Contrary to previous work (e.g., Ang & Bekaert, 2002; Wilfling, 2009; Pagan & Sossounov, 2003), we do not set a priori the number of regimes as it happens, for instance, in the bear–bull market framework ($K = 2$). Consequently, we have to address simultaneously the estimation/selection of S and K, i.e., how to decide about the number of latent classes and regimes needed. The standard model selection approach when using maximum likelihood estimation is by performing likelihood ratio tests across nested models; here the relevant tests are between models with $S - 1$ and S classes and between models with $K - 1$ and K regimes. However, in the context of latent class models this approach is problematic because the null hypothesis under test is defined on the boundary of the parameter space, and consequently the regularity condition of Cramer on the asymptotic properties of the MLE is no longer valid. As an alternative, it has been proposed to use information criteria, mostly for cross-sectional modeling. The basic principle under these criteria is parsimony, which results from the trade-off between model fit and model complexity. A number of model selection criteria have been suggested, the most prominent and widely used being the Bayesian Information Criterion (BIC) of Schwarz (1978) and the Akaike Information Criterion (AIC) of Akaike (1974). More recently, simulation studies have explored the performance of information criteria for longitudinal data (Costa & De Angelis, 2010; Dias, 2007; Dias & Vermunt, 2007). These type of simulation studies compare criteria such as Bayesian information criterion (BIC), the Akaike information criterion (AIC), the AIC3, and the consistent AIC (CAIC) in terms of performance. The overall conclusion is that AIC tends to overfit (see, e.g., Dias & Vermunt, 2007) and BIC performs better for large T (see, e.g., Costa & De Angelis, 2010). Thus, our model selection is based on the BIC. The values of S and K to be selected are the ones that minimize the value of BIC, which is defined as

$$\text{BIC}_{S,K} = -2\ell_{S,K}(\hat{\boldsymbol{\varphi}}; \mathbf{y}) + N_{S,K} \log n, \tag{4}$$

where $N_{S,K}$ is the number of free parameters of the model with S latent classes and K regimes. Note that as indicated by Dias (2007), we use sample size as n in the BIC formula, and thus not nT.

4. Results

This section reports the results obtained when applying the statistical model described in the previous section to the stock return data set at hand. Because we want to understand the forecasting ability of the model across countries, we split the sample into two parts: the information until the end of 2012 is used for estimation (in-sample), and the information from 2013 is used for forecasting (out-of-sample).

We estimated models using S and K values ranging from one to five, with 500 different sets of random starting values to minimize the effect of local maxima. A solution with two latent classes ($S = 2$) and three regimes ($K = 3$) yields the lowest BIC value (log-likelihood = -145689.56; number of free parameters = 35; and BIC = 291485.68).[3] This means that the best solution incorporates two types of regime dynamics between the three regimes. This model will therefore be treated as the final model in our analysis.

We start by characterizing the latent states or regimes. Table 2 provides information on the three regimes; that is, the average proportion of markets in regime k over time and the mean and variance of the return in each regime. The reported means show that the first regime has negative and the third expected positive returns. On the contrary, in regime 2, the mean return is not significantly different from zero. In the financial jargon regimes 1 and 3 are called bear and

---

[3] Detailed results on model selection are available from the first author upon request. Since for $n = 21 \log 21 = 3.04$, BIC $\simeq$ AIC3 and also not too far from AIC.
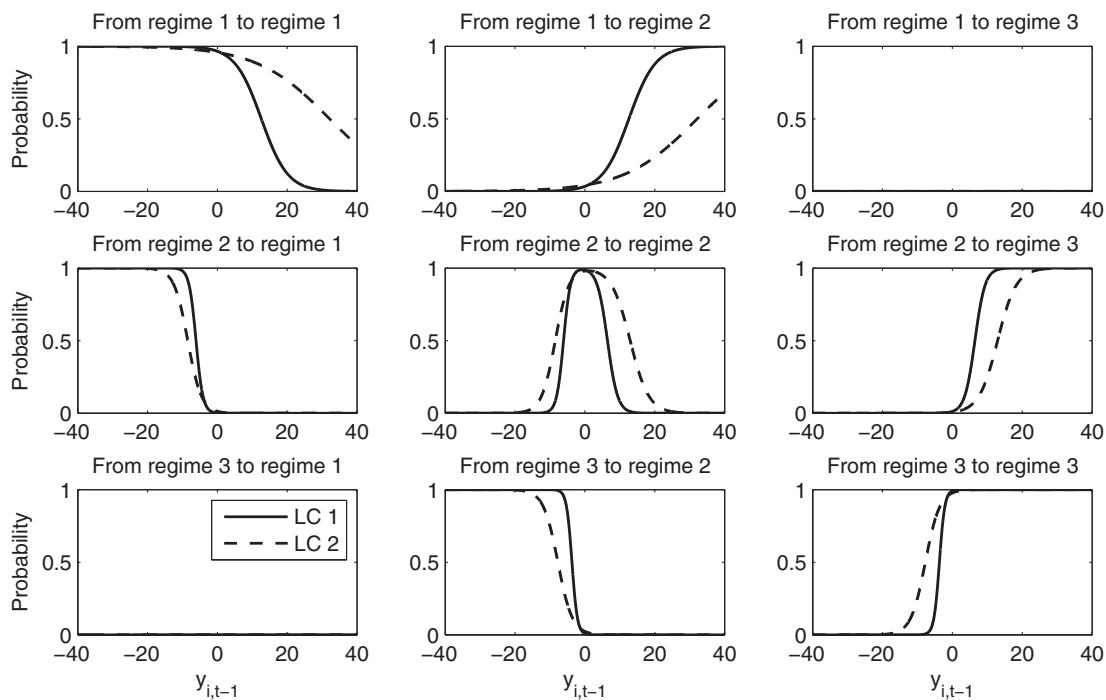
Fig. 6. Estimates of transition probabilities between regimes for both latent classes (LCs).

Table 2
Estimated state occupancies and Gaussian parameters within regimes.

Regimes	P(Z)	Return (mean: $\hat{\mu}_k$)			Risk (variance: $\hat{\sigma}_k^2$)		
		Estimate	s.e.	p-value	Estimate	s.e.	p-value
1	0.179	−0.213	0.042	0.000	15.591	0.337	0.000
2	0.532	−0.003	0.010	0.760	2.777	0.046	0.000
3	0.290	0.095	0.005	0.000	0.777	0.010	0.000

Table 3
State occupancy within latent classes.

Regimes	Latent class 1	Latent class 2
1	0.142	0.251
2	0.512	0.571
3	0.346	0.178

bull markets, while regime 2 corresponds to a stable regime in which returns are close to 0. The overall probability of being in regime 1 and 3 equals 0.18 and 0.29, respectively. The most frequently occurring regime with more than half of the time points is the one with an almost null expected return.

The asymmetry in the volatility coefficients is consistent with the pattern in stock markets: bear regimes are associated with larger volatility than bull regimes. Thus, in periods of financial crisis, market prices fall down and have higher volatility. Due to financial markets integration, such events tend to spillover to other markets, the so called *financial contagion*.[4]

Table 3 provides the probability of being in one of the three regimes for each of the two latent classes. We observe that in both latent classes the process is in regime 2 most of the time. It can be observed that the two latent classes differ with respect to the bullish regime probability. Latent class 1 tends to spend less time in the bear regime than in the bull one, whereas in latent class 2 happens the opposite; that is, class 2 countries are in the bull regime at only 17.8 percent of the time points, whereas this figure equals 34.6 percent for class 2.

Fig. 6 depicts the transition probabilities as function of y_{t-1} for both latent classes. The probability of exiting regime 1 (high volatility and negative returns) is close to zero when previous returns are negative. As returns become positive, the probability of switching to regime 2 increases. The probability of switching between regimes 1 and 3 is always 0, indicating that regime 2 is in fact an intermediate regime between the crisis and bull regimes. The transition probabilities for regime 2 are close to 0 when returns are close to 0, indicating that the market is likely to continue in regime 2. Negative returns increase the likelihood to switch to regime 1 and positive returns increase the likelihood to switch to regime 3. The probability of moving out of regime 3 is 0 when returns are positive, and becomes large when returns decrease (returns become more negative). Class 1 differs from class 2 in the size of the transition probabilities; that is, the transition probabilities are larger in class 1 than in class 2, and more strongly affected by the previous return. For instance, in latent class 1 the transition probability from regime 1 to regime 2 is higher than in latent class 2. For a value of $y_{i,t-1}$ around 20, the probability of transition from regime 1 to regime 2 is close to 1 in latent class 1, whereas latent class 2 this probability is around 0.25, i.e., regimes are more persistent in latent class 2.

Table 4 summarizes the results related to the distribution of stock markets across latent classes. The estimated prior class membership probabilities indicate that the probability that a randomly selected stock market belongs to class 1 is 0.66 and to class 2 is 0.34. A more detailed interpretation of the nature of the stock market latent variable is obtained by investigating the posterior class membership probabilities, conditional on the observed data (Table 4). As can be seen, 14 countries are assigned to class 1 and 7 to class 2. Except for Sweden, the class assignment probability is always one, indicating that the classes are fully non-overlapping. Note that even for Sweden the misclassification probability is low, assuming that we assign each stock market to the most likely latent class.

As can be seen, class 1 is composed mainly by European developed financial markets: Austria, Belgium, Czech Republic, Denmark, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, the UK and two non-E.U. countries, Norway and Switzerland. Class 2 includes Finland, Greece, Hungary, Poland, Russia, Sweden, and Turkey. While

---

[4] Some examples are the *Asian Flu Crisis* that affected Southeastern Asian countries and the *Russian Crisis* of 1998.

Table 4
Estimated prior probabilities, posterior probabilities, and modal class.

	Latent class 1	Latent class 2	Modal
Prior probabilities	0.663	0.337	
Posterior probabilities			
Austria (OE)	1.000	0.000	1
Belgium (BG)	1.000	0.000	1
Czech Rep. (CZ)	1.000	0.000	1
Denmark (DK)	1.000	0.000	1
Finland (FN)	0.000	1.000	2
France (FR)	1.000	0.000	1
Germany (BD)	1.000	0.000	1
Greece (GR)	0.000	1.000	2
Hungary (HN)	0.000	1.000	2
Ireland (IR)	1.000	0.000	1
Italy (IT)	1.000	0.000	1
Netherlands (NL)	1.000	0.000	1
Norway (NW)	1.000	0.000	1
Poland (PO)	0.000	1.000	2
Portugal (PT)	1.000	0.000	1
Russia (RS)	0.000	1.000	2
Spain (ES)	1.000	0.000	1
Sweden (SD)	0.079	0.921	2
Switzerland (SW)	1.000	0.000	1
Turkey (TK)	0.000	1.000	2
United Kingdom (UK)	1.000	0.000	1

this second group is mainly composed of Eastern European countries, it also includes three countries whose stock markets have been severely affected by economic and financial crisis: Finland, Greece,

and Sweden. Note that the countries in class 2 are the ones with more pronounced booms and slumps in volatility. The next paragraphs provide more explanations about of the reasons behind the clustering.

Figs. 7 and 8 show the regime-switching dynamics for countries belonging to the same latent class. They depict the posterior probability of being in each regime at period *t*. The figures show that the dynamic patterns of the return indices are remarkably similar for countries within each class and substantially different between the two classes. Fig. 7 shows that class 1 stock markets displayed little propensity to switch to the bear regime. We see a long period of bull markets and two main periods of bear markets: the first ranging from the end of 1999 until 2003 with episodic crises and the second representing the subprime crisis. Episodes of bear markets were not very frequent and did not last long, except during the global financial crisis. Countries in class 2 had higher propensity to switch to bear regimes (Fig. 8). Differently from class 1 markets, these stock markets seem more affected by the 2000 crisis. But the subprime crisis affected them in a similar way as the stock markets from class 1. Greece and Hungary, two countries close to bankruptcy due to the sovereign crisis and with a bailout from the IMF, spent a long period in bear markets in the beginning of 2010. The bursting of the dot-com bubble had serious economic implications for Sweden. Ericsson, a large producer of mobile telecommunications equipment and one of the main listed Swedish securities, suffered heavy losses after the dot-com crash in the early 2000s. It shed thousands of jobs, affecting many other consulting companies and start-ups in the IT sector. The stock markets of the two neighbor countries Russia and Finland showed similar
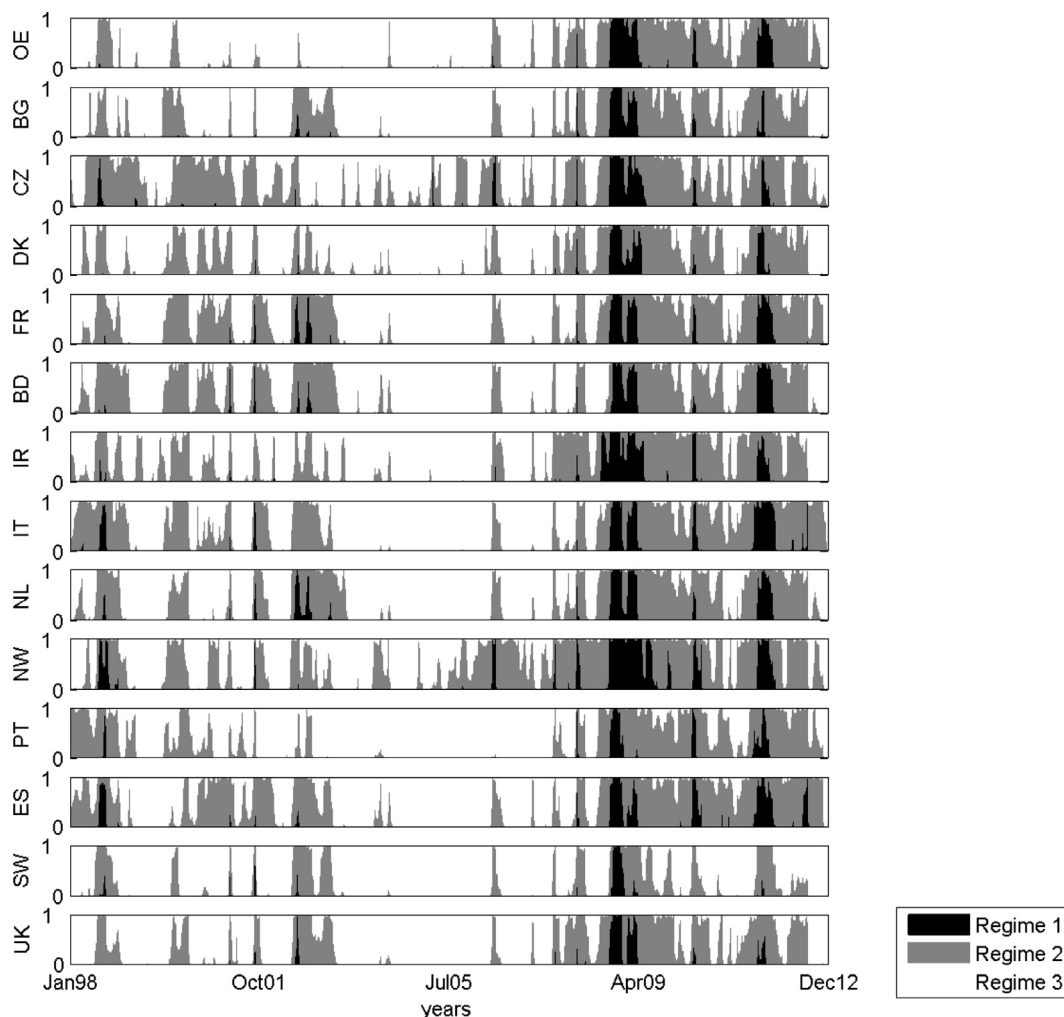
Fig. 7. Estimated posterior latent state probabilities and modal state in latent class 1.
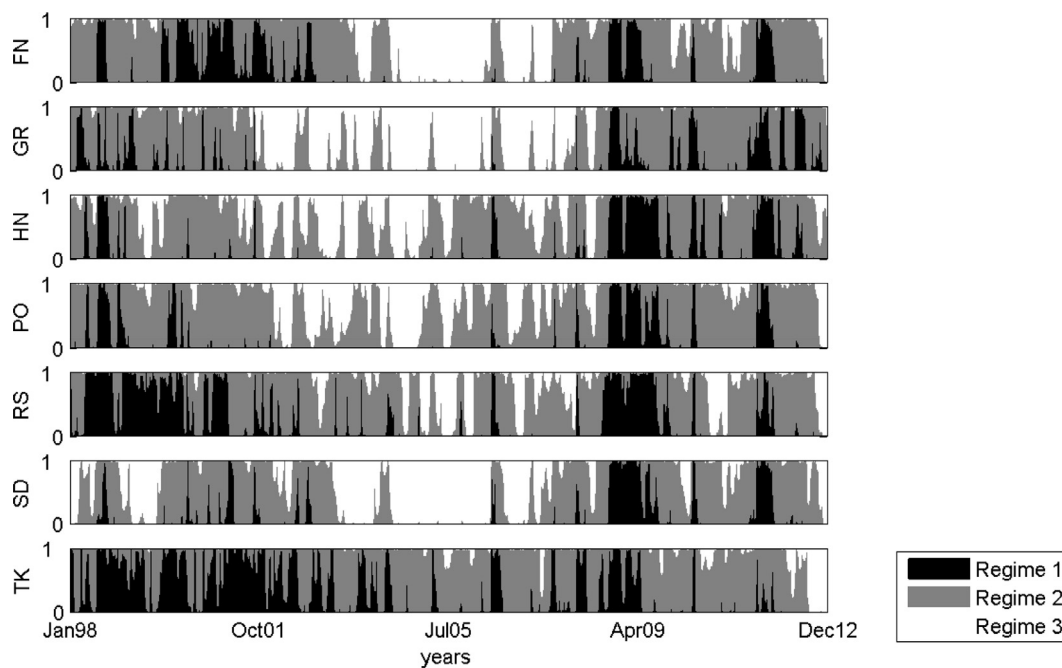
Fig. 8. Estimated posterior latent state probabilities and modal state in latent class 2.

dynamics. Between 2003 and 2007, Finland and Sweden had a period of bull market, while Russia tended to switch regularly between bull and bear markets. Finally, Turkey switched very quickly between regimes, most of the time in the intermediate regime and quite often in the bear regime.

Despite initial skepticism about the out-of-sample prediction performance of regime-switching models (Dacco & Satchell, 1999; Engel, 1994), recently De Angelis and Paas (2013) found evidence that regime-switching models may outperform the threshold GARCH model with Student-*t* innovations both in terms of in-sample and out-of-sample prediction. To evaluate the forecasting ability of the model for each country, we compare the one-day ahead forecast to the actual return. Parameters were estimated with the subsample that goes till the end of 2012 ($T = 3857$). Thus, the in-sample period is from $t = 1$ to T, whereas the out-of-sample is from $t = T + 1$ to $T^* = 4009$. We compare our results with those obtained with a model that is commonly used to analyze stock market indexes, namely the GJR-AR(1)-GARCH(1,1) model (Glosten, Jagannathan, & Runkle, 1993) with Student-*t* innovations which captures both asymmetry and non-normality. Our RSM shows better in-sample prediction than the GARCH model. In terms of out-sample prediction, both models show similar performance. Appendix B gives a detailed description of the relevant GARCH model, and the results obtained with this model.

The last important question to address is market synchronization. Measurement of synchronization of stock markets using cross-correlations of returns is rather popular, but crisis periods may yield very large outliers in returns that introduce noise and inflate similarities between markets. To measure synchronization and co-movement in the 21 stock markets, we compute the association between markets using the posterior probability of being in a given regime. In other words, synchronization is measured by the similarity of markets with respect to the likelihood of being in the same regime.

Let $\hat{\alpha}_{itk}$ be the estimated posterior probability that market i in period t will be in regime k. To obtain a number in the full range of real numbers, this probability is expressed using the logit transformation:

$$\text{logit}_{itk} = \log \left(\frac{\hat{\alpha}_{itk}}{1 - \hat{\alpha}_{itk}} \right). \tag{5}$$

Synchronization is quantified as the product-moment correlation between two markets' posterior regime logits. This transformation is

needed as the probabilities are between 0 and 1, and the Pearson correlation operates on the real line. Our logit-based measure does not suffer from distortion caused by outliers because it filters out extreme observations of returns. The measure gets close to 1 if markets tend to share the same regime.

Table 5 shows the correlations between stock markets using this measure applied to regime 2, the most frequent one. Countries belonging to class 1 show a high level of synchronization, but they are less synchronized with countries in the other latent class. Class 2 contains more heterogeneous countries: Turkey and Russia have low synchronization with all markets; Sweden and Finland are strongly synchronized; Greece shows strong synchronization with all EMU (European Monetary Union) countries, most of which are in the other group. Thus, it is worth to note that even if countries are not in the same class, returns can be synchronized in the sense that stock markets tend to be in the same regime at the same time.

5. Conclusion

The identification of groups or classes of financial assets is key in the field of portfolio management. One of the first decisions in the portfolio construction process is the asset allocation decision, which consists of allocating investments among different asset classes in order to optimize the risk/reward trade-off. Our statistical methodology is a powerful tool to identify groups of homogeneous financial assets.

This paper introduced a model-based clustering technique as a data analytic tool for financial time series analysis. The proposed model takes into account both time-constant unobserved heterogeneity between and hidden regimes within time series. Moreover, the flexible modeling of observed responses using a mixture of normal distributions makes it straightforward to capture almost any departure from the normality. For parameter estimation using maximum likelihood, a generalization of the Baum–Welch algorithm for the HMM to the mixture HMM was used.

In the analysis of a sample of 21 stock markets, the best fitting model was the one with two groups. The groups clearly reveal distinguished stock market dynamics, that is, in the switching patterns between the three regimes.

The results obtained with our approach seem to have an one-to-one correspondence with the stock market behavior. First, regimes are

Table 5
Synchronization of stock markets (correlation between posterior probabilities in regime 2).

Countries	OE	BG	CZ	DK	FN	FR	BD	GR	HN	IR	IT	NL	NW	PO	PT	RS	ES	SD	SW	TK
Austria (OE)																				
Belgium (BG)	0.71																			
Czech Rep. (CZ)	0.51	0.39																		
Denmark (DK)	0.68	0.78	0.51																	
Finland (FN)	0.37	0.43	0.25	0.31																
France (FR)	0.66	0.78	0.45	0.75	0.43															
Germany (BD)	0.55	0.74	0.41	0.69	0.41	0.90														
Greece (GR)	0.42	0.38	0.37	0.35	0.41	0.44	0.39													
Hungary (HN)	0.11	-0.07	0.28	0.03	0.07	0.08	0.00	0.22												
Ireland (IR)	0.68	0.71	0.46	0.69	0.42	0.67	0.63	0.47	0.13											
Italy (IT)	0.62	0.69	0.44	0.65	0.47	0.85	0.81	0.51	0.17	0.63										
Netherlands (NL)	0.70	0.83	0.39	0.70	0.52	0.89	0.84	0.37	-0.01	0.70	0.78									
Norway (NW)	0.48	0.38	0.37	0.43	0.22	0.42	0.34	0.20	0.40	0.48	0.43	0.41								
Poland (PO)	0.22	0.10	0.33	0.20	0.29	0.20	0.16	0.32	0.57	0.29	0.23	0.13	0.39							
Portugal (PT)	0.65	0.66	0.45	0.62	0.31	0.65	0.60	0.48	0.04	0.59	0.70	0.63	0.24	0.16						
Russia (RS)	0.03	-0.04	-0.02	-0.02	0.22	0.01	-0.02	-0.02	0.24	0.07	-0.03	0.06	0.28	0.29	-0.11					
Spain (ES)	0.56	0.61	0.42	0.64	0.42	0.84	0.76	0.53	0.14	0.60	0.84	0.72	0.34	0.29	0.68	0.05				
Sweden (SD)	0.42	0.50	0.33	0.48	0.60	0.69	0.61	0.38	0.29	0.54	0.62	0.59	0.48	0.41	0.37	0.23	0.64			
Switzerland (SW)	0.45	0.70	0.21	0.59	0.24	0.60	0.64	0.21	-0.27	0.49	0.52	0.64	0.08	-0.14	0.51	-0.14	0.49	0.23		
Turkey (TK)	0.07	-0.05	-0.01	-0.04	0.04	-0.17	-0.20	-0.14	0.10	0.04	-0.18	-0.06	0.18	0.10	-0.11	0.31	-0.17	-0.09	-0.22	
United Kingdom (UK)	0.69	0.86	0.37	0.76	0.37	0.78	0.72	0.35	-0.06	0.73	0.66	0.84	0.41	0.08	0.60	0.02	0.63	0.48	0.73	-0.06

consistent with the asymmetry of volatility: the bear regime presents higher volatility than the bull regime. Second, periods of market crises are clearly distinguished from periods of market booms and stability. Third, less developed markets tend to be assigned with a higher probability to the regime associated with higher volatility. Fourth, the transition probabilities are consistent with the persistency of volatility. Overall, the econometric approach is coherent with many stylized facts in finance.

Moreover, the specification of the correct distribution for returns has important implications. The fact that risk models do not account for non-zero higher moments might cause bias in hedging strategies and concomitantly great losses for financial institutions. For instance, it is widely acknowledged that financial institutions have gradually been adopting market risk models that rely more on historic or back simulations instead of symmetric distributions. Portfolio decisions also need to incorporate information about those higher moments. Bekaert, Erb, Harvey, and Viskanta (1998) analyzed the economic impact of taking into account skewness and kurtosis on asset allocation. Their results based on simulations show that investment weights are increased toward the asset with positive skewness (everything constant) and with higher kurtosis (holding skewness positive and constant). The semiparametric nature of the proposed MGHMM allows a flexible specification of the distribution of the observed returns that goes beyond the Gaussian distribution.

Results of the application show that our statistical methodology performs well in capturing the different regime dynamics of stock markets. It clearly distinguishes two groups of countries with different observed patterns. It also goes beyond the traditional approaches of categorizing, for instance, countries in developed and emerging market based on GDP per capita (followed by the World Bank) or credit risk ratings as it is directly based on daily market returns. This methodology highlights distinct dynamics of markets due to their idiosyncrasies, i.e., their arise endogenously. Standard approaches fail to incorporate this phenomenon in asset allocation decisions. For instance, in asset allocation decisions, investment in assets with similar "statistical" properties is redundant and can lead to the destruction of portfolio value. In other words, in the traditional mean-variance framework values do not span the mean-variance frontier and do not enhance the risk-return trade-off of the portfolio. In addition, trading costs of investing in "redundant assets" also erodes portfolio value. Additionally, understanding cycles in stock markets is key because portfolio allocation strategies are regime dependent (Ang & Bekaert, 2002).

Acknowledgments

Appendix A. The extended Baum–Welch algorithm

Because of the conditional independence assumption implied by our statistical model, in the M step of the EM algorithm one needs only the distributions $f(w_i|\mathbf{y}_i)$, $f(w_i, z_{it}|\mathbf{y}_i)$, and $f(w_i, z_{i,t-1}, z_{it}|\mathbf{y}_i)$. The Baum–Welch algorithm obtains these quantities directly rather than first computing the joint distribution of the latent variables given the observed data and parameter estimates (E step) and subsequently collapsing over the remaining dimensions as would be done in the standard EM algorithm (Baum et al., 1970). The two key components of the Baum–Welch algorithm are the forward probabilities α_{iwz_t} and the backward probabilities β_{iwz_t}. Because of our generalization to the mixture case, we need one additional quantity γ_{iw}. These three quantities are defined as follows:

$$\alpha_{iwz_t} = f(z_{it}, y_{i1}, \ldots, y_{it} | w_i, y_{i,t-1}), \tag{A.1}$$

$$\beta_{iwz_t} = f(y_{i,t+1}, \ldots, y_{iT} | z_{it}, w_i, y_{i,t}), \tag{A.2}$$

$$\gamma_{iw} = f(w_i, \mathbf{y}_i). \tag{A.3}$$

Using α_{iwz_t}, β_{iwz_t}, and γ_{iw}, one can obtain the relevant marginal posteriors as follows:

$$f(w_i | \mathbf{y}_i) = \frac{\gamma_{iw}}{f(\mathbf{y}_i)} \tag{A.4}$$

$$f(w_i, z_{it} | \mathbf{y}_i) = \frac{\alpha_{iwz_t} \beta_{iwz_t}}{f(\mathbf{y}_i)} \tag{A.5}$$

$$f(w_i, z_{i,t-1}, z_{it} | \mathbf{y}_i) = \frac{\gamma_{iw} \alpha_{iwz_{t-1}} f(z_{it} | z_{i,t-1}, w_i, y_{i,t-1}) f(y_{it} | z_{it}, w_i) \beta_{iwz_t}}{f(\mathbf{y}_i)} \tag{A.6}$$

where $f(\mathbf{y}_i) = \sum_{w_i=1}^{S} \gamma_{iw}$, and $f(z_{it} | z_{i,t-1}, w_i, y_{i,t-1})$ and $f(y_{it} | z_{it}, w_i)$ are model probabilities. The key element of the forward–backward algorithm is that the α_{iwz_t} and β_{iwz_t} are computed using recursive schemes. The forward recursion for α_{iwz_t} is:

$$\alpha_{iwz_1} = f(z_{i1} | w_i) f(y_{i1} | z_{i1}, w_i), \tag{A.7}$$

$$\alpha_{iwz_t} = \left[\sum_{z_{i,t-1}=1}^{K} \alpha_{iwz_{t-1}} f(z_{it} | z_{i,t-1}, w_i, y_{i,t-1}) \right] f(y_{it} | z_{it}, w_i), \tag{A.8}$$

for $t = 2$ up to $t = T$. The backward recursion for β_{iwz_t} is:

$$\beta_{iwz_T} = 1, \tag{A.9}$$

$$\beta_{iwz_t} = \sum_{z_{i,t+1}=1}^{K} \beta_{iwz_{t+1}} f(z_{i,t+1} | z_{it}, w_i, y_{i,t}) f(y_{i,t+1} | z_{i,t+1}, w_i), \tag{A.10}$$

for $t = T - 1$ down to $t = 1$. The quantity γ_{iw} is obtained as:

$$\gamma_{iw} = \sum_{z_{it}=1}^{K} f(w_i) \alpha_{iwz_t} \beta_{iwz_t}, \tag{A.11}$$

for any t. So, first we obtain α_{iwz_t} and β_{iwz_t} for each time point and subsequently we obtain γ_{iw}. Next, we compute $f(w_i | \mathbf{y}_i)$, $f(w_i, z_{it} | \mathbf{y}_i)$, and $f(w_i, z_{i,t-1}, z_{it} | \mathbf{y}_i)$ using Eqs. (A.4)–(A.6). In the M step, these quantities are used to obtain new estimates for the model parameters appearing in Eqs. (2) and (3) using standard methods.

Similar recursive schemes have been proposed for obtaining the gradient vector and the observed information matrix (Lystig & Hughes, 2002). These can be used to maximize the log-likelihood using the Newton–Raphson algorithm.

Appendix B. Forecasting performance of the proposed model

B.1. Forecasting using the proposed model

Let $\mathbf{e}_{it}$ be the vector containing the probability of the process being in a particular regime at time point t, conditional on model parameters and information up to $t - 1$. The predicted values $\hat{\mathbf{e}}_{it}$ and $\hat{y}_{it}$, conditional on stock market i being in latent class w, is obtained by the following recursive expressions:

$$\hat{\mathbf{e}}_{i1} = \hat{\boldsymbol{\lambda}}_w^T, \quad t = 1$$
$$\hat{\mathbf{e}}_{it} = \hat{\mathbf{e}}_{i,t-1} \hat{\mathbf{P}}_w(y_{i,t-1}), \quad t > 1$$
$$\hat{y}_{it} = \hat{\mathbf{e}}_{it} \hat{\boldsymbol{\mu}}.$$

Table B.1
Measures of forecasting accuracy.

	Stock market i	Total
In-sample		
MSE	$\frac{1}{T} \sum_{t=1}^{T} (y_{it} - \hat{y}_{it})^2$	$\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \hat{y}_{it})^2$
MAE	$\frac{1}{T} \sum_{t=1}^{T} |y_{it} - \hat{y}_{it}|$	$\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} |y_{it} - \hat{y}_{it}|$
Out-of-sample		
MSE	$\frac{1}{T^*-T} \sum_{t=T+1}^{T^*} (y_{it} - \hat{y}_{it})^2$	$\frac{1}{n(T^*-T)} \sum_{i=1}^{n} \sum_{t=T+1}^{T^*} (y_{it} - \hat{y}_{it})^2$
MAE	$\frac{1}{T^*-T} \sum_{t=T+1}^{T^*} |y_{it} - \hat{y}_{it}|$	$\frac{1}{n(T^*-T)} \sum_{i=1}^{n} \sum_{t=T+1}^{T^*} |y_{it} - \hat{y}_{it}|$

In case of a single time series and a homogeneous Markov process, this recursion reduces to the set of equations given in Hamilton (1993, 1994). Our forecasting tool adds two elements to the traditional forecasting by regime-switching models: first, the forecasts are done simultaneously for each cross-section, which adds robustness to the process; second, by allowing for an inhomogeneous process (by conditioning on the returns at time $t - 1$) the forecasts are tailored for each stock market i at time t.

B.2. Forecasting using the GJR-AR(1)-GARCH(1,1) model

The GJR-AR(1)-GARCH(1,1) model (Glosten et al., 1993) is defined by an autoregressive process of order 1

$$y_{it} = \mu_i + \phi_i y_{i,t-1} + u_{it}, \tag{B.1}$$

where $u_{it} = \sigma_{it} \varepsilon_{it}$ and the variance equation is

$$\sigma_{it}^2 = \kappa_i + \gamma_i \sigma_{i,t-1}^2 + \alpha_i u_{i,t-1}^2 + \xi_i I(u_{i,t-1} < 0) u_{i,t-1}^2, \tag{B.2}$$

where the indicator function $I_t(u_{i,t-1} < 0)$ equals 1 if $u_{i,t-1} < 0$ and 0 otherwise. Thus, the leverage coefficient gives additional weight to negative innovations. The conditional distribution of errors (ε_{it}) is a Student-t with more than 2 degrees of freedom. This GARCH model was estimated for each stock market separately. For sake of space, these results are not reported here but are available from the first author.

B.3. Accuracy measures

The in-sample and out-of-sample fit measures are the mean square error (MSE) and the mean absolute error (MAE), both described in Table B.1.

B.4. Forecasting performance

Table B.2 summarizes the in-sample and out-of-sample forecasting results. As can be seen, the model performs better for the out-of-sample period than for the in-sample period. This can be explained by the fact that the in-sample period contained the global financial crisis for which prediction errors were large. On the other hand, the calm period of first semester of 2013 (part of out-of-sample period) was described very well. We can also compare the forecasting accuracy across countries, and define a ranking of market uncertainty. For example, Switzerland is the stock market with the lowest in-sample error, whereas the United Kingdom has the lowest out-of-sample error. On the other extreme, Russia and Turkey tend to have high forecasting errors. For some countries the out-of-sample ranking is lower (relatively better prediction) than the in-sample ranking (e.g., Finland and Hungary), whereas for others the reverse applies (e.g., Greece).

In terms of competing models, we conclude that our model performs better in-sample than the AR(1)-GJR-GARCH(1,1) and shows similar out-of-sample performance. These results open promising avenues for further research and developments in terms of model forecasting.

Table B.2
Forecasting accuracy.

Stock markets	Proposed model				AR(1)-GJR-GARCH(1,1) model			
	In-sample		Out-of-sample		In-sample		Out-of-sample	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Austria (OE)	1.465	1.001	1.094	0.798	1.981	1.388	1.096	0.796
Belgium (BG)	1.391	0.983	1.270	0.929	1.884	1.341	1.269	0.930
Czech Rep. (CZ)	1.716	1.193	1.182	0.929	2.345	1.651	1.199	0.943
Denmark (DK)	1.437	1.011	1.027	0.741	1.984	1.418	1.024	0.732
Finland (FN)	2.148	1.522	1.130	0.819	3.009	2.142	1.128	0.812
France (FR)	1.545	1.096	1.204	0.898	2.162	1.559	1.204	0.894
Germany (BD)	1.503	1.069	1.048	0.777	2.084	1.503	1.046	0.771
Greece (GR)	1.953	1.373	2.162	1.575	2.613	1.874	2.177	1.583
Hungary (HN)	2.131	1.475	1.414	1.108	2.901	2.052	1.418	1.109
Ireland (IR)	1.578	1.105	0.971	0.785	2.177	1.546	0.969	0.785
Italy (IT)	1.617	1.131	1.430	1.071	2.267	1.609	1.431	1.072
Netherlands (NL)	1.532	1.067	1.011	0.737	2.130	1.502	1.009	0.733
Norway (NW)	1.894	1.305	1.171	0.851	2.660	1.836	1.169	0.847
Poland (PO)	1.961	1.406	1.294	0.923	2.630	1.957	1.305	0.930
Portugal (PT)	1.372	0.957	1.436	1.046	1.861	1.300	1.435	1.046
Russia (RS)	3.188	1.930	1.154	0.877	4.495	2.756	1.179	0.906
Spain (ES)	1.590	1.123	1.646	1.252	2.200	1.590	1.643	1.247
Sweden (SD)	1.951	1.380	1.277	0.921	2.725	1.945	1.273	0.914
Switzerland (SW)	1.219	0.877	0.936	0.707	1.695	1.236	0.934	0.704
Turkey (TK)	3.013	2.071	2.078	1.392	4.186	2.884	2.085	1.378
United Kingdom (UK)	1.405	0.977	0.882	0.638	1.990	1.394	0.880	0.637
Total	1.858	1.241	1.319	0.942	2.572	1.737	1.322	0.941

References

Akaike, H. (1974). A new look at statistical-model identification. *IEEE Transactions on Automatic Control*, *AC19*(6), 716–723.

Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, *15*(4), 1137–1187.

Bae, G. I., Kim, W. C., & Mulvey, J. M. (2014). Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research*, *234*(2), 450–458.

Basalto, N., Bellotti, R., De Carlo, F., Facchi, P., Pantaleo, E., & Pascazio, S. (2007). Hausdorff clustering of financial time series. *Physica A-Statistical Mechanics and Its Applications*, *379*(2), 635–644.

Bastos, J. A., & Caiado, J. (2014). Clustering financial time series with variance ratio statistics. *Quantitative Finance*, *14*(12), 2121–2133.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *Annals of Mathematical Statistics*, *41*(1), 164–171.

Bec, F., Rahbek, A., & Shephard, N. (2008). The ACR model: A multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics and Statistics*, *70*(5), 583–618.

Bekaert, G., Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1998). Distributional characteristics of emerging market returns and asset allocation. *Journal of Portfolio Management*, *24*(2), 102–116.

Clogg, C. (1995). Latent class models. In G. Arminger, C. Clogg, & M. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). New York: Plenum.

Costa, M., & De Angelis, L. (2010). Model selection in hidden Markov models: A simulation study. Quaderni di Dipartimento 7. Department of Statistics, University of Bologna.

Dacco, R., & Satchell, S. (1999). Why do regime-switching models forecast so badly? *Journal of Forecasting*, *18*(1), 1–16.

De Angelis, L., & Paas, L. J. (2013). A dynamic analysis of stock markets using a hidden Markov model. *Journal of Applied Statistics*, *40*(8), 1682–1700.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B-Methodological*, *39*(1), 1–38.

Dias, J. (2007). Model selection criteria for model-based clustering of categorical time series data. A Monte Carlo study. In R. Decker, & H.-J. Lenz (Eds.), *Advances in data analysis* (pp. 23–30). Berlin: Springer.

Dias, J., & Vermunt, J. (2007). Latent class modeling of website users' search patterns: Implications for online market segmentation. *Journal of Retailing and Consumer Services*, *14*(6), 359–368.

Dias, J. G., & Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing*, *14*(4), 323–332.

Elias, R., Wahab, M., & Fang, L. (2014). A comparison of regime-switching temperature modeling approaches for applications in weather derivatives. *European Journal of Operational Research*, *232*(3), 549–560.

Engel, C. (1994). Can the Markov switching model forecast exchange rates? *Journal of International Economics*, *36*(1-2), 151–165.

Esling, P., & Agon, C. (2012). Time-series data mining. *ACM Computing Surveys*, *45*(1), 12:1–12:34.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, *38*(1), 34–105.

Fu, F. J. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics*, *91*(1), 24–37.

Fu, J., Wei, J., & Yang, H. (2014). Portfolio optimization in a regime-switching market with derivatives. *European Journal of Operational Research*, *233*(1), 184–192.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*(5), 1779–1801.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27–62.

Guerry, M.-A. (2011). Hidden heterogeneity in manpower systems: A Markov-switching model approach. *European Journal of Operational Research*, *210*(1), 106–113.

Guidolin, M. (2011). Markov switching models in empirical finance. In D. M. Drukker (Ed.), *Advances in Econometrics: Vol. 27-2. Missing data methods: Time-series methods and applications* (pp. 1–86). Bingley: Emerald Group Publishing Limited.

Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics & Control*, *31*(11), 3503–3544.

Hainaut, D. (2014). Impulse control of pension fund contributions, in a regime switching economy. *European Journal of Operational Research*, *239*(3), 810–819.

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

Hamilton, J. D. (1993). Estimation, inference and forecasting of time series subject to changes in regime. In G. Maddala, C. Rao, & H. D. Vinod (Eds.), *Handbook of statistics 11: Econometrics* (pp. 231–260). Amsterdam: Elsevier.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, *64*(1-2), 307–333.

Harvey, C. R. (1995). Predictable risk and returns in emerging markets. *Review of Financial Studies*, *8*(3), 773–816.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, *109*(4), 673–748.

Kakizawa, Y., Shumway, R., & Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, *93*, 328–340.

Lange, T., & Rahbek, A. (2009). An introduction to regime switching time series models. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, & T. V. Mikosch (Eds.), *Handbook of financial time series* (pp. 871–887). Berlin Heidelberg: Springer-Verlag.

Langrock, R., MacDonald, I. L., & Zucchini, W. (2012). Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models. *Journal of Empirical Finance*, *19*(1), 147–161.

Liao, T. W. (2005). Clustering of time series data: A survey. *Pattern Recognition*, *38*(11), 1857–1874.

Lundblad, C. (2007). The risk return tradeoff in the long run: 1836-2003. *Journal of Financial Economics*, *85*(1), 123–150.

Lux, T., & Morales-Arias, L. (2010). Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. *Computational Statistics & Data Analysis*, *54*(11), 2676–2692.

Lystig, T. C., & Hughes, J. P. (2002). Exact computation of the observed information matrix for hidden Markov models. *Journal of Computational and Graphical Statistics*, *11*(3), 678–689.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*(4), 394–419.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *European Physical Journal B*, *11*, 193–197.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: John Wiley & Sons.

Pagan, A. R., & Sossounov, K. A. (2003). A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics*, *18*(1), 23–46.

Peiró, A. (1999). Skewness in financial returns. *Journal of Banking & Finance*, *23*(6), 847–862.

van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, *20*, 213–247.

Praetz, P. D. (1972). Distribution of share price changes. *Journal of Business*, *45*(1), 49–55.

Rossi, A., & Gallo, G. M. (2006). Volatility estimation via hidden Markov models. *Journal of Empirical Finance*, *13*(2), 203–230.

Schmittmann, V. D., Visser, I., & Raijmakers, M. E. J. (2006). Multiple learning modes in the development of performance on a rule-based category-learning task. *Neuropsychologia*, *44*(11), 2079–2091.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

Simkowitz, M. A., & Beedles, W. L. (1980). Asymmetric stable distributed security returns. *Journal of the American Statistical Association*, *75*(370), 306–312.

Singleton, J. C., & Wingender, J. (1986). Skewness persistence in common-stock returns. *Journal of Financial and Quantitative Analysis*, *21*(3), 335–341.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*, 213–239.

Vermunt, J. K. (2007). A hierarchical mixture model for clustering three-way data sets. *Computational Statistics & Data Analysis*, *51*(11), 5368–5376.

Vermunt, J. K., Langeheine, R., & Bockenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, *24*(2), 179–207.

Wilfling, B. (2009). Volatility regime-switching in European exchange rates prior to monetary unification. *Journal of International Money and Finance*, *28*(2), 240–270.

Zhou, Z.-J., Hu, C.-H., Xu, D.-L., Chen, M.-Y., & Zhou, D.-H. (2010). A model for real-time failure prognosis based on hidden Markov model and belief rule base. *European Journal of Operational Research*, *207*(1), 269–283.