

Latent Class Models of Time Series Data: An Entropic-Based Uncertainty Measure

José G. Dias

Abstract Latent class modeling has proven to be a powerful tool for identifying regimes in time series. Here, we focus on the classification uncertainty in latent class modeling of time series data with emphasis on entropy-based measures of uncertainty. Results are illustrated with an example.

1 Introduction

Latent class models are a powerful tool for capturing unobserved heterogeneity in a wide range of social and behavioral science data (see, for example, [McLachlan and Peel 2000](#) or [Ramos et al. 2011](#)).

Let \mathbf{y}_i denote a T -dimensional observation and $D = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ a sample of size n . Each data point is assumed to be a realization of the random variable \mathbf{Y} coming from an S -component mixture probability density function (p.d.f.)

$$f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{w=1}^S \pi_w f_w(\mathbf{y}_i; \boldsymbol{\theta}_w), \quad (1)$$

where π_w are positive mixing proportions that sum to one, $\boldsymbol{\theta}_w$ are the parameters defining the conditional distribution $f_w(\mathbf{y}_i; \boldsymbol{\theta}_w)$ for latent class w , and $\boldsymbol{\varphi} = \{\pi_1, \dots, \pi_{S-1}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_S\}$. Note that $\pi_S = 1 - \sum_{w=1}^{S-1} \pi_w$. The log-likelihood function for an LC model – assuming i.i.d. observations – has the form $\ell(\boldsymbol{\varphi}; \mathbf{y}) =$

J.G. Dias (✉)

UNIDE, ISCTE – Instituto Universitário de Lisboa, BRU – Business Research Centre, Lisbon, Portugal

Edifício ISCTE, Av. das Forças Armadas, 1649–026 Lisboa, Portugal
e-mail: jose.dias@iscte.pt

$\sum_{i=1}^n \log f(\mathbf{y}_i; \boldsymbol{\varphi})$, which is straightforward to maximize (yielding the MLE - maximum likelihood estimator) by the EM algorithm (Dempster et al. 1977).

From parameter estimates of the LC model one can derive the posterior probability that an observation belongs to a certain latent class conditional on its response pattern. From the ML parameter estimates, Bayes' theorem gives the posterior probability that observation i was generated by latent class w :

$$\hat{\alpha}_{iw} = \frac{\hat{\pi}_w f_w(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_w)}{\sum_{v=1}^S \hat{\pi}_v f_v(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_v)}, \quad (2)$$

where $\hat{\pi}_w$ and $\hat{\boldsymbol{\theta}}_w$ are the ML estimates of π_w and $\boldsymbol{\theta}_w$, respectively.

The $\hat{\alpha}_{iw}$ values define a soft partitioning/clustering of the data set, since $\sum_{w=1}^S \hat{\alpha}_{iw} = 1$ and $\hat{\alpha}_{iw} \in [0, 1]$. Let c_i represent the true cluster membership (the missing data) of observation i . Then, the optimal Bayes rule assigning observation i to the class with maximum posterior probability can be defined as follows:

$$\hat{c}_i = \arg \max_w \hat{\alpha}_{iw}, i = 1, \dots, n, \quad (3)$$

where \hat{c}_i is the estimate of c_i .

In this paper, we address the following question: How can we measure the level of uncertainty in the mapping from the $[0, 1]$ soft partition to the $\{0,1\}$ hard partition obtained by applying the optimal Bayes rule in time series analysis?

The paper is organized as follows: Sect. 2 presents the full mixture hidden Markov model used to obtain the posterior probabilities of the regimes; Sect. 3 introduces a measure of classification uncertainty; Sect. 4 studies its behavior using a synthetic example; and Sect. 5 illustrates the application of the procedure to a panel data set of twenty European stock markets. Section 6 gives concluding remarks.

2 Latent Class Modeling of Time Series Data

The mixture of hidden Markov models (MHMM-S) (Dias et al. 2008; Ramos et al. 2011) is defined by the density:

$$f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{w_i=1}^S \pi_w \sum_{z_{i1}=1}^K \cdots \sum_{z_{iT}=1}^K f(z_{i1}|w_i) \prod_{t=2}^T f(z_{it}|z_{i,t-1}, w_i) \prod_{t=1}^T f(y_{it}|z_{it}), \quad (4)$$

where the conditional distribution within each latent class is given by a hidden Markov model with K regimes. $\boldsymbol{\varphi}$ is the vector containing all parameters in the model. Thus, we assume that within latent class w the sequence $\{z_{i1}, \dots, z_{iT}\}$ is in agreement with a first-order Markov chain. Moreover, we assume that the observed value y_{it} at a particular time point depends only on the regime at this time point;

i.e, conditionally on the regime z_{it} , the response y_{it} is independent of other time points, which is often referred to as the local independence assumption. As far as the first-order Markov assumption for the latent regime switching conditional on latent class membership w is concerned, it is important to note that this assumption is not as restrictive as one may initially think. It does clearly not imply a first-order Markov structure for the responses y_{it} . The standard hidden Markov model (HMM) (Baum et al. 1970) is a special case of the MHMM-S that is obtained by eliminating the time-constant latent variable w from the model, that is, by assuming that there is no unobserved heterogeneity across time series.

The characterization of the MHMM is provided by:

- π_w is the prior probability of belonging to the latent class w ;
- $f(z_{i1}|w_i)$ is the initial-regime probability; that is, the probability of having a particular initial regime conditional on belonging to latent class w with multinomial parameter $\lambda_{kw} = P(Z_{i1} = k|W_i = w)$;
- $f(z_{it}|z_{i,t-1}, w_i)$ is a latent transition probability; that is, the probability of being in a particular regime at time point t conditional on the regime at time point $t - 1$ and latent class membership; assuming a time-homogeneous transition process, we have $p_{jkw} = P(Z_{it} = k|Z_{i,t-1} = j, W_i = w)$ as the relevant multinomial parameter. Note that the MHMM-S allows that each latent class has its specific transition or regime-switching dynamics, whereas in a standard HMM it is assumed that all cases have the same transition probabilities;
- $f(y_{it}|z_{it})$, the probability density of having a particular value y_{it} , conditional on the regime occupied at time point t , is assumed to have the form of a univariate normal (or Gaussian) density function. This distribution is characterized by the parameter vector (μ_k, σ_k^2) containing the mean (μ_k) and variance (σ_k^2) for regime k . Note that these parameters are assumed invariant across latent classes, an assumption that may, however, be relaxed.

Since $f(\mathbf{y}_i; \boldsymbol{\varphi})$, defined by Eq. (4), is a mixture of densities across S latent classes and K regimes, it defines a flexible Gaussian mixture model that can accommodate deviations from normality in terms of skewness and kurtosis. For example, for two regimes ($K = 2$), the MHMM-S has $4S + 3$ free parameters to be estimated, including $S - 1$ class sizes, S initial-regime probabilities, $2S$ transition probabilities, 2 conditional means, and 2 conditional variances.

Maximum likelihood (ML) estimation of the parameters of the MHMM-S 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 tackled by the Expectation-Maximization (EM) algorithm (Dempster et al. 1977). The E step computes the joint conditional distribution of the $T + 1$ latent variables given the 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 the estimated densities of the latent variables as weights. Since the EM algorithm requires the computation of $S \cdot 2^T$ entries in the E step, which makes this algorithm impractical or even impossible to apply with more than a few time points. However, for hidden Markov models, a special variant of the EM algorithm was proposed that

is usually referred to as the forward-backward or Baum-Welch algorithm (Baum et al. 1970). The Baum-Welch algorithm circumvents the computation of this joint posterior distribution making use of the conditional independencies implied by the model.

An important modeling issue is the setting of S and K , the number of latent classes and regimes needed to capture the unobserved heterogeneity across time series. The selection of S and K is typically based on information statistics such as the Bayesian Information Criterion (BIC) (Schwarz 1978) defined as:

$$BIC_{S,K} = -2\ell_{S,K}(\hat{\boldsymbol{\phi}}; \mathbf{y}) + N_{S,K} \log n, \quad (5)$$

where $N_{S,K}$ is the number of free parameters of the model and n is the sample size.

3 An Entropic-Based Uncertainty Measure

Classification uncertainty can be measured by the posterior probabilities $\hat{\alpha}_{is}$. An aggregate measure of classification uncertainty is the entropy. For LC models, the entropy is

$$EN(\boldsymbol{\alpha}) = - \sum_{i=1}^n \sum_{s=1}^S \alpha_{is} \log \alpha_{is}. \quad (6)$$

Its normalized version has been used as a model selection criterion, indicating the level of separation of latent classes (Dias and Vermunt 2006, 2008). The relative entropy that scales the entropy to the interval $[0,1]$ is given by

$$E = 1 - EN(\boldsymbol{\alpha}) / (n \log S). \quad (7)$$

For well-separated latent classes, $E \approx 1$; for ill-separated latent classes, $E \approx 0$. This provides a method for assessing the ‘‘fuzzyness’’ of the partition of the data under the hypothesized model. The ML estimate of $E - \hat{E}$ can be obtained using the MLE ($\hat{\alpha}_{is}$) of α_{is} in Eq. (7).

We propose an extension of the relative entropy to panel data, that we call Entropy Regime Classification Measure (ERCM). For the time series \mathbf{y}_i , the ERCM is given by

$$ERCM_i = 1 + \frac{1}{T \log K} \sum_{t=1}^T \sum_{k=1}^K \alpha_{itk} \log(\alpha_{itk}), \quad (8)$$

where $\alpha_{itk} = P(Z_{it} = k | \mathbf{y}_i)$ is the probability that time series i is in regime k at time t conditional on the observed data.

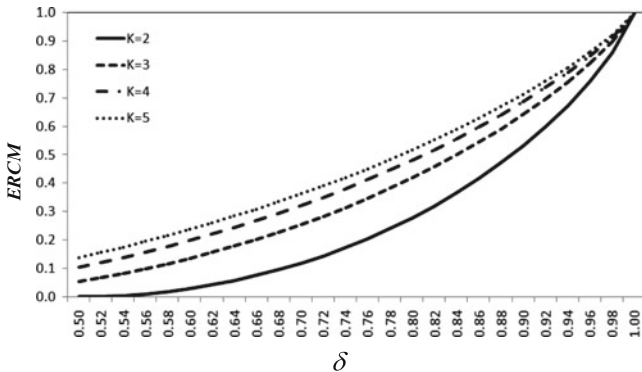


Fig. 1 ERCM as function of K and δ

4 A Synthetic Example

To understand the behavior of ERCM, let us assume that the posterior probabilities for each time series i and time $t - \alpha_{it} = (\alpha_{it1}, \dots, \alpha_{itK}) -$ are:

$$\alpha_{it} = \left(\delta, \frac{1 - \delta}{K - 1}, \dots, \frac{1 - \delta}{K - 1} \right), \tag{9}$$

i.e., it represents a regime with probability δ and the remaining $K - 1$ regimes with identical probability: $(1 - \delta)/(K - 1)$. Replacing vector α_{it} , in Eq. (8), we obtain the expression below for ERCM:

$$ERCM_{K,\delta} = 1 + \frac{1}{\log K} \left[\delta \log \delta + (1 - \delta) \log \left(\frac{1 - \delta}{K - 1} \right) \right].$$

Figure 1 depicts the ERCM as function of K and δ with values $\{2, 3, 4, 5\}$, and $[0.5, 1.0]$, respectively. For example, for $K = 2$ and $\delta = 0.5$, the classification uncertainty is maximum and then $ERCM = 0$.

As expected, the relation between δ and ERCM is nonlinear. For the same value of δ , increasing K leads to an increase of ERCM as the value of $(1 - \delta)/(K - 1)$ decreases with an increase in K , and it becomes clearer the ‘right’ regime. In the opposite case, with $\delta = 1$, then $ERCM = 1$, with $0 \cdot \log 0 = 0$.

5 Application

Modeling the dynamics of stock market returns has been an important challenge in modern financial econometrics. The statistics and dynamics of correctly specified distributions provide more accurate and detailed input for financial asset pricing

Table 1 Summary statistics

Stock market	Mean	Median	Std.			Jarque-Bera test	
			Deviation	Skewness	Kurtosis	statistics	p-value
Austria (OE)	-0.05	0.02	2.12	-0.21	6.62	564.41	0.000
Belgium (BG)	-0.05	0.04	1.89	-0.18	6.40	495.51	0.000
Czech Rep. (CZ)	-0.01	0.05	2.32	0.01	15.23	6,367.55	0.000
Denmark (DK)	-0.02	0.02	1.98	-0.23	8.46	1,275.01	0.000
Finland (FN)	-0.06	-0.04	2.14	0.08	6.12	413.55	0.000
France (FR)	-0.04	0.01	1.97	0.11	7.96	1,046.75	0.000
Germany (BD)	-0.02	0.08	1.84	0.52	11.77	3,323.29	0.000
Greece (GR)	-0.13	-0.02	2.26	0.00	5.42	247.02	0.000
Hungary (HN)	-0.05	0.00	2.81	0.04	9.21	1,640.73	0.000
Ireland (IR)	-0.11	-0.02	2.27	-0.46	6.65	599.30	0.000
Italy (IT)	-0.07	0.04	2.05	0.07	7.51	865.40	0.000
Netherlands (NL)	-0.05	0.02	2.02	-0.16	8.17	1,140.84	0.000
Norway (NW)	-0.03	0.12	2.72	-0.32	6.52	542.43	0.000
Poland (PO)	-0.04	0.04	2.39	-0.17	6.71	588.35	0.000
Portugal (PT)	-0.06	0.02	1.85	-0.05	9.03	1,545.13	0.000
Russia (RS)	-0.01	0.09	2.88	-0.17	15.06	6,197.63	0.000
Spain (ES)	-0.05	0.04	2.03	0.13	7.80	980.94	0.000
Sweden (SD)	-0.02	0.03	2.47	0.20	6.04	398.23	0.000
Switzerland (SW)	0.00	0.05	1.47	0.11	7.62	907.38	0.000
United Kingdom (UK)	-0.03	0.07	1.98	-0.04	8.61	1,338.27	0.000

and risk management. For example, investors buy or sell securities according to their expectation of the market regime. In addition, portfolio risk reduction might be achieved by procedures that take into account the synchronization of market regimes. Therefore, regime switching uncertainty is key in financial modeling.

The data set used in this article are daily closing prices from 4 July 2007 (the start of the subprime crisis) to 11 July 2011 for twenty European stock market indexes drawn from Datastream database and listed in Table 1. The series are expressed in US dollars. In total, we have 1,038 end-of-the-day observations per country. Let P_{it} be the observed daily closing price of market i on day t , $i = 1, \dots, 20$ and $t = 0, \dots, 1,037$. The daily rates of return are defined as the log-returns multiplied by 100: $y_{it} = 100 \times \log(P_{it}/P_{i,t-1})$, $t = 1, \dots, T$, with $T = 1,037$.

This period was a very harsh one for the European stock markets. Table 1 provides descriptive statistics of the time series, while Fig. 2 depicts the log-returns time series. It can be seen that the mean is not positive for all markets in this period, however only for three markets – Finland, Greece, and Ireland – the median is negative. Stock markets show, instead, very different patterns of dispersion (Fig. 2); the largest standard deviation is found in Russia, Hungary, and Norway, while the smallest is in Switzerland (1.47). Return rate distributions are diverse in terms of skewness and the kurtosis (which equals 0 for normal distributions) shows high positive values, indicating heavier tails and more peakness than the normal distribution. The Jarque-Bera test rejects the null hypothesis of normality for all twenty stock markets.

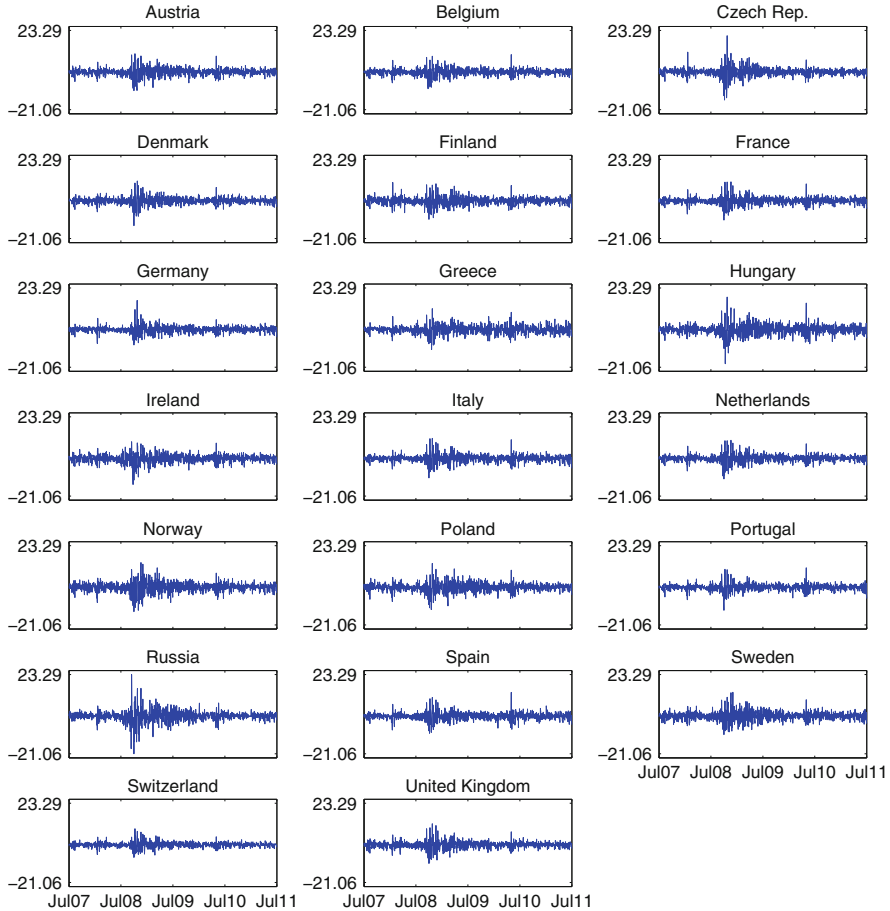


Fig. 2 Log-return time series for 20 European stock markets

Overall, these stock market features seem well suited to be modeled using MHMMs as we want to model simultaneously the 20 time series with typical cluster volatility. Given the traditional dichotomization of financial markets into “bull” and “bear” markets, we assume $K = 2$. We estimated models characterized by different number of latent classes ($S = 1, \dots, 5$). To minimize the impact of local maxima, 300 different starting values for the parameters are used for each model. The model with two latent classes ($S = 2$) yielded the lowest BIC value ($\ell_2(\hat{\phi}; \mathbf{y}) = -42,479.61$, $N_2 = 11$ and $BIC_2 = 84,992.17$).

Table 2 provides information on the two regimes that were identified ($K = 2$); that is, the average proportion of markets in regime k over time and the mean and variance of the returns in regime k . The result is in line with the common dichotomization of financial markets into “bull” and “bear” markets. Consistently, the reported means show that one of the regimes is associated with positive returns (bull market) and the other with negative returns (bear market). The probability of

Table 2 Estimated marginal probabilities of the regimes and within Gaussian parameters

	$P(Z)$		Return (mean)		Risk (variance)	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
Estimate	0.265	0.735	-0.306	0.045	12.813	2.044
Std. error	<i>0.026</i>	<i>0.026</i>	<i>0.050</i>	<i>0.012</i>	<i>0.305</i>	<i>0.031</i>

Table 3 Characterization of the switching regimes

	Latent class 1		Latent class 2	
	Regime 1	Regime 2	Regime 1	Regime 2
$P(Z W)$	0.243	0.757	0.409	0.591
	<i>0.026</i>	<i>0.026</i>	<i>0.055</i>	<i>0.055</i>
Transitions				
Regime 1	0.978	0.022	0.947	0.053
	<i>0.003</i>	<i>0.003</i>	<i>0.015</i>	<i>0.015</i>
Regime 2	0.007	0.993	0.037	0.963
	<i>0.001</i>	<i>0.001</i>	<i>0.011</i>	<i>0.011</i>
Sojourn time	44.84	133.33	18.76	26.74

Standard values are given in italics

being in the bear and bull regimes is 0.27 and 0.73, respectively. We would also like to emphasize that these results are coherent with the common acknowledgment of volatility asymmetry of financial markets. Volatility is likely to be higher when markets fall than when markets rise.

Table 3 reports the estimated probabilities of being in one of the regimes within each latent class. There is a clear distinction between these latent classes. Latent class 2 has the largest probability of being in bear regime (0.41). For latent class 1, this probability is 0.24. Moreover, Table 3 provides another key insight from our analysis. It gives the transition probabilities between the two regimes for both latent classes. First, notice that both latent classes show regime persistence. Once a stock market jumps into a regime, it is likely to remain within the same regime for a while, which is coherent with stylized facts in financial markets. Second, latent class 1 shows lower propensity to move from a bull regime to a bear regime (0.007) than latent class 2. Third, latent class 1 shows higher probability to jump from a bear to a bull regime than latent class 2.

The sojourn time is the expected number of days that a stock market stays in a given regime. For regime k and latent class w it can be obtained by $1/(1 - p_{kkw})$. As reported in Table 3, stock markets in latent class 2 stay the shortest number of days in both bear and bull markets, and consequently being the less stable group.

Table 4 summarizes the results related to the distribution of stock market across latent classes. From the posterior class membership probabilities, the probability of belonging to each of the latent classes conditional on the observed data, we found only two stock markets are more likely to belong to latent class two: Greece (1.00) and Hungary (0.99). For most of the stock markets the posterior probability is

Table 4 Estimated posterior probabilities and ERCM

Markets	Posterior probability			ERCM
	Latent class 1	Latent class 2	Modal	
Austria (OE)	1.00	0.00	1	0.876
Belgium (BG)	1.00	0.00	1	0.922
Czech Rep. (CZ)	1.00	0.00	1	0.910
Denmark (DK)	1.00	0.00	1	0.911
Finland (FN)	1.00	0.00	1	0.854
France (FR)	1.00	0.00	1	0.905
Germany (BD)	1.00	0.00	1	0.923
Greece (GR)	0.00	1.00	2	0.605
Hungary (HN)	0.01	0.99	2	0.672
Ireland (IR)	0.84	0.16	1	0.777
Italy (IT)	1.00	0.00	1	0.862
Netherlands (NL)	1.00	0.00	1	0.914
Norway (NW)	0.99	0.01	1	0.801
Poland (PO)	1.00	0.00	1	0.846
Portugal (PT)	1.00	0.00	1	0.888
Russia (RS)	1.00	0.00	1	0.893
Spain (ES)	1.00	0.00	1	0.852
Sweden (SD)	0.98	0.02	1	0.809
Switzerland (SW)	1.00	0.00	1	0.936
United Kingdom (UK)	1.00	0.00	1	0.892

precise (the probability of the most likely latent class is always one or very close to one), the exception being Ireland with probability 0.16 of belonging to latent class 2. By combining the classification information with the descriptive statistics in Table 1, latent class 1 tends to contain countries with lower volatility than latent class 2.

Based on the posterior probabilities from the estimated model, Table 4 also reports the estimate of the ERCM for each stock market. Above 0.9 we have Switzerland (0.936), Germany (0.923), Belgium (0.922), Netherlands (0.914), Denmark (0.911), Czech Republic (0.910), and France (0.905) as the least uncertain stock markets. On the other hand, Greece (0.605) and Hungary (0.672) are the most uncertain stock markets with ERCM below 0.7. The third most uncertain market is Ireland (0.777). Thus, the ERCM complements the values of posterior probabilities providing a more detailed indicator of stock market uncertainty. These results are consistent with financial market stylized facts.

6 Conclusions

This paper provides an extension to the MHMM model (Dias et al. 2008; Ramos et al. 2011) as a tool for measuring classification uncertainty in financial time series analysis. The proposed measure of uncertainty – Entropy Regime Classification

Measure (ERCM) – reveals the amount of uncertainty in a given time series. In the analysis of a sample of twenty stock markets for a period of 1,038 days, the best model was the one with two latent classes, with distinct types of regime switching. We conclude that the European market uncertainty in this period of analysis ranged from a minimum for Switzerland (ERCM = 0.936) and a maximum for Greece (ERCM = 0.605).

This approach should be further explored namely in applications where more than two regimes are needed. For instance, in modeling electricity prices it is standard to use at least three regimes as a result of abnormality in the markets and spike prices.

Acknowledgements The author would like to thank the Fundação para a Ciência e a Tecnologia (Portugal) for its financial support (PTDC/EGE-GES/103223/2008 and PEst-OE/EGE/UI0315/2011) and the three referees for their very valuable comments.

References

- 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, 164–171.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, 39, 1–38.
- Dias, J. G., & Vermunt, J. K. (2006). Bootstrap methods for measuring classification uncertainty in latent class analysis. In A. Rizzi & M. Vichi (Eds.), *COMPSTAT 2006: proceedings in computational statistics*, Rome (pp. 31–41). Heidelberg: Physica-Verlag.
- Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics*, 23(4), 643–659.
- Dias, J. G., Vermunt, J. K., & Ramos, S. B. (2008). Heterogeneous hidden Markov models. In P. Brito (Ed.), *COMPSTAT 2008: proceedings in computational statistics*, Porto (pp. 373–380). Heidelberg: Physica-Verlag.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Ramos, S. B., Vermunt, J. K., & Dias, J. G. (2011). When markets fall down: Are emerging markets all the same? *International Journal of Finance and Economics*, 16(4), 324–338.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.