Heterogeneous Hidden Markov Models

José G. Dias[1], Jeroen K. Vermunt[2] and Sofia Ramos[3]

[1] Department of Quantitative methods, ISCTE – Higher Institute of Social
   Sciences and Business Studies, Edifício ISCTE, Av. das Forças Armadas,
   1649–026 Lisboa, Portugal
   `jose.dias@iscte.pt`
[2] Department of Methodology and Statistics, Tilburg University, P.O. Box 90153,
   5000 LE Tilburg, The Netherlands,
   `J.K.Vermunt@uvt.nl`
[3] Department of Finance, ISCTE – Higher Institute of Social Sciences and
   Business Studies, Edifício ISCTE, Av. das Forças Armadas, 1649–026 Lisboa,
   Portugal
   `sofia.ramos@iscte.pt`

Abstract. Heterogeneous hidden Markov models (HHMMs) are models with
time-constant and time-varying discrete latent variables that capture unobserved
heterogeneity between and within clusters, respectively. We apply the HHMMs in
modeling financial return indexes from seven markets. The return-risk patterns of
the encountered latent states that correspond to well-known bear and bull market
states.

1 Introduction

Latent class or finite mixture modeling has proven to be a powerful tool for
analyzing unobserved heterogeneity in a wide range of social and behavioral
science data (see, for example, McLachlan and Peel (2000)). We introduce a
latent class model for time series analysis that takes into account unobserved
heterogeneity by means of time-constant and time-varying discrete latent
variables.

Here, this methodology is used to model the dynamics of the returns of
seven stock market indexes. 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 is almost always ignored in finance research. Because we selected
a heterogeneous sample of countries including both developed and emerging
countries from the American region, we expect that heterogeneity in market
returns due to country idiosyncrasies will show up in the results. For instance,
emerging market return distributions show larger deviations from normality;
i.e., are more skewed and have fat tails (Harvey, 1995).

The paper is organized as follows: Section 2 presents the full mixture hidden Markov model; Section 3 describes the seven stock market time series that are used throughout this paper. Section 4 reports HHMM estimates. The paper concludes with a summary of the main findings.

2   The heterogeneous hidden Markov model (HHMM)

We model simultaneously the time series of n stock markets. Let y_{it} represent the response of observation (stock market) i at time point t, where $i \in 1, \ldots, n$, $t \in 1, \ldots, T$, and $y_{it} \in \Re$. In addition to the observed "response" variable y_{it}, the HHMM contains two different latent variables: a time-constant discrete latent variable and a time-varying discrete latent variable. The former, which is denoted by $w \in \{1, ..., S\}$, is used to capture the unobserved heterogeneity across stock markets; that is, stock markets are clustered based on differences in their dynamics. We will refer to a model with S clusters as HHMM-S. The two-state time-varying latent variable is denoted by $z_t \in \{1, 2\}$. Changes between the two states or regimes between adjacent time points are assumed to be in agreement with a first-order Markov or first-order autocorrelation structure.

Let $f(\mathbf{y}_i; \varphi)$ be the (probability) density function associated with the index return rates of stock market i. The HHMM-S defines the following parametric model for this density:[1]

$$f(\mathbf{y}_i; \varphi) = \sum_{w=1}^{S} \sum_{z_1=1}^{2} \cdots \sum_{z_T=1}^{2} f(w) f(z_1|w) \prod_{t=2}^{T} f(z_t|z_{t-1}, w) \prod_{t=1}^{T} f(y_{it}|z_t). \quad (1)$$

As in any mixture model, the observed data density $f(\mathbf{y}_i; \varphi)$ is obtained by marginalizing over the latent variables. Because in our model these are discrete variables, this simply involves the computation of a weighted average of class-specific probability densities where the (prior) class membership probabilities or mixture proportions serve as weights (McLachlan and Peel, 2000). We assume that within cluster w the sequence $\{z_1, \ldots, z_T\}$ is in agreement with a first-order Markov chain. Moreover, we assume that the observed return at a particular time point depends only on the regime at this time point; i.e, conditionally on the latent state z_t, the response y_{it} is independent of returns at 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 cluster 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 or hidden Markov model (Baum et al., 1970) is the special

---

[1] For a detailed presentation of the model specification, we refer to Dias et al. (2007).

case of the HRSM-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.

The characterization of the HHMM is provided by:

- $f(w)$ is the prior probability of belonging to a particular latent class or cluster w with multinomial parameter $\pi_w = P(W = w)$;
- $f(z_1|w)$ is the initial-regime probability; that is, the probability of having a particular initial regime conditional on belonging to latent class w with Bernoulli parameter $\lambda_{kw} = P(Z_1 = k|W = w)$;
- $f(z_t|z_{t-1}, w)$ 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 class membership; assuming a time-homogeneous transition process, we have $p_{jkw} = P(Z_t = k|Z_{t-1} = j, W = w)$ as the relevant Bernoulli parameter. In other words, within cluster w one has the transition probability matrix

$$\mathbf{P}_w = \begin{pmatrix} p_{11w} & p_{12w} \\ p_{21w} & p_{22w} \end{pmatrix},$$

  with $p_{12w} = 1 - p_{11w}$ and $p_{22w} = 1 - p_{21w}$. Note that the HHMM-S allows that each cluster has its specific transition or regime-switching dynamics, whereas in a standard HMM it is assumed that all cases have the same transition probabilities.
- $f(y_{it}|z_t)$, 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, is assumed to have the form of a univariate normal (or Gaussian) density function. This distribution is characterized by the parameter vector $\theta_k = (\mu_k, \sigma_k^2)$ containing the mean (μ_k) and variance (σ_k^2) for regime k. Note that these parameters are assumed to be equal across clusters, an assumption that may, however, be relaxed.

Since $f(\mathbf{y}_i; \varphi)$, defined by Equation (1), is a mixture of densities across clusters w and regimes, it defines a flexible Gaussian mixture model that can accommodate deviations of normality in terms of skewness and kurtosis. The two-state HRSM-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 variances.

Maximum likelihood (ML) estimation of the parameters of the HHMM-S involves maximizing the log-likelihood function: $\ell(\varphi; \mathbf{y}) = \sum_{i=1}^n \log f(\mathbf{y}_i; \varphi)$, a problem that can be solved by means of the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). In the E step, we compute 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
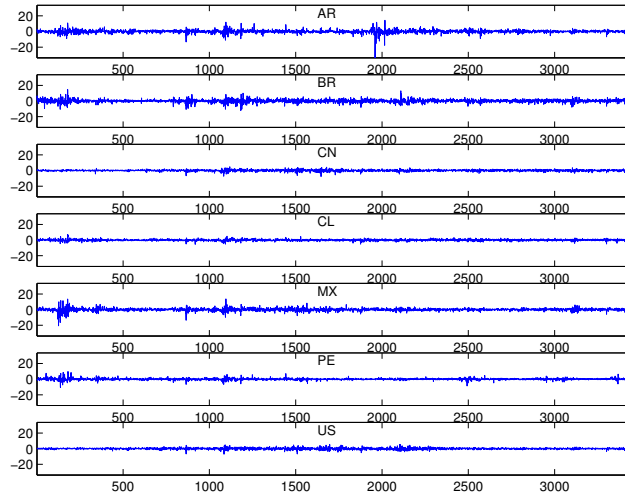
Fig. 1. Time series of index rates for seven American region stock markets

latent variables as weights. Since the EM algorithm requires us to compute and store the $S \cdot 2^T$ entries in the E step this 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 has been 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 selection of the value of S, the number of clusters needed to capture the unobserved heterogeneity across stock markets. The selection of S is typically based on information statistics such as the Bayesian Information Criterion (BIC) of Schwarz (Schwarz, 1978). In our application we select S that minimizes the BIC value defined as:

$$BIC_S = -2\ell_S(\hat{\varphi}; \mathbf{y}) + N_S \log n, \tag{2}$$

where N_S is the number of free parameters of the model concerned and n is the sample size.

3   Data set

The data set used in this article are daily closing prices from 4 July 1994 to 27 September 2007 for seven stock market indexes from the American region drawn from Datastream database and listed in Table 1. The series are denominated in US dollars. In total, we have 3454 end-of-the-day observations

Table 1. Summary statistics

| Stock market | Mean | Median | Std. Deviation | Skewness | Kurtosis | Jarque-Bera test | |
						statistics	p-value
Argentina (AR)	0.001	0.031	1.940	-1.756	34.648	173761.38	0.000
Brazil (BR)	0.055	0.077	1.961	-0.231	4.944	3527.58	0.000
Canada (CN)	0.054	0.096	0.997	-0.691	4.706	3442.42	0.000
Chile (CL)	0.027	0.000	1.024	-0.150	3.212	1487.90	0.000
Mexico (MX)	0.035	0.081	1.737	-0.807	16.347	38647.91	0.000
Peru (PE)	0.043	0.029	1.144	0.114	12.709	23139.40	0.000
United States (US)	0.038	0.042	1.040	-0.147	4.027	2332.37	0.000

per country. Let P_{it} be the observed daily closing price of market i on day t, $i = 1, ...n$ and $t = 0, ..., T$. The daily rates of return are defined as the percentage rate of return by $y_{it} = 100 \times \log(P_{it}/P_{i,t-1})$, $t = 1, ..., T$, with $T = 3454$.

Table 1 provides descriptive statistics of the time series, while Figure 1 depicts the full time series. The sample period includes periods of market instability as the Mexican crisis of 1994, the 1999 Brazilian crisis, the Argentina crises in 2001-2002, and the global stock market downturn of the 2001 Internet bubble. As can be seen, the mean return rates are all positive and close to zero. This is confirmed by the reported medians. Stock markets show very diverse patterns of dispersion, where the largest standard deviations are found in Brazil and Argentina and the smallest dispersion in Canada, Chile and the United States. Higher standard deviations are typical for emerging markets, known for their high risk. All stock market distributions of return rates are negative skewed and the kurtosis (which equals 0 for normal distributions) shows values above 0, indicating heavier tails and more peakness than the normal. The Jarque-Bera test rejects the null hypothesis of normality for each of the seven stock markets. Overall, market features seem well-suitable to apply the mixture hidden Markov model.

4   Results

This section reports the results obtained when applying the HHMM-S described before to the seven stock markets. We estimated models using different values for S ($S = 1, \ldots, 8$), where 300 different starting values were used to avoid local maxima. A solution with three latent classes ($S = 3$) yielded the lowest BIC value (log-likelihood = -38482.3136; number of free parameters = 15, and BIC = 76993.8158).

Table 2 summarizes the results related to the distribution of stock market across latent classes which gives the size of each cluster. The estimated prior class membership probability is somewhat larger for Class 1 (0.542). From the posterior class membership probabilities, the probability of belonging to each of the clusters conditional on the observed data (Table 2), we have four

Table 2. Estimated prior probabilities, posterior probabilities, and modal classes for the HHMM-3

Stock market	Latent class 1	Latent Class 2	Latent Class 3	Modal class
Prior probabilities	0.542	0.292	0.167	
Posterior probabilities				
Argentina (AR)	0.000	1.000	0.000	2
Brazil (BR)	0.000	0.000	1.000	3
Canada (CN)	1.000	0.000	0.000	1
Chile (CL)	1.000	0.000	0.000	1
Mexico (MX)	0.000	1.000	0.000	2
Peru (PE)	1.000	0.000	0.000	1
United States (US)	1.000	0.000	0.000	1

countries assigned to Class 1 (Canada, Chile, Peru and United States), two countries to Class 2 (Argentina and Mexico) and the remaining one – Brazil – to Class 3. Based on this classification one clearly has to reject the hypothesis that stock markets can clustered regionally. With the exceptions of Canada and the USA, and Peru and Chile, which share the same class, neighbor countries tend to be allocated to different classes. Notice that from the posterior probabilities the modal allocation into classes is precise (the probability of the most likely class is always one). Class 1 contains two developed countries Canada and the USA and two emerging markets. By combining the classification information with the descriptive statistics in Table 1, we conclude that Class 1 contains the countries with the lowest volatilities.

Table 3 provides information on the two regimes that were identified; that is, the average proportion of markets in regime k over time and the mean and variance of the return 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 negative returns (bear market). The probability of being in the bear and bull regimes is 0.26 and 0.74, 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 4 reports the estimated probabilities of being in one of the regimes for each latent class. There is a clear distinction between classes. Class 1 has the largest probability of being in bull regime (0.89). For Class 2 this probability becomes 0.63. In case of Brazil (Class 3) is more likely to be in

Table 3. 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	0.7438	0.2562	0.0884	-0.1251	0.7079	6.5757
Std. error	(0.0603)	(0.0603)	(0.0048)	(0.0238)	(0.0123)	(0.1715)

Table 4. Characterization of the switching regimes

Transitions	Latent class 1		Latent Class 2		Latent Class 3	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
$P(Z\|W)$	0.8903	0.1097	0.6293	0.3707	0.4683	0.5317
	(0.0120)	(0.0120)	(0.0212)	(0.0212)	(0.0299)	(0.0299)
Regime 1	0.988	0.012	0.933	0.067	0.895	0.105
	(0.0014)	(0.0976)	(0.0071)	(0.0071)	(0.0147)	(0.0147)
Regime 2	0.098	0.902	0.113	0.887	0.092	0.908
	(0.0114)	(0.0114)	(0.0125)	(0.0125)	(0.0148)	(0.0148)

a bear regime than in a bull regime during the period of analysis. Moreover, Table 4 provides another key result of our analysis. It gives the transition probabilities between the two regimes for each of the three latent classes. First, notice that all classes show regime persistence. Once a stock market jumps to a regime it is likely to continue on it for some period, which is coherent with stylized facts in financial markets. Second, Class 1 shows the lowest propensity to move from a bull regime to a bear regime. This propensity is higher for Class 2 and even higher for Class 3. Note that Class 2 and 3 were severely affected by crises during the sampled period. Third, Class 2 shows the highest probability of jumping from a bear to a bull regime.

Figure 2 shows the regime-switching dynamics of the countries within each of our three latent classes. It depicts the posterior probability of being in bull regime at period t, where the grey color identifies periods in which this probability is below 0.5 which corresponds to a higher likelihood of being in the bear state. The three clusters of countries have rather different pattern of regime switching. Class 1 is more regime persistent with short duration "bear" regimes that did not turn out to be endemic during the period of analysis. Both Class 2 and Class 3 are extremely dynamic and tend to move very fast between regimes. However, Class 3 tends to be more persistent in bear regimes.

5 Conclusion

The HHMM takes into account both time-constant unobserved heterogeneity between and hidden regimes within time series. In the analysis of a sample of seven stock markets providing observations for a period of 3454 days the best fitting model was the one with three latent classes. The three latent classes clearly distinguished three types of regime switching, which is coherent with many stylized facts in finance.
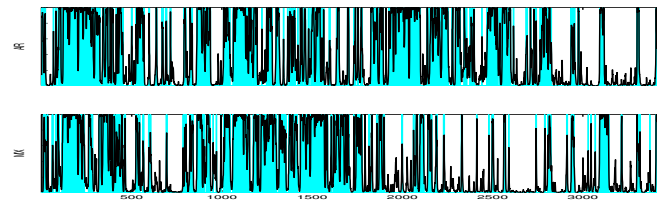
References

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.*
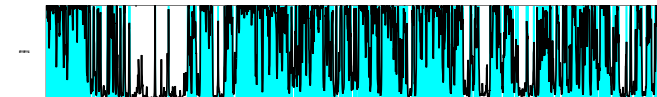
Fig. 2. Estimated posterior bull regime probability and modal regime

a. Latent Class 1
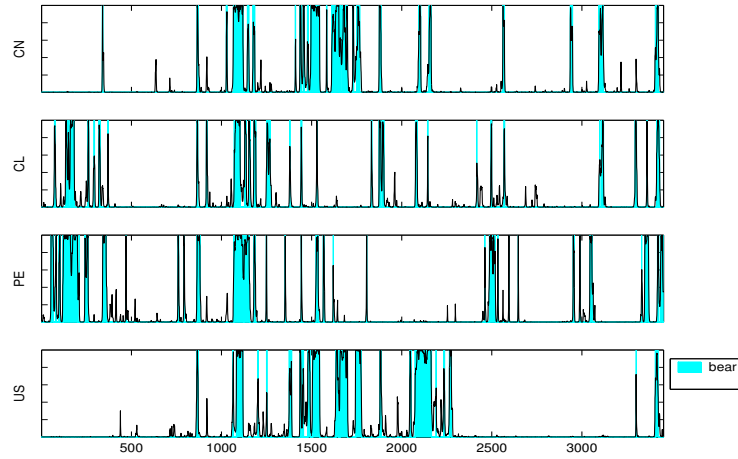
b. Latent class 2

c. Latent class 3

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., RAMOS, S. (2007): Analysis of heterogeneous financial time series using a mixture Gaussian hidden Markov model. *Working Paper, Tilburg University.*

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

McLACHLAN, G.J., PEEL, D. (2000): *Finite Mixture Models.* John Wiley & Sons, New York.

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