The dynamic analysis and prediction of stock markets through the latent Markov model

Luca De Angelis*

Department of Statistical Sciences, University of Bologna, Italy

Leonard J. Paas

Department of Marketing, Faculty of Economics and Business, VU University, Amsterdam, The Netherlands

Research Memorandum 2009-53

September 8th, 2009

* All correspondence to: Luca De Angelis, Department of Statistical Sciences, Alma Mater Studiorum University of Bologna, Via delle Belle Arti, 41, 40126, Bologna. Italy. E-mail: l.deangelis@unibo.it; Phone: +39-051-2098192.

The dynamic analysis and prediction of stock markets through the latent Markov model

Summary. In this paper we show how the latent Markov model can be used to define different conditions in the stock market, called market-regimes. Changes in regimes can be used to detect financial crises, pinpoint the end of a crisis and predict future developments in the stock market, to some degree. The model is applied to changes in monthly price indexes of the Italian and US stock market in the period from January 2000 to July 2009.

Keywords: Stock market pattern analysis; Regime-switching; Forecasting; Latent Markov model; Financial crises; Market stability periods.

1. Introduction

One of the pressing questions during a crisis concerns when the economic situation will improve. Another interesting question at such times concerns what will happen next. Since the beginning of the crisis that started in 2007, this question has also been raised many times. Below we utilize the latent Markov model (LMM) for recognizing the end of a crisis, using stock market price indexes. The LMM is also used for predicting what will happen the next month of a period of crisis.

The LMM classifies different months in a limited set of regimes on the basis of the change in stock market price indexes across these months. For example, a month characterized by a strong decline in the stock market price index may be allocated to the 'large value decrease market regime'. Contrarily, months defined by small changes may be considered as part of the stable market regime. Besides this, the model provides insight into the probability of switching from one regime to another across consecutive months. The analysis, presented in the current paper, is based on monthly changes in stock market price indexes in the period from January 2000 to July 2009, in two countries: USA which is the world leading economy and Italy which represents the fifth largest European stock market in terms of capitalization at the end of 2008 (according to the World Federation of Exchanges).

The application of the latent Markov model for the purposes outlined above is supported by the fact that financial markets are generally characterized by frequent changes in regimes. Different market regimes are characterized by different means and standard deviation values or, using the terminology of portfolio theory framework, by different risk-return profiles. For instance, during a financial crisis, the stock market experiences a strong negative mean return and the standard deviation, which is generally used as a proxy of risk, is large. During more stable phases, stock returns fluctuate around a constant mean and the standard deviation value is lower.

In Markowitz's framework and its developments, stock returns are assumed to be normally distributed. However, empirical analyses clearly show stock returns are characterized by asymmetry and larger kurtosis than the Gaussian distribution. LMM provides an effective solution to overcome these issues (Dias *et al.*, 2008), by modelling the regime changes using a mixture of normal distributions. The model pools in homogenous discrete non-observable classes (usually referred to as latent states) at every time point of the time series. Thus, LMM offers a contribution in model-based clustering of financial time series (Frühwirth-Schnatter and Kaufmann, 2008). The latent states are characterized by different profiles of mean return. Therefore, they can be interpreted as different regimes, which the stock market may experience. Morever, mixture models such as LMM provide the flexibility required for dealing with skewness and kurtosis and to capture almost any departure from normality (Dias *et al.*, 2008).

Dynamics of stock market developments can also be represented by the LMM. If the stock market price index development is subject to discrete changes in regimes, that is periods when the dynamic pattern of the series is markedly different, then it is useful to consider a nonlinear model which exploits the time path of the observed series to draw inference about a set of discrete latent states (Hamilton, 1989). For instance, the stock market may be in a fast growth, deep decline, or stable phases. The switches between these regimes may be modelled as a Markov process. The Markovian chain specification not only offers key insights into switching from one specific market phase to another. Using equations introduced in Paas *et al.* (2007), the LMM allows us to predict the future stock market dynamic pattern.

In sum, we apply the LMM for defining stock market regimes in periods of crisis and stability, to gain insight into switches between such regimes and to predict which regime will occur next. The paper thus contributes to our understanding of developments in stock market price indexes. This additional understanding will be applied to show how the end of a financial crisis can be pinpointed and how future stock market developments can be predicted.

As for the organization of the paper, Section 2 introduces the LMM that we apply. Section 3 discusses the analyzed data, the conducted analysis and results. The paper is concluded with a discussion in Section 4.

2. The Latent Markov Model

2.1. Model specification

The LMM, also known as the hidden Markov model or the regime-switching model (Hamilton and Raj, 2002), is a flexible and powerful tool for describing dynamics of a financial time series. Although this model was originally applied to categorical indicators (Van de Pol and Langeheine, 1990; Vermunt *et al.*, 1999; Bartolucci *et al.*, 2007), recent work exploits the potential of LMM for financial time series analysis of continuous variables such as the daily stock market return distribution (Rydén *et al.*, 1998; Dias *et al.*, 2008). Hamilton (1989) highlights how LMM offers a valid nonlinear alternative to linear representations such as the Box-Jenkins ARIMA specification which is the usual reference for time series analysis and forecasting. The main advantage that LMM has over ARIMA models is that LMM deals with regime-switching and structural breaks, which are common features in both economic and financial time series (Hamilton, 2008).

Denoting by z_t the return observation of a stock market index at time t (for $t = 1, \ldots, T$) the LMM analyzes $f(z)$, the probability density function of the return distribution of the market index over time, by means of a latent transition structure defined by a first-order

Markov process. For each time point t, the model defines one discrete latent variable denoted by y_t constituted by S latent classes (which are usually referred to as latent states). Thus, overall the LMM includes T latent variables.

The LMM is specified as

$$f(z) = \sum_{y_1=1}^{S} \sum_{y_2=1}^{S} \cdots \sum_{y_T=1}^{S} f(y_1,\ldots,y_T) f(z; y_1,\ldots,y_T) \tag{1}$$

where

$$f(y_1,\ldots,y_T) = f(y_1) \prod_{t=2}^{T} f(y_t \mid y_{t-1}) \tag{2}$$

and

$$f(z; y_1,\ldots,y_T) = \prod_{t=1}^{T} f(z_t \mid y_t) \tag{3}$$

From Equation (1), it can be noted that the model is a mixture with S^T latent classes (mixture components) and, as for every mixture model, $f(z)$ is obtained by marginalizing with respect to the latent variables. Since y's are discrete variables, Equation (1) is a weighted average of probability densities $f(z; y_1,\ldots,y_T)$, where the latent class membership probabilities (or prior probabilities) $f(y_1,\ldots,y_T)$ are used as weights (McLachlan and Peel, 2000). Equations (2) and (3) show the conditional independence assumption implied by the LMM, which allows the simplification of the density functions $f(y_1,\ldots,y_T)$ and $f(z; y_1,\ldots,y_T)$. Furthermore, Equation (2) implies an additional model assumption: $f(y_1,\ldots,y_T)$ follows a first-order Markov process. Thus, latent state y_t is associated with y_{t-1} and y_{t+1} only. Furthermore, $f(y_1)$ denotes the (latent) initial-state probability function. Equation (3) shows that the return observation at time t is independent of observations at other time points conditional on the latent state occupied at time t.

$f(y_t \mid y_{t-1})$ denotes the latent transition probability function which provides the probability of being in a particular latent state at time *t* conditional on the state occupied at the previous time point. Assuming a homogenous transition process with respect to time, we achieve the latent transition matrix where the generic element $p_{jk} = \Pr\text{ob}(y_t = k \mid y_{t-1} = j)$ denotes the probability of switching from latent state *j* at time *t* to latent state *k* at time *t* + 1, for *j*, *k* = 1, …, *S*.

2.2. Parameter estimation

The parameter estimation is achieved by maximizing the log-likelihood function (*LL*) through the Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977). However, the iterative procedure of the EM algorithm is often impractical to apply for estimating a LMM. In the E-step, it needs to compute and store S^T entries of the joint posterior latent distribution $f(y_1,...,y_T \mid z)$. This implies that the computational time increases exponentially with *T* and even a moderate time series length may prevent the convergence of the algorithm. Hence, we use a variant of the EM algorithm called the forward-backward or Baum-Welch algorithm (Baum *et al.*, 1970). This variant was extended by Paas *et al.* (2007) for application to data sets with multiple observed indicators and is implemented in the Latent GOLD 4.5 computer program (Vermunt and Magidson, 2007). The forward-backward algorithm exploits the conditional independence assumption of the LMM in order to compute the joint posterior latent distribution during the E-step. Basically, the E-step estimates the missing data, which in LMM are the unobserved state memberships. This is realized by computing the expected value of the log-likelihood function given the current parameter values and the observed data. The M-step uses standard maximum likelihood estimation methods for complete data to update the model parameters. The algorithm cycles between the E- and M-steps till a previously defined convergence criterion is reached. Refer to Paas et al. (2007) for a more detailed specification of the Baum-Welch algorithm that we applied.

2.3. Model selection and class membership

Model selection involves the choice of the number of latent states *S*, which in our framework represents the number of market regimes. This choice is based on the Consistent Akaike Information Criterion (CAIC):

$$CAIC = -2LL + [(\log T) + 1]NPar$$

where *T* denotes the sample size and *NPar* is the number of model parameters. This information criterion penalizes model complexity more than AIC (Bozdogan, 1987), which tends to overestimate the number of components in mixture models (Dias and Vermunt, 2007).

In our application of the LMM, monthly stock price indexes are the indicators z_t, for $t = 1, \ldots, T$. Each z_t is classified into one latent state according to the estimated posterior probabilities. That is, z_t is allocated to latent state *j* if $\hat{f}(y_t = j \mid z_t) > \hat{f}(y_t = k \mid z_t)$ for every $k = 1, \ldots, S$. This form of classification is called modal classification. Time-points with a similar development are more likely to be allocated to the same latent state than those time-points with highly divergent developments. For example, a month with a strong decline in the stock market price index is more likely to be allocated to the same state as another month with a strong decline than with a month with a positive development.

3. Empirical analysis using the latent Markov model

3.1. Data

We applied the LMM described in Section 2 for analyzing two stock market data sets: the US S&P-500 and the Italian FTSE-MIB market price indexes. The two time series consist of the monthly return distribution in percent. For example, if the stock market index on March 1st 2008 is 100 and on March 31st of 2008 it is 92 the value for March 2008 is -8%. The data cover the period from January 2000 to July 2009, thus, including $T = 115$ time-points.

As can be seen from Figure 1, the period we are considering is characterized by two market crises: the stock market crisis of 2000/01 and the crisis that started at the end of 2007. Figure 1 shows these crises are characterized by increased volatility, i.e., stronger fluctuations and rapid changes from positive to negative peaks. Between these two periods there is a more stable phase for both stock markets, from mid 2003 to the end of 2007.

As discussed in Section 1, regime switching is one of the main causes of the forecast accuracy failure of most traditional time series models such as ARCH-type or ARIMA models (see Hamilton and Susmel, 1994; Lamoureux and Lastrapes, 1993). Thus, since Goldfeld and Quandt's (1973) seminal work on regime-switching regression, time-varying parameter models based on the Markov process have much success (e.g., Turner *et al.*, 1989; Dueker, 1997; Francq and Zakoïan, 2001; Haas *et al.*, 2004).

INSERT FIGURE 1 ABOUT HERE

Table 1 displays the different values of mean returns and standard deviations of the two crisis periods and the stable regime. It is interesting to note that the mean return of the stable period is higher in absolute value than the mean return of the 2000/01 crisis. This feature underlines the fact that, after a strong downturn, stock markets tend to both recoup the losses and even create new wealth. Moreover, according to the standard deviation values, the three periods are characterized by different levels of variability. In particular, between the stable period and the crises, but also the values related to the two crises are quite different. The latter implies that each financial crisis presents its own peculiarity. Furthermore, the Jarque-Bera normality test results are significant for the entire data sets, implying a significant difference between the observed distribution and a normal distribution. However, when splitting the two time series according to crisis and stable periods, normality assumption is not rejected, according to the Jarque-Bera test. These results imply LMM may be a sound alternative to traditional financial econometric models since it accounts for both asymmetry and larger kurtosis than the normal distribution without needing to preliminary split the time series into different homogenous sub-periods.

INSERT TABLE 1 ABOUT HERE

3.2. Model estimation and class profiling

We estimate the LMM for 1 to 8 latent states (S = 1, …, 8). Table 2 shows the maximum log-likelihood function and CAIC values for the two considered stock market indexes. According to the CAIC criterion, the LMM with S = 5 latent states provides the best fit in both data sets. In our framework, these five latent states represent five different stock market regimes. According to the return means in each state, S&P-500 (US) shows three negative and two positive regimes whereas FTSE-MIB (Italy) two negative and three positive regimes, see Table 3. The profiles of the five market phases are determined by referring to the return means shown in Table 3. For example, latent state 1, in the S&P-500, has an average return of -9.07% and consists of 11.2% of the T=115 months that are analyzed.

INSERT TABLE 2 AND 3 ABOUT HERE

Table 3 shows the LMM can be used to define different regimes of the stock market. The return means are significantly different from state to state according to Wald tests (W = 345.1, df = 4, *p-value* < 0.001 and W = 285.6, df = 4, *p-value* < 0.001), which rejects the null hypothesis of equality between means. Furthermore, the dispersion within each latent state is relative low according to the similarity in standard deviation values in Table 3. Further relevant information is provided by the size of each latent state. It indicates the proportion of time points allocated to a particular latent state.

The Jarque-Bera tests in Table 3 shows all the latent states can be assumed as normally distributed, except for the first state for the S&P-500 (US). The departure from normality of this state is due to one extreme negative value (October 2008) which is particularly uncommon for the S&P-500 index. Through the LMM, the Italian data set is properly approximated by a mixture of five normal distributions, with different means and similar variances.

Figure 2 displays actual and estimated time series obtained by referring to the LMM with five latent states, for both the S&P-500 and the FTSE-MIB. The estimated series are plotted using the latent state return means. Figure 2 shows the LMM approximates the time series of indexes quite accurately. The model also detects the stable period between the two crises which is common to both stock markets. In Figure 2 it is represented by the straight lines between July 2003 and September 2007.

INSERT FIGURE 2 ABOUT HERE

3.3. Latent transition analysis

Tables 4 and 5 report the transition probability matrices estimated by the LMMs for the S&P-500 (US) and FTSE-MIB (Italy) data sets, respectively. In our framework, the transition probabilities define the stock market regime-switching. The values on the diagonals represent state persistence, i.e., the probabilities of remaining in a particular market regime. Both stock markets show one latent state with high persistence, which corresponds also to the modal state, state 4 for S&P-500 and state 3 for FTSE-MIB (p_{44} = 0.97 and p_{33} = 0.94, respectively). These latent states represent the stable market regime and, as it may be noted from Figure 2, these results underline that stock markets tend to remain in that regime for a long time (T =

$\left(1-p_{44}\right)^{-1} \approx 34$ months for S&P-500 and T = $\left(1-p_{33}\right)^{-1} \approx 18$ months for FTSE-MIB). The off-diagonal p_{jk} values indicate the probabilities of market regime-switching. For instance, it is quite likely that the Italian FTSE-MIB index switches from a period of fast growth to a very negative phase (p_{51} = 0.481), on the contrary, the same regime shift does not occur often on the US stock market (p_{51} = 0.004).

The probabilities in Tables 4 and 5 underline some differences between the two stock markets which can be attributed to their different levels of development (Demirguc-Kunt and Levine, 1996). For example, when the US S&P-500 declines (state 1) at time t, then at time t + 1 the market may continue in the negative phase (p_{11} = 0.2466 and p_{12} = 0.4970), or bounce to a positive regime (p_{15} = 0.2465). The other two states very rarely occur after state 1 (p_{13} = 0.0064 and p_{14} = 0.0034). Contrarily, the Italian FTSE-MIB remains in the most negative latent state 1 with a probability of p_{11} = 0.0962 only, it may attenuate the negative phase switching to latent state 2 (p_{12} = 0.4641), or jump to a positive period represented by state 4 (p_{14} = 0.3666). A switch from the most negative latent state 1 to the stable market regime is unlikely for the FTSE-MIB (p_{13} = 0.0695). A switch to the state with the highest returns and p_{15} = 0.0036) is even less likely to occur. Moreover, the Italian FTSE-MIB tends to change regimes more frequently than the S&P-500. Three probabilities on the FTSE-MIB matrix diagonal are less than 0.10. On the contrary, three p_{jk} when j = k for the S&P-500 index are above 0.24. When the S&P-500is in latent state 2 (-4.32% in Table 3) at time t, it is very likely that it switches to the more stable state 3 (-0.19% in Table 3) at time t + 1 (p_{23} = 0.977). On the contrary, FTSE-MIB may stay in latent state 2 (-3.75% in Table 3), p_{22} = 0.424, shift to a more negative phase (p_{21} = 0.269), or experience a positive regime at time t + 1 (p_{24} = 0.302). Thus, in the US S&P-500 the moderately negative state is very likely to be followed with a more or less neutral state, while in Italian FTSE-MIB the switch from the moderately negative state is in various directions. Note that in both countries the switch from the most negative state is in various directions. Overall, 13 transition probabilities are below 0.05 in the transition matrix of the S&P-500 index and only 10 for the FTSE-MIB, i.e. some regime switching, which are very unlikely in US S&P-500, are more probable to occur in the Italian FTSE-MIB. For this reason, the Italian stock market results more difficult to predict than the US market.

INSERT TABLES 4 AND 5 ABOUT HERE

3.4. Recognition of the stable market regime

In both stock markets, the latent state characterized by a moderate positive return mean is most common and has a high persistence probability. High persistence in this state denotes the stable market regime. In Section 3.2, we recognize latent states 4 and 3 as the stable regime states of the S&P-500 (US) and the FTSE-MIB (Italy), respectively.

In order to evaluate the model's capability to detect the stable period, we estimate the LMM with 5 latent states for shorter time series. The beginning of the stable regime, provided by the LMM applied to the entire time series, starts in July 2003. We aim to assess how many months of stability are required for detecting the end of a financial crisis using the 2000/01 crisis. For both S&P-500 and FTSE-MIB data we first estimate the model on data from January 2000 untill July 2003, then from January 2000 untill August 2003, etc. Detection of a stable period occurs when multiple consecutive months are allocated to the stable latent state.

We find that the LMM can detect the stable market regime promptly. Figure 3 compares the original time series with respect to the LMM estimates derived from the whole data sets and the estimates of a LMM with 5 latent states applied to the shorter time series. Obviously, the return means of the LMM estimates, based on the shorter time series, differ somewhat from the means of the overall LMM estimated time series. Nevertheless, latent state memberships derived from the shorter time series are almost the same as the LMM estimates achieved with the entire data sets. Figure 3 shows that we need four months for detecting the stable regime for S&P-500 data set and seven for FTSE-MIB (dotted lines). That is, with less than four respectively seven months the last few months are not allocated to the stable stock market regime latent state. Four months of stability are required for the US S&P-500 data and seven months are required for the Italian FTSE-MIB data for allocated the last months consecutively in the latent state representing the stable market regime. This is the first and only period of recovery from a crisis period in our data sets and we consider less than 50 monthly price index changes. This feature of LMM is useful for detecting when the financial crisis started in 2007 ends.

It is also interesting to note that the stable market regime in the US S&P-500, state 4 in Table 4, is almost only reached from state 5, i.e., p_{54} = 0.11 and p_{14} = p_{24} = p_{34} = 0.003. This feature underlines an interesting behaviour of the S&P-500, i.e., it tends to stabilize and consolidate after the positive regime. For the Italian FTSE-MIB, the stable period is often reached from states 4 and 1 (p_{43} = 0.21 and p_{13} = 0.07) and sometimes from state 5 (p_{53} = 0.01). The FTSE-MIB results are less predictable and its long stable regime may be misinterpreted. Figure 2 shows that the FTSE-MIB has two other shorter periods classified in the stable latent state 3 (from May 2000 to September 2000 and from December 2001 to April 2002). Such a limited number of months in the stable state 3 that are preceded and followed by months with a fluctuating stock market price index cannot be considered as a stable period. This period is much shorter the 2003 to end 2007 period of stability. Different to the FTSE-MIB is that all time points of the S&P-500 allocated to state 4 belong to the long period of stable regime from 2003 to end 2007. The latter suggests that allocation to state 4 in the S&P-500 is a strong indication for the end of a crisis in the US stock market.

INSERT FIGURE 3 ABOUT HERE

3.5. Predictive power of LMM

In Section 3.3, we reported the latent transition matrices for S&P-500 and FTSE-MIB indexes. In this section, we exploit the information provided by the transition probabilities for evaluating the forecasting accuracy of the LMM. For this we have to impose that transition probabilities do not change over time (Paas *et al.*, 2007). We check this model assumption by estimating the LMM with time-varying latent transition probabilities which, according to CAIC criterion, fits the data much worst, i.e., CAIC = 2674.10 and CAIC = 2753.85 for S&P-500 and FTSE-MIB data sets, respectively. These values of CAIC are much higher than for the models with fixed transition matrices (see Table 2), due to a large increase in the number of parameters to be estimated resulting from relaxing the assumption of fixed transition probabilities when T=115.

Tables 4 and 5 show that some regime switching can be predicted quite accurately, because their transition probabilities are high. For instance, the persistence of the stable regime of both indexes is highly predictable, as is the switching of S&P-500 index from latent state 2 to state 3. On the contrary, there are latent states for which at least three transition probabilities are above 0.10, which complicates prediction. For example, latent state 5 for S&P-500 and state 4 for FTSE-MIB have four transition probabilities higher than 0.10.

Forecasting accuracy of the LMM can also be assessed more precisely. In the LMM each regime switch has a specific probability to occur. Using these probabilities, we can determine the LMM prediction power by referring to one-step ahead forecasts (Paas et al., 2007). The forecasting results are summarized in Table 6. Here, we report the number of times the LMM is able to predict next the month market regime correctly, according to the three highest latent transition probabilities. Hence, column 1 reports the number of times that the LMM predicts the next market regime by referring to the most probable p_{jk} in the latent transition matrix, column 2 contains the amount of the times LMM forecasts correctly according to the second modal transition probability, and so on. For instance, the June 2009 observation for the S&P-500 has been classified into latent state 3 and the July 2009 observation into state 5. Since the transition probability of switching from state 3 to state 5 is p_{35} = 0.383 which is the second highest probability for latent state 3, following p_{34}, we reported it in column 2 of Table 6. The last column of table 6 provides the number of times that the model is unable to predict the next month regime by referring to the three most probable latent transition probabilities. It must be noted that the percentages of column "-" which can be considered as the proportion of times that, in a certain sense, LMM fails to predict the next market regime are quite low: 0.9% and 3.5% for the S&P-500 and FTSE-MIB, respectively. The percentages in column 1 are higher and the model prediction accuracy based on columns 1 and 2 jointly reaches or exceeds 90%.

INSERT TABLE 6 ABOUT HERE

Our findings are valuable for choosing a profitable investment strategy. A constant update of the dynamic analysis through LMM may suggest the proper investment decision for the following month. For example, if the previous month has been classified into latent state 4 for the US index then it might be wise to buy, hold, or accumulate the amount of the investments because the probability of remaining in that positive and long lasting regime is high. On the contrary, when the previous monthly return observation of S&P-500 has been classified into state 3 then it may be better to reduce the investments since the chances of switching to a negative regime are higher than a shift to a positive one ($p_{31} + p_{32}$ = 0.61 and p_{35} = 0.38). Unfortunately, our results show that choosing a good investment strategy for the Italian market is more complicated. For instance, the LMM classified a month for the FTSE-MIB in the very positive latent state 5, the following month may either slump or continue to be positive with similar probabilities (p_{51} = 0.48 and p_{54} = 0.49).

4. Discussion

In this paper, we investigate the dynamic patterns of stock markets by exploiting the potential of the LMM for defining different market regimes and providing transition probabilities of regime-switching. We find evidence of a LMM with five latent states for both the US S&P-500 and the Italian FTSE-MIB index. In our framework, the latent states represent five stock market regimes. Regimes are clearly defined and characterized by different return means. Moreover, the LMM is able to detect the 2000/01 crisis and the crisis that started in 2007. A long stable period between these crises is also detected. The stable market regime is defined by one particular latent state characterized by a moderate positive return mean and a high state persistence probability.

Regime characterization and latent transition probabilities enable us to achieve two important goals in financial analysis. First, LMM allows us to promptly recognize the beginning of stable periods within a few months. This feature may provide the opportunity of detecting the end of the financial crisis that started in 2007. Furthermore, the model highlights the fact that, despite the preceding positive months, this crisis is not over in July 2009. The LMM provides insights on when the shift to a stable period is most likely to take place, e.g., after the positive latent state 5 for the US S&P-500. Second, it allows us to predict which regime the stock market is going to experience the following month.

The LMM provides a relevant focus on the dynamics of stock price indexes, which is quite difficult to recognize by simply eyeballing the raw time series graphs. That is to say, the entire stable period in our data is characterized in the same latent state, for both the US S&P-500 and Italian FTSE-MIB indexes. This enhances the recognition of the stable period, as there are still fluctuations of the raw price indexes occurring in the stable period. The model distinguishes between relatively moderate fluctuations in the stable period and the stronger fluctuations occuring in a period of crisis. Moreover, the crisis fluctuations are characterized as four prototypes, i.e., latent states. This enhances understanding of a crisis in terms of developments in stock market indexes.

We also relevant similarities and differences between the US S&P-500 and Italian FTSE-MIB indexes. Despite the presence of five different regimes for both indexes, their characterization differs in the number of positive and negative regimes and their intensity. The FTSE-MIB is characterized by more extreme regimes. This is consistent with the fact that US market is more developed than other stock markets (Demirguc-Kunt and Levine, 1996). Also, it turns out to be more difficult to predict developments in the Italian FTSE-MIB since the number of possible regime switching is higher than for the US S&P-500. However, the close correspondence between the regimes of the two analyzed stock markets and their common definition of stable and crisis periods may imply an interesting generalization of our results to the stock market developments in other countries.

Our contribution allows improving the investment opportunities at both strategic and operative levels, by basing decision-making on an advanced methodological process. A limitation of our study is that we have analyzed a 115 month period in two countries. Future studies should apply the methodology in this paper to other periods and countries to assess whether the latent states we found and other findings reported in the paper also apply under different circumstances. A second limitation applies to the approach in general. We do not aim to assess and predict precise changes of stock prices on a daily basis. Instead we model less precise changes of regimes across monthly data, implying the model is suited for long-term investment purposes.

REFERENCES

Bartolucci, F. Pennoni, F., and Francis, B. (2007) A latent Markov model for detecting patterns of criminal activity. *J. R. Statist. Soc.* ***A***, **170**, 115-132.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.*, 41, 164-171.

Bozdogan, H. (1987) Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, **52**(**3**), 345-370.

Demirguc-Kunt, A., and Levine, R. (1996) Stock market development and financial intermediaries: Stylized facts. *World Bank Economic Review*, **10**(**2**), 291-321.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* ***B***, **39**, 1-38.

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

Dias, J.G., Vermunt, J.K., and Ramos, S. (2008) Mixture hidden Markov models in finance research. *Advances in Data Analysis, Data Handling and Business Intelligence*. Berlin: Springer. Forthcoming.

Duekel, M.J. (1997) Markov switching in GARCH processes and mean-reverting stock-market volatility. *J. of Bus. & Ec. Statist.*, American Statistical Association, **15**(**1**), 26-34.

Francq, C., and Zakoïan, J.M. (2001) Stationarity of multivariate Markov-switching ARMA models. *J. of Econometrics,* **102**, 339-364.

Frühwirth-Schnatter, S., and Kaufmann, S. (2008) Model-based clustering of multiple time series. *J. of Bus. & Econ. Statist.*, **26**, 78–89.

Goldfeld, S.M., and Quandt, R.E. (1973) A Markov model for switching regressions. *J. of Econometrics*, **1**, 3-16.

Haas, M., Mittnik, S., and Paolella, M. (2004) A new approach to Markov-switching GARCH models. *J. of Fin. Econometrics*, **2**, 27-62.

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

Hamilton, J.D. (2008) Regime-switching models. In *The New Palgrave Dictionary of Economics*, 2nd edn, (eds. S.N. Durlaf and L.E. Blume).

Hamilton, J.D., and Raj, B. (2002) *Advances in Markov-switching models*. Berlin: Springer-Verlag.

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

Lamoureux, C.G., and Lastrapes, W.D. (1993) Forecasting stock return variance: toward an understanding of stochastic implied volatilities. *Rev. of Fin. St.*, **5**, 293-326.

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

Paas, L.J., Vermunt, J.K., and Bijmolt, T.H.A. (2007) Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisitions of financial products. *J. R. Statist. Soc.* ***A***, **170**, 955-974.

Rydén, T., Teräsvirta, T., and Ǻsbrink, S. (1998) Stylized facts of daily return series and the hidden Markov model. *J. of App. Econometrics*, **13**, 217-244.

Van de Pol, F., and Langeheine, R. (1990) Mixed Markov latent class models. In *Sociological Methodology 1990* (ed. C.C. Clogg), pp. 213-247. Oxford: Basil Blackwell.

Turner, C.M., Startz, R., and Nelson, C.R. (1989) A Markov model of heteroskedasticity, risk, and learning in the stock market. *NBER Working Paper*, **W2818**.

Vermunt, J. K., Langeheine R., and Böckenholt, U. (1999) Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *J. Educ. Behav. Statist.*, **24**, 178-205.

Vermunt, J.K., and Magidson, J. (2007) *Technical guide for Latent GOLD 4.5: basic and advanced*. Belmont, MA: Statistical Innovations Inc.

TABLE 1:

Means, standard deviations, skewness, kurtosis, and Jarque-Bera tests of S&P-500 and FTSE-MIB indexes in different periods (denotes test significance at 5%)*

Period	S&P-500				
	Mean Return	Std. Dev.	Skewness	Kurtosis	Jarque-Bera Test
Crisis 2000/01 (Jan-00 - May-03)	-0.881	5.382	0.208	2.225	1.32
Current Crisis (Oct-07 - Jul-09)	-1.735	6.794	-0.255	2.489	0.48
Stable Phase (Jun-03 - Sep-07)	0.913	2.179	-0.149	2.416	0.93
Entire data set (Jan-00 - Jul-09)	-0.233	4.688	-0.533	3.778	8.34*

Period	FTSE-MIB				
	Mean Return	Std. Dev.	Skewness	Kurtosis	Jarque-Bera Test
Crisis 2000/01 (Jan-00 - May-03)	-0.781	7.360	-0.120	2.719	0.23
Current Crisis (Oct-07 - Jul-09)	-2.624	8.465	0.730	3.920	2.73
Stable Phase (Jun-03 - Sep-07)	1.031	2.846	-0.514	2.309	3.32
Entire data set (Jan-00 - Jul-09)	-0.314	6.145	-0.203	4.313	9.06*

TABLE 2:

Log-likelihood function and CAIC criterion of the LMM from 1 to 8 latent states for S&P-500 and FTSE-MIB data sets

Number of	S&P-500 Data Set		FTSE-MIB Data Set	
Latent States	*LL*	*CAIC*	*LL*	*CAIC*
1	-340.35	682.70	-371.47	744.94
2	-334.30	674.59	-368.12	742.23
3	-327.80	667.60	-358.05	728.10
4	-316.44	652.88	-350.50	721.00
5	-308.03	646.06	-340.10	710.20
6	-302.79	647.59	-336.12	714.25
7	-297.22	650.45	-331.39	718.79
8	-298.29	668.58	-328.99	729.98

TABLE 3:

Sizes, return means, standard deviations, and Jarque-Bera tests of the 5 latent states for S&P-500 and FTSE-MIB data sets (denotes test significance at 5%)*

Latent State	S&P-500				FTSE-MIB			
	Size	Return Mean	Standard Deviation	Jarque-Bera Test	Size	Return Mean	Standard Deviation	Jarque-Bera Test
1	.112	-9.07	2.82	6.99*	.106	-12.46	2.64	0.61
2	.136	-4.23	2.57	0.77	.214	-3.75	2.06	0.68
3	.180	-0.13	1.52	0.39	.522	1.11	2.70	3.83
4	.437	0.94	2.14	0.88	.123	5.01	2.74	1.38
5	.136	6.74	2.08	0.80	.036	14.68	3.76	0.58
Entire data set	1.00	-0.23	4.69	8.34*	1.00	-0.31	6.15	9.06*

TABLE 4:

Latent transition matrix for S&P-500 index

$j \setminus k$	1	2	3	4	5
1	.2466	.4970	.0064	.0034	.2465
2	.0029	.0041	.9769	.0031	.0130
3	.4603	.1501	.0039	.0029	.3828
4	.0008	.0266	.0010	.9706	.0009
5	.0035	.2758	.3351	.1103	.2753

TABLE 5:

Latent transition matrix for FTSE-MIB index

$j \setminus k$	1	2	3	4	5
1	.0962	.4641	.0695	.3666	.0036
2	.2693	.4235	.0028	.3021	.0023
3	.0013	.0530	.9433	.0018	.0007
4	.1644	.3839	.2056	.0093	.2368
5	.4811	.0119	.0107	.4868	.0095

TABLE 6:

Number of times and percentages LMMs correctly predict the next latent state according to the three highest transition probabilities

	1	2	3	-	Total
S&P-500	87	17	9	1	114
%	76.3	14.9	7.9	0.9	100
FTSE-MIB	82	20	8	4	114
%	72.0	17.5	7.0	3.5	100

FIGURE 1:

S&P-500 and FTSE-MIB monthly return distributions from January 2000 to July 2009

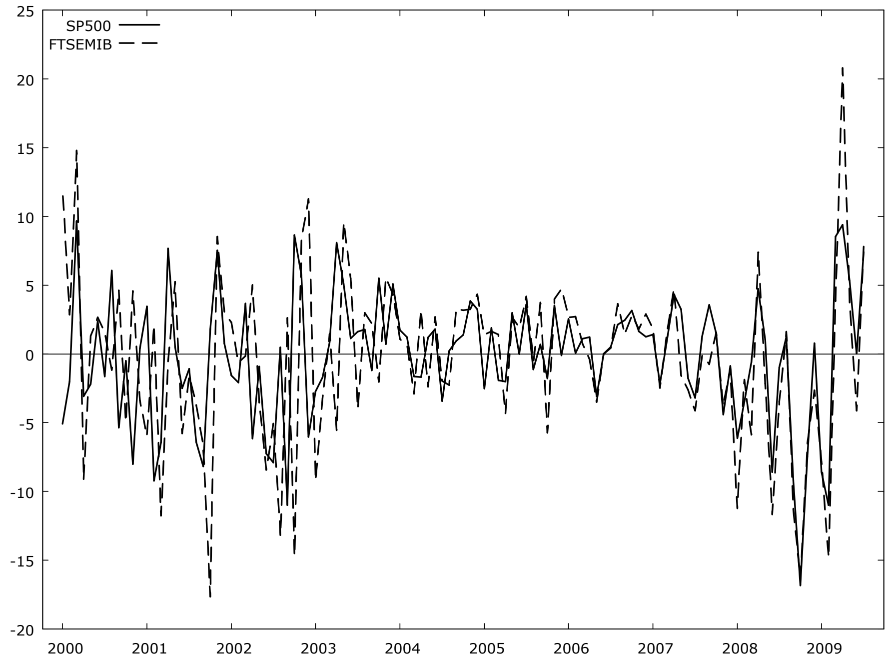
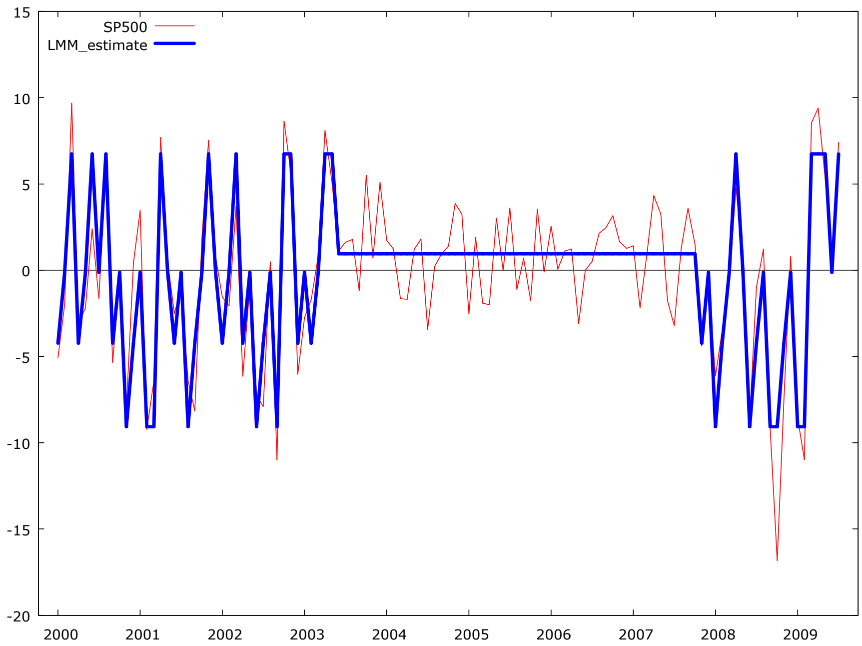


FIGURE 2:

S&P-500 (panel a), FTSE-MIB (panel b), and LMM estimates time series



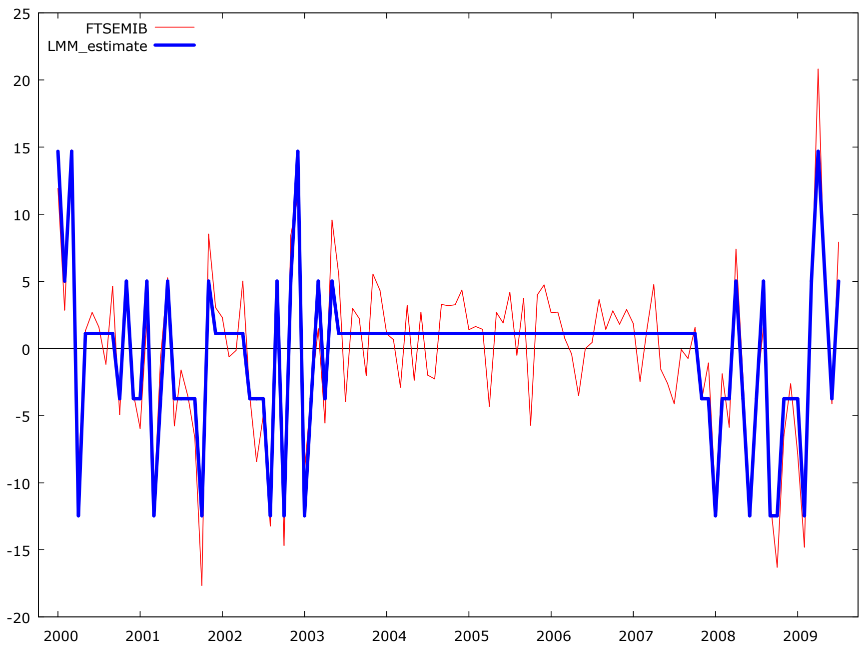


FIGURE 3:

S&P-500 (panel a), FTSE-MIB (panel b), overall LMM estimates, and LMM estimates for the stable regime (SP500_s and FTSEMIB_s)