A Duration Hidden Markov Model for the Identification of Regimes in Stock Market Returns

Christos Ntantamis *

CREATES, University of Aarhus

February 15, 2009

Abstract

When a process can be thought of as occupying one of a small number of states or regimes, the durations of those states, and factors affecting these durations, are often of particular interest. The traditional representation of equity markets as having bull and bear phases is an example; indicators of the predicted duration of these states are valuable. This paper introduces a Duration Hidden Markov Model to model such regime switches in the stock market; the duration of each state of the Markov Chain is a random variable that depends on a set of exogenous variables. The model not only allows the endogenous determination of the different regimes and but also estimates the effect of the explanatory variables on the regimes' durations. The model is estimated here on NYSE returns using the short-term interest rate and the interest rate spread as exogenous variables. The bull market regime is assigned to the identified state with the higher mean and lower variability of the returns' distribution; bull market duration is found to be negatively dependent on short-term interest rates and positively on the interest rate spread, while bear market duration depends positively the short-term interest rate and negatively on the interest rate spread.

Keywords: Hidden Markov Model; Variable-dependent regime duration; Regime Switching; Interest rate effect.

JEL Classification Numbers: C13, C22, G1

*Address: School of Economics & Management, University of Aarhus, Building 1322, 8000 Aarhus C, Denmark, telephone: +45 89422140, email: *cntantamis@creates.au.dk*. The author is grateful to John W. Galbraith for his helpful comments. I also benefited from comments at seminars at McGill University, Cornerstone Research, and CREATES. Center for Research in Econometric Analysis of Time Series, CREATES, is funded by The Danish National Research Foundation.

1 Introduction

Non-linearity of returns in the stock market is well established (Scheinkman and LeBaron 1989, Tong 1990, and McMillan 2001). The existence of different regimes in stock market returns has been proposed as one potential explanation for nonlinear conditional returns. Although in principle such processes might be represented as having any number of regimes, the existing theoretical and empirical literature has tended to concentrate on two regimes, the traditional 'bull' and 'bear' markets. Blanchard and Watson (1982) presented a model of stochastic bubbles that either survive or collapse in each period; in such a world, returns would be drawn from one of two distributions: surviving bubbles or collapsing bubbles. Cecchetti et al. (1990) considered a Lucas asset pricing model in which the endowment switches between high economic growth and low economic growth, and showed that such switching in fundamentals accounts for several features of stock market returns. Cecchetti, Lam, and Mark (2000) introduced belief distortions that vary over expansions and contractions and lead to systematic predictability in returns. Gordon and St-Amour (2000) examined a model in which risk-aversion changes according to an exogenous regime-switching process lead to the generation of bull and bear markets. Finally, Schaller and van Norden (1997) presented strong empirical evidence of regime switching in US stock market returns, evidence which was robust to different specifications of the nature of switching.

Where this division into regimes is of interest, there are a number of questions which it would be useful to answer. Can we identify these regimes? How frequent are they and when do they occur? Are regimes switches predictable? What can be said about the distributional characteristic of the returns, after accounting for regime switches?

Moreover, information about the duration of each regime is important. Investment decisions can of course be influenced, but (less evidently) policy questions may also arise. The share of stock market wealth in total household wealth has increased over recent years, and plays an important role in the household's consumption decisions [1]. The effect of these decision, when aggregated, is a determinant of economic growth and hence is of concern for governments and central banks.[2]

There are two streams of literature related to these classes of question. One stream of the econometric literature treats parameter estimation and regime identification (and therefore, duration analysis) as two independent exercises. Under this view, the estimation of the parameters of interest is performed, and the calculations of the regime of the stock market and its durations are made with the appropriately parameterized model. The identification of the regimes(states) is made exogenously with the help of data-based rules.[3] A key consequence of such a modeling strategy is that the duration of cycles is simply a by-product of the parameter estimates and not an intrinsic feature of the model.

Examples of this strain of literature include Neftçi (1984), who tested for asymmetries over the business cycles using a two-state Markov model; the identification of the different states was made using data-driven criteria. Diebold and Rudebusch (1990) modeled the hazard rates for the state duration using nonparametric methods, whereas Cohran and Defina (1995) considered a parametric approach to the state hazard rates using

[1]This accounts for 24.4 % of the total household wealth in 2006, based on the Flow of Accounts of United States reported by the Federal Reserve System

[2]Consumption accounts for 64.5% of the US Gross Domestic Product during the period 1947-2007 and for about 70 % for the period after 2001. The impact of the stock market wealth in consumption is found to be 2-5 cents increase in consumptions for every dollar increase in stock market wealth (Poterba 2000, Dynan and Maki 2001, Case, Quiegly and Shiller 2005).

[3]Hereafter the words 'regime' and 'state' will be used interchangeably.

an exponential distribution. Lunde and Timmermann (2004) modeled the hazard rates for the durations of bull and bear markets, using covariates that included exogenous variables such as interest rates, and the age of the state.

None of these papers presented models that treat data generating processes and regimes jointly. For an analyst who wishes to model the regime switches and the expected durations, the advantages of Markov switching models are clear. Since the regimes are jointly estimated with the parameters of the data generating process, the analyst can identify them without the restriction of data-based rules. The Markov switching models provide a very simple link between the transition probabilities and the expected durations. Examples, following Hamilton's (1989) seminal work, include Engel (1994) who modeled exchange rate regimes, Hamilton and Susmel (1994), Gray (1996), and Dueker (1997) who discussed regime switching in GARCH models, and Schaller and van Norden (1997) for US stock market returns.

Nevertheless, Hamilton's model does not allow one to investigate expected durations. The model assumes time-homogeneity of the transition probability matrix: as a result, expected durations of different states differ but are constant across time. The expected duration of the different regimes should be allowed to vary across time, according to some explanatory variables that describe the underlying economic and market conditions because this will be more consistent with economic intuition. Since the time-homogeneous transition probabilities model cannot capture this behavior, extensions have subsequently been introduced that allow the transition probabilities to vary across time. Diebold et al. (1994), Filardo (1994), and Schaller and van Norden (1997) allowed the state-transition probabilities to be functions of a set of explanatory variables. Maheu and McCurdy (2000) considered the transition probabilities to be logistic functions of the number of periods that the process has been in that state, during the past. Even though these approaches do not explicitly model state durations, state duration expectations are allowed to vary over time as they are now a function of the time-varying transition probabilities.

A more direct approach is proposed here: that is to model explicitly the duration of each state. Duration is now considered to be itself a random variable, with a distribution that can be modeled and its parameters estimated. Speech recognition applications have followed this approach, and there is therefore substantial existing literature. Automatic recognition of continuous speech is an attempt to use computers to transcribe naturally spoken utterances in accordance with the rules of orthography and it is done by modeling utterance production statistically, e.g. in the form of a Markov process. Even though Markov models suggest that the probability of timescale distortion at a particular instant of an utterance is reflected in the transition probabilities associated with the corresponding state of the underlying Markov process, it was found that they are not adequate in order to model the duration and temporal structure of words. Thus, the duration of the utterances were directly modeled. Ferguson (1980) considered a nonparametric approach, Russell and Moore (1985) assumed a Poisson distribution, Levinson (1986) estimated the model using a Gamma distribution, Mitchell and Jamieson(1993) suggested the use of the exponential family of distributions. The transition probability matrix is considered to be time-homogeneous with its diagonal elements to be equal to zero.

While attractive, these methods are restrictive in economic applications. Although the duration itself is now explicitly modeled, it is determined solely by the properties of the series under scrutiny without considering additional information. A set of explanatory variables must be introduced in the duration specification in order to provide more information about expected values, as in survival analysis applications for modeling survival times of patients (see e.g. Klein and Moeschberger 2002).

In this paper, a Duration Hidden Markov Model structure is proposed. The Hidden Markov Model structure, which is assumed to have time-homogeneous transition probabilities, allows for the endogenous identification of states, i.e. the identification of the regimes and the estimation of the parameters governing their distributional characteristics are jointly obtained within the model. The novelty of this model is that it assumes the duration of each state to be a random variable that depends on a set of explanatory variables, whose values are available at the time that the transition from one state to another takes place, within the Hidden Markov Model framework. The difference relative to models of survival analysis is that state durations are now not provided as part of the data set or as outcomes of data-based rules, but instead are obtained jointly with all the other parameters of the model. The functional form for the duration specification is the same across states, but the parameters are allowed to differ. Thus, expected durations are a function of explanatory variables and are allowed to vary across time but in contrast to models of time-varying transition probabilities, this effect is now direct.

Moreover, compared with the standard Markov switching literature, the observations's distributions are assumed to be a finite mixture of Normal distributions, instead of a single Normal distribution. Finite mixtures are known to provide a very flexible distributional form (MacLachlan and Peel 2000), which can accommodate many features of stock market returns (Rydén et al. 1998).

The model is applied to New York Stock Exchange (NYSE) monthly returns. The short-term interest rate and the interest rate spread, defined as the difference between long-term and short-term interest rates, are used as the explanatory variables in the duration function. The choice of these variables was made in order to capture potential information arising from the predictive value of these quantities for economic growth, and therefore profits of listed companies.

The bull market regime was identified as the state with the higher mean and the higher variance, but with lower coefficient of variability (the ratio of the standard deviation and the mean) for the returns' distribution. The bear market regime, on the other hand, had a lower mean and higher coefficient of variability for the returns. In terms of the duration specification results, the bull market's duration is found below to depend negatively on short-term interest rates and positively on the interest rate spread. The duration of bear markets is positively dependent on short-term interest rates and negatively on the interest rate spread. These results suggest that policy makers may have the ability to influence the durations of bull and bear stock markets by changing the levels of short-term interest rates.

The structure of the rest of the paper is as follows. Section 2 introduces the model and discusses the estimation procedure, the derivation of the standard errors, and how the model can be used for forecasting purposes. Section 3 discusses the data that were used for the estimation. Results from the estimation are presented in Section 4. Section 5 provides a brief conclusion and discuss some issues for further research.

2 Duration Hidden Markov Model

2.1 A Brief review of Hidden Markov Models

In what follows, only a brief description of the Hidden Markov Models is provided; for a more detailed review refer to Bilmes (2006), or Rabiner (1989).

A Hidden Markov Model (HMM) is a discrete -time stochastic process including an underlying finite-state Markov Chain (state sequence) and a sequence of random variables whose distributions depend on the state sequence of random variables only through the current state (observation sequence). The state sequence is not observable, and hence all the conclusions about the process must be made using only the observation sequence.

A more formal definition, which presents the probabilistic relationships between the state sequence and the observation sequence, is [4]:

Definition: A Hidden Markov Model (HMM) is a collection or random variables consisting of a set of T discrete scalar variables $X_{1:T}$ and a set of T other variables $Y_{1:T}$ which may be either discrete or continuous (and either scalar - or vector-valued). These variables, collectively, possess the following conditional independence properties:

$$\{X_{t:T}, Y_{t:T}\} \perp \{X_{1:t-2}, Y_{1:t-1}\} \mid X_{t-1} \tag{1}$$

and

$$Y_t \perp \{X_{\neg t}, Y_{\neg t}\} \mid X_t \tag{2}$$

for each $t \in 1:T$. No other conditional independence properties are true in general, unless they follow from (1) and (2).

An HMM will be, given the definition, any joint probability distribution over an appropriately typed set of random variables (X, Y) that obeys the stated set of conditional independence rules. The two conditional independence properties imply that, for a given T, the joint distribution over all variables may be expanded as follows (using the Chain rule and the definition equations):

$$P(y_{1:T}) = P(y_1) \prod_{t=2}^{T} P(x_t \mid x_{t-1}) \prod_{t=1}^{T} P(y_t \mid x_t) \tag{3}$$

Thus, in order to parameterize an HMM, one needs the following quantities:

(i) The distribution over the initial state variable: $P(x_1)$.

(ii) The conditional transition distributions for the first-order Markov Chain: $P(x_t \mid x_{t-1})$.

(iii) The conditional distribution for the other variables: $P(y_t \mid x_t)$.

It can be seen that these quantities correspond to the classic HMM definition, where X_t is the hidden Markov Chain and Y_t are the observed data. Specifically, the initial (not necessarily stationary, see Resnick 1992 or Hamilton 1994 for details) distribution is labeled π which is a vector of length equal to the number of the different states. Then, $\pi_i = P(X_1 = i)$, where π_i is the i^{th} element of π. The observation probability distributions are denoted $b_j(y_t) = P(Y_t = y_t \mid X_t = j)$ and the associated parameters depend on the $b_j(y)$'s family of distributions. Furthermore, the Markov Chain is typically assumed to be time-homogeneous, with stochastic transition matrix $\mathbf{A}$, where $A_{ij} = P(X_t = j \mid X_{t-1} = i)$, for all t. HMM parameters are often symbolized collectively as $\lambda \triangleq (\pi, \mathbf{A}, \mathbf{B})$, where $\mathbf{B}$ represents the parameters corresponding to all the observation distributions.

[4]The following notation will be employed: $X_{s:q}, s < q = \{X_s, X_{s+1}, \ldots, X_q\}$, $X_{<s} \triangleq \{X_1, X_2, \ldots, X_{s-1}\}$, and $X_{\neg t} \triangleq X_{1:T} \setminus X_t = \{X_1, X_2, \ldots, X_{t-1}, X_{t+1}, X_{t+2}, \ldots, X_T\}$.

The HMM parameters, $\lambda \triangleq (\pi, \mathbf{A}, \mathbf{B})$, are usually estimated with an algorithm advocated in a series of papers by Baum and co-authors (e.g. Baum et al 1970), which is known as the Baum-Welsh algorithm. This algorithm belongs to the Expectation Maximization (EM) algorithm family which was formally introduced by Dempster et al (1977), in order to feasibly estimate model parameters using maximum likelihood. The core of the Baum-Welsh algorithm is the recursive estimation of the forward and backward probabilities (α and β respectively). When these are recovered and with further assumptions regarding the form of the observations' distributions, the recursive estimation of the HMM parameters is straightforward.

2.2 Model description

The traditional HMM implies that the probability of remaining in the state i, for a duration τ, $P(\tau \mid X = i)$ is proportional to $A_{ii}^{\tau-1}$ and so state durations follow an exponential distribution. This approach is rather restrictive. Explicitly modeling the state duration as a random variable following a given distribution function, as in Ferguson (1980), Russell and Moore (1985) and Levinson (1986) was introduced in order to alleviate this problem (all of these papers are related to speech recognition applications).

Another issue is that the values of the transition probabilities (A_{ii}), are derived based only on the information set defined by the observations sequence. Extending the information set with a second observation sequence will improve the model fit on the data and may have favorable implications for forecasting. There are two ways to model this effect. The first way would be to discard the time-homogeneity of the transition probability matrix and to allow its element to depend on exogenous variables as in Diebold et al. (1994), whereas Maheu and McCurdy (2000) used the time that the chain has visited each state in the information set.

The model proposed in this paper extends the models discussed in the engineering literature by allowing the state duration to depend on a set of variables, other than the one provided by the observation sequence. This increases the information set available, and also provides intuition on how state durations may change over time. To do so we use a Hidden Markov Model with time-homogeneous transition probability. The state durations are assumed to be a probabilistic function of a set of exogenous variables, whose values are available to the researcher when the transition from one state to another takes place. The effect of these variables in the duration is allowed to differ across states. Overall, the proposed model consists of the following elements:

(I) Hidden Markov Model Structure: $\lambda = (\pi, \mathbf{A}, \mathbf{B})$.

The elements $\pi, \mathbf{A}$ are related to the hidden Markov Chain. The only assumption that will be imposed is that $A_{ii} = 0$, that is, the chain can not return immediately to the previous state. Regarding $\mathbf{B}$, the observation distributions space, a Mixture of univariate Normals is assumed for each of the underlying states. The number of the mixtures is given a priori and it is assumed to be the same across all states.[5] Hence, for $b_j \in \mathbf{B}$, we have:

[5] It is known that nonidentifiability can emerge in finite mixture models due to overfitting. The employment of an extra mixture component can be dealt with in two ways: either set one of the mixing components to zero or take two component densities to be the same. In this case, setting the number of mixtures provides an upper bound to the number of mixture components. Regardless, Redner (1981) proved that MLE remains consistent even in this case.

$$b_j(y_t) = \sum_{m=1}^{M} c_{jm} f\left(y_t; \mu_{jm}, \sigma_{jm}^2\right) \tag{4}$$

with

$$f\left(y_t; \mu_{jm}, \sigma_{jm}^2\right) = \frac{1}{\sqrt{2\pi}\sigma_{jm}} exp\left[-\frac{\left(y_t - \mu_{jm}\right)^2}{2\sigma_{jm}^2}\right] \tag{5}$$

The assumption of a mixture of Normals for the observation distributions increases the flexibility of the model since it captures more of the dynamics for the observation processes. It is well known, see McLachlan and Peel (2000), that such a mixture can replicate a large variety of distributional characteristics which are observed in financial data, such as fat tails and asymmetries. Moreover, Bilmes (2006) argues that improvement of the fitting of a Hidden Markov Model in data can be accomplished by either an increase in the number of states assumed or by considering observations' distributions that are very flexible, e.g. finite mixtures of Normal distributions with many mixtures.

(II) Duration Equation Specification: **D**

The time that the Markov Chain remains at each state, τ, is given by parametric relations, with coefficients that are different across states. The parameters are collectively summarized as $\mathbf{D} \triangleq \{\kappa_j, \zeta_j\}_{j=1}^{n}$. The general form of the model, for the state j, is:

$$\log \tau = \kappa_j' Z + \zeta_j W \tag{6}$$

where Z is the vector of the exogenous variables (of dimension k), including an intercept term. If the innovations are assumed to follow a standard Normal distribution, the density function for the duration τ is:

$$d_j(\tau; Z) = \frac{1}{\tau}\phi\left(\frac{\ln \tau - \kappa_j' Z}{\zeta_j}\right). \tag{7}$$

Assuming that the state duration is determined as soon as the process enters that particular state, the values of the variables should correspond with the time the transition to this state occurred. To put it in otherwise, if the Markov Chain (X) has just entered into state i at time t ($X_t = i$), then the amount of time that it will stay there, τ, will depend on the values of the exogenous variables Z at that particular time t, i.e. Z_t.

2.3 Model estimation

The estimation of the model is based on an extension of Levinson(1986) methodology. Although that paper modeled durations according to a gamma distribution,, the recursive equations for the $\pi, \mathbf{A}$ parameters remain the same as in the parametric specification approach of this paper. Denote the observations random vector by $Y_{1:T}$, and by $X_{1:T}$ the sequence of the hidden states that can take n discrete values $X_t = i, i = 1, \ldots, n$ (the realizations of these random variables are denoted by $y_{1:T}, x_{1:T}$ respectively).

2.3.1 Estimation of $\pi, \mathbf{A}$.

The Forward and Backward Recursions required for the calculation of the alpha and beta probabilities are:

- Forward recursion:

$$\alpha_1(j) = \pi_j b_j(y_1) d_j(1; z_1) \tag{8}$$

$$\alpha_t(j) = \sum_{\tau \leq t} \sum_{i=1, i \neq j}^{n} \alpha_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \prod_{\theta=1}^{\tau} b_j \left(y_{t-\tau+\theta}\right) \tag{9}$$

- Backward recursion:

$$\beta_T(j) = 1 \tag{10}$$

$$\beta_t(i) = \sum_{\tau \leq T-t} \sum_{j=1, j \neq i}^{n} A_{ij} d_j(\tau; z_t) \prod_{\theta=1}^{\tau} b_j \left(y_{t+\theta}\right) \beta_{t+\tau}(j) \tag{11}$$

It is apparent that the main difference between these recursions and those for the simpler Hidden Markov Model is the incorporation of the time that the chain remains in the state, which is represented by the duration probability ($d_j(\tau; Z)$) and the product of the values of the observation distribution for the data within the particular state.

The likelihood that the observed sequence is generated according to the model is then given by:

$$\mathcal{L}\left(\lambda \mid y_{1:T}\right) = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \sum_{\tau \leq T-t} \alpha_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \prod_{\theta=1}^{\tau} b_j \left(y_{t-\tau+\theta}\right) \beta_t(j) \tag{12}$$

The following quantity, corresponding to the probability that the chain is in state i at time t, can also be determined:

$$\gamma_t(i) = P\left(X_t = i \mid y_{1:T}, \lambda\right) = \frac{\alpha_t(i)\beta_t(i)}{\mathcal{L}\left(Y_{1:T} \mid \lambda\right)} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{n} \beta_t(j)\alpha_t(j)} \tag{13}$$

Using the above relations, the estimator for the elements of the transition matrix $\mathbf{A}$ is:

$$\hat{A}_{ij} = \frac{\sum_{t=1}^{T-1} \sum_{\tau \leq t} \alpha_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \prod_{\theta=1}^{\tau} b_j \left(y_{t-\tau+\theta}\right) \beta_t(j)}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)} \tag{14}$$

2.3.2 Estimation of B.

The space $\mathbf{B}$ contains the observation distributions that are assumed to be mixtures of normals. Before proceeding to the recursive formulae for the parameters, the following quantity must be introduced:

$$\gamma_t(j,k) = \left[\frac{\alpha_j(t)\beta_j(t)}{\mathcal{L}\left(Y_{1:T} \mid \lambda\right)}\right] \left[\frac{c_{jk} f\left(y_t; \mu_{jk}, \Sigma_{jk}\right)}{\sum_{m=1}^{M} c_{jm} f\left(y_t; \mu_{jm}, \Sigma_{jm}\right)}\right] \tag{15}$$

which is the probability of being at state j at time t with the k^{th} mixture accounting for y_t. The term $\gamma_t(j,k)$ generalizes to the standard $\gamma_t(j)$ in the case of a simple mixture (or a discrete density).

Moreover, the following weighting function, the ratio of the probability that the $Y_{t-\tau+\theta}$ observation is generated by the k^{th} element of the j^{th} mixture distribution to the overall probability that it is generated by the j^{th} mixture distribution, is required for the estimating equations of the means and variances of the mixtures:

$$w_\theta = \left[\frac{c_{jk} f\left(y_{t-\tau+\theta} \mid \mu_{jk}, \sigma_{jk}^2\right)}{\sum_{m=1}^{M} c_{jm} f\left(y_{t-\tau+\theta} \mid \mu_{jm}, \sigma_{jm}^2\right)} \right]. \tag{16}$$

(a) Mixing probabilities:

$$\hat{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_t(j,m)} \tag{17}$$

(b) Mean of a mixing distribution:

$$\hat{\mu}_{jk} = \frac{\sum_{t=1}^{T-1} \sum_{\tau \leq t} \alpha_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \prod_{\theta=1}^{\tau} b_j\left(y_{t-\tau+\theta}\right) \beta_t(j) \sum_{\theta=1}^{\tau} w_\theta y_{t-\tau+\theta}}{\sum_{t=1}^{T-1} \sum_{\tau \leq t} \alpha_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \prod_{\theta=1}^{\tau} b_j\left(y_{t-\tau+\theta}\right) \beta_t(j) \tau \sum_{\theta=1}^{\tau} w_\theta} \tag{18}$$

(c) Variance of mixing distribution:

$$\hat{\sigma}_{jk}^2 = \frac{\sum_{t=1}^{T-1} \sum_{\tau \leq t} \alpha_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \prod_{\theta=1}^{\tau} b_j\left(y_{t-\tau+\theta}\right) \beta_t(j) \sum_{\theta=1}^{\tau} w_\theta \left(Y_{t-\tau+\theta} - \mu_{jk}\right)^2}{\sum_{t=1}^{T-1} \sum_{\tau \leq t} \alpha_{t-\tau}(i) A_{ij} d_j(\tau) \prod_{\theta=1}^{\tau} b_j\left(y_{t-\tau+\theta}\right) \beta_t(j) \tau \sum_{\theta=1}^{\tau} w_\theta} \tag{19}$$

Compared with Hidden Markov Models without explicit duration specification, the difference in the estimation formulae lies in considering the sum of the observations only for the period that the hidden chain stays in the particular state. This is required for the correct determination of the mean; if the state does not change, more observations are generated according to its observation distribution and thus should be incorporated in the mean and variance estimates. Given the mixture approach, the observations also need to be weighted according to their probability of belonging to the particular element mixture (w_θ). The existence of the duration τ is included in the denominator (the number of the extra observations) in order to average over the extra observations due to the chain remaining in the same state. The number of these observations should also be weighted accordingly using the weighting function w_θ, so as to find the effective number of observations, within each duration τ, that have been generated by the particular element of the mixture distribution.

2.3.3 Estimation of the parameters for the duration model $d_j(\tau)$.

The parameters to be estimated are the regression coefficients κ_j and the standard deviation ζ_j. The estimating equations will be derived by taking the derivatives of the complete likelihood ($\mathcal{L}\left(Y_{1:T} \mid \lambda\right)$ with respect to each of the parameters and setting them equal to zero. Thus,

$$\frac{\vartheta \mathcal{L}}{\vartheta \kappa_j} = 0 \tag{20}$$

$$\frac{\vartheta \mathcal{L}}{\vartheta \zeta_j} = 0 \tag{21}$$

which will yield, using (12), the following relations:

$$\sum_{t=1}^{T}\sum_{\tau\leq t}\sum_{i=1,i\neq j}^{n}\alpha_{t-\tau}(i)A_{ij}\frac{\vartheta d_j(\tau;z_{t-\tau})}{\vartheta\kappa_j}\prod_{\theta=1}^{\tau}b_j\left(y_{t-\tau+\theta}\right)\beta_t(j)=0 \tag{22}$$

$$\sum_{t=1}^{T}\sum_{\tau\leq t}\sum_{i=1,i\neq j}^{n}\alpha_{t-\tau}(i)A_{ij}\frac{\vartheta d_j(\tau;z_{t-\tau})}{\vartheta\zeta_j}\prod_{\theta=1}^{\tau}b_j\left(y_{t-\tau+\theta}\right)\beta_t(j)=0 \tag{23}$$

Notice that the second summation over the states disappears; the duration model for each state j depends only on the parameters κ_j, ζ_j. The summation over time is due to fact that the derivative over a product is taken. Thus, in order to write the estimating equations for the parameters, we need to write down the partial derivatives of the duration distributions with respect to those, i.e. $\frac{\vartheta d_j(\tau;Z)}{\vartheta\kappa_j}$ and $\frac{\vartheta d_j(\tau;Z)}{\vartheta\zeta_j}$:

$$\frac{\vartheta d_j(\tau;Z)}{\vartheta\kappa_j}=\frac{\ln\tau-\kappa_j'Z}{\zeta_j^2}\frac{1}{\tau}\phi\left(\frac{\ln\tau-\kappa_j'Z}{\zeta_j}\right)Z \tag{24}$$

$$\frac{\vartheta d_j(\tau;Z)}{\vartheta\zeta_j}=\frac{\left(\ln\tau-\kappa_j'Z\right)^2}{\zeta_j^3}\frac{1}{\tau}\phi\left(\frac{\ln\tau-\kappa_j'Z}{\zeta_j}\right). \tag{25}$$

We notice that the parameters of interest can not be isolated from the derivative functions so as to obtain the estimating equations as a part of the Baum-Welsh iterative procedure. Thus, a numerical estimation will be necessary so as to retrieve the parameter values κ_j, ζ_j and use them as inputs to the iterative algorithm.

2.3.4 Numerical Instability

Although the recursive formulae given in the previous sections are correct, they suffer from the numerical problem that $\alpha_t(i), \beta_t(i)$ tend to zero as t increases. Thus, it is necessary to provide an alternative so as to keep these probabilities within the limited dynamic range of the computer.

This paper follows the Devijver and Dekesel (1988) approach to handling this issue. They proposed to replace the joint probabilities in the definitions of the forward and backward probabilities with the *a posteriori*, probabilities $(\alpha_t'(i), \beta_t'(i),)$ i.e

$$\alpha_t'(i)=P\left(X_t=i\mid y_{1:T}\right) \tag{26}$$

$$\beta_t'(i)=\frac{P\left(y_{t+1:T}\mid X_t=i\right)}{P\left(y_{t+1:T}\mid y_{1:T}\right)} \tag{27}$$

This replacement leads to the following new recursive relations:

$$\alpha_t'(i)=\frac{\alpha_t(i)}{\sum_{j=1}^{n}\alpha_t(j)} \tag{28}$$

$$\beta_t'(i)=\frac{\beta_t(i)}{\sum_{j=1}^{n}\alpha_t(j)} \tag{29}$$

These adjusted forward and backward probabilities should be used in the forward and backward recursions. This weighting scheme results in the following changes in the recursions:

- Forward recursion:

$$\alpha_t(j) = \sum_{\tau \le t} \sum_{i=1, i \ne j}^{n} \alpha'_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \left[\prod_{\theta=1}^{\tau-1} q_{t-\tau+\theta} b_j\left(y_{t-\tau+\theta}\right) \right] b_j\left(y_t\right) \tag{30}$$

- Backward recursion:

$$\beta_t(i) = \sum_{\tau \le T-t} \sum_{j=1, j \ne i}^{n} A_{ij} d_j(\tau; z_t) \left[\prod_{\theta=1}^{\tau-1} q_{t+\theta} b_j\left(y_{t+\theta}\right) \right] b_j\left(y_{t+\tau}\right) \beta'_{t+\tau}(j) \tag{31}$$

where:

$$q_t = \left[\sum_{j=1}^{n} \alpha_t(j) \right]^{-1} \tag{32}$$

The adjusted forward and backward probabilities are subsequently used in the updating functions of the model parameters. Once more, the weighting scheme needs to be considered. For example, the updating equation for the transition probabilities becomes:

$$\hat{A}_{ij} = \frac{\sum_{t=1}^{T-1} \sum_{\tau \le t} \alpha'_{t-\tau}(i) A_{ij} d_j(\tau; z_{t-\tau}) \left[\prod_{\theta=1}^{\tau-1} q_{t-\tau+\theta} b_j\left(y_{t-\tau+\theta}\right) \right] b_j\left(y_t\right) \beta'_t(j)}{\sum_{t=1}^{T-1} \alpha'_t(i) \beta'_t(i) / q_t} \tag{33}$$

2.4 Computation of standard errors

The standard errors of parameter estimates are obtained using the Dietz and Böhning (1996) approach. This approach is based on the result that in large samples from regular models for which the log likelihood is quadratic in the parameters, the likelihood ratio and the Wald test for the significance of an individual parameter are equivalent, implying that the deviance change, which is twice the change in log likelihood on omitting one variable, say λ_i, is equal to the square of the t-statistic. Thus the standard error can be calculated as the absolute value of the parameter estimate divided by the square root of the deviance change,

$$s.e.(\lambda_i) = \frac{\mid \lambda_i \mid}{\sqrt{2\left(l_\lambda - l_{\lambda_{\neg i}}\right)}}, \tag{34}$$

where l_λ is the unrestricted log-likelihood, and $l_{\lambda_{\neg i}}$ is the log-likehood for the model with the parameter λ_i being equal to zero.

It is important to emphasize that estimates of the covariance matrix of the MLE based on expected or observed information matrices are guaranteed to be valid inferentially only asymptotically.

There are other methods for deriving the confidence intervals for the parameter estimates, which asymptotically provide the same results (see Aittokallio et al. 1999 and Visser et al 2000). Nevertheless, their implementation in this model is problematic. Methods that rely on the calculation of the observed Hessian matrix, **H**, in order to derive the approximate covariance matrix **C** (Aittokallio and Uusipaikka 2000) suffer from numerical instability problems [6]. Alternatively, methods such as the construction of confidence intervals

[6] The score vector and the Hessian matrix of the log-likelihood are obtained by forward recursions of the first and the second derivatives of the forward probabilities. Since the forward recursions for the Hessian require the unweighted versions of the forward probabilities which, as has already been discussed, cannot be obtained, the computation is not feasible.

for the parameters using the profile likelihood function or a parametric bootstrap require the model being re-estimated many times, which, for the model proposed in this paper is computationally prohibiting [7].

2.5 Forecasting

When the model was introduced in the previous sections, we noted that the main idea is that the duration of each state is determined by a set of explanatory variables when the transition to this state occurs. This fact makes the task of forecasting less straightforward than with standard Hidden Markov Models.

Consider for example the case of forecasting at time $T + c$. In order to produce a forecast, two things are required: a) knowledge of the underlying state X_{T+c}, and b) all potential paths that can lead to this state, starting from the state at the end of the sample X_T, along with their probabilities of occurrence. The problem with the Duration HMM model is that the number of paths increases substantially as the forecasting horizon c increases. Imagine the process entering the state j at time $T + 1$. The time that the process will remain in the particular state, (τ_1), depends on the values of the explanatory variables at this period. It may be that $\tau_1 < c$. In such a case, the process moves to a different state (given that $A_{jj} = 0$), and a new duration, τ_2, will be specified according to the values of the explanatory variables at time $X_{T+\tau+1}$. If $\tau 1 + \tau_2 < c$ then the processes repeats itself, otherwise it is terminated. It becomes obvious that when the forecasting horizon is large, there are numerous potential paths.

One way to circumvent this problem would be to simulate a series of paths, say V, up to the forecasting horizon, and then obtain the forecasts by averaging across paths. The procedure can be summarized as follows:

1. At time $T + 1$ draw a state according to π; say j.
2. Using (6), the duration function parameters for the j^{th} state and the values of the explanatory variables Z at time $T + 1$, determine the value of the state duration τ_1.
3. The forecasted returns will be:
$$\hat{Y}_{T+1:T+\tau_1} = \sum_{m=1}^{M} c_{jm}\mu_{jm} \tag{35}$$
where $\sum_{m=1}^{M} c_{jm}\mu_{jm}$ is the mean of the observation distribution for the j^{th} state.
4. If $\tau_1 \geq c$, then the path is complete. Return to step 1.
5. If $\tau_1 \leq c$, then:
   (a) Draw the next state using the transition matrix $\mathbf{A}$; say l $(l \neq j)$.
   (b) Using (6), the duration function parameters for the l^{th} state and the values of the explanatory variables Z at time $T + \tau_1 + 1$, determine the value of the state duration τ_2.
   (c) The forecasted variable values will be:
$$\hat{Y}_{T+\tau_1+1:T+\tau_1+\tau_2} = \sum_{m=1}^{M} c_{lm}\mu_{lm} \tag{36}$$

[7] For example, a profile likelihood approach requires eight model estimations, on average, to achieve two digit accuracy for the confidence interval's end points (Visser et al 2000)

(d) If $T + \tau_1 + \tau_2 \geq T + c$, then the path is complete. Return to step 1.

(e) If $T + \tau_1 + \tau_2 \leq T + c$, the internal procedure is repeated till $T + \sum \tau_s \geq T + c$.

6. Repeat this procedure V times to obtain the paths and the corresponding forecasted values:

$$\{\hat{Y}^v_{T+1:T+c}\}^V_{v=1} \tag{37}$$

7. The mean forecasts will be provided by:

$$\hat{Y}_{T+1:T+c} = \frac{1}{V}\sum_{v=1}^{V} \hat{Y}^v_{T+1:T+c} \tag{38}$$

In practice, this approach yields a distribution of forecasts for the returns' for each time-point within the forecasting horizon, thus making it suitable for Value at Risk (VaR) applications. Moreover, one may want to obtain forecasts for the variance of the returns; instead of using the weighted sum of the means of the components of the mixture when forecasting returns, the weighted sum of the variances are to be used. This way, a distribution of forecasts for the variance will be obtained for each point in time making this approach suitable for option pricing as well.

Nevertheless, such a forecasting procedure is implausible as it assumes that the values of the explanatory variables are known during the entire forecasting period. Note that this is also a major drawback for all models where the transition probabilities are expressed in terms of exogenous variables.

However, it is still possible to perform forecasting for small horizons by concentrating on the mean state durations instead of the point outcomes. The state duration distribution function (6), implies the following mean duration value for the j^{th} state at time t:

$$E_j(\tau) = e^{\gamma_j Z_t + \zeta^2/2} \tag{39}$$

The procedure to be followed is the following:

1. Consider the mean durations for every possible state, at time $T + 1$:

$$\{E^{T+1}_j(\tau) = e^{\gamma_j Z_{T+1} + \zeta^2/2}\}^n_{j=1}$$

2. Consider the minimum of the mean durations:

$$\tau^* = \min_j E^{T+1}_j \tau$$

3. The forecasted values will be:

$$\hat{Y}_{T+1:T+\tau^*} = \sum_{j=1}^{M} \pi_j \sum_{m=1}^{M} c_{jm}\mu_{jm} \tag{40}$$

The final equation suggests that when no knowledge of future values Z is available, forecasting is possible for a horizon up to τ^*, which is the minimum of the expected state durations.

2.6 Recovering the sequence of states

The final step in model estimation is to reconstruct the sequence of hidden states. It should be clear that for all but the case of degenerate models, there is no "correct" state sequence to be found. Hence, for practical situations, an optimality criterion is usually employed to solve this problem (there are several reasonable optimality criteria that can be imposed).In this paper, the identification of the bull and bear markets is done by what is known as Bayesian Segmentation. This approach assigns data points to regimes according to probabilistic arguments for the likelihood of data point(s) belonging to a particular regime.

One potential way to do this is the Maximum Posterior Mode (MPM) approach. This approach reconstructs the state sequence by allocating at each time point the state that it is more probable, given the observations and the model, i.e.

$$\hat{x}_t = \arg \max_{i=1,\ldots,n} \gamma_t(i), \qquad 1 \leq t \leq T \tag{41}$$

Although this criterion maximizes the expected number of correct states, there could be some problems with the resulting state sequence (the "optimal" state sequence may, in fact, not even be a valid state sequence). This occurs because MPM determines the most likely state at every instant, without regard to the probability of sequences of states. In order to circumvent this problem, one may attempt to maximize the expected numbers of correct pairs of states (x_t, x_{t+1}) or triples of states (x_t, x_{t+1}, x_{t+2}). The extreme version of this approach, which is the Maximum A Posteriori (MAP) sequence, is to find the single best state sequence (path), i.e. to maximize $P(X \mid \mathbf{y}, \lambda)$ that is equivalent to maximize $P(X, \mathbf{y} \mid \lambda)$. This is accomplished by the Viterbi algorithm.

Nevertheless, the quantity $\gamma_t(i)$ is calculated at each iteration step, and thus it is readily available. Therefore, for reasons of simplicity, the MPM criterion will be employed.

3 Data

Monthly returns from 1978:03 to 2007:06, from the Center for Research in Security Prices (CRSP) value-weighted portfolio for the New York Stock/American Exchange, are considered for the application of the model.

The short term interest rate, defined as the three-month treasury bill secondary market rate, and the interest rate spread, defined as the difference between the ten-year treasury constant maturity rate and the short-term interest rate, are used as the explanatory variables in the duration function specification.[8] Interest rates have been widely documented to closely track the state of the business cycle and appear to be a key determinant of stock returns at the monthly horizon (Fama and French 1989, Whitelaw 1994, Pesaran and Timmermann 1995).

The use of short-term interest rates is an attempt to capture short-run economic outlook information in state durations. On the other hand, interest rate spread, which can be interpreted as the expectation of future rates corresponding to the period of two maturities, has been proven to have a considerable predictive ability for future economic activity, especially for longer forecasting horizons (Estrella and Hardouvelis 1991,

[8] Data for the interest rates were obtained by the Board of Governors Federal Reserve System (FRED II).

Estrella and Mishkin 1995). This permits us to incorporate a determinant in the state duration function that contains information regarding longer periods in the future.

The model is estimated for the period from 1978:03 through 2003:8. The remaining 46 observations are used for out-of-sample forecasting. Figure 1 plots the monthly series for the entire sample, and Table A.1 provides the descriptive statistics for the variables into consideration, both in-sample and out-of sample.[9] Focusing on the distributional properties of the returns, we see that: a) asymmetry exists both in-sample and out-of-sample, and b) there is excess kurtosis, compared with the normal distribution, in-sample but less out-of-sample. These results indicate, as has often been observed, that stock market returns deviate from normality; the Jarque-Bera test confirmed this observation (the p-value for the in-sample data is virtually zero). Therefore, the use of mixtures of normal distributions for the analysis for the observations' distributions is not unreasonable[10].

4 Results

The duration HMM is estimated for the case of two states for the stock market, i.e. we attempt to identify the bull and bear market regimes. The number of mixtures for the observation distributions, M, was assumed to be equal to 3.

Two more models for the stock market returns are estimated for forecast comparison purposes. The first assumes an Autoregressive model of order one (AR(1)) for the stock market returns with the innovations following a GARCH(1,1) process. The second model considers a regression of the stock market returns on the previous period's values of the short-term interest rates and of the interest rate spread. The innovations of the regressions are modeled to follow a GARCH(1,1) process as in the AR case[11].

- Model 1:

$$r_t = c_0 + c_1 r_{t-1} + \epsilon_t, \qquad \epsilon_t = \sqrt{h_t} v_t \tag{42}$$

$$h_t = b_0 + b_1 \epsilon_{t-1}^2 + b_2 h_{t-1}, \qquad v_t \sim i.i.d(0,1) \tag{43}$$

- Model 2:

$$r_t = c_0 + c_1 intsh_{t-1} + c_2 intspr_{t-1} + \epsilon_t, \qquad \epsilon_t = \sqrt{h_t} v_t \tag{44}$$

$$h_t = b_0 + b_1 \epsilon_{t-1}^2 + b_2 h_{t-1}, \qquad v_t \sim i.i.d(0,1) \tag{45}$$

where r_t is the stock market return, i_t is the T-Bill rate (in %), and s_t is the interest rate spread (in %).

The results from the estimation of the duration HMM are presented in Tables A.2 to A.5, and they can be summarized as follows:

[9] Interest rates and interest rates spread are expressed in percentages.

[10] The deviation of stock market returns from normality could also be attributed to the existence of nonlinear dynamics.

[11] One may argue that the proposed models are not appropriate for forecasting returns, thus they are poor benchmarks for evaluation purposes. Nevertheless, even though forecasting the returns is important, the main focus of the proposed model is in the explicit modeling of the duration of the states. What we need to make sure is that the provided return forecasts are not poor when compared with those obtained by relatively more naive models.

- The bull market is identified as the state with the higher mean and the lower coefficient of variability for the observations' distribution (Table A.5),i.e. state 1; the bear market is identified as having a lower mean but a higher coefficient of variability compared to the bull market, i.e. state 2.

- Although the assumption of a finite mixture of Normal distribution was made in order to allow capturing potential asymmetries or fat tail behavior of the data within each state, the results suggest that this was not necessary. The estimated mixtures do not exhibit asymmetries or fat tails; the Jarque-Bera statistic fails to reject normality at any conventional significance level. That is, the Hidden Markov structure itself has captured adequately the asymmetries and fat-tails presented in financial data; these features follow from the estimated representation. Thus, simpler distributions, like simple Normals, could have been assumed for the observations' distributions.

- The estimates for the duration parameters suggest adverse effects for the short-term interest rates and the interest rate spread. When the state enters a bull market regime, higher short-term interest rates will have a negative effect in the state duration. This may be explained by changes in portfolio holdings with shifts towards the risk-free asset when its returns are increasing. By contrast, a higher interest rate spread, which has already been employed as a signal for future economic activity, points to a larger duration for the bull market. The same reasoning applies to the effects of these variables in the bear market duration. When the future economic outlook is favorable, as suggested by the higher interest rate spread, there is a positive impact on the stock market in the long-run, leading to a reduction in the duration of the bear market regime.

- The estimated quantitative effects of the short-term interest rate and the interest rate spread on the duration of the different regimes are: a) for bull markets, an increase in the T-Bill rate of 1% will lead to a 32% reduction in the state duration, whereas a 1% increase in the spread will lead to a 42% increase in the duration, and b) for the bear markets,an increase in the T-Bill rate of 1% will lead to an approximately 44% increase in the duration, whereas a 1% increase in the spread will lead to a 49% decrease in the duration. For example, a 25 basis point reduction in the federal funds rate, all other things being equal, will prolong an expected bull market duration by a total of approximately 19% [12].

- The reconstructed state sequence is depicted in Figure 2. The resulting bull and bear stock market regimes are in line with the general consensus. The bull market during the 80's coincides with the beginning of President Reagan's administration and the subsequent economic growth. Moreover, the boom period for the stock market during the 90's is also well captured by the model. The bear stock market in the middle of the 80's contains the October 1987 stock market crash and the aftermath. However, it is interesting to notice that the stock market entered in a bear regime in the beginning of 1987, even before the crash. One more consideration should be noted. Even though the particular state sequence is one of many possible sequences (see discussion in section 2.6) and it depends on the choice of the Bayesian Segmentation method, it is still the outcome of the model's identification of the characteristics of the bull/bear markets. Different outcomes will have yielded different paths. This

[12] The total effect is the sum of 8%, which is the direct effect in the change in the short-term interest rates assuming the change in the FED rate pass completely in the 3-month T-Bill rate, and of 11%, which is the indirect change in the interest rate spread.

provides a partial explanation of the fact that the state sequence seems to contain fewer switches from one type of market to another compared to what one may have expected.

- The forecasting ability of the model is examined versus the forecasting performance of the two simpler models described above. The estimation results for these models are tabulated in A.6. The forecasting performance comparison is done in terms of the Root Mean Square Error (RMSE); the results are reported in Table A.7. The forecasted values for the duration model are obtained by simulating 5000 different paths (V=5000). Overall, the forecasting performance of the duration HMM is comparable to that of the other two models.

5 Conclusion

In this paper, I have proposed a Duration Hidden Markov Model structure for modeling processes with two or more regimes. The model combines a time-homogeneous Hidden Markov Model structure, which allows for the endogenous identification of states, with an explicit functional specification for the duration of each state. Durations are considered to be random variables that depend on a set of explanatory variables, whose values are available at the time that the transition from one state to another takes place. The functional form for the duration specification does not alter across states, but the parameters are allowed to differ.

The advantages of the model compared with the Markov regime switching currently dominant in the literature in economics are: a) the direct modeling of the expected duration of the different regimes by the introduction of covariate-dependent state durations in the Hidden Markov Model, b) the analysis of the impact of these covariates in regime's durations and the corresponding policy implications, and c) the consideration of a more general distributional form for the observations' distributions.

The issues of the estimation of the parameters and the derivation of their standard errors were discussed. The numerical instability issues were addressed by the adjustment of the forward and backward recursions. The procedure for obtaining forecasts by this model, which is based on simulating a number of paths, was developed. The model was subsequently applied in New York monthly stock market returns. The short-term interest rate and the interest rate spread, i.e. the difference between long-term interest rates and short-term interest were used as the explanatory variables for the regime's duration in order to capture the short-run and the long-run economic outlook.

In applying the model, I identified a bull market regime in which returns have a higher mean and lower variability, versus a bear market regime with lower mean returns and higher variability. Both the results from the reconstruction of the historical sequence of the states, and the estimated effect on duration of the explanatory variables were in accordance with intuition. An increase in the short-term interest rate decreases the duration of the bull market regime and increases the duration of the bear market regime. On the contrary, an increase in the interest rate spread leads to a higher duration for the bull market and a lower one for the bear market. These results not only suggest the ability of policy makers to influence the durations of the bull and bear market regimes by changing the short-term interest rates, which also indirectly affects the interest rate spread, but provide quantitative estimates of impacts. A reduction of 25 basis points in the federal funds rate is estimated to increase by 19% the duration of a bull market and to reduce by 24 % the duration of a bear market. This feature of the model has significant policy implications for central bankers and investors as well.

One of the main, unofficial, tasks that FED has undertaken the past years is the conservation of a bull stock market or the reduction of the duration of a bear market. This duration model suggests that FED can quantify the bull/bear regime durations, and adjust them by changing the short-term interest rates. Thus, if FED wants to prevent a prolonged period of bear markets, it should act immediately at the beginning and reduce interest rates so as to reduce the duration of the bear regime. Nevertheless, it should be noted that only the duration can be affected; if the economy is weak, as evidenced by a negative spread the duration of a bear market may still be considerable.

On the investors' side, the importance in determining the durations of the two regimes lies in the optimal timings of their investments. It should be expected that long or short positions on the market would be taken accordingly to anticipation of entering a bull or bear regimes. The timing of such actions depends upon the durations of the regimes. In anticipation of a reduction in the interest rate by the FED so as to reduce the duration of a bear market, investors can rebalance their positions by reducing the duration of their short positions or by entering long in the market.

Finally, the model's forecasting performance was assessed, using a new procedure introduced here that is based on simulating a number of paths. The model was compared with two simpler models and it was found to be comparable in terms of RMSE; of course, the simpler models do not provide the duration information provided by the present model, which is the main advantage of this model.

References

[1] Aitkin M., and D.B. Rubin, (1985). Estimation and hypothesis testing in finite mixture models, *J. Roy. Statist. Soc. Ser. B*, 47, 6775.

[2] Aittokallio T., V. Ahola, and E. Uusipaikka, (1999). Likelihood Based Inference in Hidden Markov Models, *Turku Centre for Computer Science*, TUCS Technical Report 260.

[3] Aittokallio T., and E. Uusipaikka, (2000). Computation of Standard Errors for Maximum-Likelihood Estimates in Hidden Markov Models, *Turku Centre for Computer Science*, TUCS Technical Report 379.

[4] Baum L.E, T. Petrie, G. Soules and N. Weiss, (1970). A Maximization Technique in the Statistical Analysis of Probabilistic Functions of Finite State Markov Chains,*Annals of Mathematical Statistics*, 41, 164-171.

[5] Bilmes J., (2006). What HMMs Can Do, *EICE Transactions on Information and Systems*, E89, 869-89.

[6] Blanchard O.J., and M. Watson, (1982). Bubbles, Rational Expectations and Financial Markets, *Crises in the Economic and Financial Structure: Bubbles, Bursts, and Shocks*, ed. P. Watchel, Lexington Books.

[7] Cappé O., E. Moulines, and T. Rydén, (2005). *Inference in Hidden Markov Models*, Springer-Verlag, New York.

[8] Case K.E., J.M. Quigley, and R.J.Shiller, (2005). Comparing Wealth Effects: The Stock Market vs. the Housing Market, *Advances in Macroeconomics*, vol. 5, 1.

[9] Cecchetti S.G., P.Lam, and N.C. Mark, (1990). Mean Reversion in Equilibrium Asset Prices, *American Economic Review*, 80, 398-418.

[10] Cecchetti S.G., P. Lam, N.C. Mark, (2000). Asset Pricing Under Distorted Beliefs: Are Equity Returns Too Good To Be True?, *American Economic Review*, 90, 787-805.

[11] Cohran S., and R. Defina, (1995). Duration Dependence in the US Stock Market Cycle: A Parametric Approach, *Applied Financial Economics*, 5, 309-318.

[12] Dempster A. P., N.M Laird, and D.B. Rubin, (1977). Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.

[13] Devijver P.A., and M. Dekesel, (1998). Champs aléatoires de Pickard et modelization d' images digitales, Traitement Du Signal, 5, 131-150.

[14] Diebold F.X., and G.D. Rudebusch, (1990). A Nonparametric Investigation of Duration Dependence in the American Business Cycle, *The Journal of Political Economy*, 98, 596-616.

[15] Diebold F.X., J-H. Lee, and G. Weinbach, (1994). Regime Switching with Time-Varying transition Probabilities, in *Nonstationary Time Series Analysis and Cointegration*, ed. C. Hargreaves (*Advanced Texts in Econometrics*, series eds. C.W.J Granger and G. Mizon), Oxford University Press.

[16] Dietz E., and D. Böhning, (1996). Statistical Inference Based On a General Model of Unobserved Heterogeneity, in *Advances in GLIM and Statistical Modeling*, eds Fahrmeir L., F. Francis, R. Gilchrist, and G. Tutz, (*Lecture Notes in Statistics*, Springer, Berlin.

[17] Dueker M.J., (1997). Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility, *Journal of Business and Economic Statistics*, 15, 26-34.

[18] Dynan K.E., and D.M. Maki, (2001). Does Stock Market Wealth Matter for Consumption?, *Finance and Economics Discussion Series*, Board of Governors of the Federal Reserve System, 23.

[19] Engel C., (1994). Can the Markov Switching Model Forecast Exchange Rates?, *Journal of International Economics*, 36, 151-165.

[20] Estrella A., and G.A. Hardouvelis, (1991). The Term Structure as a Predictor of Real Economic Activity, *Journal of Finance*, 46, 555-576.

[21] Estrella A., and F. S. Mishkin, (1995). Predicting U.S. Recessions: Financial Variables as Leading Indicators, *National Bureau of Economic Research*, Working Paper 5379.

[22] Fama E.F., and K.R. French, (1989). Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics*, 25, 23-49.

[23] Ferguson J.D., (1980). Variable Duration Models for Speech, *Proceedings of the Symposium on the Application of Hidden Markov Models to Text and Speech*, ed. J. D. Ferguson, 143-179, Princeton, New Jersey.

[24] Filardo A.J., (1994). Business-Cycle Phases and Their Transitional Dynamics, *Journal of Business and Economic Statistics*, 12, 299-308.

[25] Gordon S., and P. St-Amour, (2000). A Preference Regime Model of Bull and Bear Markets, *The American Economic Review*, 90,1019-1033.

[26] Gray S., (1996). Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process, *Journal of Financial Economics*, 42, 27-62.

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

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

[29] Hamilton J. D., and R. Susmel (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime, *Journal of Econometrics*, 64, 307-333.

[30] Klein J., and M. Moeschberger, (2002). *Survival Analysis, Techniques for Censored and Truncated Data*, 2nd edition, Springer-Verlag, New York.

[31] Leroux B.G., (1992). Maximum-likelihood estimation for hidden Markov models, *Stochastic Processes and Their Applications*, 40, 127-143.

[32] Levinson S.E., (1986). Continuously Variable Duration Hidden Markov Models for Automatic Speech Recognition, *Computer Speech and Language*, 1, 29-45.

[33] Lunde A. and A.Timmermann, (2004). Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets, *Journal of Business and Economic Statistics*, 22, 253-273.

[34] McLachlan C. and D.Peel, (2000). *Finite Mixture Models*, New York: John Wiley and Sons.

[35] McMillan D.G., (2001). Nonlinear Predictability of Stock Market Returns: Evidence from Nonparametric and Threshold Models,*International Review of Economics and Finance*, 10, 353-368.

[36] Maheu J., and T. McCurdy, (2000). Identifying Bull and Bear Markets in Stock Returns, *Journal of Business and Economic Statistics*, 18, 100-112.

[37] Mitchell C.D., and L.H. Jamieson, (1993). Modeling Duration in a Hidden Markov Model with the Exponential Family, *Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, Minnesota, April 1993, 331-334.

[38] Neftçi S.N, (1984). Are Economic Time Series Asymmetric over the Business Cycle?, *The Journal of Political Economy*, 92, 307-328.

[39] Pesaran M.H., and A. Timmermann, (1995). Predictability of Stock Returns: Robustness and Economic Significance, *Journal of Finance*, 50, 1201-1228.

[40] Poterba J.M., (2000). Stock Market Wealth and Consumption, *The Journal of Economic Perspectives*, 14, 99-118.

[41] Rabiner L.R., (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, 77, 257-286.

[42] Redner A.R., (1981). Note on the Consistency of the Maximum Likelihood Estimation for Nonidentifiable Distribution, *The Annals of Statistics*, 9, 225228.

[43] Resnick S.I, (1992).*Adventures in Stochastic Processes*, Boston: Birkhäuser.

[44] Russell M. J., and R.K. Moore, (1985). Explicit Modeling of State Occupancy in Hidden Markov Models for Automatic Speech Recognition, *Proceedings of ICASSP-85*, Tampa, Florida, March 1985, 5-8.

[45] Rydén T., T. Teräsvirta, and S. Åsbrink, (1998). Stylized Facts of Daily Return Series and the Hidden Markov Model, *Journal of Applied Econometrics*, 13, 217-244.

[46] Schaller H., and S. van Norden, (1997). Regime Switching in Stock Market Returns, *Applied Financial Economics*, 7, 177-191.

[47] Scheinkman J.A., and B. LeBaron, (1989). Nonlinear Dynamics and Stock Returns, *The Journal of Business*, 62, 311-337.

[48] Tong H., (1990). *Non-linear Time Series. A Dynamical System Approach*, Oxford University Press, Oxford.

[49] Visser I., M.E.J. Raijmakers, and P.C.M. Molenaar, (2000). Confidence Intervals for Hidden Markov Model Parameters, *British Journal of Mathematical and Statistical Psychology*, 53, 317-327.

[50] Whitelaw R.F., (1994). Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns, *Journal of Finance*, 49, 515-541.

A Tables

A.1 Data Descriptive Statistics

	Stock Returns (%)	T-Bill Rate (3 month)	Contract Rate (10 year)	Interest Rate Spread
In Sample				
mean	1.15603	6.44824	8.13170	1.68346
standard deviation	4.60555	3.10673	2.68511	1.33895
skewness	-0.78240	0.79994	0.65685	-0.59939
kurtosis	2.38846	0.71689	-0.27525	0.10507
min	-22.5336	0.9	3.33	-2.65
max	12.84833	16.3	15.32	4.42
Out of sample				
mean	1.19580	3.18354	4.49271	1.30917
standard deviation	2.39054	1.56502	0.32382	1.36019
skewness	-0.18891	-0.28328	0.29366	0.45138
kurtosis	-0.76317	-1.54113	-0.68534	-1.33067
min	-3.76809	0.88	3.83	-0.38
max	6.03314	5.03	5.11	3.7

A.2 Markov Chain Results

state	**1**	**2**
π	0.99	0.01
A		
1	0	1
2	1	0

A.3 Observation distributions results

state	**1**	**2**	**1**	**2**	**1**	**2**
mixture						
	Mixing Prob.		**Means**		**Variances**	
1	0.27509	0.7506	0.04736	0.0214	0.00072	0.0010
2	0.01311	0.1173	-0.17139	-0.0183	0.00212	0.00004
3	0.71180	0.1321	0.01182	-0.0632	0.00195	0.0003

A.4 Duration parameters results

state	1		2	
parameter	**Coefficient**	**Std. Error**	**Coefficient**	**Std. Error**
κ_{j1}	3.1996	(0.2607)	2.7696	(0.2842)
κ_{j2}	-0.3189	(0.0305)	0.4409	(0.0344)
κ_{j3}	0.4150	(0.0389)	-0.4938	(0.0457)
ζ_j	0.97	1.18		

A.5 Observation distributions characteristics

state	Mean	Variance	Variability Coef.	Skewness	Kurtosis	J-B test*
1	0.019199	0.0011045	1.684	-0.00314	3.0087	0.6210
2	0.005585	0.000569	4.27	0.0010	3.0145	0.4113
	*: p-value					

A.6 Benchmark Models

Model 1			Model 2		
	Coefficient	**Std. Error**		**Coefficient**	**Std. Error**
Mean Equation					
Const	0.010737	0.002583	**Const**	0.007970	0.008563
AR(1)	0.040318	0.067161	**irsh(-1)**	0.000241	0.000927
			irspr(-1)	0.000941	0.002004
Variance Equation					
Const	6.68E-05	6.51E-05	**Const**	6.92E-05	6.87E-05
ARCH(1)	0.039767	0.032793	**ARCH(1)**	0.040038	0.033933
GARCH(1)	0.929518	0.054305	**GARCH(1)**	0.928158	0.056670
Log-Likel.	508.4655		**Log-Likel.**	508.3253	

A.7 Forecasting Evaluations

RMSE				
Forecasting Horizons	**+12**	**+24**	**+36**	**+46**
Model				
Model 1	0.026653	0.026075	0.024494	0.023529
Model 2	0.026586	0.026082	0.024465	0.023643
Duration HMM	0.027381	0.026382	0.024924	0.023879

B Figures

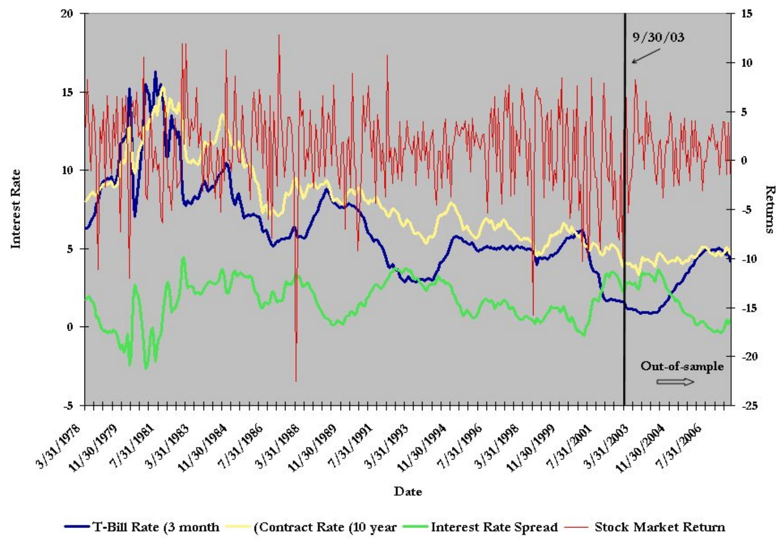

Figure 1: Time Series Plot of Variables

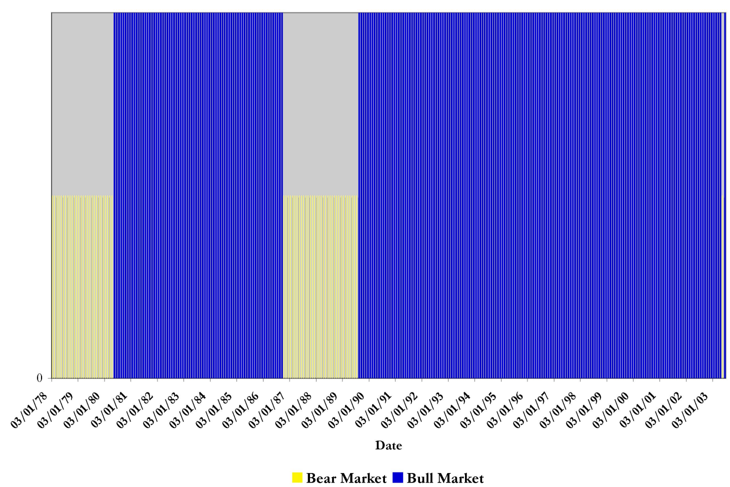

Figure 2: Time Series Plot of Variables