

**TREND DETECTION  
THROUGH SEGMENTATION  
USING DISCRETE HIDDEN  
MARKOV MODEL**

*Hacettepe University  
Journal of Economics  
and Administrative  
Sciences  
Vol. 38, Issue 2, 2020  
pp. 267-295*

**Özler ÖZGÜR**

Dr., Social Security Institution  
Directorate General for Service Delivery  
Pension Services Department  
ozlerozgur@gmail.com

**A**bstract: There are two models that benefit from the concept of hidden variables and are used frequently in practice. These are hidden Markov model and Markov switching model. Although there are quite a number of regime switching studies which applied the Markov switching model in the field of econometrics, studies that use the hidden Markov model for examining the regimes in econometric time series are sparse. In this study, we apply discrete hidden Markov model to four high-frequency time series and show that although transformed discrete data have to be used in the model structure, the model identifies two or more regimes quite well. It is concluded that the discrete hidden Markov model is defining regimes effectively, thereby, can also identify and segment trends.

**Keywords:** *Hidden Markov Model, regime switching, trend detection, segmentation, pattern finding.*

# AYRIK GİZLİ MARKOV MODELİNİ KULLANARAK SEGMENTASYON YOLUYLA TREND TESPİTİ

Hacettepe Üniversitesi  
İktisadi ve İdari Bilimler  
Fakültesi Dergisi  
Cilt 38, Sayı 2, 2020  
s. 267-295

## Özler ÖZGÜR

Dr., Sosyal Güvenlik Kurumu  
Hizmet Sunumu Genel Müdürlüğü  
Emeklilik Yazılımları Daire Başkanlığı  
ozlerozgur@gmail.com

**Ö**z: Gizli değişkenler kavramından yararlanan ve pratikte sıklıkla kullanılan iki model bulunmaktadır. Bunlar gizli Markov modeli ve Markov rejim değişikliği modelidir.

Ekonometri alanında Markov rejim değişikliği modelini uygulayan çok sayıda rejim değişikliği çalışması olmasına rağmen, ekonometrik zaman serilerindeki rejimleri incelemek için gizli Markov modelini kullanan çalışma sayısı çok azdır. Bu çalışmada, dört yüksek frekanslı zaman serisine ayrık gizli Markov modeli uygulanmış ve modellemede dönüştürülmüş ayrık veri kullanılmasına rağmen, gizli Markov modelinin iki ya da daha fazla rejimi oldukça iyi tanımladığı görülmüştür. Sonuç olarak ayrık gizli Markov modelinin rejimleri etkili bir şekilde tanımlayabildiği, böylece trendleri belirleyip seriyi segmentleyebildiği sonucuna varılmıştır.

**Anahtar Sözcükler:** Saklı Markov Modeli, rejim değişikliği, trend belirleme, segmentasyon, örüntü tanıma.

## INTRODUCTION

In the literature, there can be found an abundant number of studies that apply expectation maximization (EM) approach in fields of statistics, computational biology, electronic engineering (for various examples in related fields see Dempster *et al.* (1977), Do and Batzoglou (2008), Moon (1996) respectively). One famous model which inherits from EM is hidden Markov model (HMM) and the HMM is well known for being used in speech recognition problems and also pattern finding in computational biology. Despite this variety, there are few studies that apply discrete HMM in econometrics to determine the regime changes (or regime switchings) and explain them with the HMM results and findings.

A seminal study based on the concept of hidden variable in the literature is HMM, explained in Rabiner (1989), which was a review of several recent studies at that time. It seems that, in the literature, there are fewer studies focused on regime changes with HMM, which is relatively easy to implement. While Markov switching model (MSM) is often used in econometric applications, the number of studies applied HMM is relatively small in econometrics field. It can be surmised that it would be also difficult to directly make predictions in this type of HMM study. However, it is possible to deal with regime changes with such a model. Additionally, due to the simplicity of the model which relieves performance constraints, more hidden variables can be included in the model. Therefore, using HMM rather than MSM enables us to interpret more detailed models.

The structure of this paper is as follows: In section 2 we review the literature. In section 3, we mention the problems defined by Rabiner (1989) in HMM and we explain EM and parameter estimation procedure of HMM. In section 4, we apply HMM to four high-frequency time series in the empirical study. Section 5 covers the evaluation of the findings and discussion. Section 6 is the conclusion part of the study.

### 1. REVIEW OF LITERATURE

The concept of hidden variables appears to be very effective in modelling because it can extract additional information, which can be called a pattern, from data. In hidden Markov models, this additional information is said to be hidden or unseen. Most of the time, this hidden pattern can be thought of as showing long-term trends of the data. HMM is a very fruitful model because of that, it is used for solving a variety of problems from a variety of fields with success. In this study, for obtaining the solution, we based on a methodology that is used in voice recognition.

The HMM was intensively researched and developed for solving the problems in the voice recognition studies through the 1980's.

One of the important studies in speech recognition is Poritz (1982). In that study, Poritz used the term “hidden Markov model” apparently for the first time in the literature. Poritz presented so-called linear predictive HMMs and used predictive densities in the global model. He also gave information about techniques of parameter estimates. Similarly, Juang and Rabiner (1985a) performed autoregressive HMMs with speech recognition. Juang and Rabiner (1985a), being different from Poritz (1982), used mixture models for linear densities.

Juang and Rabiner (1985b) built up a method that measures the dissimilarities between different HMMs. They claimed that their distant measure is more advantageous from the preceding one.

Rabiner and Juang (1986) gave information on both the discrete and the continuous HMMs in their studies. They shared a general statistics that continuous models yielded a better 1% than the discrete models<sup>1</sup> according to the experiments performed in laboratories.

The model we used, which can be called discrete HMM, mostly based on Levinson *et al.* (1983) paper. It can be seen from speech recognition studies that the word predictions are performed using problem 1 solutions. In addition, problem 2 solutions are also used for voice recognition under more difficult voice or speech conditions. In this econometric study, the solution results of the second problem are used.

In this paper for the solution of the second problem, specifically the criterion where the maximum likelihood of the order of the states was considered. When this criterion is taken into consideration, the most commonly used method for solving the problem 2 is the Viterbi algorithm which was introduced by Viterbi (1967).

One of the earlier papers introducing the concept of hidden variables following the Markovian stochastic process is Lindgren (1978). Lindgren (1978) gave information on why hidden state transitions should be Markovian. He mentioned various models which have hidden variables as one of them being switching regressions and explained how to make parameter estimates for each model. He also gave information on robustness and consistency of ML estimates for models with hidden variables.

Rydén *et al.* (1998) said that Lindgren’s study has been taken as the base model for their study. Rydén *et al.* (1998) note that at any given time it is assumed that a different (or the same) regression is activated, again depending on the Markovian process. In their study, Rydén *et al.* (1998) came to a conclusion that the hidden Markov model can handle stylized facts well.

Giudici *et al.* (2000) showed that likelihood ratio test, as its asymptotic theory is problematic when used for HMM, is valid for HMM under some regularity assumptions.

One of the important steps in modelling HMM is selecting the most appropriate model. Many model selection tests used in classical methods can't be used for HMM because of the structure of HMM. While two states produce consistent and easy to assess state process in HMM modelling, it might be superficial for some types of applications. And while more than five states give detailed information related with regimes, the state process may totter too much and produce meaningless results to use in observation assessments. In that case, justification for using HMM disappears because then regime concept becomes one that is not meant to be. Because of these reasons, it is crucial to choose the appropriate model to have meaningful results.

Costa and De Angelis (2010) pointed out that, there is still no consensus on the mechanism to choose the best model. They performed a simulation study using various information criteria and likelihood ratio test. In their simulation study, they concluded that the factors which affect the information criterion most are the number of observations, conditional state dependent probabilities and latent transition matrix. They also concluded that for model selection, Akaike Information Criterion (AIC) is superior to Bayesian Information Criterion (BIC) when working with a small sample size.

De Angelis and Paas (2013) also conducted HMM for regime detection using latent variables. In that study, they also made a prediction.

Hamilton (1989) exploited the hidden variables to handle the regime changes in postwar US real GNP time series. In addition to the work carried out in this area until then, he proposed a tractable modelling of regimes that makes stochastic transitions. With that model, changes in long-run GDP forecast levels were able to be associated with growth and recession periods by a percentage proportion. Hamilton (2010) reviewed the model which is created with the Kalman filtering spirit. Hamilton and Raj (2002) conducted a review of the MSM extensions up to that time.

The model we used is discrete in time besides state and emission processes. This makes our approach different than that of Hamilton (1989), Rydén *et al.* (1998) and De Angelis and Paas (2013) where emissions are continuous. Besides that Hamilton (1989) also assumes that emissions are dependent. We didn't perform prediction as the model can not directly be used for prediction. Hence, although the discrete HMM is not a new model in literature<sup>2</sup>, this paper is the first to use discrete HMM for relating the historical events and the regimes obtained from the HMM model.

## 2. PROBLEMS DEFINED IN HMM STRUCTURE

Because HMM is well described in review paper Rabiner (1989) and other HMM related studies afterwards, all of the HMM problems and solutions were not examined in detail here. In this section, instead, fundamentals of EM and Baum-Welch algorithms were mentioned. We gave only the definition of the first problem and the procedure for the second problem.

### 2.1. Model Elements

N: number of states in the model

M: number of distinct observation symbols per state

$$\mathbf{A} = \{a_{ij}\}, a_{ij} = P(q_{t+1} = j | q_t = i), 1 \leq i, j \leq N \quad (1)$$

$$\mathbf{B} = \{b_j(k)\}, b_j(k) = P(o_t = v_k | q_t = j), 1 \leq k \leq M, j = 1, 2, \dots, N \quad (2)$$

$$\pi_i = P(q_1 = i), 1 \leq i \leq N \quad (3)$$

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}) \quad (4)$$

$O$  and  $Q$  are observation and state series respectively.

### 2.2. First Problem

Problem 1 of HMM is to obtain observation probabilities given the model parameters. Hence problem 1 is to calculate  $P(O|\lambda)$ . Rabiner and Juang (1986) said it could be considered as a model scoring or model evaluation.

### 2.3. Second Problem

Problem 2 of HMM is to determine the order of states, given the model parameters and observations, according to a predetermined optimality criterion. According to optimality criterion we have chosen, we try to find expression below (Fink, 2007):

$$\operatorname{argmax}_Q P(Q, O|\lambda) = \operatorname{argmax}_Q P(Q|O, \lambda) \quad (5)$$

Defining,

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda) \quad (6)$$

the algorithm is generally shown to have these steps:

Initialization:

$$\delta_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N \quad (7)$$

$$\psi_1(i) = 0 \quad (8)$$

Recursion:

$$\delta_t(j) = \max_{1 \leq i \leq N} \{ \delta_{t-1}(i) a_{ij} \} b_j(O_t), 2 \leq t \leq T, 1 \leq j \leq N \quad (9)$$

$$\psi_t(j) = \underset{1 \leq i \leq N}{\operatorname{argmax}} \delta_{t-1}(i) a_{ij}, 2 \leq t \leq T, 1 \leq j \leq N \quad (10)$$

Termination:

$$P^* = \max_{1 \leq i \leq N} \delta_T(i) \quad (11)$$

$$q_T^* = \underset{1 \leq i \leq N}{\operatorname{argmax}} \delta_T(i) \quad (12)$$

Back-Tracking:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), t = T-1, T-2, \dots, 1 \quad (13)$$

Interested reader can take a look at Bhar and Hamori (2004) for details.

## 2.4. Third Problem

### 2.4.1. Expectation Maximization Review:

The EM is basically a recursive optimization procedure. Initial studies on the method extend to the publication of Ceppellini *et al.* (1955) with the problem of gene frequency estimation. A more general study was presented by Hartley (1958) and Baum *et al.* (1970). Nevertheless, the algorithm is attributed to Dempster *et al.* (1977) paper (Do and Batzoglou, 2008). Dempster *et al.* (1977) provided evidence of various aspects of the algorithm and proposed a highly effective solution for a particular aspect of the EM. Another important work related to the algorithm is the 1983 article by Wu. Wu (1983) showed that there is no general convergence theorem for EM and provided evidence for a more general form of EM. Wu (1983) also proposed algorithms for more efficient operations (Kobayashi *et al.*, 2012).

An advantage of EM is that it can be applied where it is difficult to obtain Hessian matrix or it is difficult to apply the Newton-Raphson method to solve the same kind of problems. In addition, EM gives better results than Newton-Raphson method even with poor initial values. Moreover, the EM is a more stable algorithm than Newton-Raphson (Gupta, Chen, 2010). Nevertheless, the algorithm does a linear convergence different from the Newton-Raphson algorithm which performs quadratic convergence. Several methods have been proposed to speed up the EM, two of which are, based on Louis (1982), Aitken acceleration method and EM methods used in conjunction with the quasi-Newton methods (Tanner, 1996; Jamshidian, Jennrich, 1997).

The EM is used to solve problems that have missing data. This can be of two types. Either data entering the model is missing for various reasons, that is, some of the data cannot be observed or secondly the problem is such that it can be solved analytically under the assumption that some data is missing and otherwise an analytical solution cannot be obtained. For example, mixture models, HMM problems are of the second type (Bilmes, 1998). Such problems are also known as pattern recognition problems.

The problem description of the EM can be made as follows for the second kind of problems mentioned above: There are two sample spaces. One of them is an incomplete data space that is called  $\mathcal{Y}$ . The other is the complete data space, called  $\mathcal{X}$ . Observed data  $Y$  is a realization from space  $\mathcal{Y}$ . Mathematically,

$$x \rightarrow y(x) \tag{14}$$

$$y = y(x) \tag{15}$$

can be written (Dempster *et al.*, 1977). Accordingly, when an  $X$  is observed, a corresponding  $Y$  is considered to be observed. In other words,

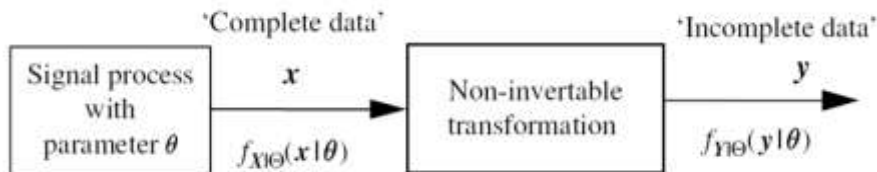
$$\mathcal{X}(Y) = \{X: f(X) = Y\} \tag{16}$$

can be written (Collins, 1997). In the model, it is assumed that there is a sample density denoted by  $f(x|\varphi)$ , and from that  $g(y|\varphi)$  values are derived. Complete data shown by  $f(\dots|\dots)$  is associated with incomplete data via,

$$g(Y|\varphi) = \int_{\mathcal{X}(Y)} f(X|\varphi) dX \tag{17}$$

This functional relationship summarized in Figure 1.

**Figure 1. Complete Data To Incomplete Data Transformation, (Vaseghi, 2007)**



EM, allows to obtain  $\theta$  value which maximizes,

$$L(\theta) = \log g(Y|\theta) \tag{18}$$

Here  $\log f(X|\theta)$  usually has an easily defined analytically solvable maximum, but  $L(\theta)$  does not have an analytical solution for maximization.  $L(\theta)$  usually has more



than one maximum (Collins, 1997). Since EM is a recursive optimization algorithm, transitions below,

$$\theta_t \rightarrow \theta_{t+1} \quad (19)$$

are defined by a serial parameter setup that associated with the relationship

$$L(\theta_{t+1}) \geq L(\theta_t) \quad (20)$$

Under certain conditions, the equality is only provided at critical points. The iteration procedure for the EM is established as below (Borman, 2009). Given,

$$L(\theta) = \ln P(X|\theta) \quad (21)$$

it is expected that the ultimate likelihood value will be greater than the values in the reiterations. This can be written as,

$$L(\theta) > L(\theta_n) \quad (22)$$

With these mathematical expressions and with the notion of incomplete data we can write inequality as below

$$L(\theta) - L(\theta_n) \geq \sum_Z P(Z|X, \theta_n) \ln \left( \frac{P(X|Z, \theta)P(Z|\theta)}{P(Z|X, \theta_n)P(X|\theta_n)} \right) \quad (23)$$

$$\triangleq \Delta(\theta|\theta_n)$$

Then with a new definition

$$L(\theta) \geq L(\theta_n) + \Delta(\theta|\theta_n) \triangleq l(\theta|\theta_n) \quad (24)$$

we can now write the iteration expression as,

$$\theta_{n+1} = \operatorname{argmax}_{\theta} l(\theta|\theta_n) \quad (25)$$

Removing the constants with respect to  $\theta$  we can obtain:

$$\begin{aligned} \theta_{n+1} &= \operatorname{argmax}_{\theta} \sum_Z \ln P(Z, X|\theta) P(Z|X, \theta_n) \\ &= \operatorname{argmax}_{\theta} E_{Z|X, \theta_n} \ln P(Z, X|\theta) \\ &= \operatorname{argmax}_{\theta} Q(\theta|\theta_n) \end{aligned} \quad (26)$$

Additionally with definitions above we'll get

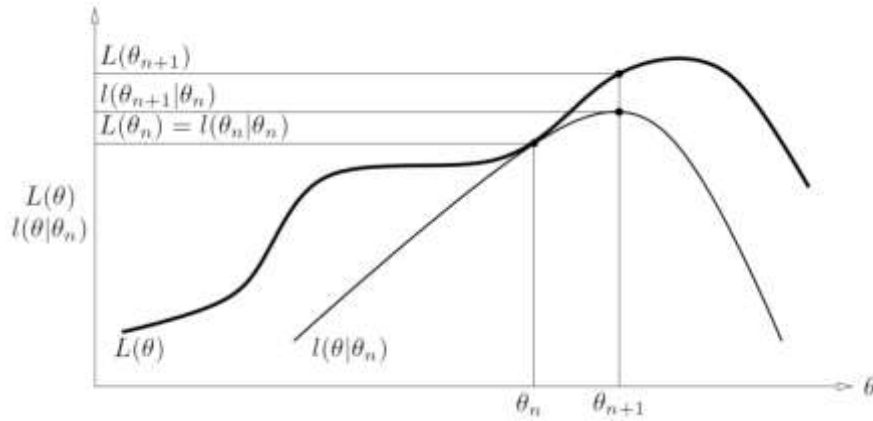
$$l(\theta_n|\theta_n) = L(\theta_n) + \Delta(\theta_n|\theta_n) = L(\theta_n) \quad (27)$$

$\theta_{n+1}$  will be chosen as to maximize  $\Delta(\theta|\theta_n)$  hence,

$$\Delta(\theta_{n+1}|\theta_n) \geq \Delta(\theta_n|\theta_n) = 0 \quad (28)$$

So that, at each iteration  $L(\theta)$  will have a non-decreasing structure. The following result is obtained:  $l(\theta|\theta_n)$  is limited by the  $L(\theta)$  function. EM selects  $\theta_{n+1}$  as to maximize the function  $l(\theta|\theta_n)$  and each maximization will increase  $L(\theta)$ . In other words, EM is maximizing by maximizing the lower limit. For this reason, it can be said that EM is an indirect maximization method (Wu, 2009). Although there is no convergence theorem that guarantees the convergence of the EM, a critical point is achieved as expected. But this convergence can result in a local value, not a global one. The iteration method of operation is shown in Figure 2.

**Figure 2. Working Principle of EM, (Borman, 2009)**



From these obtained expressions, EM's prediction and maximization steps are as below (McLachlan and Krishnan, 2008).

*Estimation (expectation) step:*

Given,

$$\log L_c(\psi) = \log g_c(x; \psi) \quad (29)$$

expression below is calculated

$$Q(\psi; \psi^{(k)}) = E_{\psi^{(k)}}\{\log L_c(\psi)|y\} \quad (30)$$

*Maximization step:*

With the following expression, the maximization is processed,

$$\psi^{(k+1)} = \underset{\psi}{\operatorname{argmax}} Q(\psi; \psi^{(k)}) \quad (31)$$

In this sense, EM is similar to a hill climbing algorithm (Koller and Friedman, 2009).

EM intuitively works as follows: If all data were available,  $\theta'$  could be estimated to maximize  $\log f(X|\theta')$  function. However, when all data are not available, the expected value of  $\log f(X|\theta')$  is maximized when observed value and the updated value of  $\theta'$  are given (Ghahramani and Jordan, 1994). EM can be thought of as a special case of a more general framework called MM (Minorization-Maximization or Majorization-Minimization) algorithm (Demidenko, 2013).

*Advantages of EM:* The EM is an algorithm that is easy to apply in practice as well as easy to express mathematically and produces successful results in solving many problems. In addition, the EM can be used without consuming a lot of computer resources. Especially with exponential family distributions, the algorithm process becomes even simpler and very practical. EM is a numerically stable algorithm. At each iteration, the likelihood is increased and this is mathematically proven. Under general conditions, the EM gives a global extremum result.

*Disadvantages of EM:* The EM will make maximization, but this maximization is not guaranteed to be at a global critical point. However, there are several methods for achieving global maximum when using EM. The EM is a slow-running algorithm, which may require many iterations for convergence. When compared to Newton's method, each of the four EM repeats, in the beginning, corresponds to one repeat of Newton's method, after then EM is even more slowly than Newton's method (Jiang, 2007; Schlattmann, 2009). The expectation step in the EM cannot be solved analytically in some problems. In such cases, the Monte-Carlo approach can be used (Demidenko, 2013). The EM algorithm is very generic and applying it differs for different problems. There is no specific way to follow for each problem structure. When the EM algorithm is applied to missing data problems, the convergence slows down as the ratio of missing data increases. In addition, the convergence rate will decrease as algorithm converges to the real parameter values (Fahrmeir, Tutz, 2001). There is no standard procedure for obtaining standard errors in the algorithm process and only approximate value finding procedures are proposed. In practice, it has been indicated that there must be a large amount of data for the algorithm to work properly (Frühwirth-Schnatter, 2006). General convergence theory of the algorithm is not currently available (Kneib, Tutz, 2010).

### 2.4.2. Baum Welch Algorithm Review:

The BW algorithm is based on the EM which is a more general approach. The algorithm was developed for models with hidden variables by Baum *et al.* (1970) with the aim of estimating the parameters. In the Baum-Welch algorithm, the method in EM is followed: The auxiliary function of Baum *et al.* (1970) is given below as,

$$Q(\lambda, \lambda') = \sum_Q \log P(O, q|\lambda)P(O, q|\lambda') \quad (32)$$

This expression can be rewritten as:

$$\begin{aligned} Q(\lambda, \lambda') &= \sum_Q \log \pi_{q_0} P(O, q|\lambda') \\ &+ \sum_Q \left( \sum_{t=1}^T \log a_{q_{t-1}q_t} \right) P(O, q|\lambda') \\ &+ \sum_Q \left( \sum_{t=1}^T \log b_{q_t}(O_t) \right) P(O, q|\lambda') \end{aligned} \quad (33)$$

In the equation above, each term can be maximized according to  $\lambda$ . In this case, the following constraints must be satisfied:

$$\sum_{i=1}^N \pi_i = 1 \quad (34)$$

$$\sum_{j=1}^N a_{ij} = 1, 1 \leq i \leq N \quad (35)$$

$$\sum_{k=1}^M b_j(k) = 1, 1 \leq j \leq N \quad (36)$$

Each term to be maximized has a  $\sum_{j=1}^N w_j \log y_j$  structure with  $\sum_{j=1}^N y_j = 1$  and  $y_j \geq 0$  constraints and will have single global maximum point:

$$y_j = \frac{w_j}{\sum_{k=1}^N w_k}, j = 1, 2, \dots, N \quad (37)$$

Thus, the values sought can be found as below. Defining,

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (38)$$

and

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (39)$$

we can write,

$$\sum_{t=1}^{T-1} \gamma_t(i): \text{Expected number of transitions from } S_i \quad (40)$$

$$\sum_{t=1}^{T-1} \xi_t(i, j): \text{Expected number of transitions from } S_i \text{ to } S_j \quad (41)$$

$$\bar{\pi}_i: \text{Expected frequency at } S_i \text{ at time } t = 1 \quad (42)$$

Given,

$$\delta(o_k, v_k) = \begin{cases} 1, & o_k = v_k \\ 0, & \text{otherwise} \end{cases} \quad (43)$$

expressions below will be valid

$$\bar{a}_{ij} = \frac{\text{Expected number of transitions from } S_i \text{ to } S_j}{\text{Expected number of transitions from } S_i} \quad (44)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^T \gamma_t(j) \delta(o_k, v_k)}{\sum_{t=1}^T \gamma_t(j)} \quad (45)$$

$$\bar{\pi}_i = \gamma_1(i) \quad (46)$$

With these variables, the left side corresponds to the update, and the right side uses the current data.

### 3. EMPIRICAL STUDY ON SELECTED TIME SERIES

We applied<sup>3</sup> HMM to four time series. These are,

- Daily closing values of BIST 100 total return index (BIST100) with 3268 observations, dated between 03.01.2005-28.12.2017
- Daily US Dollar-Turkish Lira exchange rate (DT) with 3125 observations, dated between 05.08.2005-02.01.2018
- Daily US Dollar-Euro exchange rate (DE) with 3408 observations, dated between 03.01.2005-02.01.2018
- Daily Euro-US Dollar exchange rate (ED) with 3408 observations, dated between 03.01.2005-02.01.2018

BIST100 and DT series were downloaded from (TCMB, 2018). DE series were obtained from internet address <https://www.investing.com>.

The time series ED has been obtained by dividing one to each of the corresponding element of DE series, that is, we used the formula below for obtaining ED time series:

$$\{ED_i\} = \left\{\frac{1}{DE_i}\right\}, i = 1, \dots, T \quad (47)$$

ED series simply will not contain more information than DE. Our aim to use ED series which basically have the same information but is created by inverting by multiplication is to check the behaviour of HMM algorithm as if it is consistent in this manner. We expected that, after applying HMM procedure, we would roughly have an inverse character of state sequences between these two series.

In empirical part, all of the time series have been processed against the same HMM modelling procedure and this procedure follows several steps including reviewing data properties, processing data, and training then selecting the best model

*Reviewing data properties:* After applying HMM modelling steps, we have reviewed the results. We looked at several descriptive statistics for properties of data. These are given in Table 1. From Table 1, it can be seen that return series are moderate series to work with. BIST100 can be said to be a little bit more volatile than the others. All of the series have high peakedness with respect to the normal distribution. They all have approximately zero mean.

**Table 1. Descriptive Statistics of Four time Series**

	BIST100	DT	DE	ED
Time Series Length	3268	3125	3408	3408
Min	-10.4737	-11.2508	-3.6645	-2.7427
Max	12.8932	7.2924	2.8200	3.8039
Median	0.0960	-0.0131	0.0000	0.0000
Mean	0.0691	0.0373	0.0052	-0.0014
Var	2.7567	0.6930	0.3721	0.3726
St. Dev	1.6603	0.8325	0.6100	0.6104
Skewness	-0.1539	0.0859	-0.0654	0.1419
Kurtosis	6.8017	19.5953	5.1458	5.2465

*Processing data:* The original data and data transformed with logarithm base-10 function are shown in the first and second columns of Figure 3 respectively. In the

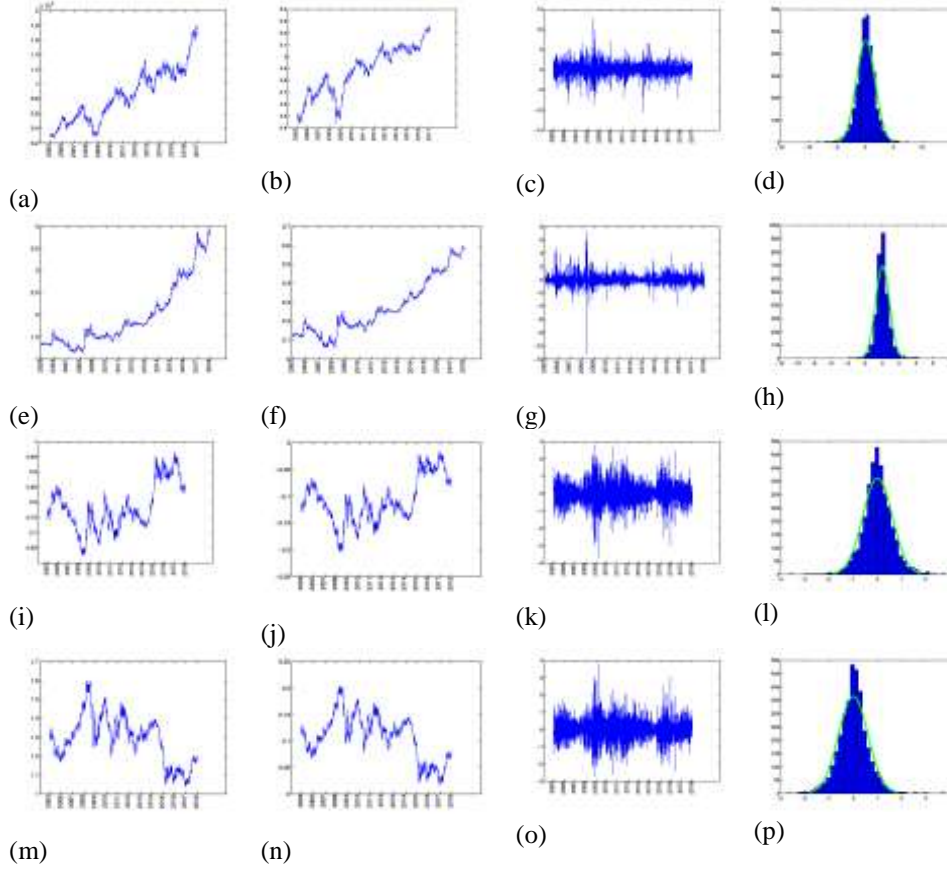
MSM and HMM studies, because of the statistical properties, the return data is used rather than original data. In this study also, return data is used. We started processing data by obtaining the return series. The formula we used for obtaining return data series is

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (48)$$

where  $P_t$  is the price or value at time  $t$  and  $R_t$  is the return ratio. Return series and histogram of return series with fitted normal can be seen at third and fourth columns in Fig. 3 respectively.

Since the model needs to be run with discrete data, we then applied vector quantization procedure to obtain the series of discrete data. In speech recognition problems where HMM mostly used, vector quantization is such a procedure that can be very complex and contains more than one data transformations (Rabiner *et al.*, 1983). Besides, it may contain other methods like k-means or neural network. In this study, we applied it for creating discrete data. This new time series has the same character as the return series' but it has less information.

**Figure 3. Data Properties of Time Series used in Empirical Analysis.**



Original data, log processed data, return data and histogram with fitted normal distribution are shown in each column for BIST100, DT, DE, ED time series respectively

The output of vector quantization stage is crucial for HMM to work properly. Vector quantization is the part for discretizing data, and it is the stage where we behaved differently from other econometric time series studies. We experienced that when quantization output is divided unequally per part, then modelling outputs in an unexpected manner. In previous studies, externally determined discreteness borders provided a better environment for assessment. Because then we could talk about 1%, 2% or 3% changes. But this approach can be criticized because of two reasons. Firstly, it is subjective since there is no reason to choose for example 2% but not 3% and secondly, for every different time series it should be decided which boundaries should



be chosen. So we design our vector quantization such that all parts have a nearly equal number of discrete data.

For this reason, a vector quantization was performed by means of first fitting a normal distribution to data and estimating parameters then using these parameters with an inverse function to obtain discrete data which falls to equally divided regions. This has allowed us to achieve a more homogeneous distribution to work with. After applying this technique we saw that model outputs became more meaningful. Equally dividing bounds can be seen in Table 2.

**Table 2. Bounds to be Used in Quantization**

Time series	1th q.	2nd q.	3rd q.	4th q.	5th q.
BIST100	-1.5371	-0.6460	0.0691	0.7843	1.6754
DT	-0.7680	-0.3212	0.0373	0.3959	0.8427
DE	-0.5850	-0.2576	0.0052	0.2679	0.5953
ED	-0.5919	-0.2643	-0.0014	0.2615	0.5891

*Training then selecting best model:*

For training data, we used Baum-Welch algorithm. Best model has been chosen according to BIC. For deciding most appropriate model, 5 different models (with 2 to 6 states) have been used, and the model with lowest BIC value has been chosen. According to Costa and De Angelis (2010) study, BIC is the most appropriate criterion for our model regarding the number of data, state-dependent probabilities and latent transition matrix.<sup>4</sup>

For the best model and the others models we conducted in the study, the most appropriate sequence is estimated by the Viterbi algorithm.

*Review of findings:* We have chosen the most appropriate model and specifically assessed this most appropriate model. Since it is informative, the outputs of models other than best model, have also been inspected and assessed. The model outputs and the assessments are given in Sect. 4.1, 4.2, 4.3, 4.4.

Within figures in these subsections, all the red lines correspond to the time series which are transformed by taking the logarithm of original series, blue lines show states, green lines are means of observations related by their states. Measurement at the left side of the graphic shows the value of states and also means of observations multiplied by ten. We signed the assessment numbers with "\*" character where we used more than one model of different states for that assessment.

For all models, the most appropriate model has been selected using BIC criterion. Log-likelihood, degrees of freedom, AIC and BIC criteria values can be seen in Tables 3, 4, 5, 6 in each subsection. Using BIC we have chosen 3-state model as the most appropriate model for all time series.

### 3.1. Important Dates and Comments for the BIST 100 Index

**Table 3. Statistics for BIST100**

#states	LL (x1.0e+03)	DF	AIC(x1.0e+04 )	BIC(x1.0e+04 )
2	-5.7043	2	1.1413	1.1425
3	-5.6823	6	1.1377	1.1413
4	-5.6769	12	1.1378	1.1451
5	-5.6656	20	1.1371	1.1493
6	-5.6734	30	1.1407	1.1590

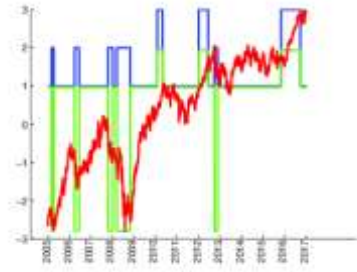
Log-likelihood func., degrees of freedom, AIC and BIC criteria for BIST100 regarding 2 to 6 hidden state models

In model with 6 states, states were not assigned to all observations

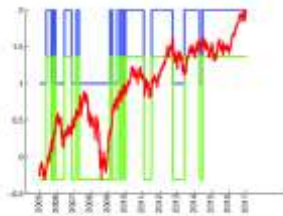
- The 3-state model which is the most appropriate model chosen by BIC seems to detect the trends well.

- According to 3-state model, the global crisis which is started to be felt in 2007 but officially associated to Lehman Brothers' fail that is at the end of 2008 is recognized by the model: There is transition to bad regime at the beginning of 2008 which is short-lived (includes January, February and March) and then there is again transition to a bad regime which persists from mid-2008 until the beginning of 2009 (persists 8 months). This second transition to the bad regime can directly be associated with the effect of the global crisis on the stock markets. Although the crisis is said to be started at the end of 2008, the first transition to the bad regime is also related to the bad news from the USA. Because of this bad news, the stock indexes in Oslo, Vienna, Prague and Istanbul markets dropped more than 10% in first half of the January 2008.

Figure 4. Outputs for BIST100



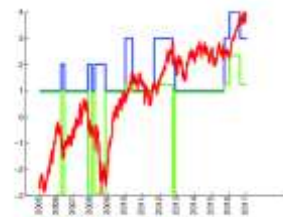
(a) 3-state model (the most appropriate model)



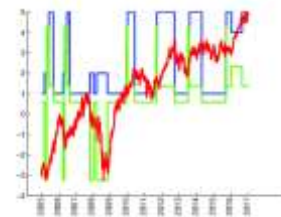
(b) 2-state model



(c) 4-state model



(d) 5-state model



(e) 6-state model

(a) being most appropriate model, 2 to 6 state models shown with log processed time series, related state numbers and means multiplied by ten

- \*The effect of Gezi protests has been seen most at the stock market index among other economic indicators in Turkey. The stock market fell during the time of protests around June 2013. When we look at the related figures from different models there is a transition to a bad regime in all of the models at that time. The effect is short in some of the models and lasts a little bit longer in others. According to the 3-state model which best fits data, we see that effect persists from mid-May to mid-July (nearly 2 months).

- \*The recent appreciation of BIST100 (probably by the effect of the expectation of the inflation), due to the recent increase in Euro and Dollar value against the Turkish

Lira seems to be recognized by the system. There is a transition to the good regime starting from the end of September 2016 according to 3-state model.

### 3.2. Important dates and Comments for the Dollar / Turkish Lira Parity

**Table 4. Statistics for DT**

#states	LL (x1.0e+03)	DF	AIC (x1.0e+04 )	BIC (x1.0e+04 )
2	-5.3434	2	1.0691	1.0703
3	-5.3040	6	1.0620	1.0656
4	-5.2877	12	1.0599	1.0672
5	-5.2815	20	1.0603	1.0724
6	-5.2727	30	1.0605	1.0787

Log-likelihood func., degrees of freedom, AIC and BIC criteria for DT regarding 2 to 6 hidden state models

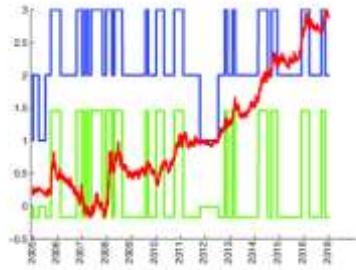
All observations assigned to all states in all models

- The 3-state model which is the most appropriate model chosen by BIC seems to detect the trends fairly.

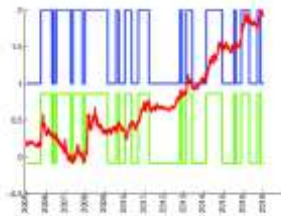
- \*In models which have four or more states, regime changes are frequent.

- It is clear that the 6-state model results are tottering. This can possibly be perceived as an indication of ineffective usage of HMM.

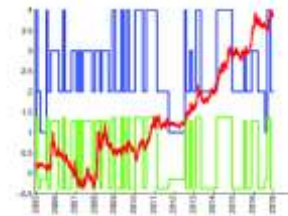
Figure 5. Outputs for DT



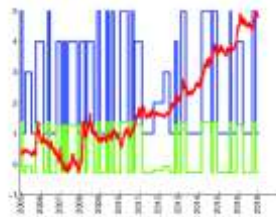
(a) 3-state model (the most appropriate model)



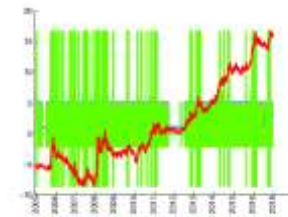
(b) 2-state model



(c) 4-state model



(d) 5-state model



(e) 6-state model

(a) being most appropriate model, 2 to 6 state models shown with log processed time series, related state numbers and means multiplied by ten

### 3.3. Important Dates and Comments for the Dollar / Euro Parity

**Table 5. Statistics for DE**

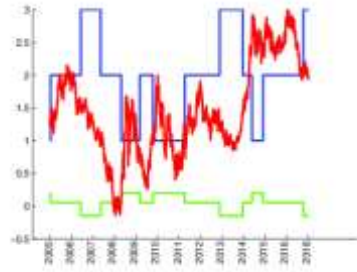
#states	LL (x1.0e+03)	DF	AIC (x1.0e+04 )	BIC (x1.0e+04 )
2	-5.9626	2	1.1929	1.1941
3	-5.9323	6	1.1877	1.1913
4	-5.9330	12	1.1890	1.1964
5	-5.9153	20	1.1871	1.1993
6	-5.9045	30	1.1869	1.2053

Log-likelihood func., degrees of freedom, AIC and BIC criteria for DE regarding 2 to 6 hidden state models

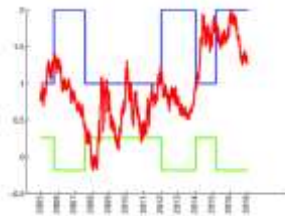
In model with 6 states, not all states were assigned to observations

- The 3-state model which is the most appropriate model chosen by BIC seems to detect the trends fairly.
- \*For the best model and other appropriate models means for each regime are almost the same, especially in models with few states including the best model which is the 3-state one.
- \*In all models numbers of regime changes are low and differences of means of regimes are small. We see that regimes are relatively stable and persistent.
- \*Generally, as the number of parameters increases, models display a harmonious detail.
- The 6-state model does not use all the states. It means that this model is inappropriate.
- \*The effect of the credit boom in the USA through years 2006 and 2007 seems to be recognized by all models as Euro gaining value. But only 2-state model consistently indicate that the regime is in favour of Euro through those years.

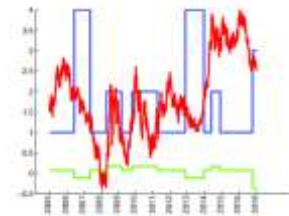
Figure 6. Outputs for DE



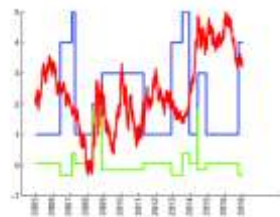
(a) 3-state model (the most appropriate model)



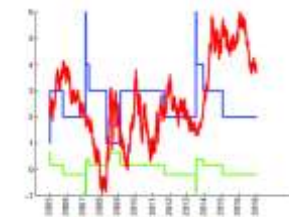
(b) 2-state model



(c) 4-state model



(d) 5-state model



(e) 6-state model

(a) being most appropriate model, 2 to 6 state models shown with log processed time series, related state numbers and means multiplied by ten

### 3.4. Important Dates and Comments for the Euro / Dollar Parity

**Table 6. Statistics for ED**

#states	LL (x1.0e+03)	DF	AIC (x1.0e+04 )	BIC (x1.0e+04 )
2	-5.9631	2	1.1930	1.1942
3	-5.9339	6	1.1880	1.1917
4	-5.9333	12	1.1891	1.1964
5	-5.9166	20	1.1873	1.1996
6	-5.9336	30	1.1927	1.2111

Log-likelihood func., degrees of freedom, AIC and BIC criteria for ED regarding 2 to 6 hidden state models

In model with 6 states, not all states were assigned to observations

- The 3-state model which is the most appropriate model chosen by BIC seems to detect the trends fairly.

- \*For the best model and other appropriate models means for each regime are almost the same, especially in models with few states including the best model which is the 3-state one.

- \*In all models numbers of regime changes are low and differences of means of regimes are small. We see that regimes are relatively stable and persistent.

- \*Generally, as the number of parameters increases, models display a harmonious detail.

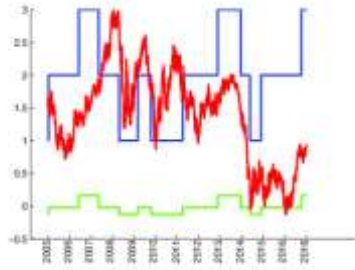
- The 6-state model does not use all the states. It means that this model is inappropriate.

- \*For the most appropriate model which is the 3-state model, Euro/Dollar and Dollar/Euro parity outputs complement each other, that is to say, they are inverse in good/bad manner. Outputs are again complementary for models with 2, 4, 5 states. For the 6-state model where not all states are used, outputs are not complementary. Actually, the model with 6 states for ED time series is inappropriate as it is for DE time series.

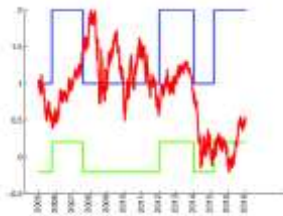
- Quantitative easing, which started on March 2015, is priced a couple of months in advance according to the model.



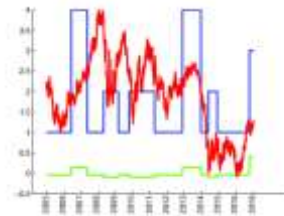
Figure 7. Outputs for ED



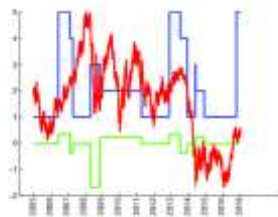
(a) 3-state model (the most appropriate model)



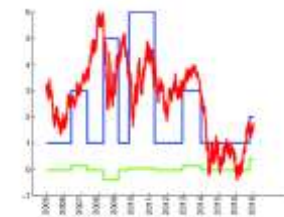
(b) 2-state model



(c) 4-state model



(d) 5-state model



(e) 6-state model

(a) being most appropriate model, 2 to 6 state models shown with log processed time series, related state numbers and means multiplied by ten

#### 4. DISCUSSION

From results of the empirical study, it can be seen that the regime switching dates are located correctly. Additionally, it should be noted that the averages of the return series are different from each other, although the difference is minor.

HMM is a very flexible algorithm and can be used in many areas. It recognizes the regime changes, which is a great advantage in obtaining information about the data. Taking regime changes into consideration in the model is one of the most successful statistical modelling ideas, as previously mentioned by Cappé *et al.* (2005). Prediction

phase also benefits from regime information. In this study it is observed that for most of the models, trends are recognized and segmented with regime changes correctly. Another advantage of HMM is that there are a number of software and package programs capable of using HMM.

A well-known disadvantage of HMM is that it is a slow running algorithm. When using HMM, modelling process may take a long time, especially where the data set is large. It took us several minutes for completing process of a single data set we used. Hence HMM is not an appropriate algorithm where the solution is needed instantly. As a second disadvantage, it is unclear which model to choose. Typically information criteria are generally used in model selection.

Thirdly, it can be seen from the figures that sometimes models with different state numbers may indicate different good/bad market conditions for the same time periods. That is to mean, as the number of states in model increases, some models may not detail harmoniously. Then their assessment can be considered subjective. We think it's a good idea to look at all the appropriate models in this case. Using the model which is the most compatible with historical events could be an alternative strategy for assessment and making decisions. Fourthly, a disadvantage we experienced in modelling HMM is that the model results were highly affected by the vector quantization phase of the data. For this reason, it is important to distribute data as evenly as possible as we noted in Sect. 4. Lastly, segments determined by the model are not created with predetermined averages (or trend properties in original series). Sometimes means of divisions are very close, which may seem to be not a very useful piece of information. Nevertheless, the model recognizes data trends (as an advantage of the model) and produces meaningful outcomes even if means of the divisions are very close.

## CONCLUSION

Markets are generally said to be in one of two conditions: good or bad (generally bull and bear terms used for indicating market condition). Sometimes it is more appropriate to know, how good or how bad it is with handling and measuring with more than two criteria.

HMM identified market conditions and did it with segmenting more than two conditions. For showing this we used discrete HMM which was densely used in speech recognition in 1980's. We conclude that this type of HMM application can handle all kinds of regime changes in general, thus can be used where different regimes are searched. The regimes, being known, will provide an improvement in the prediction of time series.

## NOTES

---

<sup>1</sup> With ratios 98.1% versus 97.1%. These values can be considered indicating almost the same performance.

<sup>2</sup> There are few studies that used discrete HMM for making predictions. But they didn't explicitly research for usability of this model directly for regimes. Although model is not suitable for predictions, we claim that discrete HMM is an excellent model for determining several level of regimes.

<sup>3</sup> Applications were implemented on Matlab R2013a program using Statistics and Machine Learning Toolbox and HMM toolbox obtained from <https://www.cs.ubc.ca/~murphyk/Software/> address.

<sup>4</sup> According to aforementioned study of simulation, for HMMs with equally distributed conditional probabilities (shown as HMM\_VI in Table 1.) BIC predicts correct state numbers with a rate of 93.3% and for HMMs with persistent transition probabilities (shown as HMM\_A in Table 2.) BIC predicts correct state numbers with a rate of 100% when T is large.

## REFERENCES

- Ailliot, P., V. Monbet (2012), "Markov-Switching Autoregressive Models for Wind Time Series", *Environmental Modelling & Software*, 30, 92–101.
- Baum, L., T. Petrie, G. Soules, N. Weiss (1970), "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", *Ann. Math. Statist.*, 41, 164–171.
- Bhar, R., S. Hamori (2004), *Hidden Markov Models: Applications to Financial Economics, Advanced Studies in Theoretical and Applied Econometrics*, v. 40, Boston, Mass. and London: Springer US.
- Bilmes, J. (1998), "A Gentle Tutorial of the Em Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models", *Technical Report ICSI-TR-97-21*, International Computer Science Institute, Computer Science Division, U.C. Berkeley.
- Borman, S. (2009), "The Expectation Maximization Algorithm - A short Tutorial", URL <http://www.seanborman.com/publications/>, Date of Access: 18.03.2018.
- Cappé, O., E. Moulines, T. Rydén (2005), *Inference in Hidden Markov Models, Springer Series in Statistics*, New York: Springer Verlag.
- Ceppellini, R., M. Siniscalco, C.A. Smith (1955), "The Estimation of Gene Frequencies in a Random-Mating Population", *Ann. Hum. Genet.*, 20, 97–115.
- Collins, M. (1997), "The EM Algorithm", *Technical Report*, Department of Computer and Information Science, University of Pennsylvania.
- Costa, M. and L. De Angelis (2010), "Model Selection in Hidden Markov Models: A Simulation Study", *Quaderni di Dipartimento 7*, Department of Statistics, University of Bologna.
- De Angelis, L., L.J. Paas (2013), "A Dynamic Analysis of Stock Markets Using a Hidden Markov Model", *Journal of Applied Statistics*, 40, 1682–1700.

- Demidenko, E. (2013), *Mixed Models: Theory and Applications With R*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2nd edition.
- Dempster, A.P., N.M. Laird, D.B. Rubin (1977), “Maximum Likelihood from Incomplete Data Via the Em Algorithm”, *Journal of the Royal Statistical Society. Series B* (Methodological), 39, 1–38.
- Do, C., S. Batzoglou (2008), “What is the Expectation Maximization Algorithm?”, *Nature Biotechnology*, 26, 897–899.
- Fahrmeir, L., G. Tutz (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Series in Statistics, New York: Springer, 2nd edition.
- Fink, G.A. (2007), *Markov Models for pattern Recognition: From Theory to Applications*, Berlin, Heidelberg: Springer-Verlag.
- Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, Springer Series in Statistics, New York: Springer.
- Ghahramani, Z.,d M.I. Jordan (1994), “Supervised Learning from Incomplete Data via an Em Approach”, in *Advances in Neural Information Processing Systems 6*, Morgan Kaufmann, 120–127.
- Giudici, P., T. Rydén, P. Vandekerkhove (2000), “Likelihood-Ratio Tests for Hidden Markov Models”, *Biometrics*, 56, 742–747.
- Gupta, M., Y. Chen (2010), “Theory and use of the EM Algorithm”, *Foundations and Trends in Signal Processing*, 4, 223–296.
- Hamilton, J.D. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”, *Econometrica*, 57, 357–384.
- Hamilton, J.D. (2010), *Regime Switching Models*, London: Palgrave Macmillan UK, 202–209.
- Hamilton, J.D., B. Raj (2002), “New Directions in Business Cycle Research and Financial Analysis”, *Empirical Economics*, 27, 149–162.
- Hartley, H.O. (1958), “Maximum Likelihood Estimation from Incomplete Data”, *Biometrics*, 14, 174–194.
- Jamshidian, M., R.I. Jennrich (1997), “Acceleration of the EM Algorithm by using Quasi-Newton Methods”, *Journal of the Royal Statistical Society. Series B* (Methodological), 59, 569–587.
- Janczura, J., R. Weron (2010), “An Empirical Comparison of Alternate Regimes Switching Models for Electricity Spot Prices”, *Energy Economics*, 32, 1059–1073.
- Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer Series in Statistics, New York and London: Springer.
- Juang, B.H., L.R. Rabiner (1985a), “Mixture Autoregressive Hidden Markov Models for Speech Signals”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33, 1404–1413.
- Juang, B.H., L.R. Rabiner (1985b), “A Probabilistic Distance Measure for Hidden Markov Models”, *AT&T Technical Journal*, 64, 391–408.
- Kneib, T., G. Tutz (2010), *Statistical Modelling and Regression Structures: Festschrift in Honour of Ludwig Fahrmeir*, Heidelberg [u.a.]: Physica-Verl.

- Kobayashi, H., B.L. Mark, W. Turin (2012), *Probability, Random Processes, and Statistical Analysis*, Cambridge and New York: Cambridge University Press.
- Koller, D., N. Friedman (2009), *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*, The MIT Press.
- Levinson, S.E., L.R. Rabiner, M.M. Sondhi (1983), “An Introduction to the Application of the Theory of Probabilistic Functions of A Markov Process to Automatic Speech Recognition”, *The Bell System Technical Journal*, 62, 1035–1074.
- Lindgren, G. (1978), “Markov Regime Models for Mixed Distributions and Switching Regressions”, *Scandinavian Journal of Statistics*, 5, 81–91.
- Louis, T.A. (1982), “Finding the Observed Information Matrix when Using the EM Algorithm”, *Journal of the Royal Statistical Society. Series B (Methodological)*, 44, 226–233.
- McLachlan, G.J., T. Krishnan (2008), *The EM Algorithm and Extensions*, Wiley Series in Probability and Statistics, Hoboken and NJ: Wiley-Interscience, 2nd edition.
- Moon, T.K. (1996), “The Expectation-Maximization Algorithm”, *IEEE Signal Processing Magazine*, 13, 47–60.
- Poritz, A. (1982), “Linear Predictive Hidden Markov Models and the Speech Signal”, in *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 7, 1291–1294.
- Rabiner, L.R. (1989), “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, *Proceedings of the IEEE*, 77, 257–286.
- Rabiner, L.R., B.H. Juang (1986), “An Introduction to Hidden Markov Models”, *IEEE ASSP Mag.*, 3, 4–16.
- Rabiner, L.R., S.E. Levinson, M.M. Sondhi (1983), “On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition”, *The Bell System Technical Journal*, 62, 1075–1105.
- Rydén, T., T. Teräsvirta, S. Åsbrink (1998), “Stylized Facts of Daily Return Series and the Hidden Markov Model”, *Journal of Applied Econometrics*, 13, 217–244.
- Schlattmann, P. (2009), *Medical Applications of Finite Mixture Models, Statistics for Biology and Health*, Berlin and Heidelberg: Springer-Verlag Berlin Heidelberg.
- Tanner, M.A. (1996), *Tools for Statistical Inference: Methods for the Explorations of Posterior Distribution and Likelihood Functions*, Springer Series in Statistics, New York: Springer, 3rd edition.
- TCMB (2018), <https://evds2.tcmb.gov.tr/index.php> E.T.: 18.05.2018.
- Vaseghi, S.V. (2007), *Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications*, Chichester and West Sussex and England and Hoboken and NJ: J. Wiley.
- Viterbi, A. (1967), “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm”, *IEEE Trans. Inf. Theor.*, 13, 260–269.
- Wu, C.F.J. (1983), “On the Convergence Properties of the EM Algorithm”, *Ann. Statist.*, 11, 95–103.
- Wu, L. (2009), *Mixed effects Models for Complex Data*, C & H/CRC Monographs on Statistics & Applied Probability, 113, v. 113, Hoboken: Chapman & Hall/CRC.