International Journal of Economic Perspectives, 2016, Volume 10, Issue 2, 117-133.

The Analysis of Foreign Direct Investment with Hidden Markov Model: Evidence from Turkey

Naciye Tuba YILMAZ*
Marmara University, Department of Econometrics, Turkey.
Email: tuba.yilmaz@marmara.edu.tr.

Tuncay CAN
Marmara University, Department of Econometrics, Turkey.
Email: tuncay.can@marmara.edu.tr.

ABSTRACT

Hidden Markov Models are based on Markov chains and enabled to forecast of hidden states and observations with data in the past. Generally economic indicators are analyzed with some econometric applications and Hidden Markov model has limited scope for analyzing them. The aim of this study is to demonstrate the applicability of Hidden Markov models could be applied on the foreign direct investment, which has an undeniable importance on economic phenomenon and contributes savings of developing countries. It also provides improvements in the macroeconomic indicators and development of the countries. The case of Turkey in 2009:09 – 2013:12 period is used in order to identify the hidden states for hidden Markov model accurately. Then, an econometric analysis is made on the determinants of a country took place in the theory. Logarithmic - linear model is estimated as it is an appropriate model with independent variables; foreign trade volume, the unemployment rate and the CPI. After the end of the data range, with the three main problems of Hidden Markov models, foreign direct investment coming into the country (observations) and its determinants (hidden states) are estimated with the previous month values. Besides, the model parameters are optimized for the desired sequence of the observation. Estimated results are compared with the actual values and they are interpreted. Under the case of having a big data and identified hidden states correctly, hidden Markov models can be said to give accurate results for longer periods.

JEL Classifications: C1; C4; C44; F20.

Key Words: Hidden Markov Models; Foreign Direct Investment; Turkey; Operations Research.

**Corresponding author; Naciye Tuba YILMAZ*
Marmara University, Department of Econometrics, Turkey; Email: tuba.yilmaz@marmara.edu.tr.

1. INTRODUCTION

Stochastic process is a mathematical process which consists of X(t) random variables as elements and allows for interpretation for future under uncertainty. Markov chain was, on the other hand, developed by Andrei Andreyevich Markov in early 20th century and refers to a method developed to estimate the condition of the process in the next step or with a long-term analysis when the current state of the situation is known. This method states that all features of the process in the past have impact on its current state for which reason the trajectory on which future prediction will be made is just dependent on the current state. In hidden Markov models, in addition to states, observation outputs generated by the signals of these states are also included. In the current time, predicting the future is aimed in the light of the values assumed by observed outputs as a result of the signals transmitted by states. States constitute series of observation with certain probabilities. In addition to the parameters of Markov chains, there are state transition probability matrixes, observation probabilities matrix which consists of symbol values taken by states and observation values that correspond to these values, and inception state probabilities in hidden Markov models. The importance of foreign direct investments (FDI) as a growth element of an economy is increasing especially for developing or under developed countries on a daily basis. FDI that a host economy attracts to its country is an important resource in that the economy can be seen as a promising one and that, when investments are evaluated well, they have positive impact on growth and development. When the theory is examined, the existence of "determinants" which attract FDI to the country attract the attention. The size of an economy, positive basic macroeconomic indicators, qualification, ratio and costs of the workforce, infrastructure facilities, transparency in administration, the level of taxes that the institutions will face as cost (tax

burden), openness and closeness of the country to foreign markets and its location against raw material sources can be listed as important factors for international companies in their selection of countries for investment. It should be noted that every investment has alternative costs in terms of the investor and country of investment. However, the purpose here is to provide optimum benefit both for the investor and country of investment.

2. LITERATURE REVIEW

Hidden Markov models has a wide area of study such as speech and character recognition, molecular biology, and engineering; they also attract attention in recent years especially in their use with economy and finance. Elliott and Hoek (1997) performed an application to asset allocation problem with hidden Markov model. In the study, a model was offered for return ratios in terms of observable and non-observable factors, an algorithm was presented using filtering and prediction techniques so as to define the model and using average-variance type benefit criteria, it was shown how this method could be applied to a strategic asset allocation problem. Pardo and Birmingham (2005) created hidden Markov models where musical scores were hidden states and musical performances were observations. Can and Oz (2009) predicted with hidden Markov model the change in USD for the first two and first three months periods of 2008 based on monthly foreign exchange rates observed in Turkey between 1997 and 2007 and balance of payments and interest rate variables which affect these exchange rates. Champion and Houghton (2015) performed an application with hidden Markov model with constant state to speech recognition problem. Ji, Quanmin and Dongya (2015) adopted hidden Markov model so as to predict probabilities for location of soccer robot's trajectories in their studies, developed an algorithm and showed the applicability of the procedure with a simulation.

Kar and Tatlısöz (2008) performed an econometric analysis with Least Square Method (LSM) on the determinants of FDIs in Turkey. They concluded that international net reserves, gross national product (GNP), ratio of openness, electrical energy production index and investment incentives had positive impact on FDIs but real exchange rate and workforce costs had negative impact on the same. Tarı and Bıdırdı (2009) analysed the factors that affected FDI in Turkey between 1990 and 2006 using Johansen multiple co-integration prediction method and predicted a model for long-term FDI entries. They also used an error correction model in order to study short-term relations. They concluded that in the long run market size of direct investments in Turkey, wages and openness index were high compared to economic stability, and that GDP and openness ratio had positive impact on FDIs and workforce costs and economic stability had negative impact on such. Sarısoy and Koç (2011) conducted panel regression analysis on 21 OECD countries between 1981-2008 in order to find the FDI impact on corporations tax collection of host countries and concluded that one or two years after FDI, it had positive impact on corporation tax collection of the host country.

Ekinci (2011) performed a causality test in order to examine the long-term relation between growth and employment and FDI and found that there is a cointegration causality with economic growth whereas they did not find any causality relation between employment and FDI. Güriş, Saçaklı and Genç (2015) analysed the factors that affect FID with Panel Torbit for the entire period so as to examine the impact of global crisis in 2008 on FDI for world countries where FDI was at high levels for the years 2005-2011, and they analysed with Tobit methods for each year in order to perform detailed analysis depending on level of development. According to the prediction results, they concluded that workforce and inflation ratios had significant impact on the estimated model for the entire period but they did not have significance for prediction results of individual years and that current account balance was the most determinant factor for countries at all development levels and that it was not affected by the crisis period.

Hidden Markov Models

Hidden Markov models were first studied in late 1960s and early 1970s; however, they gained popularity in early 1980s (Rabiner, 1989, p.257). They have a strong mathematical infrastructure for which reason they are used in a wide area of application. Hidden Markov models are layered stochastic processes which are not directly observable but produce observation series with other stochastic process cluster (Fuh, 2003, s.942). In other words, states generate observation outputs with certain probabilities and outputs come together to create an observation series. Usually hidden Markov model is interested in first-order transition probabilities between hidden states and observations are affected by hidden states but not vice versa (Ching and Ng, 2006, p.170). However, there are some interactive hidden Markov models where hidden states are affected by observations such as high-degree Markov model.

In Markov chains, states and sometimes initial state probabilities are known. If it is known that the system is in i state at t time, the observation is only i state and it is the movement of the system observed at $t+1$ time from i state to any other state $\{1,2,\dots,\mathcal{N}\}$. In hidden Markov model, the symbol $\mathcal{N}$ refers to hidden states in the model and expressed as independent states $\mathcal{S}=\{\mathcal{S}_1,\mathcal{S}_2,\dots,\mathcal{S}_{\mathcal{N}}\}$. Although these states are hidden, in most applications states have a physical meaning (Ching and Ng, 2006, p.170). There are transitions between states and the state at t time is expressed as q_t. $\mathcal{M}$ is expressed as the number of distinct observation symbols per state. Observations correspond to the physical outputs of the observed system and are expressed in the form of $\mathcal{V}=\{\mathcal{V}_1,\mathcal{V}_2,\dots,\mathcal{V}_{\mathcal{M}}\}$ which is a discrete set. The observation created by each state is independent from other observations. In hidden Markov models, the state of the system in any t time is unknown, whereas the observations generated by states are known. This means that when the system is in any state, any of $\mathcal{V}=\{\mathcal{V}_1,\mathcal{V}_2,\dots,\mathcal{V}_{\mathcal{M}}\}$ $\mathcal{M}$ number of observations occur depending on the probability function (Öz, 2009, p.37). State transition probability distribution is expressed as $\mathcal{A}=\{a_{ij}\}$ where $\mathcal{A}$ is a matrix with the size of $\mathcal{N}x\mathcal{N}$. a_{ij} probability is defined as equation [1] where $0\leq a_{ij}\leq 1 \quad i,j=1,2,\dots,\mathcal{N}$ and $\sum_{j=1}^{\mathcal{N}}a_{ij}=1;\ i=1,2,\dots,\mathcal{N}$

$$a_{ij}=\mathcal{P}(q_{t+1}=\mathcal{S}_j \setminus q_t=\mathcal{S}_i) \quad ; \quad i,j=1,2,\dots,\mathcal{N} \qquad [1]$$

As the case in Markov chains, a_{ij} defines a_{ij}; gives the possibility of a process in $\mathcal{S}$ state at t time to be in $\mathcal{S}_j$ state at $t+1$ time. State transition probabilities do not change in time and these probabilities are independent from observations (Bhar and Hamari, 2004, p.17). Observation probability distribution in j state is shown as $\mathcal{B}=\{b_j(k)\}$ where $\mathcal{B}$ is a matrix at $\mathcal{N}x\mathcal{M}$ size. B is a matrix of observation probability distribution at any state j. Its mathematical expression is that equation [2]

$$b_j(k)=\mathcal{P}(\mathcal{O}_t=v_k \setminus q_t=\mathcal{S}_j)\ ;\ j=1,2,\dots,\mathcal{N}\ and\ k=1,2,\dots,\mathcal{M} \qquad [2]$$

$b_j(k)$ refers to the possibility of a process which is known to be in $\mathcal{S}_j$ state at t time to create observation k at t time. Initial state distribution $\Pi=\{\pi_i\}$ where $\pi_i=\mathcal{P}(q_1=\mathcal{S}_i);\ i=1,2,\dots,\mathcal{N}$ showing the possibility of being in i state at the beginning. Hidden Markov model creates the observation series $\mathcal{O}=\mathcal{O}_1,\mathcal{O}_2,\dots,\mathcal{O}_{\mathcal{T}}$ for the given appropriate values of $\mathcal{N},\mathcal{M},\mathcal{A},\mathcal{B}$ and Π. $\mathcal{T}$ being the number of observations, each $\mathcal{O}_t$ observation is one of $\{v_1,v_2,\dots,v_{\mathcal{M}}\}$ observations (Rabiner, 1989, p.261). Parameter set of a hidden Markov model is expressed as $\Lambda=(\mathcal{A},\mathcal{B},\Pi)$. According to this definition, first order Markov process is mostly used to model the transitions between hidden states in a hidden Markov model (Ching and Ng, 2006, p.172). State sequence which constitutes observable data is discrete; thus, these hidden Markov models are defined as "discrete".

Five steps can be followed in order to be able to use hidden Markov model as a generator of observation series. The first step is to choose the initial state $q_1=\mathcal{S}_i$ according to initial state distribution Π. The second step is taking time, t, as 1. The third step is to choose $\mathcal{O}_t=v_k$ observation according to observation probability distribution in $\mathcal{S}_i$ state. This means that the observation created by the process in $\mathcal{S}_i$ state according to $b_j(k)$ is chosen as v_k observation. The fourth step is transition to a new state for $t+1$ time according to state transition probabilities distribution. This means that, taking the matrix $\mathcal{A}$ into consideration, there is a_{ij} probability to transit to $q_{t+1}=\mathcal{S}_j$. In the fifth and final step $t=t+1$ is taken. If $t<\mathcal{T}$, we go back to step three and repeat the other steps. Otherwise the procedure is finalised. This five-step procedure is used as both an observation generator and a model for a certain observation series created by an appropriate hidden Markov model (Rabiner, 1989, p.261). In order to define a hidden Markov model, two model parameters such as $\mathcal{N}$ and $\mathcal{M}$, observation symbols and three probability measures $\mathcal{A},\mathcal{B}$ and Π has to be defined. Some assumptions are used in the theory of hidden Markov model for convenience in mathematical and calculative procedures: Markov Assumption: in any hidden Markov model, transition probabilities are defined as equation [3].

$$a_{ij}=\mathcal{P}\{q_{t+1}=\mathcal{S}_j \setminus q_t=\mathcal{S}_i\} \quad ; \quad i,j=1,2,\dots,\mathcal{N} \qquad [3]$$

In widely used first order hidden Markov models, the process is independent from all states in the past but depends only on the current state, which is the case in Markov chains. In k order hidden Markov models the process depends on the past k state and transition probabilities are defined as equation [4] where $i_1,i_2,\dots,i_k,j=1,2,\dots\mathcal{N}$

$$a_{i1,i2,\dots,ik,j}=\mathcal{P}\{q_{t+1}=\mathcal{S}_j \setminus q_t=\mathcal{S}_{i1},q_{t-1}=\mathcal{S}_{i2},\dots,q_{t-k+1}=\mathcal{S}_{ik}\} \qquad [4]$$

Stability assumption: In this assumption, it is expressed that states are independent of time. This means that state transition probabilities are related to states but independent from time. If this is to be expressed mathematically, the equation [5] can be generated.

$$\mathcal{P}\left(q_{t1+1} = \mathcal{S}_j \setminus q_{t1} = \mathcal{S}_i\right) = \mathcal{P}\left(q_{t2+1} = \mathcal{S}_j \setminus q_{t2} = \mathcal{S}_i\right) \quad For \ \forall \ t_1 \ and \ t_2 \qquad [5]$$

The possibility that a process in $\mathcal{S}_i$ state at t_1 time to be in $\mathcal{S}_j$ state one step later is identical to the possibility of a process in $\mathcal{S}_i$ state at t_2 time to be in $\mathcal{S}_j$ state one step later. *The assumption that observations are independent:* this means that current observation is unrelated to the previous observations, that is, independent statistically. Observation series being $\mathcal{O} = \mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_\mathcal{T}$, the equation [6] can be generated according to the valid assumption.

$$\mathcal{P}(\mathcal{O} \setminus q_1, q_2, \dots, q_t, \Lambda) = \prod_{t=1}^{\mathcal{T}} \mathcal{P}(\mathcal{O}_t \setminus q_t, \Lambda) \qquad [6]$$

According to other two assumptions, the validity of this assumption is limited. In some states this assumption may not be verified which shows the strong weakness of Markov models (Öz, 2009, p.44). Using hidden Markov models is a quest for answering three basic questions. The first problem is the effort to find an answer to the question of the level of match between observation series and the model. The probability of observation series generated by the model is calculated with the given model and observation series. With second problem effort is paid to reveal the hidden part of the model meaning accurate state series. Optimality criteria are used to solve this problem. However finding what the optimality criteria and choosing among several optimality criteria is not easy. For this reason the optimum level is targeted. With the third problem it is ensured that model parameters are optimized so that the obtained observation series can be generated. This problem is critical for several hidden Markov model applications as it optimally adjusts model parameters with observed data (Rabiner, 1989, p.261). The first problem is founded on the most effective calculation of $\mathcal{P}(\mathcal{O}, \Lambda)$ probability where $\Lambda = (\mathcal{A}, \mathcal{B}, \Pi)$ is given model parameters and $\mathcal{O} = \mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_\mathcal{T}$ is observation series. It calculates the generation possibility of an observation in observation series by the model. The frist problem is also called evaluation problem. In addition it can be used for choosing the model which suits best to the observations (Blunsom, 2004, p.3).

Each state in a series of states at the length of $\mathcal{T}$ are numbered as $1,2, \dots, \mathcal{T}$ in order to solve the problem. This way the state series becomes $\mathcal{Q} = q_1, q_2, \dots, q_\mathcal{T}$ and q_1 represents initial state. For this series of state, the probability of observation series is in the form of equation [7] based on the statistical independence of observations and this equality can be written clearly with the equation [8].

$$\mathcal{P}(\mathcal{O} \setminus \mathcal{Q}, \Lambda) = \prod_{t=1}^{\mathcal{T}} \mathcal{P}(\mathcal{O}_t \setminus q_t, \Lambda) \qquad [7]$$

$$\mathcal{P}(\mathcal{O} \setminus \mathcal{Q}, \Lambda) = \mathcal{b}q_1(\mathcal{O}_1).\mathcal{b}q_2(\mathcal{O}_2) \dots \mathcal{b}q_t(\mathcal{O}_t) \qquad [8]$$

For a given model, the probability of a series of state is in the form of [9].

$$\mathcal{P}(\mathcal{Q} \setminus \Lambda) = \pi q_1 a q_1 q_2 . a q_2 q_3 \dots a q_{\mathcal{T}-1} q_\mathcal{T} \qquad [9]$$

In this case the probability of concurrence of series $\mathcal{O}$ and $\mathcal{Q}$ is expressed with [10] as a unified probability function of these two series.

$$\mathcal{P}(\mathcal{O}, \mathcal{Q} \setminus \Lambda) = \mathcal{P}(\mathcal{O} \setminus \mathcal{Q}, \Lambda).\mathcal{P}(\mathcal{Q} \setminus \Lambda) \qquad [10]$$

The probability of the series of observations given by using the equal of conditional probability expression is expressed with [11] in the form of a sum of all q states of this unified probability and expressed as equation [12] with its clear expression.

$$\mathcal{P}(\mathcal{O}, \Lambda) = \sum_{all\ \mathcal{Q}} \mathcal{P}(\mathcal{O} \setminus \mathcal{Q}, \Lambda) . \mathcal{P}(\mathcal{Q} \setminus \Lambda) \qquad [11]$$

$$\mathcal{P}(\mathcal{O} \setminus \Lambda) = \sum_{q_1,q_2,\dots,q_{\mathcal{T}}} \pi q_1 \mathcal{b} q_1(\mathcal{O}_1).\, a q_1 q_2 \mathcal{b} q_2(\mathcal{O}_2) \dots a q_{\mathcal{T}-1} q_{\mathcal{T}} \mathcal{b} q_{\mathcal{T}}(\mathcal{O}_{\mathcal{T}}) \quad [12]$$

At $t = 1$ time, meaning in the beginning, the system is at q_1 state with πq_1 probability and the model generates $\mathcal{O}_1$ observation with $\mathcal{b} q_1(\mathcal{O}_1)$ probability. One step later, at $t = 2$ time, the system transits to state q_2 from state q_1 and generates $\mathcal{O}_2$ observation with $q_2(\mathcal{O}_2)$ probability. This process continues until the system transits to the state $\mathcal{T}$ which is the final state. When the system transits to state $\mathcal{T}$ with $a q_{\mathcal{T}-1} q_{\mathcal{T}}$ probability, it will generate $\mathcal{O}_{\mathcal{T}}$ observation with $\mathcal{b} q_{\mathcal{T}}(\mathcal{O}_{\mathcal{T}})$ probability (Rabiner, 1989, p.262). For $\mathcal{O}$ observation series, as there are $\mathcal{T}$ number of observations and $\mathcal{N}$ number of states, $\mathcal{N}^{\mathcal{T}}$ different possible state series are generated. When the equation which gives the probability of generating observation series for the given model is examined, $(2\mathcal{T} - 1)\mathcal{N}^{\mathcal{T}}$ multiplication and $\mathcal{N}^{\mathcal{T}} - 1$ addition operations are needed for a solution and the number of operations increases for higher values of $\mathcal{N}$ and $\mathcal{T}$. For this reason forward and backward algorithms have been developed so that $\mathcal{P}(\mathcal{O} \setminus \Lambda)$ probability value can be calculated. These procedures are based on dynamic programming technique. In order to be able to implement dynamic programming, a repetitive feature must exist which allows for calculations to be based on current sample for future samples (Petrushin, 2000, p.4). Forward algorithm is generally used to calculate the probability of observation series generated by hidden Markov model. However, both algorithms are mostly used for finding the optimal state series and calculating hidden Markov model parameters (Petrushin, 2000, p.5).

$\alpha_t(i)$ variable was defined for forwards algorithm which was developed in order to help the hidden Markov model solve the first problem. $\alpha_t(i)$ refers to the probability of partial observation series from 1 to t generated by a probable states series ending with state i, which is given in equation [13].

$$\alpha_t(i) = \mathcal{P}(\mathcal{O} = \mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_t, q_t = i \setminus \Lambda) \quad [13]$$

Conditional probability which is the sub-series of first t observation which ends at t time and sub-series of last m hidden state is equal to $v_1, v_2, \dots, v_t$ and $\mathcal{S}_{i_{t-m+1}}, \dots, \mathcal{S}_{i_t}$ by Λ model parameter respectively. In other words, when a given Λ model is known, the probability of concurrence of last m hidden state and first t observation can be predicted by calculating $\alpha_t(i)$ forward variable and thus forward algorithm is operated and solution is sought to the first problem. If $\alpha_t(i)$ is obtained for every $i = 1,2, \dots, \mathcal{N}$, $\mathcal{P}(\mathcal{O} \setminus \Lambda)$ probability is obtained with the sum of all values of $\alpha_t(i)$. Values of $\alpha_t(i)$ are obtained by three repetitive equations called initialization, induction and termination (Petrushin, 2000, p.4).

i. *Initialization:* this is the step where unified probability of hidden states and initial observations are calculated. The probability of single symbol series as a product of initial i. state and the probability of formation of $\mathcal{O}_1$ symbol given in i. state is calculated.

$$\alpha_1(i) = \pi_i e_i(\mathcal{O}_1) \quad i = 1,2, \dots, \mathcal{N} \quad [14]$$

ii. *Induction:* this is the step for calculating $\alpha_t(i)$ variable and this is the most important and principal step of forward algorithm. In order to calculate $\alpha_{t+1}(k)$ variable, the value which corresponds to the probability of transition from i. state to k. State is multiplied with all $\alpha_t(i)$ values and the sum is taken through all i values where $i = 1,2, \dots, \mathcal{N}$. Finally, it is multiplied with the probability value of $\mathcal{O}_{t+1}$ observation. Where $k = 1,2, \dots, \mathcal{N}$ *and* $t = 1,2, \dots, \mathcal{L} - 1$,

$$\alpha_{t+1}(k) = \left[\sum_{i=1}^{\mathcal{N}} \alpha_t(i).\, a_{ik}\right] e_j(\mathcal{O}_{t+1}) \quad [15]$$

iii. *Termination:* This is the step which gives $\mathcal{P}(\mathcal{O} \setminus \Lambda)$ probability. Values of $\alpha_t(\mathcal{L})$ variables obtained in the second step are added for all states and the required probability value is reached.

$$\mathcal{P}(\mathcal{O} \setminus \Lambda) = \sum_{i=1}^{\mathcal{N}} \alpha_{\mathcal{L}}(i) \quad [16]$$

Second step of the algorithm developed for solving the first problem is backward algorithm. Similar to forward algorithm, backward algorithm is defined as $\beta_t(i)$. $\beta_t(i)$ is the conditional probability of partial observation series

from $\mathcal{O}_{t+1}$, which was generated by all state series starting with i. state, to the last observation (Petrushin, 2000, p.5).

$$\beta_t(i) = \mathcal{P}(\mathcal{O}_{t+1}\mathcal{O}_{t+2} \dots \mathcal{O}_{\mathcal{L}} \setminus q_t = i, \Lambda) \quad [17]$$

Backward algorithm goes backwards along observation series and calculates sequential backward variables (Petrushin, 2000, p.4).

$$\mathcal{P}(\mathcal{O}_{t+1}\mathcal{O}_{t+2} \dots \mathcal{O}_{\mathcal{L}} \setminus \Lambda) = \sum_{i=1}^{\mathcal{N}} \pi_i e_i(\mathcal{O}_{t+1}).\beta_t(i) \quad [18]$$

i. *Initialization*: at the starting point, the value 1 is assigned arbitrarily to $\beta_t(i)$ for all i.

$$\beta_1(i) = 1 \quad i = 1,2, \dots, \mathcal{N} \quad [19]$$

ii. *Induction*: In induction step, the probability of transition from i state to j. state, probability value of $(\mathcal{O}_{t+1})$ observation and probability value of $\beta_{t+1}(j)$ are multiplied and added through j.

$$\beta_t(i) = \sum_{j=1}^{\mathcal{N}} a_{ij} e_j(\mathcal{O}_{t+1})\beta_{t+1}(j) \quad i = 1,2, \dots, \mathcal{N} \text{ and } t = \mathcal{L} - 1, \dots, 1 \quad [20]$$

iii. *Termination*: this is the step where $\mathcal{P}(\mathcal{O} \setminus \Lambda)$ probability value is calculated.

$$\mathcal{P}(\mathcal{O} \setminus \Lambda) = \sum_{i=1}^{\mathcal{N}} \pi_i e_i(\mathcal{O}_1)\, \beta_1(i) \quad [21]$$

There are a few ways to solve the second problem. Algorithms developed for solving the first problem provide precise solutions; the reason for having more than one solutions for solving the second problem is the difficulties in defining what the optimality criterion is. The purpose of solving second problem is to find the hidden state series which best explains the given observation series. Here the criterion for choosing "best" state series has to be determined. This approach is called *backward decoding* (Petrushin, 2000, p.6). If hidden Markov model is ergodic, meaning that it is possible to transit from all states to all other states, backward decoding provides good results for solution to the problem (Petrushin, 2000, p.7). However, if the model is not ergodic, transition to some states is not allowed; thus, this algorithm will not provide a favourable series. In this case a new optimality criterion must be defined and it will be appropriate to employ Viterbi decoding or Viterbi algorithm, as the solution of equality shown as $q_t = argmax\{\lambda_t(i)\}$ gives the determinant of the optimum state ignoring the probability of formation of a series of states (Rabiner, 1989, p.264). Viterbi algorithm is the algorithm for finding the repetitive optimal solution for a state series prediction problem of Markov processes with discrete time and final state (Forney, 1973, p.268). The purpose is to choose the optimum series which will maximize the probability of single state series for a given observation series. The algorithm is divided into two parts; first $\max_{\mathcal{Q}} \mathcal{P}(\mathcal{Q} \setminus \mathcal{O})$ is found, then, in order to find the $\mathcal{Q}$ which gives this maximum, the operations are reversed backwards and the solution is solved (Ewens and Grant, 2005, p.413). Mathematically it is expressed as equation [22].

$$\mathcal{Q}^* = \underset{\mathcal{Q}}{\text{argmax}}\{\mathcal{P}(\mathcal{Q} \setminus \mathcal{O}, \Lambda)\} = \underset{\mathcal{Q}}{\text{argmax}}\{\mathcal{P}(\mathcal{Q}, \mathcal{O} \setminus \Lambda)\} \quad [22]$$

The highest probability $\delta_k(i)$ of the state series which generates first k observations, begins with i state and has a length of k for the given model is expressed as equation [23].

$$\delta_k(i) = \max_{q_1, \dots, q_{k-1}} \mathcal{P}(q_1 q_2 \dots q_k = i, o_1 o_2 \dots o_k \setminus \Lambda) \quad [23]$$

Viterbi algorithm is also a dynamic programming algorithm just like forward algorithm but with two differences. Maximization operation is used in Viterbi algorithm instead of the addition operation in the repetitive and termination steps of forward algorithm. It hides each k and i in $\mathcal{N}$x$\mathcal{L}$ matrix and chooses to maximize $\delta_k(i)$

arguments. This matrix is used to obtain optimal state series in the return step. The steps of Viterbi algorithm are as follows (Petrushin, 2000, p.7):

i. *Initialization:*

$$\delta_t(j) = \pi_j e_j(O_1) \quad j = 1,2,\dots,N \qquad [24]$$
$$\Psi_t(j) = 0 \qquad [25]$$

At this step, starting probability value of each state is multiplied with the probability value of the first observation. The variable which gives maximum j state is assigned as 0.

ii. *Recursion: Where* $j = 1,2,\dots,N$ *and* $k = 1,2,\dots,L-1$

$$\delta_{t+1}(j) = \max_{1\le i\le N}\left[\delta_k(i)a_{ij}\right] e_j(O_k) \qquad [26]$$
$$\Psi_{t+1}(j) = \operatorname*{argmax}_{1\le i\le N}\left[\delta_k(i)a_{ij}\right] \qquad [27]$$

$\delta_t(j)$ values which are obtained in the previous step separately for all states are multiplied with transition probabilities and maximum value is obtained which is then multiplied with the observation probability in the current state. Maximum variable obtained after this operation is assigned to the variable $\Psi_{t+1}(j)$.

iii. *Termination:*

$$p^* = \max_{1\le i\le N}[\delta_L(i)] \qquad [28]$$
$$q_L^* = \operatorname*{argmax}_{1\le i\le N}[\delta_L(i)] \qquad [29]$$

Maximum value is chosen among $\delta_L(i)$ values calculated for the final observation in the previous step and called p^*. q_L^* determines the state from which the chosen maximum $\delta_L(i)$ comes and finally optimum state is determined for final observation.

iv. *Path (state sequence) Backtracking:*

$$q_k^* = \Psi_{k+1}(q_{k+1}^*) \quad k = L-1,\dots,1 \qquad [30]$$

In this step, the process is backward from final observation to the first observation. With the expression $q_k^* = \Psi_{k+1}(q_{k+1}^*)$ the state from which the optimum obtained in the next step is determined. This return continues until the starting point and thus the appropriate state series is identified. If a single argmax state cannot be found at the end of these operations, one of the i states which give the maximum can be chosen arbitrarily (Ewens and Grant, 2005, p.414). The purpose in solving the third problem is to identify a method as to how model parameters can be remodelled. This means that in order to identify a desired observation series as output, $\mathcal{A}, \mathcal{B}, \Pi$ which are the model parameters that will maximize observation state probability are to be calculated. In order to maximize $\mathcal{P}(O \setminus \Lambda)$ probability locally, $\Lambda = (\mathcal{A}, \mathcal{B}, \Pi)$ parameters are to be optimized, or predicted, for which repetitive procedures such as Baum - Welch Algorithm, Viterbi Training Algorithm, K-means Clustering Algorithm have been developed. The most frequently used algorithm is Baum – Welch algorithm. It is used in the solution of third problem which is called the "training of hidden Markov model" and is a repetitive similarity maximization method founded on probabilities belonging to forward and backward variables (Vaseghi, 2007, p.364).

With the given model and observation series, the variable $\xi_t(i,j)$ which shows the probability of the process in S_i state at t time to change to S_j state at $t+1$ time is shown in equation [31] (Rabiner, 1989, p.264).

$$\xi_t(i,j) = \mathcal{P}\left(q_t = S_i, q_{t+1} = S_j \setminus O, \Lambda\right) \qquad [31]$$

If this equation is redefined using forward and backward equations, it can be defined as equation (32).

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O\backslash\Lambda)} = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{i=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \quad [32]$$

When the share of the equation is examined, it can be seen that its equal is [33].

$$P(q_t = S_i, q_{t+1} = S_j, O\backslash\Lambda) \quad [33]$$

Therefore, for the given model and observation series, the $\xi_t(i,j)$ measure which shows the probability of the process in S_i state at t time to change to S_j state at $t+1$ time is obtained by proportioning the defined share to the probability $P(O \setminus \Lambda)$. Connection can also be found between $\gamma_t(i)$ probability and $\xi_t(i,j)$ probability. If the sum of $\xi_t(i,j)$ is taken through j, it can be seen that the probability $\gamma_t(i)$ is found.

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \quad [34]$$

When the sum is taken on time index provided not that $\gamma_t(i)$ and $\xi_t(i,j)$, $t = T$, $\gamma_t(i)$ is the probability of being at S_i state for any t time for a given model and observation series, then equation [35] will give the number of expected visits to S_i state, meaning the transitions expected from state S_i.

$$\sum_{t=1}^{T-1} \gamma_t(i) \quad [35]$$

Likewise, $\xi_t(i,j)$ is the probability of the process in S_i state at t time to change to S_j state at $t+1$ time for a given model and observation series, the equation [36] will give the number of expected transition from state S_i to state S_j.

$$\sum_{t=1}^{T-1} \xi_t(i,j) \quad [36]$$

Under the light of all these information, a method has been developed to recalculate the hidden Markov model parameters.

$\gamma_1(i)$, the frequency expected in S_i state at $\bar{\pi}_i$: $t = 1$ time,

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \quad [37]$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^{T-1}\ \gamma_t(i)}{O_{t=v_k}} \quad [38]$$

The numerator of $\bar{b}_j(k)$ equation shows the number of times at j states when observable symbol is v_k whereas the expression in the denominator shows the expected number of times in j state.

Defined hidden Markov model being $\Lambda = (A, B, \Pi)$ and the new model which will be obtained for maximization of the desired observation series being $\bar{\Lambda} = (\bar{A}, \bar{B}, \bar{\Pi})$, one of the following expressions will be realised using the algorithm proved by Baum et al.

i. Initial model: Λ defines a critical point for $\Lambda = \bar{\Lambda}$ similarity function.

ii. $\bar{\Lambda}$ model makes $(O \setminus \bar{\Lambda}) > P(O \setminus \Lambda)$ inequality more probable compared to model Λ.

Therefore, according to the new $\bar{\Lambda}$ model which is found by operating the algorithm, a more appropriate model is generated compared to Λ model for the formation of observation series. If $\bar{\Lambda}$ model is used repetitively instead of

Λ model and a reprediction is made, the probability of being observed from the model of the observation series until a definite limit is reached can be improved. The result of reprediction process is called the maximum similarity prediction of hidden Markov model. As optimization level is very complex and it will have several local maximum points, in several problems backward and forward algorithms will generate only local maximum values. Re-prediction formulas can be regenerated maximizing the auxiliary function of Baum on $\overline{\Lambda}$ and using restricted optimizaiton techniques, and finally, probability function converges to the critical point: $\max_{\overline{\Lambda}} \mathcal{Q}(\Lambda, \overline{\Lambda}) \Rightarrow \mathcal{P}(\mathcal{O} \setminus \overline{\Lambda}) \geq \mathcal{P}(\mathcal{O} \setminus \Lambda)$. In addition, stochastic,c restraints of hidden Markov model such as $\sum_{i=1}^{\mathcal{N}} \bar{\pi}_i = 1$, $\sum_{j=1}^{\mathcal{N}} \bar{a}_{ij} = 1$, $1 \leq i \leq \mathcal{N}$ and $\sum_{k=1}^{\mathcal{M}} \bar{b}_j(k) = 1$, $1 \leq j \leq \mathcal{N}$ are obtained in every iteration automatically (Rabiner, 1989, p.265).

3. FOREIGN DIRECT INVESTMENT

Every country wants to meet their needs with their own dynamics and be self-sustaining without needing technology, capital, work force etc. entry from other countries. However, the facts that resources are not distributed between states evenly and differences in level of development between countries have obligated resource trade at international level. Particularly less developed and developing countries try to transfer resources from developed countries with which both countries will benefit. One of the parties has capital and technology and the other party can be stronger in terms of natural resources and manpower, and as a result both countries can benefit due to their specific powers. Capital is one of the critical resources that both kinds of countries have to possess so as to ensure development. For this reason countries turn to capital accumulation and try to attract capital to their countries in several ways. External trade, hot money entries, loans used from international financial institutions or governments or FDI are among the leading ways to create capital accumulation. FDI can be defined as long-term investments made by international companies, real or legal persons which are under the management and supervision of a main centre and active in different countries in the form of purchasing a company in another country, joining another company, opening a new company or concluding license agreements, with the purpose of diffusing the products of the company in the main centre to developing or less developed countries in a manner with which management and control will be retained.

Developed countries' investors who have the purpose of evaluating their idle capital and making more gains are willing to enter into quest for an under develop or developing country. So they evaluate the counties within certain criteria for the most suitable investment. There are some determinants that international companies take into consideration when choosing the companies for investment. These determinants are the volume of the market or its internal size, workforce ratio and cost, economic and political stability of the country, tax liabilities, openness index, interest rate, corruption indicators, infrastructure facilities, level of investment incentives and transferability of capital or profits.

4. APPLICATION

Departing from the fact that basic determinants that attract FDIs ($DYSY_t$) to a certain country are inflation rate ($TUFE_t$), real effective exchange rate ($REDK_t$), current account balance (CID_t), unemployment rate (IO_t), external trade volume (DTH_t), industrial production index (SUE_t), number of investment incentive certificates ($YTES_t$) , this study will perform econometric analysis and assumption testing with the purpose of deciding which of these determinant are more explanatory in attracting FDIs to the country in the case of Turkey based on the data between 2009:09 and 2013:12 and the analysis of variables in the suitable model as well as FID entry values in 2014:01 – 2014:01, 2014:02 – 2014:01, 2014:02, 2014:03 periods will be conducted with the help of the three basic problems of hidden Markov model and the findings after the solution of each problem will be interpreted. The initial thinking was to include corporations tax rate as tax variable and corruption index as reliability of the country indicator for obtaining the appropriate model, but they were excluded as corporations tax rate series was annual but the study would be conducted on a monthly basis and the series had low variability; likewise, corruption index was being calculated annually since 2007 taking the change in the index compared to the previous period would not be meaningful. With the purpose of determining the factors that affect FDI entries in Turkey, regression models were predicted with LSM. The most significant model among the tried ones is the logarithmic-linear regression model where logarithmic converted models of $DYSY_t$ series is explained with $LDYSY_t$; DTH_t, IO_t and $TUFE_t$ variables.

$$LDYSY_t = \beta_1 + \beta_2 DTH_t + \beta_3 IO_t + \beta_4 TUFE_t + \varepsilon_t \quad [39]$$

Table 1. LSM Estimation Results For Logarithmic-Linear Model

Dependent Variable : $LDYSY_t$				
Method : Least Squares				
Sample: 2009M09 2013M12				
Included Observations: 52				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.488548	1.708336	0.871344	0.3879
DTH_t	0.000214	0.000033	6.381691	0.0000
IO_t	0.212978	0.082689	2.575660	0.0131
$TUFE_t$	-0.017698	0.006159	-2.873505	0.0060
R-Squared	0.537996			
Adj. R-Squared	0.509121	Mean Dependent Var		6.474149
S.E.of Regression	0.503350	S.D. Dependent Var		0.718428
Sum Squ. Resid	12.16136	Akaike Info Criterion		1.538743
Log. Likelihood	-36.00732	Schwarz Criterion		1.688839
F-Statistic	18.63174	Hannan-Quinn Criter.		1.596286
Prob (F-stat.)	0.000000	Durbin-Watson Stat.		1.612934

When we examine Table 1, we can see that all coefficients are statistically significant as all predicted t-statistics of independent variable coefficients of the logarithmic-linear regression model where $LDYSY_t$ is explained are bigger than the table value which is $t_{0.05(52-4)}$ =2,021 according to 5% significance level except the constant term. The insignificance of fixed parameter is ignored. The value of specificity coefficient R^2 which expresses the ratio of explanation of $LDYSY_t$ by changes in DTH_t, IO_t and $TUFE_t$ variables is 0,53. This ratio is a valid one as it is above 50%. The result of F test turned out as 18.63 with 0.000 queue probability. As this value is bigger than table value which is $F_{(3,48)}$=2,84 according to 5% significance level, the alternative hypothesis which states that the model is generally significant could not be rejected.

Cusum and Cusum of Square tests were conducted in order to identify whether there were structural changes in the residuals of the model and it was decided that there was no structural failure in the residuals of the model. Ramsey-Reset Test was used to see whether there was a definition error in the model and the null hypothesis which claimed that there was no definition error was not rejected and it was concluded that there was no definition error. Basic assumptions of the classical regression model have to be verified so that the predicted logarithmic-linear model is a valid one (Green, 2007, s.10). For this purpose normality test was conducted to see the normal distribution of residuals of the model, autocorrelation test was done to find out whether there was an autocorrelation between the residuals of the model, and changing variance tests were made to examine the variance of residuals were fixed; finally, serious multiple linearity dependence status was examined between independent variables. As a result, it was found out that basic assumptions were verified. As a result of the regression model predicted with LSM, the effective determinants for performing future predictions so that FIDs that come to Turkey could be predicted using hidden Markov model were identified and the states of hidden Markov model were determined as DTH_t, IO_t and $TUFE_t$. Observation series of the model is provided by $LDYSY_t$. The data interval between 2009:09 – 2013:12 was studied in the application and the data were obtained from official websites of Central Bank of Republic of Turkey, UNCTAD, World Bank, OECD and Turkish Statistics Agency. The solution of the model was performed using Matlab package program.

Change rates compared to previous month for both observation and state series was calculated and "change ratio series" was created. After change ratios belonging to state and observation series was calculated each series was divided into four groups within itself so as to be able to express series in discrete manner while forming these groups changing ratios of series was checked and arithmetic mean of the years of increase was taken within themselves and the values that remain above the mean were assigned a symbol, likewise the values which remain below the mean were assigned another symbol. Again, arithmetic mean of the years of decrease in change ratios for each series was taken within themselves and the values that remain above the mean were assigned a symbol, likewise the values which remain below the mean were assigned another symbol. The mean of positive values of change ratios series of $LDYSY_t$ observation series was found as 11,10434 and mean of negative values was found

as -9,53484, and a symbol was assigned to each month depending on the mean value. The symbols of change ratios series for every month were determined according to the following criteria.

Table 2. LDYSY Description Symbol For LDYSY Observation Series

Change Ratio Range	Symbol
$DO \geq 11{,}10434$	A1
$0 \leq DO < 11{,}10434$	A2
$-9{,}53484 < DO < 0$	A3
$DO \leq -9{,}53484$	A4

A1 and A2 symbols refer to positive changes in observation series, whereas A3 and A4 symbols show negative changes. A1 was written across the months that show increase more than the mean of the series and A2 was written across the months that show increase less than the mean of the series. A3 symbol referred to the months that showed less decrease compared to the mean of negative values of the series whereas A4 define the months of decrease more than average. Therefore, the observation series of the model was created as (A1,A2,A3,A4).
The mean of positive values of change ratios series of DTH_t state was found as 8,56691 and the mean of negative values was found as -6,92869. For this reason the symbols of change ratios series for every month were determined according to the following criteria.

Table 3. Description Symbol For DTH State Series

Change Ratio Range	Symbol
$DO \geq 8{,}56691$	B1
$0 \leq DO < 8{,}56691$	B2
$-6{,}92869 < DO < 0$	B3
$DO \leq -6{,}92869$	B4

B1 and B2 symbols refer to positive changes in DTH_t state series, whereas B3 and B4 symbols show negative changes. B1 was written across the months that show increase more than the mean of the series and B2 was written across the months that show increase less than the mean of the series. B3 symbol referred to the months that showed less decrease compared to the mean of negative values of the series whereas B4 define the months of decrease more than average. The mean of positive values of change ratios series of IO state was found as 3,48897 and the mean of negative values was found as -4,44203. For this reason the symbols of change ratios series for every month were determined according to the following criteria.

Table 4. Description Symbol For IO State Series

Change Ratio Range	Symbol
$DO \geq 3{,}48897$	C1
$0 \leq DO < 3{,}48897$	C2
$-4{,}44203 < DO < 0$	C3
$DO \leq -4{,}44203$	C4

C1 and C2 symbols refer to positive changes in IO_t state series, whereas C3 and C4 symbols show negative changes. C1 was written across the months that show increase more than the mean of the series and C2 was written across the months that show increase less than the mean of the series. C3 symbol referred to the months that showed less decrease compared to the mean of negative values of the series whereas C4 define the months of decrease more than average.

Table 5 . Description Symbol For TUFE State Series

Change Ratio Range	Symbol
$DO \geq 0{,}94445$	D1
$0 \leq DO < 0{,}94445$	D2
$-0{,}49825 < DO < 0$	D3
$DO \leq -0{,}49825$	D4

The mean of positive values of change ratios series of $TUFE$ state was found as 0,94445 and the mean of negative values was found as -0,49825. For this reason the symbols of change ratios series for every month were determined according to the following criteria.

The symbols of D1 and D2 refer to positive changes in $TUFE_t$ state series, whereas D3 and D4 symbols show negative changes. D1 was written across the months that show increase more than the mean of the series and D2 was written across the months that show increase less than the mean of the series. D3 symbol referred to the months that showed less decrease compared to the mean of negative values of the series whereas D4 define the months of decrease more than average.

Symbols belonging to DTH_t, IO_t, $TUFE_t$ variables which express states are determined as (B1, B2, B3, B4) for DTH_t variable, (C1, C2, C3, C4) for IO_t variable and (D1, D2, D3, D4) for $TUFE_t$ variable. In this state transition between sub-states generated for each state are possible whereas transition generated from different states is not possible. Therefore, in order to facilitate transition between sub-states belonging to different states, organization has to be made between sub-states. For example, at a time, in one month for this study, B1 state occurs belonging to DTH_t state variable, whereas D1 states can be formed for C1 and $TUFE_t$ state variable. For this reason when a cluster of new states is created consisting of the combinations of sub-states, this time transition can be made from any state inside the cluster to any other state. This cluster consists of 64 states and is expressed as X1,X2,…,X64 cluster of states.

Interpretation on the values adopted by states by only examining observations is not possible. For application only by monitoring the entrances of $LDYSY_t$ is not sufficient for having an opinion about the states of DTH_t, IO_t, $TUFE_t$. Therefore the X1,X2,…,X64 cluster of states created here expresses "hidden states" for hidden Markov model as the determinants of $LDYSY_t$ were constructed sub-states.

Under the assumption that the state realised for each month is dependent only on the previous state, and that the impact of all states in the past are collected in the current state, a matrix for states necessary for the model has been established. State transition probabilities are in the size of 64 x 64. However, inside hidden states there are states which have not been formed throughout sample. As they violated the rule that the sum of Markov transition rows would be 1, the mentioned states are not included in the matrix.

Likewise, with signals coming from hidden states, a series of observations which correspond to these states called "observations" is generated. The observation series is generated by $LDYSY_t$. This matrix called "observation probabilities matrix" is constructed with observation values which correspond to each timeframe. This means that when any of the hidden states of observation probabilities matrix X1,X2,…,X64 will take place, any of the A1, A2, A3, A4 observations will be generated in that way. Therefore observation probabilities matrix will be one at the size of 64x4. Hidden states which have not been formed in the defined time frame will not create any observation output; thus, these situations are excluded from observation probabilities matrix.

In application, as there is no hierarchy between hidden states, initial situation states are taken equal for each state.

After all parameters required to solve these three basic problems belonging to a hidden Markov model are defined, $LDYSY_t$ which correspond to the 2014:01; 2014:01,2014:02 and 2014:01,2014:02,2014:03 months which are the three periods after the defined time interval that correspond to the solution of the first problem will be predicted. At the same time the values taken by hidden states that correspond to the observation values of $LDYSY_t$ in defined periods will be evaluated under the light of the solution of second problem. Interpretation of each symbol obtained in the solution of the first and second problems will be made according to the mean values of state series and change series of the observation series itself. Finally, the solution which corresponds to the solution of third problem and the problem which aims at optimizing model parameters will be handled.

As corresponding to observation series $\mathcal{O} = \mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \mathcal{O}_4$, the solution of the first problem has been performed for each observation value of (A1,A2,A3,A4). The probability values that correspond to 2014:01 period are as seen in table 6.

Table 6. Observation Probability Forecast for the Period of 2014:01

Observation Symbol	Probability
A1	0
A2	0
A3	0
A4	1

When table 6 is examined, it can be seen that formation probabilities of A1,A2,A3 observations for 2014:01 period is predicted as zero. For this reason these observations will not occur for the defined period. The probability of A4 observation was predicted as 1 in which case the observation which will take place for January will be 100% A4 observation. While the symbol of predicted observation is A4, meaning that a decrease over mean is expected in January 2014, it can be seen that A4 observation took place when the observation symbol is examined. This means that in reality a decrease over medium was realised in January 2014 compared to December 2013. Therefore it can be stated that the model provided a consistent result for first period. Likewise, when the observation probability values which will be formed for 2014:01, 2014:02 are examined, it can be seen that the probability values given in Table 7 are generated.

Table 7. Observation Probability Forecast for the Period of 2014:01, 2014:02 **

Observation Symbol	Probability
A4, A1	0
A4, A2	0,6667
A4, A3	0,3333
A4, A4	0

** When table 6 is examined, as A1, A2, A3 probabilities of 2014:01 period are zero, these observations were not included in table 7.

When table 7 is examined it can be seen that the formation probability of A4,A1 and A4,A4 observation series for 2014:01,2014:02 period was predicted as zero. As the probability value belonging to A4,A2 observation series has the highest probability value with 0,6667, the observation series which is expected to take place for the second period was predicted as A4,A2. This means that a decrease over mean is expected in January 2014 compared to December 2013 and an increase below mean value is expected in February 2014 compared to January 2014. When seen in terms of realised observation value, it is seen that A4,A2 observation series took place in reality for 2014:01,2014:02 period. For this reason it can be said that the model provided consistent results. When first problem is solved for 2014:01,2014:02,2014:03 period, the resulting observation probability is as seen in table 8.

When table 8 is examined the probability of occurrence of A4,A2,A2 observation series and A4,A3,A4 observation series is predicted as 0. Probability values belonging to other observation series were found as equal to each other with 0,3333. It is certain that A4,A2,A2 and A4,A3,A4 observation series will not occur definitely; however, decision cannot be made as to which one will be formed as the probability values of other observation series are equal to each other.

As the probability values of indicated observations are equal to each other, it is difficult to say that the model provides consistent results for 2014:01,2014:02,2014:03 period; however, as A4,A2,A3 observation series is among the observations which are formed is highest probability value, it can be claimed that a result close to the realised value has been obtained. When the formed observation series is examined, it is observed that a decrease above mean was witnessed in January 2014 compared to December 2013, an increase below mean was witnessed in February 2014 compared to January 2014, and a decrease below mean was witnessed in March 2014 compared to February 2014.

Table 8. Observation Probability Forecast for the Period of 2014:01, 2014:02, 2014:03***

Observation Symbol	Probability
A4, A2, A1	0,3333
A4, A2, A2	0
A4, A2, A3	0,3333
A4, A2, A4	0,3333
A4, A3, A1	0,3333
A4, A3, A2	0,3333
A4, A3, A3	0,3333
A4, A3, A4	0

***When table 7 is examined, the probability values belonging to A4, A1 and A4, A4 observation series is 0, thus, probabilities belonging to these observations are not included in table 8.

For 2014:01 period, it was predicted in the conclusion of 1st problem that A4 observation would occur 100%. While A4 observation was taking place, the 2nd problem which answered the questions of what the hidden states were which underlay this observation, it can be seen that the state prediction which will be formed for A4 observation was X54 state and, when the sub-states of this state were examined, these predicted sub-states were B4-C2-D2. This means that it was predicted that DTH_t hidden state would show a decrease above mean, but IO_t hidden state would reveal an increase below the mean value and that in $TUFE_t$ hidden state an increase below mean would take place. However in reality it can be seen that X49 state was formed and that sub-states belonging to this state are also B4-C1-D1. Sub-state belonging to DTH_t hidden state was predicted as consistent; however, it can be seen that differences occured in the other two sub-states. This means that in the realised form an increase above mean value of the sub-states were witnessed in both states, whereas the predicted state also showed an increase but that this would be below mean value. When the prediction results in the second problem with which hidden states belonging to 2014:01,2014:02 period are examined, the results given in the following table are obtained.

Table 9. State Forecast for the 2014:01, 2014:02 Observation Sequence

Observation	State Forecast	Sub-States
A4, A1	X54, X2	B4-C2-D2, B1-C1-D2
A4, A2	X54, X22	B4-C2-D2, B2-C2-D2
A4, A3	X54, X22	B4-C2-D2, B2-C2-D2
A4, A4	X54, X22	B4-C2-D2, B2-C2-D2

The observation series with the highest possibility was stated as A4,A2. An examination of table 9 shows that the hidden states occurred by the formation of A4,A2 sub-state are X54,X22. When the sub-states belonging to these states are examined, it can be seen that they are predicted as B4-C2-D2, B2-C2-D2. It was predicted that DTH_t hidden state showed a decrease below the mean value in January but that it showed an increase below the mean value in February. When IO_t hidden state is examined, it can be seen that the increase in both January and February were below the mean value. In $TUFE_t$ hidden state, however, it was predicted that the increase in January and February would be below mean value. In the real situation it can be seen that hidden state series was X49,X42 and that sub-states belonging to these states were realised B4-C1-D1,B3-C3-D2. This means that DTH_t hidden state showed an increase above its mean value and that IO_t and $TUFE_t$ hidden states showed an increase below their mean values in January; and that DTH_t and IO_t hidden states showed a decrease below their mean value and that $TUFE_t$ hidden state showed an increase below its mean values in February. When realised and predicted states are compared, although small deviations take place, it can be seen that predictions show results close to the real situation. Prediction results found in the solution of 2014:01,2014:02,2014:03 period can be seen in table 10.

Table 10. State Forecast for the 2014:01, 2014:02, 2014:03 Observation Sequence

Observation	State Forecast	Sub-States
A4, A2, A1	X54, X22, X14	B4-C2-D2, B2-C2-D2, B1-C4-D2
A4, A2, A2	X54, X22, X26	B4-C2-D2, B2-C2-D2, B2-C3-D2
A4, A2, A3	X54, X22, X14	B4-C2-D2, B2-C2-D2, B1-C4-D2
A4, A2, A4	X54, X22, X35	B4-C2-D2, B2-C2-D2, B3-C1-D3
A4, A3, A1	X54, X22, X14	B4-C2-D2, B2-C2-D2, B1-C4-D2
A4, A3, A2	X54, X22, X26	B4-C2-D2, B2-C2-D2, B2-C3-D2
A4, A3, A3	X54, X22, X14	B4-C2-D2, B2-C2-D2, B1-C4-D2
A4, A3, A4	X54, X44, X35	B4-C2-D2, B3-C3-D4, B3-C1-D3

When the probability values of 2014:01,2014:02,2014:03 observation series is compared, it was stated that the occurred observation series was A4,A2,A3 although a clear judgement could not be reached for the observation series which would occur. The states occurred belonging to this observation series are shown in table 10. It can be seen that states are predicted as X54,X22,X14 and that the sub-states belonging to these states occurred as B4-C2-D2, B2-C2-D2, B1-C4-D2. According to prediction results, it is observed that DTH_t hidden state would show a decrease below mean value in January, an increase below mean value in February and an increase above mean value in March for the defined period, that IO_t hidden state would show an increase below mean value in January, an increase below mean value in February, and a decrease above mean value in March, and that $TUFE_t$ hidden state would show an increase below the mean value in January, February and March. When the realised states are examined for the current period, it can be seen that X49,X42,X9 state occurred and that the sub-states of this state were B4-C1-D1, B3-C3-D2, B1-C3-D1. When sub-states for March are analysed, it was seen that an accurate prediction was made for DTH_t hidden state, whereas a decrease in IO_t hidden state was predicted but, although the expectation was that the decrease would be below mean value, it was observed that it turned out to be above mean value. The increase prediction for $TUFE_t$ hidden state was also a success; however, it was above the mean value although it was predicted to be below mean value.

The prediction of model parameters with the purpose of maximizing any observation or observation series belonging to 2014:01; 2014:01, 2014:02 and 2014:01, 2014:02, 2014:03 periods was performed with Baum – Welch algorithm. Model parameters for all possible observations and observation series were re-predicted and, after the prediction, the probabilities for observations or observation series with which parameters were organized by using new parameters were found as 100 percent. When the results of all three problems are examined, it is seen that first problem provided consistent results for January and January, February months; however, it is observed that it could not reach a probability value that could allow for definite judgment for January, February and March. What needs careful investigation here is that probability values found decrease as the term becomes longer in model solution. For this reason it can be claimed that prediction results of hidden Markov models provided more consistent results for short periods. However, if the time between signals sent by states and the occurred observation values is too long, it can be said that future predictions can be made for longer periods. When the results of the second problem are evaluated, it is seen that the occurred and expected situations give very close results. In the third problem, it can be seen that the parameters of the model are optimized for the desired observation or observation series.

5. CONCLUSION

In this study especially the determinants which attract foreign capital investments to a country have been determined by examining the literature. Taking into these determinants, $DYSY_t$ series was included as dependent variable and $TUFE_t$, $REDK_t$, CID_t, IO_t, DTH_t, SUE_t and $YTES_t$ variables were included as independent variables in the model and an effort was paid to find out the model which explains $DYSY_t$ best. As a result of the analysis conducted, the logarithmic-linear model where $LDYSY_t$ was dependent variable and $DTH_t, IO_t, TUFE_t$ were independent variables, which covered 2009:09 – 2013:12 period, was determined as the "optimum model". For this study which was conducted for 2009:09 – 2013:12 period in Turkey, it was found with an econometric model first that DTH_t, IO_t and $TUFE_t$ were the variables which explained FDI entries best. Then, in order to be able to make future prediction, the hidden Markov model application was assumed which constituted the main theme of this study. In this model DTH_t, IO_t and $TUFE_t$ variables constitute the states of the model whereas $LDYSY_t$ constitutes its observation. Necessary adjustments and symbol assignments were made both for observation series and state series and model parameters such as matrix for transition to hidden states, observation probabilities matrix

and initial state probabilities were determined. The first problem of hidden Markov model was solved and future predictions of FDI entries for 2014:01,2014:02 and 2014:01,2014:02,2014:03 periods were made.

As the purpose of second problem of hidden Markov model is to reveal the hidden states under an observation or observation series, second problem was solved for observations and observation series obtained at the end of the first problem and how the hidden states would occur were predicted. It is seen that when third basic problem of the model is solved, model parameters were maximized. Finally, it is observed that usually econometric studies are performed in the prediction of DYSY entries. It is possible to claim that hidden Markov models could be used as an alternative mathematical method in predicting how hidden states that underlie observation outputs and what kind of a trajectory the observation outputs would follow in the future, and that prediction results are a considerable success. In conclusion, it should be added that as there are too many factors that affect each other in social problems, there are difficulties in determining the states that explain best the observations while using mathematical models. For this reason it is essential that the optimum model is determined. Likewise it must be noted that the success of the model for future prediction depends on the multitude of data used, as the more signals are sent by states, the more accurate the prediction of observations becomes, and the longer term prediction is facilitated.

REFERENCES

Bhar, Ramaprasad and Hamari, Shigeyuki. (2004), Hidden Markov Models: Applications to Financial Economics. Springer Science & Business Media Inc.

Blunsom, Phil. (2004), Hidden Markov Models, pp.1-7. http://digital.cs.usu.edu/~cyan/CS7960/hmm-tutorial.pdf, (30.11.2014).

Champion, Colin and Houghton, S.M. (2015), Application of Continuous State Hidden Markov Models to a Classical Problem in Speech Recognition. *Computer Speech and Language*, pp. 347-364, http://www.sciencedirect.com/science/article/pii/S0885230815000443 (12.10.2015).

Ching, Wai-Ki and Ng, Michael K. (2006), Markov Chains: Models, Algorithms and Applications. Springer Science & Business Media Inc.

Can, Tuncay and Öz, Ersoy. (2009), Saklı Markov Modelleri Kullanılarak Türkiye'de Dolar Kurundaki Değişimin Tahmin Edilmesi. *Istanbul University Journal of the School of Business Administration,* Vol.38, No.1, pp. 1-23.

Ekinci, Aykut. (2011), Doğrudan Yabancı Yatırımların Ekonomik Büyüme ve İstihdama Etkisi: Türkiye Uygulaması (1980-2010), *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 6(2), pp.71-96.

Ewens, Warren and Grant, Gregory. (2005), Statistical Methods in Bioinformatics: An Introduction. Hidden Markov Models. Springer Science & Business Media Inc, 12. Section, 2.Edition.

Elliott, Robert J. and Hoek, John van der. (1997), An Application of Hidden Markov Models to Asset Allocation Problems. *Finance and Stochastics, Springer-Verlag,* pp. 229-238.

Forney, G.David. (1973). The Viterbi Algorithm. *Proceedings of the IEEE*, Vol.61, No.3.

Fuh, Cheng-Der. (2003), SPRT and Cusum in Hidden Markov Models. *The Annals of Statistics*, Vol.31, No.3, pp.942-977.

Green, W. H. (2007), Econometric Analysis (7th Edition). Prentice Hall, New Jersey.

Güriş, Selahattin, Saçaklı Saçıldı, İrem and Genç, Elif Güneren. (2015), Determining The Effects of Factors on FDI in Global Crisis Period. *International Journal of Economics and Financial Issues*, Vol.5, No.1, pp. 1-10

Ji, Qiu, Quanmin, Zhu and Dongya, Zhao. (2015), Application of Hidden Markov Model to Locate Soccer Robots. *Proceedings of the 34th Chinese Control Conference, Hangzhou Chinese*, July 28-30.

Kar, Muhsin and Tatlısöz, Fatma. (2008), Türkiye'de Doğrudan Yabancı Sermaye Hareketlerini Belirleyen Faktörlerin Ekonometrik Analizi. *KMU İİBF Dergisi*, Year.10, No.14, pp.1-23.

Öz, Ersoy. (2009), *Saklı Markov Modelleri ve Finansal Bir Uygulama.* PhD Thesis, Marmara University and Institute of Social Sciences.

Petrushin, Valery A. (2000), Hidden Markov Models: Fundamentals and Applications, Part 2: Discrete and Continuous Hidden Markov Models. *Center for Strategic Technology Research Accenture*, Northbrook, 4-7, USA, pp.1-12.

Pardo Bryan and Birmingham, William. (2005), Modeling Form For On-Line Following of Musical Performances. *Proceeding of the Twentieth National Conference on Artificial Intelligence*, Pittsburgh Pennsylvania, July 9-13.

Rabiner, Lawrence R. (1989), A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceeding of the IEEE*, Vol.77, No.2, pp. 257-265.

Sarısoy, İdris and Koç, Selçuk. (2011), Doğrudan Yabancı Sermaye Yatırımlarının Kurumlar Vergisi Gelirleri Üzerindeki Etkisinin Ekonometrik Analizi, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* No.36, pp. 133-153.

Tarı Recep and Bıdırdı, Hanife. (2009), Türkiye'de Doğrudan Yabancı Sermaye Yatırımlarının Temel Belirleyicileri: 1990-2006 Dönemine İlişkin Ekonometrik Analiz. *Dumlupınar Üniversitesi, Sosyal Bilimler Dergisi*, No.24, pp. 253-268.

Vaseghi, Saeed V. (2007), Multimedia Signal Processing Theory and Applications in Speech, Music and Comunications. John-Wiley&Sons Ltd, 364, England.