DETERMINATION OF INFLATION RATE IN A HIDDEN MARKOV MODEL FRAMEWORK: TURKEY CASE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DİLEK AYDOĞAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FINANCIAL MATHEMATICS

JUNE 2015

Approval of the thesis:

DETERMINATION OF INFLATION RATE IN A HIDDEN MARKOV MODEL FRAMEWORK: TURKEY CASE

submitted by **DİLEK AYDOĞAN** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Financial Mathematics, Middle East Technical University** by,

Prof. Dr. Bülent Karasözen
Director, Graduate School of **Applied Mathematics** ______________

Assoc. Prof. Dr. Ali Devin Sezer
Head of Department, **Financial Mathematics** ______________

Assoc. Prof. Dr. Sevtap Kestel
Supervisor, **Actuarial Sciences, METU** ______________

Assoc. Prof. Dr. Yeliz Yolcu Okur
Co-supervisor, **Financial Mathematics, METU** ______________

Examining Committee Members:

Assoc. Prof. Dr. Sevtap Kestel
Actuarial Sciences, METU ______________

Prof. Dr. Inci Batmaz
Department of Statistics, METU ______________

Assoc. Prof. Dr. Fatih Tank
Department of Statistics, Ankara University ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: DİLEK AYDOĞAN

Signature :

ABSTRACT

DETERMINATION OF INFLATION RATE IN A HIDDEN MARKOV MODEL FRAMEWORK: TURKEY CASE

Aydoğan, Dilek
M.S., Department of Financial Mathematics
Supervisor : Assoc. Prof. Dr. Sevtap Kestel
Co-Supervisor : Assoc. Prof. Dr. Yeliz Yolcu Okur

June 2015, 60 pages

Inflation is a significant issue that is cared by every segment of society but especially by economists since it plays an important role in the economic problems of countries. In Turkey, inflation has been a substantial problem since 1970s. There is no stability at inflation rates from that time to 2003. The Central Bank of the Republic of Turkey became independent and started to apply implicit inflation target regime at 2001. As a result of these improvements inflation was taken under control at 2004. Although the rates are more stable after that time, there are some fluctuations still. Hence, it is essential to study on the inflation data of 2004-2014 year interval.

In this study, a novel approach is preferred in order to analyze, model and predict the inflation data of Turkey for the mentioned period. Hidden Markov Models are applied in several fields but there are few applications of it to analyze inflation. Now, theoretical background of HMMs will be provided. Then, the advantage of performing well on autocorrelated data just like ours of the model will be taken. The monthly inflation rates that represent the price level changes at current month with respect to the same month of previous year will be accepted as observed part of the model and the hidden part is estimated by EM-algorithm approach adapted to normal-HMM. The first stage model will be applied to the data for the years 2004-2012 and the test period of 2013-2014.

Keywords : Inflation, Markov Models, Hidden Markov Models, Turkish Inflation Rates, Autocorrelation

ÖZ

SAKLI MARKOV MODEL ÇERÇEVESİNDE ENFLASYON ORANI SAPTAMASI: TÜRKİYE DURUMU

Aydoğan, Dilek
Yüksek Lisans, Finansal Matematik Bölümü
Tez Yöneticisi : Doç. Dr. Sevtap Kestel
Ortak Tez Yöneticisi : Doç. Dr. Yeliz Yolcu Okur

Haziran 2015, 60 sayfa

Enflasyon toplumun her kesimini ilgilendiren bir sorun olmanın yanında, ülkelerin ekonomik problemlerinde önemli bir rol oynadığı için özellikle ekonomistlerin ilgisini çeken bir konudur. Türkiye'de ise enflasyon 1970'lerden itibaren önemli bir sorun haline gelmiştir ve bu tarihten 2003 yılına kadar durağanlıktan uzak bir seyir izlemiştir. 2001 yılında ise Türkiye Cumhuriyeti Merkez Bankası (TCMB) bağımsız olmuş ve açık enflasyon hedeflemesi politikası kabul edilmiştir. Bu gelişmelerin sonucunda enflasyon 2004 yılında kontrol altına alınmıştır. Bu tarihten itibaren enflasyon oranları daha durağan bir seyir izlemiştir. Ancak, enflasyon oranında hala bazı dalgalanmalar devam etmektedir. Bu nedenle 2004-2014 verisi üzerine çalışmaya gerek duyulmuştur.

Bu çalışmada enflasyonu analiz etmek, modellemek ve tahminlemek amacıyla orijinal bir yaklaşım tercih edilmiştir. Saklı Markov Modellerin (SMM) farklı birçok alanda uygulamaları olmasına rağmen enflasyon verilerini modelleme konusunda çok az sayıda çalışma mevcuttur. Bu çalışmada SMM'nin teorik altyapısı anlatılacaktır. Çalışmada kullanılacak enflasyon verisinin otokorelasyonlu olduğu göz önüne alınırsa bu modelin otokorelasyonlu veriler üzerinde başarılı olma avantajı kullanılacaktır. Bir önceki yılın aynı ayına göre hesaplanan enflasyon oranları modelimizde görülebilir seri olarak kullanılacak ve saklı seri normal-SMM'ye uyarlanmış EM- algoritması ile elde edilecektir. Model 2004-2012 yılları arasındaki enflasyon verisi üzerinde uygulanacak, 2013-2014 yılları için tahminleme yapılacak ve var olan değerlerle karşılaştırılarak test edilecektir.

Anahtar Kelimeler: Enflasyon, Markov Modeller, Saklı Markov Modeller, Türkiye Enflasyon Oranları, Otokorelasyon

To My Family

ACKNOWLEDGMENTS

I would like to thank my thesis supervisor Assoc. Prof. Dr. Sevtap Kestel and my thesis co-supervisor Assoc. Prof. Dr. Yeliz Yolcu Okur for introducing me to the subject Hidden Markov Model and their patient guidance, enthusiastic encouragement and valuable advices during the development and preparation of this thesis. Their willingness to give their time and to share their experiences have brightened my path.

I thank also all members of the Institute of Applied Mathematics of Middle East Technical University for their kindness and help.

Furthermore, I would like to specially thank my friend Mr. Deniz Kenan Kılıç for his friendship, support and help.

Finally, I would like to express my pleasure to having such a great family and thank each of them for their endless love and existence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACF	Auto-correlation Function
AD	Aggregate Demand
AE	Planned Aggregate Expenditure
AIC	Akaike Information Criteria
ARIMA	Autoregressive Integrated Moving Average Model
AS	Aggregate Supply
BIC	Bayes Information Criteria
BVAR	Bayesian Vector Autoregressive Model
C	Consumption
CPI	Consumer Price Index
EM	Expectation Maximization
FED	Federal Reserve System
G	Government Purchases
HMM	Hidden Markov Model
I	Planned Investment
IHMM	Infinite Hidden Markov Model
L	Likelihood
LAS	Long-run Aggregate Supply
LRPC	Long-run Phillips Curve
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
MPE	Mean Percentage Error
PCE	Personal Consumption Expenditure
PPI	Producer Price Index
RMSE	Root Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average Model
VAR	Vector Autoregressive Model
M^d	Money Demand
M^s	Money Supply

$\mathbb{N}$	Natural Numbers
O	Observation
Pr	Probability
$\mathbb{R}$	Real Numbers
SDHP-HMM	Stick Double Hierarchical Dirichlet Process Hidden Markov Model
SRPC	Short-run Phillips Curve
T	Net Taxes
X	State
Γ	Transition Probability Matrix
δ	Initial Distribution
μ	Mean
σ	Standard Deviation

CHAPTER 1

INTRODUCTION

Inflation can be described as an increase in the general price level just to clarify with the most basic and commonly used definition of it. It is a common problem facing both developing and developed countries. While some countries cope with the high inflation, some others complain about the very low inflation rates.

For developing countries as Turkey, high inflation is an important macroeconomic problem that continues for long periods. [2] Because, high inflation causes the uncertainty over future inflation as Cukierman and Meltzer presented [10]. This uncertainty may discourage investments and savings also it may cause shortages of goods if consumers stock with the expectation that prices will increase [22]. Moreover, the distribution of income may be destroyed since the ability of purchasing power of people living with a fixed income falls.

Therefore, it is important to analyze, model and forecast the inflation. In order to support the importance of this issue the following quote from a report of FED is presented [19]:

”Control of inflation is at the core of monetary policymaking and, consequently, central bankers have a great interest in reliable inflation forecasts to help them achieving this aim. For other agents in the economy accurate inflation forecasts are likewise of importance, either to be able to assess how policymakers will act in the future or to help them in forming their inflation expectations when negotiating about wages, price contracts and so on. And in the academic literature inflation predictability is assessed to get a gauge on the characteristics of inflation dynamics in general.”.

Therefore, in this thesis the inflation data of Turkey is analyzed, modeled and predicted. For this purpose Normal Hidden Markov models are preferred since the inflation data of Turkey is non-stationary and HMMs can deal with this type of data.

1.1 General Information about the Hidden Markov Model

There are several definitions of HMM. One of the best of them is represented by Rabiner [38]. According to him the HMMs are the extended types of Markov models. While in Markov models each state corresponds to an observable event, in HMMs observations are probabilistic functions of states. An HMM is embedded stochastic process with an underlying and unobservable stochastic process. It can be observed by only a stochastic process that produce the same observation sequence. Cheng-Der Fuh [17] says that "A hidden Markov model is defined as a parametrized Markov chain in a Markovian random environment,with the underlying environmental Markov chain viewed as missing data.". The key idea of HMM is to decribe a probability distribution over an infinite number of possible sequences of observations [12].

Since HMMS are general-purpose models they can be used for modeling various types of time series such as continuous-valued, circular, multivariate, as well as binary data, bounded and unbounded counts and categorical observations [28]. Some of the applications of HMMs are speech recognition [37], DNA sequence analysis [9], molecular biology [26], stock market forecasting [21] and economics [20].

The EM algorithm and the Viterbi algorithm are the solutions to the main problems of HMM. The EM algorithm is the most famous algorithm for getting the parameters that maximize the likelihood of the model. On the other hand, the best sequence of hidden states are estimated by the Viterbi algorithm.

1.2 A Brief History of Inflation in Turkey

The inflation of Turkey started to grow up in the 1960s. However, Turkish economy came up against conspicuously high inflation rates firstly in the 1970s. Between 1970 and 1980 the price per barrel of crude oil was increased from 2.74 to 11.65 dollars because of the oil crisis in the world. Moreover, in this period there were exporting problems in the economies. As a result of these problems, in 1979 the inflation rate reached to 62%. Although the 24 January decisions had entered into the applications in 1980 because of the second petrol crisis the inflation rate exceed the percentage of 101 in 1980. However, after a while the new program showed result by decreasing the inflation rate to 28% in 1982. In the upcoming years inflation again increased and reached to 73% in 1988. It is also significant that the inflation reached to top points in the general election periods 1983,1987 and 1991 [13].

The near inflation history of Turkey can be divided into four periods according to the inflation values and the characteristics of the economy. In the first period that is between 1989-1993 , prices were influenced by financial liberalization and the inflation rates were around 60%. The second period was started with the 1994 crisis and continued till 1999. Inflation reached to an extreme value of 125% in 1994 because of the exchange market crisis. The austerity plan introduced in 1994 decreased the inflation rates temporarily but it could not avoid the imbalances in the economy. Although it decreased to 72% in 1995, it again peaked to 100% in 1997. Consequently, it reduced to 60% in the end of this period. The exchange-rate based stabilization program is applied in the third period that is between 2000 and 2001 [25] [11]. The aim of the Central Bank with this policy was preventing exchange rate rises. Although another target of the Central Bank is to deal with inflation in this period, constant exchange rate policy did not allow it.

The most important part of the inflation history of Turkey for this thesis is the last period. This period can be also called as the inflation targeting period. The Central Bank determined its main target to maintain price stability. The inflation rate that was 70% in 2000 decreased to one digit values in 2004 by the affect of the new policy and after that time although there have been still fluctuations, they are not as big as in previous periods. The differences between the last period and the previous ones can be also seen in Figure 1.1. Moreover, in order to view the data that will be used from now on more detailed Figure 1.2 is represented.

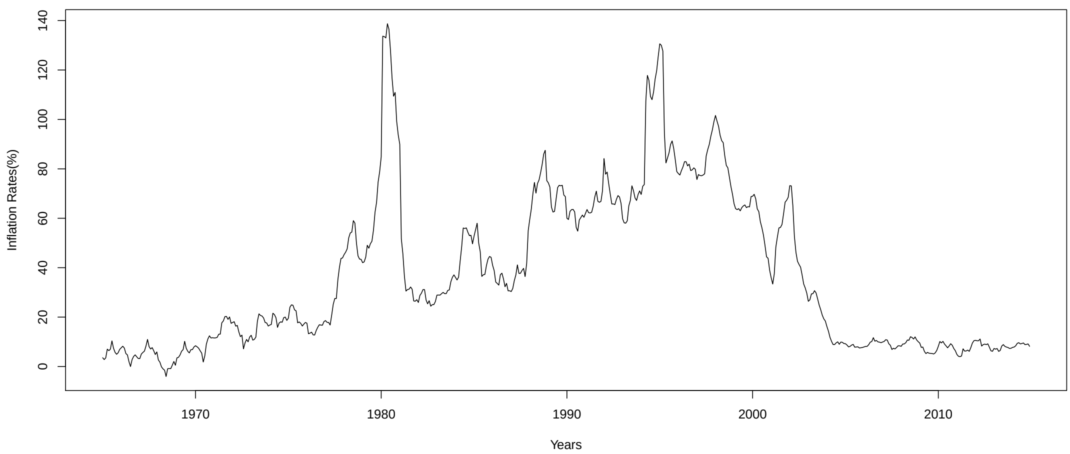


Figure 1.1: Inflation rates of Turkey between 1965-2014

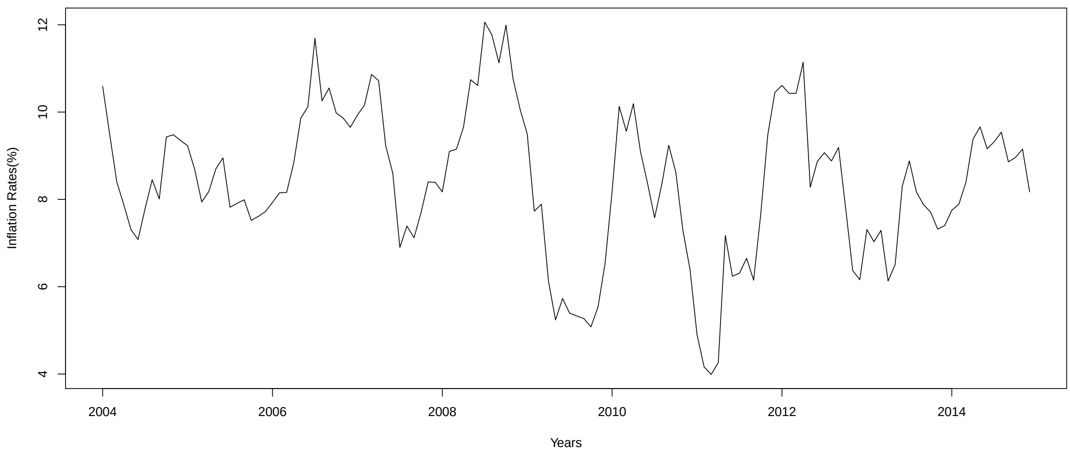


Figure 1.2: Inflation rates of Turkey between 2004-2014

1.3 Aim of the Study

The main aim of this study is to model, analyze and forecast the inflation rates of Turkey between 2004-2014. The HMM is preferred to apply on inflation rates since the data is autocorrelated. In fact, this application is a novel approach. Because, there are a few studies about the modeling the inflation in Turkey. These studies use VAR, Bayesian VAR, ARIMA and SARIMA models. However, there are a few studies in the world about the modeling the inflation rates by using HMM. These studies categorize the inflation rates and then use the Poisson-HMMs. This study benefits from the advantage of working of HMMs on non-stationary data as the inflation data of Turkey. Moreover, in order to study on the original data Normal-HMM is used.

In this study, general structure of HMMs is explained by taken "Hidden Markov Models for Time Series" by Walter Zucchini and Iain L. MacDonald as a reference and examples are added in order to give a chance to the readers for understanding the topic more clearly. The R-codes that are adapted to continuous and normally distributed data in this thesis are created by inspiring from the codes in the book of Zucchini and MacDonald [45] and they can be helpful for future studies on similar data. The codes of the algorithm is written in R and the algorithm is presented in the Appendix A. Moreover, it is aimed to give a brief explanation about the economical structure of the inflation. Especially for non-economists these synoptic and basic explanations can be useful.

This master thesis is comprised of five chapters. Chapter 1 draws a frame for Hidden Markov Model and represents the inflation history of Turkey. Then, introduces the aim of the study and summarizes the literature survey. Chapter 2 explains the inflation as an economic indicator. It explains the computing methods via price indices and gives the relation of inflation with the aggregate supply and aggregate demand. Moreover, the Philips curve which shows the relationship between inflation and unemployment is explained in this chapter. The methodology of HMM is explained in Chapter 3. Firstly, the Markov chain is mentioned since it is one of the main part (hidden part) of the HMM. Then, EM-algorithm which estimates the parameters of the model and Viterbi algorithm which finds the most probable states of the model are explained. The HMM, which is explained in detail in this chapter , is applied to the 2004-2014 inflation data of Turkey and the results of the application are represented in Chapter 4. Lastly, Chapter 5 finalizes this thesis with a conclusion and comments.

1.4 Literature Survey

Although there is no study about modeling inflation data of Turkey by using HMM, there are some studies that use other methods. Meçik and Karabacak(2011) [31] models inflation of Turkey also. They practice upon the CPIs between 2003-2011 period. In contrast with our study they use the seasonally adjusted data. The aim of their study is determining the best ARIMA model for the data and forecasting the CPIs of 2010-2012 period. SARIMA(1,0,0) is selected as the best model according to AIC and BIC. Then, the forecasting results and accuracy of the model are represented in that study. Lastly, it is indicated that the model is succeeding according to these performance indicators. Saz, G. (2011) [41] models the inflation rates in Turkey between 2003 and 2009. The monthly CPIs are used as the inflation measure unlike our study . He indicates that the seasonality in the Turkish inflation rate is both deterministic and stochastic and he uses SARIMA models for forecasting the inflation rates. Hyndman-Khandakar (HK) algorithm is used to derive the proposed SARIMA models. The most appropriate model chosen by the study is SARIMA(0,0,0)(1,1,1), i.e., a seasonal autoregressive moving average model with one degree of seasonal integration, one seasonal autoregressive and one seasonal moving average part. Akdoğan et al. (2012) [33] make forecasts for the inflation in Turkey by using several econometric models. They use the percentage changes in the CPI as in our study. However, they use quarterly inflation rates while we use monthly ones. In this study, the inflation rates between 2003:Q1 and 2009:Q3 are used as the training sample while the rates between 2009:Q4 and 2011:Q2 are used as the forecasting sample. They use univariate models, decomposition based approaches, a Phillips curve motivated time varying parameter model, VAR, BVAR and dynamic factor model. Consequently, the study concludes that the individual models that include more economic information model the inflation better than the benchmark random walk model. In particular, BVAR model is the best fitting model of the inflation rates of Turkey.

There are several studies that focus on modeling inflation by using HMM in the world. Hossain, Ahmed and Rabbi (2012) [22] analyzes inflation of different time periods. They use monthly inflation rates. Firstly, they convert continuous inflation data to discrete data by separating it into classes. Three hidden states are chosen as increasing, no change and decreasing to model the inflation rates. Then, they train the HMM for different time periods and test them on 2010-2011 inflation data in order to detect the similarities of inflation behavior among previous years. Andre Inge (2013) [23] in his thesis uses the Sweden 1831-2012 inflation data to build the HMM on. He uses the yearly percentage changes in CPI as an inflation measure. He also estimates parameters by using EM-algorithm and determines the most suitable HMM by using AIC and BIC as in our study. The most suitable HMM for the Sweden inflation rates are chosen as 5-state HMM in the study. Then, he forecasts the 2013 inflation rate and the result is in the confidence interval of the expectations of the Sweden Central Bank. Jochman (2010) [24] applies the Infinite Hidden Markov Model (IHMM) to the U.S. inflation data in order to find the structural breaks. In this study IHMM is preferred because it does not predetermine the state number before but learns it from the data and so it allows modeling time series with unknown number of structural breaks. He uses personal consumption expenditure (PCE) deflator as the price level and computes inflation quarterly by the formula of $\pi_t = 400ln\frac{P_t}{P_{t-1}}$ where π_t is inflation rate and P_t is price level at time t. The sample of this study is from 1953:Q1 to 2009:Q3. The study catches the structural breaks during financial crises. Song Yong (2011) [42] uses also an IHMM to model the U.S. inflation data. In his study in addition to study of Jochman(2010), a structure is presented to allow getting information about the parameter of the conditional data density in each state. This structure is named as the sticky double hierarchical Dirichlet process hidden Markov model (SDHDP-HMM). He makes comparison between this model with the existing alternative regime switching and structural change models. Then, he concludes that SDHDP-HMM is robust to model uncertainty and makes better forecasts than regime switching and structural break models according to the results of the application to U.S. inflation data.

CHAPTER 2

INFLATION AS AN ECONOMIC INDICATOR

In this chapter, the economical basis of inflation is explained, briefly. Firstly, the definition of inflation is given. Then, price indices that are used to calculate inflation are presented. After the indices, the relationship between the inflation and the pair of aggregate supply and aggregate demand are explained with the structure of them. Lastly, in order to mention the exchange relation between unemployment and inflation the Phillips curve is examined.

2.1 Definition of Inflation

There is no consensus about the definition of inflation but there are several alternative definitions. In this chapter, the most accepted ones are presented. The following definition is one of the most commonly used one:"Inflation is a process of continuously rising prices, or equivalently, of continuously falling value of money." [27]. Bronfenbrenner and Holzman also define inflation as a rise in the price level,i.e, depreciation of the monetary unit [7]. Namely, it is the reducement of purchasing power of consumers. Although this definition has information about signs of inflation, it does not present any of causes and effects of it. Hence, Bronfenbrenner and Holzman explain inflation by four more detailed definitions:

- Inflation is the situation of buying few goods of too much money because of the generalized excess demand.
- Inflation is a rise of money stock or money income.
- Inflation is a rise in price levels with additional characteristics or conditions: it is incompletely anticipated; it leads (via cost increases) to further rises; it does not increase employment and real output; it is faster than some "safe" rate; it arises "from the side of money"; it is measured by prices net of indirect taxes and subsidies; and/or it is irreversible.
- Inflation is a fall in the external value of money as measured by foreign exchange rates, by the price of gold, or indicated by excess demand for gold or foreign exchange at official rates.

According to monetarist definition of Friedman [14]: ”Inflation is always and everywhere a monetary phenomenon and can be produced only by a more rapid increase in the quantity of money than in output.”.

Turvey [43] explains inflation as ”the process resulting from competition in attempting to maintain total real income,total real expenditure, and/or total output at a level which has become physically impossible, or attempting to increase any of them to a level which is physically impossible.”.

It is possible to display much more definitions according to the different views of streams of thoughts, but these distinctions are already examined at following parts in detail.

2.2 Measurement of Inflation: Price Indices

Measuring inflation requires measuring price changes of goods and services that do not result form changes in the value such as volume, quality or performance. Since single price changes cannot represent general inflation in an overall economy, the price of a large ”basket” of representative goods and services is compared over time. The price indices are used for this purpose [32].

A price index is a weighted average of the prices of a selected basket of goods and services relative to their prices in some base-year. The ratio of the goods in the basket are mentioned as weight. One of the things to take account is updating the basket regularly in order to prevent losing representative quality of it. The other significant point is that the base year should not be an unordinary year. Because, in such years for example in a wartime prices can be extremely high.

Laspeyres and Paasche Indices are the two main price indices that are used to measure inflation. These two indices diverge from each other at the point of choosing quantities of the basket as the base year or current year quantity. These indices are explained in detail in the following parts.

2.2.1 Laspeyres Index

Laspeyres index LP is a base year weighted index and it shows the relative change in the cost of a basket originally purchased in a base period [16]. If there are n goods in the basket and the price of good i in period t and the quantity of good i in period t denoted respectively as p_i^t and x_i^t, then the formula of Laspeyres Index at time period t (LP_t) is as follows,

$$LP_t = \frac{\sum_{i=1}^{n} p_i^t x_i^0}{\sum_{i=1}^{n} p_i^0 x_i^0} \cdot 100 \tag{2.1}$$

The Laspeyres index overestimates the rise in the general price level. Because it does not take into account that firms and households demand smaller quantities of goods whose relative prices increases according to their substitutions and they prefer these substitute goods. Hence it gives too much weight to relatively expensive goods and gives too little weight to relatively cheap ones. Although it has such a weakness, it is the most common price index in statistics. Furthermore, it is also used when calculating TÜFE(GPI), YI-ÜFE(Domestic PPI) and YD-ÜFE (Foreign PPI) indices which are used in order to measure the inflation in Turkey.

2.2.2 Paasche Index

The Paasche Index PP is a current year based index because while it determines the basket of goods, it takes the end period as the reference point. Hence, it does not have the same weakness of Laspeyres index. However, Paasche index overestimates total expenditures in the base year and underestimates the rise in the general price level [16]. The Paasche Index at time period t (PP_t) can be summarized by the following formula,

$$PP_t = \frac{\sum_{i=1}^{n} p_i^t x_i^t}{\sum_{i=1}^{n} p_i^0 x_i^t} \cdot 100 \tag{2.2}$$

where n is the number of goods in the basket, t is the time period, p_i^t is the price of good i in period t and x_i^t is the quantity of good i in period t.

2.2.3 Fisher Ideal Price Index

The Fisher ideal price index (IP) is the geometric mean of Laspeyres and Paasche indices,

$$IP_t = \sqrt{LP_t \cdot PP_t} \tag{2.3}$$

It is clear that this index uses quantities of base and current period as weight. Hence, the value of the index is between the value of Laspeyres and Paasche index and it is closer to the actual course of inflation. However, it does not have common usage because it has no direct economic interpretation [1].

Example 2.1. Suppose our basket of goods includes only three items: butter, bread, milk. The base year is determined as 2008 and the current year is 2015. The prices and quantities of the items according to base and current years are displayed in the following table,

	price(TL)		quantity	
commodity	2008	2015	2008	2015
butter	10	15	10	20
bread	2	3	35	55
milk	2	4	45	40

The following table is constructed in order to ease the index computation,

Now, Laspeyres, Paasche and Fisher Indices are computed for this scenario as follows,

	2008		2015					
commodity	p^0	x^0	p^t	x^t	p^0x^0	p^tx^0	p^0x^t	p^tx^t
butter	10	10	15	20	100	150	200	300
bread	2	35	3	55	70	105	110	165
milk	2	45	4	40	90	180	80	160
					260	435	390	625

- Laspeyres Index: $LP_t = \frac{\sum_{i=1}^{n} p_i^t x_i^0}{\sum_{i=1}^{n} p_i^0 x_i^0} \cdot 100 = \frac{435}{260} \cdot 100 = 167$
- Paasche Index:$PP_t = \frac{\sum_{i=1}^{n} p_i^t x_i^t}{\sum_{i=1}^{n} p_i^0 x_i^t} \cdot 100 = \frac{625}{390} \cdot 100 = 160$
- Fisher Index:$IP_t = \sqrt{L_P \cdot P_P} = \sqrt{167 \cdot 160} = 163$

2.3 Aggregate Demand, Aggregate Supply and Inflation

In this section in order to determine the equilibrium price level in the economy, aggregate demand and aggregate supply curves are put together. Moreover, the affect of the economy on prices and inversely the affects of the prices on the economy are examined.

2.3.1 Aggregate Demand

Aggregate demand is the total demand for all goods and services in an economy. The aggregate demand curve comprises of the points that represents different price and output combinations associated with simultaneous equilibrium in the money and output markets [3]. When aggregate demand curve is derived, it is assumed that the government does not intervene in the economy. Namely, government purchases (G), net taxes (T) and money supply (M^S) variables are assumed as constant.

Before deriving the AD curve, for non-economists the following brief descriptions can be needed,

- Money Supply (M^S): It is a vertical line because central banks of countries are authorized controlling the amount of money in the economy and they do not prefer to associate money supply to interest rate, generally.
- Money Demand (M^D): Firms and households hold less money in order to take the advantage of bonds when interest rates are high. Namely, there is an inverse relationship between the interest rate and the quantity of money demanded. Hence, the money demand curve has negative slope.
- Planned Aggregate Expenditure (AE): It is the sum of aggregate consumption spending by households (C), planned investments by business firms (I) and government purchases of good and services(G).

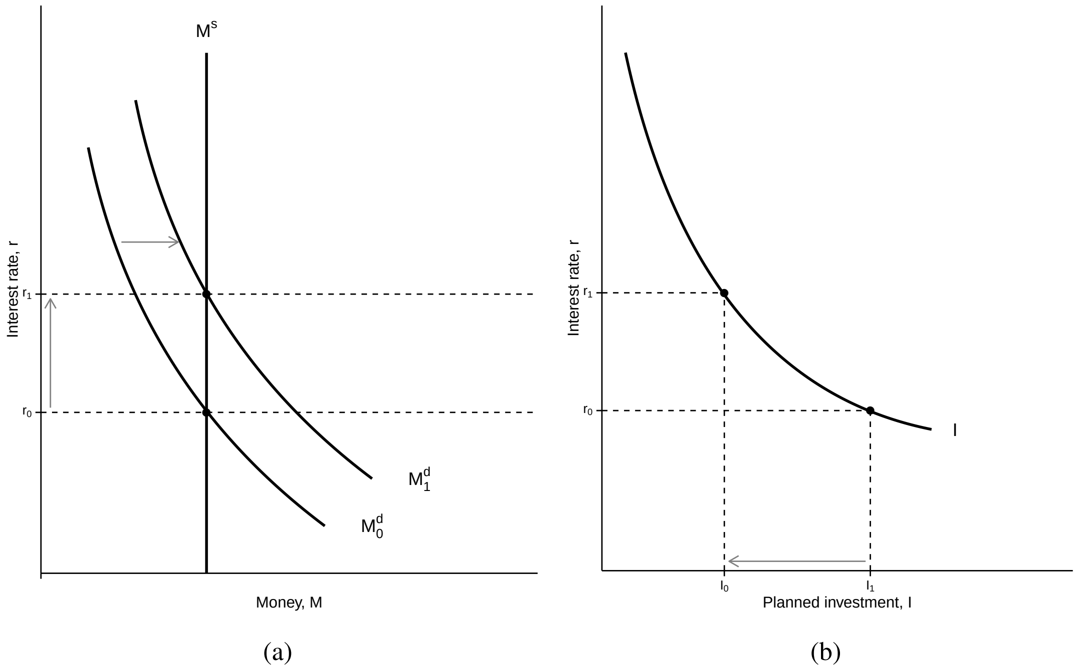


Figure 2.1: Aggregate expenditure curve.

When the price level is increased, demand for money (M^D) will increase. Because households and firms prefer to keep more money in high price levels. In Figure 2.1a increase at the money demand is displayed by shifting M^D curve to the right. Since money supply does not change, interest rate will increase. This rise in interest rate decreases the investment as displayed in Figure 2.2b As a result of the fall in the planned investment planned aggregate expenditure (AE) decreases and equilibrium output reduces from Y_0 to Y_1 [8].

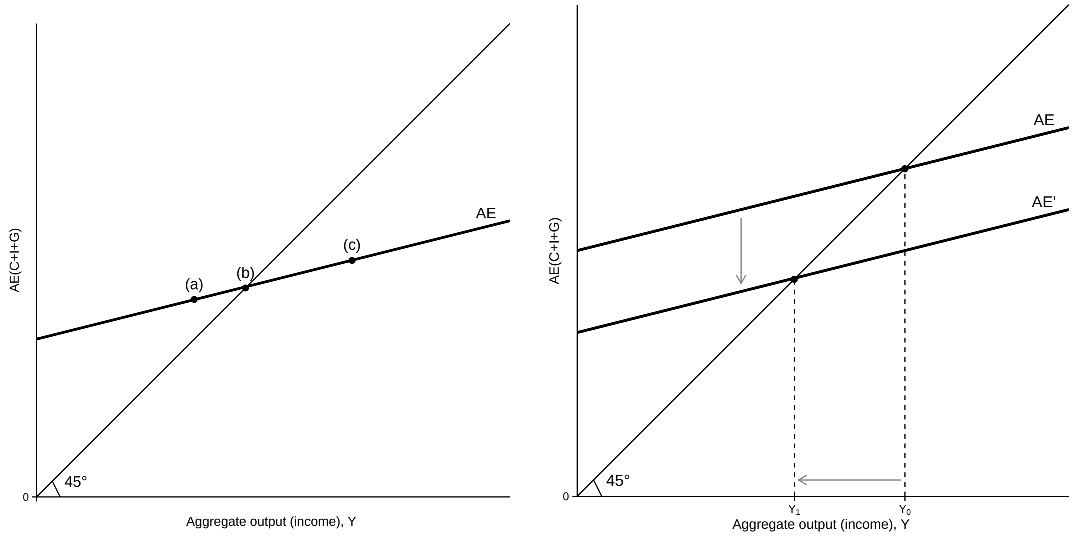
In Figure 2.2a at the points on the 45 degree line, planned aggregate expenditure is equal to aggregate output. Hence, point (b) is the equilibrium point while there is an inflationary gap at point (a) and deflationary gap at point (c).

It becomes reversed when there is a fall in price level. The money demand and correspondingly interest rates decrease . This situation causes a increase in the planned investment. Hence, planned aggregate expenditures increase. Consequently, equilibrium output rises.

Finally, it can be said that when the price level increases, the output decreases. However, a decrease in price level causes a rise in output level.

The negative relationship between price level and aggregate output is displayed with aggregate demand (AD) curve, as seen from Figure 2.3. The money and good market are in the equilibrium at all the points on the AD curve.

When there is a change in the variables G, T and M^s that are assumed as constant at the beginning of this part, the aggregate demand curve will shift:



(a) (b)

Figure 2.2: Aggregate expenditure curve.

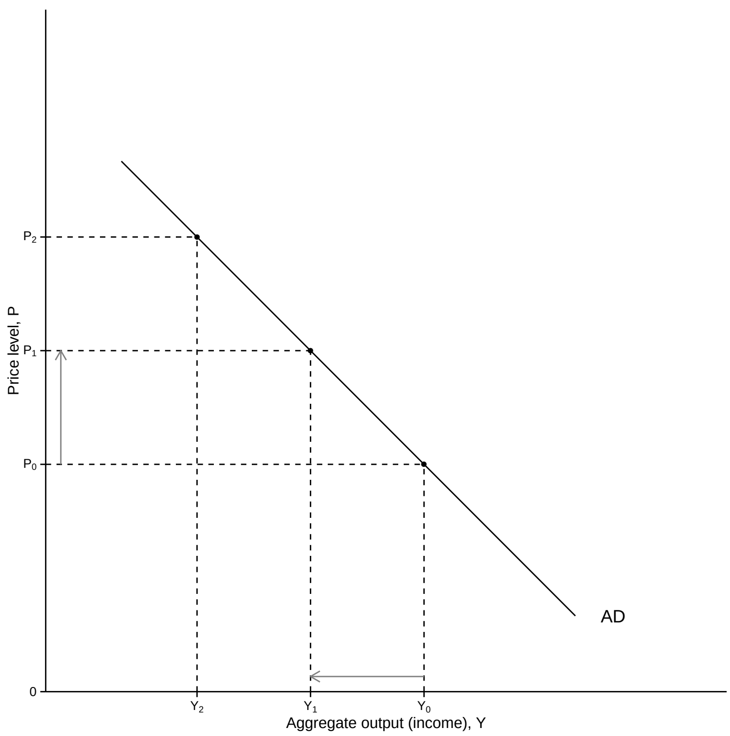


Figure 2.3: Aggregate demand curve

- Change in money supply (M^s): When the quantity of money is expanded at a given price level, M^s curve will shift to right and interest rates will fall. Then, planned investment and relatedly planned aggregate expenditure will rise. Hence, equilibrium output will increase at the given price level. Namely, increased money supply will shift the AD curve to the right. When the quantity of money is restricted, the situation is reversed.

- Change in government purchases (G): When the government purchases increase at a given price level, the planned aggregate will rise and so output will increase. In other words, an increase in G shifts the AD curve to the right.

- Change in net taxes (T): A decrease in net taxes at a given price level causes an increase in the consumptions, which will raise the planned aggregate expenditure and output. Namely, an decrease in T also shifts the AD curve to the right.

Table 2.1 summarize the situations that shifts AD curve to the right or left.

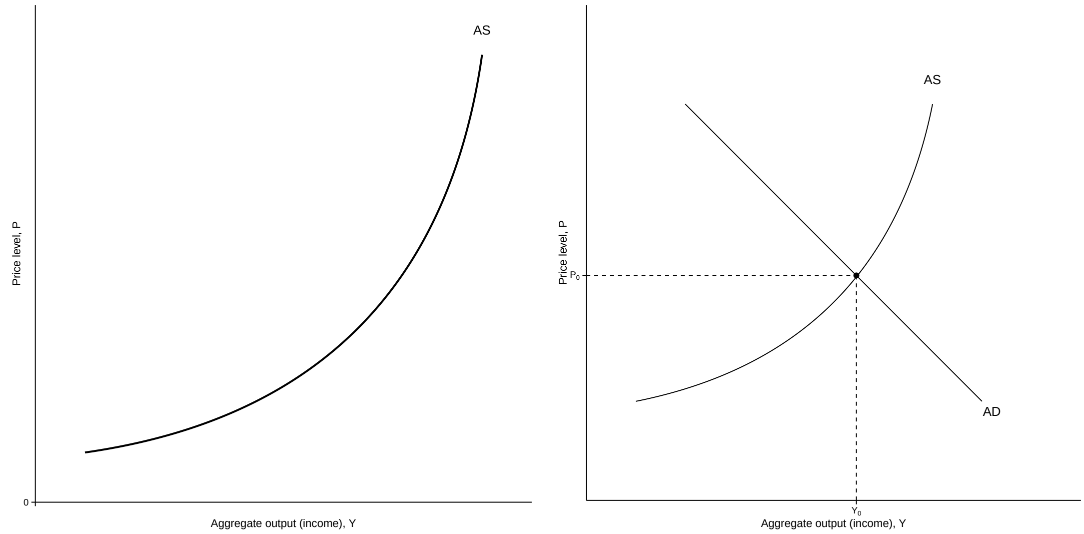
Table 2.1: The factors shifting AD curve

AD curve shifts to the right	AD curve shifts to the left
increase in M^s	decrease in M^s
increase in G	decrease in G
decrease in T	increase in T

2.3.2 Aggregate Supply

Aggregate supply (AS) is the total supply of goods and services in an economy for a given period of time. The aggregate supply curve shows the output is supplied by the economy at different price levels. There is disagreement about the shape of the AS curve but many says that for a short term period it has a positive slope and for low output level the curve is nearly flat, while for high level of outputs it is nearly horizontal as in Figure 2.4a. Because, when the output in the economy is low, the economy and firms are in the excess capacity. Moreover, there is cyclical unemployment in this output level. Hence, firms can satisfy the increasing demand without extra costs and get enough labor force without much increase in wage rates since there are unemployed people that want to get a job. Namely, the economy can produce more output with little increase in price level when it is at low output level. However, at a level of output the economy comes to the full employment point. Then, it cannot produce any further output and the increasing demand only rises the prices. In other words, the AS curve becomes vertical.

Any factor that changes the firms decisions shifts the AS curve. Some of them are displayed at Table 2.2 [8]. Lower costs and wages enable firms to produce more at same price level. Economic growth is also rises the maximum capacity of the economy. Moreover, supply-side public policies prompt people to work and entrepreneurship. Lastly, the better conditions met during production like good weather increases supply.



(a) (b)

Figure 2.4: Aggregate supply curve and the equilibrium price

Hence, all of these factors shift AS curve to the right while the exact opposite situations cause left shift.

The equilibrium price in the economy is at the intersection points of AD and AS curves. P_0 and Y_0 in Figure 2.4b are the coordinates of the equilibrium point. This point corresponds to firms price output decisions since it is on the AS curve. Moreover, at this point the money and good markets are in the equilibrium since it is on the AD curve.

The AS curve has a positive slope in the short term because a gap occurs when the prices get higher since the wages are sticky and cannot rise as much as prices. Therefore, firms produce more output when the price level rises. Although the cost gap increases the output level for a short-term, many economists agree that in the long term these gap closes. For example, wages rise if the inflation is stable in the long period since the costs increase again the output return previous level. For this reason, the long-term aggregate supply (LAS) curve is vertical as in the Figure 2.5. In the figure the equilibrium point A removes to point B and so output rises from Y_0 to Y_1 by increasing price level from P_0 to P_1. However, after a while the costs increase such as wages and supply curve shifts to the left. The new equilibrium point moves to point C. Hence, the output go back to its previous level but the prices rise.

The output point that the long-term aggregate supply curve corresponds name as $potential$ $output$. It is the output level that the economy stand there in the long run without inflation.

Table 2.2: The factors shifting the AS curve

AS curve shifts to the right	AS curve shifts to the left
Lower Costs lower input prices lower wage rates	**Higher Costs** higher input prices higher wage rates
Economic Growth more capital more labor technological change	**Stagnation** capital deterioration
Public Policy supply-side policies tax cuts deregulations	**Public Policy** waste and inefficiency over-regulation
Good Weather	**Bad weather**

2.3.3 Types of Inflation

According to causes of inflation, two types of it can be mentioned as demand-pull and cost-push inflation.

2.3.3.1 Demand-Pull Inflation

Inflation ensues from the increasing of demand is called as demand-pull inflation. If the economy is in the full capacity when the demand increase occurs, these demand rise only will cause a rise in price levels but the output level will not change. Furthermore, as mentioned before if the long-term AS curve is vertical, even the economy is not in full capacity the output level will not increase but only the price level will increase in the long term.

2.3.3.2 Cost-Push Inflation

The inflation is the result of increases of costs is called as cost-push inflation. When the costs rise, the AS curve shifts to the left. Then, price level increases and beside that the output level decrases. Appearence of these two problems together is called as *stagflation*. Even though the government intervenes the economy (increase in G or M^s or decrease in T) the output come to its back level but the prices increase more. Therefore, the costs shocks are significant problems for policy makers.

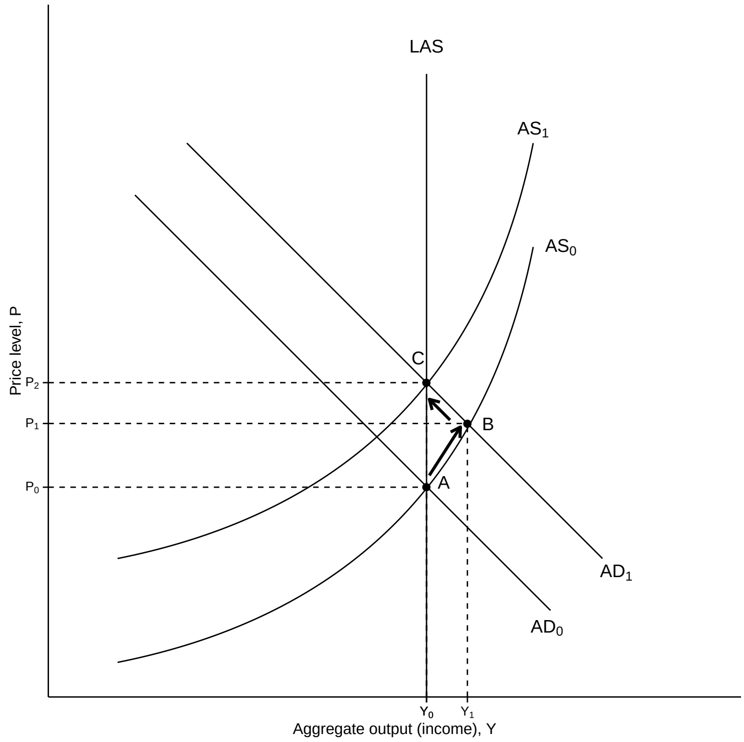


Figure 2.5: Long-run aggregate supply curve (LAS)

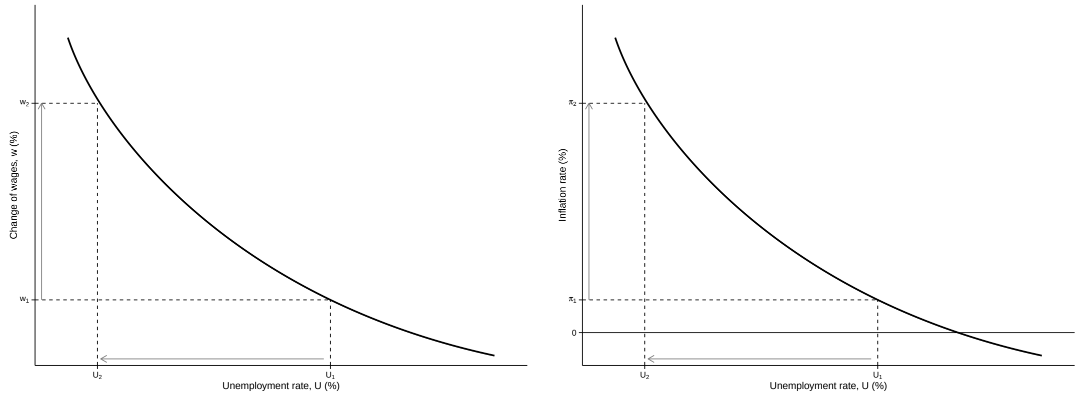
2.4 Phillips Curve

The remarkable article of A.W. Phillips that is about the relationship between unemployment and inflation is issued at 1958 and from that time it is one of the most important source of the macroeconomic studies [35]. Many economists use the Philips curve concept in order to explain the relationship between demand expansion, inflation, unemployment, wages and prices. In this part, the classical type of Phillips curve which examines the relation of unemployment with the wage changes and the developed types adapted to inflation relation in short and long term are presented.

2.4.1 Classical Version of Phillips Curve

According to classical Philips curve there is a negative relationship between wages and unemployment which is displayed in Figure 2.6a . In the plots of Phillips the curves that shows the relation between unemployment and the percentage changes in wages are non-linear and have negative slope. It is explained by Philips (1958) [35] as follows:

”When the demand for labour is high and there are very few unemployed we should expect employers to bid wage rates up quite rapidly, each firm and each industry being continually tempted to offer a little above the prevailing rates to attract the most



(a) (b)

Figure 2.6: Classical version of Phillips curve

suitable labour from other firms and industries. On the other hand it appears that workers are reluctant to offer their services at less than prevailing rates when demand for labour is low and unemployment is high so that wage rates fall only very slowly. The relationship between unemployement and the rate of change of wage rates is therefore likely to be highly non-linear."

Samuelson and Solow (1960) [40] presented the Phillips curve that is the representer of the relation between the price-level and unemployment firstly. This relation is seen in Figure 2.6b. Now, it is the most popular version of the Phillips curve. Because it allows economic policy makers to formulate policy programs with alternative combinations of unemployment and inflation rates [16].

2.4.2 Long-term Phillips Curve

The long-term Phillips curve is first posed by Milton Friedman [15]. He agrees with the classical Philips curve in the short term. He describes the long-term Phillips curve as vertical line in his study in 1977. He explains the reason of that by giving the following example:

In the case of an unexpected acceleration in the aggregate nominal demand, the producer thinks that this increase is special to his product. Then, he wants to produce more since he expects that the price of his product will be higher than the future market price. Hence, he is willing to pay higher nominal wages to attract additional workers. Because the real wage according to his expectations are lower since he perceives that price as higher than before. From the view point of workers the situation is different. A rise in nominal wages is perceived by the workers as a rise in the real wages. However, when the aggregate nominal demand and prices continue to increase, the initial effect disappears and is reversed for both producers and workers. They find themselves locked into inappropriate contracts. Ultimately, the unemployment level goes back to previous level but the prices do not go back.

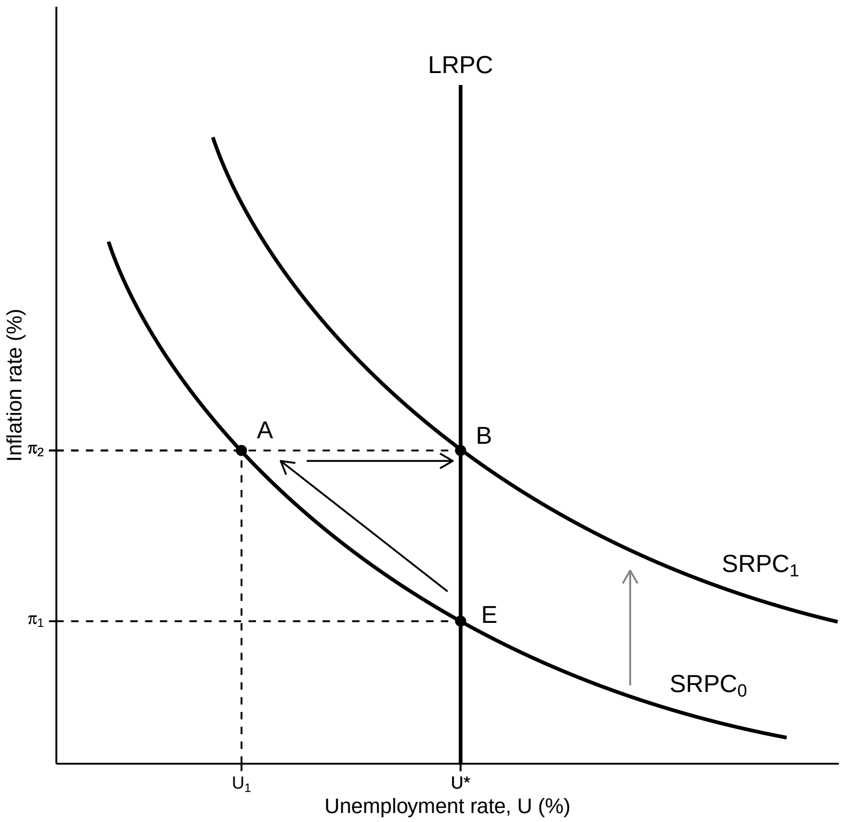


Figure 2.7: Long-run Phillips curve

Figure 2.7 summarizes that in the beginning the market is in the equilibrium at the point E. When the aggregate demand rises , prices increase and the unemployment decreases. Namely, the economy moves to point A. However, when the workers and producers realize that they are mistaken, the unemployment moves to its natural rate U* at the last price level. In other words, short term Phillips curve shifts to the right and the equilibrium point moves from point A to B.

In this chapter, in order to give a general information about inflation the measurement method and the economical structure of it are explained. Now, the hidden Markov model will be presented in the next chapter since it is the other component of the major subject of this thesis.

CHAPTER 3

HIDDEN MARKOV MODEL

In this chapter, firstly the Markov Chain will be introduced since the unobservable part of the Hidden Markov Model is a Markov chain. After the explanation of Markov Chain, the structure of Hidden Markov Model will be presented. Then, Expectation Maximization (EM) algorithm that find the parameters maximizing the likelihood of the model will be introduced. Lastly, Viterbi Algorithm will be explained in order to decode the hidden states.

3.1 Markov Chain

Markov chain is a special type of stochastic processes and it is named after the series of papers about theories of finite state Markov chain starting in 1906 that are published by Russian mathematician Andrei Andreyevich Markov [30]. Today, it is very useful to capture the nature of stochastic processes of financial and economic variables. Furthermore, it is in use of solving the problems in physics, chemistry, speech recognition, internet applications, social sciences, music, genetics and many other fields.

Definition 3.1. *(Markov Chain)* It is a stochastic process $X = \{X_t; t = 0, 1, ...\}$ that provides following property for all $t \in N$ and for each $j \in E$ where E is a discrete state space,

$$Pr\{X_{t+1} = j|X_0 = i_0, X_1 = i_1, ..., X_t = i_t\} = Pr\{X_{t+1} = j|X_t = i_t\}. \qquad (3.1)$$

Let $X^{(t)}$ denote $\{X_0 = i_0, X_1 = i_1, ..., X_t = i_t\}$ for convenience, then the Equation (3.6) becomes as follows:

$$Pr\{X_{t+1} = j|X^{(t)}\} = Pr\{X_{t+1} = j|X_t = i_t\} \qquad (3.2)$$

Namely, the next outcome of the system only depends on the current outcome not on previous ones.

To give an example, assume that there is a game between A and B. Each of them have five fair and unbiased coins and they toss the coin again and again. If it comes up tails A gives B a coin, if heads vice versa. It ends when one of them has no coin. Observe that if A has 4 coins currently, next time the probability of holding 5 coins is 0.5 regardless of past outcomes. Hence, future is independent from past given present. Then, it can be clearly seen that the sequence of heads and tails patterns an i.i.d. stochastic process and the sequence of total number of coins that A or B hold patterns a Markov chain. [29]

Definition 3.2. *(Transition Probabilities)* They are the probabilities p_{ij}'s that identify the probability of moving to state j from current state i.

The Markov chain is called $homogenous$ if the probability of transition from state i to state j is always constant. The property of being a homogenous Markov chain is displayed also in the following equation,

$$Pr(X_t = j | X_{t-1} = i) = p_{ij}, \quad t = 2, 3, \cdots$$

That is, one-step transition probabilities do not change as time goes on.

Definition 3.3. *(Transition Probability Matrix)* It is the square matrix with (i, j) element p_{ij} and it is essential to observe that its row sums equal to 1:

$$\Gamma = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix} \tag{3.3}$$

where m denotes the number states of the Markov chain.

Also, note that in fact Γ is one-step transition probability matrix and it can be implied by $\Gamma(1)$. If t-step transition probability matrix is required, Chapman-Kolmogorov equations imply that it can be calculated by taking the t th power of $\Gamma(1)$:

$$\Gamma(t) = \Gamma(1)^t$$

Definition 3.4. *(Unconditional Probabilities)* They are the elements of the row vectors that are denoted by:

$$\pi(t) = (Pr(X_t = 1)...Pr(X_t = m)), \quad t \in N \tag{3.4}$$

where $Pr(X_t = j)$ is the probability of being in a given state at a given time t.

$\pi(1)$ is the *initial distribution* of Markov chain. It will be denoted as δ in this study. The distribution at time $t + 1$ can be find out by multiplying the distribution at time t by the transition probability matrix Γ:

$$\pi(t + 1) = \pi(t)\Gamma \tag{3.5}$$

The Markov chain is said to have *stationary* transition probabilities if it is homogenous and additionally provides following property:

$$\Gamma\delta = \delta \quad and \quad \delta 1^{'} = 1$$

where δ is a row vector with non negative elements. In order to illustrate the definitions of transition probability matrix, initial probabilities and unconditional probabilities Example 3.1 is represented.

Example 3.1. Consider that there is a town that has had only one market (Market A) and now another market (Market B) is ready to provide service after today. Then, define state 0 as a customer trades with old market A and state 1 as a customer trades with new market B, and the transition probabilities are given by the below table.

It can be seen from the table that the probability of remaining loyal to market A is 1/4 and to market B is 2/3 . Thus, the number of customers of the markets is a two-state Markov chain with the transition probability matrix Γ:

$$\Gamma = \begin{bmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{bmatrix}$$

	day t+1	
day	Market A	Market B
Market A	1/4	3/4
Market B	1/3	2/3

Now, assume that at the beginning there are 240 customers of market A ($N_0(1) = 240$) and market B has no customer ($N_1(1) = 0$) i.e., the distribution of this day's customer is

$$\begin{aligned} \pi(1) &= (Pr(X_1 = 0) \quad Pr(X_1 = 1)) \\ &= (1 \quad 0) \end{aligned}$$

This distribution is the initial distribution of the Markov chain.

The distribution of next day's customer and the day after next day and so on, can be computed as follows:

$$\begin{aligned} \pi(2) &= (Pr(X_2 = 0) \quad Pr(X_2 = 1)) \\ &= \pi(1)\Gamma \\ &= (1/4 \quad 3/4) \end{aligned}$$

$$\begin{aligned} N_0(2) &= p_{00}N_0(1) + p_{10}N_1(1) \\ &= 60 \end{aligned}$$

$$\begin{aligned} N_1(2) &= p_{01}N_0(1) + p_{11}N_1(1) \\ &= 180 \end{aligned}$$

$$\begin{aligned} \pi(3) &= (Pr(X_3 = 0) \quad Pr(X_3 = 1)) \\ &= \pi(2)\Gamma \\ &= (5/16, 11/16) \end{aligned}$$

$$
\begin{aligned}
N_0(3) &= p_{00}N_0(2) + p_{10}N_1(2) \\
&= 75
\end{aligned}
$$

$$
\begin{aligned}
N_1(3) &= p_{01}N_0(2) + p_{11}N_1(2) \\
&= 165
\end{aligned}
$$

These distributions represent the unconditional probabilities of states for times 2 and 3.

3.1.1 Estimating Transition Probabilities

In order to estimate transition probabilities EM-Algorithm will be used in this thesis.A basic way for this aim is represented in this section.

One basic way of estimating transition probabilities is by counting the transitions. To illustrate this method a modified version of an example from Zucchini and MacDonald (2009) [45] is used.

Example 3.2. Assume that there is a 4-state Markov chain as follows:

3434313414 1241221113 2134241434 4222332413 2143221331 1344242213 2244422424 2223231312 1224331431 1222433242

Then, the matrix of transition counts as below:

$$
\begin{pmatrix}
4 & 5 & 8 & 4 \\
6 & 12 & 3 & 10 \\
6 & 7 & 4 & 6 \\
5 & 8 & 7 & 4
\end{pmatrix}
$$

To illustrate, observe that the entry of (2,4) is equal to the number of transitions from state 2 to state 4 that is 10. Now, the transition probabilities can be calculated by dividing these transition counts by total number of transitions from related state.For example, $\Gamma(2, 4) = 10/31$ since the total number of transitions from 2 is 31. Hence, the transition probability matrix Γ is

$$
\Gamma = \begin{pmatrix}
4/21 & 5/21 & 8/21 & 4/21 \\
6/31 & 12/31 & 3/31 & 10/31 \\
6/23 & 7/23 & 4/23 & 6/23 \\
5/24 & 8/24 & 7/24 & 4/24
\end{pmatrix}.
$$

3.2 Hidden Markov Model

Hidden Markov Model is a statistical tool for modeling a stochastic process by associating them to a hidden (unobservable) process. The hidden process follows a Markov

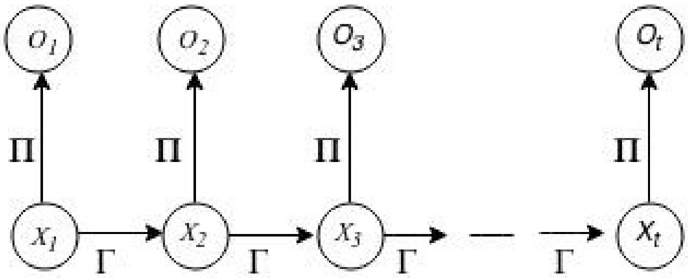


Figure 3.1: Dependence structure for a hidden Markov model

Chain and the observed data depends only on current state of the hidden process. This structure can be expressed as follows:

$$Pr(X_t|X^{(t-1)}) = Pr(X_t|X_{t-1}), \quad t = 2, 3, ... \tag{3.6}$$

$$Pr(O_t|O^{(t-1)}, X^{(t)}) = Pr(O_t|X_{t-1}), \quad t \in N \tag{3.7}$$

where O_t is the observation and X_t is the state at time t.

It can be clearly seen from this structure that HMM does not have a long memory.

In order to construct HMM's hidden part the initial probability δ and transition probability matrix Γ that are explained before are used. However, for constructing the relation between observations and the hidden states Definition 3.5 is provided.

Definition 3.5. *(Emission Probabilities)* They are the probabilities of observing a particular value provided that the system is in one of the hidden states [4].

$$p_i(o) = Pr(O_t = o|X_t = i) \tag{3.8}$$

m distributions p_i's refer to the $state - dependent\ distributions$ of the model. In this study, $P(o)$ will refer to the following diagonal matrix consisting of emission probabilities of observation o:

$$P(o) = \begin{pmatrix} p_1(o) & & 0 \\ & \ddots & \\ 0 & & p_m(o) \end{pmatrix}$$

3.2.1 Distribution of Observations

The distribution of observations O_t is as follows,

$$Pr(O_t = o) = \sum_{i=1}^{m} Pr(X_t = i)Pr(O_t = o|X_t = i).$$

or equally, it can be represented as,

$$Pr(O_t = o) = \sum_{i=1}^{m} \pi_i(t) p_i(o)$$

where $\pi_i(t) = Pr(X_t = i)$.

Then, it can be rewritten in the matrix form:

$$\begin{aligned} Pr(O_t = o) &= (\pi_1(t)...\pi_m(t)) \begin{pmatrix} p_1(o) & & 0 \\ & \ddots & \\ 0 & & p_m(o) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \pi(t) P(o) 1'. \end{aligned} \tag{3.9}$$

Lastly, the distribution takes its final form by substituting $\pi(1)\Gamma^{t-1}$ or equally $\delta\Gamma^{t-1}$ for $\pi(t)$ by the help of Equation 3.5 and representing the diagonal matrix with P(o),

$$Pr(O_t = o) = \delta\Gamma^{t-1} P(o) 1'. \tag{3.10}$$

For higher-order marginal distributions, for example for bivariate and trivariate distributions the equation can be extended to,

$$Pr(O_t = v, O_{t+k} = w) = \pi(t) P(v) \Gamma^k P(w) 1', \tag{3.11}$$

$$Pr(O_t = v, O_{t+k} = w, O_{t+k+l} = z) = \pi(t) P(v) \Gamma^k P(w) \Gamma^l P(z) 1' \tag{3.12}$$

respectively.

3.2.2 The likelihood

The likelihood of a HMM is the probability of observing the sequence $O = o_1, o_2, ..., o_T$ with respect to the parameters of the model,i.e.,

$$\begin{aligned} L_t &= Pr(O^{(T)} = o^{(T)}) \\ &= \sum_{x_1,...,x_T=1}^{m} Pr(O^{(T)} = o^{(T)}, X^{(T)} = x^{(T)}) \end{aligned}$$

Now, note that the joint distribution of random variables is as follows:

$$Pr(Z_1, Z_2, ..., Z_n) = \prod_{i=1}^{n} Pr(Z_i | parents(Z_i)) \tag{3.13}$$

where parents of random variable Z_i is a minimal set of predecessors of Z_i in the total ordering such that the other predecessors of Z_i are conditionally independent of Z_i given parents(Z_i) [36].

Then, by Equation 3.13 the likelihood is

$$
\begin{aligned}
L_T &= Pr(O^{(T)}, X^{(T)}) \\
&= Pr(X_1) \prod_{k=2}^{T} Pr(X_k|X_{k-1}) \prod_{k=1}^{T} Pr(O_k|X_k)
\end{aligned} \tag{3.14}
$$

$$
\begin{aligned}
L_T &= \sum_{x_1,\dots,x_T=1}^{m} (\delta_{x_1}\gamma_{x_1,x_2}\gamma_{x_2,x_3}...\gamma_{x_{T-1},x_T})(p_{x_1}(o_1)p_{x_2}(o_2)...p_{x_T}(o_T)) \\
&= \sum_{x_1,\dots,x_T=1}^{m} \delta_{x_1}p_{x_1}(o_1)\gamma_{x_1,x_2}p_{x_2}(o_2)\gamma_{x_2,x_3}...\gamma_{x_{T-1},x_T})...p_{x_T}(o_T)
\end{aligned}
$$

Hence, the final form of the likelihood is as follows:

$$
L_T = \delta P(o_1)\Gamma P(o_2)\Gamma P(o_3)...\Gamma P(o_T)1'. \tag{3.15}
$$

In order to ease the understanding of computation of HMM's likelihood example 3.3 is given.

Example 3.3. Assume that there is a class and the teacher of mathematics sets a game in his mind. He thinks that hidden states of students are to understand (U) and not to understand (N) the subject at that moment. Additionally, the observation symbols are looking at the teacher (L) and writing on notebook (W). He also sets the initial distribution, transition and emission matrix as follows,

$$
\delta = (\delta_U \quad \delta_N) = \begin{pmatrix} 0.95 & 0.05 \end{pmatrix}
$$

Transition Matrix

	U	N
U	0.7	0.3
N	0.4	0.6

Emission Matrix

	U	N
L	0.6	0.9
W	0.4	0.1

Notice that,

$$
P(o_1) = P(L) = \begin{pmatrix} p_1(L) & 0 \\ 0 & p_2(L) \end{pmatrix} = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.9 \end{pmatrix}
$$

Similarly,

$$
P(o_2) = P(W) = \begin{pmatrix} 0.4 & 0 \\ 0 & 0.1 \end{pmatrix}
$$

and, $P(o_3) = P(L), P(o_4) = P(L), P(o_5) = P(W)$

Additionally, $\Gamma P(L) = \begin{pmatrix} 0.42 & 0.27 \\ 0.24 & 0.54 \end{pmatrix}, \Gamma P(W) = \begin{pmatrix} 0.28 & 0.03 \\ 0.16 & 0.06 \end{pmatrix}$

Then, he observes a particular student per minutes along 5-minute and gets the observation sequence L,W,L,L,W. By computing the probability of observing that sequence according to his model ,which is actually an HMM, he comes to the likelihood of the model,

$$\begin{aligned}
L_T &= \delta P(L)\Gamma P(W)\Gamma P(L)\Gamma P(L)\Gamma P(W)1' \\
&= \begin{pmatrix} 0.95 & 0.05 \end{pmatrix} \begin{pmatrix} 0.6 & 0 \\ 0 & 0.9 \end{pmatrix} \begin{pmatrix} 0.28 & 0.03 \\ 0.16 & 0.06 \end{pmatrix} \begin{pmatrix} 0.42 & 0.27 \\ 0.24 & 0.54 \end{pmatrix} \begin{pmatrix} 0.42 & 0.27 \\ 0.24 & 0.54 \end{pmatrix} \begin{pmatrix} 0.28 & 0.03 \\ 0.16 & 0.06 \end{pmatrix} 1' \\
&= 0.025
\end{aligned}$$

3.3 HMM Parameter Estimation

A commonly used method to estimate parameters of HMM is EM algorithm, which stands for expectation maximization. It is also known as Baum-Welch algorithm. This algorithm is rather suitable for the data that have missing parts like HMM's hidden states. Moreover, it enables to estimate parameters of models with not stationary but homogenous Markov chains. The key idea here is to start with an initial distribution and by iterating getting new distributions. At each step the distribution is better or the same. However, this does not mean that the algorithm produces the best distribution since it can reach the local maximum.

In order to apply this estimation method we need the backward and forward probabilities. Hence, forward and backward probabilities are clarified firstly. Then, EM estimation is explained.

3.3.1 Forward and Backward Probabilities

The forward-backward algorithm is used to find the probability $Pr(X_t = k|O^{(t)})$ of being in a specific state k at a particular time t for a given sequence of observations.

3.3.1.1 Forward Probabilities

Forward probabilities are defined as the elements of the vector α_t displayed in the equation (3.16).

$$\alpha_t = \delta P(o_1)\Gamma P(o_2)\Gamma P(o_3)...\Gamma P(o_t) \tag{3.16}$$

Now, remember the likelihood function of HMM at equation (3.15). Then, the likelihood formula is,

$$L_T = \alpha_T 1'.$$

Additionally, note that it can be deduced from the definition of forward probabilities that $\alpha_t = \alpha_{t-1}\Gamma P(o_t)$ for $t = 1, ..., T$

Moreover, forward probabilities also refer to the probabilities of producing $O^{(t)}$ while ending up in state j:

For t=1,2,...,T and j=1,2,...,m

$$\alpha_t(j) = Pr(O^{(t)} = o^{(t)}, X_t = j) \tag{3.17}$$

3.3.1.2 Backward Probabilities

Backward probability also affects $Pr(X_t = k|O^{(t)})$ as well as forward probability and is defined by

$$\beta_t' = \Gamma P(o_{t+1})\Gamma P(o_{t+2})...\Gamma P(o_T)1' \tag{3.18}$$

for $t = 1, ..., T$ From this definition, it can be concluded that $\beta_t' = \Gamma P(o_{t+1})\beta_{t+1}'$ for $t = 1, ..., T-1$.

It is also equal to the probability of producing the observations $o_{t+1}, ..., o_T$ given that the system is at state i at time t.

For t=1,2,...,T-1 and i=1,2,...,m,

$$\beta_t(i) = Pr(O_{t+1} = o_{t+1}, O_{t+2} = o_{t+2}, ..., O_T = o_T|X_T = i) \tag{3.19}$$

For convenience, the vector $(O_k, O_{k+1}, ..., O_l)$ is denoted by O_k^l, then the equation (3.19) becomes as follows:

$$\beta_t(i) = Pr(O_{t+1}^T = o_{t+1}^T|X_T = i) \tag{3.20}$$

Observe that backward probabilities are conditional probabilities while the forwards are the joint ones.

Moreover, note that since there is no observation after time T, β_T can be regarded as the probability of observing a empty set given the system is at ith state at time T. Hence, $\beta_T(i) = 1$.

3.3.1.3 Inferences of Forward and Backward Probabilities

Let's multiply the two types of probabilities for state i:

$$\alpha_t(i)\beta_t(i) = Pr(O_1^{(t)}, x_T = i)Pr(O_{t+1}^{(T)}|X_t = i)$$

$$= Pr(X_t = i)Pr(O_1^t|X_t = i)Pr(O_{t+1}^T|X_t = i)$$

Then, by using the conditional independence of O_1^t and O_{t+1}^T given X_t, the equation is evaluated as follows:

$$\alpha_t(i)\beta_t(i) = Pr(X_t = i)Pr(O_1^t, O_{t+1}^T|X_t = i).$$

Equally, it can be rewritten as,

$$\alpha_t(i)\beta_t(i) = Pr(O^{(T)}, X_t = i) \tag{3.21}$$

By summation of equation (3.21), we get the following inference also,

$$\begin{aligned}\alpha_t\beta_t' &= Pr(O^{(T)} = o^{(t)}) \\ &= L_T\end{aligned} \tag{3.22}$$

The other inferences essential for applying EM algorithm are as follows

$$Pr(X_t = j|O^{(T)} = o^{(t)}) = \frac{\alpha_t(j)\beta_t(j)}{L_T} \tag{3.23}$$

$$Pr(X_{t-1} = j, X_t = k|O^{(T)} = o^{(t)}) = \frac{\alpha_{t-1}(j)\gamma_{jk}p_k(o_t)\beta_t(k)}{L_T} \tag{3.24}$$

Now, an example will be given in order to make the computations of forward and backward probabilities more clear.

Example 3.4. The forward and backward probabilities of Example 3.3 are computed in this example. Forward probabilities are as follows,

$$\begin{aligned}\alpha_1 &= \delta P(o_1) = \begin{pmatrix}0.57 & 0.045\end{pmatrix} \\ \alpha_2 &= \alpha_1\Gamma P(o_2) = \begin{pmatrix}0.167 & 0.020\end{pmatrix} \\ \alpha_3 &= \alpha_2\Gamma P(o_3) = \begin{pmatrix}0.075 & 0.056\end{pmatrix} \\ \alpha_4 &= \alpha_3\Gamma P(o_4) = \begin{pmatrix}0.045 & 0.050\end{pmatrix} \\ \alpha_5 &= \alpha_4\Gamma P(o_5) = \begin{pmatrix}0.021 & 0.004\end{pmatrix}\end{aligned}$$

and, the backward ones are,

$$\beta_5' = \begin{pmatrix}1 \\ 1\end{pmatrix}$$

$$\beta_4' = \Gamma P(o_5)1' = \begin{pmatrix}0.31 \\ 0.22\end{pmatrix}$$

$$\beta_3' = \Gamma P(o_4)\beta_4 = \begin{pmatrix}0.190 \\ 0.193\end{pmatrix}$$

$$\beta_2' = \Gamma P(o_3)\beta_3 = \begin{pmatrix}0.132 \\ 0.150\end{pmatrix}$$

$$\beta_1' = \Gamma P(o_2)\beta_2 = \begin{pmatrix}0.041 \\ 0.030\end{pmatrix}$$

Now, observe that

$$\alpha_1\beta_1' = \alpha_2\beta_2' = ... = \alpha_5\beta_5' = 0.025 = L_T.$$

It is clearly seen that this result supports the inference in the equation (3.22).

3.3.2 The EM Algorithm

The EM algorithm is an iterative method that is used to find the maximum-likelihood estimation of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values [5]. The logic of the algorithm is after giving initial values to parameters, applying two main steps of the model repeatedly. In the first step, which is the expectation (E) step, for given observations and current estimation of parameters the conditional expectations of missing data are computed. In fact, in this step the likelihood of the complete data that already includes conditional expectations of hidden data is computed for convenience. The following step is maximization (M) step and in this step the log-likelihood of the complete data is maximized. Then, the parameters that are found at maximization step is used at the next expectation step and this iteration method is repeated until the convergence. In the case of HMM, by identifying u and v as follows,

$u_j(t) = 1$ if and only if $c_t = j$, for t=1,...,T and

$v_{jk}(t) = 1$ if and only if $c_{t-1} = j$ and $c_t = k$ for t=2,...,T.

the log-likelihood of the complete data that consists of the observed and missing ones

$$\begin{aligned} log(Pr(o^{(T)}, x^{(T)})) &= log(\delta_{x_1} \prod_{t=2}^{T} \gamma_{x_{t-1},x_t} \prod_{t=1}^{T} p_{x_t}(o_t)) \\ &= log\delta_{x_1} + \sum_{t=2}^{T} log\gamma_{x_{t-1},x_t} + \sum_{t=1}^{T} logp_{x_t}(o_t) \end{aligned}$$

can be rewritten as,

$$log(Pr(o^{(T)}, x^{(T)})) = \sum_{j=1}^{m} u_j(1) log\delta_j + \sum_{j=1}^{m}\sum_{k=1}^{m}(\sum_{t=2}^{T} v_{jk}(t)) log\gamma_{jk} + \sum_{j=1}^{m}\sum_{t=1}^{T} u_j(t) logp_j(o_t) \tag{3.25}$$

In the expectation step instead of $v_{jk}(t)$ and $u_j(t)$ use the conditional expectations of being in a state at a particular time given the observations:

$$\begin{aligned} \hat{u}_j(t) &= Pr(X_t = j | o^{(T)}) \\ &= \frac{\alpha_t(j)\beta_t(j)}{L_T} \end{aligned} \tag{3.26}$$

$$\begin{aligned} \hat{v}_{jk}(t) &= Pr(X_{t-1} = j, X_t = k | o^{(T)}) \\ &= \frac{\alpha_{t-1}(j)\gamma_{jk} p_k(o_t)\beta_t(k)}{L_T} \end{aligned} \tag{3.27}$$

Then, in the maximization step maximize the complete data likelihood by maximizing each term of equation (3.25) and observe that the first term depends only on initial distribution δ. Here, the second term depends on transition probability matrix Γ and the third term depends on the parameters of state dependent distributions.

Namely, maximize the following components of the complete data log-likelihood,

1. $\sum_{j=1}^{m} \hat{u}_j(1) log\delta_j$ with respect to δ,
2. $\sum_{j=1}^{m} \sum_{k=1}^{m} (\sum_{t=2}^{T} \hat{v}_{jk}(t)) log\gamma_{jk}$ with respect to Γ,
3. $\sum_{j=1}^{m} \sum_{t=1}^{T} \hat{u}_j(t) logp_j(o_t)$ with respect to state-dependent parameters.

The maximizing values of relevant parameters as follows,

1. $\delta_j = \frac{\hat{u}_j(1)}{\sum_{j=1}^{m} \hat{u}_j(1)} = \hat{u}_j(1)$
2. $\gamma_{jk} = \frac{\sum_{t=2}^{T} \hat{v}_{jk}(t)}{\sum_{k=1}^{m} \sum_{t=2}^{T} \hat{v}_{jk}(t)}$
3. Since the third term is on state-dependent distributions, this part varies by the type of the distribution. For give some examples, the maximizing values of this term for Poisson-HMM are,

$$\hat{\lambda}_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t) o_t}{\sum_{t=1}^{T} \hat{u}_j(t)}$$

and for Normal-HMM are,

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t) o_t}{\sum_{t=1}^{T} \hat{u}_j(t)}$$

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} \hat{u}_j(t)(o_t - \mu_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)}$$

3.4 Conditional Distributions

In this section, the conditional distribution of the observation at a particular time t given observations at all other times is computed. The notation $O^{(-t)}$ is used at computations for convenience as,

$$O^{(-t)} \equiv (O_1, ..., O_{t-1}, O_{t+1}, ..., O_T)$$

The conditional distribution is the ratio of the likelihood of the observations with o instead of o_t and the likelihood of observations except for the observation o_t that is treated as missing [45]:

$$Pr(O_t = o | O^{(-t)} = o^{(-t)}) = \frac{Pr(O_t = o_t, O^{(-t)} = o^{(-t)})}{Pr(O^{(-t)} = o^{(-t)})} \tag{3.28}$$

$$= \frac{\delta P(o_1)\Gamma P(o_2)...\Gamma P(o_{t-1})\Gamma P(o)\Gamma P(o_{t+1})...\Gamma P(o_T)1'}{\delta P(o_1)\Gamma P(o_2)...\Gamma P(o_{t-1})\Gamma\Gamma P(o_{t+1})...\Gamma P(o_T)1'}$$

$$= \frac{\alpha_{t-1}\Gamma P(o)\beta_t'}{\alpha_{t-1}\Gamma\beta_t'}$$

Moreover, equation (3.28) can be written as the the mixture of the m state-dependent distributions,

$$Pr(O_t = o|O^{(-t)} = o^{(-t)}) = \sum_{i=1}^{m} \frac{f_i(t)}{\sum_{j=1}^{m} f_j(t)} p_i(o) \tag{3.29}$$

where $f_i(t)$ is the product of the ith entry of the vector $\alpha_{t-1}\Gamma$ and the ith entry of the vector β_t .

3.5 Forecast Distributions

Forecast distributions of HMM, which is another type of conditional distributions, are illustrated in this chapter. For this purpose the conditional distribution of O_{t+h} given observations O^T where h is the forecast horizon is computed as follows,

$$Pr(O_{T+h} = o|O^{(T)} = o^{(T)}) = \frac{Pr(O^{(T)} = o^{(T)}, O_{T+h} = o)}{Pr(O^{(T)} = o^{(T)})}$$

$$= \frac{\delta P(o_1)\Gamma P(o_2)\Gamma P(o_3)...\Gamma P(o_T)\Gamma^h P(o)1'}{\delta P(o_1)\Gamma P(o_2)\Gamma P(o_3)...\Gamma P(o_T)1'}$$

$$= \frac{\alpha_T\Gamma^h P(o)1'}{\alpha_T 1'}$$

$$= \rho_T\Gamma^h P(o)1'$$

where $\rho_T = \alpha_T/\alpha_T 1'$. Similar to conditional distributions, forecast distributions can also be written in the mixed form of state-dependent probability distributions,

$$Pr(O_{T+h} = o|O^{(T)} = o^{(T)}) = \sum_{i=1}^{m} \tau_i(h)p_i(o) \tag{3.30}$$

where $\tau_i(h)$ is the ith entry of the vector $\rho_T\Gamma^{(h)}$.

3.6 Decoding

In this chapter, two types of decoding method is examined. First one is local decoding which determines the most likely state at time t and the second one is global decoding which determines the sequence of most likely states.

3.6.1 State Probabilities and Local Decoding

The conditional distribution of state X_t given observations $O^{(T)}$ can be evaluated as

$$\begin{aligned} Pr(X_t = i | O^{(T)} = o^{(T)}) &= \frac{Pr(X_t = i, O^{(T)} = o^{(T)})}{Pr(O^{(T)}) = o^{(T)})} \\ &= \frac{\alpha_t(i)\beta_t(i)}{L_T}. \end{aligned} \tag{3.31}$$

The most likely state at time t is the state that maximize the conditional distribution of states given observations and it is represented in the following equation:

$$\begin{aligned} i* &= \underset{i=1,\dots,m}{argmax} Pr(X_t = i | O^{(T)} = o^{(T)}) \\ &= \underset{i=1,\dots,m}{argmax} \frac{\alpha_t(i)\beta_t(i)}{L_T}. \end{aligned} \tag{3.32}$$

3.6.2 Global Decoding and Viterbi Algorithm

Although the results of global decoding is very similar to local decoding, they are not identical. Global decoding has more common usage then local decoding. It does not deal with estimating most likely state for each separate time but estimates the most likely sequence of states. Namely, global decoding search for the state sequence that maximizes the conditional probability

$$Pr(X^{(T)} = x^{(T)} | O^{(T)} = o^{(T)}). \tag{3.33}$$

Equivalently, the joint probability

$$\frac{Pr(X^{(T)}, O^{(T)})}{Pr(O^{(T)} = o^{(T)})} = \delta_{x_1} \prod_{t=2}^{T} \gamma_{x_{t-1},x_t} \prod_{t=1}^{T} p_{x_t}(o_t) \tag{3.34}$$

can be used in order to determine the most likely sequence of HMM.

Recognize that there are m^T possible sequences of $(x_1, x_2, ..., x_T)$. Therefore, m^T function evaluations will be needed if the equation (3.33) is used in order to find the most likely state sequence. Hence, the decoding will not be feasible for large T. Therefore, a dynamic programming algorithm is needed for global decoding. Viterbi algorithm is used for that purpose as a dynamic programming algorithm. It can be applied to both stationary or not stationary underlying Markov chains.

The Viterbi Algorithm was developed to solve the decoding of convolutional codes by Andrew J. Viterbi in 1967 [44]. Then, Omura showed that it can be interpreted as a dynamic programming algorithm [34].Moreover, Omura and Forney [34] , [18] showed that the Viterbi algorithm is a maximum likelihood decoder [39].

In order to use the Viterbi algorithm in HMM, the following terms are defined firstly,

$$\begin{aligned}\varphi_{1i} &= Pr(X_1 = i, O_1 = o_1) \\ &= \delta_i p_i(o_1)\end{aligned} \tag{3.35}$$

and,

$$\varphi_{ti} = \max_{x_1, x_2, \dots, x_{t-1}} Pr(X^{(t-1)} = x^{(t-1)}, X_t = i, O^{(T)} = o^{(T)}) \tag{3.36}$$

for $t = 2, 3, ..., T$.

The relation between successive φ 's can be shown by following equation,

$$\varphi_{tj} = (\max_i(\varphi_{t-1,i}\gamma_{ij}))p_j(o_t) \tag{3.37}$$

for $t = 2, 3, ..., T$ and $i = 1, 2, ..., m$.

Then the most likely state sequence is estimated by

$$i_T = \underset{i=1,\dots,m}{argmax} \varphi_{Ti} \tag{3.38}$$

and,

$$i_t = \underset{i=1,\dots,m}{argmax}(\varphi_{ti}\gamma_{i,i_{t+1}}). \tag{3.39}$$

The following example illustrates the usage of Viterbi algorithm for global decoding.

Example 3.5. Assume that there are two coins with the property that one of them is fair and the other one is biased. 5 times a randomly chosen coin was tossed and all of them appear to be head (H). Moreover, the following information about the process is provided.

The initial distribution is

$$\delta = \begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$$

and, if it is assumed that choosing biased coin is state 1 while choosing fair one is state 2, the transition probability matrix is

$$\gamma_{ij} = \begin{pmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{pmatrix}$$

Lastly, the emission probabilities are as follows:

$$p_F(H) = p_F(T) = 0.5, p_B(H) = 0.8, p_B(T) = 0.2$$

Namely, the probability of coming head or tail is equal and it is 0.5 if the coin is fair. However, if the coin is biased, the probability of coming head is 0.8 while coming tail is 0.2.

The following table computes the multiplier of equation (3.37) only for observation H, since all of the observations are H in this example.

Table 3.1: Multipliers used for passing successive φ

i		γ_{ij}	$p_j(o_t)$	$\gamma_{ij}p_j(o_t)$
B	$\rightarrow$	0.4	0.8	0.32
	$\searrow$	0.6	0.5	0.30
F	$\nearrow$	0.3	0.8	0.24
	$\rightarrow$	0.7	0.5	0.35

Then, the φ 's can be computed as follows,

$$\varphi_{1i} = \delta_i p_i(o_1) = \begin{cases} \varphi_{11} = \delta_1 p_1(H) = 0.5 \times 0.8 = 0.40 \\ \varphi_{12} = \delta_2 p_2(H) = 0.5 \times 0.5 = 0.25 \end{cases}$$

$$\varphi_{21} = \max_{x_1}(\varphi_{1i}\gamma_{i1})p_1(o_2) = max \begin{cases} \varphi_{11}\gamma_{11}p_1(H) = 0.40 \times 0.32 = 0.128 \\ \varphi_{12}\gamma_{21}p_2(H) = 0.25 \times 0.24 = 0.06 \end{cases} = 0.128$$

$$\varphi_{22} = \max_{x_1}(\varphi_{1i}\gamma_{i2})p_2(o_2) = max \begin{cases} \varphi_{11}\gamma_{21}p_1(H) = 0.40 \times 0.30 = 0.12 \\ \varphi_{12}\gamma_{22}p_2(H) = 0.25 \times 0.35 = 0.087 \end{cases} = 0.12$$

Similarly,

$$\varphi_{31} = max \begin{cases} 0.128 \times 0.32 = 0.041 \\ 0.12 \times 0.24 = 0.029 \end{cases} = 0.041$$

$$\varphi_{32} = max \begin{cases} 0.128 \times 0.30 = 0.038 \\ 0.12 \times 0.35 = 0.042 \end{cases} = 0.042$$

$$\varphi_{41} = max \begin{cases} 0.041 \times 0.32 = 0.0131 \\ 0.042 \times 0.24 = 0.0101 \end{cases} = 0.0131$$

$$\varphi_{42} = max \begin{cases} 0.041 \times 0.30 = 0.0123 \\ 0.042 \times 0.35 = 0.0147 \end{cases} = 0.0147$$

$$\varphi_{51} = max \begin{cases} 0.0131 \times 0.32 = 0.0042 \\ 0.0147 \times 0.24 = 0.0035 \end{cases} = 0.0042$$

$$\varphi_{52} = max \begin{cases} 0.0131 \times 0.30 = 0.0039 \\ 0.0147 \times 0.35 = 0.0051 \end{cases} = 0.0051$$

The summary of the equations is in the table

Table 3.2: φ's with respect to time t

	H		H		H		H		H
B	0.40	$\rightarrow$	0.128	$\rightarrow$	0.041	$\rightarrow$	0.131	$\rightarrow$	0.0042
F	0.25	$\searrow$	0.12	$\rightarrow$	0.042	$\rightarrow$	0.0147	$\rightarrow$	0.0051

In order to find the best state sequence, the most likely state for time $T = 5$ is determined as state F since φ_{52} is higher than φ_{51}, then the most likely state for time $T - 1 = 4$ is found by searching for the producer of the predetermined state F of T which is state F again. When this process is applied backwardly, the most likely sequence will be (B,F,F,F,F).

3.7 State Prediction

The conditional distributions of state X_t given observations $o^{(T)}$ for time $t \leqslant T$ are mentioned before. Now, the conditional probabilities of state X_t for $t > T$ will be determined in order to perform state prediction. Remember that h is the time horizon, i.e., $h = t - T$. Then,

$$Pr(X_{T+h} = i|O^{(T)} = o^{(T)}) = \frac{\alpha_T \Gamma^h(,i)}{L_T} \tag{3.40}$$

where $\Gamma^h(,i)$ is the ith column of the Γ^h matrix.

3.8 Model Selection

The fit of the model is better if the state number m is higher but it requires high number of parameters. However, parsinomy theory suggest as to work with less parameters. Box and Jenkins explains this subject in their study published in 1970 [6] as follows: "We have seen that the mathematical models we need to employ contain certain constants or parameters whose values must be estimated from the data. It is important, in practice, that we employ the smallest possible number of parameters for adequate representations. The central role played by this principle of parsimony in the use of parameters will become clearer as we proceed.

In order to decide the number of the states m of the model a criterion that consider the advantages and disadvantages of higher m is needed. In this part, firstly the AIC and BIC are introduced in order to choose a more convenient model and then, the pseudo-residuals are described in order to check whether the selected model is adequate.

3.8.1 Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

AIC is a function of the likelihood and the number of parameters of the model. It has a negative relationship with the likelihood of the model and a positive relationship with the number of parameters. Therefore, the model with lower AIC value is better one to choose. The equation of AIC is,

$$AIC = -2logL + 2p$$

The first term decreases with the increase in m. However, the second term, which is a penalty term, increases with the increase of m.

BIC is also a function of the likelihood and the number of parameters of the model. It is similar to AIC but the penalty term of BIC is different from AIC. The formulation of it,

$$BIC = -2logL + plogT$$

where T is the observation number of the model. It can be clearly seen from the equation that when $T > e^2$,which is the case in most applications, the BIC gives more weight to penalty term. Hence, BIC often selects the models with fewer parameters than does AIC.

3.8.2 Model Checking with Pseudo-residuals

The AIC and BIC choose the best model but this does not mean that it is an adequate model. For this reason, the pseudo-residuals are used to check the general goodness of fit of the selected model and to identify outliers relative to the model.

Before presenting uniform pseudo-residuals firstly notice that if X is a random variable with continuous function F, $F(X)$ is uniformly distributed on the unit interval. Under the fitting model, the probability of observing a value below the observation o_t is called as *uniform pseudo-residual* of the observation o_t from a continuous random variable O_t. If the uniform pseudo-residual is denoted as u_t, the equation of it will be as follows,

$$\begin{aligned} u_t &= Pr(O_t \leq o_t) \\ &= F_{O_t}(x_t) \end{aligned}$$

If the model is correct, u_t distributed as uniformly and extreme values are close to 0 or 1. However, this closeness is not a net definition and so uniform pseudo-residuals are not good at identifying outliers. The normal pseudo-residuals are used in order to overcome this problem.

Before introducing normal pseudo-residuals note that $Z \equiv \Phi^{-1}(F(X))$ is distributed standard normal where Φ is the distribution function standard normal and X is a random variable with distribution function F. Now, the *(*normal pseudo-residuals) can be described as follows,

$$\begin{aligned} z_t &= \Phi^{-1}(u_t) \\ &= \Phi^{-1}(F_{O_t}(o_t)) \end{aligned}$$

When the normal pseudo-residuals are adapted to continuous normal HMMs, it will be,

$$z_t = \Phi^{-1}(Pr(O_t \leq o_t | O^{(-t)} = o^{(-t)}))$$

Normal pseudo-residuals are distributed standard normal if the the related model is adequate.

CHAPTER 4

APPLICATION OF HMM TO THE INFLATION DATA OF TURKEY

In this chapter, the methods that are explained in Chapter 3 are applied to inflation data of Turkey between 2004 and 2012. Then, the model is tested on the data of 2013-2014. When the method is applied, the R codes in appendix A that are adapted to normal-HMM are used.

4.1 Description of Data

The targeting inflation policy that started in 2001 changed the structure of inflation in Turkey in one or two year. In order not to deal with extreme changes that preexisting in the inflation history of Turkey, analyzing the new period of inflation is preferred in this thesis. Hence, the Hidden Markov Model is applied to the inflation data of Turkey between 2004 and 2014 that is represented in Figure 4.1.

The data is retrieved from the database of the Central Bank of the Republic of Turkey (CBRT). Inflation rates are monthly and they are percentage changes that are obtained by comparing the price level of the current month to the same month of the previous year. The mentioned price levels are calculated by Laspeyres index and the base year of them is 2003.

The descriptive statistics of the inflation rates between 2004 and 2012 are displayed in Table **??**. Moreover, histogram, QQ-plot and test of normality results of them represented in the Figure 4.3, Figure 4.4 and Table 4.1 respectively. It seems from the QQ-plot that the data is normally distributed. Even so according to the Shapiro-Wilk results in Table 4.1 the data is normally distributed since the sigma value is bigger then 5%

A part of the dataset belonging to 2002-2012 horizon is used as training set, the model is build on this time horizon. The other part, i.e. 2013-2014 interval, is used as the test set.

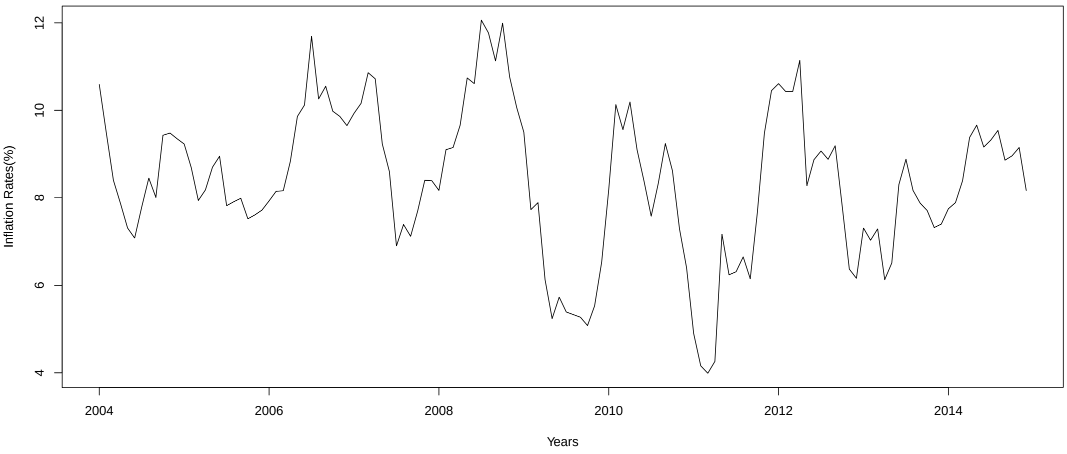


Figure 4.1: Inflation rates of Turkey between 2004-2014

Descriptives

			Statistic	Std. Error
inflation	Mean		8.4272	.17403
	95% Confidence Interval for Mean	Lower Bound	8.0822	
		Upper Bound	8.7722	
	5% Trimmed Mean		8.4625	
	Median		8.4000	
	Variance		3.271	
	Std. Deviation		1.80861	
	Minimum		3.99	
	Maximum		12.06	
	Range		8.07	
	Interquartile Range		2.39	
	Skewness		-.335	.233
	Kurtosis		-.282	.461

Figure 4.2: Histogram of inflation rates between 2004-2014

4.2 Steps in Normal-HMM

The steps that are followed while modeling, analyzing and forecasting the inflation rates are itemized as follows:

- Checking normality of data
- Choosing initial parameters (Γ, μ, σ, δ)
- Applying EM Algorithm to estimate parameters
- Choosing most suitable model
- Estimating the most likely states
- Making prediction about the future states
- Estimating forecasting distributions and forecasting values
- Calculating the conditional distributions of observations
- Checking the model by using normal pseudo-residuals

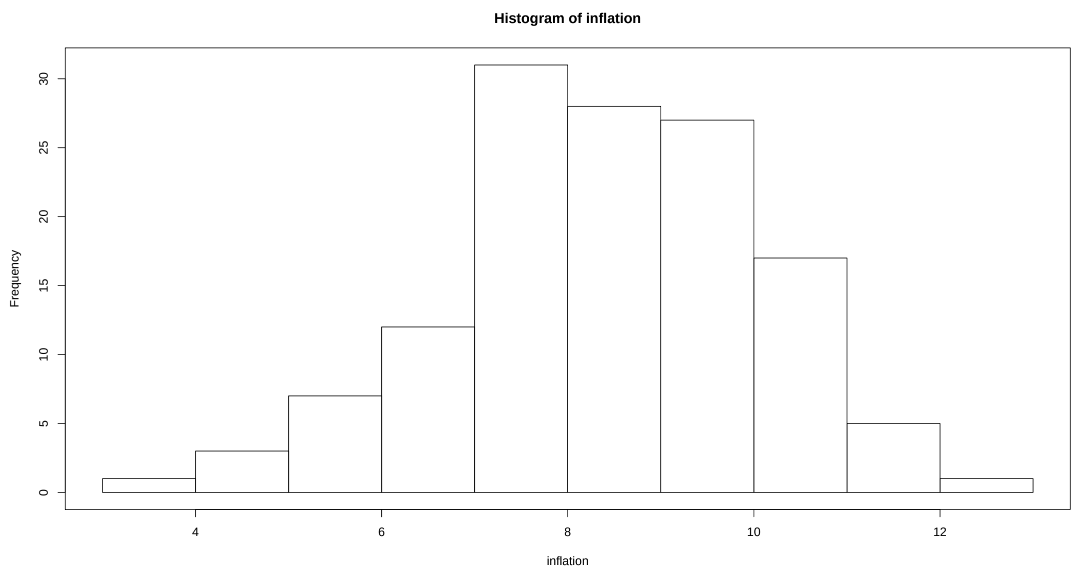


Figure 4.3: Histogram of inflation rates between 2004-2014

Table 4.1: Normality test results of inflation rates for years 2004-2014

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
inflation	0.055	132	0.200	0.988	132	0.299

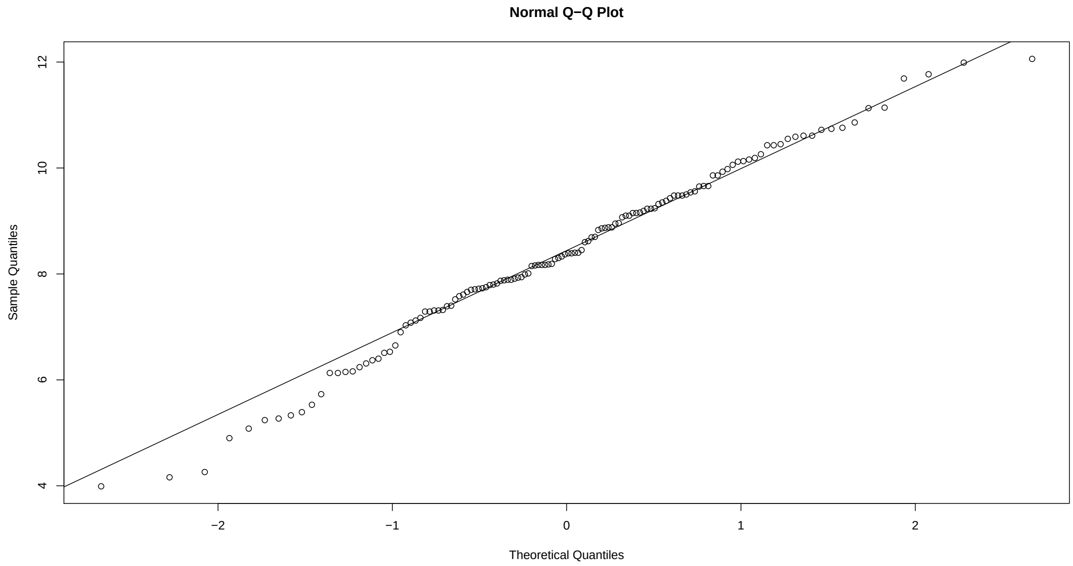


Figure 4.4: QQ-plot of inflation rates between 2004-2014

4.3 Analyses

The normality of data is examined in the beginning of this chapter and it is concluded that our data is normally distributed. Hence, the mean mu and standard deviation σ vectors will be needed besides the parameters transition probability matrix Γ and initial distribution δ of the HMM. EM algorithm is the starting point of the analysis of this thesis. Firstly, EM algorithm is applied to the 2004-2012 inflation data of Turkey. Because of convergency constraints,the choice of initial values are tested for different values and the optimal selection is derived from the data set. The initial means are selected around the mean of original data and the initial trasformation matrix is constructed by giving the biggest probabilities to the transitions from a state to the same state. Moreover, it can be beneficial to remember that the EM algorithm can produce results for local maximums. In order to deal with this problem also several initial values were tried and the ones that gives the minimal negative log-likelihood accepted as the actual initial values. The initial parameters of 3-state HMM and 4-state HMM that are obtained after several trying as follows respectively:

$$\Gamma = \begin{pmatrix} 0.93 & 0.05 & 0.02 \\ 0.01 & 0.94 & 0.05 \\ 0.02 & 0.06 & 0.92 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 5 & 8 & 11 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.6 & 0.8 & 1.1 \end{pmatrix}$$

$$\delta = \begin{pmatrix} 0.95 & 0.025 & 0.025 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 0.92 & 0.05 & 0.02 & 0.01 \\ 0.01 & 0.87 & 0.07 & 0.05 \\ 0.02 & 0.05 & 0.92 & 0.01 \\ 0.01 & 0.05 & 0.07 & 0.87 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 4 & 5 & 10 & 11 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.10 & 0.64 & 0.68 & 0.70 \end{pmatrix}$$

$$\delta = \begin{pmatrix} 0.95 & 0.025 & 0.020 & 0.005 \end{pmatrix}$$

The initial parameters of 3-state HMM and 4-state HMM are represented only because it will be seen that they are the most suitable HMMs for our data.

These parameters are used to start the EM-algorithm with expectation step. In this step the forward and backward probabilities are computed by the initial parameters. Then, the conditional expectations of states are computed by using the forward and backward probabilities. The complete data log-likelihood is estimated at the last of the first estimation step. After that the complete data is maximized in maximization step. The maximizing values are used again in the estimation step and the same processes are repeated. This loop continues until to convergence. Now, it is essential to indicate that the tolerance for the convergence is accepted as 10^{-6} in this study. Consequently, mean and standard deviation vectors of states,transition probability matrix, initial distribution, negative log-likelihood, AIC and BIC values of the model are obtained. Remember that the formulas of AIC and BIC as follows:

$$AIC = -2logL + 2p$$

$$BIC = -2logL + plogT$$

where the p is the number of parameters. Since the Normal-HMMs are used in this study, the number of parameters of Normal-HMMs computed by using the following equation:

$$p = m^2 + 2m - 1.$$

These parameters' $m - 1$ is coming from delta, m is from mean, m is from sigma,

Table 4.2: Comparison of several Normal-HMMs by means of AIC and BIC.

no.of states	p	(-)log L	AIC	BIC
2	7	187.21	388.42	407.2
3	14	158.46	**344.93**	**382.48**
4	23	152.06	350.13	411.82
5	34	142.11	352.23	443.42
6	47	132.46	358.91	484.97

$m^2 - m$ is from the transition probability matrix.

Table 4.2 displays the number of parameters, negative log-likelihood, AIC and BIC of normal-HMMs with several different state numbers. Since HMMs with more states fit

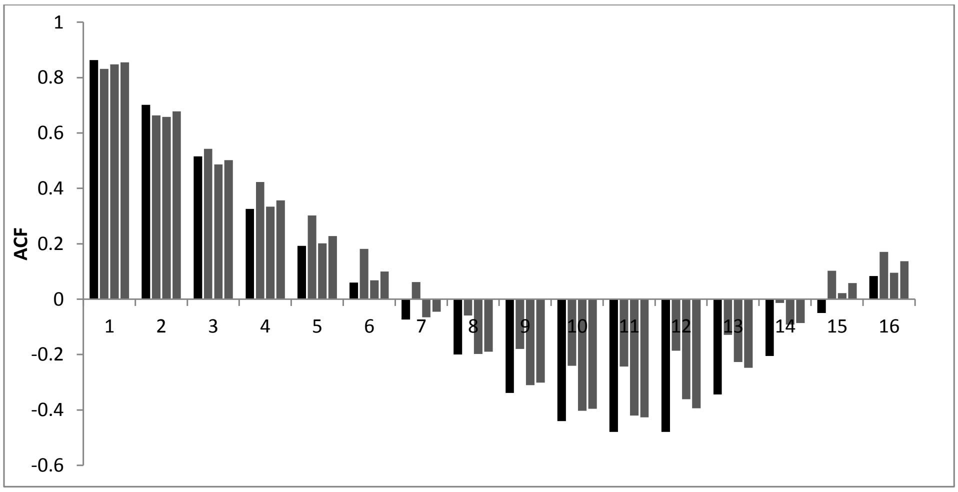


Figure 4.5: Autocorrelation functions in two, three, four-state HMMs.

better the data, negative log-likelihood is getting smaller while the number of states getting higher. However, both AIC and BIC has the smallest value for the three state HMM. Hence, three state HMM is the most appropriate one for modelling Turkey inflation data. 5 and 6 state HMMs are not preferable unsuprisingly since 34 and 47 number of parameters are excessive for only 108 observations.

The auto-correlation functions of different HMMs also can be used in order to choose the most appropriate model. In Figure 4.5 ACFs of the inflation data, two-state, three-state and four-state model are represented, respectively. By the help of this figure and Table 4.3, it can be clearly seen that three and four state HMMs correspond well to the inflation data.

Table 4.3: Autocorrelations of inflation rates and two, three, four-state HMMs.

k	1	2	3	4	5	6	7	8
observations	0.86	0.70	0.52	0.33	0.19	0.06	-0.07	-0.20
2-state HMM	0.83	0.66	0.54	0.42	0.30	0.18	0.06	-0.06
3-state HMM	0.85	0.66	0.49	0.33	0.20	0.07	-0.06	-0.20
4-state HMM	0.86	0.68	0.50	0.36	0.23	0.10	-0.04	-0.19

After that point, 3-state and 4-state models are used. Because of that only their estimated mean vectors, standard deviation vectors and the transition probability matrices that are the results of EM-algorithm are displayed as follows:

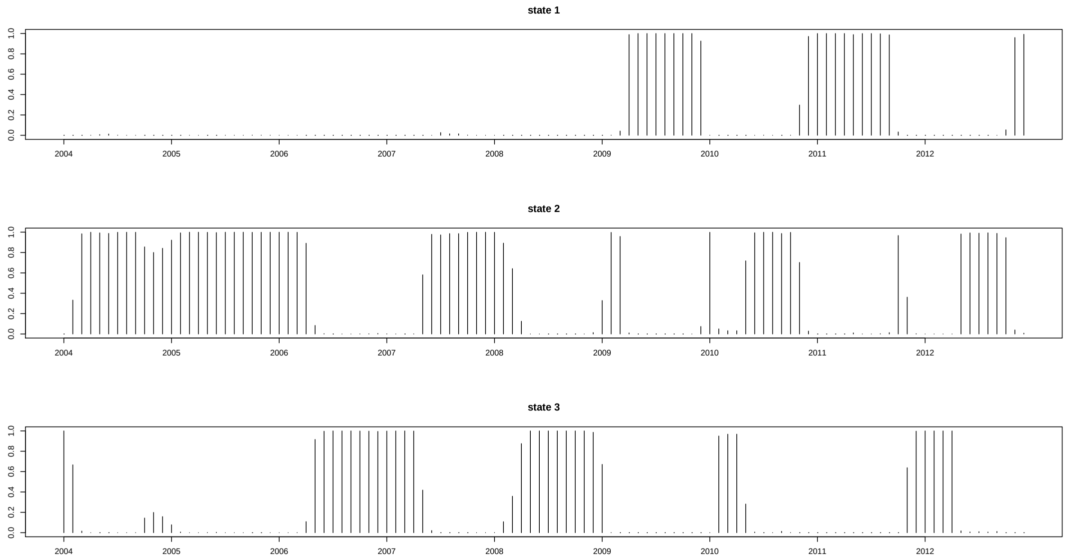


Figure 4.6: State probabilities for three-state HMM.

- 3-state model:

$$\Gamma = \begin{pmatrix} 0.90 & 0.10 & 0 \\ 0.06 & 0.86 & 0.08 \\ 0 & 0.16 & 0.84 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 5.70 & 8.27 & 10.40 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.88 & 0.69 & 0.74 \end{pmatrix}$$

$$\delta = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$$

- 4-state model:

$$\Gamma = \begin{pmatrix} 0.66 & 0.34 & 0 & 0 \\ 0.06 & 0.82 & 0.12 & 0 \\ 0 & 0.06 & 0.86 & 0.08 \\ 0 & 0 & 0.16 & 0.84 \end{pmatrix}$$

$$\mu = \begin{pmatrix} 4.14 & 5.94 & 8.27 & 10.39 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.11 & 0.64 & 0.68 & 0.74 \end{pmatrix}$$

$$\delta = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$$

Now, the most suitable states will be determined but before that the state probabilities will be estimated since they are essential for local decoding method. Because in local decoding method the most likely states will be determined by choosing the states with biggest probabilities for each time point separately.

Hence, the state probabilities are determined by using the equation (3.31) and the results for 3-state HMM are displayed in Figure 4.6.

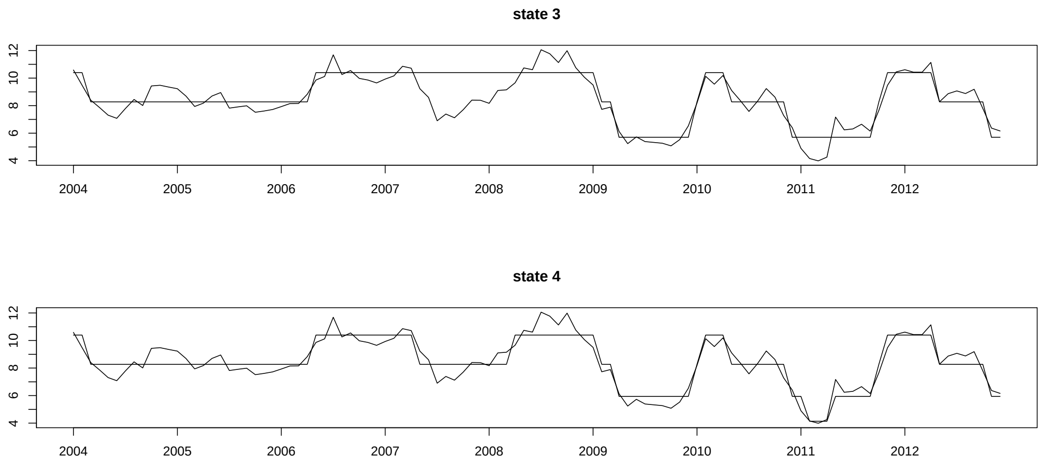


Figure 4.7: Local decoding for three and four state HMM.

For three and four state HMM the results of local decoding are displayed in Figure 4.7. The horizontal lines that indicate state dependent means are drawn on related inflation data of Turkey. In order to illustrate the relation between state probabilities and the local decoding, observe that in the case of 3-state HMM the states with highest probabilities are 3,3,2,2 and 2 for first five time points and the local decoding selects these states as the most likely states for the related time points. The other remarkable point that as seen from the figure that 4-state fits the data better. However, as mentioned before more parameters can cause a less adequate model, three state model is selected in analyses.

Similarly, Figure 4.8 displays global decoding results of three and four state HMMs. This most likely state sequence is achieved by the Viterbi Algorithm. Although it is mentioned in Chapter 3.6.2 that results of local and global decoding do not have to be similar, in this study they are same accidentally.

In order to see the success of predicted states in the case of 3-state HMM, the inflation rate values are simulated according to these states by using their mean and standard deviation values. The simulation is done by the Excel with the following formula,

IF(”state”=1,NORMINV(RAND(),”mean(1)”,”sd(1)”),

IF(”state”=2,NORMINV(RAND(),”mean(2)”,”sd(2)”),

NORMINV(RAND(),”mean(3)”,”sd(3)”)))

where $mean(i)$ and $sd(i)$ means the mean and standard deviation of ith state,respectively. The terms that are in the quotes should be replaced by the cells of these terms. They are stated in quotation marks in order to give a general definition.

The formulation produces a random number for the probability and for the given values of mean and standard deviation it gives the inverse of cumulative normal distribution

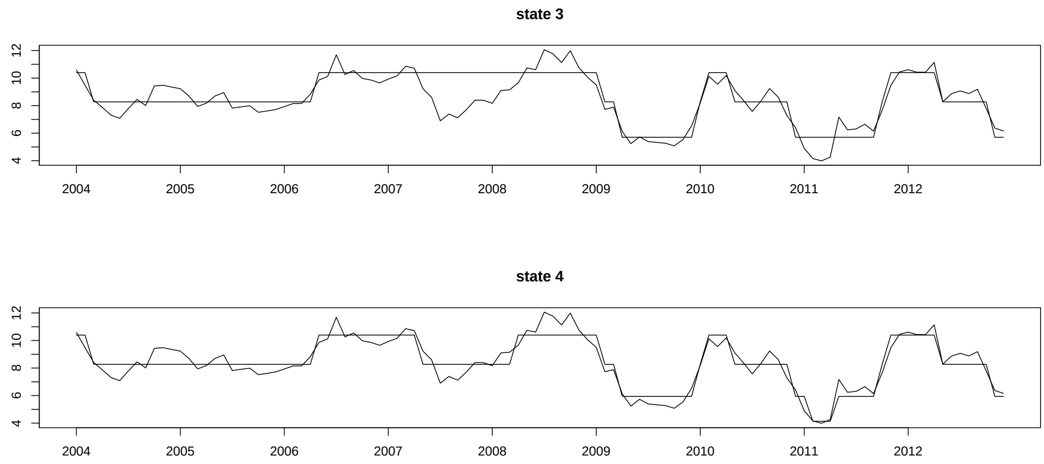


Figure 4.8: Global decoding for three and four state HMM.

function. The results of application of it to the inflation data in Figure 4.9.

It can be seen from the Figure 4.9 that the simulation values are very close to the original inflation rates. Mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE) of the simulation is displayed in Table 4.4 in order to observe the accuracy.

Table 4.4: Accuracy of simulation

ME	RMSE	MAE	MPE	MAPE
-0.360	1.038	0.829	-6.495	12.172

Table 4.5 gives the state predictions that are the probabilities of being in a given state in specified year for three state HMM. These prediction probabilities are achieved by using equation (3.31).

It can be deduced from the Table 4.5 that there is a prediction of a positive trend in the future inflation rates. Because the probabilities of 2 and 3 states, which have higher means than state 1, increase when the probability of state 1 decrease.

Table 4.5: State prediction with using a three-state normal HMM.

state	Jan-13	Feb-13	May-13	Oct-13	Apr-14	Dec-14
1	0.89	0.81	0.63	0.46	0.37	0.31
2	0.11	0.18	0.32	0.41	0.45	0.47
3	0	0.01	0.05	0.13	0.19	0.22

The forecast distributions of six time points including the limiting distribution are represented in Figure 4.10. Firstly, distributions are computed for continuous data. However, in order to work with R the interval between successive numbers are divided into 10 parts and distribution of each integer is recomputed by averaging the distributions of

Figure 4.9: Simulation results for 3-state HMM

following parts till the next integer. Observe that, the forecast distribution approaches to its limiting distribution.

Then, means which are displayed in Table 4.6 are calculated for these six time points.

Table 4.6: Forecasts of inflation rates with respect to three-state Normal-HMM.

date	Jan-13	Feb-13	May-13	Oct-13	Apr-14	Dec-14
horizon	1	2	5	10	16	24
mean	6.43	6.66	7.22	7.81	8.17	8.38

The means are accepted as the forecasting values in this study. These forecasting results for all months in 2013-2014 interval are displayed in Figure 4.11. The fluctuating line shows the original inflation rates of Turkey between 2004 and 2014, the dashed line represents the forecasting results and the bold area is the 95 % confidence interval. It can be clearly seen from the figure that the forecast rates are in the confidence interval.

The accuracy of the forecasting is displayed in the Table 4.7:

Table 4.7: Accuracy of forecasting

ME	RMSE	MAE	MPE	MAPE
-0.385	0.784	0.679	-4.866	8.761

In Figure 4.12 the conditional distributions of observations for some randomly selected years are presented. The points on the distribution curves are the conditional probability of real observation of relative year. Observe that each distribution curve has different shape. Moreover, it is obvious that probabilities of some observations are ex-

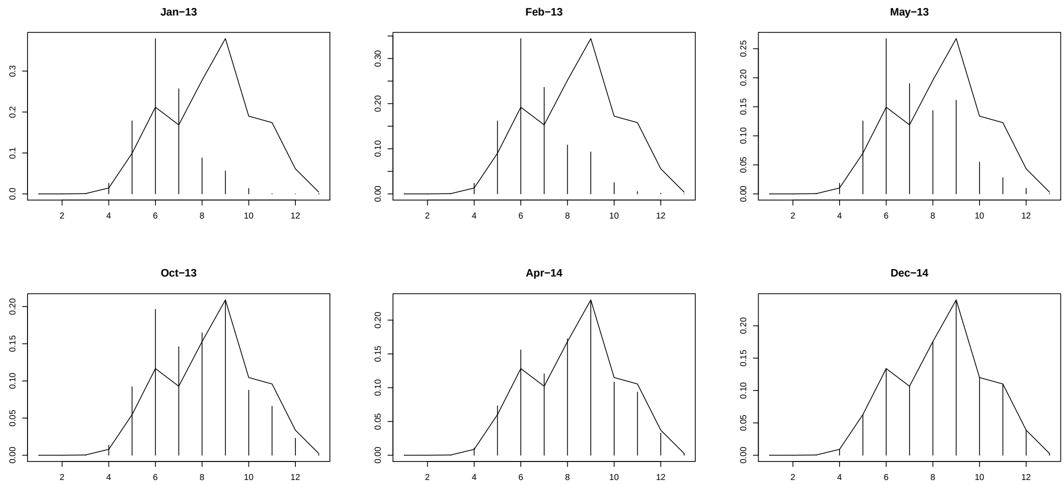


Figure 4.10: Inflation forecasts with respect to three-state normal HMM.

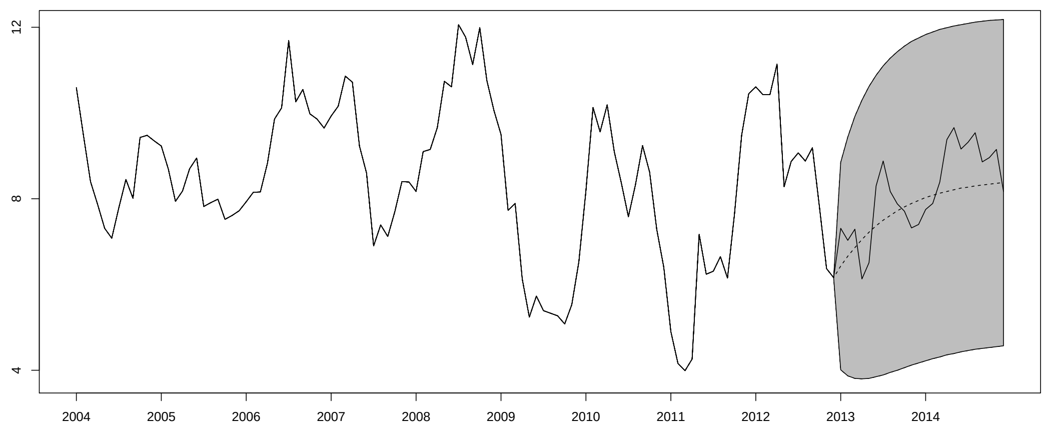


Figure 4.11: Forecasting

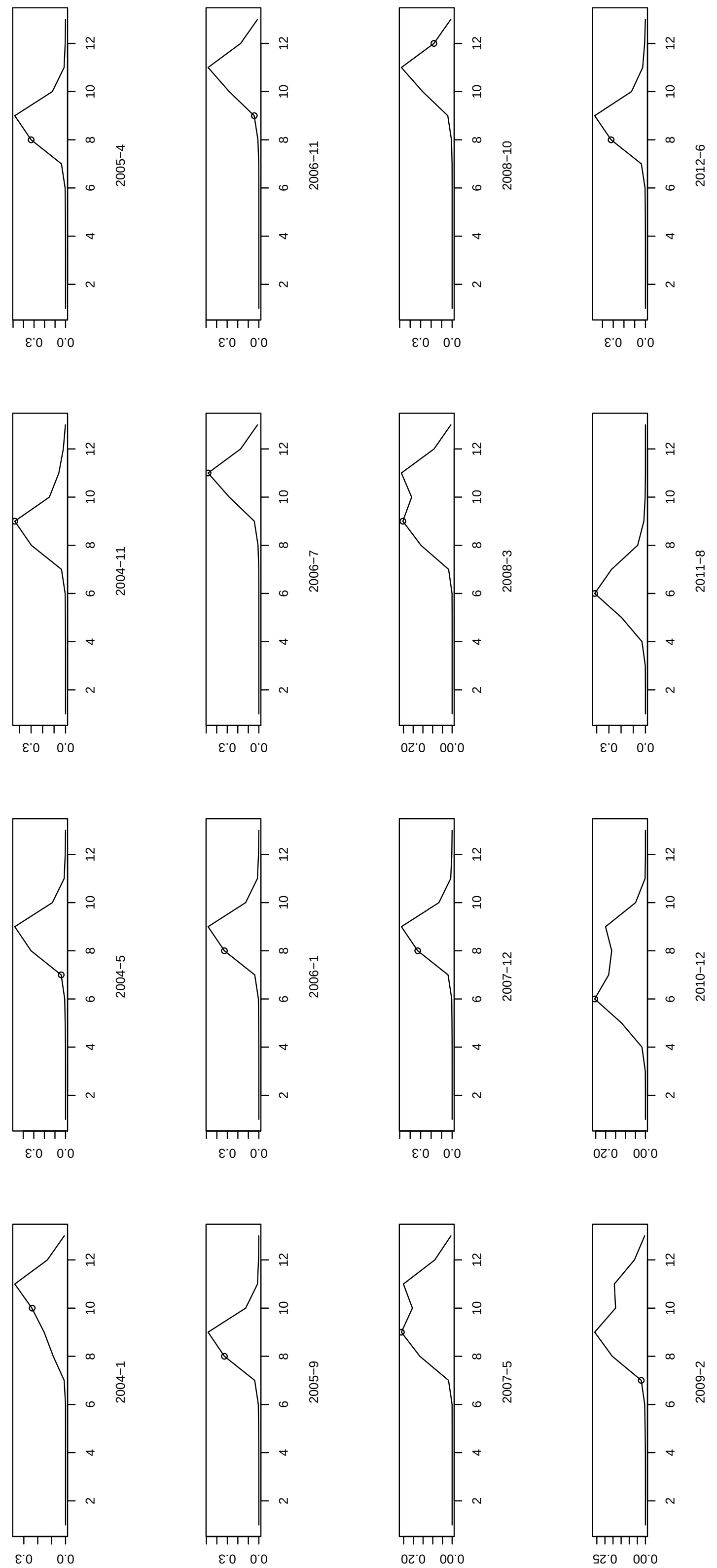


Figure 4.12: Conditional distribution of 3-state HMM for randomly selected years.

treme relative to their distributions. This observation gives the opinion that conditional distributions can be used for outlier checking.

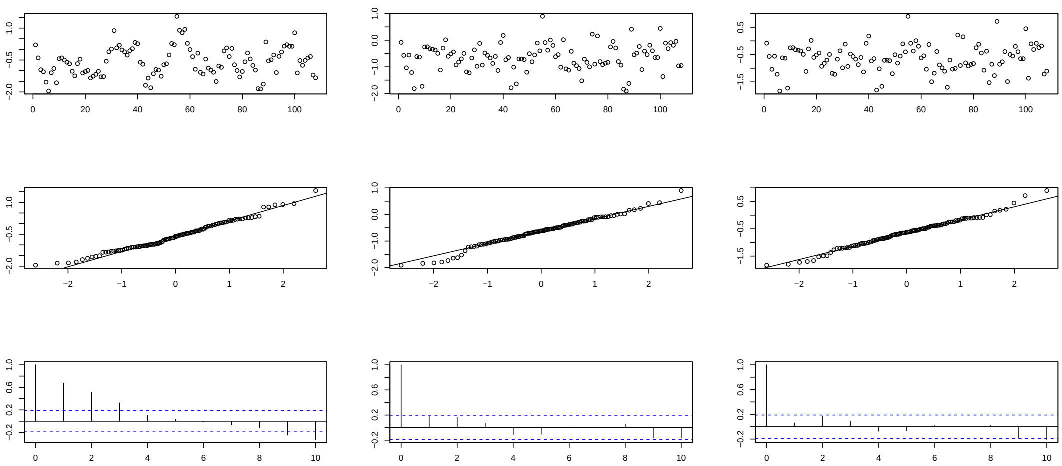


Figure 4.13: Normal pseudo-residuals plot

By looking at conditional distribution, the pseudo-residual is used for checking the suitability of the HMM for our data and the results are displayed in Figure 4.13. The columns 1,2 and 3 are relate to 1,2,3 state HMMs, respectively. The top row shows the values of normal pseudo-residuals. Moreover, the second row displays the QQ-plot of pseudo-residuals. When the QQ-plots are taken as a reference, it can be said that the normal pseudo residuals are distributed normally for all of 1,2,3 state HMMs. However, Shapiro-Wilk test is applied to the normal pseudo-residuals and it also give the same result since the p-values are bigger than 0.05. Lastly, the last row shows the auto-correlations of normal pseudo-residuals. It can be clearly seen that 2-state is an adequate model but auto-correlated. Hence, it has evidential value for the explanation of Zucchini and MacDonald (2009) [45] that if a model is true, the pseudo-residuals do not have to be auto-correlated necessarily.

CHAPTER 5

CONCLUSION AND OUTLOOK

Inflation has been a significant problem since 1970s in Turkey. Although it dropped to one-digit levels after the price level targeting regime of Central Bank of Republic of Turkey, it still is not stable. Hence, it still worth to model, analyze and forecast.

This study is the first application of HMM to Turkish inflation rates. However, there are a few studies that model it with another methods. These studies use vector autoregressive models (VAR), Bayesian autoregressive models (BVAR), autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) models. Comparison of existing studies on Turkish data is out of scope of this thesis as the index in most of the studies are selected as CPI. This part is taken as the first step for a future work. Moreover, best of our knowledge there are a few studies on modeling inflation by using HMM in the world. These studies categorize the inflation rates and then use the Poisson-HMMs.

In this thesis, the advantage of working of HMM on autocorrelated data is used. Moreover, the Normal-HMM is preferred since our data is normal and continuous. The inflation rates are not categorized in contrast with the studies using HMM to model inflation rates but the parameters of Normal-HMM is considered. This model is applied to the inflation rates of Turkey between 2004-2014.

Before the applications, the HMM is introduced very briefly in the introduction part. Moreover, the inflation history of Turkey between 1965-2014 is displayed. The reasons of the fluctuations in the history are explained. It is concluded that there are less fluctuations after 2004 because of the effects of the inflation targeting regime. However, it is also indicated that although the fluctuations are lessen in the last period, the rates still are not stable. Hence, the aim of the thesis represented as the modeling, analyzing and forecasting the inflation rates between 2004-2014 by using HMM.

In the first chapter, the economical structure behind the inflation theory is represented firstly. The computation method of price levels which is necessary to compute the inflation rates, the relation of inflation with the aggregate demand and aggregate supply and the Philips curve in short and long term are explained.

In the second chapter, HMM is explained in detail before applying it to inflation rates. The Markov chain is clarified firstly since it is the based on structure of HMM. It is also explained that HMM consists of hidden states and observations depending on these states. Then, the parameters of the HMM is introduced as initial probabilities δ, transition probability matrix Γ and the parameters of related distribution (mean μ and standard deviation σ in our case). The most remarkable method EM-algorithm that estimates these parameters of the HMM is introduced and the estimation of forward and backward probabilities that are used in EM-algorithm is explained. After that the methods of selecting the most likely states are represented as local decoding and global decoding. Since the local decoding selects most likely states as the most probable states of each time, the estimation of state probabilities is explained also. Morever, Viterbi algorithm is introduced in order to benefit from it in global decoding method. Then, the estimation of forecasting distributions are explained. Lastly, AIC,BIC, conditional distributions of observations and pseudo-residuals are introduced in order to use them to choose the most suitable model.

In the application part of the thesis, the 2004-2014 inflation data of Turkey is used to analyze, model and forecast with HMM. R language is preferred in this study while applying the HMM. The following results are for the 2004-2012 part of the data which is selected as the training part. The parameters of mean μ, standard deviation σ , transition probability matrix Γ and the initial distribution δ are estimated by the EM-algorithm. The AIC and BIC values are also computed during the estimation of parameters. According to the both of the criteria 3-state and 4-state HMMs are more appropriate than others, in fact the 3-state HMM is the best one. Therefore, The hidden states of 3-state HMM are estimated by Viterbi algorithm and the states of 2013-2014 interval are predicted. Then, the forecast distributions of the every month in 2013-2014 interval are computed. The means of them are accepted as the forecasting values for this study and they are in the %95 confidence interval successfully. Lastly, in order to be sure that the selected model is adequate it is checked whether the pseudo-residuals of the model are distributed normally and the result of this checking brings us to a successful conclusion.

REFERENCES

[1] R. Allen, *Index Numbers in Economic Theory and Practice*, Aldine Pub. Co., 1975.

[2] I. O. Badurlar, *Türkiye'de Enflasyon Hedeflemesi Strateji Uygulaması ve Para Politikasına Etkisi:2002-2008*, Anadolu Uluslararası İktisat Kongresi, 2009.

[3] J. M. Barron and G. J. Lynch, The aggregate demand curve: A defense, The Journal of Economic Education, 18(1), pp. 41–46, 1987.

[4] R. Bhar and S. Hamori, *Hidden Markov Models Applications to Financial Economics*, Kluwer Academic Publishers, 2004.

[5] J. A. Bilmes, A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models, International Computer Science Institute, 1998.

[6] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control, Fourth Edition*, John Wiley and Sons, 1970.

[7] M. Bronfenbrenner and F. D. Holzman, Survey of inflation theory, The American Economic Review, 53(4), pp. 593–661, 1963.

[8] K. E. Case and R. C. Fair, *Principles of Economics*, Pearson Education, 2006.

[9] L. W.-K. Cheung, Use of run statistics for pattern recognation in genomic dna sequences, Journal of Computational Biology, 11(1), pp. 107–124, 2004.

[10] A. Cukierman and A. Meltzer, A theory of ambiguity, credibility, and inflation under discretion and asymmetric information, Econometrica, 54(4), pp. 1099–1128, 1986.

[11] İ. Domaç, Explaining and forecasting inflation in turkey, World Bank Policy Research Working Paper, (3287), 2004.

[12] S. R. Eddy, Hidden markov models, Current Opinion in Structural Biology, 6(3), pp. 361–365, 1996.

[13] E. Ermişoğlu, Enflasyon hedeflemesi rejiminin türkiye'deki başarısının değerlendirilmesi, Türkiye Cumhuriyet Merkez Bankası, 2011.

[14] M. Friedman, The counter-revolution in monetary theory, IEA Occasional Paperl, (33), 1970.

[15] M. Friedman, Nobel lecture: Inflation and unemployment, Journal of Political Economy, 50(2), 1977.

[16] H. Frisch, *Theories of Inflation*, Cambridge University Press, 1983.

[17] C.-D. Fuh, Sprt and cusum in hidden markov models, The Annals of Statistics, 31(3), pp. 942–977, 2003.

[18] J. G. David Forney, *Convolutional Codes II. Maximum-Likehood Decoding*, Elsevier, 1974.

[19] J. J. J. Groen, R. Paap, and F. Ravazzolo, Real-time inflation forecasting in a changing world, Federal Reserve Bank of New York Staff Reports, (388), 2009.

[20] J. D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica: Journal of the Econometric Society, pp. 357–384, 1989.

[21] M. R. Hassan and B. Nath, Stock market forecasting using hidden markov model: a new approach, in *Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference on*, pp. 192–196, IEEE, 2005.

[22] B. Hossain, M. Ahmed, and M. F. Rabbi, A novel approach for inflation analysis using hidden markov model, International Journal of Computer Science Issues,, 9(2), pp. 619–623, 2012.

[23] A. Inge, Hidden markov models: Theory and simulation, Mathematical Statistics Stockholm University, 2013.

[24] M. Jochmann, Modeling us inflation dynamics: A bayesian nonparametric approach, 2010.

[25] A. Kibritçioğlu, L. Rittenberg, and F. Selçuk, *Inflation and Disinflation in Turkey*, Ashgate Publisher Limited, 2001.

[26] A. Krogh, M. Brown, I. S. Mian, K. Sjölander, and D. Haussler, Hidden markov models in computational biology: Applications to protein modeling, Journal of molecular biology, 235(5), pp. 1501–1531, 1994.

[27] D. Laidler and J. Parkin, Inflation - a survey, Economic Journal, pp. 741–809, 1975.

[28] R. Langrock, Some applications of nonlinear and non-gaussian state space modelling by means of hidden markov model, Journal of Applied Statistics, 38(12), pp. 2955–2970, 2011.

[29] R. M. Feldman and C. Valdez-Flores, *Applied Probability and Stochastic Processes*, Springer, 2010.

[30] A. A. Markov, Extension of the law of large numbers to dependent quantities, Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser), 15, pp. 135–156, 1906.

[31] O. Meçik and M. Karabacak, Arima modelleri ile enflasyon tahminlemesi: Türkiye uygulaması, SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi, 22, pp. 177–198, 2011.

[32] E. Muljadi, *Economics and Banking*, Paul Muljadi, 2011.

[33] F. Öğünç, K. Akdoğan, S. Başer, M. G. Chadwick, D. Ertuğ, T. Hülagü, S. Kösem, M. U. Özmen, and N. Tekatlı, Short-term inflation forecasting models for turkey and a forecast combination analysis, Economic Modelling, 33, pp. 312–325, 2013.

[34] J. Omura, *On the Viterbi Decoding Algorithm*, IEEE Transactions on Information Theory, 1969.

[35] A. Phillips, The relation between unemployment and the rate of change of money wage rates in the united kingdom,1861-1957, Economica, 25(100), pp. 283–299, 1958.

[36] D. L. Poole and A. K. Mackworth, *Artificial Intelligence: Foundations of Computational Agents*, Cambridge University Press, 2010.

[37] L. Rabiner and B.-H. Juang, Fundamentals of speech recognition, 1993.

[38] L. R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, Proc. IEEE, 77(2), pp. 257–286, 1989.

[39] M. S. Ryan and G. Nudd, The viterbi algorithm, Department of Computer Science, University of Warwick, 1993.

[40] P. A. Samuelson and R. M. Solow, Analytical aspects of anti-inflation policy, The American Economic Review, 50(2), pp. 177–194, 1960.

[41] G. Saz, The efficacy of sarima models for forecasting inflation rates in developing countries: The case for turkey, International Research Journal of Finance and Economics, 62, pp. 111–142, 2011.

[42] Y. Song, *Development and Application of Hidden Markov Models in the Bayesian Framework*, Ph.D. thesis, University of Toronto, 2011.

[43] R. Turvey, Some aspects of the theory of inflation in a closed economy, The Economic Journal, 61(243), pp. 531–543, 1951.

[44] A. Viterbi, *Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm*, IEEE Transactions on Information Theory, 1967.

[45] W. Zucchini and I. L. MacDonald, *Hidden Markov Models for Time Series An Introduction Using R*, CRC Press, 2009.

APPENDIX A

ALGORITHM OF R CODES

The codes used in this thesis are created by adapting the codes of Zucchini and MacDonald (2009) [45] to the Normal-HMMs.

Firstly, an algorithm that computes the logarithms of forward and backward probabilities is created since it is required for the EM-algorithm. In this algorithm the observation data, mean vector, standard deviation vector, transition probability matrix and initial probabilities are the inputs. The only parameters that are needed to calculate the logarithm of forward and backward probabilities are $P(O_i)$'s and they are computed in R by using "$dnorm$" function.

Secondly, the function of EM-algorithm is constructed. This function estimates the mean (μ), standard deviation (σ), initial distribution (δ) vectors and the transition probability matrix (Γ) and gives the negative log-likelihood, AIC and BIC values of the model.The algorithm starts with the initial values of μ, σ and δ.Then, it computes logarithms of forward and backward probabilities in each iteration by using mentioned forward and backward probability algorithm. The next μ, σ, δ and Γ are computed by taken the items in 3.3.2 as references. This iteration continues until the sum of the differences between the previous parameters and the current ones is smaller than the tolerance that is accepted as 10^{-6} in this thesis. However, in this algorithm it is concluded that there is no convergence if the criteria does not hold until the $1000th$ iteration. If it satisfies the criteria it produces the estimations of μ, σ, δ and Γ. Moreover, the number of parameters for the model is computed by using the formula $m^2 + 2m - 1$. These parameters' $m - 1$ is from delta, m is from mean, m is from sigma, $m^2 - m$ is from the transition probability matrix. Lastly, the AIC, BIC and negative log-likelihood values are computed. Remember that the estimated parameters can be the parameters of local maximums. In order to deal with that problem this algorithm is applied to several initial values and the ones that reach to the least negative log-likelihood at the end of the algorithm are accepted as the initial values. This algorithm is applied to several HMMs with different number of states. By considering the AIC and BIC values of these HMMs, the most appropriate state number can be chosen for the data.

Thirdly, the Viterbi algorithm is represented in R by calculating the φ_{ti}s, determining the state that maximizes φ_{Ti} as in equation (3.38) and continuing backwardly to find the states that maximize $\varphi_{t_i}\gamma_{i,i_{t+1}}$ as in equation (3.39).

Although Viterbi algorithm is preferred to find the most likely states, we also use local decoding, alternatively. Hence, the algorithm of it also constructed but firstly the function of conditional state probabilities created since it will be used in the algorithm of local decoding. Remember that the formula of conditional state probabilities is $Pr(X_t = i|O^{(T)} = o^{(t)}) = \frac{\alpha_t(i)\beta_t(i)}{L_T}$. Since this formula includes just forward, backward probabilities and the likelihood, representing it with R-codes is easy by using forward-backward algorithm.Then, the local decoding is just the function that finds the states that maximize the conditional state probabilities.

Then, state predictions are turned into R codes that compute the α_T/L_T firstly and then multiply it with the ith coloumn of Γ^h matrix as in equation (3.31).

In order to compute the forecast distributions a range is determined by using minimum and maximum of mean and paying attention to taking into account all of the observations with significant distributions. Moreover, the interval between successive numbers in this range divided into 10 parts. The ratio of $\frac{\alpha_T}{\alpha_T 1'}$ is computed. Then, by using a loop this ratio is multiplied with Γ^h. Consequently, the forecast distributions of the predetermined x values are calculated. However, in order to find the expected value of the forecasting, distribution of each integer is recomputed by averaging the distributions of following percentiles till the next integer. Hence, the distribution of integers are produced lastly.

While computing the distributions of observations at time t given the other observations, the intervals between successive integers are distributed into percentiles as in this algorithm of forecast distributions. Then, by using equation (3.29) the conditional distributions are computed for the percentiles. Lastly, by averaging of them the final distributions are produced for integers.

Finally, the algorithm of pseudo-residuals is constructed by using conditional distributions of observations that are formulated previously. In the algorithm the cumulative distributions for integers are computed firstly. Then, uniform pseudo-residuals are computed by taking into account the residues of the observations from the integers. Lastly, these uniform pseudo-residuals are transformed into normal pseudo-residuals.