Markov Regime Switching Models

Application on Annual GDP Growth of the Netherlands

Chanan Noordhuis
6909655

Universiteit Utrecht

A thesis presented for the Bachelor's degree in Mathematics

Supervised by Dr. K. (Karma) Dajani

Utrecht University
2022-2023

Contents

1 Introduction

In this thesis, we are going to explore the concept of Markov regime switching models. In general, models are interesting since it allows us to describe a certain process. Therefore, it gives us a better understanding how certain variables behave. It even allows us to make some predictions on the future based on the given model. Markov regime-switching models have a very unique feature, which is that these type of models allow for different values of parameters across different states. We can consider the business cycles: an economy can be either a rising economy or it is an economy in recession. It would be reasonable to expect that in these different states the values of the parameters underlying the model are different, since consumer behaviour and economic policies are different across these states. We will now briefly introduce the Markov regime-switching model. This model gives us a specification for a time-series process Y_t with some unobservable underlying process S_t. In our example, the underlying process S_t reflects the business cycles. Thus S_t gives whether the economy is rising or whether the economy is in a recession. The variable Y_t is specified as follows

$$Y_t = \begin{cases} \alpha_0 + \beta_0 Y_{t-1} + \sigma_0 \epsilon & \text{if } S_t = 0 \\ \alpha_1 + \beta_1 Y_{t-1} + \sigma_1 \epsilon & \text{if } S_t = 1 \end{cases},$$

where $\epsilon \sim N(0, 1)$. The goal of this thesis is to understand all the basics of these type of models and we want to estimate the parameters in the model. We will also apply our model on the annual GDP growth rates of the Netherlands during the period 2001-2021. This thesis is mainly based on the papers [Ham16], [Kol19] and [SW21]. Note that Kole proposes a model, where $\beta_0 = \beta_1 = 0$. This will become more apparent in the next chapter. However, we will still follow his method, since he thoroughly discusses the estimation procedure. Hamilton has proposed a model, where both β and σ did not depend on the regimes. We let these parameters vary across states such that we do not limit ourselves to certain applications. In [SW21] the same model as our model is introduced. However, they do not go into much detail of estimating the parameters. We try to combine the papers [Kol19], [Ham16] and [SW21], and give more details on some derived expressions.

In Chapter 2, we will discuss the key concepts used in Markov regime switching models. There, we will introduce Markov chains, first-order autoregressions and maximum likelihood estimations. In Chapter 3, we introduce Markov regime switching models, in which Markov chains and first-order autoregressions are combined. Thereafter, we will derive useful expressions in Chapter 4. Chapter 5 uses these expressions to estimate the parameters in the models. We will make use of the Expectation-Maximization algorithm. Chapter 6 will be a brief overview of the derived results, and shows how to implement this algorithm. Lastly, we will end this thesis by an application on the Netherlands. Here, we are going to discuss whether our found results are in accordance with our expectations. Our model will be able to estimate the probability of two different states. We will use a rising economy and a shrinking economy as the two different states. Our model will be able to find a probability whether a country is in a rising economy or in a recession. In the 21st century, the Netherlands had to deal with several economic crises, namely the early 2000s recession (2000-2001), the global financial crisis of 2007-2008, and also the recession as a consequence of the COVID-19 pandemic in 2020-2021. We will show that our model roughly recognizes these periods as a "bad" state economy. Furthermore, we will discuss the found parameters.

2 Introduction to the Key Concepts of Markov Switching Models

In this chapter, we are going to introduce some of the key concepts that are important for regime-switching models. These concepts will be used in the other chapters when considering our model specification. The three key concepts that we are going to discuss are Markov chains, first-order autoregressions, and maximum likelihood estimations.

2.1 Markov Chains

In our model, we want to be able to switch between different regimes (also called states). We refer to the next chapter, where we will specify our Markov regime-switching model. For now, we introduce the concept of Markov chains, which will be the concept allowing regime switching. In this section, we will use the definitions provided in [Nor09].

We assume that our model has a finite number of regimes. Therefore, let I be a finite set, such that $i \in I$ denotes a specific regime. We could also use similar definitions with $|I| = \infty$. However, we limit ourselves to the finite case. Using infinitely many different regimes makes our model proposed in the next chapter more complicated and difficult to interpret. Our model will use discrete-time data, thus we consider discrete-time Markov chains. Before we give the formal definition of discrete-time Markov chains, we need two definitions.

Definition 2.1. *A vector $\lambda = (\lambda_i)_{i \in I}$, where I is the state-space, is called a probability vector if $0 \leq \lambda_i \leq 1$ and $\sum_{i \in I} \lambda_i = 1$.*

Definition 2.2. *A matrix $P = (p_{ij})_{i,j \in I}$, where I is the state-space, is called a stochastic matrix if for all $i \in I$ $\sum_{j \in I} p_{ij} = 1$ (i.e. every row is a distribution), and $p_{ij} \geq 0$ for all $i, j \in I$.*

With these definitions, we are able to define a discrete-time Markov chain.

Definition 2.3. *A discrete time stochastic process $\{S_t : t \geq 1\}$ with values in the set I is a Markov chain with initial distribution $\lambda = (\lambda_i)_{i \in I}$ (probability vector) and transition matrix $P = (p_{ij})_{i,j \in I}$ (stochastic matrix) if*

1. *S_1 has distribution λ, i.e. $\mathbb{P}(S_1 = i) = \lambda_i$ for all $i \in I$.*
2. *$\mathbb{P}(S_t = j | S_1 = k_1, ..., S_{t-2} = k_{t-2}, S_{t-1} = i) = \mathbb{P}(S_t = j | S_{t-1} = i) = p_{ij}$ (Markov property)*

If these conditions are satisfied, then $(S_t)_{t \geq 0}$ is Markov(λ, P).

We call p_{ij} with $i, j \in I$ the transition probabilities. The Markov property states that the S_t only depends on the previous observation S_{t-1}. Thus, p_{ij} gives the probability that the stochastic process S_t changes from state i to state j. In the following chapters, we will see that the transition probabilities play an important role in Markov-switching models. The initial distribution is also relevant for when we have to consider S_1.

2.2 First-Order Autoregression

In the following chapters, we consider the case where $|I| = 2$. Note that setting $|I| = 2$ is an assumption of our model. It can be easily generalized n different regimes. The model is a slight modification of a standard first-order autoregression. Our model considers two first-order autoregressions, with different parameters. The parameters will depend on the state of an underlying process, thus we will switch between two first-order regression using Markov chains with two states. Again, we refer to the next chapter where our model will be specified. For now, we consider the standard form of a first-order regression. The papers [Ham94] and [Ham16] provide us the definition.
A first-order regression is a model that can be used to describe a time-series process. The model states that a variable Y_t depends linearly on Y_{t-1} and a stochastic term, i.e.

$$Y_t = \alpha + \beta Y_{t-1} + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$ and $t = 1, 2, ..., T$. Note that we can rewrite this as

$$Y_t = \alpha + \beta Y_{t-1} + \sigma \epsilon,$$

where $\epsilon \sim N(0,1)$ and $t = 1, 2, ..., T$. It is assumed that we know the value of Y_0. We are interested in Y_t, since we want to predict a stochastic process by analyzing a specific model. In our case, we will take the annual growth rates of the GDP of the Netherlands as Y_t.

2.3 Maximum Likelihood Estimation

In this section, we will introduce the method of maximum likelihood estimation. This method allows us to estimate the values of parameters of some probability distribution based on some observed data. Maximizing the likelihood function will give us the estimated values. These generated values are values for which the observed data is most probable under our proposed model. We follow the definitions provided in [Ric07].

In our model, we will have to make use of conditional likelihood functions. These definitions will be similar. These conditional likelihood functions will be used in order to estimate the values of the parameters of the model for the two different states. In Chapter 3, we will see that the parameters in our models are $\lambda_0, p_{00}, p_{11}, \alpha_0, \alpha_1, \beta_0, \beta_1, \sigma_0$ and σ_1. Maximum likelihood estimation will be used to determine what values are most probable for these parameters under our proposed model given the data on the annual GDP growth rate of the Netherlands during the period 2001-2021.

The random variables $X_1, ..., X_n$ have a distribution depending on some unknown parameter γ. Given the observations, we could estimate the value of γ, which is the most probable under the observations. Note that the likelihood function is a function where we consider γ to be the variable. In other words, we are assuming that $X_1, ..., X_n$ follow a specific family of distributions depending on the value of γ

Definition 2.4. *Suppose that the random variables $X_1, ..., X_n$ have a joint density function $f(x_1, ..., x_n|\gamma)$. Given observed values $X_i = x_i$ $(i = 1, .., n)$ the likelihood of γ as a function of $x_1, ..., x_n$ is defined as*

$$\mathcal{L}(\gamma) = f_{(X_1,...,X_n)}(x_1, ..., x_n|\gamma).$$

The likelihood function simplifies if $X_1, ..., X_n$ are independent and identically distributed. Then, $\mathcal{L}(\gamma) = \prod_{i=1}^{n} f_{X_i}(x_i|\gamma)$. Rather than maximizing the likelihood itself, it is usually easier to maximize its natural logarithm. For an independent and identically distributed sample, the log-likelihood function is given by

$$\ell(\gamma) = \sum_{i=1}^{n} \log(f_{X_i}(x_i|\gamma)).$$

3 Model Specification for a Markov-switching Model with Two Regimes

In this chapter, we will introduce our model, which is a Markov regime-switching model with two states. We will discuss the differences of some models proposed in [Kol19], [SW21] and [Ham16].

We want to give a specification for a variable Y_t, with $t = 1, ..., T$, for which we assume it follows a first-order autoregression. However, we allow the model to switch between two regimes. Let $0, 1$ be the two regimes. Now let S_t be the latent process, which captures in which regime we are. However, since we assume S_t to be a latent process, we cannot observe in which state we are in. Through the variable Y_t, we can infer some information whether we are in state 0 or 1. We assume that the latent process S_t is a Markov chain with initial distribution $\lambda = (\lambda_0, \lambda_1)'$, where $'$ denotes the transpose, and transition matrix $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$. Note that λ is a column vector and a probability vector, which implies that $\lambda_0 + \lambda_1 = 1$ and $\lambda_i \geq 0$, thus

$$\lambda = \begin{pmatrix} \mathbb{P}(S_1 = 0) \\ \mathbb{P}(S_1 = 1) \end{pmatrix} = \begin{pmatrix} \lambda_0 \\ 1 - \lambda_0 \end{pmatrix}. \tag{1}$$

Similarly, P is a stochastic matrix, thus $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$. We get

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}. \tag{2}$$

We want the parameters of the first-order autoregression to depend on the regimes. Let Y_t be described by the following equation

$$Y_t = \alpha_{S_t} + \beta_{S_t} Y_{t-1} + \sigma_{S_t} \epsilon, \tag{3}$$

where $\epsilon \sim N(0, 1)$, $t = 1, ..., T$, and $\alpha_{S_t}, \beta_{S_t}, \sigma_{S_t}$ are constants depending on the latent process S_t. Rewriting equation 3, we have

$$Y_t = \begin{cases} \alpha_0 + \beta_0 Y_{t-1} + \sigma_0 \epsilon & \text{if } S_t = 0 \\ \alpha_1 + \beta_1 Y_{t-1} + \sigma_1 \epsilon & \text{if } S_t = 1 \end{cases}, \tag{4}$$

where again $\epsilon \sim N(0, 1)$. We assume that the actual value of Y_0 is known, and we do not consider the regime of Y_0, i.e. we start with an initial distribution at S_1. Our model is similar to the models introduced in [Kol19] and [Ham16]. Note that Kole proposes a model, where $\beta_0 = \beta_1 = 0$. This will become more apparent in the next chapter. However, we will still follow his method, since he thoroughly discusses the estimation procedure. Hamilton has proposed a model, where both β and σ did not depend on the regimes. We let these parameters vary across states such that we do not limit ourselves to certain applications. In [SW21] the same model as our model is introduced. However, they do not go into much detail of estimating the parameters. We try to combine the papers [Kol19], [Ham16] and [SW21], and give more details on some derived expressions.

We now have a model, which allows for different parameters in different states. Therefore, as introduced in Chapter 1, we have a model which allows us to distinguish between a rising economy and an economy which is in a recession. In Chapter 7, we will exploit this model to analyse the annual GDP growth of the Netherlands.

4 Conditional Probabilities and Densities

In this chapter, we will introduce some notation and start deriving some useful expressions. These expressions will be needed in the next chapter, where we will start estimating the parameters. We will slightly modify the notation provided in [Ham16], in order to make the notation slightly more intuitive.

4.1 Conditional Densities

Let $\Omega_t = \{y_t, y_{t-1} ..., y_1, y_0\}$, i.e. Ω_t is the set which contains the value y_0 and the actual values of the random variables $Y_1, ..., Y_t$. Furthermore, let θ be the vector containing all parameters, thus $\theta = (\sigma_0, \sigma_1, \alpha_0, \alpha_1, \beta_0, \beta_1, p_{00}, p_{11}, \lambda_0)'$. With these notations, we can consider some conditional probabilities and densities. First of all, we are interested in the probability that the latent process S_t is in state 0 or 1 given the parameters and the observations of Y up to time t, i.e. conditional on Ω_t and θ. We denote

$$\xi_{t,j|t} = \mathbb{P}(S_t = j|\Omega_t, \theta) \quad \text{for } j = 0, 1, \text{ and } t = 1, ..., T. \tag{5}$$

Later, we will also see similar expressions, for example $\xi_{t,j|T}$. Then t will be related to S_t and T to the information of Y. Thus, $\xi_{t,j|T} = \mathbb{P}(S_t = j|\Omega_T, \theta)$ for $j = 0, 1$. We are also interested in the density of Y_t conditional on the observations till time $t-1$ and that the parameters are given. Thus we are interested in $f(Y_t|\Omega_{t-1}, \theta)$. Before we can determine an expression for this density, we have to condition on the latent process as well. From equation 3 and 4, we see that

$$\begin{aligned}\mathbb{E}[Y_t|S_t = j, \Omega_{t-1}, \theta] &= \mathbb{E}[\alpha_j + \beta_j Y_{t-1} + \sigma_j \epsilon|S_t = j, \Omega_{t-1}, \theta] \\ &= \mathbb{E}[\alpha_j|S_t = j, \Omega_{t-1}, \theta] + \mathbb{E}[\beta_j Y_{t-1}|S_t = j, \Omega_{t-1}, \theta] + \mathbb{E}[\sigma_j \epsilon|S_t = j, \Omega_{t-1}, \theta].\end{aligned}$$

Now, since we know whether the latent process S_t is in 0 or 1, and the parameter vector is known, α_j is a constant. Therefore, $\mathbb{E}[\alpha_j|S_t = j, \Omega_{t-1}, \theta] = \alpha_j$ for $j = 0, 1$. Similarly, since $S_t = j, \theta$ and Ω_{t-1} is given, we know the value of β_j and Y_{t-1}. Thus, $\mathbb{E}[\beta_j Y_{t-1}|S_t = j, \Omega_{t-1}, \theta] = \beta_j y_{t-1}$. For the third term, we know the value of σ_j since θ is known and the state of S_t is known. We also know that $\epsilon \sim N(0, 1)$, thus the mean of ϵ is equal to zero. We conclude that $\mathbb{E}[\sigma_j \epsilon|S_t = j, \Omega_{t-1}, \theta] = 0$. Combining these results, we get

$$\mathbb{E}[Y_t|S_t = j, \Omega_{t-1}, \theta] = \mathbb{E}[\alpha_j + \beta_j Y_{t-1} + \sigma_j \epsilon|S_t = j, \Omega_{t-1}, \theta] = \begin{cases} \alpha_0 + \beta_0 y_{t-1} & \text{if } S_t = 0 \\ \alpha_1 + \beta_1 y_{t-1} & \text{if } S_t = 1. \end{cases}$$

The argument for the variance is similar. Again conditional on the state, the previous observations and the parameter vector, we know that $\alpha_j + \beta_j Y_{t-1}$ can be considered as a constant. Therefore, it does not change the variance. Note that we also know σ_j. We get

$$\begin{aligned}Var(Y_t|S_t = j, \Omega_t, \theta) &= Var(\alpha_j + \beta_j Y_{t-1} + \sigma_j \epsilon|S_t = j, \Omega_{t-1}, \theta) \\ &= \sigma_j^2 Var(\epsilon) = \sigma_j^2,\end{aligned}$$

for $j = 0, 1$. Note that Y_t conditioned on the event $\{S_t = j, \Omega_{t-1}, \theta\}$, is a linear function of the normal distributed random variable ϵ, where $\alpha_j + \beta_j y_{t-1}$ can be considered as a constant under the conditions. Thus Y_t conditioned on $\{S_t = j, \Omega_{t-1}, \theta\}$ is also normally distributed with mean $\alpha_0 + \beta_j y_{t-1}$ and variance σ_0^2 if $j = 0$, and with mean $\alpha_1 + \beta_j y_{t-1}$ and variance σ_1^2 if $j = 1$. Therefore, we get conditioned on the event that

$$\eta_{jt} = f_{Y_t}(y_t|S_t = j, \Omega_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\Big[\frac{-(y_t - \alpha_j - \beta_j y_{t-1})^2}{2\sigma_j^2}\Big], \tag{6}$$

with $j = 0, 1$, and $t = 1, ..., T$.

We are now able to determine an expression for the density of Y_t conditional on Ω_{t-1} and θ. We will give the result as Corollary. The proof is mainly based on the the law of total probability.

Theorem 4.1. *(Law of Total Probability) Let B_n be a partition of the sample space, then for any event A on the probability space, we have*

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A, B_n).$$

Note that $(A, B_n) \equiv (A \cap B_n)$. The proof of this theorem is very simple, and is therefore not included. The Law of Total Probability is similar for conditional probabilities. This will be used in the proof of the next theorem.

Theorem 4.2. *Let $p_{ij}, \xi_{t-1,i|t-1}$ and η_{jt} be defined as above, then*

$$f_{Y_t}(y_t|\Omega_{t-1}, \theta) = \sum_{i=0}^{1}\sum_{j=0}^{1} p_{ij}\xi_{t-1,i|t-1}\eta_{jt}, \quad \text{for } t = 2, ..., T \tag{7}$$

$$f_{Y_1}(y_1|\Omega_0, \theta) = \sum_{j=0}^{1} \eta_{j1}\lambda_j.$$

Proof. We start with the case that $t = 2, ..., T$. Consider the left-hand side. Note that for every t, the latent process S_t only has two possible states, namely 0 or 1. Therefore, we can sum over all states using the Law of Total Probability. This gives us

$$\begin{aligned} f_{Y_t}(y_t|\Omega_{t-1}, \theta) &= \sum_{j=0}^{1} f_{(Y_t,S_t)}(y_t, S_t = j|\Omega_{t-1}, \theta), \\ &= \sum_{j=0}^{1} f_{Y_t}(y_t|S_t = j, \Omega_{t-1}, \theta)\mathbb{P}(S_t = j|\Omega_{t-1}, \theta). \end{aligned}$$

Again by using the Law of Total Probability, we can rewrite

$$\begin{aligned} \mathbb{P}(S_t = j|\Omega_{t-1}, \theta) &= \sum_{i=0}^{1} \mathbb{P}(S_t = j, S_{t-1} = i|\Omega_{t-1}, \theta), \\ &= \sum_{i=0}^{1} \mathbb{P}(S_t = j|S_{t-1} = i, \Omega_{t-1}, \theta)\mathbb{P}(S_{t-1} = i|\Omega_{t-1}, \theta). \end{aligned}$$

Combining these results, give us

$$\begin{aligned} f_{Y_t}(y_t|\Omega_{t-1}, \theta) &= \sum_{j=0}^{1}\sum_{i=0}^{1} f_{Y_t}(y_t|S_t = j, \Omega_{t-1}, \theta)\mathbb{P}(S_t = j|S_{t-1} = i, \Omega_{t-1}, \theta)\mathbb{P}(S_{t-1} = i|\Omega_{t-1}, \theta), \\ &= \sum_{j=0}^{1}\sum_{i=0}^{1} \eta_{jt}\mathbb{P}(S_t = j|S_{t-1} = i, \theta)\xi_{t-1,i|t-1}, \\ &= \sum_{j=0}^{1}\sum_{i=0}^{1} \eta_{jt}p_{ij}\xi_{t-1,i|t-1}. \end{aligned}$$

Note that $\mathbb{P}(S_t = j|S_{t-1} = i, \Omega_{t-1}, \theta) = \mathbb{P}(S_t = j|S_{t-1} = i, \theta) = p_{ij}$, since S_t is a Markov chain and depends on the previous state and θ. Note that θ contains the value of p_{ij}. Furthermore, recall that $\eta_{jt} = f_{Y_t}(y_t|S_t = j, \Omega_{t-1}, \theta)$ and $\xi_{t-1,i|t-1} = \mathbb{P}(S_{t-1} = i|\Omega_{t-1}, \theta)$. Rearranging the last equality gives us the desired result.

It remains to proof the case of $t = 1$. Again, we start with the left-hand side and apply the law of total probability. This gives us

$$\begin{aligned} f_{Y_1}(y_1|\Omega_0, \theta) &= \sum_{j=0}^{1} f_{Y_1}(y_1, S_1 = j|\Omega_0, \theta), \\ &= \sum_{j=0}^{1} f_{Y_1}(y_1|S_1 = j, \Omega_0, \theta)\mathbb{P}(S_1 = j|\Omega_0, \theta), \\ &= \sum_{j=0}^{1} \eta_{j1}\lambda_j, \end{aligned}$$

where the last equality holds since $\mathbb{P}(S_1 = j|\Omega_0, \theta) = \mathbb{P}(S_1 = j|\theta) = \lambda_j$. Since again, θ contains the value of λ_j. □

4.2 Inference and Forecast Probabilities

As a consequence of the result of the previous section, one can find an expression for $\xi_{t,j|t}$ in terms of $\xi_{t-1,0|t-1}$ and $\xi_{t-1,1|t-1}$.

Theorem 4.3. *Let everything be as above, we have*

$$\begin{aligned}\xi_{t,j|t} &= \frac{\sum_{i=0}^{1} p_{ij}\xi_{t-1,i|t-1}\eta_{jt}}{f_{Y_t}(y_t|\Omega_{t-1},\theta)} \quad \textit{for } j=0,1, \ \textit{and } t=2,...,T, \\ \xi_{1,j|1} &= \frac{\eta_{j1}\lambda_j}{f_{Y_1}(y_1|\Omega_0,\theta)} \quad \textit{for } j=0,1.\end{aligned} \tag{8}$$

Proof. Again, we start with the case $t=2,...,T$. First, we recall the definition of $\xi_{t,j|t}$, we get

$$\xi_{t,j|t} = \mathbb{P}(S_t = j|\Omega_t,\theta) = \mathbb{P}(S_t = j|y_t,\Omega_{t-1},\theta),$$

where in the last equality we used that $\Omega_t = \Omega_{t-1} \cap \{y_t\}$. Using the definition of the conditional density, we get

$$\xi_{t,j|t} = \frac{f_{(Y_t,S_t)}(S_t = j, y_t|\Omega_{t-1},\theta)}{f_{Y_t}(y_t|\Omega_{t-1},\theta)}.$$

Recall that in the proof of equation 7, we have found an expression for $\sum_{j=0}^{1} f_{(Y_t,S_t)}(y_t, S_t = j|\Omega_{t-1},\theta)$, thus also an expression for $f_{(Y_t,S_t)}(y_t, S_t = j|\Omega_{t-1},\theta)$. We get that

$$f_{(Y_t,S_t)}(y_t, S_t = j|\Omega_{t-1},\theta) = \sum_{i=0}^{1} p_{ij}\xi_{t-1,i|t-1}\eta_{jt}.$$

Therefore, we conclude that

$$\xi_{t,j|t} = \frac{\sum_{i=0}^{1} p_{ij}\xi_{t-1,i|t-1}\eta_{jt}}{f_{Y_t}(y_t|\Omega_{t-1},\theta)}.$$

It remains to prove the case where $t=1$. The proof is very similar. We have

$$\begin{aligned}\xi_{1,j|1} &= \mathbb{P}(S_1 = j|\Omega_1,\theta), \\ &= \mathbb{P}(S_1 = j|y_1,\Omega_0,\theta), \\ &= \frac{f_{(Y_1,S_1)}(y_1, S_1 = j|\Omega_0,\theta)}{f_{Y_1}(y_1|\Omega_0,\theta)}, \\ &= \frac{f_{Y_1}(y_1|S_1 = j,\Omega_0,\theta)\mathbb{P}(S_1 = j|\Omega_0,\theta)}{f_{Y_1}(y_1|\Omega_0,\theta)} = \frac{\eta_{j1}\lambda_j}{f_{Y_1}(y_1|\Omega_0,\theta)}.\end{aligned}$$

□

Corollary 4.3.1. *We have that*

$$\begin{aligned}\xi_{t,j|t} &= \frac{\sum_{i=0}^{1} p_{ij}\xi_{t-1,i|t-1}\eta_{jt}}{\sum_{i=0}^{1}\sum_{j=0}^{1} p_{ij}\xi_{t-1,i|t-1}\eta_{jt}} \quad \textit{for } j=0,1, \ \textit{and } t=2,...,T, \\ \xi_{1,j|1} &= \frac{\eta_{j1}\lambda_j}{\sum_{j=0}^{1}\eta_{j1}\lambda_j} \quad \textit{for } j=0,1.\end{aligned} \tag{9}$$

Proof. Combining the results of 4.2 and 4.3 gives us the desired result. □

We call the vector $(\xi_{t,0|t}, \xi_{t,1|t})'$ the inference vector at time t, since based on the observations at time t, we determine the probability of both regimes at that time. We can also find expressions for $\xi_{t+1,j|t}$ and $\xi_{t,j|T}$. Furthermore, $\xi_{t+1,j|t}$ can be seen as a forecast, since based on the observations at time t, we consider the probability of both regimes at time $t+1$. Lastly, $\xi_{t,j|T}$ is also a sort of inference probability. We call this expression the smoothed inference, since now we determine the probability of the regimes at time t based on all the observations. We formulate these results as two theorems.

Theorem 4.4. *Using the notation introduced, we have*

$$\xi_{t+1,j|t} = \sum_{i=0}^{1} p_{ij}\xi_{t,i|t}, \quad \textit{for } t = 1, .., T-1. \tag{10}$$

Proof. Recall that $\xi_{t+1,j|t} = \mathbb{P}(S_{t+1} = j|\Omega_t, \theta)$. Applying the law of total probability gives us

$$\begin{aligned}\mathbb{P}(S_{t+1} = j|\Omega_t, \theta) &= \sum_{i=0}^{1} \mathbb{P}(S_{t+1} = j, S_t = i|\Omega_t, \theta) \\ &= \sum_{i=0}^{1} \mathbb{P}(S_{t+1} = j|S_t = i, \Omega_t, \theta)\mathbb{P}(S_t = i|\Omega_t, \theta).\end{aligned}$$

Note that S_{t+1} only depends on its own previous value, thus $\mathbb{P}(S_{t+1} = j|S_t = i, \Omega_t, \theta) = \mathbb{P}(S_{t+1} = j|S_t = i) = p_{ij}$. Recall that $\mathbb{P}(S_t = i|\Omega_t, \theta) = \xi_{t,i|t}$. Thus, we have derived equation 10. □

Theorem 4.5. *Again, using the notation introduced, we find*

$$\xi_{t,j|T} = \sum_{i=0}^{1} \xi_{t,j|t} p_{ji} \frac{\xi_{t+1|T,i}}{\xi_{t+1|t,i}}, \quad \textit{for } t = 1, .., T-1. \tag{11}$$

Proof. Our proof is similar to the proof in [Ham94]. However, we do not consider the impact of exogenous variables. This proof consists of several steps:

1. Show that $\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_T, \theta) = \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta)$.
2. Show that $\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta) = \frac{p_{ji}\mathbb{P}(S_t=j|\Omega_t,\theta)}{\mathbb{P}(S_{t+1}=i|\Omega_t,\theta)}$.
3. Show that
$$\mathbb{P}(S_t = j, S_{t+1} = i|\Omega_T, \theta) = \mathbb{P}(S_{t+1} = i|\Omega_T, \theta)\frac{p_{ji}\mathbb{P}(S_t = j|\Omega_t, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)}. \tag{12}$$
4. Find an expression for $\xi_{t,j|T}$.

Step 1: It is enough to show that

$$\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_{t+1}, \theta) = \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta), \quad \text{for any } t < T.$$

Recall that $\Omega_{t+1} = \Omega_t \cap y_{t+1}$. Therefore, rewriting the left side gives us

$$\begin{aligned}\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_{t+1}, \theta) &= \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, y_{t+1}, \theta), \\ &= \frac{\mathbb{P}(S_t = j, y_{t+1}|S_{t+1} = i, \Omega_t, \theta)}{f(y_{t+1}|S_{t+1} = i, \Omega_t, \theta)}, \\ &= \frac{f(y_{t+1}|S_t = j, S_{t+1} = i, \Omega_t, \theta)\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta)}{f(y_{t+1}|S_{t+1} = i, \Omega_t, \theta)} \\ &= \frac{f(y_{t+1}|S_{t+1} = i, \Omega_t, \theta)}{f(y_{t+1}|S_{t+1} = i, \Omega_t, \theta)}\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta).\end{aligned}$$

The last equality follows from the definition of our model. The value of Y_{t+1} only depends on the current state, thus $f(y_{t+1}|S_t = j, S_{t+1} = i, \Omega_t, \theta) = f(y_{t+1}|S_{t+1} = i, \Omega_t, \theta)$. Therefore, we can conclude that $\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_{t+1}, \theta) = \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta)$, for any $t < T$. We can now show that the $\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_T, \theta) = \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta)$ holds by increasing one unit at a time until it reaches T. This concludes the first step.

Step 2: Starting by the left-hand side, we get

$$\begin{aligned}\mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta) &= \frac{\mathbb{P}(S_t = j, S_{t+1} = i|\Omega_t, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)},\\ &= \frac{\mathbb{P}(S_{t+1} = i|S_t = j, \Omega_t, \theta)\mathbb{P}(S_t = j|\Omega_t, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)},\\ &= \frac{p_{ji}\mathbb{P}(S_t = j|\Omega_t, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)},\end{aligned}$$

where the last equality follows from the fact that S_{t+1} only depends on S_t, i.e. $\mathbb{P}(S_{t+1} = i|S_t = j, \Omega_t, \theta) = \mathbb{P}(S_{t+1} = i|S_t = j) = p_{ji}$.

Step 3: We have

$$\begin{aligned}\mathbb{P}(S_t = j, S_{t+1} = i|\Omega_T, \theta) &= \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_T, \theta)\mathbb{P}(S_{t+1} = i|\Omega_T, \theta),\\ &= \mathbb{P}(S_t = j|S_{t+1} = i, \Omega_t, \theta)\mathbb{P}(S_{t+1} = i|\Omega_T, \theta),\\ &= \frac{p_{ji}\mathbb{P}(S_t = j|\Omega_t, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)}\mathbb{P}(S_{t+1} = i|\Omega_T, \theta).\end{aligned}$$

The second equality follows from the first step, and the third equality follows from the second step.

Step 4: It remains to get an expression for $\xi_{t|T,j}$. By the law of total probability, we have

$$\begin{aligned}\xi_{t,j|T} = \mathbb{P}(S_t = j|\Omega_T, \theta) &= \sum_{i=0}^{1} \mathbb{P}(S_t = j, S_{t+1} = i|\Omega_T, \theta),\\ &= \sum_{i=0}^{1} \frac{p_{ji}\mathbb{P}(S_t = j|\Omega_t, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)}\mathbb{P}(S_{t+1} = i|\Omega_T, \theta),\\ &= \sum_{i=0}^{1} \mathbb{P}(S_t = j|\Omega_t, \theta)p_{ji}\frac{\mathbb{P}(S_{t+1} = i|\Omega_T, \theta)}{\mathbb{P}(S_{t+1} = i|\Omega_t, \theta)},\\ &= \sum_{i=0}^{1} \xi_{t,j|t}p_{ji}\frac{\xi_{t+1,i|T}}{\xi_{t+1,i|t}}.\end{aligned}$$

The second equality follows from step 3. In the third equality, we have rearranged the terms. □

We note that $\xi_{T,j|T}$ can be derived using the formula of the inference vectors provided in equation 9. We now have everything needed to start estimating the parameters using the Maximum Likelihood Estimator (MLE), and start applying a method called "Expectation-Maximization Algorithm". This will be discussed in the next chapter. In Chapter 6, we will see how these derived results can be used in the expectation-maximization algorithm.

5 Estimation Procedure of the Parameters

In this chapter, we will consider some likelihood functions, which will be necessary to find estimations of the unknown parameter vector θ using maximum likelihood estimation. We will make use of a concept called "Expectation-Maximization Algorithm", which guarantees that a maximum can be found using the joint likelihood function.

5.1 Likelihood Functions

We have introduced the concept of maximum likelihood estimation in section 2.3. We are considering a likelihood function conditioned on some set of observations, which makes $Y_1, ..., Y_T$ independent of each other conditioned on the observations. This likelihood function is then given by

$$\mathcal{L}(\Omega_T, \theta) = \prod_{t=1}^{T} f_{Y_t}(y_t|\Omega_{t-1}, \theta). \tag{13}$$

It is often preferred to use the log-likelihood function instead, since it simplifies the determination of the optimal value such that the likelihood function is maximized. We get

$$\begin{aligned} \ell(\Omega_T, \theta) = \log(\prod_{t=1}^{T} f_{Y_t}(y_t|\Omega_{t-1}, \theta) &= \sum_{t=1}^{T} \log(f_{Y_t}(y_t|\Omega_{t-1}, \theta)), \\ &= \log(\sum_{j=0}^{1} \eta_{j1}\lambda_j) + \sum_{t=2}^{T} \log(\sum_{j=0}^{1}\sum_{i=0}^{1} \eta_{jt} p_{ij} \xi_{t-1,i|t-1}), \end{aligned} \tag{14}$$

where the last equality follows from Theorem 4.2. We now have an expression for a likelihood function, which we would like to maximize. This function is problematic due to many local optima, which makes it complicated to find the global maximum. We take a different approach, which is used in [Kol19],[Ham16], and is first introduced by [DLR77]. This approach is called the Expectation-Maximization algorithm.

For this algorithm, we have to make use of the joint log-likelihood function. Therefore, we need to introduce some notation. Let $\mathcal{S}_t := \{s_1, s_2, ..., s_t\}$, where we suppose that we can observe the realizations of the latent process S_t for now. We now consider the joint density function of (y_t, s_t) conditional on $\Omega_{t-1}, \mathcal{S}_{t-1}$ and θ, for $t = 2, ..., T$. We consider the special case $t = 1$ afterwards. We get

$$\begin{aligned} f_{(Y_t,S_t)}(y_t, s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta) &= f_{Y_t}(y_t|\Omega_{t-1}, S_t = s_t, \mathcal{S}_{t-1}, \theta)\mathbb{P}(S_t = s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta) \\ &= f_{Y_t}(y_t|S_t = s_t, \Omega_{t-1}, \theta)\mathbb{P}(S_t = s_t|S_{t-1} = s_{t-1}, \theta), \end{aligned}$$

where the last equality follows from the fact that Y_t only depends on the current regime, instead of all past regimes as well, and the fact that S_t is a Markov chain and thus only depends on S_{t-1}. Furthermore, it depends on θ, since the transition probabilities are contained in θ. Substituting gives us

$$\begin{aligned} f_{(Y_t,S_t)}(y_t, s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta) &= \eta_{s_t t} p_{s_{t-1} s_t} \\ &= \begin{cases} \eta_{0t}p_{00} & \text{if } s_t = 0, s_{t-1} = 0; \\ \eta_{0t}p_{10} & \text{if } s_t = 0, s_{t-1} = 1; \\ \eta_{1t}p_{01} & \text{if } s_t = 1, s_{t-1} = 0; \\ \eta_{1t}p_{11} & \text{if } s_t = 1, s_{t-1} = 1. \end{cases} \end{aligned}$$

Recall that $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$, thus we can rewrite the equation above. We get

$$f_{(Y_t,S_t)}(y_t, s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta) = (\eta_{0t}p_{00})^{(1-s_t)(1-s_{t-1})}(\eta_{0t}(1-p_{11}))^{(1-s_t)s_{t-1}}(\eta_{1t}(1-p_{00}))^{s_t(1-s_{t-1})}(\eta_{1t}p_{11})^{s_t s_{t-1}}.$$

We are interested in determining the log-likelihood of the joint density of (y_t, s_t). First, we take the log of the previous equation. This gives us

$$\begin{aligned} \log(f_{(Y_t,S_t)}(y_t, s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta)) &= (1 - s_t)(1 - s_{t-1}) \log(\eta_{0t}p_{00}) + (1 - s_t)s_{t-1} \log(\eta_{0t}(1 - p_{11})) \\ &\quad + s_t(1 - s_{t-1}) \log(\eta_{1t}(1 - p_{00})) + s_t s_{t-1} \log(\eta_{1t}p_{11}). \end{aligned}$$

It remains to determine an expression for $\log(f_{(Y_t,S_t)}(y_t, s_t|\Omega_{t-1}, \theta))$ for $t = 1$. Recall that S_t is a Markov chain with initial distribution λ, thus $\mathbb{P}(S_1 = 0) = \lambda_0$ and $\mathbb{P}(S_1 = 1) = \lambda_1 = 1 - \lambda_0$. Note that

$$\begin{aligned} f_{(Y_1,S_1)}(y_1, s_1|\Omega_0, \theta) &= f_{Y_1}(y_1|S_1 = s_1, \Omega_0, \theta)\mathbb{P}(S_1 = s_1|\theta), \\ &= \eta_{s_1 1}\mathbb{P}(S_1 = s_1) = \eta_{s_1 1}\lambda_{s_1}, \\ &= \begin{cases} \eta_{01}\lambda_0 & \text{if } s_1 = 0, \\ \eta_{11}(1-\lambda_0) & \text{if } s_1 = 1. \end{cases} \end{aligned}$$

Thus, we get $f_{(Y_1,S_1)}(y_1, s_1|\Omega_0, \theta) = (\eta_{01}\lambda_0)^{1-s_1} + (\eta_{11}(1-\lambda_0))^{s_1}$, and therefore

$$\log(f_{(Y_1,S_1)}(y_1, s_1|\Omega_0, \theta)) = (1 - s_1)\log(\eta_{01}\lambda_0) + s_1 \log(\eta_{11}(1 - \lambda_0)).$$

We can now determine the joint log-likelihood function. We have

$$\begin{aligned} \ell(\Omega_T, \mathcal{S}_T, \theta) &= \sum_{t=1}^{T} \log(f_{(Y_t,\mathcal{S}_t)}(y_t, s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta)), \\ &= \log(f_{(Y_1,S_1)}(y_1, s_1|\theta)) + \sum_{t=2}^{T} \log(f_{(Y_t,\mathcal{S}_t)}(y_t, s_t|\Omega_{t-1}, \mathcal{S}_{t-1}, \theta)). \end{aligned}$$

In order to rewrite this, we want to get similar terms together. Therefore, we rewrite $\log(\eta_{0t}p_{00}) = \log(\eta_{0t}) + \log(p_{00})$, $\log(\eta_{01}\lambda_0) = \log(\eta_{01}) + \log(\lambda_0)$ and all the other terms are rewritten in a similar manner. Let $\ell_{\eta_{0t}}(\Omega_T, \mathcal{S}_T, \theta)$ denote the terms of $\ell(\Omega_T, \mathcal{S}_T, \theta)$ containing η_{0t}, and all other terms are defined similar. Then we have

$$\ell(\Omega_T, \mathcal{S}_T, \theta) = \ell_{\eta_{0t}}(\Omega_T, \mathcal{S}_T, \theta) + \ell_{\eta_{1t}}(\Omega_T, \mathcal{S}_T, \theta) + \ell_{p_{00}}(\Omega_T, \mathcal{S}_T, \theta) + \ell_{p_{11}}(\Omega_T, \mathcal{S}_T, \theta) + \ell_{\lambda_0}(\Omega_T, \mathcal{S}_T, \theta),$$

since there are no terms containing a product of two different terms. Now, starting by adding the terms containing η_{0t}, we get

$$\begin{aligned} \ell_{\eta_{0t}}(\Omega_T, \mathcal{S}_T, \theta) &= (1 - s_1)\log(\eta_{01}) + \sum_{t=2}^{T}(1 - s_t)(1 - s_{t-1})\log(\eta_{0t}) + (1 - s_t)s_{t-1}\log(\eta_{0t}) \\ &= (1 - s_1)\log(\eta_{01}) + \sum_{t=2}^{T}(1 - s_t)\log(\eta_{0t}) = \sum_{t=1}^{T}(1 - s_t)\log(\eta_{0t}), \end{aligned}$$

where the second equality follows from the fact that $(1 - s_t)(1 - s_{t-1}) + (1 - s_t)s_{t-1} = 1 - s_t$, since s_{t-1} can be either 0 or 1. All other terms are done similarly. We conclude that

$$\begin{aligned} \ell(\Omega_T, \mathcal{S}_T, \theta) = \sum_{t=1}^{T} \big((1 - s_t)\log(\eta_{0t}) + s_t \log(\eta_{1t})\big) + \sum_{t=2}^{T} \big((1 - s_t)(1 - s_{t-1})\log(p_{00}) + (1 - s_t)s_{t-1}\log(1 - p_{11}) \\ s_t(1 - s_{t-1})\log(1 - p_{00}) + s_t s_{t-1}\log(p_{11})\big) + (1 - s_1)\log(\lambda_0) + s_1 \log(1 - \lambda_0). \end{aligned} \tag{15}$$

5.2 Expectation-Maximization Algorithm

In the previous section, we have assumed that the observations on S_t are known, which was a violation of the assumption that S_t is a latent process. However, in the expectation-maximization algorithm, the expectation of the joint log-likelihood will be maximized. Therefore, we do not need to know the actual observations of S_t, and will only need the expectations of S_t conditional on Ω_T and θ. We consider the EM-algorithm. In this algorithm, we start with a given parameter vector $\theta^{(0)}$, for which we then calculate the expectations. Then we treat the expectations as known, and estimate the parameter vector again by maximizing the expected joint log-likelihood function. This gives us $\theta^{(1)}$. We repeat this process. Therefore, we have a series of maximization's, thus we have a series of parameter estimates. We denote the $k-$th estimated parameter vector by $\theta^{(k)}$, which satisfies

$$\theta^{(k)} = \text{argmax}_\theta \mathbb{E}[\ell(\Omega_T, \mathcal{S}_T, \theta)|\Omega_T, \theta^{(k-1)}].$$

Therefore, we have to slightly modify our notation of $\xi_{t,j|t}, \xi_{t+1,j|t}$ and $\xi_{t,j|T}$. We will add a superscript to indicate which θ we use. For example, $\xi_{t,j|t}^{(k-1)} = \mathbb{P}(S_t = j|\Omega_t, \theta^{(k-1)})$. We will show that the sequence $\theta^{(k)}$ has a 'nice' property. First, denote

$$\ell_{EM}(\Omega_T, \theta, \theta^{(k-1)}) = \mathbb{E}[\ell(\Omega_T, \mathcal{S}_T, \theta)|\Omega_T, \theta^{(k-1)}], \tag{16}$$

thus $\theta^{(k)} = \text{argmax}_\theta \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)})$.

Theorem 5.1. *The loglikelihood function (equation 14) increases at each iteration, i.e.*

$$\ell(\Omega_T, \theta^{(k)}) \geq \ell(\Omega_T, \theta^{(k-1)}).$$

Proof. This proof is inspired by [Ham90] and mainly by [Tab21]. Let $y = (y_1, ..., y_T)'$ and $s = (s_1, ..., s_T)'$. Note that if y is given, we can use our notation introduced earlier, namely Ω_T. The conditional probability formula gives

$$f(y|\theta) = \frac{f(s, y|\theta)}{f(s|y, \theta)} = \frac{f(s, y|\theta)}{f(s|\Omega_T, \theta)}.$$

Taking the logarithm on both sides gives us

$$\log(f(y|\theta)) = \log(\frac{f(s, y|\theta)}{f(s|\Omega_T, \theta)}) = \log(f(s, y|\theta)) - \log(f(s|\Omega_T, \theta)).$$

Now, we multiply both sides with $f(s|\Omega_T, \theta^{(k-1)})$. We get

$$f(s|\Omega_T, \theta^{(k-1)}) \log(f(y|\theta)) = f(s|\Omega_T, \theta^{(k-1)}) \log(f(s, y|\theta)) - f(s|\Omega_T, \theta^{(k-1)}) \log(f(s|\Omega_T, \theta)).$$

We can now take the summation over all possible outcomes of s, i.e.

$$\sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(y|\theta)) = \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s, y|\theta)) - \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s|\Omega_T, \theta)).$$

Note that $\log(f(y|\theta))$ does not depend on s, and

$$\sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) = 1.$$

Therefore, it follows that

$$\log(f(y|\theta)) = \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s, y|\theta)) - \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s|\Omega_T, \theta)).$$

Recall that $\ell(\Omega_T, \theta) = \log(f(y|\theta))$ and $\log(f(s, y|\theta)) = \ell(\Omega_T, \mathcal{S}_T, \theta)$, thus

$$\sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s, y|\theta)) = \mathbb{E}[\ell(\Omega_T, \mathcal{S}_T, \theta)|\Omega_T, \theta^{(k-1)}] = \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)}).$$

Therefore,

$$\ell(\Omega_T, \theta) = \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)}) - \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s|\Omega_T, \theta)),$$

and for $\theta = \theta^{(k-1)}$, we get

$$\ell(\Omega_T, \theta^{(k-1)}) = \ell_{EM}(\Omega_T, \theta^{(k-1)}, \theta^{(k-1)}) - \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(f(s|\Omega_T, \theta^{(k-1)})).$$

Subtracting gives us

$$\ell(\Omega_T, \theta) - \ell(\Omega_T, \theta^{(k-1)}) = \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)}) - \ell_{EM}(\Omega_T, \theta^{(k-1)}, \theta^{(k-1)}) - \sum_{s\in\{0,1\}^T} f(s|\Omega_T, \theta^{(k-1)}) \log(\frac{f(s|\Omega_T, \theta)}{f(s|\Omega_T, \theta^{(k-1)})}).$$

Recall that the Jensen's inequality for a concave function φ gives us that $\varphi(\mathbb{E}[X]) \geq \mathbb{E}[\varphi(X)]$. Since the logarithm is indeed a concave function, we have

$$\sum_{s\in\{0,1\}^T} f(s|\Omega_T,\theta^{(k-1)})\log(\frac{f(s|\Omega_T,\theta)}{f(s|\Omega_T,\theta^{(k-1)})}) \leq \log(\sum_{s\in\{0,1\}^T} f(s|\Omega_T,\theta^{(k-1)})\frac{f(s|\Omega_T,\theta)}{f(s|\Omega_T,\theta^{(k-1)})}),$$
$$= \log(\sum_{s\in\{0,1\}^T} f(s|\Omega_T,\theta)) = \log(1) = 0.$$

Therefore, we conclude that

$$\ell(\Omega_T,\theta) - \ell(\Omega_T,\theta^{(k-1)}) \geq \ell_{EM}(\Omega_T,\theta,\theta^{(k-1)}) - \ell_{EM}(\Omega_T,\theta^{(k-1)},\theta^{(k-1)}).$$

Setting $\theta = \theta^{(k)}$, we get

$$\ell(\Omega_T,\theta^{(k)}) - \ell(\Omega_T,\theta^{(k-1)}) \geq \ell_{EM}(\Omega_T,\theta^{(k)},\theta^{(k-1)}) - \ell_{EM}(\Omega_T,\theta^{(k-1)},\theta^{(k-1)}).$$

By construction of $\theta^{(k)}$, we know that $\theta^{(k)}$ maximizes $\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)})$. Therefore,

$$\ell_{EM}(\Omega_T,\theta^{(k)},\theta^{(k-1)}) \geq \ell_{EM}(\Omega_T,\theta^{(k-1)},\theta^{(k-1)}),$$

or equivalently

$$\ell_{EM}(\Omega_T,\theta^{(k)},\theta^{(k-1)}) - \ell_{EM}(\Omega_T,\theta^{(k-1)},\theta^{(k-1)}) \geq 0.$$

We can conclude that

$$\ell(\Omega_T,\theta^{(k)}) - \ell(\Omega_T,\theta^{(k-1)}) \geq 0,$$

and thus

$$\ell(\Omega_T,\theta^{(k)}) \geq \ell(\Omega_T,\theta^{(k-1)}).$$

□

Therefore, we have shown that the likelihood function is an increasing function with respect to the series $\theta^{(k)}$. Each iteration gives parameter estimates that are more likely. Since the likelihood function is increasing, the likelihood converges to the most likely (local) maximum. However, we still can not be sure whether we converge to the global maximum. There are several methods, which can be used to find global maxima with many local maxima, for example an algorithm called Simulated Annealing (See [Rut89]).

5.3 First Order Conditions of the Parameters

We are interested in maximizing $\mathbb{E}[\ell(\Omega_T,\mathcal{S}_T,\theta)|\Omega_T,\theta^{(k-1)}]$ with respect to all the components of θ. Recall equation 15. The first order conditions are given by

$$\frac{\partial\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)})}{\partial\theta}\Big|_{\theta=\theta^{(k)}} = 0. \tag{17}$$

First, we rewrite $\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)})$. We have

$$\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)}) = \sum_{t=1}^{T}\big(\mathbb{E}[1-S_t|\Omega_T,\theta^{(k-1)}]\log(\eta_{0t}) + \mathbb{E}[S_t|\Omega_T,\theta^{(k-1)}]\log(\eta_{1t})\big)$$
$$+\sum_{t=2}^{T}\big(\mathbb{E}[(1-S_t)(1-S_{t-1})|\Omega_T,\theta^{(k-1)}]\log(p_{00}) + \mathbb{E}[(1-S_t)S_{t-1}|\Omega_T,\theta^{(k-1)}]\log(1-p_{11})$$
$$+\mathbb{E}[S_t(1-S_{t-1})|\Omega_T,\theta^{(k-1)}]\log(1-p_{00}) + \mathbb{E}[S_tS_{t-1}|\Omega_T,\theta^{(k-1)}]\log(p_{11})\big)$$
$$\mathbb{E}[1-S_1|\Omega_T,\theta^{(k-1)}]\log(\lambda_0) + \mathbb{E}[S_1|\Omega_T,\theta^{(k-1)}]\log(1-\lambda_0),$$

since expectation is a linear operator and $\log(\eta_{jt}), \log(p_{ij})$ are all constants. We start by rewriting the expectation parts. We get

$$\begin{aligned}
\mathbb{E}[1-S_t|\Omega_T,\theta^{(k-1)}] &= 1\cdot\mathbb{P}(1-S_t=1|\Omega_T,\theta^{(k-1)}) = \mathbb{P}(S_t=0|\Omega_T,\theta^{(k-1)}) = \xi_{t,0|T}^{(k-1)},\\
\mathbb{E}[S_t|\Omega_T,\theta^{(k-1)}] &= 1\cdot\mathbb{P}(S_t=1|\Omega_T,\theta^{(k-1)}) = \xi_{t,1|T}^{(k-1)},\\
\mathbb{E}[(1-S_t)(1-S_{t-1})|\Omega_T,\theta^{(k-1)}] &= 1\cdot\mathbb{P}(S_t=0,S_{t-1}=0|\Omega_T,\theta^{(k-1)}),\\
\mathbb{E}[(1-S_t)S_{t-1}|\Omega_T,\theta^{(k-1)}] &= 1\cdot\mathbb{P}(S_t=0,S_{t-1}=1|\Omega_T,\theta^{(k-1)}),\\
\mathbb{E}[S_t(1-S_{t-1})|\Omega_T,\theta^{(k-1)}] &= 1\cdot\mathbb{P}(S_t=1,S_{t-1}=0|\Omega_T,\theta^{(k-1)}),\\
\mathbb{E}[S_tS_{t-1}|\Omega_T,\theta^{(k-1)}] &= 1\cdot\mathbb{P}(S_t=1,S_{t-1}=1|\Omega_T,\theta^{(k-1)}).
\end{aligned}$$

We introduce some notation. Let $\tilde{p}_{ji,t+1}^{(k-1)} = \mathbb{P}(S_{t+1}=i,S_t=j|\Omega_T,\theta^{(k-1)})$. Equation 12 provides that

$$\begin{aligned}
\tilde{p}_{ji,t+1}^{(k-1)} &= \mathbb{P}(S_t=j,S_{t+1}=i|\Omega_T,\theta^{(k-1)}) \qquad (18)\\
&= \mathbb{P}(S_{t+1}=i|\Omega_T,\theta^{(k-1)})\frac{p_{ji}^{(k-1)}\mathbb{P}(S_t=j|\Omega_t,\theta^{(k-1)})}{\mathbb{P}(S_{t+1}=i|\Omega_t,\theta^{(k-1)})} = \xi_{t,j|t}^{(k-1)}p_{ji}\frac{\xi_{t+1,i|T}^{(k-1)}}{\xi_{t+1,i|t}^{(k-1)}}.
\end{aligned}$$

We conclude that

$$\begin{aligned}
\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)}) &= \sum_{t=1}^{T}\big(\xi_{t,0|T}^{(k-1)}\log(\eta_{0t}) + \xi_{t,1|T}^{(k-1)}\log(\eta_{1t})\big) \qquad (19)\\
&+\sum_{t=2}^{T}\big(\tilde{p}_{00,t}^{(k-1)}\log(p_{00}) + \tilde{p}_{10,t}^{(k-1)}\log(1-p_{11}) + \tilde{p}_{01,t}^{(k-1)}\log(1-p_{00}) + \tilde{p}_{11,t}^{(k-1)}\log(p_{11})\big)\\
&\xi_{1,0|T}^{(k-1)}\log(\lambda_0) + \xi_{1,1|T}^{(k-1)}\log(1-\lambda_0).
\end{aligned}$$

Recall that $\theta = (\sigma_0,\sigma_1,\alpha_0,\alpha_1,\beta_0,\beta_1,p_{00},p_{11},\lambda_0)'$. We now determine the optimal parameters.

5.3.1 Estimation of the Parameters in η_{0t} and η_{1t}

Recall that equation 6 gives us that

$$\eta_{jt} = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\Big[\frac{-(y_t-\alpha_j-\beta_jy_{t-1})^2}{2\sigma_j^2}\Big],$$

thus

$$\log(\eta_{0t}) = \log(\frac{1}{\sqrt{2\pi}\sigma_0}) - \frac{(y_t-\alpha_0-\beta_0y_{t-1})^2}{2\sigma_0^2}.$$

The only parameters in η_{0t} are α_0,β_0 and σ_0. These parameters are in not any other terms of $\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)})$. We start with the first-order condition for α_0, i.e.

$$\frac{\partial\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)})}{\partial\alpha_0}\Big|_{\theta=\theta^{(k)}} = 0.$$

Note that

$$\frac{\partial}{\partial\alpha_0}\log(\eta_{0t}) = \frac{\partial}{\partial\alpha_0}(\log(\frac{1}{\sqrt{2\pi}\sigma_0}) - \frac{(y_t-\alpha_0-\beta_0y_{t-1})^2}{2\sigma_0^2}) = \frac{2(y_t-\alpha_0-\beta_0y_{t-1})}{2\sigma_0^2},$$

thus

$$\begin{aligned}
\frac{\partial\ell_{EM}(\Omega_T,\theta,\theta^{(k-1)})}{\partial\alpha_0} &= \frac{\partial}{\partial\alpha_0}\sum_{t=1}^{T}\xi_{t,0|T}^{(k-1)}\log(\eta_{0t}) = \sum_{t=1}^{T}\xi_{t,0|T}^{(k-1)}\frac{\partial}{\partial\alpha_0}\log(\eta_{0t}),\\
&= \sum_{t=1}^{T}\xi_{t,0|T}^{(k-1)}\frac{2(y_t-\alpha_0-\beta_0y_{t-1})}{2\sigma_0^2}.
\end{aligned}$$

Splitting this summation into two parts, taking $\theta_0 = \theta_0^{(k)}$ and setting it equal to 0 gives us

$$\sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \frac{y_t - \beta_0^{(k)} y_{t-1}}{(\sigma_0^{(k)})^2} - \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \frac{\alpha_0^{(k)}}{(\sigma_0^{(k)})^2} = 0.$$

Rewriting gives

$$\alpha_0^{(k)} = \frac{\sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} (y_t - \beta_0^{(k)} y_{t-1})}{\sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)}}. \tag{20}$$

Note that equation 20 still contains $\beta_0^{(k)}$. We will need to substitute this expression into the first-order condition of β_0. To avoid confusion, we will make use of a different index of the summation of $\alpha_0^{(k)}$. First, note that

$$\frac{\partial}{\partial \beta_0} \log(\eta_{0t}) = \frac{2y_{t-1}(y_t - \alpha_0 - \beta_0 y_{t-1})}{2\sigma_0^2}.$$

This gives us

$$\begin{aligned}\frac{\partial \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)})}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \log(\eta_{0t}) = \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \frac{\partial}{\partial \beta_0} \log(\eta_{0t}), \\ &= \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \frac{2y_{t-1}(y_t - \alpha_0 - \beta_0 y_{t-1})}{2\sigma_0^2}.\end{aligned}$$

Evaluating at $\theta = \theta^{(k)}$ and setting the expression equal to 0 gives us

$$\frac{\partial \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)})}{\partial \beta_0}\Big|_{\theta=\theta^{(k)}} = \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \frac{y_{t-1}(y_t - \alpha_0^{(k)} - \beta_0^{(k)} y_{t-1})}{(\sigma_0^{(k)})^2} = 0,$$

thus

$$\sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} y_{t-1}(y_t - \frac{\sum_{n=1}^{T} \xi_{n,0|T}^{(k-1)} (y_n - \beta_0^{(k)} y_{n-1})}{\nu_0^{(k-1)}} - \beta_0^{(k)} y_{t-1}) = 0,$$

where $\nu_0^{(k-1)} = \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)}$. Rewriting this equation gives us

$$\sum_{t=1}^{T} \big(\xi_{t,0|T}^{(k-1)} y_{t-1}(y_t \nu_0^{(k-1)} - \sum_{n=1}^{T} \xi_{n,0|T}^{(k-1)} y_n)\big) - \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} y_{t-1}(\nu_0^{(k-1)} \beta_0^{(k)} y_{t-1} + \sum_{n=1}^{T} \xi_{n,0|T}^{(k-1)} \beta_0^{(k)} y_{n-1}) = 0.$$

Therefore, we conclude that

$$\beta_0^{(k)} = \frac{\sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} y_{t-1}(y_t \nu_0^{(k-1)} - \sum_{n=1}^{T} \xi_{n,0|T}^{(k-1)} y_n)}{\sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} y_{t-1}(\nu_0^{(k-1)} y_{t-1} + \sum_{n=1}^{T} \xi_{n,0|T}^{(k-1)} y_{n-1})}. \tag{21}$$

Thus we have found an expression for $\beta_0^{(k)}$ without an $\alpha_0^{(k)}$ term, which allows us to substitute the value of $\beta_0^{(k)}$ in equation 20. It remains to find an expression for $\sigma_0^{(k)}$. We have that

$$\begin{aligned}\frac{\partial}{\partial \sigma_0} \log(\eta_{0t}) &= \frac{\partial}{\partial \sigma_0}(\log(\frac{1}{\sqrt{2\pi}\sigma_0}) - \frac{(y_t - \alpha_0 - \beta_0 y_{t-1})^2}{2\sigma_0^2}), \\ &= \sqrt{2\pi}\sigma_0 \frac{-\sqrt{2\pi}}{2\pi\sigma_0^2} - \frac{(y_t - \alpha_0 - \beta_0 y_{t-1})^2}{2} \frac{-2}{\sigma_0^3} = \frac{-1}{\sigma_0} + \frac{(y_t - \alpha_0 - \beta_0 y_{t-1})^2}{\sigma_0^3}.\end{aligned}$$

As a result,

$$\begin{aligned}\frac{\partial \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)})}{\partial \sigma_0} &= \frac{\partial}{\partial \sigma_0} \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \log(\eta_{0t}) = \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} \frac{\partial}{\sigma_0} \log(\eta_{0t}), \\ &= \sum_{t=1}^{T} \xi_{t,0|T}^{(k-1)} (\frac{-1}{\sigma_0} + \frac{(y_t - \alpha_0 - \beta_0 y_{t-1})^2}{\sigma_0^3}).\end{aligned}$$

Evaluating at $\theta = \theta^{(k)}$, and setting it equal to zero gives

$$-\sum_{t=1}^{T}(\xi_{t,0|T}^{(k-1)}(\sigma_0^{(k)})^2) + \sum_{t=1}^{T}\xi_{t,0|T}^{(k-1)}(y_t - \alpha_0^{(k)} - \beta_0^{(k)}y_{t-1})^2 = 0.$$

Therefore, we conclude that

$$\sigma_0^{(k)} = \sqrt{\frac{\sum_{t=1}^{T}\xi_{t,0|T}^{(k-1)}(y_t - \alpha_0^{(k)} - \beta_0^{(k)}y_{t-1})^2}{\sum_{t=1}^{T}\xi_{t,0|T}^{(k-1)}}}. \tag{22}$$

The derivations of η_{1t} are done in exactly the same manner. We only have the parameters with subscript 1, and $\xi_{t,1|T}^{(k-1)}$ will appear instead of $\xi_{t,0|T}^{(k-1)}$. We get

$$\alpha_1^{(k)} = \frac{\sum_{t=1}^{T}\xi_{t,1|T}^{(k-1)}(y_t - \beta_1^{(k)}y_{t-1})}{\sum_{t=1}^{T}\xi_{t,1|T}^{(k-1)}}, \tag{23}$$

$$\beta_1^{(k)} = \frac{\sum_{t=1}^{T}\xi_{t,1|T}^{(k-1)}y_{t-1}(y_t\nu_1^{(k-1)} - \sum_{n=1}^{T}\xi_{n,1|T}^{(k-1)}y_n)}{\sum_{t=1}^{T}\xi_{t,1|T}^{(k-1)}y_{t-1}(\nu_1^{(k-1)}y_{t-1} + \sum_{n=1}^{T}\xi_{n,1|T}^{(k-1)}y_{n-1})}, \tag{24}$$

and

$$\sigma_1^{(k)} = \sqrt{\frac{\sum_{t=1}^{T}\xi_{t,1|T}^{(k-1)}(y_t - \alpha_1^{(k)} - \beta_1^{(k)}y_{t-1})^2}{\sum_{t=1}^{T}\xi_{t,1|T}^{(k-1)}}}. \tag{25}$$

5.3.2 Estimation of the Initial Distribution Parameter λ_0

Equation 19 gives us

$$\begin{aligned}\frac{\partial \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)})}{\partial \lambda_0} &= \frac{\partial}{\partial \lambda_0}(\xi_{1,0|T}^{(k-1)}\log(\lambda_0) + \xi_{1,1|T}^{(k-1)}\log(1-\lambda_0)),\\ &= \frac{\xi_{1,0|T}^{(k-1)}}{\lambda_0} - \frac{\xi_{1,1|T}^{(k-1)}}{1-\lambda_0}.\end{aligned}$$

Evaluating at $\lambda_0 = \lambda_0^{(k)}$ and setting this equal to 0, gives us

$$\lambda_0^{(k)} = \frac{\xi_{1,0|T}^{(k-1)}}{\xi_{1,0|T}^{(k-1)} + \xi_{1,1|T}^{(k-1)}} = \xi_{1,0|T}^{(k-1)}, \tag{26}$$

where the last equality follows from the fact that S_1 can be either 0 or 1, thus $\mathbb{P}(S_1 = 0|\Omega_T, \theta^{(k-1)}) + \mathbb{P}(S_1 = 1|\Omega_T, \theta^{(k-1)}) = 1$.

5.3.3 Estimation of the Transition Probabilities p_{00} and p_{11}

We start with the transition probability p_{00}. In equation 19, we see that the second equation has two terms containing p_{00}. Therefore, we get

$$\begin{aligned}\frac{\partial \ell_{EM}(\Omega_T, \theta, \theta^{(k-1)})}{\partial p_{00}} &= \frac{\partial}{\partial p_{00}}\Big(\sum_{t=2}^{T}\tilde{p}_{00,t}^{(k-1)}\log(p_{00}) + \tilde{p}_{01,t}^{(k-1)}\log(1-p_{00})\Big),\\ &= \sum_{t=2}^{T}\frac{\tilde{p}_{00,t}^{(k-1)}}{p_{00}} - \frac{\tilde{p}_{01,t}^{(k-1)}}{1-p_{00}}.\end{aligned}$$

Evaluating at $p_{00} = p_{00}^{(k)}$ and setting the expression equal to 0 gives us

$$(1-p_{00}^{(k)})\sum_{t=2}^{T}(\tilde{p}_{00,t}^{(k-1)}) - p_{00}^{(k)}\sum_{t=2}^{T}(\tilde{p}_{01,t}^{(k-1)}) = 0.$$

Therefore, rewriting this equation gives us

$$p_{00}^{(k)} = \frac{\sum_{t=2}^{T} \tilde{p}_{00,t}^{(k-1)}}{\sum_{t=2}^{T} \tilde{p}_{00,t}^{(k-1)} + \tilde{p}_{01,t}^{(k-1)}} = \frac{\sum_{t=2}^{T} \tilde{p}_{00,t}^{(k-1)}}{\sum_{t=2}^{T} \xi_{t-1|T,0}^{(k-1)}}, \tag{27}$$

where the last equality follows from the fact that

$$\tilde{p}_{00,t}^{(k-1)} + \tilde{p}_{01,t}^{(k-1)} = \mathbb{P}(S_{t-1} = 0, S_t = 0|\Omega_T, \theta^{(k-1)}) + \mathbb{P}(S_{t-1} = 0, S_t = 1|\Omega_T, \theta^{(k-1)}) = \mathbb{P}(S_{t-1} = 0|\Omega_T, \theta^{(k-1)}) = \xi_{t-1|T,0}^{(k-1)}.$$

Similarly, we find an expression for $p_{11}^{(k)}$, namely

$$p_{11}^{(k)} = \frac{\sum_{t=2}^{T} \tilde{p}_{11,t}^{(k-1)}}{\sum_{t=2}^{T} \xi_{t-1|T,1}^{(k-1)}}. \tag{28}$$

We have now derived an estimation for all parameters of θ. In the next chapter, we will apply these estimations, and find the optimal values for given data.

6 Implementation

In this chapter, we will give an overview of our found expressions and how all these expressions are used to derive the estimated parameters and probability vectors.

First of all, it is important to remember that we use data, which means that the values of Y_t are known, i.e. Ω_T is given. Now, we will clarify the order of calculations, which we have derived in Chapter 4 and 5.

1. For starting the EM-algorithm, we have to pick $\theta^{(0)}$. The values of $\theta^{(0)}$ are not really important. If we pick a "bad" $\theta^{(0)}$, then the EM-algorithm will need more steps to converge. There is no result, which states that the likelihood function will indeed converge to a global maximum. It can also converge to a local maximum. Therefore, it is useful to consider several different starting values $\theta^{(0)}$. This will give a good indication whether we indeed have obtained the global maximum. Another alternative is to use specific algorithms such as simulated annealing to find the global maximum. We will not explore this concept.

2. Given $\theta^{(0)}$ and Ω_T , we can determine the value of $\eta_{jt}^{(0)}$ (The superscript refers to which θ we use.) for all $j = 0, 1, t = 1, ..., T$ using equation 6 for $k = 0$, i.e.

$$\eta_{jt}^{(k)} = f_{Y_t}(y_t|S_t = j, \Omega_{t-1}, \theta^{(k)}) = \frac{1}{\sqrt{2\pi}\sigma_j^{(k)}} \exp\left[\frac{-(y_t - \alpha_j^{(k)} - \beta_j^{(k)} y_{t-1})^2}{2(\sigma_j^{(k)})^2}\right], \quad \text{for } j = 0, 1, t = 1, ..., T.$$

3. Now, we can use the values of $\eta_{jt}^{(0)}$ and $\theta^{(0)}$ to determine the inference vectors. Recall Corollary 4.3.1, which allows us to find an expression for $\xi_{t,j|t}^{(k)}$ with $k = 0$.

$$\xi_{t,j|t}^{(k)} = \frac{\sum_{i=0}^{1} p_{ij}^{(k)} \xi_{t-1,i|t-1}^{(k)} \eta_{jt}^{(k)}}{\sum_{i=0}^{1}\sum_{j=0}^{1} p_{ij}^{(k)} \xi_{t-1,i|t-1}^{(k)} \eta_{jt}^{(k)}} \quad \text{for } j = 0, 1, \text{ and } t = 2, ..., T,$$

$$\xi_{1,j|1}^{(k)} = \frac{\eta_{j1}^{(k)} \lambda_j^{(k)}}{\sum_{j=0}^{1} \eta_{j1}^{(k)} \lambda_j^{(k)}} \quad \text{for } j = 0, 1.$$

First, we determine $\xi_{1,j|1}^{(k)}$. Then, this value can be used to determine $\xi_{2,j|2}^{(k)}$, and so on. It is a forward recursion.

4. After that, we determine the values of the forecast vectors. Theorem 4.4 with $k = 0$ gives us

$$\xi_{t+1,j|t}^{(k)} = \sum_{i=0}^{1} p_{ij}^{(k)} \xi_{t,i|t}^{(k)}, \quad \text{for } t = 1, .., T-1.$$

5. Using step 3 and 4, we can now determine the smoothed inferences. Recall that Theorem 4.5 shows us that for $k = 0$

$$\xi_{t,j|T}^{(k)} = \sum_{i=0}^{1} \xi_{t,j|t}^{(k)} p_{ji}^{(k)} \frac{\xi_{t+1|T,i}^{(k)}}{\xi_{t+1|t,i}^{(k)}}, \quad \text{for } t = 1, .., T-1.$$

Note that $\xi_{T,0|T}, \xi_{T,1|T}$ is found using step 3, thus we can determine $\xi_{T-1,j|T}$. Again, this can be used to determine the $\xi_{T-2,j|T}$, and so on. In other words, we note that this equation is a backwards recursion.

6. We have now enough information to determine the loglikelihood of $\theta^{(0)}$ using equation 14, which gives us

$$\ell(\Omega_T, \theta^{(k)}) = \log(\sum_{j=0}^{1} \eta_{j1}^{(k)} \lambda_j^{(k)}) + \sum_{t=2}^{T} \log(\sum_{j=0}^{1}\sum_{i=0}^{1} \eta_{jt}^{(k)} p_{ij}^{(k)} \xi_{t-1,i|t-1}^{(k)}).$$

7. Now, we can start estimating the next parameter vector in the series. Thus we can find the values of $\theta^{(1)}$ using the following formulas for $k = 1$, which we have found in chapter 5,

$$\beta_j^{(k)} = \frac{\sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)} y_{t-1}(y_t \nu_j^{(k-1)} - \sum_{n=1}^{T} \xi_{n,j|T}^{(k-1)} y_n)}{\sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)} y_{t-1}(\nu_j^{(k-1)} y_{t-1} + \sum_{n=1}^{T} \xi_{n,j|T}^{(k-1)} y_{n-1})}, \quad \text{with } \nu_j^{(k-1)} = \sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)},$$

$$\alpha_j^{(k)} = \frac{\sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)} (y_t - \beta_j^{(k)} y_{t-1})}{\sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)}},$$

$$\sigma_j^{(k)} = \sqrt{\frac{\sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)} (y_t - \alpha_j^{(k)} - \beta_j^{(k)} y_{t-1})^2}{\sum_{t=1}^{T} \xi_{t,j|T}^{(k-1)}}},$$

$$\lambda_0^{(k)} = \xi_{1,0|T}^{(k-1)},$$

$$p_{jj}^{(k)} = \frac{\sum_{t=2}^{T} \tilde{p}_{jj,t}^{(k-1)}}{\sum_{t=2}^{T} \xi_{t-1|T,j}^{(k-1)}},$$

where

$$\tilde{p}_{jj,t}^{(k-1)} = \xi_{t-1,j|t-1}^{(k-1)} p_{ji}^{(k-1)} \frac{\xi_{t,j|T}^{(k-1)}}{\xi_{t,j|t-1}^{(k-1)}}.$$

Note that the order is important, since we need $\beta^{(k)}$ and $\alpha^{(k)}$ for determining the value of $\sigma^{(k)}$. Thus we first determine $\beta^{(k)}$, then $\alpha^{(k)}$, and finally $\sigma^{(k)}$. The order of the other parameters do not matter, since there is now dependency on each other.

8. We can now go back to step 2, and increase the value of k. We continue this recursive process until the difference in the loglikelihood of two consective parameter estimates is smaller than some prescribed value ϵ, i.e. we stop after the k-th iteration if

$$\ell(\Omega_T, \theta^{(k)}) - \ell(\Omega_T, \theta^{(k-1)}) < \epsilon.$$

Recall that the loglikelihood function is an increasing function. Therefore, $\ell(\Omega_T, \theta^{(k)}) - \ell(\Omega_T, \theta^{(k-1)})$ is always positive.

This method gives us, theoretically, a nice and easy way to find the optimal parameter vector for θ. However, there are some problems with the implementation of this procedure. We have to deal with numerical errors due to the precision of the calculations. Python has a standard precision of 18 places for a float type. Therefore, we do not have exact values for the parameter vector and the probability vectors. Since, we have to do many calculations, these small errors can lead to a significant error. We consider the following table:

Iteration	17	18	19	20	21
Loglikelihood	-39.60850	-39.60836	-39.60840	-39.60846	-39.60850

Figure 1: Table of the loglikelihood at iteration 17-21.

First, note that this table differs if we use different values for $\theta^{(0)}$. For this table, we have used $\theta^{(0)}$ as introduced in the next chapter. The point is that in this case that the loglikelihood decreases at the 19th iteration, and afterwards. This contradicts the obtained result of section 5.2, which shows that the loglikelihood function is an increasing function for increasing iterations. However, due to the numerical errors, we see that this is not the case. Therefore, step 8 of the described process can not be used, which implies that we need a different stopping condition. An alternative is to stop whenever the absolute value of the difference in the loglikelihood of two consecutive parameter estimates is smaller than ϵ, i.e.

$$|\ell(\Omega_T, \theta^{(k)}) - \ell(\Omega_T, \theta^{(k-1)})| < \epsilon.$$

However, we do need to be aware that the given parameter vector $\theta^{(k)}$ at the last iteration is not necessarily the optimal solution. Therefore, we do need to find the maximum value of the loglikehood function.

We can implement this using Python. The code of the implementation can be found in Appendix A. In the next chapter, we will see how this implementation can be applied to a data set on GDP growth.

7 Application on the GDP Growth of the Netherlands

In this chapter, we are going to apply the derived theory on a data set, provided by The World Bank, about the GDP growth rate of the Netherlands from 2001 till 2021. The data can be seen in the following table:

Year	Observation	Annual GDP growth in percentages
2000	y0	4,195642498
2001	y1	2,326955087
2002	y2	0,217273595
2003	y3	0,155645898
2004	y4	1,984945714
2005	y5	2,050876108
2006	y6	3,460988954
2007	y7	3,772842521
2008	y8	2,170324851
2009	y9	-3,666883937
2010	y10	1,342739336
2011	y11	1,551189312
2012	y12	-1,030353991
2013	y13	-0,130175288
2014	y14	1,423395395
2015	y15	1,959169721
2016	y16	2,191713719
2017	y17	2,910902513
2018	y18	2,360915095
2019	y19	1,955588416
2020	y20	-3,798635993
2021	y21	5,035902024

Figure 2: Table of the annual GDP growth rate of the Netherlands.

We assume that the GDP growth rate of the Netherlands is a process, which follows our model introduced in section 3.1. In other words, we assume that the annual GDP growth rate is a process that follows a first-order autoregression in which we allow regime-switching. The underlying process S_t is based on the general state of the economy, i.e. we let the underlying process describe the business cycles. This assumption is based on our expectations that the parameters of a first-order autoregression in a "good" and "rising" economy are different from the parameters in a economic recession, due to different consumer behaviour and also different political behaviour in these different states. Thus, in short, the annual GDP growth rates are our observations, and therefore denoted by y_t. Then, the latent process S_t describes the behaviour of the economy, let $S_t = 0$ denote that at time t the economy is rising, and $S_t = 1$ means that at time t the economy is shrinking.

We can now use the implementation procedure described by the previous chapter. We discussed the consequences of numerical errors. Therefore, in our tables, we will only show the results until the likelihood function reaches the maximum. As noted in the previous chapter, we have to start with $\theta^{(0)}$. The importance of the choice of the initial parameter vector has been discussed as well. Note that taking "extreme" starting values will also result in problems, because of the possibility of too small numbers such that we have to divide by 0 in the EM-algorithm. We start with

$$\begin{array}{llllll} \alpha_0 = 2 & \beta_0 = 1 & \sigma_0 = 0.5 & p_{00} = 0.9 & p_{11} = 0.7 & \lambda_0 = 0.5 \\ \alpha_1 = -0.5 & \beta_1 = 0.7 & \sigma_1 = 1 & p_{01} = 0.1 & p_{10} = 0.3 & \lambda_1 = 0.5. \end{array}$$

Based on economic expectations, the chosen parameters can be explained. First of all, in a recession we would expect more volatility. That is why we have chosen $\sigma_1 > \sigma_0$. Secondly, a business cycle tends to stay in the same state as a previous time period, thus $p_{00} > p_{01}$ and $p_{11} > p_{10}$.

We can now run the EM-algorithm. The following table provides us with the estimated parameters at each iteration. As we have seen in the previous chapter, the maximum is obtained after the 18th iteration. Therefore, we will only give the parameter estimates up to the 18th iteration.

Iteration	alpha0	alpha1	beta0	beta1	sigma0	sigma1	p00	p11	lambda0	Loglikelihood
0	2,00000	-0,50000	1,00000	0,70000	0,50000	1,00000	0,90000	0,70000	0,50000	-107,39111
1	1,55455	1,42172	0,89208	-0,12156	0,35699	2,16657	0,12996	0,90023	0,00000	-45,48557
2	1,54397	1,42124	0,80728	-0,12557	0,36392	2,14296	0,06870	0,91538	0,00000	-45,07706
3	1,59756	1,40748	0,69266	-0,13603	0,36117	2,13543	0,08109	0,91172	0,00000	-44,57511
4	1,61724	1,38275	0,63348	-0,15472	0,33302	2,13306	0,11743	0,89328	0,00000	-44,27605
5	1,60259	1,35955	0,62024	-0,16667	0,32591	2,14040	0,14829	0,87216	0,00000	-44,14581
6	1,57546	1,34077	0,61866	-0,17355	0,33811	2,15065	0,18091	0,85583	0,00000	-44,03045
7	1,53251	1,32083	0,62027	-0,17908	0,36830	2,16348	0,23632	0,84431	0,00000	-43,76060
8	1,45391	1,28857	0,62608	-0,18783	0,41691	2,18651	0,34235	0,83416	0,00000	-43,00859
9	1,33009	1,22501	0,63923	-0,21043	0,47294	2,23481	0,50605	0,81968	0,00000	-41,62202
10	1,21536	1,12472	0,64688	-0,25789	0,52197	2,30847	0,65341	0,79785	0,00000	-40,43592
11	1,16105	1,01982	0,63744	-0,30826	0,56122	2,36742	0,72719	0,77610	0,00000	-39,93395
12	1,15295	0,94081	0,61754	-0,34467	0,59168	2,39399	0,75515	0,75902	0,00000	-39,74390
13	1,16344	0,88263	0,59591	-0,37171	0,61230	2,40094	0,76641	0,74620	0,00000	-39,65954
14	1,17664	0,83977	0,57890	-0,39142	0,62458	2,39992	0,77162	0,73704	0,00000	-39,62493
15	1,18732	0,81088	0,56763	-0,40437	0,63139	2,39702	0,77416	0,73101	0,00000	-39,61294
16	1,19483	0,79308	0,56069	-0,41211	0,63506	2,39453	0,77541	0,72733	0,00000	-39,60938
17	1,19979	0,78272	0,55655	-0,41650	0,63703	2,39285	0,77601	0,72521	0,00000	-39,60850
18	1,20294	0,77685	0,55411	-0,41894	0,63809	2,39179	0,77630	0,72403	0,00000	-39,60836

Figure 3: Table of the parameter estimates for each iteration.

Therefore, we have that the annual GDP growth rate based on the data is most likely to satisfy the following equation

$$GDP_t = \begin{cases} 1,20294 + 0,55411 \cdot GDP_{t-1} + 0,63809 \cdot \epsilon & \text{if } S_t = 0 \\ 0,77685 - 0,41894 \cdot GDP_{t-1} + 2,39179 \cdot \epsilon & \text{if } S_t = 1 \end{cases}, \tag{29}$$

where S_t is a Markov chain with initial distribution $\lambda = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and transition matrix

$$P = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix} = \begin{pmatrix} 0,77630 & 0,22370 \\ 0,27597 & 0,72403 \end{pmatrix}.$$

We conclude that $0,77630 = p_{00} > p_{01} = 0,22370$, and $0,72403 = p_{11} > p_{10} = 0,27597$, which is indeed in line with our expectations that the economy tends to stay in the same state as the year before. We also notice that $p_{00} > p_{11}$. This is due to the fact that the economy tends to stay longer in the same state in a rising economy than during recessions. Note that λ_0 is equal to 0 after one iteration. Therefore, we can conclude that the latent process is in state 1 at $t = 1$ with probability 1. In other words, in 2001 the economy has probability 1 to be in a shrinking state. Furthermore, we see that $\sigma_0 < \sigma_1$ as expected, since we would expect more volatility in a recession than in a rising economy. At last, we see that $\alpha_0 + \beta_0 \cdot Y_{t-1} > \alpha_1 + \beta_1 \cdot Y_{t-1}$ for $Y_{t-1} > -0,4379$. Thus, for most values of Y_{t-1}, the expected value of Y_t is higher in a rising economy, which we would expect since a rising economy tends to have higher annual GDP growth rates. In our data, there are only three values smaller than $-0,4379$. However, it is still questionable if for these values we would indeed expect that the expected value of Y_t is higher in a bad economy than a rising economy.

Now we consider the derived probability vectors at the 18th iteration. We only look at the probabilities on state 0 for $t = 1, ..., 21$. We get the following values:

Probabilities on State 0 given the Estimates of the Parameters			
Time	Inference Probability for S_t=0	Forecast Probability for S_t=0	Smoothed Inference Probability for S_t=0
1	6,40E-79		2,03E-79
2	0,00251	0,27597	0,00301
3	0,21640	0,27723	0,35610
4	0,59796	0,38424	0,77382
5	0,87367	0,57515	0,93240
6	0,85621	0,71309	0,93070
7	0,96754	0,70436	0,95616
8	0,84530	0,76006	0,62801
9	5,60E-19	0,69890	1,80E-19
10	0,00471	0,27597	0,00498
11	0,58235	0,27833	0,30113
12	4,37E-05	0,56734	8,40E-05
13	0,45020	0,27599	0,64375
14	0,77759	0,50122	0,89193
15	0,90740	0,66503	0,95940
16	0,93941	0,72997	0,97376
17	0,94852	0,74599	0,97455
18	0,94561	0,75055	0,95000
19	0,92033	0,74909	0,78113
20	6,41E-19	0,73644	1,98E-19
21	4,18E-19	0,27597	4,18E-19

Figure 4: Table of the inference, forecast and smoothed inference probabilities on $S_t = 0$.

Note that the forecast probabilities are not close to 0 or 1. Therefore, we cannot really be sure about which state will occur in the next time period. In contrast, the smoothed inference vector often has values close to 0 or 1, which means that we have a good indication whether it was likely for the latent process to be in state 0 or 1 at the corresponding time. Plotting these smoothed inferences in a graph gives us the following:

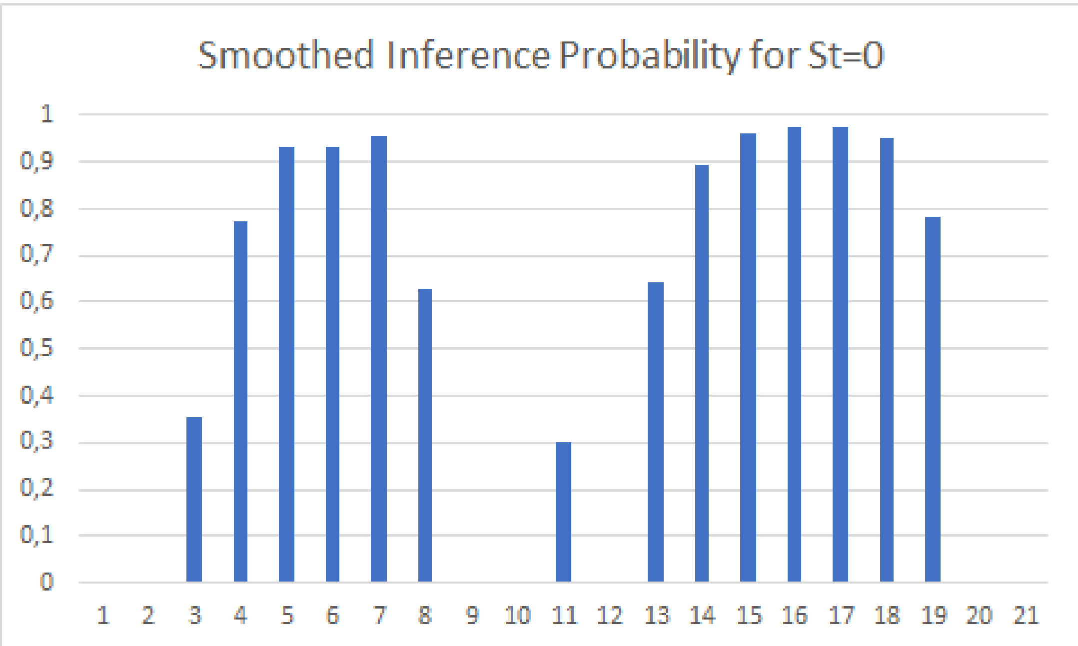

Figure 5: Graph of the smoothed inference probabilities at time $t = 1, ..., 21$.

Based on this graph, we can conclude that it was almost certain that at time $t = 1, 2, 9, 10, 12, 20, 21$ the latent process was in state 1. In other words, in the years 2001,2002,2009,2010,2012,2020,2021 the economy was in a recession. We note that this is not exactly in line with our expectations. Recall that the Netherlands had seen several recession in the 21st century, namely the early 2000s recession (2000-2001), the global financial crises of 2007-2008 and also the recession as a consequence of the COVID-19 pandemic in 2020-2021. However, the annual GDP growth is considered as a lagging indicator of the business cycles, which means that the growth GDP growth rate does not change directly when the business cycle changes, but it changes afterwards. We refer to [Shi19] for further discussion on different types of indicators. Therefore, we can conclude that our graph indeed is able to approximately determine the recessions. However, the model fails to exactly determine the correct years.

8 Conclusion

In this thesis, we have explored the properties of a Markov regime-switching model with two regimes. First, we have discussed the key concepts of Markov switching models, namely Markov chains, first-order autoregressions and maximum likelihood concepts. The last concept is used to estimated the parameters of the Markov regime-switching model. Furthermore, we have introduced some useful notation and derived expressions for some conditional densities and probabilities, which were needed for determining the likelihood functions. Afterwards we have introduced the Expectation-maximization algorithm. This algorithm was used to derive expressions for determining the most likely value of the parameters. Afterwards, we used these expression to implement the algorithm in Python. Finally, we have applied the theory on the annual GDP growth rate of the Netherlands during the period 2000-2021.

We have seen that our model roughly estimate the recessions (early 2000s recession, global financial crises and COVID-19 recession). However it fails to exactly determine the correct years. An alternative is to use composite indicators, which are more used recently. These indicators are composed of several observable variables, with different types. It can consist of leading, lagging and coincident indicators. A leading indicator points towards future changes. Coincident indicators are indicators, which happen in real time, i.e. a change of this indicator means that the business cycle has changed at that exact moment. Even though, composite indicators will be able to be more accurate in terms of estimating the states of the latent process, there is also a downside. We are not able to interpret the estimated parameters corresponding to a model describing the changes of a composite indicator, since composite indicators have no unit. Our goal was to describe a process and understand the estimated parameters and also give an estimation on whether the latent process is in a specific state. The estimated parameters of our proposed model on the annual GDP growth rate of the Netherlands can be interpreted. We have found that the annual GDP growth rate of the Netherlands follows the following specification:

$$GDP_t = \begin{cases} 1,20294 + 0,55411 \cdot GDP_{t-1} + 0,63809 \cdot \epsilon & \text{if } S_t = 0 \\ 0,77685 - 0,41894 \cdot GDP_{t-1} + 2,39179 \cdot \epsilon & \text{if } S_t = 1 \end{cases}, \tag{29}$$

where S_t is a Markov chain with initial distribution $\lambda = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and transition matrix

$$P = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix} = \begin{pmatrix} 0,77630 & 0,22370 \\ 0,27597 & 0,72403 \end{pmatrix}.$$

and $\epsilon \sim N(0, 1)$.

References

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977. With discussion.

[Ham90] James D. Hamilton. Analysis of time series subject to changes in regime. *J. Econometrics*, 45(1-2):39–70, 1990.

[Ham94] James D. Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

[Ham16] James D. Hamilton. *Regime Switching Models*, pages 1–7. Palgrave Macmillan UK, 2016.

[Kol19] E. Kole. Markov switching models: An example for a stock market index. 2019.

[Nor09] J.R. Norris. *Markov Chains.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 15 edition, 2009.

[Ric07] J.A. Rice. *Mathematical Statistics and Data Analysis.* Brooks/Cole, Cengage Learning, 3 edition, 2007.

[Rut89] R.A. Rutenbar. Simulated annealing algorithms: An overview, 1989.

[Shi19] B. Shipley. Understanding leading and lagging indicators. https://arnerichmassena.com/blog-assets/Understanding-Leading-and-Lagging-Indicators$_jan - 2019.pdf$, 2019.

[SW21] Y. Song and T. Woźniak. *Markov Switching.* Oxford University Press, 03 2021.

[Tab21] M. Taboga. Em algorithm. https://www.statlect.com/fundamentals-of-statistics/EM-algorithm, 2021.

A Python Code

This appendix gives the code of the implementation of the EM-algorithm:

```
import pandas as pd
import math


df =pd.read_excel ("data2.xls")
print (df)
print("-----------------------------")
print(df.keys())
print("-----------------------------")
print(df['Country Name'])
print("-----------------------------")
print(df.index)
print("-----------------------------")
print(df.at[176, 'Country Name'])
print("-----------------------------")

nld=[]
#array with annual growth of GDP w.r.t. the previous year (value of 2000 is y0).
year=2000
T=21
for i in range(T+1):
    nld.append(df.at[176,str(year)])
    year+=1

print(nld)
print(len(nld))
print("-----------------------------")

#Give start values of theta
alpha=[2,-0.5]
beta=[1,0.7]
sigma=[0.5,1]
p00= 0.9
p01=1-p00
p11=0.7
p10=1-p11
initial_lambda=[0.5, 0.5]

#Construct an array in which we will give the value of the likelihood function ...
(equation 14) at each iteration
likelihood= []
hulp=[]
difference=10
epsilon=1e-6


iterationcounter =0
#while iterationcounter<19: #this stopping condition can be useful to make tables
while abs(difference) > epsilon:
    #Construct an array with eta vector
    eta=[]
    for i in range(1,T+1):
        state0=1/(math.sqrt(2*math.pi)*sigma[0])*math.exp((-1*(nld[i]-alpha[0]...
        -beta[0]*nld[i-1])**2)/(2*sigma[0]**2))
        state1=1/(math.sqrt(2*math.pi)*sigma[1])*math.exp((-1*(nld[i]-alpha[1]...
```

```
    -beta[1]*nld[i-1])**2)/(2*sigma[1]**2))
    eta.append([state0,state1])

print("List with eta vectors:")
print(eta) #Recall that inference at time 1 has index 0, similarly inference ...
at time t, can be found using t-1
print("------------------------------")


#Construct an array with the inference vector xi_{t|t}
inference=[]
denom =  eta[0][0]*initial_lambda[0]+ eta[0][1]*initial_lambda[1]
inf0= (eta[0][0]*initial_lambda[0])/denom
inference.append([inf0,1-inf0])
for i in range(2,T+1):
    denom= p00*inference[i-2][0]*eta[i-1][0] + p10*inference[i-2][1]*eta[i-1][0] ...
    + p01*inference[i-2][0]*eta[i-1][1] + p11*inference[i-2][1]*eta[i-1][1]
    inf0 = (p00*inference[i-2][0]*eta[i-1][0]...
    + p10*inference[i-2][1]*eta[i-1][0])/denom
    inference.append([inf0,1-inf0])

print("List with inference vectors at iteration " + str(iterationcounter))
print(inference)
print("------------------------------")


#Construct an array with the forecast vectors xi_{t+1|t}
#IMPORTANT NOTE, the length of this list is 20 (not 21)
#We start with xi_{2|1}
forecast=[]
for i in range(1, T):
    fore0=p00*inference[i-1][0]+p10*inference[i-1][1]
    forecast.append([fore0,1-fore0])

print("List with forecast vectors at iteration " + str(iterationcounter))
print(forecast)
print("------------------------------")


#Construct an array with the smoothed inference vectors xi_{t|T=21}
sm_inf=[]
for i in range(T):
    sm_inf.append([0,0]) #backwards recursion, therefore we first create 21 arrays
sm_inf[T-1]=inference[T-1]
for i in range(2,T+1):
    sm_inf[-i][0]=((inference[-i][0]*p00*sm_inf[-i+1][0])/(forecast[-i+1][0]))+...
    ((inference[-i][0]*p01*sm_inf[-i+1][1])/(forecast[-i+1][1]))
    sm_inf[-i][1]=1-sm_inf[-i][0]

print("List with smoothed inference vectors at iteration " + str(iterationcounter))
print(sm_inf)
print("------------------------------")

#Determine the loglikelihood
value=0
value+= math.log(eta[0][0]*initial_lambda[0]+eta[0][1]*initial_lambda[1])
for i in range(1,T):
    value+=math.log(eta[i][0]*p00*inference[i-1][0]+eta[i][0]*p10*inference[i-1][1]
```

```
    +eta[i][1]*p01*inference[i-1][0]+eta[i][1]*p11*inference[i-1][1])
likelihood.append(value)


print("The list of values of the likelihood function is given by:")
print(likelihood)
print(len(likelihood))
print("———————————————")


#Check the difference
if iterationcounter == 0:
    difference= 10 #random value greater then epsilon
else:
    difference= likelihood[iterationcounter]-likelihood[iterationcounter -1]

#print("The difference is given by " + str(difference))
#print("———————————————")

iterationcounter+=1
#Determine new values of theta using the formulas we have found

#But first we determine the useful expression nu
nu0=0
nu1=0
for i in range(len(sm_inf)):
    nu0+=sm_inf[i][0]
    nu1+=sm_inf[i][1]

#starting with beta

#crossterms of smoothed inference with present values of gdp
sm_val0=0
sm_val1=1
sm_prev_val0=0
sm_prev_val1=0
for i in range(T):
    sm_val0+=sm_inf[i][0]*nld[i+1]
    sm_val1+=sm_inf[i][1]*nld[i+1]
    sm_prev_val0+= sm_inf[i][0]*nld[i]
    sm_prev_val1+= sm_inf[i][1]*nld[i]

num0=0
num1=0
denom0=0
denom1=0
for i in range(T):
    num0+=sm_inf[i][0]*nld[i]*(nld[i+1]*nu0-sm_val0)
    num1+=sm_inf[i][1]*nld[i]*(nld[i+1]*nu1-sm_val1)
    denom0+=sm_inf[i][0]*nld[i]*(nld[i]*nu0-sm_prev_val0)
    denom1+=sm_inf[i][1]*nld[i]*(nld[i]*nu1-sm_prev_val1)
beta[0]=num0/denom0
beta[1]=num1/denom1

print("New beta at iteration " + str(iterationcounter) + " are given by")
print(beta)
print("———————————————")
```

```
#now alpha
num0=0
num1=0
for i in range(T):
    num0+=sm_inf[i][0]*(nld[i+1]-beta[0]*nld[i])
    num1+=sm_inf[i][1]*(nld[i+1]-beta[1]*nld[i])
alpha[0]=num0/nu0
alpha[1]=num1/nu1

print("New alpha at iteration " + str(iterationcounter) + " are given by")
print(alpha)
print("---------------------------------")



#sigma
sum0=0
sum1=0
for i in range(T):
    sum0+=sm_inf[i][0]*(nld[i+1]-alpha[0] - beta[0]*nld[i])**2
    sum1+=sm_inf[i][1]*(nld[i+1]-alpha[1] - beta[1]*nld[i])**2
sigma[0]=math.sqrt(sum0/nu0)
sigma[1]=math.sqrt(sum1/nu1)

print("New sigma at iteration " + str(iterationcounter) + " are given by")
print(sigma)
print("---------------------------------")


#initial lambda
initial_lambda[0]=sm_inf[0][0]
initial_lambda[1]=sm_inf[0][1]

print("New Initial_lambda at iteration " + str(iterationcounter) + " are given by")
print(initial_lambda)
print("---------------------------------")



#transition probabilities
sum0=0
sum1=0
for i in range(2,T+1):
    sum0+=(inference[i-2][0]*p00*sm_inf[i-1][0]) /forecast[i-2][0]
    sum1+=(inference[i-2][1]*p11*sm_inf[i-1][1])/forecast[i-2][1]
p00=sum0/(nu0-sm_inf[0][0])
p11=sum1/(nu1-sm_inf[0][1])
p01=1-p00
p10=1-p11


print("New transition probabilities at iteration " + str(iterationcounter)...
+ " are given by")
print("p00: " +str(p00))
print("p01: " +str(p01))
print("p10: " +str(p10))
print("p11: " +str(p11))
print("---------------------------------")
```

```
print(max(likelihood))
print(likelihood[-1])
```