

UNIVERZITA PALACKÉHO V OLMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Skryté Markovovy řetězce - spojitý případ



Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **Mgr. Kamila Fačevicová, Ph.D.**
Vypracovala: **Bc. Michaela Křišťanová**
Studijní program: N1103 Aplikovaná matematika
Studijní obor: Aplikace matematiky v ekonomii
Forma studia: prezenční
Rok odevzdání: 2020

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Michaela Křišťanová

Název práce: Skryté Markovovy řetězce - spojitý případ

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: Mgr. Kamila Fačevicová, Ph.D.

Rok obhajoby práce: 2020

Abstrakt: Skryté Markovovy řetězce představují speciální případ Markovových řetězců, kde jednotlivé stavy řetězce nejsou přímo viditelné. V práci se zaměříme na spojitý případ, kdy pozorování odpovídá realizaci spojitě náhodné veličiny. Cílem diplomové práce bude seznámení s teorií spojitých skrytých Markovových řetězců a s jejich aplikací.

Klíčová slova: Skryté Markovovy řetězce, Markovovy řetězce, spojitá náhodná veličina, lineární regrese, Viterbiho algoritmus, Baum-Welchův algoritmus

Počet stran: 65

Počet příloh: 1 CD

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Michaela Křišťanová

Title: Hidden Markov chains - a continuous case

Type of thesis: Master's

Department: Department of Mathematical Analysis and Applications of Mathematics

Supervisor: Mgr. Kamila Fačevicová, Ph.D.

The year of presentation: 2020

Abstract: Hidden Markov chains represent a special case of Markov chains where the individual states are not directly observable. In the thesis, we focus on a continuous case when we observe realizations of a continuous random variable. The aim of the thesis is to gain knowledge about the theory of hidden Markov chains and their application.

Key words: Hidden Markov chains, Markov chains, continuous random variable, linear regression, Viterbi algorithm, Baum-Welch algorithm

Number of pages: 65

Number of appendices: 1 CD

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením paní Mgr. Kamily Fačevicové, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	7
1 Náhodný proces, Markovovy řetězce	8
1.1 Náhodný proces	8
1.2 Markovovy řetězce	9
2 Skryté Markovovy řetězce	13
2.1 Diskrétní a spojité skryté Markovovy řetězce	14
2.2 Tři základní úlohy skrytých Markovových řetězců	16
2.2.1 Úloha ohodnocení	16
2.2.2 Úloha určení optimální posloupnosti stavů	21
2.2.3 Úloha učení	24
2.3 Skryté Markovovy regresní modely	33
3 Aplikace skrytých Markovových řetězců	37
3.1 HMM - Počet nově sjednaných hypotéčních úvěrů	38
3.2 HMM - Míra nezaměstnanosti	46
3.3 HMMR - Vztah nezaměstnanosti a počtu nově sjednaných hypotéčních úvěrů	52
3.4 Shrnutí a porovnání výsledků s tempem růstu HDP	58
Závěr	63
Literatura	64

Poděkování

Na tomto místě bych ráda poděkovala vedoucí mé diplomové práce, paní Mgr. Kamile Fačevicové, Ph.D., za pomoc, ochotu a především čas strávený odborným vedením mé diplomové práce.

Úvod

Tématem diplomové práce je představit teorii skrytých Markovových řetězců, které jsou rozšířením teorie Markovových řetězců. Markovovy řetězce se zabývají chováním náhodného procesu v čase, kdy předpokládáme, že v jeho vývoji v čase je jistá neurčitost. V teorii skrytých Markovových řetězců předpokládáme, že realizace tohoto náhodného procesu nejsou přímo pozorovatelné a můžeme je určit prostřednictvím jiné náhodné veličiny. V práci se zaměříme na případ spojitých skrytých Markovových řetězců, tedy situaci, kdy je předmětem našeho pozorování spojitá náhodná veličina.

Cílem diplomové práce je nastudovat, popsat a v neposlední řadě aplikovat teorii skrytých Markovových řetězců.

První kapitola diplomové práce představuje seznámení s Markovovými řetězci a dává čtenáři nezbytný základ pro porozumění další teorie, která je předmětem práce.

Druhá kapitola, která se týká samotných skrytých Markovových řetězců, je rozdělena do tří částí. V první podkapitole je čtenář seznámen se základními pojmy a rozdíly oproti dříve popsané teorii. Druhá část se týká tří úloh, které jsou ve spojitosti se skrytými Markovovými řetězci řešeny. Závěrečnou třetí podkapitolou je rozšíření teorie o případ regresní analýzy, tedy na skryté Markovovy regresní modely. V průběhu textu jsou jednotlivé problémy pro názornost aplikovány na jednoduchém příkladu a doplněny příslušnými funkcemi ve statistickém softwaru R.

Třetí (a poslední) kapitola je věnována aplikaci teoretických poznatků na reálném datovém souboru.

Kapitola 1

Náhodný proces, Markovovy řetězce

V této kapitole se seznámíme se základy Markovových řetězců, které jsou nezbytné k porozumění problematice skrytých Markovových řetězců. Základy teorie markovských řetězců položil ruský matematik A. A. Markov (1856-1922), který své poznatky aplikoval například na románu A. S. Puškina Evžen Oněgin, na kterém provedl statistický výzkum výskytu souhlásek a samohlásek. Tato teorie se zabývá chováním náhodného procesu v čase, kdy předpokládáme, že v jeho vývoji v čase je jistá neurčitost, přičemž vývoj procesu v čase budoucím je ovlivněn pouze současností nikoli minulostí. Tato kapitola je zpracována s použitím zdrojů [8], [13], [15], [17].

1.1. Náhodný proces

Jak již bylo zmíněno výše, Markovův řetězec je speciálním případem náhodného procesu. Definujme si tedy nejprve náhodný proces.

Definice. *Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní prostor, nechť $\mathbb{T} \subset \mathbb{R}$. Rodina reálných náhodných veličin $\{Q_t, t \in \mathbb{T}\}$ definovaných na (Ω, \mathcal{A}, P) se nazývá náhodný proces.*

Náhodné procesy můžeme dělit dle času na:

- proces s diskretním časem (pokud $\mathbb{T} = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ nebo $\mathbb{T} = \mathbb{N}_0 = \{0, 1, \dots\}$);
- proces se spojitým časem (pokud $\mathbb{T} = \langle a, b \rangle$, kde $-\infty \leq a < b \leq \infty$).

Stejně tak je můžeme dělit dle hodnot náhodných veličin Q_t , které značí stav, ve kterém se systém nachází v čase t , na:

- proces s diskretními stavy (náhodné veličiny Q_t nabývají diskretních hodnot);
- proces se spojitými stavy (náhodné veličiny Q_t nabývají hodnot z nějakého intervalu).

Dále se v práci budeme zabývat pouze procesy s diskretními stavy, které nazýváme náhodné posloupnosti.

1.2. Markovovy řetězce

Uvažujme posloupnost náhodných veličin $\{Q_t, t \in \mathbb{N}_0\}$ definovanou na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Množinou stavů S rozumíme množinu celých čísel i , pro která platí, že $i \in S$ právě tehdy, když existuje $t \in \mathbb{N}_0$ tak, že $P(Q_t = i) > 0$. Prvky takovéto množiny budeme nazývat stavy. V práci budeme předpokládat, že $S = \{1, \dots, N\}$, kde N je počet stavů.

Definice. *Posloupnost diskretních náhodných veličin $\{Q_t, t \in \mathbb{N}_0\}$ se nazývá Markovův řetězec s diskretním časem a množinou stavů S , jestliže je splněna markovská vlastnost, tj. jestliže*

$$P(Q_{t+1} = j | Q_t = i, Q_{t-1} = i_{t-1}, \dots, Q_0 = i_0) = P(Q_{t+1} = j | Q_t = i)$$

pro každé $t = 0, 1, \dots$ a každé $i_0, i_1, \dots, i_{t-1}, i, j \in S$ taková, že $P(Q_t = i, Q_{t-1} = i_{t-1}, \dots, Q_0 = i_0) > 0$.

Markovská vlastnost nám říká, že vývoj náhodného procesu do budoucnosti je dán pouze stavem současným a nezávisí na tom, jak se do tohoto stavu proces

dostal. Říkáme také, že se jedná o systém bez paměti, jelikož „zapomíná“ stavy předešlé.

Markovovy řetězce jsou (spolu s množinou stavů) určeny také počátečním rozdělením pravděpodobnosti a pravděpodobnostmi přechodu. Počáteční rozdělení pravděpodobnosti určuje pravděpodobnosti jednotlivých stavů v čase 0 a je dáno vektorem $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_N)$, jehož složky jsou definovány jako

$$\pi_i = P(Q_0 = i) \text{ pro každé } i \in S.$$

Současně také platí vztahy

$$0 \leq \pi_i \leq 1, \quad i \in S, \quad \sum_{i \in S} \pi_i = 1.$$

Tyto vztahy udávají, že vektor $\boldsymbol{\pi}$ můžeme nazvat pravděpodobnostním vektorem. Pravděpodobností přechodu ze stavu i v čase t do stavu j v čase $t + 1$ nazýváme podmíněnou pravděpodobnost

$$a_{ij}(t, t + 1) = P(Q_{t+1} = j | Q_t = i), \quad i, j \in S, t \in \mathbb{N}_0.$$

Definice. *Markovův řetězec s diskrétním časem, pro který platí*

$$a_{ij}(t, t + 1) = a_{ij}, \text{ pro všechna } i, j \in S, t \in \mathbb{N}_0$$

tj. pravděpodobnost přechodu nezávisí na t , se nazývá homogenní Markovův řetězec.

Dále se budeme zabývat pouze homogenními řetězci. Pro takovéto řetězce platí

$$0 \leq a_{ij} \leq 1, \quad i, j \in S, \quad \sum_{j \in S} a_{ij} = 1, \quad i \in S.$$

Pravděpodobnosti přechodu můžeme uspořádat do tzv. matice pravděpodobností přechodu

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ a_{31} & a_{32} & \cdots & a_{3N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

Tato matice je stochastická. To znamená, že se jedná o čtvercovou matici o rozměru odpovídajícímu počtu stavů v množině S a splňující

$$a_{ij} \geq 0, \sum_{j \in S} a_{ij} = 1, i, j \in S,$$

tj. součet každého řádku je roven jedné.

Pomocí matice pravděpodobností přechodu se však posuneme pouze o jeden krok. Mohou nás však dále zajímat pravděpodobnosti přechodu po n krocích.

Definice. *Nechť n je libovolné celé nezáporné číslo. Podmíněnou pravděpodobnost*

$$a_{ij}^{(n)} = P(Q_{t+n} = j | Q_t = i), i, j \in S, t \in \mathbb{N}_0$$

nazýváme pravděpodobností přechodu ze stavu i do stavu j za n kroků.

Pro takovéto pravděpodobnosti platí vztahy

$$0 \leq a_{ij}^{(n)} \leq 1, \sum_{j \in S} a_{ij}^{(n)} = 1, i, j \in S, n \in \mathbb{N}_0.$$

Jelikož $n \in \mathbb{N}_0$, pro kompletnost dodefinujeme

$$a_{ij}^0 = \begin{cases} 1 & \text{pro } i = j, \\ 0 & \text{pro } i \neq j. \end{cases}$$

Nyní lze pomocí zavedených pravděpodobností přechodu za n kroků vyjádřit tzv. *Chapman-Kolmogorovovy rovnosti*, které vypadají následovně

$$\begin{aligned} a_{ij}^{(n+1)} &= \sum_{s \in S} a_{is}^{(n)} a_{sj}, \\ a_{ij}^{(n+k)} &= \sum_{s \in S} a_{is}^{(n)} a_{sj}^{(k)}, \\ a_{ij}^{(n_+)} &= \sum_{s \in S} a_{is}^{(r)} a_{sj}^{(n_+ - r)}, \end{aligned}$$

kde $n, k \in \mathbb{N}_0$, $i, j \in S$ a $n_+, r \in \mathbb{N}$, pro která platí vztah $1 \leq r \leq n_+ - 1$.

Stejně jako pravděpodobnosti přechodu můžeme pravděpodobnosti přechodu za n kroků uspořádat do stochastické matice, kterou značíme

$$\mathbf{A}_n = (a_{ij}^{(n)})_{i, j \in S}.$$

Jelikož $\mathbf{A}_1 = \mathbf{A}$ a $\mathbf{A}_n = \mathbf{A} \cdot \mathbf{A}_{n-1}$, lze matici pravděpodobností přechodu za n kroků vyjádřit jako

$$\mathbf{A}_n = \mathbf{A} \cdot \mathbf{A}_{n-1} = \mathbf{A}^2 \cdot \mathbf{A}_{n-2} = \dots = \mathbf{A}^{n-1} \cdot \mathbf{A}_{n-(n-1)} = \mathbf{A}^n,$$

kde \mathbf{A}^n značí n -tou mocninu matice pravděpodobností přechodu.

V kapitole jsme se již seznámili s pravděpodobnostmi přechodu a pravděpodobnostmi přechodu za n kroků. Na závěr kapitoly Markovových řetězců nám zbývá určit pravděpodobnosti „bytí v určitém stavu“, tedy absolutní pravděpodobnosti. Takovéto pravděpodobnosti nám popíše následující definice.

Definice. *Absolutními pravděpodobnostmi jednotlivých stavů v čase t nazýváme pravděpodobnosti toho, že se systém v čase t nachází ve stavu i , tj. $a_i^{(t)} = P(Q_t = i)$, $i \in S, t \in \mathbb{N}_0$.*

Pro takovéto pravděpodobnosti opět platí

$$0 \leq a_i^{(t)} \leq 1, \sum_{i \in S} a_i^{(t)} = 1, i \in S, t \in \mathbb{N}_0.$$

Pro vyjádření absolutních pravděpodobností dále platí vztahy

$$\begin{aligned} a_j^{(t)} &= \sum_{s \in S} \pi_s a_{sj}^{(t)}, \\ a_j^{(t+1)} &= \sum_{s \in S} a_s^{(t)} a_{sj}, \end{aligned}$$

kde $t \in \mathbb{N}$, $j \in S$. Absolutní pravděpodobnosti můžeme zapsat jako pravděpodobnostní vektor $\mathbf{a}^{(t)} = (a_0^{(t)}, a_1^{(t)}, a_2^{(t)}, \dots, a_N^{(t)})$. Jestliže známe počáteční pravděpodobnosti a pravděpodobnosti přechodu za n kroků, lze pomocí nich tento vektor absolutních pravděpodobností vyjádřit následujícím způsobem

$$\begin{aligned} \mathbf{a}^{(t)} &= \boldsymbol{\pi} \cdot \mathbf{A}^t, \\ \mathbf{a}^{(t+1)} &= \mathbf{a}^{(t)} \cdot \mathbf{A}. \end{aligned}$$

Kapitola 2

Skryté Markovovy řetězce

V této kapitole se seznámíme se skrytými Markovovými řetězci (HMM), které představují rozšíření teorie Markovových řetězců na případ, kdy jednotlivé stavy nejsou přímo pozorovatelné. To znamená, že v daném čase nevidíme, v jakém stavu se řetězec nachází, ale známe pouze hodnoty pozorování jiné veličiny. To je důvod, proč takovéto řetězce nazýváme skrytými. Skryté Markovovy řetězce se skládají ze dvou náhodných procesů. Prvním je již zmíněný Markovův řetězec, který je charakterizován jednotlivými stavy a pravděpodobnostmi přechodu mezi nimi. Druhý proces charakterizují rozdělení pravděpodobnosti námi sledované náhodné veličiny, která jsou závislá na stavech. Tedy každému stavu odpovídá jiné rozdělení pravděpodobnosti. V práci se zaměříme na situaci, kdy je sledovaná veličina spojitá. Tyto spojité skryté Markovovy řetězce nacházejí řadu uplatnění. Jejich nejčastější a nejznámější aplikací je rozpoznávání mluvené řeči. Dále bývají aplikovány například při rozpoznávání fyzické aktivity či obrazu, v marketingu, bankovníctví, biomedicínském inženýrství k analýze signálů EKG a EEG či analýze seismické aktivity. V neposlední řadě bývají aplikovány v souvislosti s analýzou akciových trhů. Tato kapitola je zpracována za pomoci literatury [4], [6], [7], [9], [10], [12], [16].

2.1. Diskrétní a spojitě skryté Markovovy řetězce

Skryté Markovovy řetězce jsou definovány množinou stavů, počátečním rozdělením, pravděpodobnostmi přechodu a rozdělením pravděpodobnosti náhodné veličiny O_t , tedy námi sledované náhodné veličiny jejíž realizacím odpovídají pozorované hodnoty. Pro definování skrytých markovských řetězců je nejprve nutné se seznámit s jeho jednotlivými složkami. Důležité je také zmínit, že stejně jako v kapitole zabývající se markovskými řetězci, budeme i nadále uvažovat homogenní markovské řetězce s diskrétním časem.

Složky, se kterými jsme se seznámili již v teorii Markovových řetězců jsou:

- množina stavů $S = \{1, \dots, N\}$, kde N značí počet stavů,
- počáteční rozdělení pravděpodobnosti $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_N)$, kde

$$\pi_i = P(Q_0 = i) \text{ pro každé } i \in S,$$

- matice pravděpodobností přechodu $\mathbf{A} = (a_{ij})_{i,j \in S}$, pro jejíž složky platí

$$a_{ij} = P(Q_{t+1} = j | Q_t = i), \quad t = 0, 1, \dots, T, i, j \in S.$$

Složky, které se vyskytují nově u skrytých Markovových řetězců:

- množina možných pozorování V ,
- rozdělení pravděpodobnosti náhodné veličiny O_t .

Podívejme se nejprve na množinu možných pozorování V . Je-li množina možných pozorování konečná, hovoříme o diskrétních skrytých Markovových řetězcích a tato množina má podobu $V = \{v_0, v_1, \dots, v_Z\}$, kde Z je počet prvků množiny. Naopak v případě, kdy je množina možných pozorování nespočetná, $V \subseteq \mathbb{R}$, hovoříme o spojitých skrytých Markovových řetězcích.

Další nově se vyskytující složkou řetězců je rozdělení pravděpodobnosti náhodné veličiny O_t . V případě, že náhodná veličina O_t je diskrétní (tj. nabývá hodnot z $V = \{v_0, v_1, \dots, v_Z\}$), má rozdělení pravděpodobnosti následující tvar

$$b_j(k) = P(O_t = v_k | Q_t = j),$$

a platí tedy

$$b_j(k) \geq 0, \sum_{k=0}^Z b_j(k) = 1,$$

kde $t = 0, 1, \dots, T, j \in S, 0 \leq k \leq Z$. V případě spojitých HMM, kdy náhodná veličina O_t nabývá hodnot z $V \subseteq \mathbb{R}$, použijeme spojitě rozdělení pravděpodobnosti náhodné veličiny O_t (respektive jeho hustotu). Požadavek na spojitě rozdělení pravděpodobnosti je, aby jeho hustota byla log-konkávní. V práci tedy zvolíme hustotu normálního rozdělení pravděpodobnosti a $b_j(o_t)$ můžeme vyjádřit jako

$$b_j(o_t) = f(o_t, \mu_j, \sigma_j^2), t = 0, 1, \dots, T, j \in S$$

kde f je hustota normálního rozdělení pravděpodobnosti v bodě o_t s příslušnými parametry. Tedy μ_j je střední hodnota normálního rozdělení příslušující stavu j , σ_j^2 značí rozptyl příslušující stavu j a o_t je realizací náhodné veličiny O_t (tedy naším pozorováním). Jelikož se jedná o hustotu, musí být splněna podmínka

$$\int_{-\infty}^{\infty} b_j(O_t) dO_t = 1.$$

Je důležité si uvědomit, že narozdíl od diskrétního případu, kde $b_j(k)$ vyjadřuje pravděpodobnost generování pozorování v_k za předpokladu, že se nacházíme ve stavu j , není u spojitého případu $b_j(o_t)$ vyjádřením pravděpodobnosti, ale pouze hodnotou hustoty příslušného rozdělení v bodě o_t . Dále se v práci budeme zabývat případem spojitým.

Pozn. V některých situacích může být pro jednotlivé stavy vhodné místo jedné hustoty normálního (nebo jiného log-konkávního) rozdělení pravděpodobnosti použít hustot více. V našem případě hovoříme o směsi normálních rozdělení pravděpodobnosti, kterou vyjádříme jako lineární kombinaci dílčích hustot následovně

$$b_j(o_t) = \sum_{m=1}^M c_{jm} f(o_t, \mu_{jm}, \sigma_{jm}^2),$$

kde $f(o_t, \mu_{jm}, \sigma_{jm}^2)$ je hodnota dílčí (m -té) hustoty normálního rozdělení v bodě $o_t, t = 0, 1, \dots, T$ příslušející stavu $j \in S$, c_{jm} je váha dílčí (m -té) hustoty příslušející stavu j a M je počet dílčích hustot. Pro jednotlivé váhy platí

$$c_{jm} \geq 0, \sum_{m=1}^M c_{jm} = 1$$

a opět tedy musí být splněna podmínka

$$\int_{-\infty}^{\infty} b_j(O_t) dO_t = 1.$$

V práci se však omezíme pouze na základní případ, tedy bez směsí.

Nyní jsme se seznámili s veškerými parametry modelu a pro další použití si model, který je dán parametry $\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$ zapíšeme jako

$$\lambda = (\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2),$$

kde $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)$ a $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_N^2)$.

2.2. Tři základní úlohy skrytých Markovových řetězců

V následující kapitole si představíme tři úlohy, se kterými se při studiu Markovových řetězců můžeme setkat. Jsou jimi úloha ohodnocení, úloha určení optimální posloupnosti stavů a úloha učení.

2.2.1. Úloha ohodnocení

První úlohou, kterou si představíme, je úloha ohodnocení (*Evaluation problem*). Nechť máme dán model λ a posloupnost pozorování $\mathbf{O} = (o_0, o_1, \dots, o_T)$. Naším úkolem je získat míru věrohodnosti, že daná posloupnost pozorování \mathbf{O} je generována modelem λ , tj $P(\mathbf{O}|\lambda)$. Nejprve uvažujeme jednu posloupnost stavů $\mathbf{Q} = (q_0, q_1, \dots, q_T)$. Pravděpodobnost této posloupnosti stavů za předpokladu, že máme model λ je rovna

$$P(\mathbf{Q}|\lambda) = \pi_{q_0} \prod_{t=1}^T a_{q_{t-1}q_t}.$$

Dále potřebujeme znát míru věrohodnosti posloupnosti \mathbf{O} za předpokladu, že máme dānu posloupnost stavů \mathbf{Q} a model λ . Tuto míru věrohodnosti získáme jako součin hodnot hustot normálních rozdělení odpovídajících stavům q_0, q_1, \dots, q_T v bodech o_0, o_1, \dots, o_T , tj.

$$P(\mathbf{O}|\mathbf{Q}, \lambda) = \prod_{t=0}^T P(o_t|q_t, \lambda) = \prod_{t=0}^T b_{q_t}(o_t).$$

Míru věrohodnosti současného pozorování posloupnosti \mathbf{O} a posloupnosti stavů \mathbf{Q} , pokud máme dán model λ , vyjádříme jako

$$P(\mathbf{O}, \mathbf{Q}|\lambda) = \pi_{q_0} b_{q_0}(o_0) \prod_{t=1}^T a_{q_{t-1}q_t} b_{q_t}(o_t).$$

Sečtením přes všechny možné posloupnosti stavů získáme námi hledanou míru věrohodnosti $P(\mathbf{O}|\lambda)$

$$P(\mathbf{O}|\lambda) = \sum_{\mathbf{Q}} P(\mathbf{O}, \mathbf{Q}|\lambda).$$

Příklad. Dva biatlonisté se připravují na závod. Při každém tréninku je jim změřen čas, za který zajedou závodní trať. Jak tomu občas bývá, při ukládání měřených časů nastala chyba a k jednotlivým časům nejsou přiřazena jména závodníků. Tudíž trenér nemůže zanalyzovat výsledky závodníků a získat tak představu o jejich kondici. Rozhodl se tedy, že se na základě výsledků z předešlé sezóny pokusí jména závodníků k naměřeným časům přiřadit sám. Při výpočtu předpokládá, že pravděpodobnosti přechodu (střídání závodníků na tréninku) jsou určeny tak, že absolvování dvou po sobě jdoucích tréninků jedním závodníkem je méně pravděpodobné. Pravděpodobnost absolvování prvního tréninku, tj. prvního naměřeného času, je pro oba závodníky stejnā. Závodník 1 je zkušenější, proto jeho výsledky z loňské sezóny nemají příliš velký rozptyl $\sigma_1^2 = 5152,567$ a $\mu_1 = 1529,833$. Oproti tomu je závodník 2 poněkud zbrklejší, což se občas výrazněji projeví na čase, jehož střední hodnota a rozptyl jsou $\mu_2 = 1650,167$ a

$\sigma_2^2 = 11553,77$. Zobražíme si prvních šest zaznamenaných časů. Náš vektor pozorování tedy bude $\mathbf{O} = (1511, 1715, 1634, 1540, 1460, 1780)$ a počáteční rozdělení pravděpodobnosti $\boldsymbol{\pi} = (0, 5; 0, 5)$. Matice pravděpodobností přechodu bude mít následující tvar

$$\mathbf{A} = \begin{pmatrix} 0,3 & 0,7 \\ 0,7 & 0,3 \end{pmatrix}$$

Pro začátek si určíme věrohodnost, že první naměřený čas zaběhl závodník 1, která bude mít následující tvar

$$\begin{aligned} P(o_0, Q_0 = 1|\lambda) &= \pi_1 b_1(o_0) = \pi_1 \cdot \frac{1}{\sigma_1 \sqrt{2\pi}} \cdot e^{-\frac{(o_0 - \mu_1)^2}{2\sigma_1^2}} \\ &= 0,5 \cdot 0,00537 \\ &= 0,00269. \end{aligned}$$

Naopak, věrohodnost závodníka 2 pro první čas by byla

$$\begin{aligned} P(o_0, Q_0 = 2|\lambda) &= \pi_2 b_2(o_0) = \pi_2 \cdot \frac{1}{\sigma_2 \sqrt{2\pi}} \cdot e^{-\frac{(o_0 - \mu_2)^2}{2\sigma_2^2}} \\ &= 0,5 \cdot 0,00161 \\ &= 0,00081. \end{aligned}$$

Věrohodnost $P(\mathbf{O}, \mathbf{Q}|\lambda)$, tedy současného pozorování posloupnosti \mathbf{O} a posloupnosti stavů $\mathbf{Q} = (1, 2, 1, 1, 2, 2)$, můžeme určit jako

$$\begin{aligned} P(\mathbf{O}, \mathbf{Q}|\lambda) &= \pi_1 \cdot b_1(o_0) \cdot a_{12} \cdot b_2(o_1) \cdot a_{21} \cdot b_1(o_2) \cdot a_{11} \cdot b_1(o_3) \cdot a_{12} \cdot b_2(o_4) \cdot a_{22} \cdot b_2(o_5) \\ &= 0,5 \cdot 0,00537 \cdot 0,7 \cdot 0,00309 \cdot 0,7 \cdot 0,00194 \cdot 0,3 \cdot 0,00550 \cdot \\ &\quad \cdot 0,7 \cdot 0,00078 \cdot 0,3 \cdot 0,00179 \\ &= 3,8 \cdot 10^{-18} \end{aligned}$$

Pozn. Příklady jsou inspirovány reálnými výsledky biatlonových závodů dostupných z [5].

Jelikož v každém čase t můžeme dosáhnout jednoho z N možných stavů, existuje N^{T+1} různých posloupností stavů. Takovýto přístup by byl výpočetně náročný, proto se seznámíme s algoritmy, které nám výpočet $P(\mathbf{O}|\lambda)$ zjednoduší.

Prvním takovýmto algoritmem je **dopředný algoritmus** (*forward algorithm*). Zavedeme si pomocnou proměnnou

$$\alpha_t(i) = P(o_0, o_1, \dots, o_t, Q_t = i | \lambda),$$

kterou nazveme proměnnou dopřednou (*forward variable*) a která vyjadřuje míru věrohodnosti, s jakou pozorujeme posloupnost o_0, o_1, \dots, o_t a zároveň jsme v čase t ve stavu i za předpokladu modelu λ . Algoritmus má následující podobu:

1. Inicializace

$$\alpha_0(i) = \pi_i b_i(o_0) \text{ pro } i = 1, \dots, N$$

Na počátku algoritmu určíme pro každý stav míru věrohodnosti, se kterou generujeme pozorování o_0 . Tu určíme pomocí počátečního rozdělení pravděpodobnosti a rozdělení pravděpodobnosti veličiny O_t pro jednotlivé stavy.

2. Indukce (krok $t + 1$)

$$\alpha_{t+1}(j) = b_j(o_{t+1}) \sum_{i=1}^N \alpha_t(i) a_{ij} \text{ pro } j = 1, \dots, N, t = 0, \dots, T - 1$$

Při určení dopředné proměnné v čase následujícím ($t + 1$) pro j -tý stav uvažujeme míru věrohodnosti, se kterou generujeme pozorování o_{t+1} ve stavu j a věrohodnost všech „cest“, kterými jsme se do stavu j dostali. Tuto věrohodnost cest určíme jako součet dopředných proměnných v čase současném přes všechny možné stavy se zohledněním pravděpodobností přechodů z jednotlivých stavů do stavu uvažovaného (j -tého).

3. Závěr

$$P(\mathbf{O} | \lambda) = \sum_{i=1}^N \alpha_T(i)$$

Až se pomocí předchozích kroků dostaneme do času T , sečteme všechny dopředné proměnné příslušející tomuto času (tedy N proměnných), které

představují věrohodnosti všech možných cest, kterými jsme se mohli do jednotlivých koncových stavů dostat a získáme hledanou míru věrohodnosti $P(\mathbf{O}|\lambda)$.

Obdobně můžeme získat námi hledanou míru věrohodnosti $P(\mathbf{O}|\lambda)$ pomocí kombinace dopředné a zpětné proměnné (*backward variable*). Zpětná proměnná vyjadřuje míru věrohodnosti, s jakou budeme pozorovat posloupnost $o_{t+1}, o_{t+2}, \dots, o_T$ za předpokladu, že jsme v čase t ve stavu i a uvažujeme model λ . Zpětná proměnná bude tedy definována následovně

$$\varphi_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | Q_t = i, \lambda).$$

Narozdíl od proměnné dopředné postupujeme při výpočtu odzadu, tedy od času T . Výpočet zpětné proměnné v čase t bude vypadat následovně

$$\varphi_t(i) = \sum_{j=1}^N \varphi_{t+1}(j) a_{ij} b_j(o_{t+1}), \quad t = T-1, \dots, 1, 0,$$

kde

$$\varphi_T(i) = 1, \quad i = 1, \dots, N.$$

Vynásobením výše zmíněných proměnných vyjádřených v čase t a stavu i získáme

$$\alpha_t(i) \varphi_t(i) = P(\mathbf{O}, Q_t = i | \lambda)$$

a součtem přes všechny stavy získáme hledanou míru věrohodnosti

$$P(\mathbf{O}|\lambda) = \sum_{i=1}^N P(\mathbf{O}, Q_t = i | \lambda) = \sum_{i=1}^N \alpha_t(i) \varphi_t(i).$$

Příklad. Pokud se opět vrátíme k našemu příkladu biatlonových závodníků, vypočítáme hodnoty dopředných proměnných

$$\begin{aligned} \alpha_0(1) &= \pi_1 b_1(o_0) &= 0,00269, & \alpha_0(2) &= \pi_2 b_2(o_0) &= 0,00081 \\ \alpha_1(1) &= b_1(o_1) \sum_{i=1}^2 \alpha_0(i) a_{i1} &= 2,73 \cdot 10^{-7}, & \alpha_1(2) &= b_2(o_1) \sum_{i=1}^2 \alpha_0(i) a_{i2} &= 6,56 \cdot 10^{-6} \\ \alpha_2(1) &= b_1(o_2) \sum_{i=1}^2 \alpha_1(i) a_{i1} &= 9,06 \cdot 10^{-9}, & \alpha_2(2) &= b_2(o_2) \sum_{i=1}^2 \alpha_1(i) a_{i2} &= 7,92 \cdot 10^{-9} \\ &\vdots & & &\vdots & \\ \alpha_5(1) &= b_1(o_5) \sum_{i=1}^2 \alpha_4(i) a_{i1} &= 6,21 \cdot 10^{-19}, & \alpha_5(2) &= b_2(o_5) \sum_{i=1}^2 \alpha_4(i) a_{i2} &= 1,33 \cdot 10^{-16}. \end{aligned}$$

Následně určíme věrohodnost $P(\mathbf{O}|\lambda)$ jako

$$P(\mathbf{O}|\lambda) = \sum_{i=1}^2 \alpha_5(i) = 1,34 \cdot 10^{-16}.$$

Výsledná $P(\mathbf{O}|\lambda)$ tedy říká, s jakou věrohodností je daná posloupnost pozorování \mathbf{O} generována modelem λ .

2.2.2. Úloha určení optimální posloupnosti stavů

Druhou úlohou, kterou můžeme u skrytých Markovových řetězců řešit, je určení optimální posloupnosti stavů (*Decoding problem*) při dané posloupnosti pozorování \mathbf{O} . Nechť máme dán model λ a posloupnost nezávislých pozorování $\mathbf{O} = (o_0, o_1, \dots, o_T)$. Nyní máme dva přístupy k určení optimální posloupnosti stavů. V prvním z nich určíme optimální stavy v jednotlivých časech individuálně. Tedy v čase t bude optimálním stavem ten stav q_t , který bude v tomto čase nejvíce věrohodný.

Definujeme si novou pomocnou proměnnou, která vyjadřuje pravděpodobnost, se kterou se v čase t nacházíme ve stavu i za předpokladu, že máme danu posloupnost pozorování \mathbf{O} a model λ

$$\gamma_t(i) = P(Q_t = i | \mathbf{O}, \lambda).$$

Tuto pomocnou proměnnou lze vyjádřit také jako

$$\gamma_t(i) = \frac{P(\mathbf{O}, Q_t = i | \lambda)}{P(\mathbf{O} | \lambda)} = \frac{P(\mathbf{O}, Q_t = i | \lambda)}{\sum_{i=1}^N P(\mathbf{O}, Q_t = i | \lambda)}$$

a s využitím již dříve definované dopředné a zpětné proměnné můžeme napsat

$$\gamma_t(i) = \frac{\alpha_t(i) \varphi_t(i)}{\sum_{i=1}^N \alpha_t(i) \varphi_t(i)}.$$

Nejpravděpodobnější stav v čase t určíme pomocí $\gamma_t(i)$ jako

$$q_t^* = \arg \max_{1 \leq i \leq N} \gamma_t(i), t = 0, \dots, T.$$

Ne vždy však může být tento přístup, kdy určujeme optimální stav v každém čase individuálně, vhodný. Problém může nastat v případě, kdy některá z pravděpodobností přechodu je nulová ($a_{ij} = 0$), tudíž by mohlo být nemožné se do námi určeného optimálního stavu vůbec dostat. Na řadu přichází druhý přístup k určení optimální posloupnosti stavů, který vyřeší zmíněný problém. Nazývá se **Viterbiho algoritmus**.

Na rozdíl od výše popsaného přístupu hledá Viterbiho algoritmus optimální posloupnost stavů jako celek, nikoli pro jednotlivé časy individuálně. Jedná se tedy o úlohu maximalizace $P(\mathbf{Q}|\mathbf{O}, \lambda)$. Jelikož $P(\mathbf{Q}|\mathbf{O}, \lambda) = \frac{P(\mathbf{O}, \mathbf{Q}|\lambda)}{P(\mathbf{O}|\lambda)}$, je tato maximalizace ekvivalentní maximalizaci $P(\mathbf{Q}, \mathbf{O}|\lambda)$.

Opět si zavedeme novou pomocnou proměnnou $\delta_t(i)$, která bude vyjadřovat nejlepší skóre (nejvyšší míru věrohodnosti), které můžeme v čase t obdržet tak, že se nacházíme v čase t ve stavu i a pozorujeme posloupnost o_0, o_1, \dots, o_t . Posloupnost stavů odpovídající nejlepšímu skóre v čase t bude $q_0, q_1, \dots, q_{t-1}, q_t = i$

$$\delta_t(i) = \max_{q_0, \dots, q_{t-1}} P(q_0, \dots, q_{t-1}, q_t = i, o_0, \dots, o_t | \lambda).$$

Pro úplnost si zavedeme ještě pomocnou proměnnou $\psi_t(j)$, která bude sloužit ke zpětnému trasování stavů (konkrétně bude odpovídat stavu q_{t-1} z něhož přechod do stavu $q_t = j$ byl nejvíce věrohodný). Nyní můžeme přistoupit k samotnému Viterbiho algoritmu, který má následující podobu

1. Inicializace

$$\delta_0(i) = \pi_i b_i(o_0) \text{ pro } i = 1, \dots, N.$$

Stejně jako u dopředného algoritmu, začneme u Viterbiho algoritmu tím, že určíme pro každý stav míru věrohodnosti, se kterou generujeme pozorování o_0 , tj. skóre v čase $t = 0$.

2. Rekurze (krok t)

$$\delta_t(j) = b_j(o_t) \max_{i=1, \dots, N} \delta_{t-1}(i) a_{ij} \text{ pro } j = 1, \dots, N, t = 1, \dots, T.$$

$$\psi_t(j) = \arg \max_{i=1, \dots, N} \delta_{t-1}(i) a_{ij} \quad \text{pro } j = 1, \dots, N, t = 1, \dots, T.$$

Při určení skóre v čase „současném“ (t) pro j -tý stav uvažujeme míru věrohodnosti, s jakou generujeme pozorování o_{t+1} ve stavu j a „nejvěrohodnější cestu“, kterou se do tohoto stavu můžeme dostat. Cestou zde nazýváme posloupnost stavů a její věrohodnost získáme tak, že pro každý stav provedeme součin skóre v čase předchozím ($t - 1$) a pravděpodobnosti přechodu do stavu j . Nejvěrohodnější cestu získáme jako maximum z těchto součinů přes všechny stavy. Zároveň si do proměnné $\psi_t(j)$ „uložíme“ stav, ze kterého tato nejvěrohodnější cesta vedla. V každém kroku algoritmu tedy spočítáme $2N$ proměnných.

3. Závěr

$$P^* = \max_{i=1, \dots, N} \delta_T(i)$$

$$q_T^* = \arg \max_{i=1, \dots, N} \delta_T(i)$$

Až se pomocí předchozích kroků dostaneme do času T , určíme maximální skóre P^* , které odpovídá míře věrohodnosti nejvěrohodnější cesty (tj. posloupnosti stavů). Stav, který odpovídá tomuto největšímu skóre, je nami hledaným optimálním stavem q_T^* v čase T .

4. Zpětné trasování

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad \text{pro } t = T - 1, \dots, 0.$$

Poslední fází Viterbiho algoritmu je zpětné trasování (*backtracking*). V této fázi určíme pomocí proměnné $\psi_t(j)$ optimální stavy v jednotlivých časech, až získáme kompletní posloupnost optimálních stavů. Zpětné trasování tedy začínáme v čase $T - 1$ a končíme v čase 0.

Příklad. Viterbiho algoritmus opět aplikujeme na příklad biatlonových závodníků. Vypočítáme hodnoty proměnných $\delta_t(j)$, které vyjadřují nejvyšší míru věrohodnosti, kterou můžeme v čase t obdržet a určíme hodnoty proměnných $\psi_t(j)$, které představují příslušné nejvěrohodnější stavy.

$$\begin{aligned}
\delta_0(1) &= \pi_1 b_1(o_0) &= 0,00269, & \delta_0(2) &= \pi_2 b_2(o_0) &= 0,00081 \\
\delta_1(1) &= b_1(o_1) \max_{i=1,2} \delta_0(i) a_{i1} &= 1,61 \cdot 10^{-7}, & \delta_1(2) &= b_2(o_1) \max_{i=1,2} \delta_0(i) a_{i2} &= 5,82 \cdot 10^{-6} \\
\delta_2(1) &= b_1(o_2) \max_{i=1,2} \delta_1(i) a_{i1} &= 7,89 \cdot 10^{-9}, & \delta_2(2) &= b_2(o_2) \max_{i=1,2} \delta_1(i) a_{i2} &= 6,40 \cdot 10^{-9} \\
\delta_3(1) &= b_1(o_3) \max_{i=1,2} \delta_2(i) a_{i1} &= 2,47 \cdot 10^{-11}, & \delta_3(2) &= b_2(o_3) \max_{i=1,2} \delta_2(i) a_{i2} &= 1,21 \cdot 10^{-11} \\
\delta_4(1) &= b_1(o_4) \max_{i=1,2} \delta_3(i) a_{i1} &= 2,94 \cdot 10^{-14}, & \delta_4(2) &= b_2(o_4) \max_{i=1,2} \delta_3(i) a_{i2} &= 1,34 \cdot 10^{-14} \\
\delta_5(1) &= b_1(o_5) \max_{i=1,2} \delta_4(i) a_{i1} &= 1,20 \cdot 10^{-19}, & \delta_5(2) &= b_2(o_5) \max_{i=1,2} \delta_4(i) a_{i2} &= 3,68 \cdot 10^{-17}
\end{aligned}$$

$$\begin{aligned}
\psi_1(1) &= 1, & \psi_1(2) &= 1 \\
\psi_2(1) &= 2, & \psi_2(2) &= 2 \\
\psi_3(1) &= 2, & \psi_3(2) &= 1 \\
\psi_4(1) &= 2, & \psi_4(2) &= 1 \\
\psi_5(1) &= 2, & \psi_5(2) &= 1
\end{aligned}$$

Dalším krokem je stanovit nejvyšší skóre v čase $t = 5$ a na základě toho i optimální stav v čase 5.

$$\begin{aligned}
P^* &= \max_{i=1,2} \delta_5(i) &= 3,68 \cdot 10^{-17} \\
q_5^* &= \arg \max_{i=1,2} \delta_5(i) &= 2
\end{aligned}$$

Na závěr určíme zpětným trasováním optimální stavy pro všechny předešlé časové okamžiky $t = 0, \dots, 4$, tj.

$$\begin{aligned}
q_4^* &= \psi_5(q_5^*) = 1 \\
q_3^* &= \psi_4(q_4^*) = 2 \\
q_2^* &= \psi_3(q_3^*) = 1 \\
q_1^* &= \psi_2(q_2^*) = 2 \\
q_0^* &= \psi_1(q_1^*) = 1
\end{aligned}$$

Pro určení optimální posloupnosti stavů pomocí softwaru R lze použít funkci *viterbicont* z knihovny **HMMCont** [2].

2.2.3. Úloha učení

V části nazývané „učení“ se seznámíme s metodou, pomocí které je možné odhadnout optimální parametry modelu λ . To znamená, že hledáme parametry

$\lambda^* = (\boldsymbol{\pi}^*, \mathbf{A}^*, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*})$ tak, aby příslušná věrohodnost $P(\mathbf{O}|\lambda^*)$ byla maximální, tj. $\lambda^* = \arg \max_{\lambda} P(\mathbf{O}|\lambda)$. Touto metodou je Baum-Welchův algoritmus, který je speciálním případem EM (*expectation-maximalization*) algoritmu. Jedná se o iterativní proces, který po zadání počátečního odhadu parametrů tyto odhady iterativně zpřesňuje dokud nebude splněno příslušné ukončovací kritérium. Ukončovacím kritériem může být jednak maximální počet iterací $I_{max} \in \mathbb{N}$, a dále pak minimální rozdíl Akaikeho kritérií (AIC) vypočtených ve dvou po sobě jdoucích iteracích (tj. $|AIC^{(l+1)} - AIC^{(l)}| < \epsilon$, kde $AIC^{(l)} = -2 \sum_{i=1}^N \alpha_T^{(l)}(i) + 2k$ značí hodnotu Akaikeho kritéria v l -té iteraci, $l = 0, 1, \dots, I_{max}$, kde $l = 0$ značí počáteční odhad parametrů (viz. dále), $k \in N$ je počet parametrů modelu a $\epsilon \in \mathbb{R}^+$ označuje toleranci). V [3] je místo rozdílu AIC použit rozdíl logaritmu věrohodnosti, tj. $\log P(\mathbf{O}|\lambda)$ ve dvou po sobě jdoucích iteracích. Jednoduše lze vyjádřit vzájemný vztah těchto dvou ukazatelů, ze kterého plyne, že jako vhodné ukončovací kritérium lze použít i rozdíl Akaikeho kritérií ve dvou po sobě jdoucích iteracích l a $l + 1$, kde $l = 0, \dots, I_{max} - 1$.

$$|AIC^{(l+1)} - AIC^{(l)}| = 2 \cdot |\log P(\mathbf{O}|\lambda)^{(l+1)} - \log P(\mathbf{O}|\lambda)^{(l)}|.$$

Nejjednodušším případem, se kterým se můžeme setkat, je situace, kdy máme k dispozici jak pozorování $\mathbf{O} = (o_0, o_1, \dots, o_T)$, tak příslušné stavy $\mathbf{Q} = (q_0, q_1, \dots, q_T)$ (např. trénovací data). V takovém případě lze parametry modelu odhadnout přímo.

Příklad. Představme si náš modelový příklad za předpokladu, že by ke ztrátě dat nedošlo a trenér by měl kompletní informace o časech obou závodníků. V takovém případě získáme posloupnosti pozorování a příslušných stavů

$$\begin{aligned} \mathbf{O} &= (1511 , 1715 , 1634 , 1540 , 1460 , 1780) \\ \mathbf{Q} &= (1 , 2 , 1 , 1 , 2 , 2) \end{aligned}$$

Z dat získaných prostřednictvím tohoto měření odhadneme příslušné parametry modelu.

V první řadě určíme odhad počátečního rozdělení pravděpodobnosti

$$\hat{\pi}_i = \frac{\text{počet výskytů ve stavu } i \text{ v čase } 0}{\text{počet výskytů v čase } 0}, i = 1, 2.$$

V našem případě

$$\hat{\pi} = (1, 0).$$

Dalším odhadovaným parametrem modelu je matice pravděpodobností přechodu $\hat{\mathbf{A}}$, jejíž prvky odhadneme jako

$$\hat{a}_{ij} = \frac{\text{počet přechodů ze stavu } i \text{ do stavu } j}{\text{počet přechodů ze stavu } i}$$

Matice bude v našem případě vypadat následovně

$$\hat{\mathbf{A}} = \begin{pmatrix} \frac{1}{3}, & \frac{2}{3} \\ \frac{1}{2}, & \frac{1}{2} \end{pmatrix}.$$

Na závěr odhadneme příslušné střední hodnoty a rozptyly rozdělení pro jednotlivé stavy $i = 1, 2$

$$\hat{\mu}_i = \frac{\text{součet hodnot pozorování ve stavu } i}{\text{počet pozorování ve stavu } i}, \quad \hat{\sigma}_i^2 = \frac{\text{součet kvadrátů odchylek hodnot pozorování ve stavu } i \text{ od } \hat{\mu}_i}{(\text{počet pozorování ve stavu } i) - 1}$$

Po dosazení našich pozorování získáme

$$\begin{aligned} \hat{\mu}_1 &= 1561,67, & \hat{\sigma}_1^2 &= 4134,33, \\ \hat{\mu}_2 &= 1651,67, & \hat{\sigma}_2^2 &= 28608,33. \end{aligned}$$

Běžně to však tak jednoduché nemáme a posloupnost stavů neznáme. Zde přichází na řadu již zmíněný Baum-Welchův algoritmus. V první řadě je nutné stanovit počáteční odhady parametrů modelu. Počáteční odhady parametrů můžeme stanovit buď expertně, nebo použít jinou vhodnou metodu. Takovou metodou může být například metoda K-průměrů. Jedná se o metodu shlukové analýzy a tedy jejím cílem je rozdělit pozorování do K shluků tak, aby byla co nejméně vzdálena od středu shluku, který odpovídá průměru pozorování v daném shluku. Předpokládáme tedy, že shluky odpovídají stavům řetězce. Na základě zařazení

pozorování do stavů můžeme určit počáteční odhady parametrů skrytých Markovových řetězců, které jsou dány následujícími vzorci

$$\begin{aligned}\hat{\pi}_i^{(0)} &= \frac{\text{počet výskytů ve stavu } i \text{ v čase } 0}{\text{počet výskytů v čase } 0}, \\ \hat{a}_{ij}^{(0)} &= \frac{\text{počet přechodů ze stavu } i \text{ do stavu } j}{\text{počet přechodů ze stavu } i}, \\ \hat{\mu}_i^{(0)} &= \frac{\text{součet hodnot pozorování ve stavu } i}{\text{počet pozorování ve stavu } i}, \\ \hat{\sigma}_i^2(0) &= \frac{\text{součet kvadrátů odchylek hodnot pozorování ve stavu } i \text{ od } \hat{\mu}_i}{(\text{počet pozorování ve stavu } i) - 1},\end{aligned}$$

kde $i, j = 1, 2, \dots, N$.

Dále si zavedeme $\xi_t(i, j)$ pro $t = 0, \dots, T - 1$. Jestliže máme dán model λ a posloupnost pozorování \mathbf{O} , je pravděpodobnost, že se systém v čase t nachází ve stavu i a zároveň v čase $t + 1$ ve stavu j definována jako

$$\begin{aligned}\xi_t(i, j) &= P(Q_t = i, Q_{t+1} = j | \mathbf{O}, \lambda) \\ &= \frac{P(Q_t = i, Q_{t+1} = j, \mathbf{O} | \lambda)}{(\mathbf{O} | \lambda)} \\ &= \frac{\alpha_t(i) \varphi_{t+1}(j) a_{ij} b_j(o_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) \varphi_{t+1}(j) a_{ij} b_j(o_{t+1})}.\end{aligned}$$

Již dříve jsme si zavedli pomocnou proměnnou $\gamma_t(i)$, kterou jsme definovali následovně

$$\gamma_t(i) = \frac{\alpha_t(i) \varphi_t(i)}{\sum_{i=1}^N \alpha_t(i) \varphi_t(i)}, \quad t = 0, \dots, T,$$

a která vyjadřuje pravděpodobnost, se kterou se v čase t nacházíme ve stavu i za předpokladu, že máme dānu posloupnost pozorování \mathbf{O} a model λ . Vzájemný vztah těchto dvou pravděpodobností můžeme (pro $t = 0, \dots, T - 1$) vyjádřit jako

$$\gamma_t(i) = P(Q_t = i | \mathbf{O}, \lambda) = \sum_{j=1}^N P(Q_t = i, Q_{t+1} = j | \mathbf{O}, \lambda) = \sum_{j=1}^N \xi_t(i, j).$$

Pokud na tyto pravděpodobnosti uplatníme součet přes všechny časové okamžiky, bude výsledkem příslušný očekávaný počet přechodů. Konkrétně

$$\sum_{t=0}^{T-1} \gamma_t(i) = \text{očekávaný počet přechodů ze stavu } i,$$

$$\sum_{t=0}^{T-1} \xi_t(i, j) = \text{očekávaný počet přechodů ze stavu } i \text{ do stavu } j.$$

Využitím těchto vztahů odhadneme parametry modelu následovně

$$\begin{aligned} \hat{\pi}_i &= \frac{\text{očekávaný počet výskytů ve stavu } i \text{ v čase } 0}{\text{očekávaný počet výskytů v čase } 0} \\ &= \gamma_0(i) \\ \hat{a}_{ij} &= \frac{\text{očekávaný počet přechodů ze stavu } i \text{ do stavu } j}{\text{očekávaný počet přechodů ze stavu } i} \\ &= \frac{\sum_{t=0}^{T-1} \xi_t(i, j)}{\sum_{t=0}^{T-1} \gamma_t(i)} \end{aligned}$$

V námi uvažovaném spojitém případě potřebujeme dále odhadnout parametry μ_i a σ_i^2 . Představme si, že $N = 1$ (to znamená, že uvažujeme pouze jeden možný stav). V takové situaci bychom odhady těchto parametrů určili jednoduše jako

$$\hat{\mu} = \frac{1}{T+1} \sum_{t=0}^T o_t,$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=0}^T (o_t - \hat{\mu})^2.$$

Pokud však tuto myšlenku rozšíříme na náš případ, ve kterém můžeme mít stavů více, provedeme výpočty těchto parametrů váženě. Jako váhy zvolíme pravděpodobnosti, tedy pravděpodobnosti že se v čase t nacházíme ve stavu i . Potom

odhady paramterů μ_i a σ_i^2 budou mít následující podobu

$$\hat{\mu}_i = \frac{\sum_{t=0}^T \gamma_t(i) o_t}{\sum_{t=0}^T \gamma_t(i)},$$

$$\hat{\sigma}_i^2 = \frac{\sum_{t=0}^T \gamma_t(i) (o_t - \hat{\mu}_i)^2}{\sum_{t=0}^T \gamma_t(i)}.$$

Na základě výše zavedených vyjádření můžeme přejít k samotnému Baum-Welchovu iteračnímu algoritmu, který lze popsat následovně:

1. Krok 1 (Inicializace)

V první fázi stanovíme počáteční odhady parametrů modelu.

$$\hat{\lambda}^{(0)} = (\hat{\boldsymbol{\pi}}^{(0)}, \hat{\mathbf{A}}^{(0)}, \hat{\boldsymbol{\mu}}^{(0)}, \hat{\boldsymbol{\sigma}}^2{}^{(0)})$$

Položíme $l = 0$.

2. Krok 2

Dále dopočítáme hodnoty proměnných $\xi_t^{(l)}(i, j)$ a $\gamma_t^{(l)}(i)$ pro všechna $i, j \in S$ následovně

$$\xi_t^{(l)}(i, j) = \frac{\alpha_t^{(l)}(i) \varphi_{t+1}^{(l)}(j) \hat{a}_{ij}^{(l)} b_j^{(l)}(o_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t^{(l)}(i) \varphi_{t+1}^{(l)}(j) \hat{a}_{ij}^{(l)} b_j^{(l)}(o_{t+1})}, \quad t = 0, 1, \dots, T-1,$$

$$\gamma_t^{(l)}(i) = \frac{\alpha_t^{(l)}(i) \varphi_t^{(l)}(i)}{\sum_{i=1}^N \alpha_t^{(l)}(i) \varphi_t^{(l)}(i)}, \quad t = 0, 1, \dots, T,$$

kde

$$\begin{aligned}
 b_j^{(l)}(o_t) &= f(o_t, \hat{\mu}_j^{(l)}, \hat{\sigma}_j^2{}^{(l)}), \\
 \alpha_0^{(l)}(i) &= \hat{\pi}_i^{(l)} b_i^{(l)}(o_0), \\
 \alpha_t^{(l)}(i) &= b_i^{(l)}(o_t) \sum_{j=1}^N \alpha_{t-1}^{(l)}(j) \hat{a}_{ji}^{(l)}, \\
 \varphi_T^{(l)}(j) &= 1, \\
 \varphi_t^{(l)}(i) &= \sum_{j=1}^N \varphi_{t+1}^{(l)}(j) \hat{a}_{ij}^{(l)} b_j^{(l)}(o_{t+1}).
 \end{aligned}$$

Vypočítáme hodnotu Akaikeho kritéria pro model s počtem parametrů odpovídajícím k .

$$AIC^{(l)} = -2 \sum_{i=1}^N \alpha_T^{(l)}(i) + 2k.$$

Pokud $l = 0$, přejdeme na krok 4.

Pokud $l > 0$, přejdeme na krok 3.

3. Krok 3

Pokud $l = I_{max}$ nebo $|AIC^{(l)} - AIC^{(l-1)}| < \epsilon$, přejdeme na krok 5.

V opačném případě přejdeme na krok 4.

4. Krok 4

Pro všechna $i, j \in S$ odhadneme nové parametry modelu jako

$$\begin{aligned}\hat{\pi}_i^{(l+1)} &= \gamma_0^{(l)}(i), \\ \hat{a}_{ij}^{(l+1)} &= \frac{\sum_{t=0}^{T-1} \xi_t^{(l)}(i, j)}{\sum_{t=0}^{T-1} \gamma_t^{(l)}(i)}, \\ \hat{\mu}_i^{(l+1)} &= \frac{\sum_{t=0}^T \gamma_t^{(l)}(i) o_t}{\sum_{t=0}^T \gamma_t^{(l)}(i)}, \\ \hat{\sigma}_i^{2(l+1)} &= \frac{\sum_{t=0}^T \gamma_t^{(l)}(i) (o_t - \hat{\mu}_i^{(l+1)})^2}{\sum_{t=0}^T \gamma_t^{(l)}(i)}.\end{aligned}$$

Položíme $l = l + 1$ a přejdeme na krok 2.

5. Krok 5

Optimálními odhady parametrů modelu jsou

$$\begin{aligned}\hat{\pi}_i^* &= \hat{\pi}_i^{(l)}, \\ \hat{a}_{ij}^* &= \hat{a}_{ij}^{(l)}, \\ \hat{\mu}_i^* &= \hat{\mu}_i^{(l)}, \\ \hat{\sigma}_i^{2*} &= \hat{\sigma}_i^{2(l)}.\end{aligned}$$

Příklad. Pro odhad optimálních parametrů modelu v našem modelovém příkladě použijeme tedy Baum-Welchův algoritmus. Vzhledem k výpočetní náročnosti provedeme Baum-Welchův algoritmus prostřednictvím softwaru R, a to pomocí funkce *baumwelchcont* z knihovny **HMMCont**.

Naše počáteční odhady parametrů jsou

$$\begin{aligned}\hat{\pi}^{(0)} &= (0, 5; 0, 5) \\ \hat{\mu}^{(0)} &= (1529, 83; 1650, 17) \\ \hat{\sigma}^{2(0)} &= (5152, 57; 11553, 77)\end{aligned}$$

$$\hat{\mathbf{A}}^{(0)} = \begin{pmatrix} 0, 3; 0, 7 \\ 0, 7; 0, 3 \end{pmatrix}.$$

Po jedné iteraci metodou Baum-Welchova algoritmu získáme

$$\begin{aligned}\hat{\boldsymbol{\pi}}^{(1)} &= (0, 87; 0, 13), \\ \hat{\boldsymbol{\mu}}^{(1)} &= (1526, 49; 1632, 79), \\ \hat{\boldsymbol{\sigma}}^2^{(1)} &= (4181, 25; 7378, 35),\end{aligned}$$

$$\hat{\mathbf{A}}^{(1)} = \begin{pmatrix} 0, 23; 0, 77 \\ 0, 62; 0, 38 \end{pmatrix}.$$

Hodnota Akaikeho kritéria bude v tomto případě rovna $AIC^{(1)} = 93, 10$. Pokud chceme získat optimální parametry modelu, je třeba provést iterací více. Nejprve si však zvolíme ukončovací kritéria. Jedním z kritérií bude maximální počet iterací $I_{max}=50$. Dalším ukončovacím kritériem bude tolerance, tedy minimální rozdíl Akaikeho kritérií ve dvou po sobě jdoucích iteracích, kterou položíme rovnu $\epsilon = 0, 5$. Při použití těchto kritérií odhadneme optimální parametry modelu po pěti iteracích. Takto odhadnuté parametry modelu budou mít následující podobu

$$\begin{aligned}\hat{\boldsymbol{\pi}}^{(5)} &= (1; 0), \\ \hat{\boldsymbol{\mu}}^{(5)} &= (1497, 02; 1638, 19), \\ \hat{\boldsymbol{\sigma}}^2^{(5)} &= (963, 26; 5463, 86),\end{aligned}$$

$$\hat{\mathbf{A}}^{(5)} = \begin{pmatrix} 0, 16; 0, 84 \\ 0, 36; 0, 64 \end{pmatrix}.$$

Akaikeho kritérium po pěti iteracích je rovno $AIC^{(5)} = 90, 74$. Po následné aplikaci Viterbiho algoritmu s nově odhadnutými parametry pomocí funkce *viterbi-cont* získáme novou posloupnost optimálních stavů $\hat{\mathbf{Q}}^{*(5)} = (\hat{q}_0^{*(5)}, \dots, \hat{q}_5^{*(5)})$. Pro porovnání zobrazíme i posloupnost optimálních stavů získanou Viterbiho algorit-

mem na základě počátečního odhadu parametrů $\hat{Q}^{*(0)} = (\hat{q}_0^{*(0)}, \dots, \hat{q}_5^{*(0)})$.

$$\begin{aligned} \hat{q}_0^{*(0)} &= 1, & \hat{q}_0^{*(5)} &= 1 \\ \hat{q}_1^{*(0)} &= 2, & \hat{q}_1^{*(5)} &= 2 \\ \hat{q}_2^{*(0)} &= 1, & \hat{q}_2^{*(5)} &= 2 \\ \hat{q}_3^{*(0)} &= 2, & \hat{q}_3^{*(5)} &= 2 \\ \hat{q}_4^{*(0)} &= 1, & \hat{q}_4^{*(5)} &= 1 \\ \hat{q}_5^{*(0)} &= 2, & \hat{q}_5^{*(5)} &= 2 \end{aligned}$$

Na základě výsledků se tedy zdá, že oba závodníci dosahují lepších a stabilnějších výkonů, než tomu bylo v minulém roce. Můžeme také říci, že závodník číslo 2 trénuje častěji než závodník číslo 1.

2.3. Skryté Markovovy regresní modely

Skryté Markovovy regresní modely (HMMR) jsou rozšířením teorie skrytých markovských modelů na případ regresní analýzy. Tato teorie byla zpracována s použitím zdrojů [9], [12].

V případě skrytých Markovových regresních modelů předpokládáme, že jednotlivým stavům markovského řetězce odpovídají různé regresní parametry. Tedy každý stav je určen jednou sadou regresních parametrů, které charakterizují vztah mezi pozorovanou závisle proměnnou (Y) a jednou či více nezávisle proměnnými (X_1, \dots, X_p). Necht' máme danu množinu stavů $S = \{1, \dots, N\}$. Vícenásobnou lineární regresi za předpokladu, že se systém nachází ve stavu i zapíšeme jako

$$Y_t = \beta_{0i} + \beta_{1i}X_{t1} + \beta_{2i}X_{t2} + \dots + \beta_{pi}X_{tp} + \epsilon_{ti}, \quad 0 \leq t \leq T, i \in S,$$

kde Y_t značí hodnotu vysvětlované proměnné v čase t , X_{t1}, \dots, X_{tp} jsou hodnoty nezávisle proměnných v čase t , kde p je jejich počet. Dále jako $\beta_{0i}, \dots, \beta_{pi}$ označujeme regresní parametry za předpokladu, že se nacházíme ve stavu i a ϵ_{ti} reprezentuje chybu modelu v čase t opět za předpokladu, že se nacházíme ve stavu i . O chybách předpokládáme, že jsou nezávislé náhodné veličiny s normálním rozdělením $\mathcal{N}(0, \sigma_i^2)$, kde σ_i^2 značí rozptyl chyb ve stavu i .

Lineární regresní model pro i -tý stav bude mít následující podobu

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i,$$

$$\mathbf{Y} = (Y_0, \dots, Y_T)', \boldsymbol{\epsilon}_i = (\epsilon_{0i}, \dots, \epsilon_{Ti})', \boldsymbol{\beta}_i = (\beta_{0i}, \dots, \beta_{pi})', \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T1} & \cdots & X_{Tp} \end{pmatrix}.$$

Matici \mathbf{X} nazýváme maticí plánu a předpokládáme, že její rozměr je $T \times (p + 1)$, tj. předpokládáme model s absolutním členem. Naše pozorování nyní tvoří posloupnost hodnot závisle proměnných $\mathbf{Y} = (Y_0, \dots, Y_T)$, které jsou normálně rozdělené se střední hodnotou odpovídající $\beta_{0i} + \beta_{1i}X_{t1} + \beta_{2i}X_{t2} + \cdots + \beta_{pi}X_{tp}$ pro $t = 0, \dots, T$ a rozptylem σ_i^2 , za předpokladu, že jsme ve stavu i . V takovémto případě bude mít hustota normálního rozdělení v bodě Y_t následující podobu

$$b_i(Y_t) = f(Y_t, \mathbf{X}_t' \boldsymbol{\beta}_i, \sigma_i^2), t = 0, 1, \dots, T, i \in S,$$

kde \mathbf{X}_t je vektor, který odpovídá t -tému řádku matice \mathbf{X} .

Jelikož se jedná o hustotu, musí být opět splněna podmínka

$$\int_{-\infty}^{\infty} b_i(k_t) dk_t = 1.$$

Další změnou oproti klasické teorii skrytých Markovových modelů, na kterou při použití skrytých Markovových regresních modelů narazíme jsou parametry modelu. Parametry, které mají stejnou podobu jako doposud jsou:

- počáteční rozdělení pravděpodobnosti $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_N)$
- matice pravděpodobností přechodu $\mathbf{A} = (a_{ij})_{i,j \in S}$
- rozptyl $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$

Dalším parametrem modelu byla doposud střední hodnota $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)$. Jelikož nyní můžeme střední hodnotu v čase t a stavu i vyjádřit jako $\mathbf{X}_t' \boldsymbol{\beta}_i$, lze za nový parametr modelu (místo parametru $\boldsymbol{\mu}$) považovat matici regresních

koeficientů β , kterou můžeme zapsat následujícím způsobem

$$\beta = \begin{pmatrix} \beta_{01} & \beta_{11} & \cdots & \beta_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0N} & \beta_{1N} & \cdots & \beta_{pN} \end{pmatrix}.$$

Zápis modelu λ bude mít v případě skrytých Markovových regresních modelů následující podobu

$$\lambda = (\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2).$$

Dopředné ($\alpha_t(i)$), zpětné ($\varphi_t(i)$) a další pomocné proměnné ($\gamma_t(i), \delta_t(i)$) jsou vypočteny obdobně jako v případě skrytých Markovových řetězců bez použití regrese.

V případě úlohy učení, respektive odhadu nových parametrů modelu

$\hat{\lambda} = (\hat{\boldsymbol{\pi}}, \hat{\mathbf{A}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2)$, postupujeme tak, aby věrohodnost $P(\mathbf{Y}|\hat{\lambda})$ byla lokálně maximální. Věrohodnostní funkce, která je funkcí parametru λ bude mít tvar

$$L(\lambda) = \sum_{q_t \in S} \pi_{q_0} b_{q_0}(Y_0) \prod_{t=1}^T a_{q_{t-1}q_t} b_{q_t}(Y_t)$$

a odhad $\hat{\lambda}$ určíme jako $\hat{\lambda} = \arg \max_{\lambda} L(\lambda)$. Platí tedy, že $L(\hat{\lambda}) \geq L(\lambda)$. Odhady nových parametrů v případě Baum-Welchova algoritmu můžeme pro případ skrytých Markovových regresních modelů vyjádřit následovně

$$\begin{aligned} \hat{\pi}_i &= \gamma_0(i), \\ \hat{a}_{ij} &= \frac{\sum_{t=0}^{T-1} \xi_t(i, j)}{\sum_{t=0}^{T-1} \gamma_t(i)}, \\ \hat{\sigma}_i^2 &= \frac{\sum_{t=0}^T \gamma_t(i) (Y_t - \mathbf{X}_t' \hat{\boldsymbol{\beta}}_i)^2}{\sum_{t=0}^T \gamma_t(i)}, \end{aligned}$$

$$\hat{\boldsymbol{\beta}}_i = \begin{pmatrix} \omega_{x_0, x_0} & \omega_{x_0, x_1} & \cdots & \omega_{x_0, x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{x_p, x_0} & \omega_{x_p, x_1} & \cdots & \omega_{x_p, x_p} \end{pmatrix}^{-1} \times \begin{pmatrix} \omega_{x_0, y} \\ \vdots \\ \omega_{x_p, y} \end{pmatrix},$$

kde $\omega_{X_k, X_l} = \sum_{t=0}^T \alpha_t(i) \varphi_t(i) X_{tk} X_{tl}$, $\omega_{X_k, Y} = \sum_{t=0}^T \alpha_t(i) \varphi_t(i) X_{tk} Y_t$

a kde $k, l = 0, \dots, p$, $X_{t0} = 1$.

Pro výpočet parametrů $\hat{\beta}_i$ byla použita metoda vážených nejmenších čtverců s váhami odpovídajícími $\alpha_t(i) \varphi_t(i)$.

Kapitola 3

Aplikace skrytých Markovových řetězců

V následující části práce aplikujeme skryté Markovovy řetězce na reálný datový soubor. Datový soubor popisuje počet nově sjednaných hypotéčních úvěrů a míru nezaměstnanosti ve Spojeném království (UK), které sledujeme na měsíční bázi v časovém období od 01/2001 do 02/2020. Datový soubor má tedy 230 pozorování, která jsou popsána následujícími třemi proměnnými:

- *Date* : datum,
- *Mortgages* : počet nově sjednaných hypotéčních úvěrů v UK,
- *Unemployment* : míra nezaměstnanosti v UK.

Datový soubor byl vytvořen s použitím zdrojů [1], [14]. Hodnoty proměnných *Mortgages* a *Unemployment* jsou již očištěny od sezónních vlivů.

V následujících podkapitolách se pokusíme pomocí skrytých Markovových řetězců odhalit skryté stavy v ekonomice, a to prostřednictvím sledovaných ukazatelů. Jako sledované ukazatele jsou v práci zvoleny počty nově sjednaných hypotéčních úvěrů a míra nezaměstnanosti ve Spojeném království (UK) v daném časovém období, a to z důvodu jejich možné souvislosti s výkonem ekonomiky. V první podkapitole budou naše pozorování odpovídat počtu nově sjednaných hypotéčních úvěrů ve Spojeném království (UK) v daném časovém intervalu. V podkapitole druhé uvažujeme pozorování ve stejném časovém období, avšak

sledovanou veličinou bude nezaměstnanost ve Spojeném království. Ve třetí podkapitole aplikujeme skryté Markovovy regresní modely na zmíněná data, kdy budeme sledovat vztah mezi nezaměstnaností a nově sjednanými hypotéčními úvěry v UK. Pokusíme se tedy zjistit, zda se v ekonomice vyskytují nějaké skryté stavy, které by souvisely se vzájemným vztahem výše zmíněných proměnných. Na závěr zhodnotíme výsledky porovnáním s tempem růstu hrubého domácího produktu (HDP) v UK.

3.1. HMM - Počet nově sjednaných hypotéčních úvěrů

Jak již bylo zmíněno, v této části aplikujeme spojitě skryté Markovovy řetězce na datech nově sjednaných hypotéčních úvěrů.

Nejprve budeme uvažovat skrytý řetězec tvořený posloupností stavů, kde stavy systému nabývají dvou hodnot, které by dle předpokladu mohly odpovídat růstu, případně poklesu ekonomiky. Vzhledem k interpretaci stavů je vhodné provést transformaci dat následujícím způsobem:

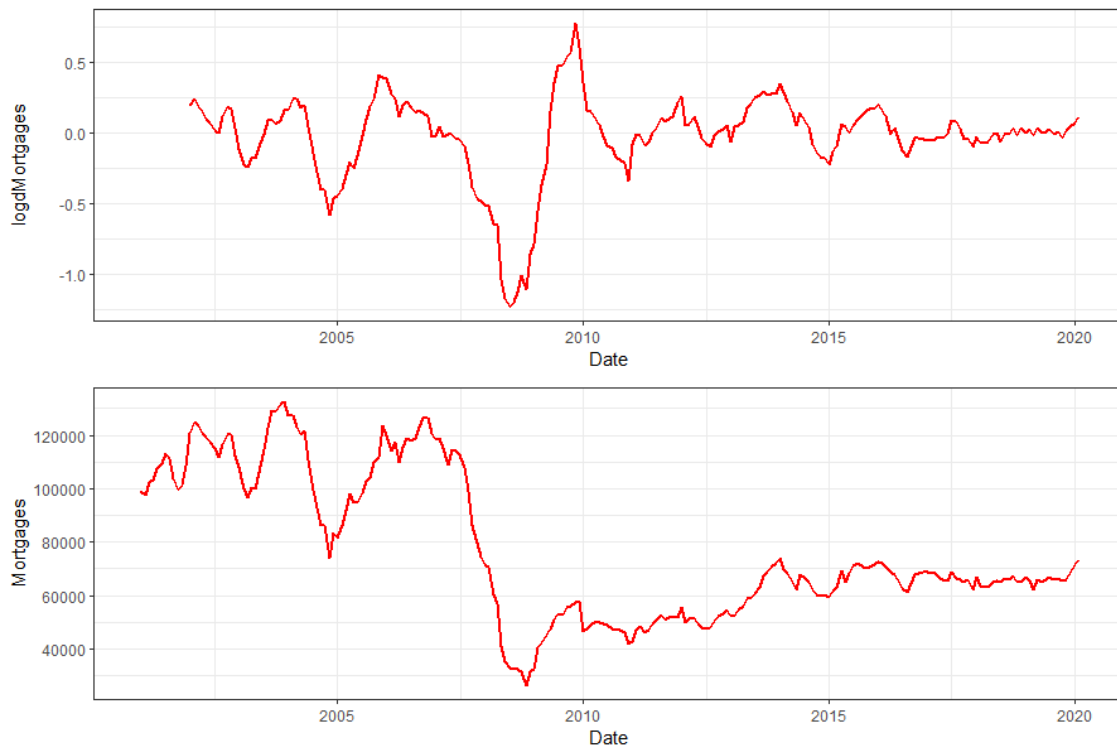
$$\text{logdMortgages}[t] = \log \frac{\text{Mortgages}[t]}{\text{Mortgages}[t - 12]}, \quad t = 12, \dots, 230$$

Tedy log-transformaci meziročního poměru nově sjednaných hypoték. Použití log-transformace je vhodné vzhledem k nově stanovenému poměrovému vyjádření. Interpretace nových hodnot je tedy následující:

- $\text{logdMortgages}[t] = 0$ v případě, že počet nově sjednaných hypoték v čase t je stejný jako v čase $t - 12$ (tj. meziročně je počet nově sjednaných hypoték stejný),
- $\text{logdMortgages}[t] > 0$ v případě, že počet nově sjednaných hypoték v čase t je větší než v čase $t - 12$ (tj. meziročně je počet nově sjednaných hypoték vyšší),

- $\text{logdMortgages}[t] < 0$ v případě, že počet nově sjednaných hypoték v čase t je menší než v čase $t - 12$ (tj. meziročně je počet nově sjednaných hypoték nižší).

Na obrázku 3.1 jsou zobrazeny hodnoty původní proměnné (*Mortgages*) a proměnné po transformaci (*logdMortgages*).



Obrázek 3.1: Hodnoty proměnné *logdMortgages* a proměnné *Mortgages*.

Nyní přejdeme k samotné aplikaci skrytých Markovových řetězců na transformovaných datech. Aplikaci provedeme v softwaru R použitím knihovny **HMM-Cont**. V první řadě vytvoříme model a provedeme nastavení jeho parametrů pomocí funkce *hmmsetcont*. Nejprve parametry stanovíme expertně dle očekávání. Předpokládáme, že skrytý stav 1 bude odpovídat růstu ekonomiky, naopak skrytý stav 2 poklesu ekonomiky. Počáteční odhady parametrů jsou stanoveny expertně tak, že v počátečním čase jsou oba stavy stejně pravděpodobné. Dále je pravděpodobnější, že systém v dalším časovém období setrvá v aktuálním stavu, než přechod do stavu jiného s tím, že pro období růstu ekonomiky je tato prav-

děpodobnost ještě vyšší než pro období poklesu. Co se týče středních hodnot sledované proměnné, předpokládáme v období růstu kladnou střední hodnotu a naopak v období poklesu hodnotu zápornou. Rozptyl je stanoven v případě obou stavů shodně. Nastavení počátečních hodnot parametrů je zobrazeno v tabulce 3.1.

Následně pomocí funkce *baumwelchcont* provedeme jednu iteraci Baum-Welchova

$\hat{\pi}_1$	$\hat{\pi}_2$	\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{21}	\hat{a}_{22}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
0,5	0,5	0,8	0,2	0,35	0,65	0,5	-0,5	0,5	0,5

Tabulka 3.1: Počáteční odhady parametrů stanovené na základě expertního odhadu.

algoritmu a odhadneme nové parametry modelu. Dále odhady iterativně zpřesňujeme, dokud nebude splněno ukončovacím kritérium. Ukončovacím kritériem v našem případě bude jednak maximální počet iterací I_{max} , který nastavíme na 50 a dále pak minimální rozdíl Akaikeho kritérií (AIC) vypočtených ve dvou po sobě jdoucích iteracích, který stanovíme jako 0,1. (Toto nastavení se na základě provedení více výpočtů s různými hodnotami ukončovacího kritéria projevilo jako dostačující). Výsledky jednotlivých iterací Baum-Welchova algoritmu jsou popsány následujícím kódem, kde po osmi iteracích získáme výsledné odhady parametrů. Výsledné odhady parametrů můžeme interpretovat tak, že v počátečním čase bude systém ve stavu 1 a s největší pravděpodobností (0,99) bude ve stavu 1 i v dalším časovém období. Stejně tak pokud se systém dostane do stavu 2, tak s vysokou pravděpodobností (0,94) bude ve stavu 2 i v dalším časovém období. Hodnoty pozorované proměnné, které přísluší stavu 1 se řídí normálním rozdělením se střední hodnotou 0,05 a rozptylem 0,02. Naopak hodnoty, které přísluší stavu 2 mají nižší střední hodnotu, ale vyšší rozptyl. Řídí se tedy normálním rozdělením se střední hodnotou -0,34 a rozptylem 0,29.


```

> hmmM<-baumwelchcont(hmmM)
> hmmM<-baumwelchcont(hmmM)
> i=1
> while((abs(hmmM$Results[i+1,4]-hmmM$Results[i,4])>0.1) & (i<50))
+ {hmmM=baumwelchcont(hmmM)
+   i=i+1}
> print(hmmM)

```

The number of Baum-welch iterations: 8

The parameters accumulated so far:

	Pi1	Pi2	A11	A12	A21	A22	Mu1	Mu2	Var1	Var2
[1,]	0.50	0.50	0.80	0.20	0.35	0.65	0.50	-0.50	0.50	0.50
[2,]	0.67	0.33	0.82	0.18	0.28	0.72	0.08	-0.19	0.04	0.13
[3,]	0.92	0.08	0.91	0.09	0.24	0.76	0.07	-0.29	0.02	0.20
[4,]	0.99	0.01	0.96	0.04	0.13	0.87	0.06	-0.32	0.02	0.25
[5,]	1.00	0.00	0.98	0.02	0.09	0.91	0.05	-0.31	0.02	0.27
[6,]	1.00	0.00	0.98	0.02	0.07	0.93	0.05	-0.32	0.02	0.28
[7,]	1.00	0.00	0.99	0.01	0.06	0.94	0.05	-0.33	0.02	0.29
[8,]	1.00	0.00	0.99	0.01	0.06	0.94	0.05	-0.34	0.02	0.29
[9,]	1.00	0.00	0.99	0.01	0.06	0.94	0.05	-0.34	0.02	0.29

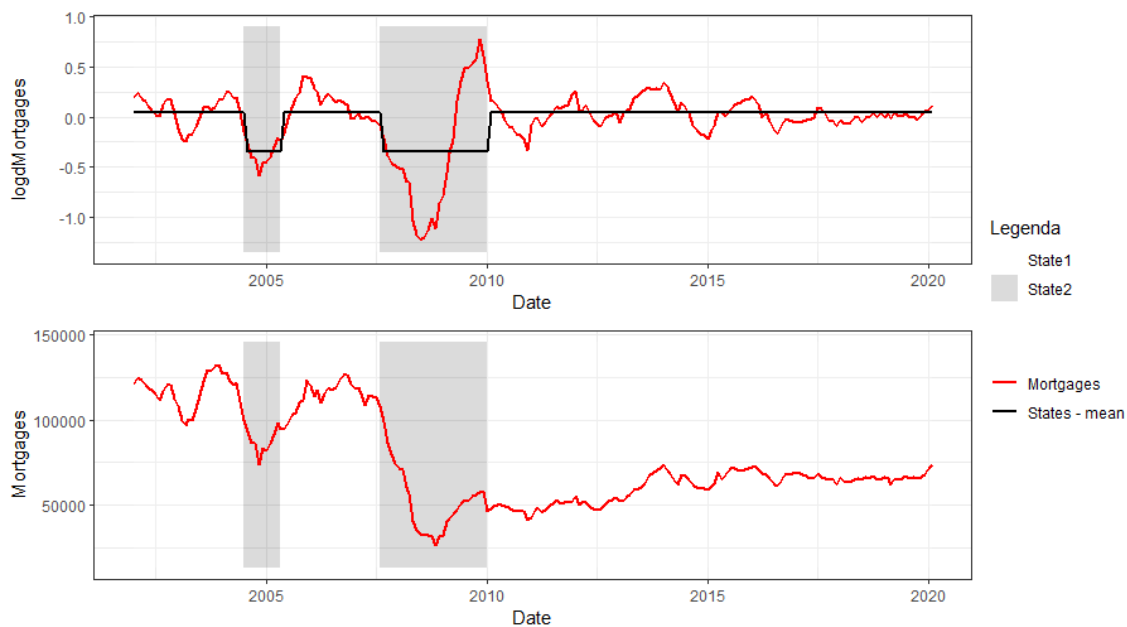
The results accumulated so far:

	P	AIC	SBIC
[1,]	2.564355e-80	386.530201	474.220102
[2,]	7.410158e+03	2.178786	89.868687
[3,]	1.009786e+20	-72.122880	15.567021
[4,]	4.758084e+24	-93.643775	-5.953873
[5,]	2.871500e+25	-97.238924	-9.549023
[6,]	5.081858e+25	-98.380609	-10.690707
[7,]	6.069034e+25	-98.735654	-11.045752
[8,]	6.291148e+25	-98.807542	-11.117641

The viterbi algorithm was not yet executed

Nyní konečně můžeme přejít k samotnému určení optimální posloupnosti stavů. K tomu nám slouží Viterbiho algoritmus, který aplikujeme pomocí funkce *viterbicont*. Výsledky si nejlépe zobrazíme vykreslením grafu, který lze pozorovat na obrázku 3.2. Z prvního grafu je patrné, že Viterbiho algoritmus odhalil dvě časová období, ve kterých se proměnná *logdMortgages* vyvíjí odlišně na rozdíl od období ostatních, což je znázorněno odlišnou barvou pozadí grafu. Dále je na grafu znázorněna střední hodnota μ_i příslušející danému stavu $i = 1, 2$ (States - mean). Na grafu druhém je zobrazen stejný výsledek Viterbiho algoritmu, avšak vzhledem k původní proměnné *Mortgages*. Pro přehlednost si dále výsledky Viterbiho algoritmu zaznamenáme do tabulky 3.2.

Pokud se nespokojíme s expertním odhadem počátečních parametrů modelu, lze dle [7] provést odhad například pomocí metody K-průměrů (*K-means*). Jedná se o metodu shlukové analýzy a tedy jejím cílem je rozdělit pozorování do K shluků tak, aby byla co nejméně vzdálena od středu shluku, který odpovídá průměru pozorování v daném shluku. Pokud nemáme počet shluků (K) stanoven expertně, můžeme jej určit například pomocí siluetového grafu, který nalezneme



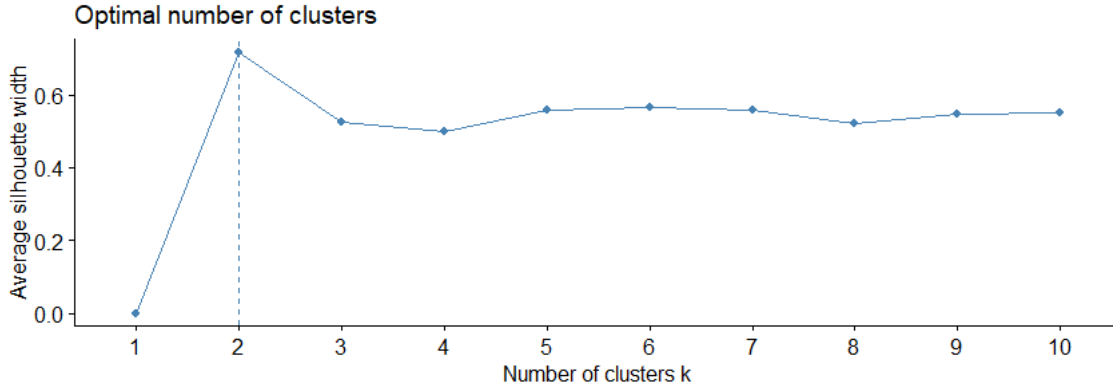
Obrázek 3.2: Optimální stavy zobrazené na hodnotách proměnné *logdMortgages* a proměnné *Mortgages*.

Date		State
Start	End	
2002-01-01	2004-07-01	1
2004-08-01	2005-05-01	2
2005-06-01	2007-08-01	1
2007-09-01	2010-01-01	2
2010-02-01	2020-02-01	1

Tabulka 3.2: Zápis výsledků Viterbiho algoritmu.

v knihovně **factoextra** [11] jako funkci *fviz_nbclust*. Siluetový graf je zobrazen na obrázku 3.3. Optimální počet shluků dle siluetového grafu je $K = 2$. Tedy co se týče počtu shluků, byl expertní odhad shodný a opět zvolíme počet shluků roven dvěma.

Výsledek metody K-průměrů zobrazuje obrázek 3.4. Na základě zařazení pozorování do shluků můžeme určit počáteční odhady parametrů skrytých Marko-



Obrázek 3.3: Siluetový graf proměnné *logdMortgages* při použití metody K-průměrů.

vových řetězců, které jsou dány následujícími vzorci

$$\hat{\pi}_i = \frac{\text{počet výskytů ve stavu } i \text{ v čase } 0}{\text{počet výskytů v čase } 0},$$

$$\hat{a}_{ij} = \frac{\text{počet přechodů ze stavu } i \text{ do stavu } j}{\text{počet přechodů ze stavu } i},$$

$$\hat{\mu}_i = \frac{\text{součet hodnot pozorování ve stavu } i}{\text{počet pozorování ve stavu } i},$$

$$\hat{\sigma}_i^2 = \frac{\text{součet kvadrátů odchylek hodnot pozorování ve stavu } i \text{ od } \hat{\mu}_i}{(\text{počet pozorování ve stavu } i) - 1},$$

kde $i=1, 2$. Hodnoty těchto počátečních odhadů parametrů v případě jejich stanovení pomocí metody K-průměrů jsou zobrazeny v tabulce 3.3.

$\hat{\pi}_1$	$\hat{\pi}_2$	\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{21}	\hat{a}_{22}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
1	0	0,98	0,02	0,12	0,88	0,06	-0,68	0,03	0,09

Tabulka 3.3: Počáteční odhady parametrů stanovené pomocí metody K-průměrů.

Po aplikaci skrytých Markovových řetězců získáme po třech iteracích Baum-Welchova algoritmu následující hodnoty parametrů:

```

> hmmM<-baumwelchcont(hmmM)
> hmmM<-baumwelchcont(hmmM)
> i=1
> while((abs(hmmM$Results[i+1,4]-hmmM$Results[i,4])>0.1) & (i<50))
+ {hmmM=baumwelchcont(hmmM)
+   i=i+1}
> print(hmmM)

```

The number of Baum-welch iterations: 3

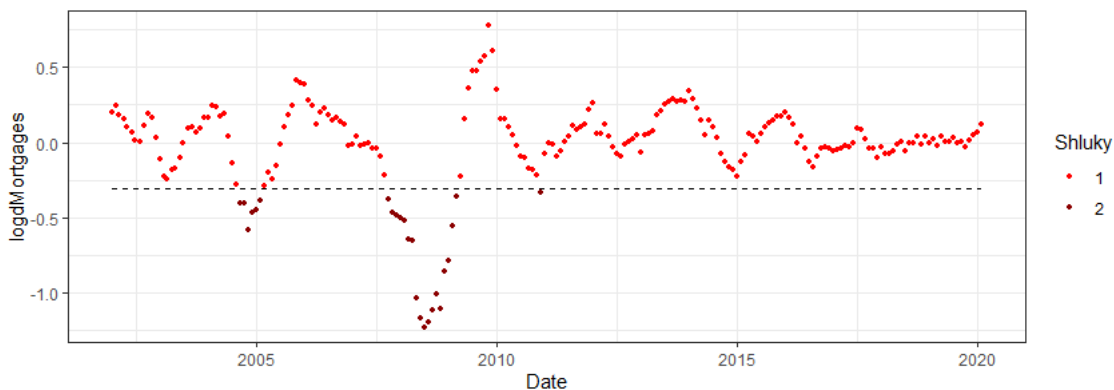
The parameters accumulated so far:

	Pi1	Pi2	A11	A12	A21	A22	Mu1	Mu2	Var1	Var2
[1,]	1	0	0.98	0.02	0.12	0.88	0.06	-0.68	0.03	0.09
[2,]	1	0	0.99	0.01	0.08	0.92	0.06	-0.65	0.03	0.10
[3,]	1	0	0.99	0.01	0.07	0.93	0.06	-0.64	0.03	0.10
[4,]	1	0	0.99	0.01	0.07	0.93	0.06	-0.63	0.03	0.10

The results accumulated so far:

	P	AIC	SBIC
[1,]	1.405244e+21	-77.38900	10.300905
[2,]	3.042887e+21	-78.93419	8.755714
[3,]	3.183469e+21	-79.02452	8.665384

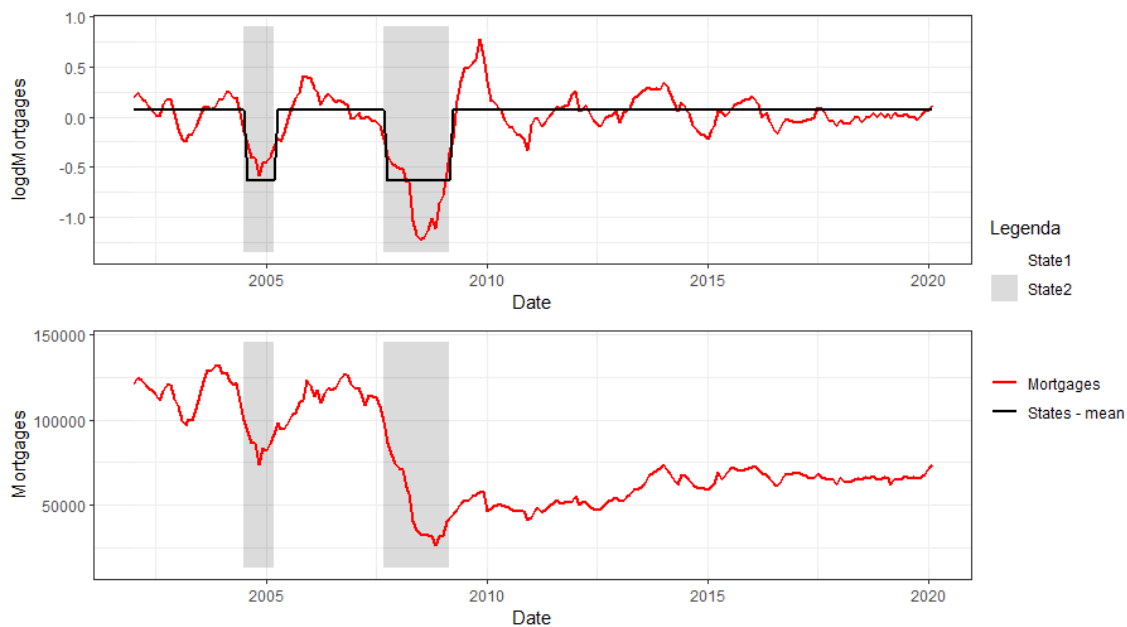
The viterbi algorithm was not yet executed



Obrázek 3.4: Přiřazení pozorování proměnné *logdMortgages* do shluků metodou K-průměrů.

Výsledek Viterbiho algoritmu při počátečních odhadech parametrů určených metodou K-průměrů zobrazuje obrázek 3.5.

Jak je na výsledcích jednotlivých iterací Baum-Welchova algoritmu vidět, prostřednictvím metody K-průměrů jsme odhadli počáteční parametry modelu poměrně přesně a Baum-Welchův algoritmus tedy zkonvergoval již po třech iteracích. Pokud si opět výsledky Viterbiho algoritmu zaznamenáme do tabulky 3.4, lze pozorovat mírnou změnu oproti předchozím výsledkům. Období, ve kterém se systém nachází ve stavu 2 se více zúžilo a porovnáním grafů 3.2 a 3.4 lze říci, že v přístupu, ve kterém jsou počáteční parametry modelu odhadnuty metodou K-



Obrázek 3.5: Optimální stavy zobrazené na hodnotách proměnné $\log dMortgages$ a proměnné $Mortgages$ při počátečních odhadech parametrů metodou K-průměrů.

průměrů, stav 2 více reflektuje výrazný meziroční pokles počtu nově sjednaných hypotéčních úvěrů a tudíž se můžeme domnívat, že by mohl souviset s poklesem ekonomiky.

Date		State
Start	End	
2002-01-01	2004-07-01	1
2004-08-01	2005-03-01	2
2005-04-01	2007-09-01	1
2007-10-01	2009-03-01	2
2009-04-01	2020-02-01	1

Tabulka 3.4: Zápis výsledků Viterbiho algoritmu při počátečních odhadech parametrů metodou K-průměrů.

3.2. HMM - Míra nezaměstnanosti

Stejně jako v předchozí kapitole budeme aplikovat skryté Markovovy modely na datech z reálného datového souboru. Jak již název kapitoly napovídá, budou naším pozorováním odpovídat hodnoty proměnné *Unemployment*. Opět provedeme transformaci dat následujícím způsobem:

$$\text{logdUnemployment}[t] = \log \frac{\text{Unemployment}[t]}{\text{Unemployment}[t - 12]}, \quad t = 12, \dots, 230$$

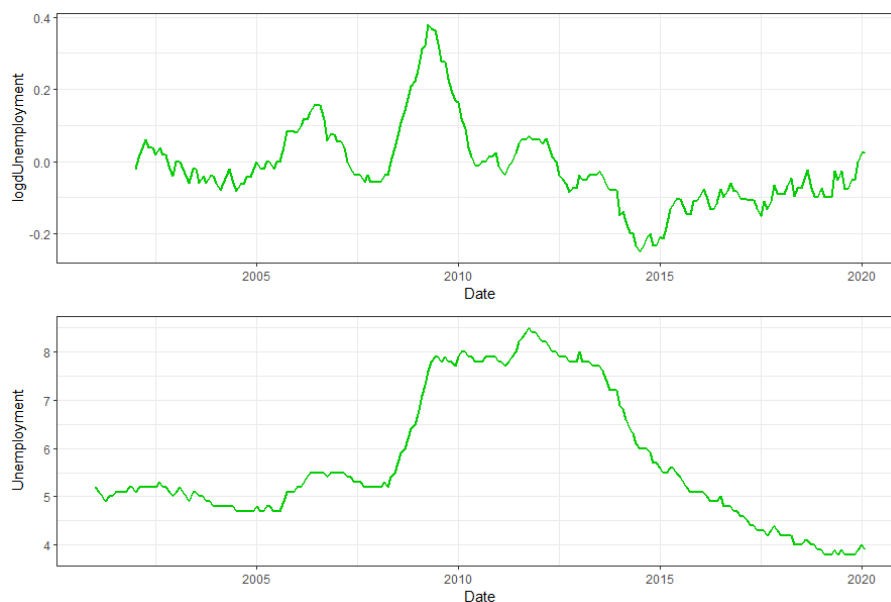
Interpretace nových hodnot je tedy následující:

- $\text{logdUnemployment}[t] = 0$ v případě, že míra nezaměstnanosti v čase t je stejná jako v čase $t - 12$,
- $\text{logdUnemployment}[t] > 0$ v případě, že míra nezaměstnanosti v čase t je větší než v čase $t - 12$,
- $\text{logdUnemployment}[t] < 0$ v případě, že míra nezaměstnanosti v čase t je menší než v čase $t - 12$.

Dále si graficky zobrazíme hodnoty proměnných *Unemployment* a *logdUnemployment*, což můžeme sledovat na obrázku 3.6. Počáteční odhady parametrů modelu opět stanovíme nejprve expertně a to obdobným způsobem jako v případě, kdy jsme sledovali hodnoty proměnné *logdMortgages*. Rozdíl spočívá pouze ve stanovení středních hodnot sledované proměnné *logdUnemployment*, kdy předpokládáme v období růstu zápornou střední hodnotu a naopak v období poklesu hodnotu kladnou. Hodnoty expertních odhadů parametrů jsou zobrazeny v tabulce 3.5.

$\hat{\pi}_1$	$\hat{\pi}_2$	\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{21}	\hat{a}_{22}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
0,5	0,5	0,8	0,2	0,35	0,65	-0,5	0,5	0,5	0,5

Tabulka 3.5: Počáteční odhady parametrů stanovené na základě expertního odhadu.



Obrázek 3.6: Hodnoty proměnné *logdUnemployment* a proměnné *Unemployment*.

Jednotlivé iterace Baum-Welchova algoritmu jsou zobrazeny v následujícím kódu, kde bylo použito stejné ukončovací kritérium jako dříve (tj. maximální počet iterací a rozdíl Akaikeho kritérií).

```
> hmmU<-baumwelchcont(hmmU)
> hmmU<-baumwelchcont(hmmU)
> i=1
> while((abs(hmmU$Results[i+1,4]-hmmU$Results[i,4])>0.1) & (i<50))
+ {hmmU=baumwelchcont(hmmU)
+ i=i+1}
> print(hmmU)
```

The number of Baum-welch iterations: 7

The parameters accumulated so far:

	Pi1	Pi2	A11	A12	A21	A22	Mu1	Mu2	Var1	Var2
[1,]	0.50	0.50	0.80	0.20	0.35	0.65	-0.50	0.50	0.50	0.50
[2,]	0.50	0.50	0.81	0.19	0.34	0.66	-0.04	0.02	0.01	0.02
[3,]	0.58	0.42	0.87	0.13	0.28	0.72	-0.05	0.07	0.01	0.02
[4,]	0.74	0.26	0.94	0.06	0.16	0.84	-0.06	0.11	0.00	0.02
[5,]	0.78	0.22	0.97	0.03	0.08	0.92	-0.07	0.13	0.00	0.01
[6,]	0.74	0.26	0.98	0.02	0.06	0.94	-0.07	0.13	0.00	0.01
[7,]	0.69	0.31	0.98	0.02	0.06	0.94	-0.07	0.13	0.00	0.01
[8,]	0.63	0.37	0.98	0.02	0.06	0.94	-0.07	0.13	0.00	0.01

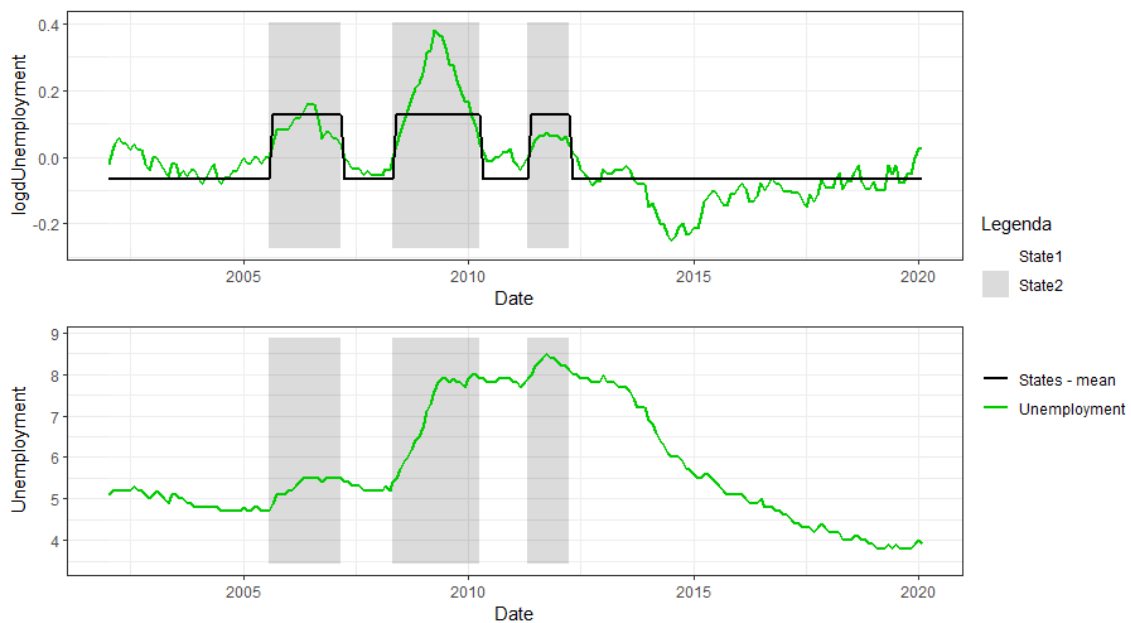
The results accumulated so far:

	P	AIC	SBIC
[1,]	5.308219e-78	375.8648	463.5547
[2,]	9.947462e+81	-357.6134	-269.9235
[3,]	4.399532e+96	-425.0593	-337.3694
[4,]	2.585174e+105	-465.4425	-377.7526
[5,]	9.534121e+108	-481.8681	-394.1782
[6,]	1.245639e+109	-482.4028	-394.7129
[7,]	1.263182e+109	-482.4308	-394.7409

The viterbi algorithm was not yet executed

Je důležité poznamenat, že rozptyl hodnot stavu 1 není, jak by se dle výsledků Baum-Welchova algoritmu mohlo zdát, nulový. Je však vzhledem k použitému měřítku velmi nízký (0,0038). Dle výsledků Baum-Welchova algoritmu můžeme říci, že stavu 1 odpovídají spíše nižší hodnoty proměnné $\log dUnemployment$ ($\mu_1 = -0,07$). Naopak stavu 2 odpovídají hodnoty spíše vyšší ($\mu_2 = 0,13$).

Výsledek Viterbiho algoritmu, který zobrazuje obrázek 3.7, určil optimální posloupnost stavů tak, že ve sledovaném časovém úseku odhalil tři období, kdy se systém nachází ve stavu 2, kterému odpovídají vysoké hodnoty meziročního nárůstu nezaměstnanosti. Tato období jsou detailněji zobrazena v tabulce 3.6

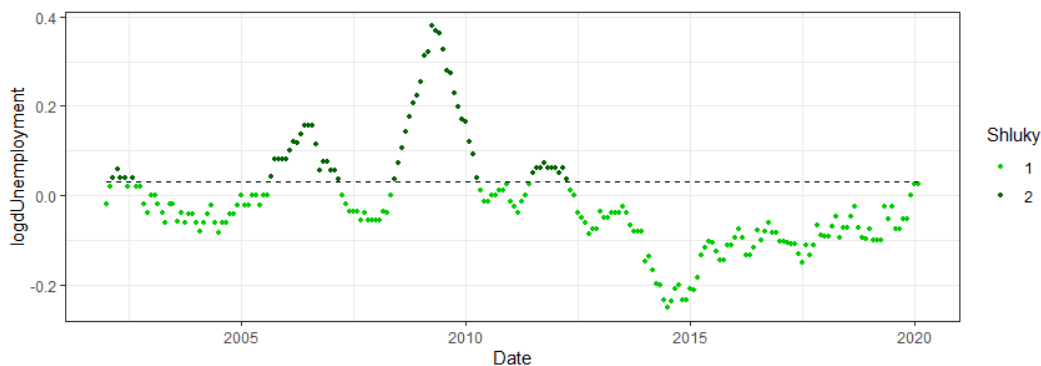


Obrázek 3.7: Optimální stavy zobrazené na hodnotách proměnné $\log dUnemployment$ a proměnné $Unemployment$.

Date		State
Start	End	
2002-01-01	2005-08-01	1
2005-09-01	2007-03-01	2
2007-04-01	2008-05-01	1
2008-06-01	2010-04-01	2
2010-05-01	2011-06-01	1
2011-07-01	2012-04-01	2
2012-05-01	2020-02-01	1

Tabulka 3.6: Zázpis výsledků Viterbiho algoritmu.

Pokud pro počáteční odhad parametrů modelu použijeme metodu K-průměrů a počet stavů určíme prostřednictvím siluetového grafu (zobrazen na obrázku 3.9), získáme opět optimální počet stavů roven dvěma. Shluky vytvořené na základě metody K-průměrů jsou vyobrazeny na obrázku 3.8. Počáteční odhady



Obrázek 3.8: Přiřazení pozorování proměnné $\log dUnemployment$ do shluků metodou K-průměrů.

parametrů v případě použití metody K-průměrů jsou zobrazeny v tabulce 3.7.

$\hat{\pi}_1$	$\hat{\pi}_2$	\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{21}	\hat{a}_{22}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
1	0	0,97	0,03	0,09	0,91	-0,07	0,13	0,004	0,01

Tabulka 3.7: Počáteční odhady parametrů stanovené pomocí metody K-průměrů.

Odhady parametrů modelu při jednotlivých iteracích Baum-Welchova algoritmu jsou zobrazeny níže. Kromě počátečního rozdělení pravděpodobnosti zkonvergoval Baum-Welchův algoritmus k téměř shodnému řešení jako v případě, kdy

byly počáteční odhady parametrů stanoveny expertně.

```
> hmmU<-baumwelchcont(hmmU)
> hmmU<-baumwelchcont(hmmU)
> i=1
> while((abs(hmmU$Results[i+1,4]-hmmU$Results[i,4])>0.1) & (i<50))
+ {hmmU=baumwelchcont(hmmU)
+ i=i+1}
> print(hmmU)
```

The number of Baum-welch iterations: 4

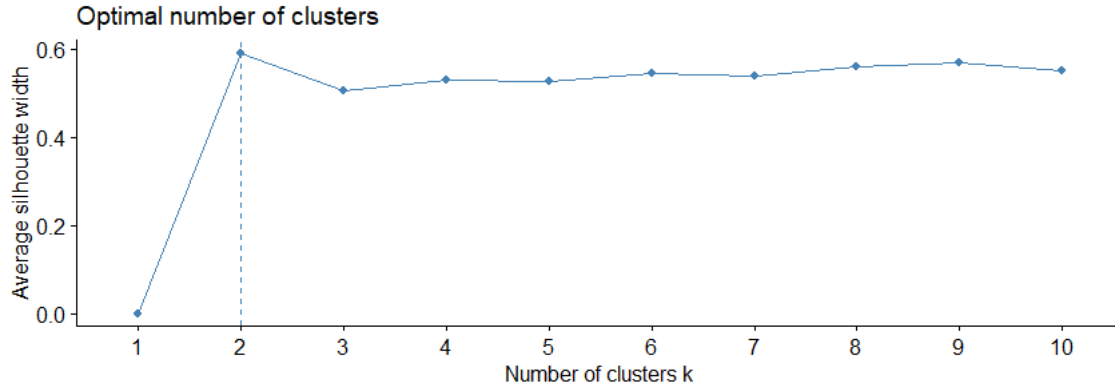
The parameters accumulated so far:

	Pi1	Pi2	A11	A12	A21	A22	Mu1	Mu2	Var1	Var2
[1,]	1	0	0.97	0.03	0.09	0.91	-0.07	0.13	0	0.01
[2,]	1	0	0.98	0.02	0.07	0.93	-0.07	0.13	0	0.01
[3,]	1	0	0.98	0.02	0.06	0.94	-0.07	0.13	0	0.01
[4,]	1	0	0.98	0.02	0.06	0.94	-0.06	0.13	0	0.01
[5,]	1	0	0.98	0.02	0.06	0.94	-0.06	0.13	0	0.01

The results accumulated so far:

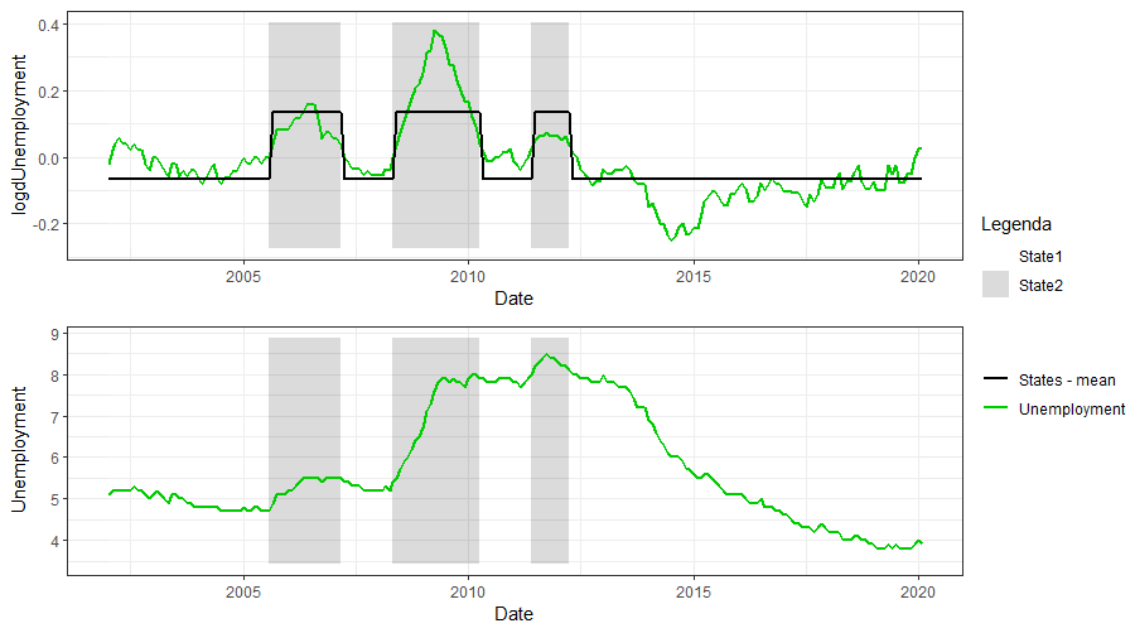
	P	AIC	SBIC
[1,]	6.095399e+108	-480.9734	-393.2835
[2,]	1.168123e+109	-482.2743	-394.5844
[3,]	1.298136e+109	-482.4854	-394.7955
[4,]	1.361711e+109	-482.5810	-394.8911

The viterbi algorithm was not yet executed



Obrázek 3.9: Siluetový graf proměnné *logdUnemployment* při použití metody K-průměrů.

Graf zobrazující výsledek Viterbiho algoritmu 3.10 i souhrnná tabulka 3.8 jsou téměř shodné jako při použití expertního odhadu pro stanovení počátečních parametrů.



Obrázek 3.10: Optimální stavy zobrazené na hodnotách proměnné *logdUnemployment* a proměnné *Unemployment* při počátečních odhadech parametrů metodou K-průměrů.

Date		State
Start	End	
2002-01-01	2005-08-01	1
2005-09-01	2007-03-01	2
2007-04-01	2008-05-01	1
2008-06-01	2010-04-01	2
2010-05-01	2011-05-01	1
2011-06-01	2012-04-01	2
2012-05-01	2020-02-01	1

Tabulka 3.8: Zápis výsledků Viterbiho algoritmu při počátečních odhadech parametrů metodou K-průměrů..

3.3. HMMR - Vztah nezaměstnanosti a počtu nově sjednaných hypotéčních úvěrů

V následující kapitole aplikujeme skryté Markovovy regresní modely na datový soubor, kdy budeme sledovat vztah mezi nezaměstnaností a počtem nově sjednaných hypotéčních úvěrů v UK. Konkrétně budeme sledovat proměnné $\log dMortgages$ a $\log dUnemployment$. Předpokládáme, že hodnoty proměnné $\log dMortgages$ lze vyjádřit prostřednictvím hodnot proměnné $\log dUnemployment$. Jejich vztah popíšeme lineárním regresním modelem, kde $\log dMortgages$ zastává roli závisle proměnné a $\log dUnemployment$ roli nezávisle proměnné. Ve skrytých Markovových regresních modelech předpokládáme, že každému stavu odpovídá jiný regresní model. Tedy regresní parametry se v závislosti na stavu liší. Odpovídající lineární regresní model můžeme tedy zapsat následujícím způsobem:

$$\log dMortgages_t = \beta_{0i} + \beta_{1i} \log dUnemployment_t + \epsilon_{ti}, \quad 0 \leq t \leq T, i \in S,$$

kde t značí čas, i stav a ϵ_{ti} chybu.

Nejprve budeme předpokládat dva stavy modelu. Pro počáteční odhad parametrů modelu použijeme metodu K-průměrů. Jelikož nás zajímá především vzájemný vztah proměnných, provedeme nejprve odhad regresních parametrů tak, abychom je mohli vyčíslit v každém čase t a následně je rozdělit do dvou shluků metodou K-průměrů. Odhad regresních parametrů v čase t provedeme po klouzavých ročních úsecích následujícím způsobem:

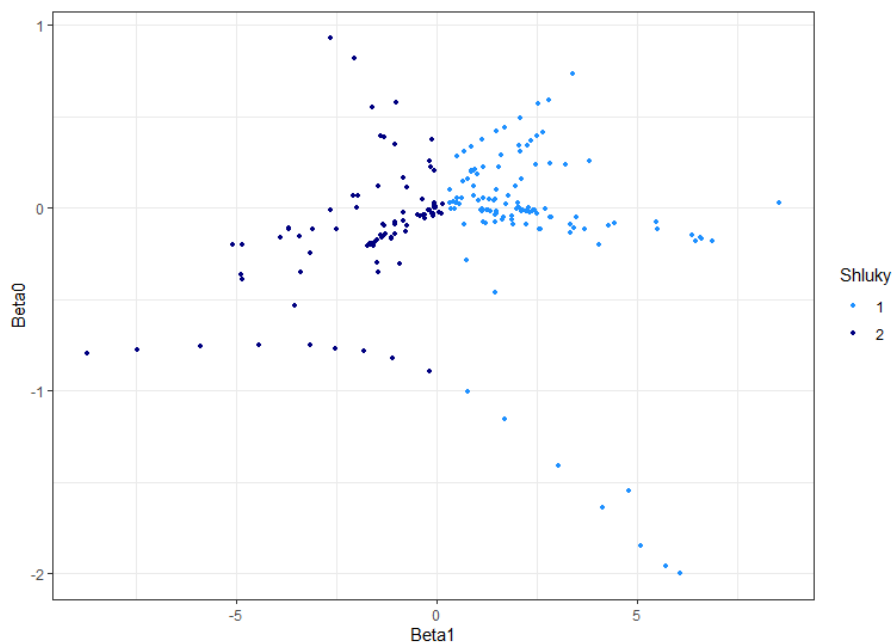
$$\hat{\beta}[h] \Leftarrow \log dMortgages_t = \beta_0 + \beta_1 \log dUnemployment_t + \epsilon_t, \quad h - 6 \leq t \leq h + 6,$$

kde $\hat{\beta}[h]$ značí odhad regresních koeficientů v h -tém časovém úseku a $h = 6, \dots, T - 6$.

Tedy odhadované regresní koeficienty v čase t odpovídají lineárnímu regresnímu modelu, který byl vytvořen z dat za období odpovídající intervalu $\langle t - 6, t + 6 \rangle$ a jsou odhadnuty metodou nejmenších čtverců. Jinými slovy, při konstrukci regresního modelu uvažujeme data, která jsou vzdálena nejvíce 6 měsíců od uvažovaného času t . Hodnoty regresních parametrů, které na začátku a na konci celkového

sledovaného období touto metodou nestanovíme, nahradíme konstantami, které budou odpovídat první (resp. poslední) dopočtené hodnotě.

Na takto určené hodnoty aplikujeme metodu K-průměrů, která rozdělí hodnoty regresních koeficientů do dvou shluků. Výsledek je zobrazen na obrázku 3.11. Dle vytvořených shluků určíme odhady počátečních parametrů skrytého Marko-



Obrázek 3.11: Přiřazení regresních koeficientů do shluků při použití metody K-průměrů.

vova regresního modelu. Odhady počátečních parametrů jsou zobrazeny v tabulce 3.9.

$\hat{\pi}_1$	$\hat{\pi}_2$	\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{21}	\hat{a}_{22}
1	0	0,92	0,08	0,11	0,89
$\hat{\beta}_{01}$	$\hat{\beta}_{11}$	$\hat{\beta}_{02}$	$\hat{\beta}_{12}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
0	-0,42	-0,05	0,63	0,07	0,11

Tabulka 3.9: Počáteční odhady parametrů stanovené pomocí metody K-průměrů.

Skryté Markovovy regresní modely již nejsou součástí knihovny **HMMCont**, tudíž jsou doposud použité funkce z této knihovny modifikovány tak, aby byly

vhodné i pro tento případ. Odhady počátečních parametrů modelu následně zpřesníme pomocí Baum-Welchova algoritmu. Výsledky jednotlivých iterací algoritmu zobrazuje následující kód:

```
> hmm<-baumwelchcontR(hmm)
> i=1
> while((abs(hmm$Results[i+1,4]-hmm$Results[i,4])>0.1) & (i<50))
+ {hmm=baumwelchcontR(hmm)
+ i=i+1}
> print(hmm)
```

The number of Baum-welch iterations: 8

The parameters accumulated so far:

	Pi1	Pi2	A11	A12	A21	A22	b01	b11	b02	b12	Var1	Var2
[1,]	1	0	0.92	0.08	0.11	0.89	0.00	-0.42	-0.05	0.63	0.07	0.11
[2,]	1	0	0.95	0.05	0.12	0.88	0.01	-0.34	-0.17	0.86	0.03	0.20
[3,]	1	0	0.98	0.02	0.08	0.92	0.05	0.26	-0.35	1.10	0.02	0.27
[4,]	1	0	0.98	0.02	0.07	0.93	0.07	0.45	-0.45	1.19	0.02	0.26
[5,]	1	0	0.98	0.02	0.07	0.93	0.07	0.49	-0.46	1.17	0.02	0.24
[6,]	1	0	0.98	0.02	0.07	0.93	0.07	0.51	-0.46	1.14	0.02	0.24
[7,]	1	0	0.98	0.02	0.07	0.93	0.08	0.51	-0.45	1.12	0.02	0.24
[8,]	1	0	0.98	0.02	0.07	0.93	0.08	0.52	-0.45	1.10	0.02	0.24
[9,]	1	0	0.98	0.02	0.08	0.92	0.08	0.52	-0.44	1.08	0.02	0.24

The results accumulated so far:

	P	AIC	SBIC
[1,]	9.401558e-15	88.59580	193.82368
[2,]	3.072604e+14	-42.71743	62.51045
[3,]	5.197267e+28	-108.24103	-3.01315
[4,]	2.512176e+30	-115.99740	-10.76952
[5,]	3.192741e+30	-116.47686	-11.24898
[6,]	3.448018e+30	-116.63070	-11.40282
[7,]	3.630041e+30	-116.73359	-11.50571
[8,]	3.753177e+30	-116.80031	-11.57243

The viterbi algorithm was not yet executed

Hodnoty proměnné $\log dMortgages_t$, které odpovídají stavu 1 se tedy řídí normálním rozdělením se střední hodnotou $0,08 + 0,52 \log dUnemployment_t$ a rozptylem $0,02$. Naopak hodnoty, které odpovídají stavu 2 se řídí normálním rozdělením se střední hodnotou $-0,44 + 1,08 \log dUnemployment_t$ a rozptylem $0,24$.

Následně určíme optimální posloupnost stavů pomocí Viterbiho algoritmu, která je zobrazena na obrázku 3.12. V případě skrytého Markovova regresního modelu odhalil Viterbiho algoritmus tři období, ve kterých je systém ve stavu 2. Podrobnější výsledky jsou zobrazeny v tabulce 3.10.

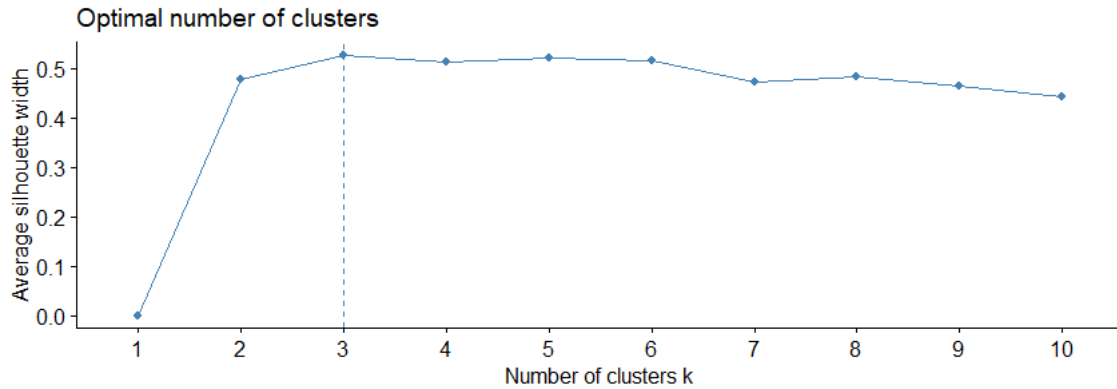


Obrázek 3.12: Optimální stavy zobrazené na hodnotách regresních koeficientů $\hat{\beta}_0$ a $\hat{\beta}_1$ při počátečních odhadech parametrů metodou K-průměrů.

Date		State
Start	End	
2002-01-01	2004-07-01	1
2004-08-01	2005-06-01	2
2005-07-01	2007-08-01	1
2007-09-01	2009-12-01	2
2010-01-01	2010-08-01	1
2010-09-01	2010-12-01	2
2011-01-01	2020-02-01	1

Tabulka 3.10: Zápís výsledků Viterbiho algoritmu při počátečních odhadech parametrů metodou K-průměrů.

V případě, že budeme chtít stanovit optimální počet stavů pomocí siluetového grafu 3.13, obdržíme stavy tři. V případě skrytého Markovovova regresního mo-



Obrázek 3.13: Siluetový graf v případě použití regresního modelu a metody K-průměrů.

delu s třemi stavy jsou počáteční odhady parametrů modelu stanovené pomocí metody K-průměrů zobrazeny v tabulce 3.11.

$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$						
0	1	0						
\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{13}	\hat{a}_{21}	\hat{a}_{22}	\hat{a}_{23}	\hat{a}_{31}	\hat{a}_{32}	\hat{a}_{33}
0,72	0,24	0,04	0,05	0,88	0,06	0	0,14	0,86
$\hat{\beta}_{01}$	$\hat{\beta}_{11}$	$\hat{\beta}_{02}$	$\hat{\beta}_{12}$	$\hat{\beta}_{03}$	$\hat{\beta}_{13}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$
-0,13	-1,35	0,05	0,25	-0,08	0,74	0,08	0,06	0,12

Tabulka 3.11: Počáteční odhady parametrů stanovené pomocí metody K-průměrů.

Pokud uvažujeme skrytý Markovův regresní model s třemi stavy získáme použitím Baum-Welchova algoritmu následující odhady parametrů:


```

> hmmR3c1<-baumwelchcontr3c1(hmmR3c1)
> hmmR3c1<-baumwelchcontr3c1(hmmR3c1)
> i=1
> while((abs(hmmR3c1$Results[i+1,4]-hmmR3c1$Results[i,4])>0.1) & (i<50))
+ {hmmR3c1=baumwelchcontr3c1(hmmR3c1)
+ i=i+1}
> print(hmmR3c1)

The number of Baum-welch iterations: 7
The parameters accumulated so far:
  Pi1 Pi2 Pi3  A11  A12  A13  A21  A22  A23  A31  A32  A33  b01  b11  b02  b12  b03  b13 Var1 Var2 Var3
[1,]  0   1   0  0.72  0.24  0.04  0.05  0.88  0.06  0  0.14  0.86 -0.13 -1.35  0.05  0.25 -0.08  0.74  0.08  0.06  0.12
[2,]  0   1   0  0.82  0.16  0.02  0.04  0.93  0.03  0  0.20  0.80 -0.37 -1.94  0.07  0.52 -0.07  1.78  0.14  0.02  0.07
[3,]  0   1   0  0.90  0.09  0.01  0.02  0.96  0.02  0  0.17  0.83 -0.49 -1.20  0.08  0.52 -0.09  2.41  0.13  0.01  0.04
[4,]  0   1   0  0.93  0.05  0.01  0.01  0.97  0.02  0  0.11  0.89 -0.61 -0.60  0.08  0.49 -0.11  2.80  0.12  0.01  0.03
[5,]  0   1   0  0.95  0.05  0.00  0.01  0.97  0.02  0  0.09  0.91 -0.69 -0.24  0.08  0.48 -0.12  2.97  0.11  0.01  0.03
[6,]  0   1   0  0.95  0.05  0.00  0.01  0.97  0.02  0  0.08  0.92 -0.71 -0.16  0.08  0.49 -0.12  3.04  0.11  0.01  0.03
[7,]  0   1   0  0.95  0.05  0.00  0.01  0.97  0.02  0  0.08  0.92 -0.71 -0.14  0.08  0.50 -0.12  3.09  0.11  0.01  0.03
[8,]  0   1   0  0.95  0.05  0.00  0.01  0.97  0.02  0  0.08  0.92 -0.71 -0.14  0.08  0.50 -0.12  3.12  0.11  0.01  0.03

The results accumulated so far:
      P      AIC      SBIC
[1,] 7.793269e-03  51.70899 235.85778
[2,] 4.229067e+33 -112.85458  71.29421
[3,] 4.831262e+41 -149.96219  34.18660
[4,] 2.686853e+44 -162.60423  21.54456
[5,] 1.103068e+45 -165.42885  18.71994
[6,] 1.325001e+45 -165.79549  18.35331
[7,] 1.370139e+45 -165.86248  18.28631

The Viterbi algorithm was not yet executed

```

Stavu 1 odpovídají hodnoty proměnné $\log dMortgages_t$, které se řídí normálním rozdělením se střední hodnotou $-0,71 - 0,14 \log dUnemployment_t$ a rozptylem 0,11. Hodnoty, které odpovídají stavu 2 se řídí normálním rozdělením se střední hodnotou $0,08 + 0,5 \log dUnemployment_t$ a rozptylem 0,01. Stav 3 odpovídají hodnoty, které se řídí normálním rozdělením se střední hodnotou $-0,12 + 3,12 \log dUnemployment_t$ a rozptylem 0,03. Poté, co na model s odhadnutými parametry aplikujeme Viterbiho algoritmus, získáme výstup o třech stavech, který lze pozorovat na obrázku 3.14

Vzhledem k výsledkům Viterbiho a Baum-Welchova algoritmu se můžeme domnívat, že v makroekonomickém kontextu by jednotlivé stavy mohly být klasifikovány následovně: Stav 1 nejspíše bude znázorňovat pokles ekonomiky, stav 2 stagnaci a na závěr stav 3 ekonomický růst. Při porovnání s modelem, ve kterém jsme uvažovali pouze dva stavy, můžeme říci, že model se třemi stavy lépe vystihuje data, jelikož i přes nárůst počtu parametrů modelu je AIC nižší.



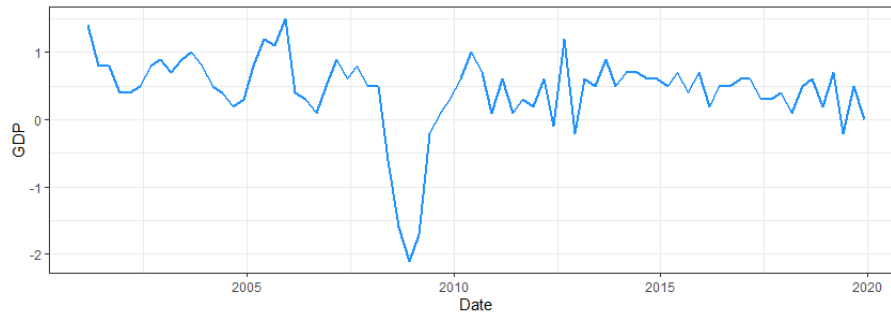
Obrázek 3.14: Optimální stavy zobrazené na hodnotách regresních koeficientů $\hat{\beta}_0$ a $\hat{\beta}_1$ při počátečních odhadech parametrů metodou K-průměrů.

3.4. Shrnutí a porovnání výsledků s tempem růstu HDP

V této závěrečné kapitole porovnáme dříve získané výsledky s tempem růstu HDP coby ukazatelem výkonnosti ekonomiky. Datový soubor obsahuje tempo růstu HDP v UK, které je vyjádřeno „Q-Q“, tedy kvartálně, v období od 12/2001 do 12/2019. Datový soubor byl vytvořen na základě dat získaných z [14]. Hodnoty tempa růstu HDP jsou již očištěny od sezónních vlivů. Proměnné v datovém souboru jsou tedy následující:

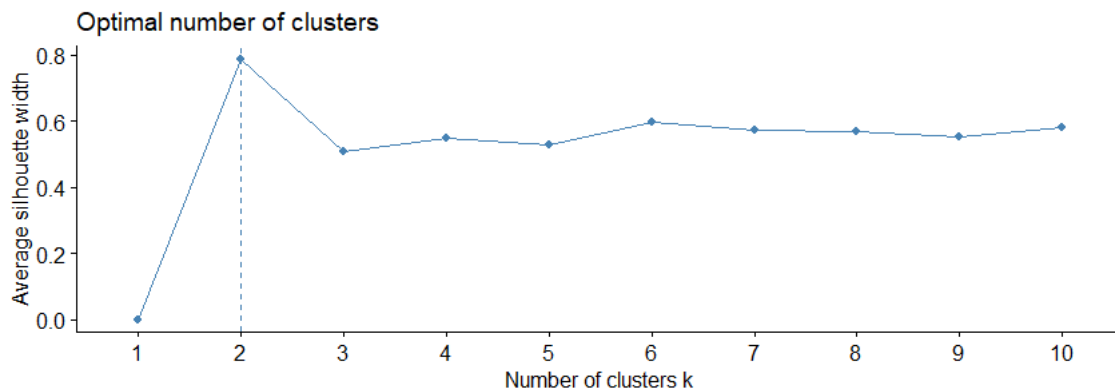
- *Date* : datum,
- *GDP* : HDP v UK.

Opět začneme tím, že si data zobrazíme graficky na obrázku 3.15. Již z grafu je patrný výrazný pokles tempa růstu v letech 2008 až 2009. Stejně jako dříve aplikujeme na datový soubor skryté Markovovy modely s cílem určit optimální posloupnost stavů, které budou symbolizovat období poklesu resp. růstu ekono-



Obrázek 3.15: Hodnoty proměnné *GDP*.

miky. Siluetovým grafem, který je zobrazen na obrázku 3.16, ověříme, že optimální počet stavů je opravdu dva. Počáteční odhad parametrů určíme na základě

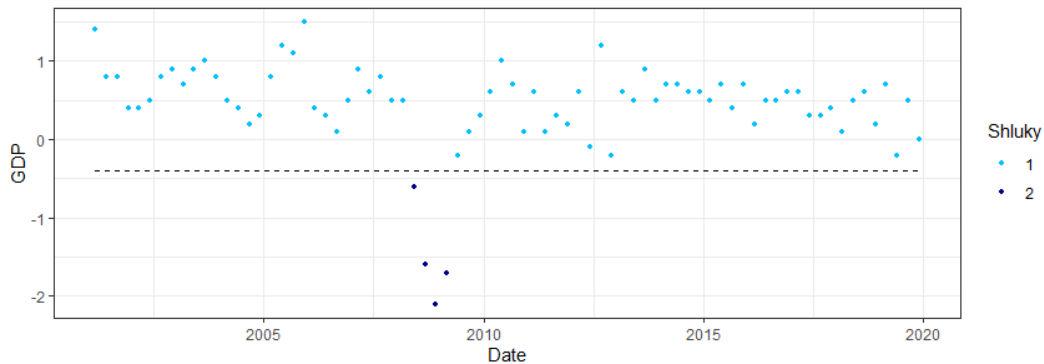


Obrázek 3.16: Siluetový graf v případě použití metody K-průměrů.

rozdělení dat do dvou shluků pomocí metody K-průměrů. Shluky jsou zobrazeny na obrázku 3.17. Pomocí funkcí z knihovny **HMMCont** upravíme parametry modelu a odhadneme optimální posloupnost stavů. Počáteční odhad parametrů modelu pomocí metody K-průměrů je zobrazen v tabulce 3.12.

$\hat{\pi}_1$	$\hat{\pi}_2$	\hat{a}_{11}	\hat{a}_{12}	\hat{a}_{21}	\hat{a}_{22}	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
1	0	0,99	0,01	0,25	0,75	0,53	-1,5	0,12	0,41

Tabulka 3.12: Počáteční odhady parametrů stanovené pomocí metody K-průměrů.



Obrázek 3.17: Přiřazení pozorování proměnné *GDP* do shluků metodou K-průměrů.

Odhad nových parametrů modelu provedeme Baum-Welchovým algoritmem při počátečních odhadech stanovených pomocí metody K-průměrů. Výsledky jsou zobrazeny následujícím kódem:

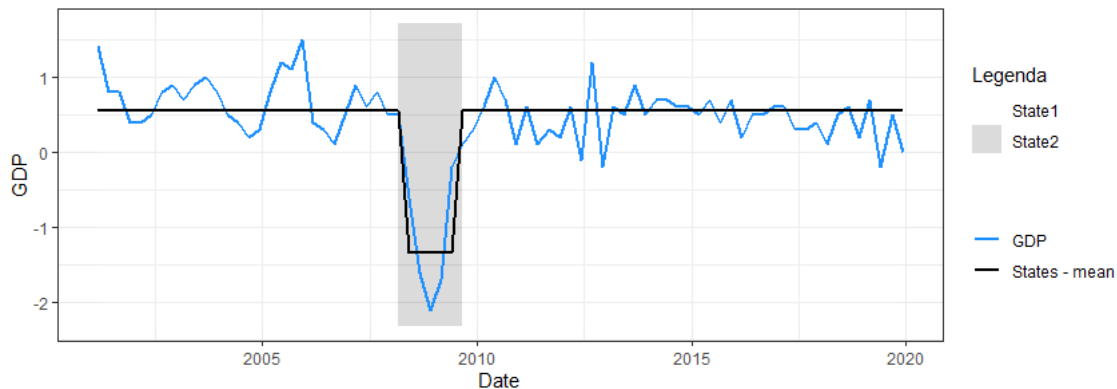
```
> hmmM<-baumwelchcont(hmmM)
> hmmM<-baumwelchcont(hmmM)
> i=1
> while((abs(hmmM$Results[i+1,4]-hmmM$Results[i,4])>0.1) & (i<50))
+ {hmmM=baumwelchcont(hmmM)
+ i=i+1}
> print(hmmM)

The number of Baum-welch iterations: 3
The parameters accumulated so far:
      Pi1 Pi2 A11 A12 A21 A22 Mu1 Mu2 Var1 Var2
[1,]  1  0 0.99 0.01 0.25 0.75 0.53 -1.50 0.12 0.41
[2,]  1  0 0.99 0.01 0.23 0.77 0.55 -1.40 0.12 0.41
[3,]  1  0 0.99 0.01 0.23 0.77 0.55 -1.36 0.12 0.45
[4,]  1  0 0.99 0.01 0.22 0.78 0.55 -1.33 0.11 0.47

The results accumulated so far:
      P      AIC      SBIC
[1,] 1.604896e-16 92.73660 159.3513
[2,] 1.749887e-16 92.56362 159.1783
[3,] 1.784266e-16 92.52471 159.1394

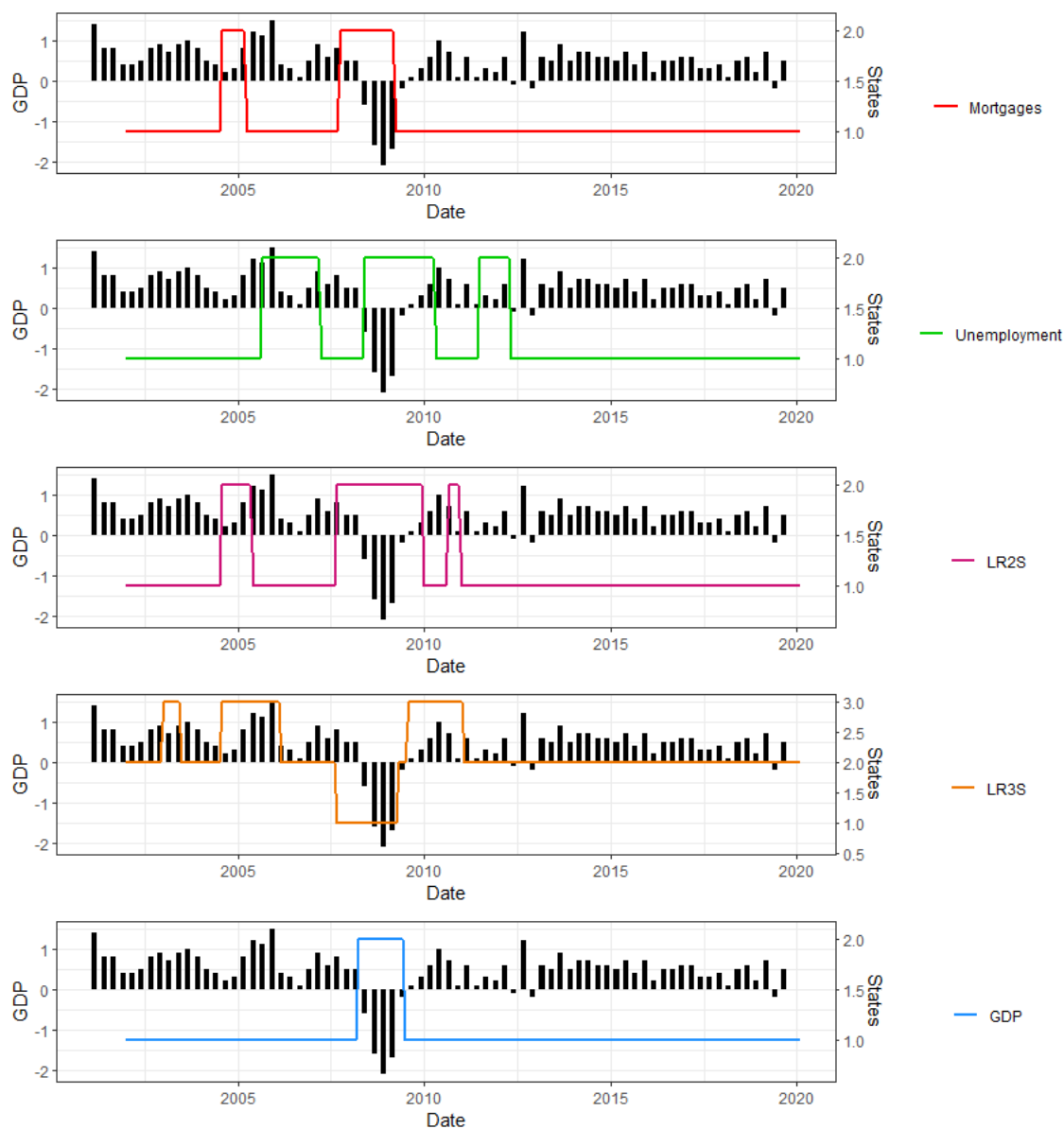
The viterbi algorithm was not yet executed
```

Do splnění ukončovacího kritéria byly provedeny tři iterace, avšak odhady parametrů se oproti počátečním odhadům příliš nezměnily. Pomocí odhadnutých parametrů určíme Viterbiho algoritmem optimální posloupnost stavů, která je zobrazena spolu s hodnotami proměnné *GDP* na obrázku 3.18. Jak jsme tedy tušili, období, kdy je systém ve stavu 2 (pokles ekonomiky), určil Viterbiho algoritmus mezi 2008-04-01 a 2009-06-01.



Obrázek 3.18: Optimální stavy zobrazené na hodnotách proměnné GDP při počátečních odhadech parametrů metodou K-průměrů.

Na závěr si pro lepší porovnání shrneme obdržené výsledky do jednoho obrázku (3.19). Na grafech pozorujeme vývoj tempa růstu HDP v čase (černě). Barevnou linkou jsou potom vyznačeny výsledné stavy v jednotlivých případech. Názvem LR2S zde zjednodušeně označujeme lineární regresní model z kapitoly 3.3., kdy uvažujeme řetězec se dvěma stavy. LR3S potom značí stejný případ, avšak se stavy třemi. Z grafů můžeme zaznamenat, že ve všech případech byl na přelomu let 2008 a 2009 identifikován stav, ve kterém se ekonomika vyvíjí odlišně od většiny sledovaného období. Tomuto období odpovídá období ekonomické krize, což lze pozorovat přímo na hodnotách GDP [14]. Dále byla identifikována další období, ve kterých dle srovnání s proměnnou GDP bylo nižší tempo růstu HDP než v předchozích obdobích, avšak k poklesu ekonomiky nedošlo. Naopak stav 3 u proměnné označené LR3S odhalil období, ve kterých bylo tempo růstu HDP převážně vysoké.



Obrázek 3.19: Jednotlivé výsledky optimálních stavů získaných Viterbiho algoritmem zobrazené na hodnotách proměnné GDP při počátečních odhadech parametrů metodou K-průměrů.

Závěr

Cílem diplomové práce bylo čtenáře seznámit s teorií skrytých Markovových řetězců se zaměřením na jejich spojitý případ.

Práce se skládá z teoretické a praktické části. V první kapitole byl čtenář seznámen se základními pojmy teorie Markovových řetězců, jejichž znalost je nezbytná pro pochopení teorie skrytých Markovových řetězců. Kapitola druhá byla věnována samotné teorii skrytých Markovových řetězců, která byla doplněna názornými, avšak jednoduchými příklady. Součástí druhé kapitoly bylo taktéž rozšíření teorie na případ skrytých Markovových regresních modelů, které umožňují další zajímavá využití.

Třetí a poslední kapitola odpovídá praktické části práce, ve které jsme aplikovali získané znalosti na reálný datový soubor vybraných ekonomických ukazatelů vykazovaných ve Spojeném království s cílem odhalit skryté stavy britské ekonomiky. Výsledky dle jednotlivých ukazatelů a použitých metod se lehce lišily, ovšem ve všech případech bylo modelem odhaleno období ekonomické krize v letech 2008 až 2009. Při porovnání výsledků s hodnotami tempa růstu HDP jsme shledali možnou podobu.

Z mého pohledu vidím přínos práce především v získání znalostí této teorie, která nachází spoustu zajímavých uplatnění. Dalším velkým přínosem a výzvou pro mě bylo vlastní zpracování algoritmů, které nebyly součástí jinak používané knihovny **HMMCont**, v softwaru R.

Literatura

- [1] Bank of England: *Bank of England Statistical Interactive Database* [online]. 2020, [cit. 2020-06-09]. dostupné z: <http://www.bankofengland.co.uk/boeapps/iadb/>.
- [2] Beketov, M. A.: *Package 'HMMCont'*. [online]. 2015, [cit. 2020-06-20]. dostupné z: <https://cran.r-project.org/web/packages/HMMCont/HMMCont.pdf>.
- [3] Berchtold, A.: *Markov chains computation for homogeneous and non-homogeneous data* [online]. 2001, [cit. 2020-07-05]. dostupné z: <https://www.jstatsoft.org/article/view/v006i03>.
- [4] Černocký, J.: *Zpracování řečových signálů — studijní opora*. [online]. 2006, [cit. 2018-03-11]. dostupné z: <https://docplayer.cz/12788883-Http-www-fit-vutbr-cz-cernocky.html>.
- [5] Český svaz biatlonu: *Evidence Český Biatlon*. [online]. 2020, [cit. 2020-05-29]. dostupné z: <https://evidence.biatlon.cz>.
- [6] Dymarski, P.: *Hidden Markov Models, Theory and Applications*. [online]. 2011, [cit. 2018-11-06]. dostupné z: https://repo.pw.edu.pl/.../Hidden_Markov_Models_Theory_and_Applications.pdf.
- [7] Ejnarsson, M., Nilsson, M.: *Speech Recognition using Hidden Markov Model*. [online]. 2002, [cit. 2019-11-20]. dostupné z: <https://www.diva-portal.org/smash/get/diva2:831263/FULLTEXT01.pdf>.
- [8] Fačevicová, K., Hron, K., Kunderová, P.: *Markovovy řetězce a jejich aplikace*. Univerzita Palackého v Olomouci, Olomouc, 2018.
- [9] Fridman, M.: *Hidden Markov Model Regression*. [online]. 1993, [cit. 2020-03-15]. dostupné z: <https://www.researchgate.net/publication/2791408HiddenMarkovModelRegression>.
- [10] Jurafsky, D., Martin, J. H.: *Speech and Language Processing*. [online]. 2019, [cit. 2020-02-10]. dostupné z: https://web.stanford.edu/~jurafsky/slp3/edbook_ct162019.pdf.

- [11] Kassambara, A., Mundt, F.: *Package 'factoextra'*. [online]. 2020, [cit. 2020-07-03]. dostupné z: <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf>.
- [12] Leeflang, P. S. H., Wieringa, J. E., Bijmolt, T. H. A., Pauwels, K. H.: *Advanced Methods for Modeling Markets*. New York City: Springer, 2017.
- [13] Maixner, P.: *Markovovy procesy a jejich aplikace*. Vydavatelství Univerzity Palackého, Olomouc, 1991.
- [14] Office for National Statistics: *Office for National Statistics* [online]. 2020, [cit. 2020-06-09]. dostupné z: <https://www.ons.gov.uk/>.
- [15] Prášková, Z., Lachout, P.: *Základy náhodných procesů*. Karolinum, Praha, 2005.
- [16] Rabiner, LR.: *Theory and implementation of Hidden Markov Models*. Karolinum, Praha, 2005.
- [17] Sedlačiková, B.: *Historie matematické lingvistiky*. [online]. 2010, [cit. 2019-03-04]. dostupné z: https://is.muni.cz/th/ihe3t/DP_Sedlacikova.pdf.