BIVARIATE HIDDEN MARKOV MODEL TO CAPTURE THE DEPENDENCY IN CLAIM ESTIMATE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

ZARINA OFLAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

JULY 2016

Approval of the thesis:

BIVARIATE HIDDEN MARKOV MODEL TO CAPTURE THE DEPENDENCY IN CLAIM ESTIMATE

submitted by **ZARINA OFLAZ** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Ayşen Dener Akkaya
Head of Department, **Statistics** ______________

Assoc. Prof. Dr. Ceylan Talu Yozgatlıgil
Supervisor, **Department of Statistics, METU** ______________

Assoc. Prof. Dr. A. Sevtap Kestel
Co-supervisor, **Institute of Applied Mathematics, METU** ______________

Examining Committee Members:

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

Assoc. Prof. Dr. Ceylan Yozgatlıgil
Department of Statistics, METU ______________

Assoc. Prof. Dr. A. Sevtap Kestel
Institute of Applied Mathematics, METU ______________

Prof. Dr. Fatih Tank
Department of Insurance and Actuarial Sciences,
Ankara University ______________

Assist. Prof. Dr. Ceren Vardar Acar
Department of Statistics, METU ______________

Date: ______________

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

Name, Last Name: ZARINA OFLAZ

Signature :

ABSTRACT

BIVARIATE HIDDEN MARKOV MODEL TO CAPTURE THE DEPENDENCY IN CLAIM ESTIMATE

Oflaz, Zarina
M.S., Department of Statistics
Supervisor : Assoc. Prof. Dr. Ceylan Talu Yozgatlıgil
Co-Supervisor : Assoc. Prof. Dr. A. Sevtap Kestel

July 2016, 59 pages

Most actuarial models rely on an assumption that both claim counts and aggregate claim amounts are serially independent, that simplifies the study of many risk quantities. However, this hypothesis does not always reflect the reality and is too restrictive in different frameworks. Some weather or economic conditions reasonably affect the claim-causing events, as a result, it influences both the claim number and the claim amount distributions. The unobservable background factor can be characterized by a hidden finite state Markov chain. In our study, we propose a novel approach for modeling claim dependence, Bivariate Hidden Markov Model (BHMM), which to our knowledge has not been studied before. We assume that the claim counts and the aggregate claim amounts are mutually dependent and serially dependent through an underlying hidden state. We construct three different Bivariate Hidden Markov Models, namely Poisson-Normal HMM, Poisson-Gamma HMM and Negative Binomial-Gamma HMM. To fit the model EM algorithm is used. In order to maximize the state-dependent part of complete-data log-likelihood of bivariate HMMs, we established and proved three propositions. In application part of our thesis, we fit the Poisson-Normal HMM with a different number of states to vehicle insurance observations for Istanbul taken from Traffic Insurances Information and Monitoring Center (TRAMER) for the years 2007-2009. In addition, we performed forecasting of distributions and state prediction, obtained the most likely sequence of states.

Keywords: Claim modeling, Dependency, Bivariate Hidden Markov Model, EM algorithm, Viterbi Algorithm

ÖZ

TALEP TAHMİNİNDEKİ BAĞIMLILIĞIN İKİ DEĞİŞKENLİ SAKLI MARKOV MODELİ İLE ÇÖZÜMLENMESİ

Oflaz, Zarina
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi : Doç. Dr. Ceylan Talu Yozgatlıgil
Ortak Tez Yöneticisi : Doç. Dr. A. Sevtap Kestel

Temmuz 2016 , 59 sayfa

Aktüerya biliminde, riskin genellikle bağımsız olduğu varsayılır ancak bu varsayım her zaman gerçeği yansıtmamaktadır ve farklı çerçevelerde çok kısıtlayıcıdır. Hava ya da ekonomik koşullara bağlı olarak talep sayısı ve miktarı dağılımları zamana bağlı bir özellik gösterebilir. Gözlemlenemeyen bu tarz faktörler saklı sonlu durumlu Markov zinciri ile karakterize edilebilir. Bu çalışmada talepteki bağımlılığı iki değişkenli saklı Markov Modeli (BHMM) ile modelleyecek yeni bir yaklaşım önerilmektedir. Toplam talep sayısı ve toplam talep miktarını saklı durumlar aracılığıyla karşılıklı bağımlı ve zamana bağlı varsayarak üç farklı saklı Markov Modeli; Poisson Normal SMM, Poisson-Gama SMM ve Negatif Binom-Gama SMM geliştirildi ve EM algoritması ile model parametreleri tahminlendi. İki değişkenli SMMin log-olabilirlik fonksiyonunu maksimize etmek için üç önerme kanıtlanmıştır. Elde edilen sonuçları gerçek veriye uygularken 2007-2009 yılları arasında Trafik Sigortaları Bilgi ve Gözetim Merkezi'nden (TRAMER) alınan

İstanbul için araç sigorta talep miktarı ve sayısı arasındaki bağımlılık Poisson-Normal SMM ile farklı durum sayıları gözönüne alınarak modellemesi. Ayrıca ileriye dönük en olabilir durum zinciri oluşturulmuştur.

Anahtar Kelimeler: Talep modellemesi, Bağımlılık, İki değişkenli saklı Markov Modeli, EM algoritması, Viterbi algoritması

To My Family

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my thesis supervisor Assoc.Prof.Dr.Ceylan Talu Yozgatlıgil for the continuous support of my graduate study, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.

I would also like to thank my thesis co-supervisor Assoc.Prof.Dr.A.Sevtap Kestel. The door to Prof. Kestel office was always open whenever I ran into a trouble spot or had a question about my research or writing. She consistently allowed this paper to be my own work but steered me in the right the direction whenever she thought I needed it.

Last but not the least, I must express my very profound gratitude to my spouse and to my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ACF	autocorrelation function
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
BHMM	Bivariate Hidden Markov Model
CDLL	complete-data log-likelihood
HMM	Hidden Markov Model
p.d.f.	probability density function
pf	probability function
QQplot	quantile-quantile plot
TRAMER	Traffic Insurances Information and Monitoring Center

CHAPTER 1

INTRODUCTION

The fundamental objectives for insurance companies include safeguard policyholders against potential losses by apportioning the risk with others and compensate the loss, [1]. In order to be solvent over a certain time horizon, an insurer must adequately price the premiums to be charged and have sufficient amount of capital and reserves. Hence, predicting the distribution of the total claims amount in a given time period is important as it directly related to the equity and reserving requirements for an insurance company, [2].

The classical approach in modeling aggregate claims amount of the portfolio, consisting of n insurance policies, is to sum all amounts payable during a certain time period. It is assumed that the number of claims follows a particular discrete distribution and the monetary amount of each claim follows a continuous distribution, [3] Generally, individual risks are assumed to be independent, that simplifies the study of many risk quantities. Hence, the aggregate claims distribution is assessed under the independence assumption, [4]. Despite its simplicity and accessibility to actuaries, this assumption sometimes far from reality and is too restrictive in different frameworks. Recently, the impact of dependencies between risks have received increasing attention in the literature, see [5] and [6]. For example, according to Dhaene and Goovaerts [7], some type of dependency between individuals may produce the riskiest aggregate claims and cause the largest stop-loss premiums.

Claim modeling with dependence has been mentioned by some authors. For instance, Dhaene and Goovaerts [8] concern conditional independence of claim amounts; generalized linear models assuming the dependence between the claim counts and amounts have been constructed, [9]. Since dependence modeling using copulae was introduced by Frees and Valdez [10], copulae have become a very popular tool, e.g.various Levy copula models [11] have been applied. Copulae have been used for the modeling bivariate loss distributions [12], a joint copula-based model [13] has been suggested to capture the dependence in frequency and in severity.

Some researchers have been considered models allowing dependence among aggregate claims, see [14], [15] and [16].

In order to relax the assumption of serial independence of observations, we allow the parameter process to be serially dependent. An optimal way is to assume that the parameter process must satisfy the Markov property. The resulting model for the observations is a Hidden Markov Model. The main reason for selecting an HMM for modeling claim dependence is that unobservable background factors, which affect claim-causing events can be characterized by a hidden parameter process. That seems both claim amounts and claim numbers may behave similarly under some economic or weather conditions, consequently, we suppose they might be dependent on each other. The [17] has also considered the dependence of claim counts and the claim amounts on a common random environment. The researcher gives an overview of models where unobservable information was described by exogenous variables, using fixed and random effects models.

HMMs have been applied in various fields, namely speech recognition [18], molecular biology [21], analysis of DNA sequence [22], stock market forecasting [23]. In claim modeling, Hidden Markov model is considered to be a relatively new tool. For instance, Poisson Hidden Markov Model has been used to model the dynamics of claim counts in non-life insurance, [27], while [24] has generated the intensity function of the claim arrival process by a hidden Markov model (HMM) with Erlang state-dependent distributions.

The main work of our study involves introducing a novel approach for modeling claim dependence, Bivariate Hidden Markov Model (BHMM). We make two conditional independence assumptions, namely contemporaneous and longitudinal, i.e. we assume that the claim counts and the aggregate claim amounts are dependent and both are serially dependent via an underlying hidden state, see Figure 3.2. Multivariate HMM with considered two assumptions have been constructed to fit the multisite precipitation by Zucchini and Guttorp (1991) [40]. Also, the model has been fitted to the bivariate series, where one component is linear and other is circular [29]. Theory related to the proposed model structure is discussed by Zucchini and MacDonald [29], [20].

In the thesis, we construct three different Bivariate Hidden Markov Models, namely Poisson-Normal HMM, Poisson-Gamma HMM and Negative Binomial-Gamma HMM. In order to estimate model parameters, EM algorithm is conducted.

This thesis is organized as follows: in Chapter 2, an overview of the theoretical framework of the claim modeling is given, Chapter 3 presents the Bivariate Hidden Markov Model and related definitions and propositions,the basic theory of Hidden Markov models is considered in this chapter as well. Chapter 4 includes an application of the Poisson-Normal HMM to the vehicle insurance data for Istanbul, 2007-2009. Data description and results of analysis are considered.

CHAPTER 2

CLAIM MODELING

Insurance companies are primarily interested in assessing the likelihood of claim occurrence, as well as the monetary loss of the claim. Evaluating total payments in a given time period is pivotal for calculating premiums and reserves, pricing of insurance contracts and preventing insolvency of the company [19].

An *aggregate loss* S is the sum of the monetary losses of all the claim in a certain period of time $(0, t]$. The number of claims, N, called the *frequency* random variable and the monetary amount of each claim, X, called the *severity* are combined to model the total loss, S. Obviously, N is assumed to be a nonnegative discrete random variable,while X is continuously distributed.

Aggregate loss distributions have been discussed in the most actuarial literature, e.g. see [2] and [3].

There are two major approaches in modeling aggregate loss: the **individual risk model** and the **collective risk model**.

The individual risk model specifies the aggregate loss as follows:

$$S = \sum_{i=1}^{n} X_i, \tag{2.1}$$

where n is a *fixed number* of individual risks in the portfolio and the X_i's are independent random variables for the individual losses [26].

The use of collective risk theory provides an alternative way to the above approach. The aggregate claim amount is assumed to be a random process. The model is specified by

$$S = \sum_{i=1}^{N} X_i, \tag{2.2}$$

where X_i is the size of claim i and N is the number of claims in a time period [3]. In contrast to the individual risk model, the number of claims N is a *random variable.*

In the compound distribution, $X_1, X_2, ..., X_N$ are identically distributed, and $N_t, X_1, X_2, ..., X_N$ are assumed to be mutually independent.

Now, we introduce the *aggregate claim process* S_t which is widely used in actuarial modeling. It is defined by the summation of each policies' claim amount X_i's in a certain time period $t \geq 0$:

$$S_t = \sum_{i=1}^{N_t} X_i, \tag{2.3}$$

where the $\{N_t : \quad t \geq 0\}$ is called the *claim number process.* In contrast to the collective risk model, which is interested in claim modeling for a single period, here we are concerned in the number of claims and aggregate claims for a long time period, i.e. we would like to obtain the distributions at all times $t \geq 0$. [2]

Modeling the claim frequency distribution and the claim severity distribution separately has been found advantegous in many aspects, see [19].

2.1 Claim frequency distribution

As the number of claims can only take nonnegative integer values, counting distributions can be used for modeling the claim frequency distribution. In practice, the most commonly used distributions are the Poisson, Binomial, and Negative Binomial distributions. A classic choice for modeling claim counts is the Poisson distribution, since

Nonnegative discrete variable N follows a Poisson distribution with parameter λ, if the pf of N is defined as follows

$$p_n = \frac{e^{-\lambda}\lambda^n}{n!}, \qquad n = 0, 1, 2, ...$$

The Poisson distribution is characterized as equidispersed distribution since its mean and variance are equal. In the following, we present two useful properties of the Poisson distribution [3].

Theorem 2.1 *If $N_1, N_2, ..., N_n$ are independently distributed as a Poisson distribution with parameter λ_i, for $i = 1, ..., n$, then $N = N_1 + ... + N_n$ has a Poisson distribution with parameter $\lambda_1 + ... + \lambda_n$.*

Theorem 2.2 *Suppose that the number of events N distributed as a Poisson distribution with mean λ. Let each event be classified into one of m types with probabilities $p_1, ..., p_m$. Events are mutually independent. Then $N_1, ..., N_m$ are mutually independent random variables distributed as a Poisson with parameters $\lambda p_1, ..., \lambda p_m$ respectively.*

The Negative Binomial distribution might be an optimal candidate for overdispersed data, since its variance exceeds its mean. Compare to the Poisson distribution, the Negative Binomial distribution is more flexible in shape, because it has two parameters.

The probability function of the Negative Binomial distribution with parameters $r > 0, p \in (0, 1)$ is given by

$$p_n = \binom{n-1}{r-1} p^r (1-p)^{n-r}.$$

The Binomial distribution differs from other counting distributions, its variance is smaller than its mean. N is said to have Binomial distribution with parameters p, θ, if the pf of N, for $n = 0, 1, ..., p$ is given by

$$p_n = \binom{p}{n} \theta^n (1-\theta)^{p-n}.$$

2.2 Claim severity distribution

The claim severity is usually distrbuted as a nonegative continuous random variable. Here we present the common claim severity distributions, however, it may also be modelled by a mixture of distributions or by a modification of existing distributions.

The Gamma distribution is usually used if the cumulative distribution function has not too heavy tail, for instance, in motor insurance, where a claim event causes injury to an insured vehicle [2]. The p.d.f. of the Gamma distribution is defined as follows

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \qquad \text{for} \quad \beta > 0, \alpha > 0.$$

Fire insurance, where the claim event creates a severe loss, requires modeling claim severity with heavy-tailed distributions. Generally, the Lognormal distribution and the Pareto distribution are suggested to use in this type of insurance [2].

However, the choice of an adequate distribution depends on a given data and an experience of a researcher.

CHAPTER 3

BIVARIATE HIDDEN MARKOV MODEL

The purpose of this chapter is to provide a short review of Hidden Markov models (HMMs) and then to introduce new models of our study, Bivariate Hidden Markov models (BHMMs). First, in Section 3.1 we give an account of Markov chains because the unobserved 'parameter process' of hidden Markov model satisfies the Markov property. Second, in Section 3.2 an HMM and related definitions are introduced. In Section 3.3, we discuss Bivariate Hidden Markov models; propositions necessary for parameter estimation and their proofs, EM algorithm and forward-backward algorithms, lastly, Viterbi algorithm to decode hidden states is presented. "Hidden Markov Models for Time Series" by Walter Zucchini and Iain L. MacDonald [29] is taken as the main reference of our study. Additionally, we modified the theory of classic HMM for bivariate case.

3.1 Markov chains

We will consider a stochastic process $\{C_t\}$ in discrete time $t = 1, 2, ...$, referring to the value C_t as the *state* of the process at time t, and C_1 indicates the initial state.

Definition 3.1 (Markov chains) *A sequence of random variables* $\{C_t : t = 1, 2, ...\}$ *is called a* ***Markov chain*** *if for all* $t \in \mathbf{N}$ *it follows a* ***Markov property***

$$P(C_{t+1}|C^{(t)}) = P(C_{t+1}|C_t), \tag{3.1}$$

where $C^{(t)}$ *is defined as the history* $(C_1, C_2, ..., C_t)$.

Thus, the probability distribution of the next state depends only on the current state and not on previous ones.

Definition 3.2 (Matrix of transition probabilities) *The matrix* $\Gamma(1)$ *,abbreviated as* Γ*, is a square matrix of probabilities with row summing up to one*

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix},$$

where $\gamma_{ij}(t) = P(C_{k+t} = j | C_k = i)$ *are* ***transition probabilities*** *and* m *denotes the number of states of the Markov chain.*

Transition probability $\gamma_{ij}(t)$ can be expressed as the probability of moving from state i to state j at time t. If these probabilities do not depend on k, the Markov chain is said to be **homogeneous**. Finite state-space homogeneous Markov chains fulfill the **Chapman-Kolmogorov equations** [29] :

$$\Gamma(t + u) = \Gamma(t)\Gamma(u), \tag{3.2}$$

which implies $\Gamma(t) = \Gamma(1)^t$.

Probabilities of a Markov chain being in a given state at a given time t can be defined by **unconditional probabilities**

$$u(t) = (P(C_t = 1), ..., P(C_t = m)). \tag{3.3}$$

$u(1)$ is considered as **initial distribution** of the Markov chain, which specifies the starting state.

In our study, we consider a homogeneous nonstationary Markov chain. However, in order to define a starting value of initial distribution we use a stationary distribution of a Markov chain.

Definition 3.3 *A Markov chain with transition probability Gamma has a stationary probability* δ *if* $\delta\Gamma = \delta$ *and* $\delta 1' = 1$*.*

A stationary distribution δ can be found by the following expression,[29]

$$\delta(I_m - \Gamma + U) = 1, \tag{3.4}$$

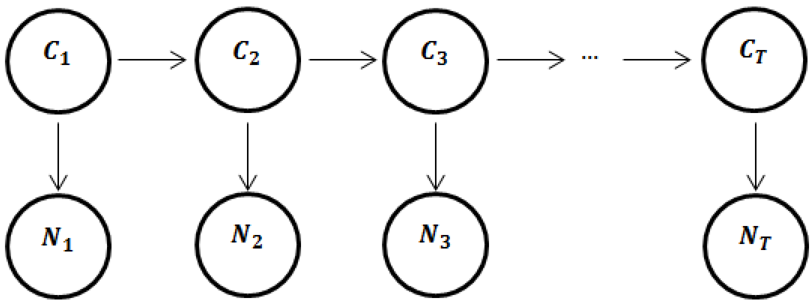

Figure 3.1: Directed graph of basic HMM.

where δ is the stationary distribution, I_m is the $m \times m$ identity matrix, U is the $m \times m$ matrix of ones and 1 is a row vector of ones.

3.2 Hidden Markov Model

A Hidden Markov Model (HMM) is a powerful statistical tool for modeling time series data. Let us N_t denote as the observation at time t, $t \in \mathbf{N}$. The model assumes that process generating N_t depends on the *hidden* state C_t which satisfies the Markov property.

Thus, an HMM can be determined by hidden 'parameter process' $\{C_t : t = 1, 2, ...\}$ and the 'state-dependent process' $\{N_t : t = 1, 2, ...\}$, [29] satisfying

$$P(C_{t+1}|C_t, ...C_1) = P(C_{t+1}|C_t), \quad t = 2, 3, ...$$

$$P(N_t|N^{(t-1)}, C^{(t)}) = P(N_t|C_t), \quad t \in N. \tag{3.5}$$

The structure of HMM is displayed in the following Figure 3.1.

Defined above the initial distribution $u(1)$ and matrix of transition probabilities $\gamma_{ij}(t)$ are necessary to construct a probability distribution over sequences of observations. Additionally, we need to specify the **state-dependent distribution** $p_i(n)$, that defines the relation between the observation and an unobserved state. For discrete-valued observations $p_i(n)$ is defined as follows, for $i = 1, 2, ..., m$:

$$p_i(n) = P(N_t = n|C_t = i).$$

For continuous case p_i is defined to be the probability *density* function of N_t if the Markov chain is in state i at time t.

We indicate $P(n)$ as the diagonal matrix of state-dependent distributions of observation n

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

3.3 Bivariate Hidden Markov Models

3.3.1 Model specification

Let $\{N_t : t = 1, 2, ...\}$ be the number of claims and $\{S_t : t = 1, 2, ...\}$ be the aggregate claim amount reported by policyholders during the time period $t = 1, 2, ...$ Most actuarial models rely on the assumption that both N_t and S_t are serially independent, that simplifies the study of many risk quantities. However, this hypothesis does not always reflect the reality and is too restrictive in different frameworks. The sample autocorrelation function of claim counts and aggregate amounts, displayed in Figure 4.3 and Figure 4.4,respectively, indicates that the values at different times have a dependency among each other. Moreover, it is necessary to remark that according to Equation 2.3 from Chapter 2 N_t and S_t are also dependent.

Some weather or economic conditions reasonably affect the claim-causing events, as a result, it influences both the claim number and claim amount distributions [27]. The unobservable background factor can be characterized by hidden finite state Markov chain. In our study, we propose a new approach for modeling claim dependence, ***Bivariate Hidden Markov Model***, which to our knowledge has not been studied in literature before. We assume that N_t and S_t are mutually dependent and serially dependent through an underlying hidden state $\{C_t : t = 1, 2, ...\}$. In our study, we consider that the Markov chain of the bivariate model is homogeneous and non-stationary. The model's structure is displayed in Figure 3.2.

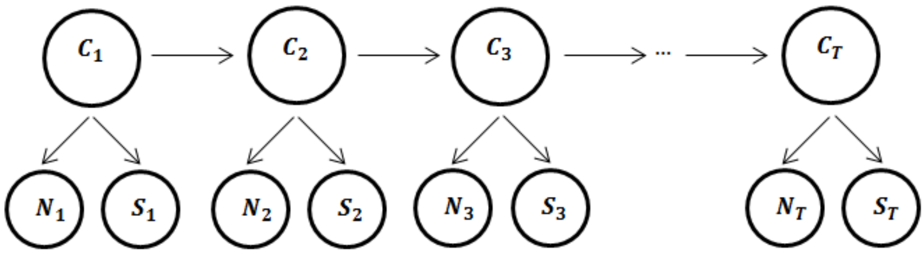

Figure 3.2: Directed graph of bivariate HMM.

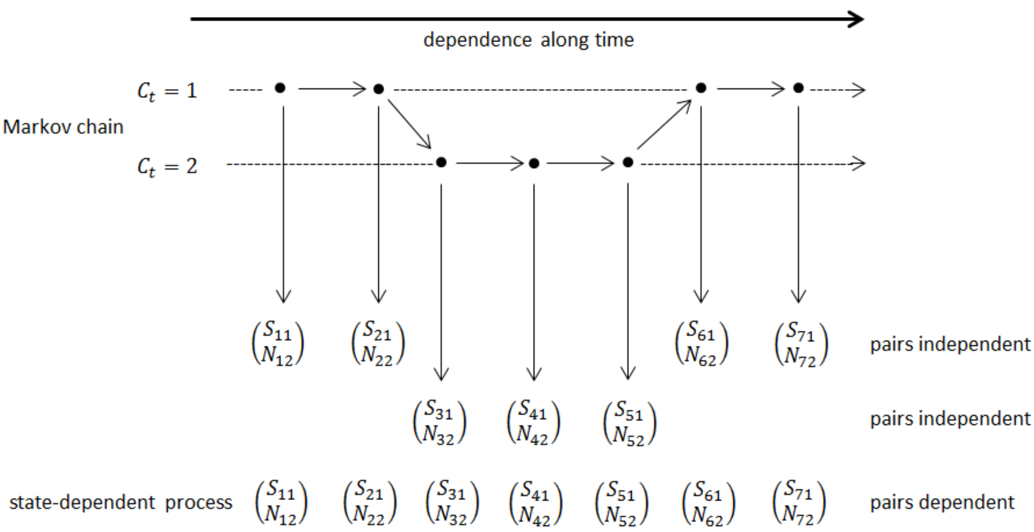

Figure 3.3: Contemporaneous conditional independence.

Obviously, claim numbers N_t and aggregate claim amounts S_t are reported at same time t, therefore in our study information given by bivariate observations (S_t, N_k), $t \neq k$ is insignificant.

We assume *longitudinal conditional independence*, i.e. conditional on the underlying hidden state $\{C_t : t = 1, 2, ...\}$ the claim counts at time t and the aggregate amounts at time t are assumed to be independent. In addition, we admit *contemporaneous conditional independence*, which is interpreted in Figure 3.3, the scheme was taken from [29]. These two conditional independence assumptions do not imply the serial independence of N_t and S_t or that the component series are mutually independent, hence that N_t and S_t are dependent. [29]

To specify the bivariate model it is necessary to postulate a joint state-dependent distribution for $t = 1, 2, ..., T$,$i = 1, 2, ..., m$ and all relevant s, n

$$p_i(s_t, n_t) = P((S_t, N_t) = (s_t, n_t)|C_t = i).$$

According to the contemporaneous conditional independence, the state-dependent probabilities are given by a product of the corresponding marginal probabilities [29]:

$$\begin{aligned} p_i(s_t, n_t) &= P((S_t, N_t) = (s_t, n_t)|C_t = i) \\ &= P(S_t = s_t|C_t = i)P(N_t = n_t|C_t = i). \end{aligned} \tag{3.6}$$

In our study we construct three different bivariate models. These are: the Poisson-Normal Hidden Markov Model, the Poisson-Gamma Hidden Markov Model and the Negative Binomial-Gamma Hidden Markov Model. The Poisson-Normal Hidden Markov Model applied to the real insurance data.

We define a joint state-dependent distribution for the ***Poisson-Normal Hidden Markov Model***, for $n \in N, s > 0, \lambda > 0, \mu > 0, \sigma^2 > 0$, as follows:

$$p_i(s_t, n_t) = (2\pi\sigma_i^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_i^2}(s_t-\mu_i)^2-\lambda_i} \frac{\lambda_i^{n_t}}{n_t!}. \tag{3.7}$$

We consider that N_t follows the Poisson distribution and S_t the Normal distributions with underlying unobservable stochastic process C_t.

Similarly, we present the ***Poisson-Gamma Hidden Markov Model*** with marginal distributions, the Poisson distribution for N_t and the Gamma distribution for S_t. The joint state-dependent distribution, for $n \in N, s > 0, \lambda > 0, \alpha > 0, \beta > 0$, is given by

$$p_i(s_t, n_t) = \frac{\beta_i^{\alpha_i} s_t^{\alpha_i - 1} \lambda_i^{n_t} e^{-\beta_i s_t - \lambda_i}}{\Gamma(\alpha_i) n_t!}. \tag{3.8}$$

For the ***Negative Binomial-Gamma Hidden Markov Model*** the state-dependent distribution is of the form

$$p_i(s_t, n_t) = \frac{\binom{n_i-1}{r_i-1} \beta_i^{\alpha_i} s_t^{\alpha_i-1} e^{-\beta_i s_t} p_i^{r_i} (1-p_i)^{n_t - r_i}}{\Gamma(\alpha_i)}. \tag{3.9}$$

for $\quad n \in N, s > 0, r > 0, p \in (0, 1), \alpha > 0, \beta > 0.$

3.3.2 The likelihood and marginal distributions

The following definitions and propositions are modified for bivariate case, based on the classic theory of HMM [29]. We suppose there is an observation sequence $s_1, s_2, ..., s_T, n_1, n_2, ..., n_T$. An m-state BHMM has an initial distribution $u(1)$, the transition probability matrix Γ and matrix of joint state-dependent probabilities $P(s, n)$.

Proposition 3.1 *The likelihood of bivariate Hidden markov model, for* $t = 1, 2, ..., T$ *and all relative* s, n, *is given by*

$$L_T = u(1)P(s_1, n_1)\Gamma P(s_2, n_2)\Gamma P(s_3, n_3)...\Gamma P(s_T, n_T)1'.$$

We define the marginal distribution , for $t = 1, 2, ..., T$ as follows:

$$\begin{aligned} P((S_t, N_t) = (s, n)) &= \sum_{i=1}^{m} P(C_t = i)P((S_t, N_t) = (s, n)|C_t = i) \\ &= \sum_{i=1}^{m} u_i(t)p_i(s, n). \end{aligned} \tag{3.10}$$

Also, it can be represented in a matrix form:

$$\begin{aligned} P((S_t, N_t) = (s, n)) &= (u_1(t), ...u_m(t)) \begin{pmatrix} p_1(s, n) & & 0 \\ & \ddots & \\ 0 & & p_m(s, n) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \\ &= u(t)P(s, n)1'. \end{aligned} \tag{3.11}$$

It is of interest to obtain the marginal distribution separately for N_t and S_t. Considering, that N_t is a discrete variable, the marginal distribution $P(N_t = n)$ is given as follows:

$$\begin{aligned} P(N_t = n) &= \sum_{i=1}^{m} P(C_t = i)P(N_t = n|C_t = i) \\ &= \sum_{i=1}^{m} u_i(t)p_i(n). \end{aligned} \tag{3.12}$$

Likewise, in matrix form:

$$\begin{aligned} P(N_t = n) &= (u_1(t), ...u_m(t)) \begin{pmatrix} p_1(n) & & 0 \\ & \ddots & \\ 0 & & p_m(n) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \\ &= u(t)P(n)1'. \end{aligned} \tag{3.13}$$

For continuous variable S_t, which state-dependent distribution can be described as a complete p.d.f. over the continuous observation space for each state, the marginal distribution is given by

$$\begin{aligned} P(S_t) &= \sum_{i=1}^{m} P(C_t = i)P(S_t|C_t = i) \\ &= \sum_{i=1}^{m} u_i(t)p_i(s). \end{aligned} \tag{3.14}$$

The expression can be represented in matrix notation:

$$\begin{aligned} P(S_t = n) &= (u_1(t), ...u_m(t)) \begin{pmatrix} p_1(s) & & 0 \\ & \ddots & \\ 0 & & p_m(s) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \\ &= u(t)P(s)1'. \end{aligned} \tag{3.15}$$

3.4 Parameter Estimation in Bivariate HMM

To construct the model it is necessary to estimate transition probabilities, initial probability, and parameters of the joint state-dependent probabilities. In order to fit the Bivariate HMMs, the EM algorithm is used. In the context of HMMs, the algorithm is also known as the Baum-Welch algorithm. EM algorithm performs maximum likelihood estimation of parameters having missing value in the data [34]. We treat hidden states as missing data [33]. In addition, the algorithm enables estimation of the parameters of an HMM whose Markov chain is homogeneous but not necessarily stationary [29]. In order to maximize the state-dependent part of complete-data log-likelihood of bivariate HMMs, we establish and prove three propositions.

3.4.1 Forward-Backward Probabilities

The tools we need to apply the EM algorithm are the forward and the backward probabilities. In this section, we give definitions of the forward and the backward probabilities and present propositions necessary for maximization part of EM estimation [35].

Definition 3.4 (Forward probabilities) *For $t = 1, 2, ..., T$ forward probabilities, α_t, are defined as follows:*

$$\alpha_t = \delta P(s_1, n_1)\Gamma P(s_2, n_2)...\Gamma P(s_t, n_t) = \delta P(s_1, n_1) \prod_{k=2}^{t} \Gamma P(s_k, n_k).$$

Definition 3.5 (Backward probabilities) *Backward probabilities, β_t', for $t = 1, 2, ..., T$ are defined by*

$$\beta_t' = \Gamma P(s_{t+1}, n_{t+1})\Gamma P(s_{t+2}, n_{t+2})...\Gamma P(s_T, n_T)1' = (\prod_{k=t+1}^{T} \Gamma P(s_k, n_k))1'.$$

$$\beta_T = 1.$$

The following proposition identifies $\alpha_t(j)$ as the joint probability of the observations $s_1, s_2, ..., s_t, n_1, n_2, ..., n_t$ and hidden state j at time t.

Proposition 3.2 *For $t = 1, 2, ..., T$ and $j = 1, 2, ..., m$*

$$\alpha_t(j) = P((S^{(t)}, N^{(t)}) = (s^{(t)}, n^{(t)}), C_t = j).$$

The following proposition defines $\beta_t(i)$ as the probability of the observations being $s_{t+1}, s_{t+2}, ..., s_T, n_{t+1}, n_{t+2}, ..., n_T$, given that the Markov chain is in state i at time t.

Proposition 3.3 *For $t = 1, 2, ..., T$ and $j = 1, 2, ..., m$*

$$\beta_t(i) = P((S_{t+1}^T, N_{t+1}^T) = (s_{t+1}^T, n_{t+1}^T), C_t = i)$$

where Z_a^b denotes the vector $(Z_a, Z_{a+1}, ..., Z_b)$.

We now establish the propositions concerning the forward and backward probabilities useful in applying the EM algorithm to bivariate HMMs.

Proposition 3.4 *Firstly, for* $t = 1, 2, ..., T$

$$P(C_t = j | (S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = \alpha_t(j)\beta_t(j)/L_T,$$

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

$$P(C_{t-1} = j, C_t = k | (S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = \alpha_{t-1}(j)\gamma_{jk} p_k(s_t, n_t)\beta_t(k)/L_T.$$

Proposition 3.5 *For* $t = 1, 2, ..., T$ *and* $i = 1, 2, ..., m$

$$\alpha_t(i)\beta_t(i) = P((S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)}), C_t = j),$$

and therefore

$$\alpha_t\beta_t' = P((S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = L_T, \quad \textit{for each such} \quad t.$$

3.4.2 EM algorithm

The EM algorithm is an efficient iterative procedure to compute the maximum likelihood estimation of the parameters of an underlying distribution from a given dataset in the presence of missing or hidden data. The expectation maximization algorithm alternates between two phases. In the E-step conditional expectations of the missing data given the observed data and a current estimate of the model parameters are estimated. In the M-step, the complete-data log-likelihood function is maximized under the assumption that the missing data are known. Iterations are repeated until a convergence is satisfied [34].

The complete-data log-likelihood of a bivariate HMM, i.e. the log-likelihood of observed variables and hidden states, is defined as follows [29]:

$$\log(P(s^{(T)}, n^{(T)}, c^{(T)}) = \log \delta_{c_1} + \sum_{t=2}^{T} \log \delta_{c_{t-1}, c_t} + \sum_{t=1}^{T} \log p_{c_t}(s_t, n_t). \tag{3.16}$$

Defining the zero-one random variables, we have

$$\begin{aligned} \log(P(s^{(T)}, n^{(T)}, c^{(T)}) = & \sum_{j=1}^{m} u_j(1) \log \delta_j + \sum_{j=1}^{m} \sum_{k=1}^{m} (\sum_{t=2}^{T} v_{jk}(t)) \log \gamma_{jk} \\ & + \sum_{j=1}^{m} \sum_{t=1}^{T} u_j(t) \log p_j(s_t, n_t), \end{aligned} \tag{3.17}$$

where $u_j(t) = 1$ if and only if $c_t = j$,$(t = 1, 2, ..., T)$,$v_{jk} = 1$ if and only if $c_{t-1} = j$ and $c_t = 1k$ $(t = 2, 3, ..., T)$.

The EM algorithm for a bivariate HMM [29]:
In **E part** $v_{jk}(t)$ and $u_j(t)$ are replaced by the conditional expectations of being in a state j at time t given the observations $s^{(T)}, n^{(T)}$:

$$\hat{u}_j(t) = P(C_t = j|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = \alpha_t(j)\beta_t(j)/L_T;$$

and

$$\hat{v}_{jk}(t) = P(C_{t-1} = j, C_t = k|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = \alpha_{t-1}(j)\gamma_{jk}p_k(s_t, n_t)\beta_t(k)/L_T.$$

M part: Each term of the CDLL is maximized with respect to the related set of parameters, i.e. the initial distribution $u(1)$, the transition probability matrix Γ, and the parameters of the joint state-dependent distributions. Observing the CDLL of bivariate HMM, we indicate that three separate maximizations in the M-step are required. Thus:
1. Setting $u_j(1) = \hat{u}_j(1)/\sum_{j=1}^{m}\hat{u}_j(1) = \hat{u}_j(1)$, maximize $\sum_{j=1}^{m} u_j(1)\log\delta_j$ with respect to initial distribution $u(1)$;
2. Setting $\gamma_{jk} = \sum_{t=2}^{T} v_{jk}(t)/\sum_{k=1}^{m}(\sum_{t=2}^{T} v_{jk}(t))$, maximize

$$\sum_{j=1}^{m}\sum_{k=1}^{m}(\sum_{t=2}^{T} v_{jk}(t))\log\gamma_{jk}$$

with respect to Γ;
3. Depending on the nature of the joint state-distributions assumed, the maximization of the third term can be performed analytically,i.e. closed-form solutions are available, or numerical estimation will be required.

In the next sections, we present propositions related to our new models in order to maximize the third term of CDLL of bivariate HMM.

3.4.3 Poisson-Normal Hidden Markov Model

Proposition 3.6 *Given two random variables, S and N having Normal (μ_j, σ_j^2) and Poisson(λ_j) distributions,respectively, the EM estimate of joint state-dependent* *distribution are*

$$\hat{\lambda}_j = \frac{\sum_{t=1}^T \hat{u}_j(t) n_t}{\sum_{t=1}^T \hat{u}_j(t)},$$
$$\hat{\mu}_j = \frac{\sum_{t=1}^T \hat{u}_j(t) s_t}{\sum_{t=1}^T \hat{u}_j(t)}, \tag{3.18}$$
$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^T \hat{u}_j(t)(s_t - \hat{\mu}_j)^2}{\sum_{t=1}^T \hat{u}_j(t)}.$$

Proof: The joint state-dependent probability for the Poisson-Normal HMM is given by

$$p_j(s_t, n_t) = (2\pi\sigma_j^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_j^2}(s_t-\mu_j)^2 - \lambda_j} \frac{\lambda_j^{n_t}}{n_t!}.$$

M step of EM algorithm requires the maximization of the state-dependent part of the CDLL

$$\sum_{j=1}^m \sum_{t=1}^T \hat{u}_j(t) \log p_j(x_t, n_t) \tag{3.19}$$

with respect to the parameters of the joint state-dependent distribution. Defining $F = \sum_{t=1}^T \hat{u}_j(t) \log p_j(s_t, n_t)$, we have

$$F = \sum_{t=1}^T \hat{u}_j(t)[-\frac{1}{2}\log(2\pi\sigma_j^2) - \frac{(s_t - \mu_j)^2}{2\sigma_j^2} - \lambda_j + n_t \log \lambda_j - \log(n_t!)]. \tag{3.20}$$

Maximizing values of the state-dependent parameters λ_j, μ_j and σ_j^2 can be computed by setting the derivative to zero with respect to corresponding parameters:

$$\frac{dF}{d\lambda_j} = \sum_{t=1}^T \hat{u}_j(t)[-1 + \frac{n_t}{\lambda_j}] = 0$$

and hence,that

$$\hat{\lambda}_j = \frac{\sum_{t=1}^T \hat{u}_j(t) n_t}{\sum_{t=1}^T \hat{u}_j(t)}. \tag{3.21}$$

Maximization of the state-dependent part of CDLL with respect to μ_j proceeds as follows:

$$\frac{dF}{d\mu_j} = \sum_{t=1}^T \hat{u}_j(t)[\frac{s_t}{\sigma_j^2} - \frac{\mu_j}{\sigma_j^2}] = 0,$$

then

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

Analogically for σ_j^2:

$$\frac{dF}{d\sigma_j^2} = \sum_{t=1}^T \hat{u}_j(t)[-\frac{1}{2\sigma_j^2} + \frac{(s_t - \mu_j)^2}{2(\sigma_j^2)^2}] = 0,$$

then,
$$\sum_{t=1}^{T} \hat{u}_j(t) \frac{(s_t - \mu_j)^2}{2(\sigma_j^2)^2} = \sum_{t=1}^{T} \hat{u}_j(t) \frac{1}{2\sigma_j^2},$$
and hence,that
$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \hat{\mu}_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)}. \tag{3.22}$$
For confidence that the estimated parameters maximize the state-dependent part of CDLL, we check second derivatives of F with respect to parameters. The second derivative of F with respect to λ_j is
$$\frac{d^2F}{d\lambda_j^2}\Big|_{\lambda_j=\hat{\lambda}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{n_t}{\lambda_j^2}\Big|_{\lambda_j=\hat{\lambda}_j} < 0,$$
since $n_t > 0$ and $\hat{u}_j(t) = \{0, 1\}$ by definition. It is obvious, that the following satisfies
$$\frac{d^2F}{d\mu_j^2}\Big|_{\mu_j=\hat{\mu}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{1}{\sigma_j^2}\Big|_{\mu_j=\hat{\mu}_j} < 0.$$
Finally, we check the second derivative of F with respect to σ_j^2.
$$\begin{aligned} \frac{d^2F}{d\sigma_j^2}\Big|_{\sigma_j^2=\hat{\sigma}_j^2} &= \sum_{t=1}^{T} \hat{u}_j(t)[-\frac{1}{2\sigma_j^2} + \frac{(s_t - \mu_j)^2}{2(\sigma_j^2)^2}]\Big|_{\sigma_j^2=\hat{\sigma}_j^2} \\ &= \sum_{t=1}^{T} \hat{u}_j(t)[\frac{\sigma_j^2 - 2(s_t - \mu_j)^2}{2(\sigma_j^2)^3}]\Big|_{\sigma_j^2=\hat{\sigma}_j^2} \\ &= \frac{\sigma_j^2 \sum_{t=1}^{T} \hat{u}_j(t) - 2\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \mu_j)^2}{2(\sigma_j^2)^3}\Big|_{\sigma_j^2=\hat{\sigma}_j^2} \end{aligned} \tag{3.23}$$
It is suffcicient to prove that
$$\sigma_j^2 \sum_{t=1}^{T} \hat{u}_j(t) - 2\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \mu_j)^2\Big|_{\sigma_j^2=\hat{\sigma}_j^2} < 0.$$
Transforming the above expression, we derive:
$$\begin{aligned} \sigma_j^2 - \frac{2\sum_{t=1}^{T} \hat{u}_j(t)(s_t-\mu_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)}\Big|_{\sigma_j^2=\hat{\sigma}_j^2} &= \frac{\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \mu_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)} - \frac{2\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \mu_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)} \\ &= -\frac{\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \mu_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)} < 0. \end{aligned} \tag{3.24}$$
This finalizes the proof of the estimated parameters maximizing the state-dependent part of CDLL introduced in this study. □

3.4.4 Poisson-Gamma Hidden Markov Model

Proposition 3.7 *Given two random variables, S and N having Gamma(α_j, β_j) and Poisson(λ_j) distributions,respectively, the EM estimate of joint state-dependent distribution are*

$$\hat{\lambda}_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t) n_t}{\sum_{t=1}^{T} \hat{u}_j(t)}, \qquad \hat{\beta}_j = \frac{\hat{\alpha}_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} s_t}. \tag{3.25}$$

To estimate $\hat{\alpha}_j$ numerical maximization is required.

Proof: The joint state-dependent probability for the Poisson-Gamma HMM is given by

$$p_j(s_t, n_t) = \frac{\beta_i^{\alpha_j} s_t^{\alpha_j - 1} \lambda_j^{n_t} e^{-\beta_j s_t - \lambda_j}}{\Gamma(\alpha_j) n_t!}.$$

M step of EM algorithm requires the maximization of the state-dependent part of the CDLL

$$\sum_{j=1}^{m} \sum_{t=1}^{T} \hat{u}_j(t) \log p_j(x_t, n_t)$$

with respect to the parameters of the joint state-dependent distribution. Defining $F = \sum_{t=1}^{T} \hat{u}_j(t) \log p_j(s_t, n_t)$, we have

$$F = \sum_{t=1}^{T} \hat{u}_j(t)[\alpha_j \log \beta_j + (\alpha_j - 1) \log s_t + n_t \log \lambda_j - \log \Gamma(\alpha_j) - \log(n_t!) - \beta_j s_t - \lambda_j].$$

Maximizing values of the state-dependent parameters λ_j, μ_j and σ_j^2 can be computed by setting the derivative to zero with respect to corresponding parameter:

$$\frac{dF}{d\lambda_j} = \sum_{t=1}^{T} \hat{u}_j(t)[-1 + \frac{n_t}{\lambda_j}] = 0$$

and hence,that

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

Analogically for β_j:

$$\frac{dF}{d\beta_j} = \sum_{t=1}^{T} \hat{u}_j(t)[\frac{\alpha_j}{\beta_j} - s_t] = 0,$$

and hence that

$$\hat{\beta}_j = \frac{\alpha_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t) s_t}.$$

Maximization of the state-dependent part of CDLL with respect to α_j proceeds as follows

$$\frac{dF}{d\alpha_j} = \sum_{t=1}^{T} \hat{u}_j(t)[-\frac{d}{d\alpha_j} \log \Gamma(\alpha_j) + \log \beta_j +_t] = 0,$$

then replacing β_j by $\hat{\beta}_j$, we get

$$\frac{dF}{d\alpha_j} = \sum_{t=1}^{T} \hat{u}_j(t)[-\frac{d}{d\alpha_j} \log \Gamma(\alpha_j) + \log \frac{\alpha_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t) s_t} + \log s_t] = 0$$

In order to estimate the above equation numerical maximization is required. For confidence that the estimated parameters maximize the state-dependent part of CDLL, we check second derivatives of F with respect to parameters. The second derivative of F with respect to λ_j is

$$\frac{d^2 F}{d\lambda_j^2}\Big|_{\lambda_j=\hat{\lambda}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{n_t}{\lambda_j^2}\Big|_{\lambda_j=\hat{\lambda}_j} < 0,$$

since $n_t > 0$ and $\hat{u}_j(t) = \{0, 1\}$ by definition.

For the parameters of Gamma distribution, we have

$$\frac{d^2 F}{d\alpha_j^2}\Big|_{\alpha_j=\hat{\alpha}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{d^2}{d\alpha_j^2} \log \Gamma(\alpha_j)\Big|_{\alpha_j=\hat{\alpha}_j} < 0,$$

since the trigamma function, defined as the sum of the series, is positive

$$\frac{d^2}{d\alpha_j^2} \log \Gamma(\alpha_j) = \sum_{k=0}^{\infty} \frac{1}{(\alpha_j + k)^2} > 0.$$

Finally, we check the second derivative of F with respect to β_j^2

$$\frac{d^2 F}{d\beta_j^2}\Big|_{\beta_j=\hat{\beta}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{\alpha_j}{\beta_j^2} < 0,$$

since $\alpha_j > 0$. □

3.4.5 Negative Binomial-Gamma Hidden Markov Model

Proposition 3.8 *Given two random variables, S and N having Gamma (α_j, β_j) and Negative Binomial (r_j, p_j) distributions,respectively, the EM estimate of joint state-dependent distribution are*

$$\begin{aligned} \hat{\beta}_j &= \frac{\hat{\alpha}_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} s_t}, \\ \hat{p}_j &= \frac{\hat{r}_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - \hat{r}_j) + \hat{r}_j \sum_{t=1}^{T} \hat{u}_j(t)}. \end{aligned} \tag{3.26}$$

To estimate $\hat{\alpha}_j$ and $\hat{r}_j$ numerical maximization is required.

Proof: The joint state-dependent probability for the Negative Binomial-Gamma HMM is given by

$$p_j(s_t, n_t) = \frac{\binom{n_j-1}{r_j-1}\beta_j^{\alpha_j} s_t^{\alpha_j-1} e^{-\beta_j s_t} p_j^{r_j}(1-p_j)^{n_t-r_j}}{\Gamma(\alpha_j)}$$

M step of EM algorithm requires the maximization of the state-dependent part of the CDLL

$$\sum_{j=1}^{m}\sum_{t=1}^{T} \hat{u}_j(t) \log p_j(x_t, n_t)$$

with respect to the parameters of the joint state-dependent distribution. Defining $F = \sum_{t=1}^{T} \hat{u}_j(t) \log p_j(s_t, n_t)$, we have

$$F = \textstyle\sum_{t=1}^{T} \hat{u}_j(t) \log p_j(s_t, n_t) = \sum_{t=1}^{T} \hat{u}_j(t)[\log \binom{n_j-1}{r_j-1} + \alpha_j \log \beta_j + (\alpha_j - 1) \log s_t$$
$$- \beta_j s_t + r_j \log p_j + (n_t - r_j) \log(1 - p_j) - \log \Gamma(\alpha_j)]. \quad (3.27)$$

By setting the derivative to zero with respect to β_j we derive

$$\frac{dF}{d\beta_j} = \sum_{t=1}^{T} \hat{u}_j(t)[\frac{\alpha_j}{\beta_j} - s_t] = 0,$$

and hence that

$$\hat{\beta}_j = \frac{\alpha_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t) s_t}.$$

Maximization of the state-dependent part of CDLL with respect to α_j performs as follows

$$\frac{dF}{d\alpha_j} = \sum_{t=1}^{T} \hat{u}_j(t)[-\frac{d}{d\alpha_j} \log \Gamma(\alpha_j) + \log \beta_j + \log s_t] = 0,$$

then replacing β_j by $\hat{\beta}_j$, we derive

$$\frac{dF}{d\alpha_j} = \sum_{t=1}^{T} \hat{u}_j(t)[-\frac{d}{d\alpha_j} \log \Gamma(\alpha_j) + \log \frac{\alpha_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t) s_t} + \log s_t] = 0.$$

In order to estimate the above equation numerical maximization is required. Below, derivative of F with respect to p_j is obtained

$$\frac{dF}{dp_j} = \sum_{t=1}^{T} \hat{u}_j(t)[\frac{r_j}{p_j} - \frac{n_t - r_j}{1 - p_j}] = 0,$$

and so

$$\frac{r_j}{p_j}\sum_{t=1}^{T}\hat{u}_j(t) = \sum_{t=1}^{T}\frac{n_t - r_j}{1-p_j}\hat{u}_j(t),$$

then

$$\frac{1-p_j}{p_j} = \frac{\sum_{t=1}^{T}\hat{u}_j(t)(n_t - r_j)}{r_j\sum_{t=1}^{T}\hat{u}_j(t)}.$$

The parameter $\hat{p}_j$:

$$\hat{p}_j = \frac{r_j\sum_{t=1}^{T}\hat{u}_j(t)}{\sum_{t=1}^{T}\hat{u}_j(t)(n_t - r_j) + r_j\sum_{t=1}^{T}\hat{u}_j(t)}. \tag{3.28}$$

Maximization of the state-dependent part of CDLL with respect to r_j proceeds as follows

$$\frac{dF}{dr_j} = \sum_{t=1}^{T}\hat{u}_j(t)[\frac{d}{dr_j}\log\binom{n_j-1}{r_j-1} + \log p_j - \log(1-p_j)] = 0$$

then replacing r_j by $\hat{r}_j$, we get

$$\begin{aligned}\frac{dF}{dr_j} = \sum_{t=1}^{T}\hat{u}_j(t)[\frac{d}{dr_j}\log\binom{n_j-1}{r_j-1} + \log\frac{r_j\sum_{t=1}^{T}\hat{u}_j(t)}{\sum_{t=1}^{T}\hat{u}_j(t)(n_t - r_j) + r_j\sum_{t=1}^{T}\hat{u}_j(t)} \\ - \log(1 - \frac{r_j\sum_{t=1}^{T}\hat{u}_j(t)}{\sum_{t=1}^{T}\hat{u}_j(t)(n_t - r_j) + r_j\sum_{t=1}^{T}\hat{u}_j(t)})] = 0,\end{aligned} \tag{3.29}$$

which requires numerical tools to maximize the expression. For confidence that the estimated parameters maximize the state-dependent part of CDLL, we check second derivatives of F with respect to parameters. The second derivative of F with respect to the parameters of Gamma distribution, we have

$$\frac{d^2F}{d\alpha_j^2}\Big|_{\alpha_j=\hat{\alpha}_j} = -\sum_{t=1}^{T}\hat{u}_j(t)\frac{d^2}{d\alpha_j^2}\log\Gamma(\alpha_j)\Big|_{\alpha_j=\hat{\alpha}_j} < 0,$$

since the trigamma function, defined as the sum of the series, is positive

$$\frac{d^2}{d\alpha_j^2}\log\Gamma(\alpha_j) = \sum_{k=0}^{\infty}\frac{1}{(\alpha_j+k)^2} > 0.$$

Then, we check the second derivative of F with respect to β_j^2

$$\frac{d^2F}{d\beta_j^2}\Big|_{\beta_j=\hat{\beta}_j} = -\sum_{t=1}^{T}\hat{u}_j(t)\frac{\alpha_j}{\beta_j^2} < 0, \quad \text{since} \quad \alpha_j > 0.$$

Next, the second derivative of F with respect to parameters of Negative Binomial distribution is considered.

$$\frac{d^2F}{dp_j^2}\Big|_{p_j=\hat{p}_j} = \sum_{t=1}^{T}\hat{u}_j(t)[-\frac{r_j}{p_j^2} - \frac{n_t-r_j}{(1-p_j)^2}] \tag{3.30}$$

In order to maximize the state-dependent term with respect to p_j it is necessary to prove the following inequality

$$\frac{(1-p_j)^2}{p_j^2} > -\frac{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j)}{r_j \sum_{t=1}^{T} \hat{u}_j(t)}.$$

According to the Equation 3.28, we have

$$-\frac{[\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j)]^2}{r_j^2[\sum_{t=1}^{T} \hat{u}_j(t)]^2} < \frac{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j)}{r_j \sum_{t=1}^{T} \hat{u}_j(t)}.$$

Therefore,

$$\frac{\sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j)} + 1 > 0$$

According to estimated $\hat{p}_j$, it follows, that

$$\frac{\hat{p}_j}{r_j(1-\hat{p}_j)} + 1 > 0,$$

which is true, since $r_j > 0$ and $\hat{p}_j \in (0,1)$. In the following, we examine r_j

$$\frac{dF^2}{dr_j^2}\Big|_{r_j=\hat{r}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{d^2}{dr_j^2} \log \binom{n_t - 1}{r_j - 1}$$

Using the first derivative of $\log \binom{n_t-1}{r_j-1}$, we have

$$\frac{d}{dr_j} \log \binom{n_t - 1}{r_j - 1} = \sum_{i=0}^{r_j-2} \frac{1}{n_t - 1 - i}$$

The series sum is equal to

$$\frac{1}{n_t - 1} + \frac{1}{n_t - 2} + ... + \frac{1}{n_t - r_j - 3}$$

So the second derivative is equal to

$$\frac{d^2}{dr_j^2} \log \binom{n_t - 1}{r_j - 1} = ... + \frac{r_j}{(n_t - 3 - r_j)^2},$$

and it is obvious that the expression is positive. □

3.5 Conditional distribution

In this section, we give an account of conditional distributions, that are convenient for assessing outliers or forecasting.

We refer to $N^{(-t)}$ and $S^{(-t)}$ as the observations at all times other than t, defining

$$N^{(-t)} \equiv (N_1, ..., N_{t-1}, N_{t+1}, ..., N_T),$$

and similarly $S^{(-t)}, n^{(-t)}, s^{(-t)}$.

Conditional distribution of (S_t, N_t) given all the other observations of bivariate HMM can be computed as follows:

$$\begin{aligned} &P((S_t, N_t) = (s, n) | (S^{(-t)}, N^{(-t)}) = (s^{(-t)}, n^{(-t)})) \\ &\quad = \frac{P((S_t, N_t) = (s, n), (S^{(-t)}, N^{(-t)}) = (s^{(-t)}, n^{(-t)}))}{P((S^{(-t)}, N^{(-t)}) = (s^{(-t)}, n^{(-t)}))}. \end{aligned} \tag{3.31}$$

According to the likelihood of a bivariate HMM and the definition of the forward and backward probabilities, for $t = 2, 3, ..., T$, it follows, that

$$\begin{aligned} &P((S_t, N_t) = (s, n) | (S^{(-t)}, N^{(-t)}) = (s^{(-t)}, n^{(-t)})) \\ &= \frac{u(1)P(s_1, n_1)\Gamma P(s_2, n_2) \cdots \Gamma P(s_{t-1}, n_{t-1})\Gamma P(s, n)\Gamma P(s_{t+1}, n_{t+1}) \cdots \Gamma P(s_T, n_T)1'}{u(1)P(s_1, n_1)\Gamma P(s_2, n_2) \cdots \Gamma P(s_{t-1}, n_{t-1})\Gamma P(s_{t+1}, n_{t+1}) \cdots \Gamma P(s_T, n_T)1'} \\ &= \frac{\alpha_{t-1}\Gamma P(s, n)\beta_t^{'}}{\alpha_{t-1}\beta_t^{'}}. \end{aligned} \tag{3.32}$$

For the case $t = 1$,

$$\begin{aligned} &P((S_1, N_1) = (s, n) | (S^{(-1)}, N^{(-1)}) = (s^{(-1)}, n^{(-1)})) \\ &\quad = \frac{u(1)P(s, n)\Gamma P(s_2, n_2) \cdots \Gamma P(s_T, n_T)1'}{u(1)I\Gamma P(s_2, n_2) \cdots \Gamma P(s_T, n_T)1'} \\ &\quad = \frac{u(1)P(s, n)\beta_1^{'}}{u(1)I\beta_1^{'}}. \end{aligned} \tag{3.33}$$

The conditional distribution can be expressed as the mixture of the m joint state-dependent distributions in the following form [29]

$$P((S_t, N_t) = (s, n) | (S^{(-t)}, N^{(-t)}) = (s^{(-t)}, n^{(-t)})) = \sum_{i=1}^{m} \frac{f_i(t)}{\sum_{j=1}^{m} f_i(t)} p_i(s, n),$$

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

Additionaly, we are interested in conditional distributions of S_t and N_t. The derivation is presented only for discrete variable, the continuous case can be performed analogously.

Conditional distribution of N_t given $N^{(-t)}$ can be derived as:

$$P(N_t = n|N^{(-t)} = n^{(-t)}) = \frac{P(N_t = n, N^{(-t)} = n^{(-t)})}{P(N^{(-t)} = n^{(-t)})},$$

and hence that, for $t = 2, 3, ..., T$

$$\begin{aligned} P(N_t &= n|N^{(-t)} = n^{(-t)}) \\ &= \frac{u(1)P(n_1)\Gamma P(n_2)\cdots\Gamma P(n_{t-1})\Gamma P(n)\Gamma P(n_{t+1})\cdots\Gamma P(n_T)1'}{u(1)P(n_1)\Gamma P(n_2)\cdots\Gamma P(n_{t-1})\Gamma P(n_{t+1})\cdots\Gamma P(n_T)1'} \\ &= \frac{\alpha_{t-1}\Gamma P(n)\beta_t'}{\alpha_{t-1}\beta_t'}. \end{aligned} \tag{3.34}$$

For the case $t = 1$,

$$\begin{aligned} P(N_1 = n^{(-1)} = n^{(-1)}) &= \frac{u(1)P(n)\Gamma P(n_2)\cdots\Gamma P(n_T)1'}{u(1)I\Gamma P(n_2)\cdots\Gamma P(n_T)1'} \\ &= \frac{u(1)P(n)\beta_1'}{u(1)I\beta_1'}. \end{aligned} \tag{3.35}$$

3.6 Forecast distributions

Using a bivariate HMM it is possible to make forecasts. Applicable expression for conditional distribution of (S_{T+h}, N_{T+h}) with forecast horizon h given all observations of the model is available. It can be computed as a ratio of likelihoods [29]:

$$\begin{aligned} P((S_{T+h}, N_{T+h}) &= (s, n)|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) \\ &= \frac{(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})), P((S_{T+h}, N_{T+h}) = (s, n)}{P((S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)}))} \\ &= \frac{u(1)P(s_1, n_1)\Gamma P(s_2, n_2)\cdots\Gamma P(s_T, n_T)\Gamma^h P(s, n)1'}{u(1)P(s_1, n_1)\Gamma P(s_2, n_2)\cdots\Gamma P(s_T, n_T)1'} \\ &= \frac{\alpha_T\Gamma^h P(s, n)1'}{\alpha_T 1'}. \end{aligned} \tag{3.36}$$

Moreover, the forecast distribution can be determined as the mixture of the joint state-dependent distributions:

$$P((S_{T+h}, N_{T+h}) = (s, n)|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = \sum_{i=1}^{m} \psi_i(h)p_i(s, n),$$

where the $\psi_i(h)$ is the ith entry of the vector $\frac{\alpha_T\Gamma^h}{\alpha_T 1'}$.

3.7 Decoding

In this section we consider two types of the decoding problem: the *local decoding* indicates the most probable state at a particular time, while the *global decoding* determines the most likely sequence of states.

3.7.1 State probabilities and local decoding

Here we define the conditional distribution of C_t given the observations, for $i = 1, 2, ..., m$, as

$$\begin{aligned} P(C_t = i|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) &= \frac{P(C_t = i, (S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)}))}{P((S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)}))} \\ &= \frac{\alpha_t(i)\beta_t(i)}{L_T}. \end{aligned} \tag{3.37}$$

For each $t = 1, 2, ...T$ the most likely state i_t^* given $S^{(T)}, N^{(T)}$ can be obtained as

$$i_t^* = \underset{i=1,...,m}{\operatorname{argmax}} P(C_t = i|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})).$$

3.7.2 Global decoding

Global decoding is more applicable than local decoding. It detects the sequence of hidden states $c_1, c_2, ..., c_T$ which maximizes the conditional probability

$$P(C^{(T)} = c^{(T)}|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)}));$$

or the joint probability:

$$P(C^{(T)}, (S^{(T)}, N^{(T)})) = \delta_{c_1} \prod_{t=2}^{T} \gamma_{c_{t-1},c_t} \prod_{t=1}^{T} p_{c_t}(s_t, n_t).$$

Since the decoding is not executable for large T, the ***Viterbi algorithm*** [36], [37] is used instead. First, we define

$$\psi_{1i} = P(C_1 = i, (S_1, N_1) = (s_1, n_1)) = \delta_i p_i(s_1, n_1,)$$

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

$$\psi_{ti} = \max_{c_1, c_2, ..., c_{t-1}} P(C^{(t-1)} = c^{(t-1)}, C_t = i, (S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})).$$

The recursion formula for ψ_{ti}, for $t = 2, 3, ..., T$ and $i = 1, 2, ..., m$ is given below

$$\psi_{tj} = (\max_i(\psi_{t-1,i}\gamma_{ij}))p_j(s_t, n_t).$$

Then, the most likely state sequence can be determined recursively from

$$i_T = \operatorname*{argmax}_{i=1,...,m} \psi_{Ti}$$

and, for $t = T-1, T-2, ..., 1$, from

$$i_t = \operatorname*{argmax}_{i=1,...,m}(\psi_{ti}\gamma_{i,i_{t+1}}).$$

3.8 State prediction

The state prediction can be performed by the following expression [29],for $i = 1, 2, ..., m$

$$P(C_{T+h} = i|(S^{(T)}, N^{(T)}) = (s^{(T)}, n^{(T)})) = \alpha_T \Gamma^h(, i)/L_T,$$

where $\Gamma^h(, i)$ denotes the ith column of the matrix Γ^h and the time horizon h is equal to $t - T$.

3.9 Model selection

According to the likelihood of the bivariate HMM, the increasing number of states m yields a better fit of the model, whereas it may cause a quadratic increase in the number of parameters to be estimated. Consequently, a criterion for model selection is necessary. In this section, we give a brief summary of model selection and the use of pseudo-residuals.

3.9.1 AIC and BIC criterions

The Akaike Information Criterion (AIC) is a method of selecting an appropriate model from a set of models. The model which minimizes the Kullback-Leibler distance between the model and the truth is assumed to be at some point the superior model. However, AIC does not provide an information about the general quality of selected model. The Akaike Information Criterion is defined as [38]:

$$\text{AIC} = -2\log L + 2p, \tag{3.38}$$

where $\log L$ is the log-likelihood of the fitted model and p is the number of free parameters in the model. The first term rewards goodness of fit of the model and decreases with increasing m, while the second term, defined as the penalty term, is an increasing function of the number of states m. The penalty prevents overfitting of the model. The preferred model is the one with the minimum AIC value.

The Bayesian Information Criterion (BIC) was developed by Gideon E. Schwarz as the approach to model selection among a set of models. As the Akaike Information Criterion BIC resolves the problem of overfitting. BIC differs from AIC in the penalty term [39]:

$$\text{BIC} = -2\log L + p\log T, \tag{3.39}$$

where $\log L$ is the log-likelihood of the fitted model, p and T denotes the number of parameters and the of observations in the model, respectively.

3.9.2 Pseudo-residuals

Despite the fact that the model opted by AIC or BIC criterion is supposed to be the most appropriate model, the goodness of fit of the model in an absolute sense is not assessed. An optimal way to do so is to obtain pseudo-residuals, which are also able to identify outliers relative to the model. We consider ***ordinary pseudo-residuals***.

The ordinary pseudo-residuals are based on the conditional distribution given all other observations [29]. The normal pseudo-residual is defined as

$$z_t = \Phi^{-1}(P(S_t \le s_t | S^{-t} = s^{-t})).$$

Normal pseudo-residuals are standard normally distributed if the related model is correct. The conditional probabilities are given by Equations 3.31 and 3.32 in Section 3.5.

CHAPTER 4

APPLICATION OF THE BIVARIATE HMM: AUTOMOBILE INSURANCE

Although the total claim amounts distribution in literature is not commonly taken as normal, the theory developed on HMM concentrates on the Poisson-normal case. For this reason, this assumption is taken as the first choice. However, surprisingly, the claim data analyses also supports the assumption on normality. Therefore, the case study is done on the Poisson-normal case at the first sight.

The Poisson-Normal HMM described in the previous chapter can be applied in many forms of insurance, where dependence among observations exists. For example, in a private household or motor insurance. In our study, we compare the Poisson-Normal HMMs with different state numbers and fit the most optimal model to the vehicle insurance data. According to the selected model, we derive forecast distributions, conduct local and global decoding and predict states. In this chapter, statistical analysis on the dataset is provided as well.

4.1 Data description

The dataset, which is analyzed in this thesis, is provided by the Traffic Insurances Information and Monitoring Center (TRAMER). The dataset contains compulsory auto insurance recordings from all over Turkey. Every policy is registered to the system with details: policy number (anonymous), insured id number, start date of policy, end date of policy, vehicle tariff group code (car, minibus, taxi, etc.), country licence code, vehicle id number, vehicle age, usage of vehicle (private, commercial), passenger capacity, nationality information of insured, damage date, claim reason, claim amount.

As the number of policies is numerous and not easy to handle, Istanbul is taken as the sample province to apply the analyses. The choice on Istanbul is due to its highly populated, industrialized and high rate of insurance penetration position compared to other cities in Turkey.

The dataset includes information about claims, which policies starts from 2006 to 2009, and accident year varies from 2005 to 2011. In order to examine the dataset, we constructed two tables. Table 4.1 provides a total number of claims for each year and Table 4.2 gives details about a total claim amount for each year. Cells of the tables contain information about claims occurred in a certain accident year, which policies start in a certain year, e.g. there were 210,612 accidents resulted in total loss of 337,879,194 TL, which policies started in 2008 and accidents occurred in 2009. The observations, which are significant to be used in our study, are indicated in bold.

Table4.1: Claim numbers

Start date of a policy\Accident year	2005	2006	2007	2008	2009	2010	2011
2006	NA	196,046	**191,902**	314	128	12	NA
2007	NA	NA	**187,360**	**199,467**	239	31	NA
2008	NA	NA	NA	**198,368**	**210,612**	276	12
2009	NA	NA	NA	NA	**204,166**	196,746	59

As a result, the database used in the application provides information on automobile insurance portfolios from Istanbul over the period January 2007 to December 2009. In our study, we consider only non-zero claims. We focus on the monthly aggregate claim amounts, therefore, for each month t, we aggregate all the claim amounts occurring at the same month. Likewise, we obtain the

Table4.2: Aggregate claim amounts

Start date of a policyyear	2005	2006	2007	2008	2009	2010	2011
2006	NA	290,861,223	**257,040,552**	444,279	132,221	9,098	NA
2007	NA	NA	**290,537,933**	**291,431,736**	505,044	51,846	NA
2008	NA	NA	NA	**320,698,956**	**337,879,194**	543,344	21,553
2009	NA	NA	NA	NA	**502,834,660**	336,400,909	207,676

total number of claims occurred in each month.

For convenience during this chapter, we refer to the *'total number of claims'* and the *'monthly aggregate claim amounts'* as the *'claim numbers'* and the *'claim amounts'*, respectively.

The individual claim amounts were adjusted for inflation. Annual inflation rates, which are taken from [31], are 9,67% for 2008 and 6,21% for 2009. Having adjusted the dataset for inflation, we truncated the individual claim amounts, considering only variables greater than 250.

Having removed the duplicated observations, observations with same policy numbers are not detected.

The individual claim amount valued at 99,000,500 extremely differs from the other observations in the dataset and therefore, it was replaced by the mean of individual claim amounts reported in June 2009. Additionally, on 8th and 9th of September in 2009 Istanbul exposed to massive flooding, which caused the enormous number of claims. According to the fact that such flooding occurs only once a hundred years, [32], the observations at these days are recognized as an outlier of the dataset. Thus, the number of claims occurred in these days were decreased to the average number of claims occurred in September 2009 and total monetary value of the claims was replaced by the average claim amounts of the month. Also, in next three months, we notice an unusual pattern in claim behaves. We suppose, that those have been affected by the flooding, therefore, we modified these observations.

Having modified the dataset, as a result, 809,327 policies are used in our study.

In the following, we present some summary statistics, the distribution plots and the scatter plots of the claim amounts and the claim numbers. Furthermore, the corresponding autocorrelations are analyzed.

Figure 4.1 displays a histogram of the claim amounts, variables are given in thousands. We observe that the amounts are accumulated between above 44,000 and 52,000, the data has a roughly symmetric shape. Moreover, the Shapiro-

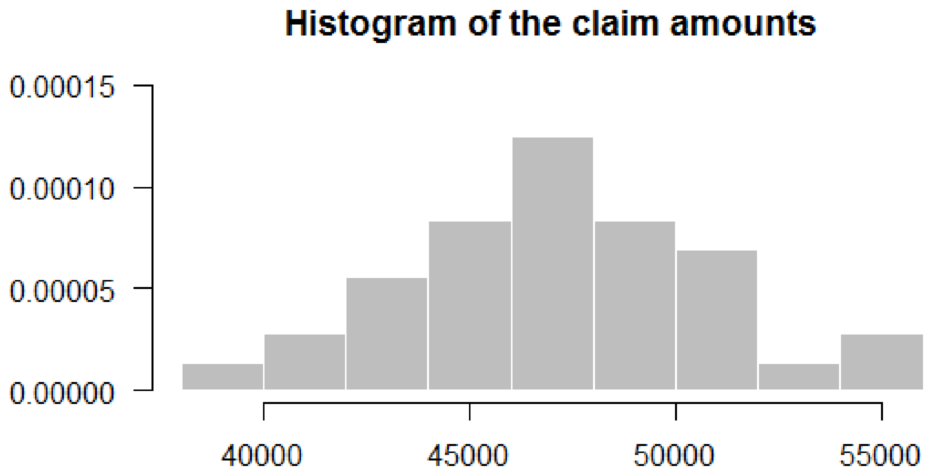

Figure 4.1: Histogram of the claim amounts for Istanbul, 2007-2009 (in thousands).

Wilk normality test indicates that the claim amounts are normally distributed with p-value equal to 0.8414.

According to Figure 4.2, the histogram of the claim numbers indicates a right-skewed distribution.

In the Table 4.3 we provide a descriptive statistics of the claim amounts and the claim numbers aggregated monthly. Mean,mode and median values of the claim amounts are almost equal, that is indicative for a symmetric distribution. The summary statistics of the claim numbers infers that the data follows a right-skewed distribution, since the mean value greater than the median value, and the mode value less than the median value.

Table4.3: Descriptive statistics of the claim amounts and the claim numbers aggregated monthly

Observations	Mean	Median	Minimum	Maximum	Standard deviation	Mode
Claim amounts	46,957,468	47,367,916	38,844,907	55,226,741	3,811,343	46,957,470
Number of claims	22,481.31	22,173	19,429	25,316	1,612.77	21,550

Autocorrelation functions detect the presence of serial dependence in claim counts as well as in claim sizes. See Figure 4.3 and Figure 4.4.

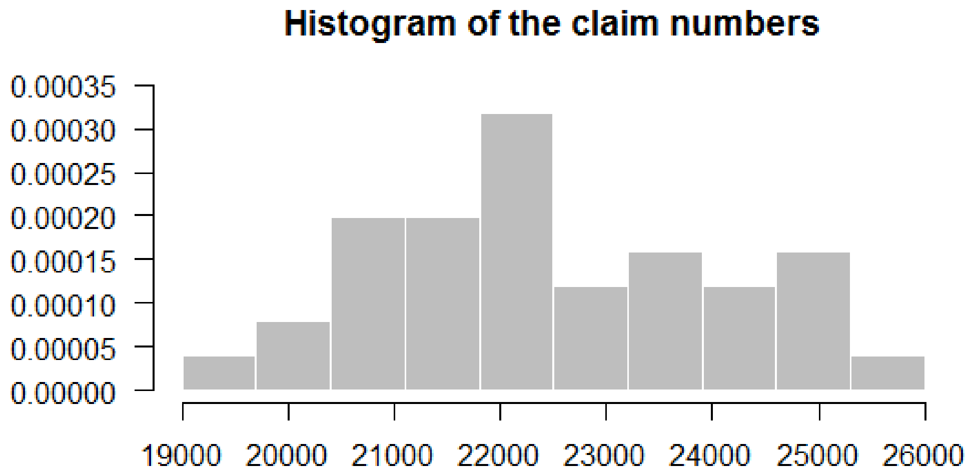

Figure 4.2: Histogram of the claim numbers for Istanbul, 2007-2009.

According to the fact that the ACFs of claim numbers and claim amounts drop to zero relatively slowly, we infer that the observations are non-stationary. Additionally, for the confidence of our conclusions Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and Augmented Dickey-Fuller (ADF) test were performed.

The null-hypothesis for an ADF test is that the data are non-stationary. So large p-value is indicative of non-stationarity, and small p-value suggests stationarity. Using the 5% threshold, we determine that time series of claim numbers (p-value=0.3986) and claim amounts (p-value=0.2081) are non-stationary. The alternative hypothesis for KPSS test is that the data are non-stationary, large p-value specifies stationarity, and conversely small p-value indicates non-stationarity. Performing KPSS test we also reject stationarity since both p-values for claim size and claim counts less than 0.01.

Figures 4.8 and 4.9 depict plots of the claim amounts and the claim numbers, respectively, having time points on horizontal axis. In both graphs we observe a weak sinusoidal behaviour, which might indicate seasonality. Additionally, slight increasing trend is indicated.

The Table 4.4 provides an information about aggregated claim amounts and claim numbers in each month, that we use as the observed variables in bivariate

Table4.4: Claim amounts and claim numbers reported in Istanbul during 2007-2009

Month-Year	**Claim amounts**	**Number of claims**
January, 2007	44771818	21213
February, 2007	40610503	19429
March, 2007	41727487	20940
April, 2007	42391254	21406
May, 2007	44716296	22712
June, 2007	44892121	22181
July, 2007	47209622	22300
August, 2007	47575857	21288
September, 2007	44063532	21336
October, 2007	50647142	22165
November, 2007	44193298	20895
December, 2007	47526209	20049
January, 2008	48004924	21400
February, 2008	42781346	20398
March, 2008	38844907	20686
April, 2008	42448010	21846
May, 2008	46063021	23291
June, 2008	47667936	23721
July, 2008	47053227	21934
August, 2008	48005758	22445
September, 2008	48554476	22160
October, 2008	48219700	23154
November, 2008	44724944	21810
December, 2008	48162470	20782
January, 2009	48399958	23982
February, 2009	42908270	20860
March, 2009	46708121	23158
April, 2009	46014566	23876
May, 2009	51292119	25231
June, 2009	51553338	25316
July, 2009	55226741	24403
August, 2009	52834942	24519
September, 2009	54577870	24615
October, 2009	50835988	25000
November, 2009	51645853	24926
December, 2009	47615231	23900

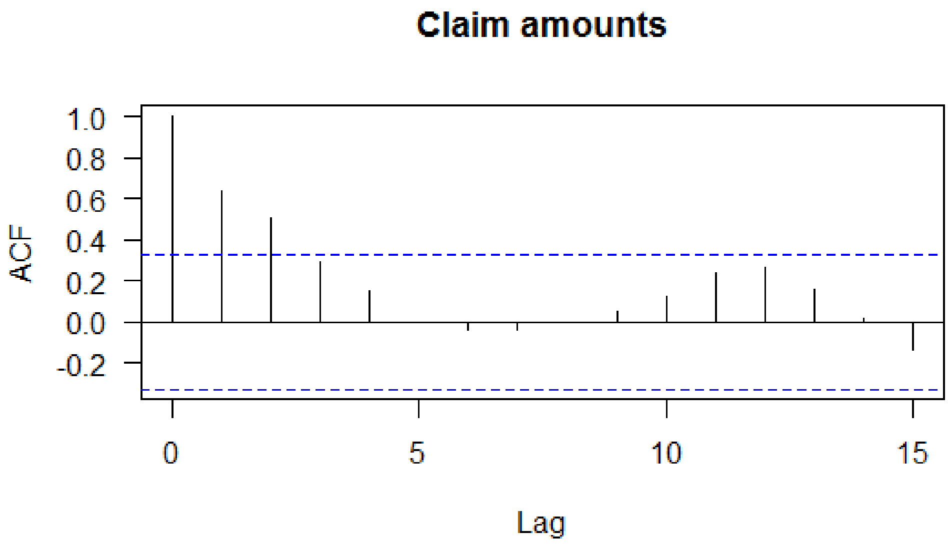

Figure 4.3: Autocorrelation function of claim amounts for Istanbul, 2007-2009.

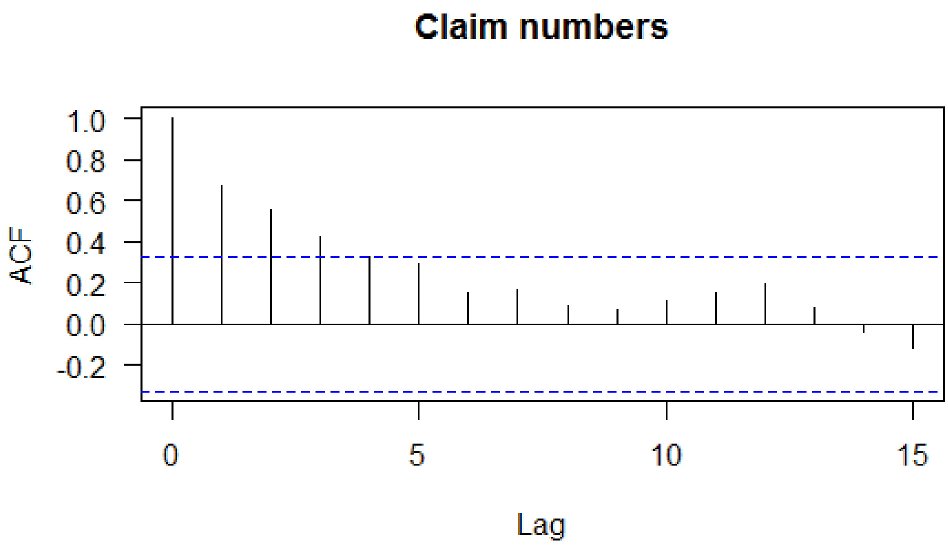

Figure 4.4: Autocorrelation function of claim numbers for Istanbul, 2007-2009.

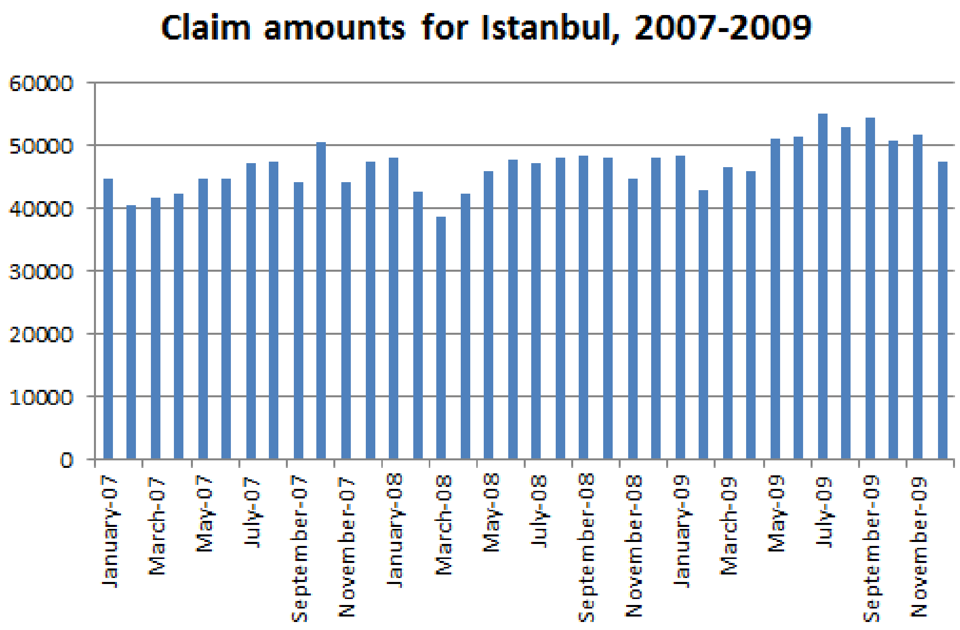

Figure 4.5: The monthly aggregate claim amounts (in thousands).

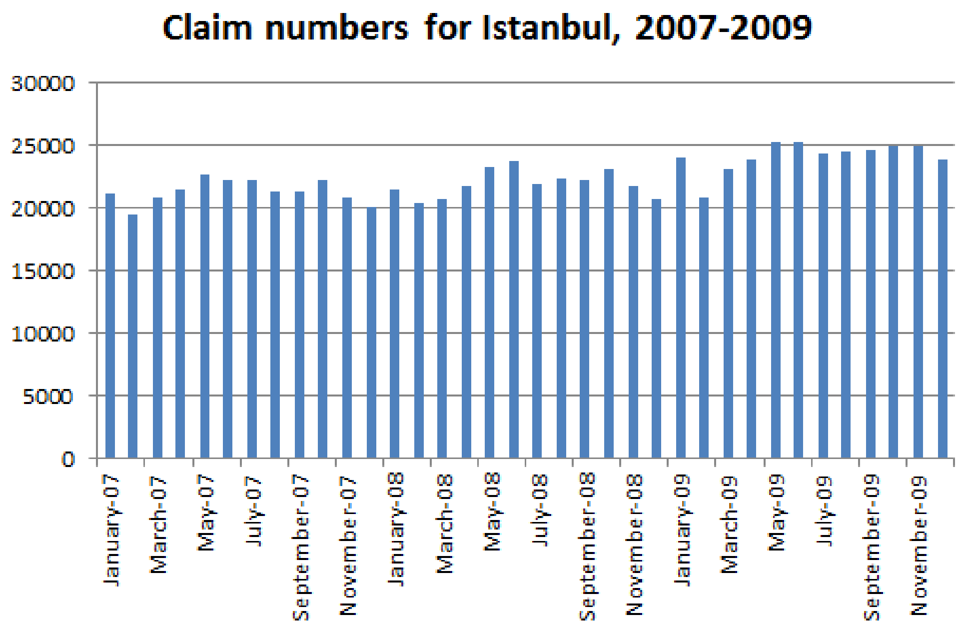

Figure 4.6: The monthly aggregate claim numbers.

Table4.5: Descriptive statistics of individual claim amounts

Month-Year	Mean	Standard deviation	Variance	Minimum	Maximum	Median
January, 2007	2110.58	5509.23	30351602	250	421475	942
February, 2007	2090.20	4628.33	21421412	250	169535	952
March, 2007	1992.72	4597.80	21139733	250	301617	930
April, 2007	1980.34	4025.99	16208572	250	161406	906
May, 2007	1968.84	4464.77	19934202	250	234138	881
June, 2007	2023.90	4709.87	22182903	250	335764	896
July, 2007	2117.02	4906.99	24078527	250	215153	900
August, 2007	2234.87	4530.28	20523393	250	121457	973
September, 2007	2065.22	4155.56	17268664	250	92000	937
October, 2007	2285.01	5331.28	28422502	250	346430	1000
November, 2007	2115.02	4672.96	21836555	250	218079	994
December, 2007	2370.50	5907.35	34896730	250	295000	1020
January, 2008	2243.22	5153.35	26557032	250	168290	977
February, 2008	2097.33	4522.43	20452372	250	207727	949
March, 2008	1877.84	3876.24	15025267	250	100535	881
April, 2008	1943.06	4303.13	18516917	250	140429	899
May, 2008	1977.72	4462.17	19910916	250	180998	912
June, 2008	2009.53	4682.30	21923894	250	230133	912
July, 2008	2145.22	4917.63	24183069	250	194219	912
August, 2008	2138.82	4556.68	20763344	250	116340	920
September, 2008	2191.09	4662.06	21734826	250	159096	956
October, 2008	2082.57	4518.34	20415431	250	197866	941
November, 2008	2050.66	4135.15	17099428	250	135562	922
December, 2008	2317.51	5100.18	26011875	250	164314	1011
January, 2009	2018.18	4937.95	24383303	250	276148	910
February, 2009	2056.96	4899.80	24008046	250	163609	943
March, 2009	2016.93	4779.61	22844622	250	180908	892
April, 2009	1927.23	4253.52	18092444	250	194067	878
May, 2009	2032.90	5090.75	25915742	250	185538	870
June, 2009	2036.39	4952.04	24522647	250	202796	871
July, 2009	2263.11	6003.62	36043442	250	365024	905
August, 2009	2154.86	5199.91	27039104	250	206124	916
September, 2009	2206.72	5779.25	33399674	250	329809	934
October, 2009	2033.44	5198.26	27021897	250	273127	927
November, 2009	2071.97	4983.31	24833383	250	185294	936
December, 2009	1992.27	4188.98	17547541	250	170866	924

HMM.

In the Table 4.5, we present summary statistics of the individual claim amounts for each month, from January, 2007 to December, 2009. Monthly mean values are higher than monthly median values, that indicate right-skewed monthly distributions, even though total claim amounts are symmetric. As can be seen in the table, standard deviation values are higher than mean values because of the high variaton in data, the individual claim amounts rise from 250 to 421,475.

4.2 Analyses

In this section, we report the results of the application of the Poisson-Normal hidden Markov model to the vehicle insurance data. The iterative procedure of the algorithm is implemented in R code. For convenience, we altered the observations of the data by dividing the claim amounts by 1000 and the claim numbers by 100.

Table4.6: Comparison of the Poisson-Normal HMMs

Number of states	Number of parameters	logL	AIC	BIC
1	4	-656.00	1320.02	1326.35
2	10	-534.22	1088.44	1104.28
3	18	-501.39	**1038.77**	**1067.28**
4	28	-500.04	1056.07	1100.41
5	40	-498.83	1077.67	1141.01
6	54	-498.12	1104.24	1189.75

We present the Poisson-Normal HMMs with one to six states fitted by EM algorithm and compare them on the basis of two criteria, namely AIC and BIC. Also, the maximum log-likelihoods are provided. The detailed information is presented in Table 4.6.

By a one-state Poisson-Normal HMM we mean a model with independence assumption, i.e. the observations are realizations of the product of independent

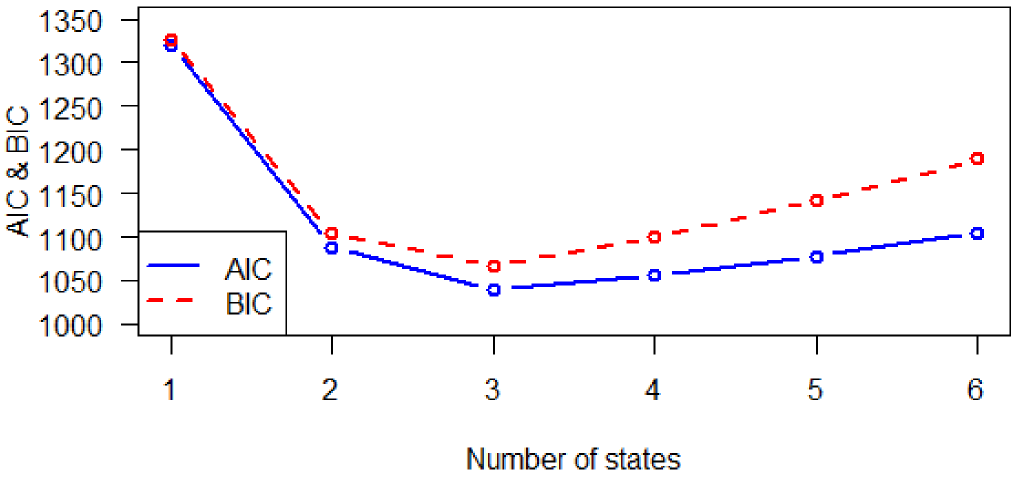

Figure 4.7: Comparison of AIC and BIC results for Poisson-Normal HMMs with different state numbers.

Poisson random variables and Normal random variables with common parameters for all time points. As regards the results in Table 4.6, the one-state model shows the weakest goodness of fit to the insurance data. Despite an increasing number of states gives a better result for log-likelihood, however, it demands more parameters to estimate. Therefore both AIC and BIC indicate that the model with three states is the most suitable, compared to other models, see Figure 4.7.

We present here a three-state Poisson-Normal HMM. For the model the stationary distribution is computed by expression 3.4 and used as the starting value of initial distribution $u(1)$.

The initial values of the off-diagonal transition probabilities are taken to be 0.1. As the starting values of the state-dependent means we use the lower quartile, median and upper quartile of the observations, for claim counts is $(212.8, 222, 239)$ and claim amounts is $(44590, 47370, 48440)$. However, it was challenging to find an optimal initial value for σ, we performed EM estimation with several starting values and selected the ones that give the maximum log-likelihood, as a result, $(5000, 2000, 1300)$ has been used.

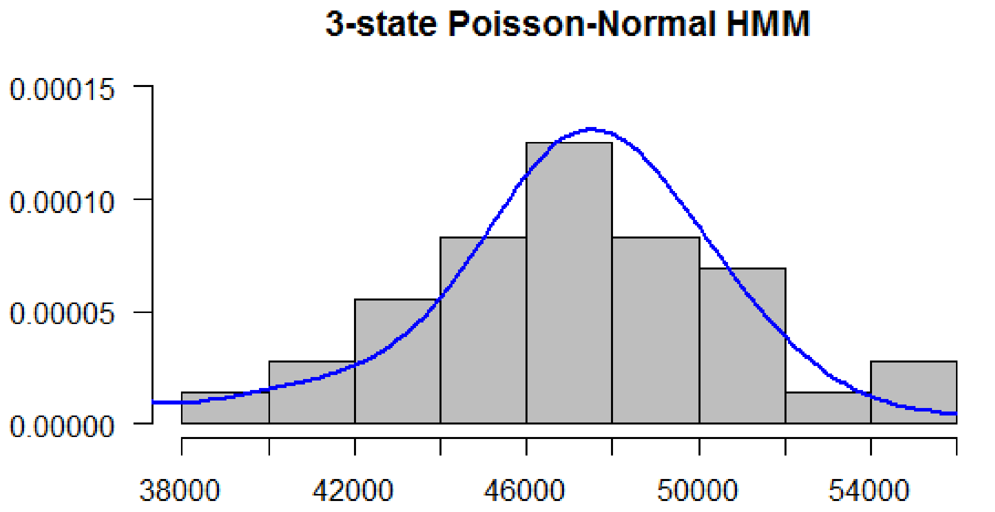

Figure 4.8: Three-state Poisson-Normal HMM: the marginal distribution for the claim amounts.

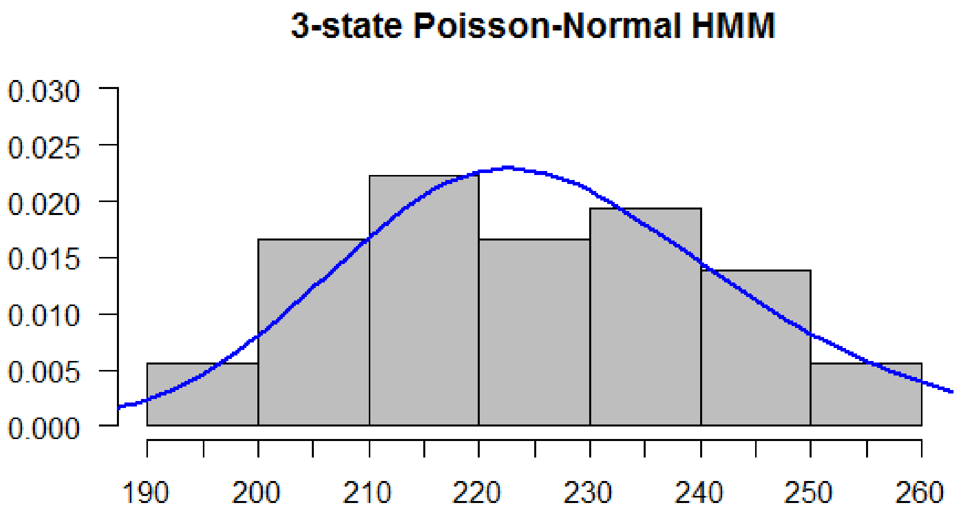

Figure 4.9: Three-state Poisson-Normal HMM: the marginal distribution for the claim numbers.

The estimated three-state model is

$$\Gamma = \begin{pmatrix} 0.7010 & 0.1961 & 0.1029 \\ 0.1213 & 0.8217 & 0.0570 \\ 0.2899 & 0.0000 & 0.7101 \end{pmatrix}$$

with initial probabilities $u(1) = (0.8565, 0.1405, 0.0030)$, parameters of joint state-dependent distribution $\lambda = (218.8, 223, 243.7)$, $\mu = (45799.45, 46991.43, 49617.07)$ and $\sigma = (5020.22, 2022.55, 2092.90)$. The estimated log-likelihood is $l = -501.387$. The marginal distributions of the selected model, compared with histograms of observations are displayed in Figure 4.8 and 4.9.

Although the 3-state model has been selected as the most appropriate model, the goodness of fit of the model in an absolute sense is not assessed.

According to Figures 4.10 and 4.11, which depict the conditional distributions for the claim numbers and the claim amounts, we observe that the shape of the conditional distributions may change significantly from one time point to another. Due to the fact that some of the observations are extreme relative to their conditional distributions, we infer that using the conditional distributions to check outliers is reasonable.

As it was mentioned in the previous chapter, we use conditional distributions to compute ordinary normal pseudo-residuals, see Figure 4.12 and 4.13. Regarding the residual plots, it is obvious that the selected model provides an optimal fit to the data. In addition, we apply Shapiro-Wilk normality test to pseudo-residuals, which confirms that those are normally distributed, p-values are 0.6546 for claim numbers and 0.4537 for claim amounts. Quantile-quantile plots of the normal pseudo-residuals provide the same result as well, see Figure 4.14 and Figure fig:qq ps-resid number .

For fitted 3-state Poisson-Normal HMM we derive state probabilities, that are necessary for performing local decoding, see Figure 4.20. Having applied 3-state model we are interested in defining hidden states that are most probable to have given rise to the sequence of observed values. We conducted local and global decodings both for claim amounts and claim numbers, definitions of methods

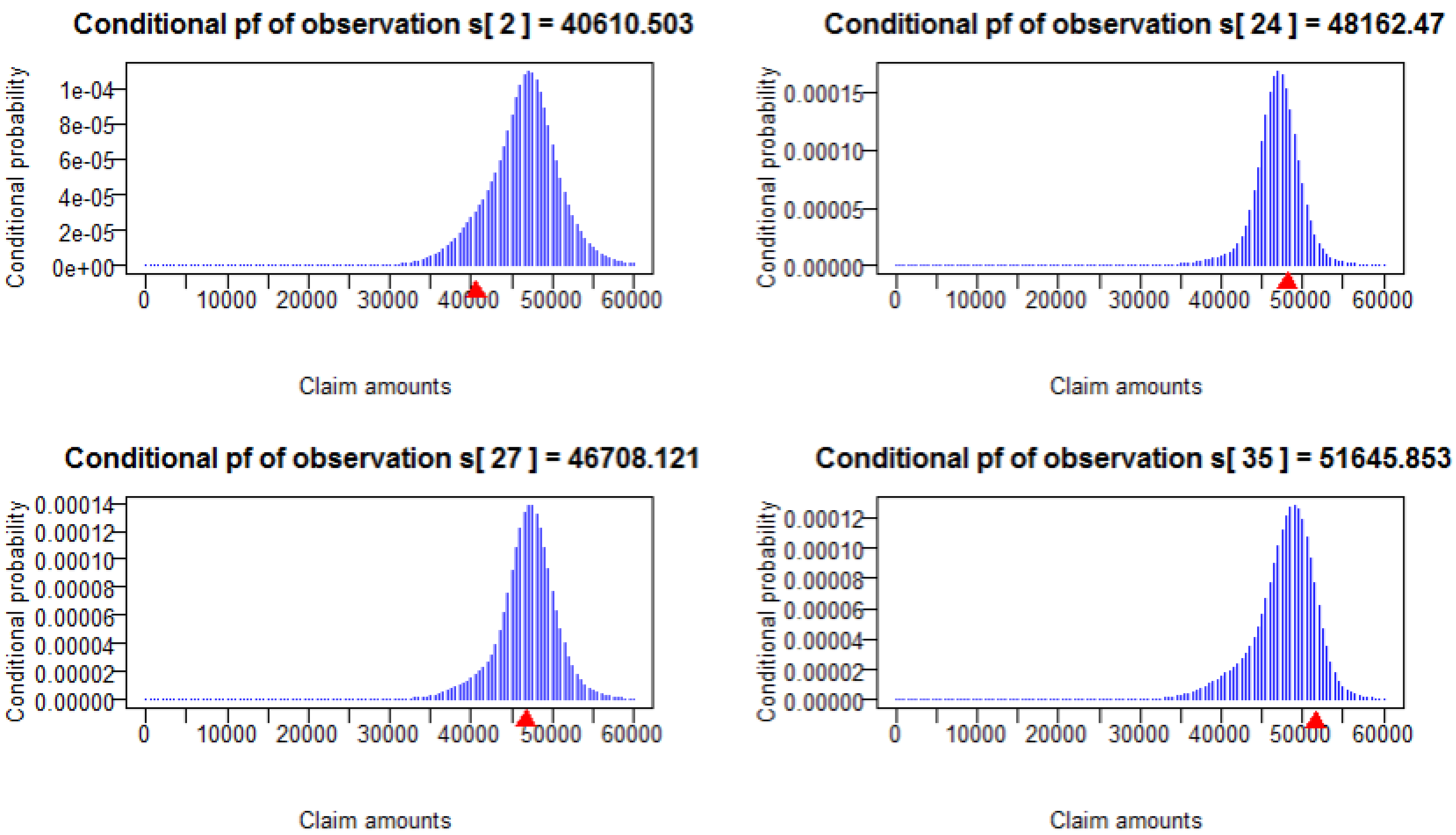

Figure 4.10: Three-state Poisson-Normal HMM: conditional distribution of the claim amounts in February 2007, December 2008, March 2009 and November 2009, given all the other observations. The triangle symbol corresponds to the actual claim amount in that month.

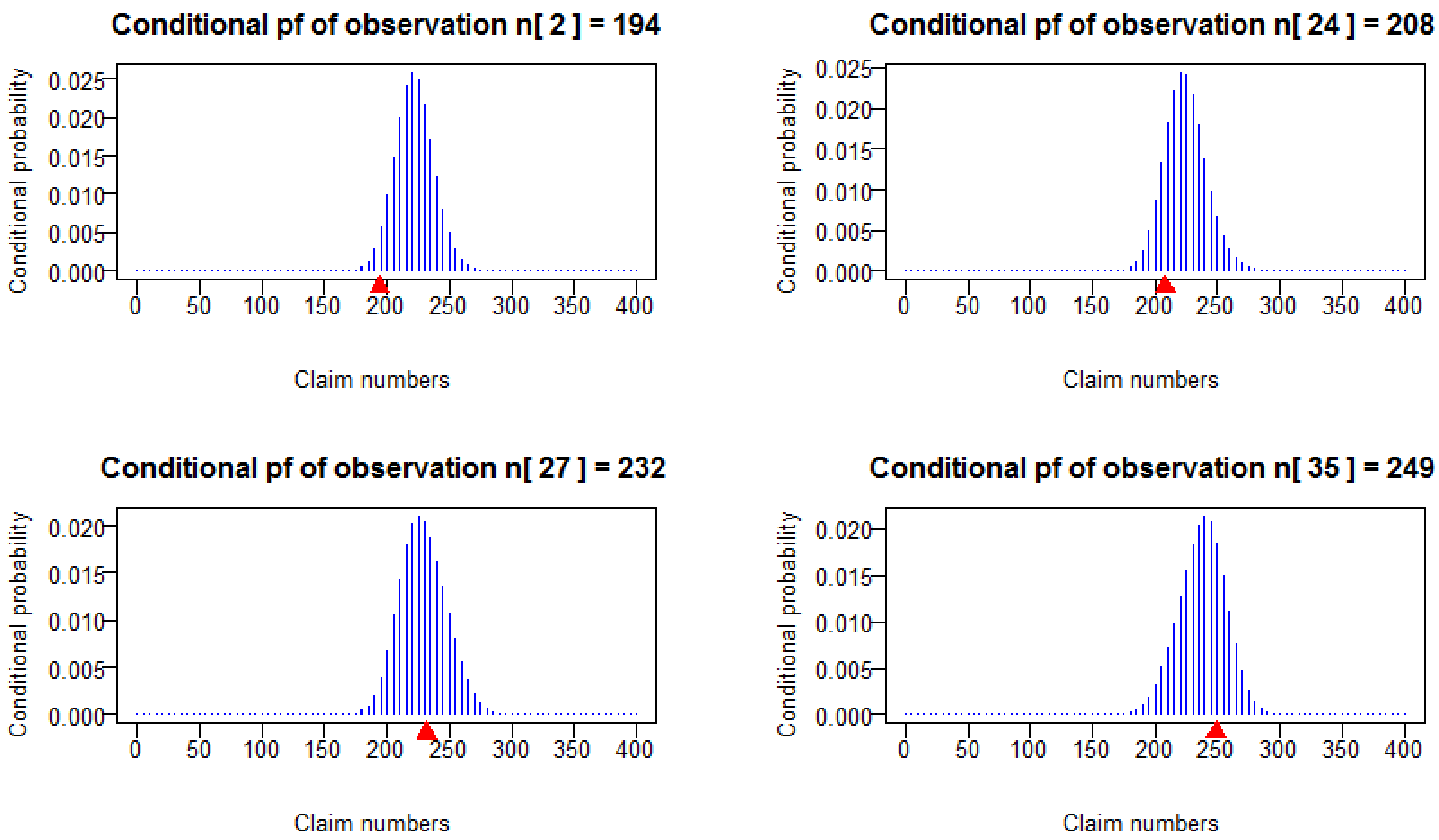

Figure 4.11: Three-state Poisson-Normal HMM: conditional distribution of the claim numbers in February 2007, December 2008, March 2009 and November 2009, given all the other observations. The triangle symbol corresponds to the actual claim number in that month.

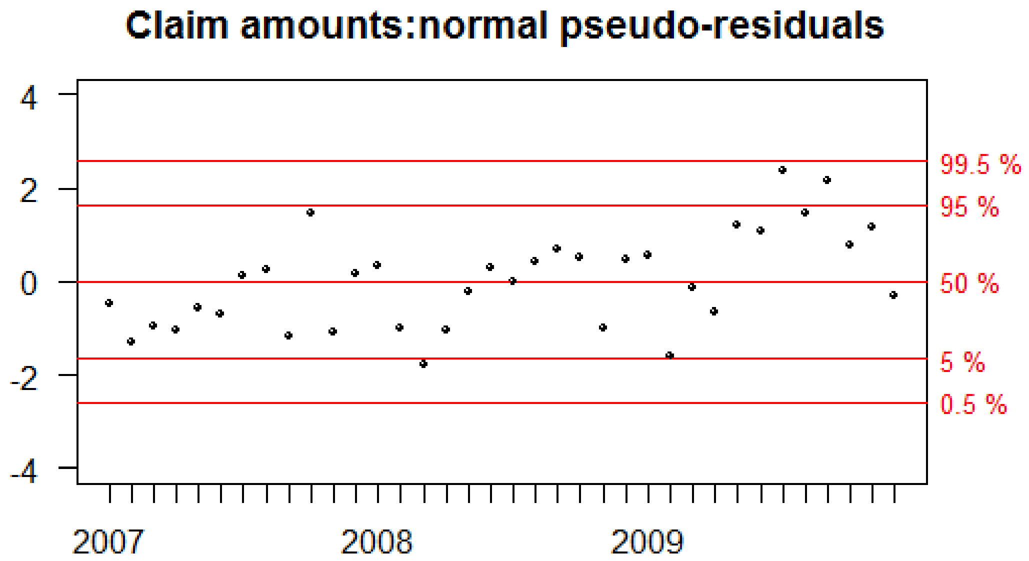

Figure 4.12: Claim amounts: ordinary pseudo-residuals. Index plot of the normal pseudo-residuals,with horizontal lines at $0, \pm 1.96, \pm 2.58$.

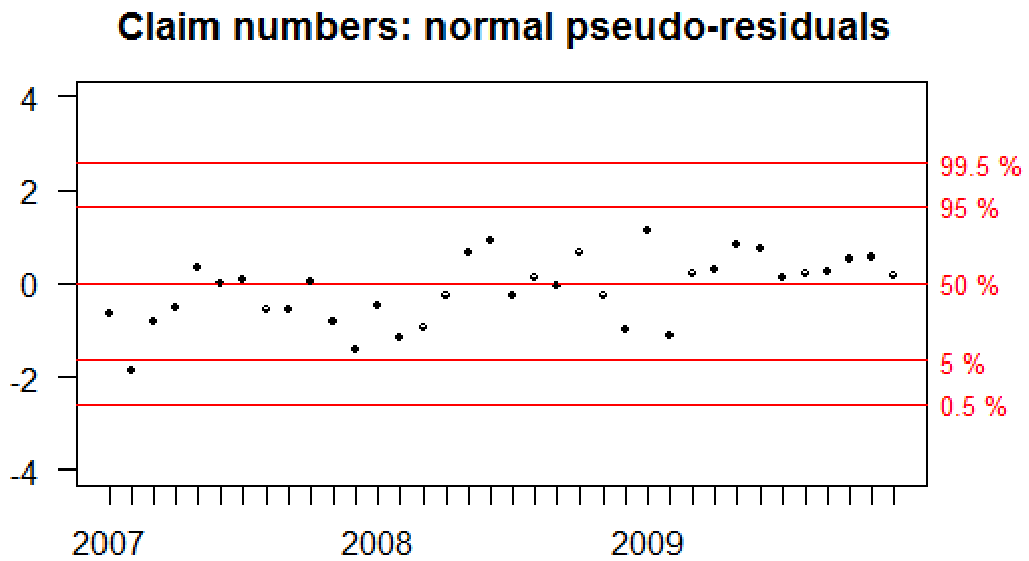

Figure 4.13: Claim numbers: ordinary pseudo-residuals. Index plot of the normal pseudo-residuals,with horizontal lines at $0, \pm 1.96, \pm 2.58$.

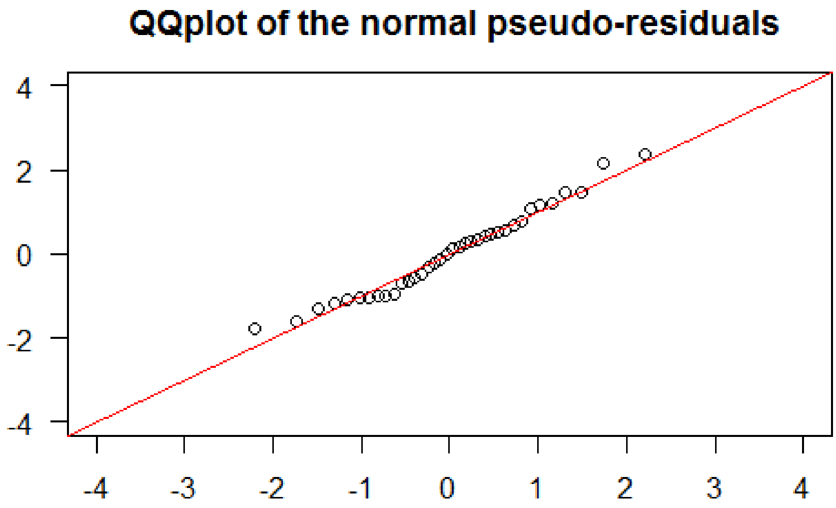

Figure 4.14: Claim amounts: QQplot of the ordinary normal pseudo-residuals.

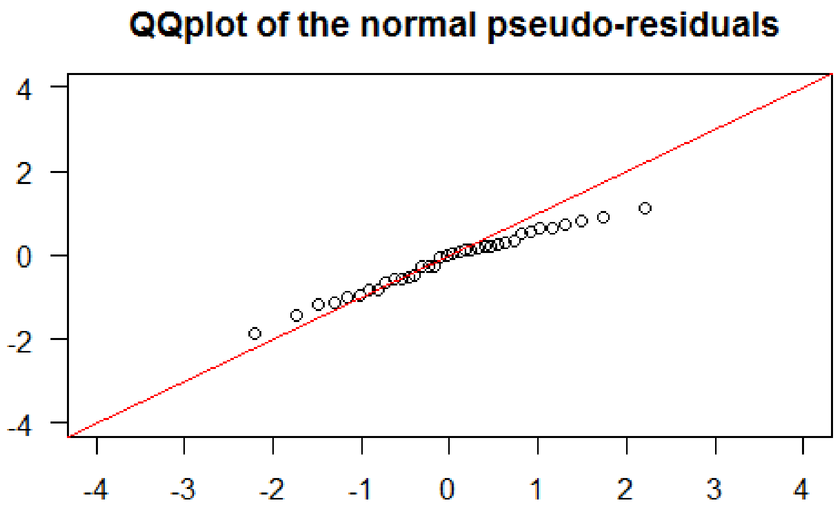

Figure 4.15: Claim numbers: QQplot of the ordinary normal pseudo-residuals.

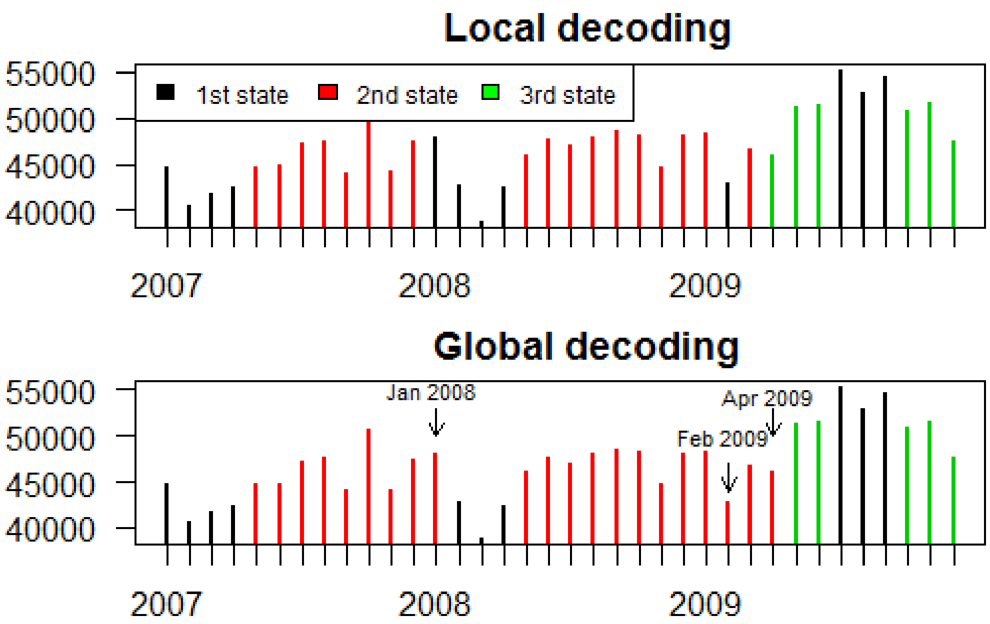

Figure 4.16: Claim amounts: local vs. global decoding according to three-state Poisson-Normal HMM.

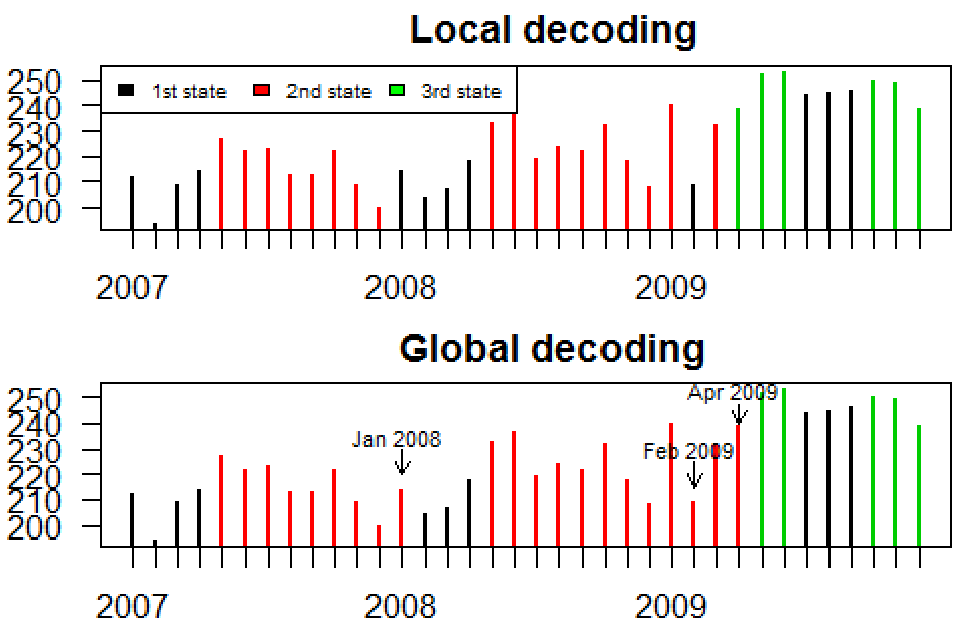

Figure 4.17: Claim numbers: local vs. global decoding according to three-state Poisson-Normal HMM.

are given in Section 3.7. In order to derive the most likely sequence of states, the Viterbi algorithm is applied. Observing Figure 4.16 and 4.17 we note that decoding results are very similar but differ in January 2008, February 2009, and April 2009.

Additionally, we obtained state probabilities for three years ahead, that can be used for analysis of further claim behavior. We suppose, that in the beginning of 2010 observations will be dependent on the third state, in the following few months those will continue with the first state, and during two years the first and second states have almost equal probabilities and are dominant compared to the third state, see Figure 4.21.

Four of the forecast distributions for claim amounts and claim numbers are displayed in Figures 4.18 and 4.19. The distributions are compared with the limiting distributions, i.e. the marginal distributions of the Poisson-Normal HMM. It is clear that the forecast distributions approach the limiting distribution relatively fast.

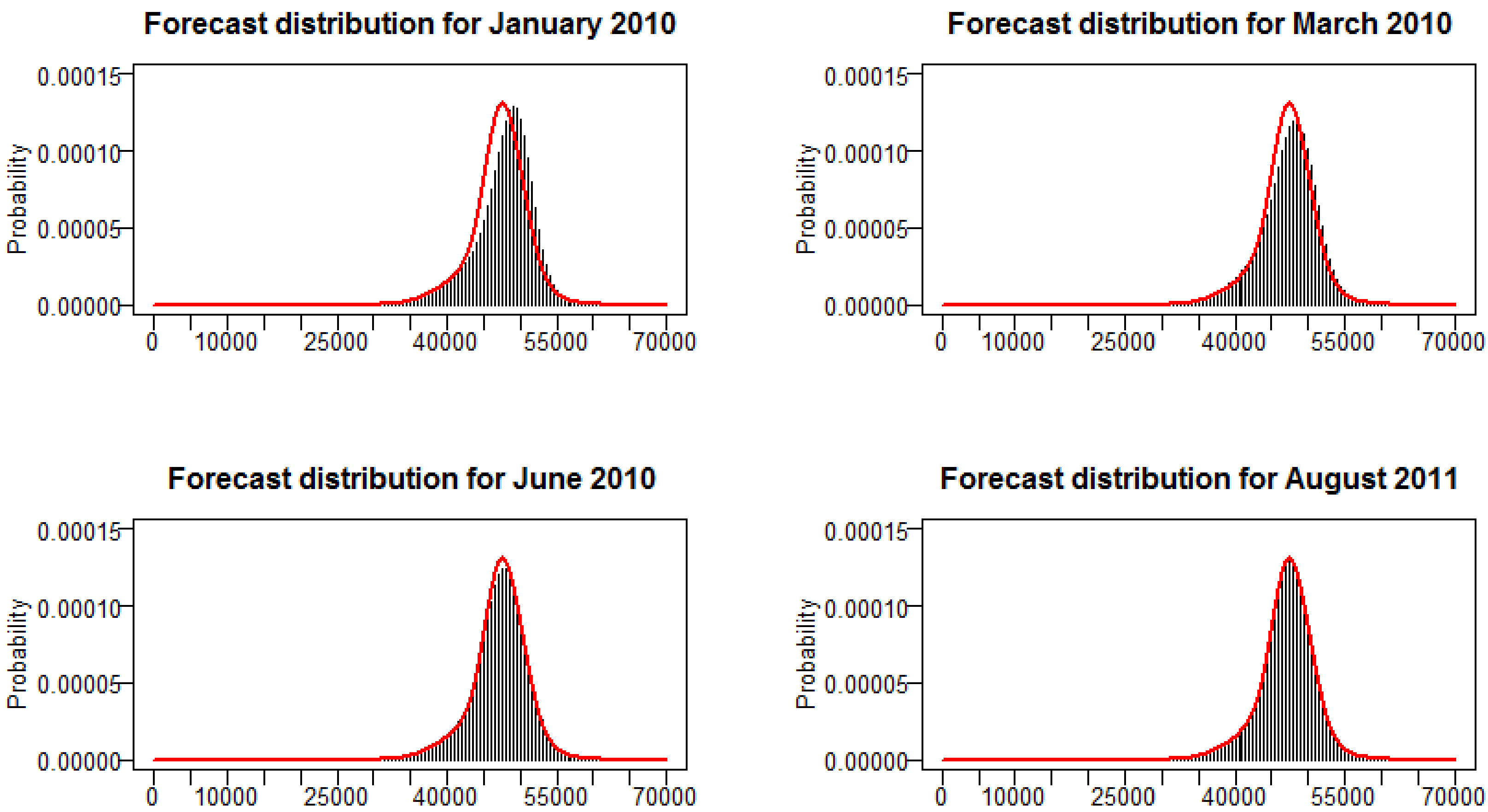

Figure 4.18: Claim amounts: forecast distributions for 1 to 20 months ahead. Red line shows a limiting distribution.

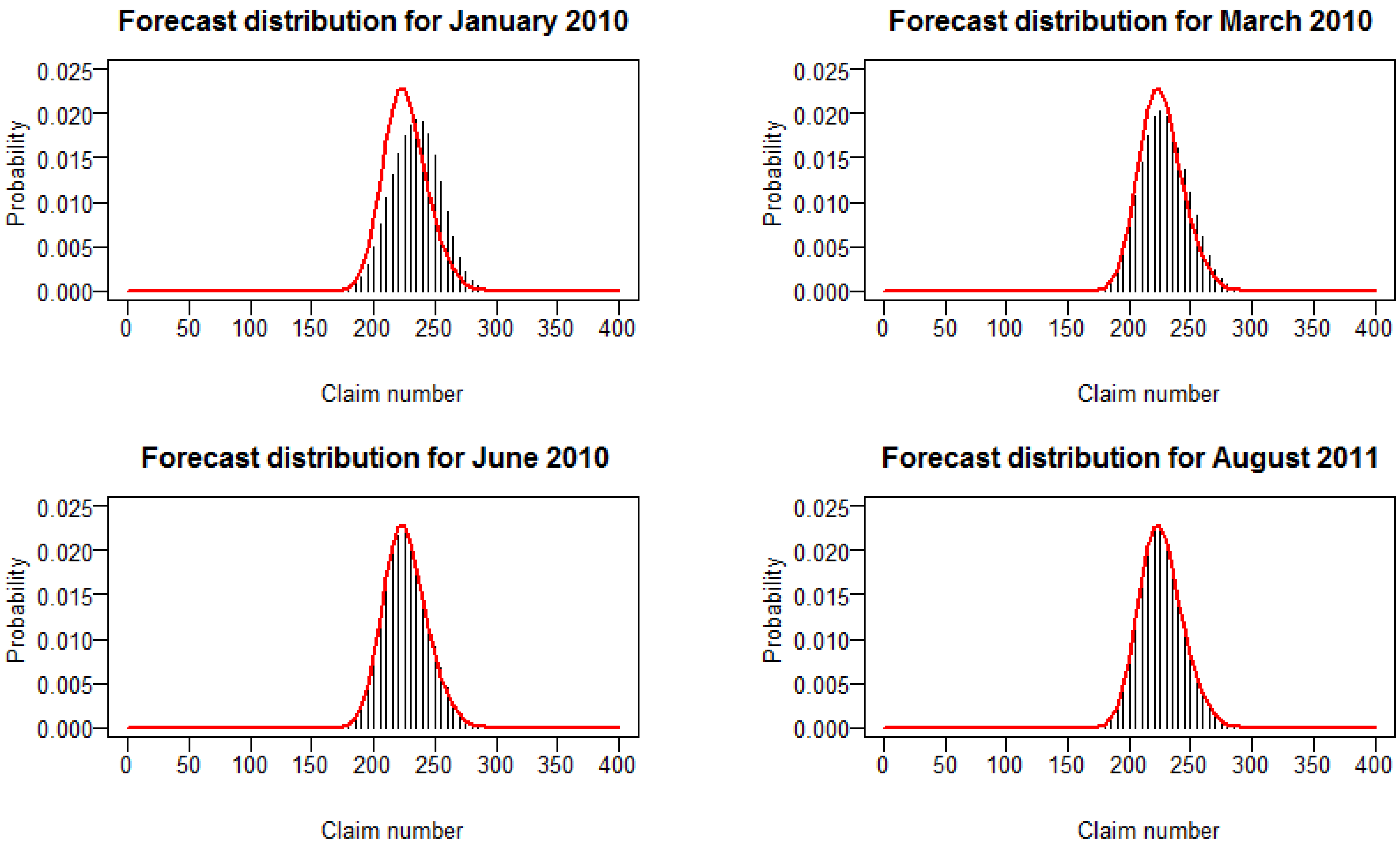

Figure 4.19: Claim numbers: forecast distributions for 1 to 20 months ahead. Red line shows a limiting distribution.

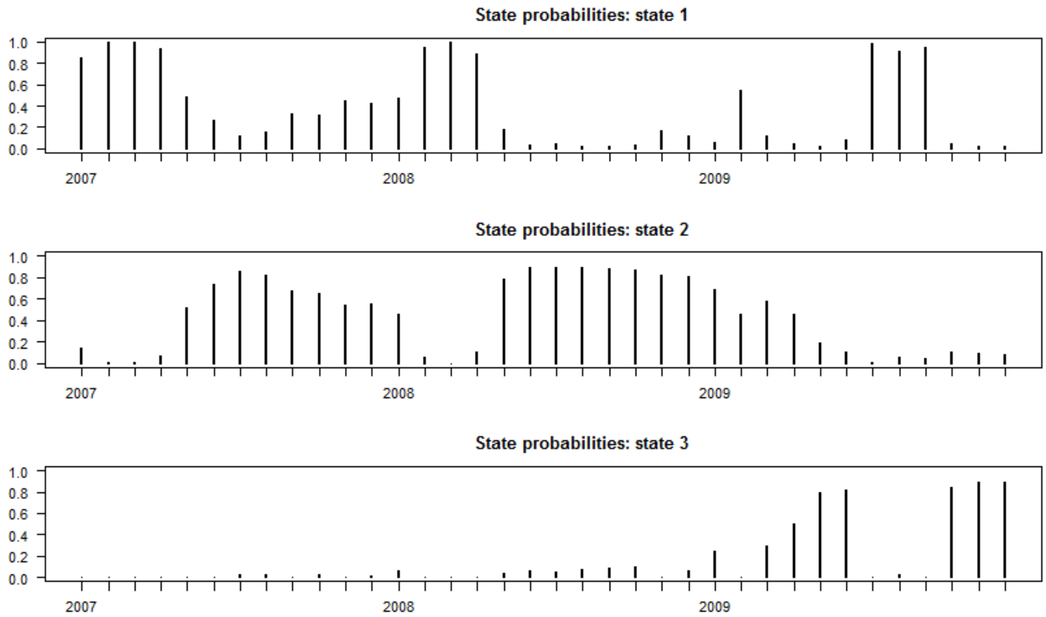

Figure 4.20: State probabilities for fitted three-state Poisson-Normal HMM.

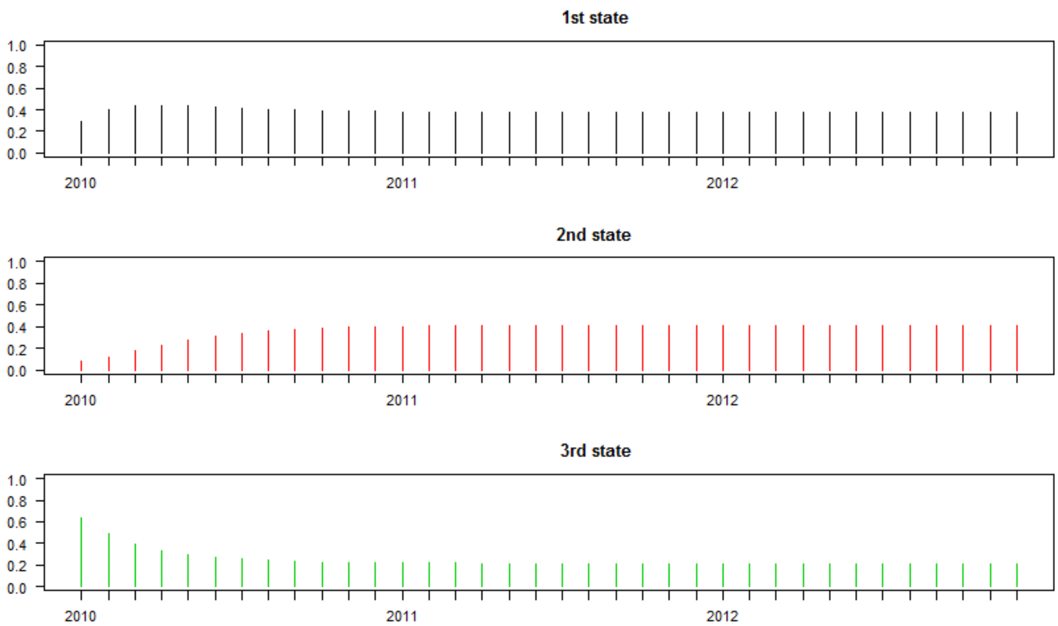

Figure 4.21: State prediction for fitted three-state Poisson-Normal HMM.

CHAPTER 5

CONCLUSION

We propose Bivariate Hidden Markov Model as a novel approach in modeling claim dependence. The model allows claim numbers and aggregate claim amounts to be mutually and serially dependent through an underlying hidden state. We modify the classical HMM definitions and propositions to bivariate case. Three different BHHMs are presented, namely Poisson-Normal HMM, Poisson-Gamma HMM and Negative Binomial-Gamma HMM. For parameter estimation of the model, we conducted EM algorithm. To perform the algorithm, we acknowledge and proved three propositions, which maximize the state-dependent part of complete-data log-likelihood of proposed models.

To examine the performance of our model, we apply the Poisson-Normal HMM with the different number of states to the vehicle insurance data for Istanbul taken from Turkish Motor Insurance Center (TRAMER) for the years 2007-2009.

Three-state Poisson-Normal HMM is selected as the most suitable model by comparing Akaike and Bayesian information criterions. In order to determine whether indeed the model performs adequate, we obtain and assess ordinary normal pseudo-residuals. Shapiro-Wilk normality test and quantile-quantile plots confirm the goodness of fit of the model. According to the selected model we conduct Viterbi algorithm to derive the most likely sequence of states which underly and affect the observations. Using these results the specialists may specify names of the states. Additionally, we derive forecast distributions both for claim numbers and claim amounts and performed state prediction.

The main advantage of the model is a flexibility in a sense of accommodating different types of data. In our study, we modeled a bivariate series with one discrete and one continuous variable. Moreover, proposed model is applicable in various fields of life and non-life insurance, where the serial dependence and mutual dependence among observations exist. Remarkably, that information provided by the model, such as the most likely sequence of hidden states, can be used for further analysis by the experts, like doctors, biologists or actuaries. It allows determining the character of events or factors influencing the observations.

In future work, we plan to apply the Poisson-Gamma HMM and the Negative Binomial-Gamma HMM to the motor insurance data. Additionally, some exploratory variables can be added to the model.

REFERENCES

[1] Giulio Rocca, Insurance Objectives, studio D, http://work.chron.com/insurance-objectives-23793.html, last accessed date 12.07.2016.

[2] N.L. Bowers, H.U. Gerber, J.C. Hickman, D.A. Jones, C.J. Nesbitt, *Actuarial Mathematics*, Society of Actuaries, Itacsa, IL, 1997.

[3] Yiu-Kuen Tse, Nonlife Actuarial Models Theory,Methods and Evaluatio, Cambridge University Press, Cambridge CB2 8RU, UK, 2009.

[4] S. Wang, J. Dhaene, Comonotonicity, correlation order and premium principles, Insurance: Mathematics and Economics,22,pp. 235-242, 1998.

[5] N. Bäuerle, A. Müller, Modeling and comparing dependencies in multivariate risk portfolios, Astin Bulletin, 28, pp. 59–76, 1998.

[6] Michel Denuita, Claude Lefèvreb, Sergey Utevc, Measuring the impact of dependence between claims occurrences, Insurance: Mathematics and Economics,30,1, pp. 1–19, February 2002.

[7] J. Dhaene, M.J. Goovaerts, On the dependency of risks in the individual life model, Insurance Mathematics and Economics, 19, pp. 243–253, 1997.

[8] M.J. Goovaerts, J. Dhaene, The compound Poisson approximation for a portfolio of dependent risks, Insurance: Mathematics and Economics, 18, pp. 81–85, 1996.

[9] J. Garridoa, C. Genestb, J. Schulzb, Generalized linear models for dependent frequency and severity of insurance claims, Insurance: Mathematics and Economics, June 2016.

[10] E. Frees, E. Valdez, Understanding relationships using copulas, North American Actuarial Journal, 2, pp. 1–25,1998.

[11] Benjamin Avanzi, Luke C. Cassar, Bernard Wong, Modeling Dependence in Insurance Claims Processes with Lévy Copulas, Australian School of Business Research Paper, No. 2011ACTL01, 2011 .

[12] S. Klugman, R. Parsa, Fitting bivariate loss distributions with copulas, Insurance: Mathematics and Economics, 24, pp. 139–148, 1999.

[13] Nicole Kramer, Eike C. Brechmann, Daniel Silvestrini, Claudia Czado, Total loss estimation using copula-based regression models, Insurance: Mathematics and Economics,53,pp. 829–839, 2013.

[14] G. Wanga, K. C. Yuen, On a correlated aggregate claims model with thinning-dependence structure, Insurance: Mathematics and Economics,36, 3, pp. 456–468, June 2005.

[15] K. C. Yuen, J. Guob, X. Wua, On a correlated aggregate claims model with Poisson and Erlang risk, Insurance: Mathematics and Economics, 31, 2, pp. 205–214, October 2002.

[16] R. S. Ambagaspitiya, On the distributions of two classes of correlated aggregate claims, Insurance: Mathematics and Economics, 24(3), pp. 301–308, May 1999.

[17] J. Pinquet, Experience rating through heterogeneous models. In: Handbook of Insurance, Kluwer Academic Publishers, Boston, pp. 459–500, 2000.

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

[19] S.A.Klugman, H.H.Panjer,G.E.Willmot, *Loss Models from Data to Decisions*, 4th Edition, John Wiley&Sons,Inc.,Hoboken,New Jersey, 2012.

[20] I.L.MacDonald,W.Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series*, 1st Edition, Chapman&HAll, 2-6 Boundary Row, London SE1 8HN, UK, 1997.

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

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

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

[24] Andrei L. Badescu, X. Sheldon Lin, , Dameng Tang, A marked Cox model for the number of IBNR claims: Theory, Insurance: Mathematics and Economics, 69, pp. 29–37, July 2016.

[25] G.H. Givens, J.A. Hoeting, *Computational Statistics*, John Wiley&Sons,Inc.,Hoboken,New Jersey, 2013.

[26] R.Kaas,M.Goovaerts,J.Dhaene,M.Denuit, *Modern Actuarial Risk Theory*, Springer Heidelberg Dordrecht London New York, 2nd ed. 2008.

[27] Paroli. R.,Redaelli.G.,Spezia.L. Poisson hidden Markov models for time series of overdispersed insurance counts.

[28] Hanna Arvidsson and Sofie Francke, *Dependence in non-life insurance*, Uppsala University, 2007 .

[29] Walter Zucchini and Iain L.MacDonald *Hidden Markov Models for Time Series*, 2009.

[30] D.Lamberton, B.Lapeyre, *Introduction to Stochastic Calculus Applied to Finance*, Chapman&HAll/CRC, Boca Raton, Florida, 2000.

[31] Inflation rates, http://www.bigpara.com/haberler/enflasyon-verileri/, last accessed date 13.07.2016.

[32] Flooding in Istanbul, 8-9 Eylül 2009 İstanbul sel felaketı üzerine kamuoyu açıklaması, Sukru Ersoy, https://www.researchgate.net/publication/, last accessed date 13.07.2016

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

[34] R.J.A. Little and D.B.Rubin, *Statistical Analysis with Missing Data*, second edition. Wiley, New York.

[35] L.R.Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proc. IEEE, 77(2), pp. 257–286, 1989.

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

[37] G.D.Forney, The Viterbi algorithm, Proc. IEEE 61, 268–278, 1973.

[38] H.Akaike, Information theory and an extension of the maximum likelihood principle, in Petrov, B.N.; Csáki, F., 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, Budapest: Akadémiai Kiadó, pp. 267-281, September,1971.

[39] K. P.Burnham, D. R.Anderson, Multimodel inference: understanding AIC and BIC in Model Selection, Sociological Methods Research 33,pp. 261–304, 2004.

[40] W.Zucchini and P.Guttorp, A hidden Markov model for space-time precipitation, Water Resour. Res. 27, 1917–1923, 1991