AGGREGATE CLAIM ESTIMATION USING BIVARIATE HIDDEN MARKOV MODEL

Article *in* Astin Bulletin · November 2018

DOI: 10.1017/asb.2018.29

Zarina Oflaz
KTO Karatay University

Ceylan Yozgatligil
Middle East Technical University

Sevtap Kestel
Middle East Technical University

AGGREGATE CLAIM ESTIMATION USING BIVARIATE HIDDEN MARKOV MODEL

BY

ZARINA NUKESHTAYEVA OFLAZ, CEYLAN YOZGATLIGIL AND
A. SEVTAP SELCUK-KESTEL

ABSTRACT

In this paper, we propose an approach for modeling claim dependence, with the assumption that the claim numbers and the aggregate claim amounts are mutually and serially dependent through an underlying hidden state and can be characterized by a hidden finite state Markov chain using bivariate Hidden Markov Model (BHMM). We construct three different BHMMs, namely Poisson–Normal HMM, Poisson–Gamma HMM, and Negative Binomial–Gamma HMM, stemming from the most commonly used distributions in insurance studies. Expectation Maximization algorithm is implemented and for the maximization of the state-dependent part of log-likelihood of BHMMs, the estimates are derived analytically. To illustrate the proposed model, motor third-party liability claims in Istanbul, Turkey, are employed in the frame of Poisson–Normal HMM under a different number of states. In addition, we derive the forecast distribution, calculate state predictions, and determine the most likely sequence of states. The results indicate that the dependence under indirect factors can be captured in terms of different states, namely low, medium, and high states.

KEYWORDS

Claim estimation, bivariate Hidden Markov model, EM algorithm, Viterbi algorithm, MTPL

1. INTRODUCTION

The fundamental objectives for insurance companies include safeguard policyholders against potential losses by apportioning the risk with others and compensate the loss (Rocca, 2016). In order to be solvent over a certain time horizon, an insurer must adequately price the premiums to be charged

Astin Bulletin, 1-27. doi:10.1017/asb.2018.29

and have sufficient amount of capital and reserves. Hence, predicting the distribution of the total claim amounts in a given time period is important as it is directly related to the equity and reserving requirements for an insurance company (Bowers *et al.*, 1997). The classical approach in modeling aggregate claim amounts 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 (Tse, 2009). Most actuarial models rely on an assumption that both claim counts and aggregate claim amounts are independent which yields nice outcomes analytically. However, some conditions such as climate, economical, and financial factors affect the claim-causing events, resulting in the interaction between the claim number and the total claim amount distributions. Despite its simplicity and accessibility, independence assumption is too restrictive under different frameworks.

The impact of dependencies between claims has gained increasing attention in recent years (Äuerle and Müller, 1998; Denuit *et al.*, 2002). For example, Dhaene and Goovaerts (1997) state that some types of dependence between individuals may produce the riskiest aggregate claims and cause the largest stop-loss premiums. Sufficiently many studies have developed and applied models with claim dependency assumption. In particular, models with dependence among the reserving claim amounts (Hudecová and Pešta, 2013; Pešta and Okhrin, 2014), serial dependence in claim counts (Avanzi *et al.*, 2016), and dependence between claim frequency and severity (Boudreault *et al.*, 2014; Garrido *et al.*, 2016) have been explored. Furthermore, models assuming the dependence among aggregate claims have also been widely studied (Ambagaspitiya, 1999; Yuen *et al.*, 2002; Wang and Yuen, 2005).

The main work of our study involves a novel approach for modeling claim dependence, introducing bivariate Hidden Markov Model (BHMM). In order to relax the assumption on 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 leading to HMM due to its nature. HMMs have been applied in various fields, namely speech recognition (Rabiner and Juang, 1993), molecular biology (Krogh *et al.*, 1994), analysis of DNA sequence (Cheung, 2004), and stock market forecasting (Hassan and Nath, 2005). The main reason for selecting an HMM for modeling the claim dependence is that unobservable background factors triggering the claim-causing events can be characterized and captured by a hidden parameter process. That seems both total claim amounts and claim numbers may react similarly to some exogenous conditions, consequently, resulting in dependence. This approach is also introduced and modeled where unobservable information is described by exogenous variables, using fixed and random effects models (Pinquet, 2000). Similarly, claim numbers and claim amounts dependency is designed using approach based on Markov-switching model (Ren, 2012). In claim modeling, HMM is considered to be a relatively new tool. For instance, Poisson HMM is used to model the dynamics of claim counts in nonlife insurance (Paroli *et al.*, 2000), while Badescu *et al.* (2016) generate the intensity function of the claim arrival process by a HMM with Erlang state-dependent distributions.

In this paper, we assume two conditional independence, namely contemporaneous and longitudinal. We assume that the claim counts and the aggregate claim amounts are dependent and both are serially dependent via an underlying hidden state. Multivariate HMM with these two conditional assumptions is constructed. Also, the model is theoretically adopted to the bivariate series, where one component is continuous and other is taken as discrete (Zucchini and Guttorp, 1991; Zucchini and MacDonald, 2009). We propose and construct three different BHMMs, under Poisson–Normal, Poisson–Gamma, and Negative Binomial–Gamma distributions whose parameters are estimated using EM algorithm. The conditions required to implement the EM are proposed and analytically proved. The numerical illustrations are employed to Turkish motor liability third party (MTPL) data for Istanbul province collected monthly between years 2007 and 2009.

The organization of the paper is as follows: Section 2 includes a theoretical framework for the claim modeling and HMM. Furthermore, we present BHMM, and related definitions and theorems. Parameter estimation method is introduced in Section 3. Next section includes decoding techniques. State prediction and model selection are in Section 5. In Section 6, we establish three different BHMMs. We apply Poisson–Normal BHMM to the vehicle insurance claims data in Section 7. The conclusion summarizes the findings with the remarks.

2. Bivariate Hidden Markov Model

A HMM assumes that process generating N_t depends on the *hidden* state C_t which satisfies the Markov property. Therefore, an HMM can be determined by hidden "parameter process" $\{C_t : t = 1, 2, \ldots\}$ and the "state-dependent process" $\{N_t : t = 1, 2, \ldots\}$ satisfying (Zucchini and MacDonald, 2009)

$$\begin{aligned} P(C_{t+1}|C_t, \ldots C_1) &= P(C_{t+1}|C_t), \quad t = 2, 3, \ldots \\ P(N_t|N^{(t-1)}, C^{(t)}) &= P(N_t|C_t), \quad t \in N. \end{aligned} \tag{2.1}$$

Transition probability, $\gamma_{ij}(t)$, can be expressed as the probability of moving from state i to state j at time t:

$$\gamma_{ij}(t) = P(C_{k+t} = j | C_k = i).$$

If these probabilities do not depend on k, Markov chain is said to be homogeneous. Finite state-space homogeneous Markov chains fulfill the Chapman–Kolmogorov equations (Zucchini and MacDonald, 2009). Probabilities of a

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

$$u(t) = (P(C_t = 1), \ldots, P(C_t = m)) \, . \tag{2.2}$$

Here, $u(1)$ is considered as initial distribution of the Markov chain which specifies the starting state. The initial distribution, $u(1)$, and matrix of transition probabilities, $\gamma_{ij}(t)$, are necessary to construct a probability distribution over sequence of observations. Additionally, the state-dependent distribution, $p_i(n)$; $i = 1, 2, \ldots, m$, where m is the number of states, that defines the relation between observation and an unobserved state is

$$p_i(n) = P(N_t = n | C_t = i), i = 1, \ldots, m.$$

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. In case when an initial distribution, $u(1)$, is not supplied, Zucchini and MacDonald suggest to use an initial distribution of a stationary Markov chain, δ, which can be found from Zucchini and MacDonald (2009):

$$\delta(I_m - \Gamma + U) = 1, \tag{2.3}$$

where I_m is the $m \times m$ identity matrix, U is the $m \times m$ matrix of 1, and the row vector of 1. Therefore, we use the calculated δ as the starting value for an initial distribution of BHMM.

Let $\{N_t: t = 1, 2, \ldots\}$ denote the number of claims and $\{S_t: t = 1, 2, \ldots\}$ be the aggregate claim amount reported by policyholders during $t = 1, 2, \ldots$ which is expressed as

$$S_t = \sum_{i=1}^{N_t} X_i, \tag{2.4}$$

where X_i denotes the ith claim amount.

Using BHMM in which unobservable background factor can be characterized by hidden finite-state Markov chain under insurance claims setup, we assume that N_t and S_t are mutually and serially dependent through an underlying hidden state $\{C_t: t = 1, 2, \ldots\}$. We consider that the Markov chain of the bivariate model is homogeneous and nonstationary. It is obvious that the claim numbers N_t and S_t are reported at the same time, t; therefore, we collect the information given by bivariate observations (S_t, N_k), $t = k$, as illustrated in Figure 1.

Besides longitudinal conditional independence, that is, conditional on the underlying hidden state $\{C_t: t = 1, 2, \ldots\}$ which refers to the claim counts at time t and the aggregate amounts at time t is independent, we consider also contemporaneous conditional independence as shown in Figure 2. These two conditional independence assumptions neither imply the serial independence of N_t and S_t nor the mutually independence of component series, concluding that N_t and S_t are dependent, as stated in Zucchini and MacDonald (2009).

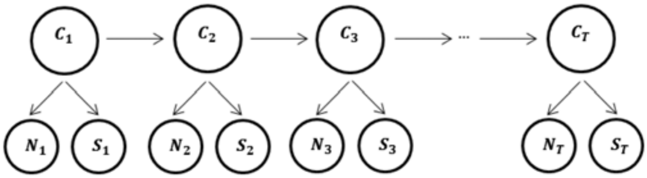

FIGURE 1: Directed graph of BHMM (longitudinal).

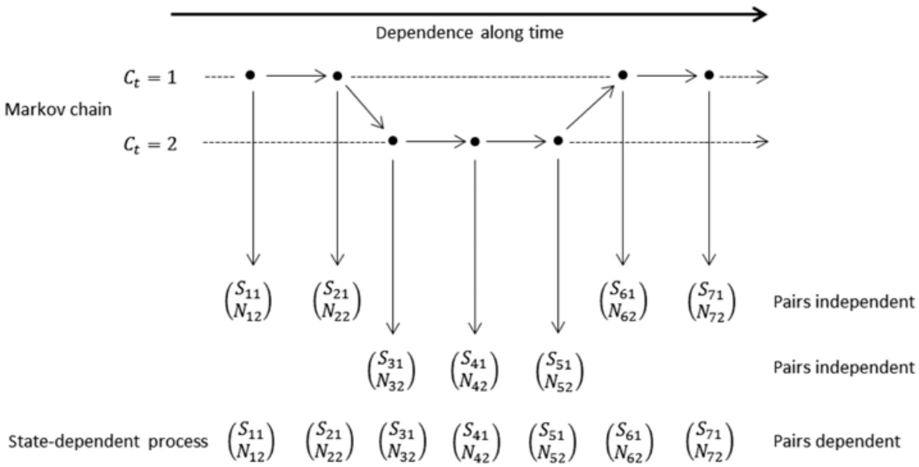

FIGURE 2: Contemporaneous conditional independence in BHMM with two states Zucchini and MacDonald (2009).

To specify the bivariate model, it is necessary to postulate a joint state-dependent distribution of the pair (s, n) for $t = 1, 2, \ldots, T, i = 1, 2, \ldots, m$, which is defined as

$$p_i(s_t, n_t) = P((S_t, N_t) = (s_t, n_t)|C_t = i). \tag{2.5}$$

According to the contemporaneous conditional independence, the state-dependent probabilities are given by a product of the corresponding marginal probabilities (Zucchini and Guttorp, 1991; Zucchini and MacDonald, 2009) as follows:

$$\begin{aligned} p_i(s_t, n_t) &= P\big((S_t, N_t) = (s_t, n_t)\big| C_t = i\big) \\ &= P(S_t = s_t|C_t = i)P(N_t = n_t|C_t = i). \end{aligned} \tag{2.6}$$

3. Parameter estimation in BHMM

To construct BHMM, we need to estimate transition probabilities, initial probability, and parameters of the joint state-dependent probabilities. Regarding the complexity in likelihood function, EM (namely Baum–Welch) algorithm is employed. EM which performs maximum likelihood estimation of parameters having missing value in the data functions well in HMM estimation as the hidden states are treated as missing information (Dempster *et al.*, 1977; Little and Rubin, 2014). In addition, the algorithm enables estimation of the parameters of an HMM whose Markov chain is homogeneous but not necessarily stationary (Zucchini and MacDonald, 2009). The *EM* algorithm alternates between two phases. In the E-step, conditional expectations of the hidden states that are treated as missing data, given the observed data and a current estimate of the model parameters, are computed. In the M-step, the complete-data log-likelihood (CDLL) function is maximized under the assumption that the missing data are known. Iterations are repeated until a convergence is satisfied (Little and Rubin, 2014). In order to maximize the state-dependent part of CDLL of BHMMs, we establish and prove three theorems (Theorems 1, 2, and 5.6).

The forward (α_t) and the backward (β_t') probabilities are needed for the maximization part of EM estimation (Rabiner, 1990). For $t = 1, 2, \ldots, T$,

$$\alpha_t = \delta P(s_1, n_1)\Gamma P(s_2, n_2), \ldots, \Gamma P(s_t, n_t) = \delta P(s_1, n_1) \prod_{k=2}^{t} \Gamma P(s_k, n_k) \tag{3.1}$$

and

$$\beta_t' = \Gamma P(s_{t+1}, n_{t+1})\Gamma P(s_{t+2}, n_{t+2}), \ldots, \Gamma P(s_T, n_T)1' = \left(\prod_{k=t+1}^{T} \Gamma P(s_k, n_k) \right) 1' \tag{3.2}$$

with $\beta_T = 1$, define the forward and backward probabilities, respectively.

The CDLL of BHMM, that is, the log-likelihood of observed variables and hidden states, is defined as (Zucchini and MacDonald, 2009)

$$\begin{aligned} \log\left(P(s^{(T)}, n^{(T)}, c^{(T)})\right) &= \log \delta_{c_1} + \sum_{t=2}^{T} \log \delta_{c_{t-1},c_t} + \sum_{t=1}^{T} \log p_{c_t}(s_t, n_t) \\ &= \sum_{j=1}^{m} u_j(1) \log \delta_j + \sum_{j=1}^{m}\sum_{k=1}^{m} \left(\sum_{t=2}^{T} v_{jk}(t) \right) \log \gamma_{jk} \\ &\quad + \sum_{j=1}^{m}\sum_{t=1}^{T} u_j(t) \log p_j(s_t, n_t), \end{aligned} \tag{3.3}$$

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

In the **E** part of EM algorithm, $v_{jk}(t)$ and $u_j(t)$ are replaced by the conditional expectations of being in a state j at time t given the observations $s^{(T)}, n^{(T)}$ (Zucchini and MacDonald, 2009):

$$\hat{u}_j(t) = P\big(C_t = j \big| \big(S^{(T)}, N^{(T)}\big) = \big(s^{(T)}, n^{(T)}\big)\big) = \frac{\alpha_t(j)\beta_t(j)}{L_T} \tag{3.4}$$

and

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

respectively. The likelihood $L_T = u(1)P(s_1, n_1)\Gamma P(s_2, n_2)\Gamma P(s_3, n_3), \ldots, \Gamma P(s_T, n_T)1'$. On the other side, the **M** part maximizes the each term of CDLL with respect to the related set of parameters, that is, the initial distribution $u(1)$, the transition probability matrix Γ, and the parameters of joint state-dependent distributions. The CDLL of BHMM is found to require three separate maximizations which are stated, as in Zucchini and MacDonald (2009):

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 $\gamma_{jk} = \sum_{t=2}^T v_{jk}(t)/\sum_{k=1}^m \big(\sum_{t=2}^T v_{jk}(t)\big)$,

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

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

4. Model selection

According to the likelihood of BHMM, 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 selection criteria such as Akaike (AIC) and Bayesian (BIC) Information Criteria and the use of pseudo-residuals are taken into account to determine the best fitting model.

Despite the fact that the model opted by AIC or BIC criterion is supposed to select the most optimal model, we also wish to assess the goodness of fit of the model in an absolute sense. An optimal way to do so is to obtain pseudo-residuals, which are also able to identify outliers relative to the model. In this aspect, we consider **ordinary pseudo-residuals** which are based on the conditional distribution (Zucchini and MacDonald, 2009). The normal pseudo-residual is defined as

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

Pseudo-residuals are distributed as standard normal if the related model is correct.

5. BHMM MODELS

In this study, we construct three different bivariate models. The details of these models can be found in Oflaz (2016). These are Poisson–Normal HMM, Poisson–Gamma HMM, and Negative Binomial–Gamma HMM. We also introduce the theorems related to our models in order to maximize the third term of CDLL of BHMM.

5.1. Poisson–normal HMM

We consider that N_t follows Poisson and S_t the normal distributions with underlying unobservable stochastic process, C_t. A joint state-dependent probability distribution for $n \in N, s \in (-\infty, +\infty), \lambda > 0, \mu, \sigma^2$ is expressed as follows:

$$p_i(s_t, n_t) = (2\pi\sigma_i^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_i^2}(s_t-\mu_i)^2-\lambda_i} \frac{\lambda_i^{n_t}}{n_t!}. \tag{5.1}$$

We derive the parameters of Poisson–Normal BHMM based on steps explained earlier.

Theorem 1. *Given two random variables, S and N having normal* (μ_j, σ_j^2) *and Poisson*(λ_j) *distributions, respectively, the EM estimates of joint state-dependent distribution are* (*Oflaz,* 2016)

$$\begin{aligned}
\hat{\lambda}_j &= \frac{\sum_{t=1}^{T} \hat{u}_j(t) n_t}{\sum_{t=1}^{T} \hat{u}_j(t)}, \\
\hat{\mu}_j &= \frac{\sum_{t=1}^{T} \hat{u}_j(t) s_t}{\sum_{t=1}^{T} \hat{u}_j(t)}, \\
\hat{\sigma}_j^2 &= \frac{\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \hat{\mu}_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)}.
\end{aligned} \tag{5.2}$$

Here, $\hat{u}_j(t)$ *is given in Equation* (3.4)*. These estimates are consistent with the ones given by Zucchini and MacDonald* (2009)*.*

The Proof of Theorem 1 is summarized in the Appendix.

5.2. Poisson–gamma HMM

Similarly, having the marginal distributions for N_t Poisson and for S_t Gamma, the joint state-dependent distribution, for $n \in N, s > 0, \lambda > 0, \alpha > 0, \beta > 0$, is given by

$$p_i(s_t, n_t) = \frac{\beta_i^{\alpha_i} s_t^{\alpha_i - 1} \lambda_i^{n_t} e^{-\beta_i s_t - \lambda_i}}{\Gamma(\alpha_i) n_t!}. \tag{5.3}$$

We derive the parameter estimates of BHMM with Poisson–Gamma assumptions.

Theorem 2. *Given two random variables, S and N having Gamma(α_j, β_j) and Poisson(λ_j) distributions, respectively, the EM estimates of joint state-dependent distribution are found as* (*Oflaz,* 2016)

$$\hat{\lambda}_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t) n_t}{\sum_{t=1}^{T} \hat{u}_j(t)},$$
$$\hat{\beta}_j = \frac{\hat{\alpha}_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} s_t}. \tag{5.4}$$

$\hat{u}_j(t)$ *is given in Equation* (3.4)*. To estimate* $\hat{\alpha}_j$ *numerical maximization is required (Equation* (A.2) *in the Appendix).*

The proof of the theorem is given in the Appendix.

5.3. Negative Binomial–Gamma HMM

The state-dependent distribution is of the form

$$p_i(s_t, n_t) = \frac{\binom{n_i - 1}{r_i - 1} \beta_i^{\alpha_i} s_t^{\alpha_i - 1} e^{-\beta_i s_t} p_i^{r_i} (1 - p_i)^{n_t - r_i}}{\Gamma(\alpha_i)} \tag{5.5}$$

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

Theorem 3. *Given two random variables, S and N having Gamma (α_j, β_j) and Negative Binomial (r_j, p_j) distributions, respectively, the EM estimates of joint state-dependent distribution are derived as* (*Oflaz,* 2016)

$$\hat{\beta}_j = \frac{\hat{\alpha}_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} s_t},$$
$$\hat{p}_j = \frac{\hat{r}_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - \hat{r}_j) + \hat{r}_j \sum_{t=1}^{T} \hat{u}_j(t)}. \tag{5.6}$$

The parameter estimates of $\hat{\alpha}_j$ *and* $\hat{r}_j$ *require numerical maximization, whose derivations are presented in the Appendix (Equations* (A.4) *and* (A.7)*, respectively).*

6. Case study: Motor third-party liability claims

The implementation of BHMM model under the dependence assumption between claim numbers and the total claim amounts is performed to the MTPL data collected from Turkish insurance sector. According to the Turkish Insurance Market Outlook 2016–2017, having greater share in total premium (25.5%), MTPL is one of the most accounted insurance business line in Turkey (JLT Sigorta ve Reasurans Brokerligi A.S., 2017). The growth of MTPL business is about 35% from 2014 (JLT Sigorta ve Reasurans Brokerligi A.S., 2017) resulting also (30%) increase in the loss ratio in 5 years from 2010 (Garanti Securities, 2016). The recent regulations and reforms require insurance companies to allocate sufficient reserves for the claims which appear highly from MTPL losses. For this reason, insurance companies are in seek of methods yielding good estimates on their future losses.

The insurance companies are required by regulation to report the policy-based information every year to the Insurance Information Center (Turkish synonym SBM) which does have separate sections for each line of businesses. Traffic Insurances Information and Monitoring Center (TRAMER) is the section which collects and processes the MTPL and Casco details. The data set used in this paper is provided by TRAMER which contains the vehicle insurance recordings from Istanbul in Turkey.

6.1. Data processing

The data set includes information on the policies written in Istanbul where the policies starting years vary between the years 2006 and 2009 having the accident years varying from 2006 to 2011. Every policy registered to the system includes coded policy number, starting and end dates of policy, vehicle tariff group code (car, minibus, taxi, etc.), registered city code, vehicle ID number, vehicle age, usage type (private, commercial), passenger capacity, nationality of insured, damage date, claim reason, and individual claim amount. The data set were not available and accessible after 2011. As Istanbul has the highest rate of insurance penetration position compared to other cities in Turkey, we expect it to capture the model better.

In data processing we go through each claim records and tabulate the resulting outputs to detect the appropriate set of data over years. Table 1 provides the total number of claims and aggregate claim amount for each year. It can be noticed that not all years enable us sufficient data to be processed, the ones which are taken into account in this study are marked as bold.

As a result of data processing we utilize the monthly information on automobile insurance portfolios from Istanbul over the period January 2007–December 2009, which has only nonzero claims. For convenience, we refer to the *monthly total number of claims* and the *monthly aggregate claim amounts* as the *claim numbers* and the *total claim amounts*, respectively. The individual

TABLE 1

MONTHLY TOTAL CLAIM NUMBERS AND AGGREGATE CLAIM AMOUNTS (TL) BETWEEN 2006 AND 2011.

		Accident year					
	Policy start	2006	2007	2008	2009	2010	2011
Claim no.	2006	196,046	**191,902**	314	128	12	NA
	2007	NA	**187,360**	**199,467**	239	31	NA
	2008	NA	NA	**198,368**	**210,612**	276	12
	2009	NA	NA	NA	**204,166**	196,746	59
Aggregate claim amount (TL)	2006	290,861,223	**257,040,552**	444,279	132,221	9098	NA
	2007	NA	**290,537,933**	**291,431,736**	505,044	51,846	NA
	2008	NA	NA	**320,698,956**	**337,879,194**	543,344	21,553
	2009	NA	NA	NA	**502,834,660**	336,400,909	207,676

TABLE 2

DESCRIPTIVE STATISTICS OF MONTHLY AGGREGATE CLAIM AMOUNTS (TL) AND NUMBERS.

	Mean	Median	Minimum	Maximum	St. dev	Mode
Total claim amount	46,957,468	47,367,916	38,844,907	55,226,741	3,811,343	46,957,470
Claim no.	22,481.31	22,173	19,429	25,316	1612.77	21,550

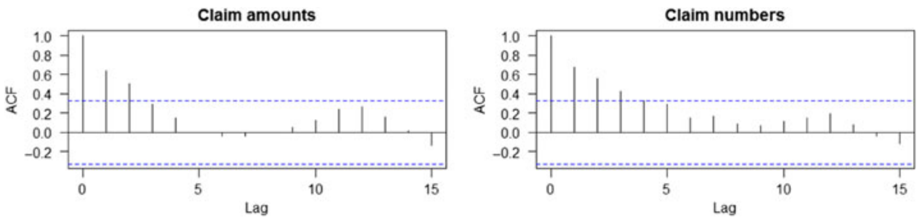

FIGURE 3: ACF plots of total claim amounts and numbers.

claim amounts are inflation adjusted with respect to the rates for the years 2008 (9.67%) and 2009 (6.21%) (Inflation rates, 2016). The claim amounts less than 250 TL (1USD = 1.537 TL in January 2011) are taken as deductible as the contribution of these observations is found to be insignificant on the whole set.

The summary statistics, the scatter plots, and furthermore, autocorrelations of the total claim amounts and numbers are analyzed as the first step in the data description. Table 2 exposes that the mean, mode, and median values of the total claim amounts are almost equal that is indicative for a symmetric distribution. We observe that the total claim amounts are accumulated between 44 and 52 billion TL. Moreover, the Shapiro–Wilk (SW) normality test indicates that the total claim amounts are normally distributed with p-value of 0.8414. The summary statistics of the claim numbers infers that the data follow a right-skewed distribution.

Significant spikes in the graphs of autocorrelation functions (ACF) detect the presence of serial dependence in both variables (Figure 3). Additionally, we check the stationarity of both series by performing Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and Augmented Dickey–Fuller (ADF) test. The results of KPSS conclude the nonstationarity of total claim amounts and claim numbers ($p < 0.01$). ADF test is performed for claim numbers ($p = 0.3986$) and total claim amounts ($p = 0.2081$) resulting in rejecting stationarity in both series.

6.2. BHMM fitting and analyses

Although the total claim amounts distribution in literature is not commonly taken as normal, the theory developed on HMM concentrates mostly on the

TABLE 3

THE SELECTION CRITERIA FOR THE BEST FITTING POISSON–NORMAL HMM.

No. of states	No. of parameters	log-likelihood	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

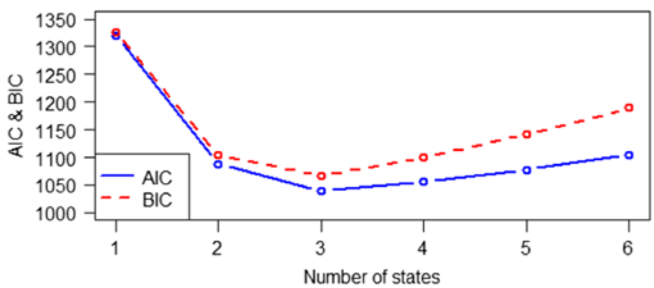

FIGURE 4: Poisson–Normal BHMM fit criteria for different state numbers.

Poisson distribution and normal distribution (Zucchini and Guttorp, 1991; Zucchini and MacDonald, 2009; Dias and Ramos, 2014). On the other hand, goodness-of-fit tests on the claim data support the assumption on normality. For this reason, the case study is applied to the Poisson–Normal BHMM. The iterative procedure of the algorithm is implemented in *R*. For calculation simplicity and convenience, the total claim amounts and the claim numbers are adjusted by 1000 and 100, respectively.

Poisson–Normal HMMs fit with respect to one to six states using EM algorithm are performed. The goodness-of-fit results in Table 3 show that the one-state model has the weakest fit to the insurance data. Despite an increasing number of states gives a better result for log-likelihood, yet it demands more parameters to estimate. Therefore, based on both AIC and BIC values, the model with three states is chosen to be the most suitable, compared to other choices (Figure 4).

The initial values of the off-diagonal transition probabilities are arbitrarily taken to be 0.1. As the starting values of the state-dependent means we use the lower quartile, the median quartile, and upper quartile of the observations, which are found to be for claim counts as (212.8, 222, 239), respectively. Similarly, these three starting values for the total claim amounts of quartiles are (44, 590, 47, 370, 48, 440), respectively. However, it is challenging to find an optimal initial value for σ. Hence, we perform EM estimation with several starting values and select the ones that give the maximum log-likelihood,

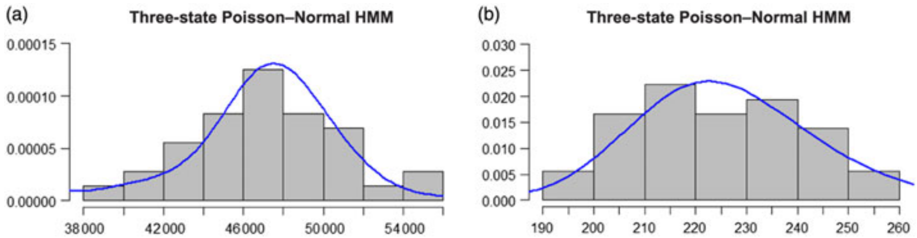

FIGURE 5: Marginal distributions of three-state Poisson–Normal HMM: (a) total claim amounts and (b) claim numbers.

leading us to the values (5000, 2000, 1300) as initial ones. Based on these, we estimate the transition probability matrix of three-state model as follows:

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

The transition probability matrix displays that the insurance claims are likely to remain in the same state. According to the probabilities of moving between the other states, it is most likely to shift from third state to the first with probability 0.289. The movement from state 1 to state 2 is likely with probability 0.196. The probability of transition between the other states is fairly small. Furthermore, it is extremely unlikely for the insurance claims to move from state 3 to state 2.

We find the estimated initial probabilities of the model as $u(1) = (0.8565, 0.1405, 0.0030)$, the parameters of joint state-dependent distribution as $\hat{\lambda} = (218.8, 223, 243.7)$, $\hat{\mu} = (45799.45, 46991.43, 49617.07)$, and $\hat{\sigma} = (5020.22, 2022.55, 2092.90)$. The estimated log-likelihood is $l = -501.387$. The marginal distributions are displayed with histograms in Figure 5.

The stationary distribution of three-state Poisson–Normal HMM is computed using Equation (2.3) and taken as the starting value of initial distribution, $u(1)$.

We observe that the shape of the conditional distributions may change significantly from one time point to another, which is quoted to the fact that some of the observations are extreme relative to their conditional distributions. The conditional distribution of the total claim amounts and numbers in February 2007, December 2008, and March 2009, given all the other observations, is compared with the actual values in the related month which are marked with a triangle symbol (Figure 6) corresponding to the actual total claim amount in that month. This justifies that employing the conditional distributions to check outliers is reasonable in our case. Regarding the residual plots (Figure 7, upper row), it is obvious that the selected model provides an optimal fit to the data. In addition, we apply SW normality test to pseudo-residuals, which confirms normality assumption (p-value 0.6546 for claim numbers, 0.4537 for total claim amounts) verified by Q–Q plots of pseudo-residuals (Figure 7, bottom row).

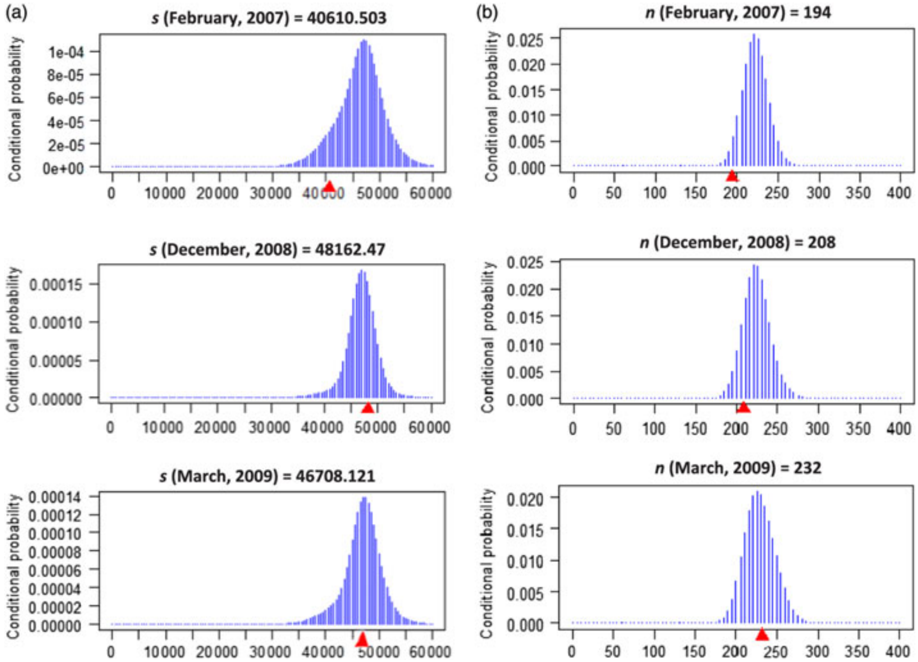

FIGURE 6: Conditional distributions for certain months (in brackets) given other observations: (a) total claim amounts and (b) claim numbers.

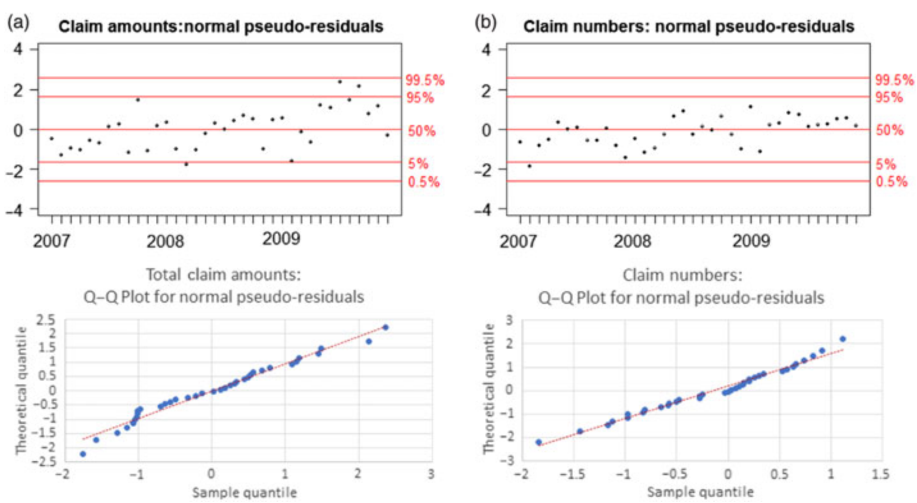

FIGURE 7: The graph of ordinary pseudo-residuals with certain confidence levels and their Q–Q plots: (a) total claim amounts and (b) claim numbers.

For the fitted three-state Poisson–Normal HMM, we derive state probabilities that are necessary for performing local decoding. Utilizing Figure 8, we are interested in defining hidden states that are most probable to have given rise to the sequence of observed values. We conduct local and global decoding both for total claim amounts and claim numbers to derive the most likely sequence of states by employing Viterbi algorithm. The theoretical inferences on the state probabilities, state prediction, global, and local decoding are given in Zucchini and MacDonald (2009).

Observing Figures 8 and 9, we state that:

(i) From January 2007 to April 2007 the total claim amounts and claim numbers were in state 1. Afterwards, the insurance claims went into state 2 and lasted for about 9 months (May 2007 to January 2008).
(ii) In February 2008, the claims went back to state 1 and remained there until April 2008. Notably, that in January 2008, the probabilities of being in state 1 (0.478) and state 2 (0.462) are fairly close to each other. It also confirms the difference of global and local decoding results in January 2008.
(iii) According to Figure 8, from May 2008 to January 2009, the highest state probabilities were in state 2. Once again, we notice that the local and global decoding present different results in February 2009 and April 2009. According to the local decoding the claims are in state 1 in February 2009, then switch to state 2 in the next month, and move to state 3 in April 2009. On the other hand, the global decoding results state that the claims remain in state 2 for almost 1 year from May 2008 to April 2009.
(iv) From May 2009 during 2 months the claims are in state 3, then those switch to state 1 and stay there until September 2009 and went back to state 3, remaining there until the end of 2009.

In order to analyze the nature of three states, we derive the claim severity for each month, which is the rate of total claim amounts and claim numbers, the results are presented in Table 4. We use the result of local decoding since it could capture the seasonal behavior of the claims during 2007 and 2008. We group the claim severity according to the local decoding results and derive the mean value of claim severity values in each state, whose results are in Table 5.

Based on these severity results, we assign state definitions as follows:

(i) State 3 presents the lowest mean value of the claim severity; thus, we call state 3 as the low state.
(ii) State 2 shows the highest mean value; therefore, it is named as the high state.
(iii) State 1 is called as the medium state.

We suppose that there is a seasonal impact on the insurance claims, since the local decoding shows that from January to April the claims were in the medium state, and from May to January the claims were in the high state during 2 years.

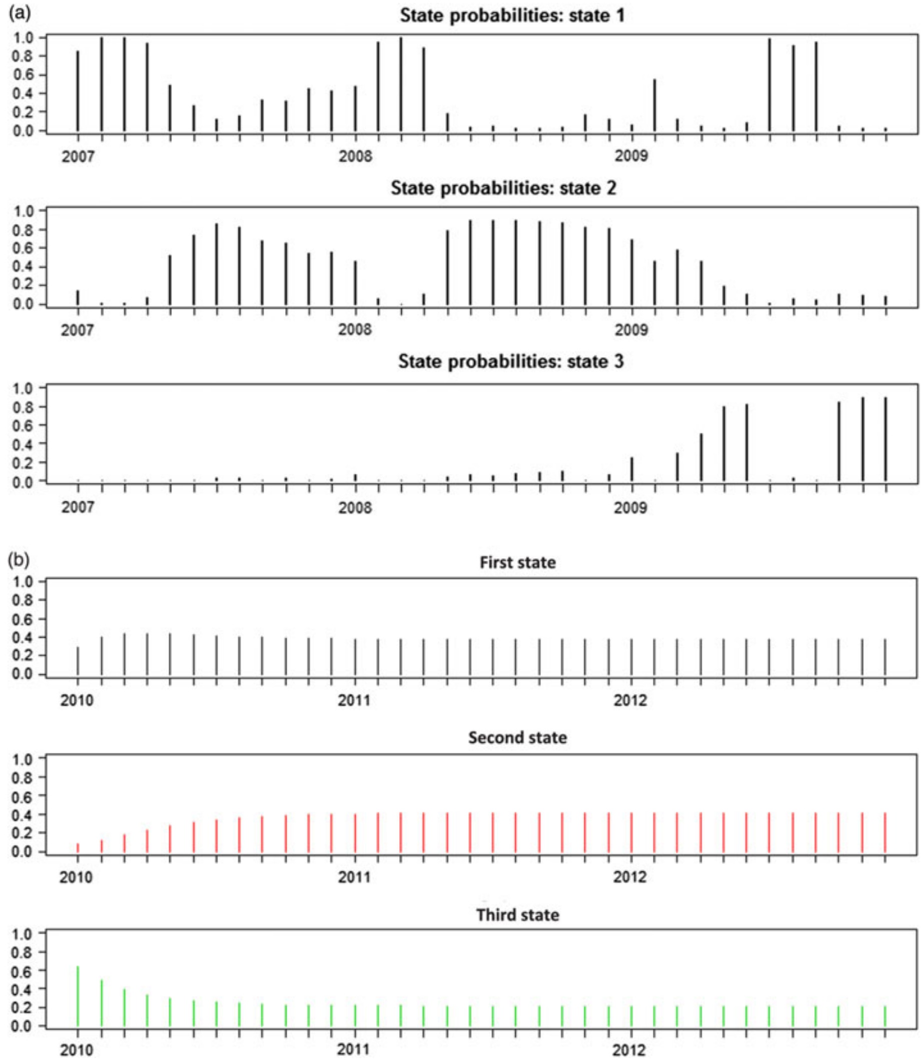

FIGURE 8: State prediction fitted three-state Poisson–Normal HMM: (a) state probabilities and (b) state prediction.

In 2009, the claims behave differently in comparison with previous 2 years, moving to the low state. The low state is characterized by the high total claim amounts and, at the same time, by the high claim numbers, resulting at the lowest claim severity.

Additionally, we obtain state probabilities for 3 years ahead that can be used for analysis of further claim behavior. We interpret that in the beginning of 2010, the observations will be dependent on the low state and the following few months those will continue with the medium state; during 2 years the

TABLE 4

CLAIM SEVERITY (TL) BETWEEN JANUARY 2007 AND DECEMBER 2009.

Month	Claim severity	Month	Claim severity
January 2007	2110.584	July 2008	2145.219
February 2007	2090.200	August 2008	2138.817
March 2007	1992.717	September 2008	2191.086
April 2007	1980.344	October 2008	2082.565
May 2007	1968.840	November 2008	2050.662
June 2007	2023.900	December 2008	2317.509
July 2007	2117.023	January 2009	2018.179
August 2007	2234.867	February 2009	2056.964
September 2007	2065.220	March 2009	2016.932
October 2007	2285.005	April 2009	1927.231
November 2007	2115.018	May 2009	2032.901
December 2007	2370.503	June 2009	2036.394
January 2008	2243.221	July 2009	2263.113
February 2008	2097.330	August 2009	2154.857
March 2008	1877.836	September 2009	2217.261
April 2008	1943.056	October 2009	2033.440
May 2008	1977.718	November 2009	2071.967
June 2008	2009.525	December 2009	1992.269

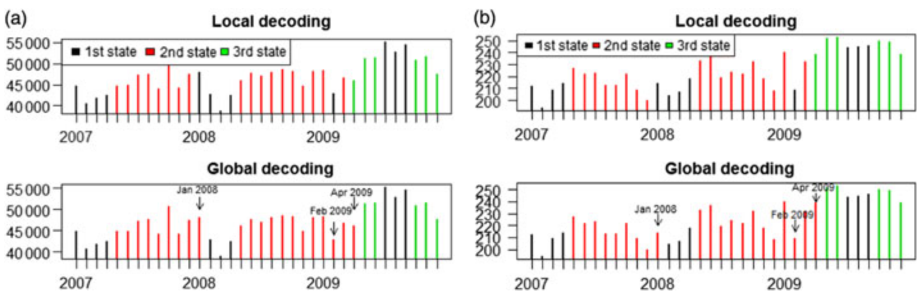

FIGURE 9: Local and global decoding for three-state Poisson–Normal case: (a) claim amounts and (b) claim numbers.

medium and high states have almost equal probabilities and are dominant compared to the low state, based on Figure 8.

Four forecast distributions for total claim amounts and claim numbers are displayed in Figure 10 for the times January 2010, March 2010, June 2010, and August 2011. The distributions are compared with the limiting distributions, that is, the marginal distributions of the Poisson–Normal HMM. It is clear that the forecast distributions approach the limiting distribution shown with continuous (red) line in the graphs. These results are found based on the theoretical framework given earlier (Zucchini and MacDonald, 2009).

TABLE 5

MEAN OF CLAIM SEVERITY IN THREE STATES.

	Mean of the claim severities
State 1	2073.440
State 2	2118.255
State 3	2033.394

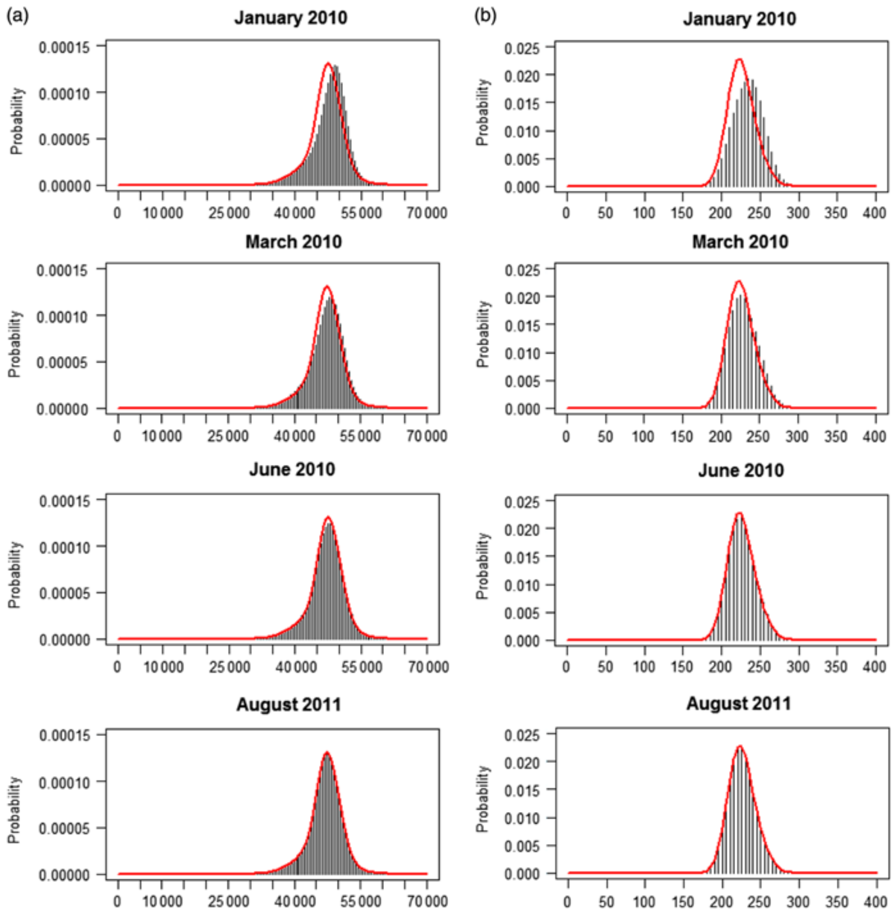

FIGURE 10: Forecast distributions for four periods: (a) total claim amounts and (b) claim numbers (red line is limiting distribution).

7. CONCLUSION

It is proposed to illustrate BHMM capturing the claim dependence, which allows claim numbers and aggregate claim amounts to be mutually and serially dependent through an underlying hidden state. We modify and derive the classical HMM definitions and theorems to the bivariate case. Three different models, Poisson–Normal HMM, Poisson–Gamma HMM, and Negative Binomial–Gamma HMM are studied and their parameter estimates are derived. For parameter estimation of the model, we conduct EM algorithm which requires also further analytical modification. We utilize an algorithm with required derivations which maximizes the state-dependent part of CDLL of the proposed models. Real-life application of the proposed model is applied on MTPL data collected from Turkish insurance sector (Istanbul) to examine the performance of our model. Three-state Poisson–Normal HMM is selected as the most suitable model whose results are used to determine forecast distributions both for claim numbers and for total claim amounts and perform state prediction. According to the local and global decoding results, we determine the claim dependency on low, medium, and high states, which allows us to estimate claim severity based on the dependence level.

The main advantage of the model is a flexibility in a sense of accommodating different types of data, such as modeling a bivariate series with one discrete and one continuous variable. Moreover, the proposed model is applicable in various fields of life and nonlife 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 specialists allowing them to distinguish the character of events or factors influencing the claim behavior.

References

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

Äuerle, N. and Müller, A. (1998) Modeling and comparing dependencies in multivariate risk portfolios. *ASTIN Bulletin: The Journal of the IAA*, **28**(1), 59–76.

Avanzi, B., Wong, B. and Yang, X. (2016) A micro-level claim count model with overdispersion and reporting delays. *Insurance: Mathematics and Economics*, **71**, 1–14.

Badescu, A.L., Lin, X.S. and Tang, D. (2016) A marked Cox model for the number of IBNR claims: Theory. *Insurance: Mathematics and Economics*, **69**, 29–37.

Boudreault, M., Cossette, H. and Marceau, Ã. (2014) Risk models with dependence between claim occurrences and severities for Atlantic hurricanes. *Insurance: Mathematics and Economics*, **54**, 123–132.

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997) *Actuarial Mathematics*. Shaumburg, IL: The Society of Actuaries.

Cheung, L.W.K. (2004) Use of runs statistics for pattern recognition in genomic DNA sequences. *Journal of Computational Biology*, **11**(1), 107–124.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38.

Denuit, M., Lefèvre, C. and Utev, S. (2002) Measuring the impact of dependence between claims occurrences. *Insurance: Mathematics and Economics*, **30**(1), 1–19.

Dhaene, J. and Goovaerts, M.J. (1997) On the dependency of risks in the individual life model. *Insurance: Mathematics and Economics*, **19**(3), 243–253.

Dias, J.G. and Ramos, S.B. (2014) Dynamic clustering of energy markets: An extended hidden Markov approach. *Expert Systems with Applications*, **41**(17), 7722–7729.

Forney, G.D. (1973) The Viterbi algorithm. *Proceedings of the IEEE*, **61**(3), 268–278.

Garanti Securities (2016) *Turkish Non-life Insurance Sector - Waiting for the Dust to Settle*. Istanbul: Garanti Securities.

GARRIDO, J., GENEST, C. and SCHULZ, J. (2016) Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, **70**, 205–215.

HASSAN, M.R. and NATH, B. (2005) Stock market forecasting using hidden Markov model: A new approach. *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications, 2005, ISDA'05*, pp. 192–196. IEEE, Warshaw, Poland.

HUDECOVÁ, Š. and PEŠTA, M. (2013) Modeling dependencies in claims reserving with GEE. *Insurance: Mathematics and Economics*, **53**(3), 786–794.

Inflation rates (2016), http://www.bigpara.com/haberler/enflasyon-verileri/, last accessed 13 July 2016.

JLT SIGORTA VE REASURANS BROKERLIGI A.S. (2017) *Turkish Insurance Market Outlook 2016-17.* Istanbul: JLT Sigorta ve Reasurans Brokerligi A.S.

KROGH, A., BROWN, M., MIAN, I.S., SJÖLANDER, K. and HAUSSLER, D. (1994) Hidden Markov models in computational biology: Applications to protein modeling. *Journal of molecular biology*, **235**(5), 1501–1531.

LITTLE, R.J. and RUBIN, D.B. (2014) *Statistical Analysis with Missing Data*. Hoboken: John Wiley & Sons.

OFLAZ, Z. (2016) *Bivariate Hidden Markov model to capture the claim dependency*. Unpublished M.Sc. Thesis, METU, Department of Statistics.

PAROLI, R., REDAELLI, G. and SPEZIA, L. (2000) Poisson hidden Markov models for time series of overdispersed insurance counts. *Proceedings of the XXXI International ASTIN Colloquium (Porto Cervo, 18–21 September 2000)*. Roma: Instituto Italiano degli Attuari.

PEŠTA, M. and OKHRIN, O. (2014) Conditional least squares and copulae in claims reserving for a single line of business. *Insurance: Mathematics and Economics*, **56**, 28–37.

PINQUET, J. (2000) Experience rating through heterogeneous models. In *Handbook of Insurance*, pp. 459–500. Netherlands: Springer.

RABINER, L.R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

RABINER, L.R. and JUANG, B.H. (1993) *Fundamentals of Speech Recognition*, Vol. 14. Englewood Cliffs: PTR Prentice Hall.

REN, J. (2012) A multivariate aggregate loss model. *Insurance: Mathematics and Economics*, **51**(2), 402–408.

ROCCA, G. (2016) Insurance Objectives, studio D, http://work.chron.com/insurance-objectives-23793.html, last accessed 12 July 2016.

TSE, Y.K. (2009) *Nonlife Actuarial Models. Theory, Methods and Evaluation*. New York: Cambridge University Press.

VITERBI, A. (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, **13**(2), 260-269.

WANG, G. AND YUEN, K.C. (2005) On a correlated aggregate claims model with thinning-dependence structure. *Insurance: Mathematics and Economics*, **36**(3), 456–468.

YUEN, K.C., GUO, J. and WU, X. (2002) On a correlated aggregate claims model with Poisson and Erlang risk processes. *Insurance: Mathematics and Economics*, **31**(2), 205–214.

ZUCCHINI, W. and GUTTORP, P. (1991) A hidden Markov model for space -time precipitation. *Water Resources Research*, **27**(8), 1917–1923.

ZUCCHINI, W. and MACDONALD, I. L. (2009) *Hidden Markov Models for Time Series: An Introduction Using R*. Boca Raton: Chapman and Hall, CRC.

ZARINA NUKESHTAYEVA OFLAZ
Department of Statistics
Middle East Technical University
Ankara, Turkey
Department of Insurance and Social Security
KTO Karatay University
Konya, Turkey
E-Mail: nukesh.zar@mail.ru

CEYLAN YOZGATLIGIL
Department of Statistics
Middle East Technical University
Ankara, Turkey
E-Mail: ceylan@metu.edu.tr

A. SEVTAP SELCUK-KESTEL (Corresponding author)
Institute of Applied Mathematics, Actuarial Sciences
Middle East Technical University
Ankara, Turkey
E-Mail: skestel@metu.edu.tr

APPENDIX

The proofs of Theorems 1–3 are given in this part.

Proof of Theorem 1. The joint state-dependent probability for the Poisson–Normal HMM is given by

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

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

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

with respect to the parameters of the joint state-dependent distribution. $\hat{u}_j(t)$ is given in Equation (3.4).

The state-dependent part of the CDLL for the Poisson–Normal bivariate HMM is defined as follows:

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

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:

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

and hence that

$$\hat{\lambda}_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t) n_t}{\sum_{t=1}^{T} \hat{u}_j(t)}. \tag{A.2}$$

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

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

then

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

Analogously for σ_j^2:

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

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

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \hat{\mu}_j)^2}{\sum_{t=1}^{T} \hat{u}_j(t)}. \tag{A.3}$$

For confidence that the estimated parameters maximize the state-dependent part of CDLL, we check second derivatives of F with respect to parameters:

$$\left.\frac{d^2 \ln L}{d\lambda_j^2}\right|_{\lambda_j=\hat{\lambda}_j} = -\sum_{t=1}^{T} \hat{u}_j(t)\left.\frac{n_t}{\lambda_j^2}\right|_{\lambda_j=\hat{\lambda}_j} < 0,$$

since $n_t > 0$ and $\hat{u}_j(t) = \{0, 1\}$ by definition. It is obvious that the following satisfies:

$$\left.\frac{d^2 \ln L}{d\mu_j^2}\right|_{\mu_j=\hat{\mu}_j} = -\sum_{t=1}^{T} \hat{u}_j(t)\left.\frac{1}{\sigma_j^2}\right|_{\mu_j=\hat{\mu}_j} < 0.$$

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

$$\begin{aligned}\left.\frac{d^2 \ln L}{d\sigma_j^2}\right|_{\sigma_j^2=\hat{\sigma}_j^2} &= \sum_{t=1}^{T} \hat{u}_j(t)\left.\left[-\frac{1}{2\sigma_j^2} + \frac{(s_t - \mu_j)^2}{2(\sigma_j^2)^2}\right]\right|_{\sigma_j^2=\hat{\sigma}_j^2} \\ &= \left.\frac{\sigma_j^2 \sum_{t=1}^{T} \hat{u}_j(t) - 2\sum_{t=1}^{T} \hat{u}_j(t)(s_t - \mu_j)^2}{2(\sigma_j^2)^3}\right|_{\sigma_j^2=\hat{\sigma}_j^2}.\end{aligned}$$

It is sufficient 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:

$$\sigma_j^2 - \frac{2\sum_{t=1}^{T}\hat{u}_j(t)(s_t-\mu_j)^2}{\sum_{t=1}^{T}\hat{u}_j(t)}\Bigg|_{\sigma_j^2=\hat{\sigma}_j^2} = \frac{\sum_{t=1}^{T}\hat{u}_j(t)(s_t-\mu_j)^2}{\sum_{t=1}^{T}\hat{u}_j(t)} - \frac{2\sum_{t=1}^{T}\hat{u}_j(t)(s_t-\mu_j)^2}{\sum_{t=1}^{T}\hat{u}_j(t)}$$

$$= -\frac{\sum_{t=1}^{T}\hat{u}_j(t)(s_t-\mu_j)^2}{\sum_{t=1}^{T}\hat{u}_j(t)} < 0. \qquad \blacksquare$$

Proof of Theorem 2. The joint state-dependent probability for the Poisson–Gamma HMM is given by

$$p_j(s_t, n_t) = \frac{\beta_i^{\alpha_j} s_t^{\alpha_j-1}\lambda_j^{n_t} e^{-\beta_j s_t-\lambda_j}}{\Gamma(\alpha_j) n_t!} \qquad s_t > 0, n \in N$$

and the state-dependent part of CDLL:

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

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:

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

and hence that

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

Analogously for β_j:

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

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:

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

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

$$\frac{d\ln L}{d\alpha_j} = \sum_{t=1}^{T} \hat{u}_j(t)\left[-\frac{d}{d\alpha_j}\log\Gamma(\alpha_j) + \log\frac{\alpha_j\sum_{t=1}^{T}\hat{u}_j(t)}{\sum_{t=1}^{T}\hat{u}_j(t)s_t} + \log s_t\right] = 0 \tag{A.4}$$

In order to estimate the above equation, numerical maximization is required.

Finally, we check the second derivatives of F with respect to the parameters. For λ_j, see Proof of Theorem 1:

$$\left.\frac{d^2\ln L}{d\alpha_j^2}\right|_{\alpha_j=\hat{\alpha}_j} = -\sum_{t=1}^{T}\hat{u}_j(t)\frac{d^2}{d\alpha_j^2}\log\Gamma(\alpha_j)\bigg|_{\alpha_j=\hat{\alpha}_j} < 0,$$

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:

$$\left.\frac{d^2\ln L}{d\beta_j^2}\right|_{\beta_j=\hat{\beta}_j} = -\sum_{t=1}^{T}\hat{u}_j(t)\frac{\alpha_j}{\beta_j^2} < 0,$$

since $\alpha_j > 0$. ■

Proof of Theorem 3. The joint state-dependent probability for the Negative Binomial–Gamma HMM is given by

$$p_j(s_t, n_t) = \frac{\binom{n_j-1}{r_j-1}\beta_j^{\alpha_j}s_t^{\alpha_j-1}e^{-\beta_j s_t}p_j^{r_j}(1-p_j)^{n_t-r_j}}{\Gamma(\alpha_j)} \qquad s_t > 0, n \in N$$

and the state-dependent part of CDLL:

$$\ln L = \sum_{t=1}^{T}\hat{u}_j(t)\log p_j(s_t, n_t) = \sum_{t=1}^{T}\hat{u}_j(t)\left[\log\binom{n_j-1}{r_j-1} + \alpha_j\log\beta_j + (\alpha_j-1)\log s_t\right.$$
$$\left. - \beta_j s_t + r_j\log p_j + (n_t-r_j)\log(1-p_j) - \log\Gamma(\alpha_j)\right]. \tag{A.5}$$

Estimation of the parameters, α_j and β_j, and the second derivative of F with respect to the mentioned parameters are provided in the Proof of Theorem 2. In the following, the derivation of $\hat{p}_j$ and $\hat{r}_j$ is shown:

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

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

Then

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

$$\hat{p}_j = \frac{r_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j) + r_j \sum_{t=1}^{T} \hat{u}_j(t)}. \tag{A.6}$$

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

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

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

$$\begin{aligned}\frac{d\ln L}{dr_j} = \sum_{t=1}^{T} \hat{u}_j(t)\Bigg[&\frac{d}{dr_j}\log\binom{n_j - 1}{r_j - 1} + \log\frac{r_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j) + r_j \sum_{t=1}^{T} \hat{u}_j(t)} \\ &- \log\left(1 - \frac{r_j \sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t)(n_t - r_j) + r_j \sum_{t=1}^{T} \hat{u}_j(t)}\right)\Bigg] = 0,\end{aligned} \tag{A.7}$$

which requires numerical tools to maximize the expression:

$$\left.\frac{d^2\ln L}{dp_j^2}\right|_{p_j = \hat{p}_j} = \sum_{t=1}^{T} \hat{u}_j(t)\left[-\frac{r_j}{p_j^2} - \frac{n_t - r_j}{(1-p_j)^2}\right].$$

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 Equation (A.5), we have

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

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:

$$\left.\frac{d \ln L^2}{dr_j^2}\right|_{r_j=\hat{r}_j} = -\sum_{t=1}^{T} \hat{u}_j(t) \frac{d^2}{dr_j^2} \log \binom{n_t - 1}{r_j - 1},$$

where

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

It is obvious that the expression above is positive. This completes the proof. ■