AIMS Mathematics, 10(10): 23411–23428.
DOI: 10.3934/math.20251040
Received: 17 July 2025
Revised: 21 September 2025
Accepted: 30 September 2025
Published: 15 October 2025

https://www.aimspress.com/journal/Math

Research article

Nonstationary transition Poisson-Lindley Hidden Markov model for community-based disaster insurance claim

Hilda Azkiyah Surya[1], Sukono[2,*], Herlina Napitupulu[2] and Noriszura Ismail[3]

[1] Doctoral Program in Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia
[2] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia
[3] Department of Mathematical Sciences, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

* **Correspondence:** Email: sukono@unpad.ac.id.

Abstract: This study proposes a nonstationary Poisson-Lindley Hidden Markov Model (PL-HMM) as a novel framework for modeling the frequency of community-based disaster insurance claims. The model accounts for both serial dependence and overdispersion in claim counts through hidden risk states, while nonstationary transition probabilities are introduced via a sliding-window mechanism. Parameters are estimated using the Generalized Expectation-Maximization (GEM) algorithm, supported by a theoretical foundation to ensure a monotonic improvement of the complete log-likelihood. The model was simulated using monthly claim frequency data from West Java Province, Indonesia. A comparative analysis against nonstationary Poisson HMMs with varying numbers of hidden states showed that the two-state nonstationary PL-HMM achieved the lowest Bayesian information criterion (BIC), thus indicating the best fit. A sensitivity analysis of sliding-window horizons (12, 24, and 36 months) demonstrated that persistence patterns of claim risk-states remained robust, with horizon changes reflecting alternative risk measurement periods. The results highlight that the proposed model effectively captures time-varying claim risks, particularly the alternation between low- and high-claim periods, while realistically reflecting the empirical dominance of high-claim regimes. Beyond the simulation data, a nonstationary PL-HMM is flexible and applicable to other regions that exhibit overdispersed claim data, making it a valuable framework for adaptive premium design and disaster risk financing in community-based insurance schemes.

Keywords: nonstationary; Poisson-Lindley; hidden Markov model; community-based; claim
Mathematics Subject Classification: 62M05, 62P05, 91B05, 91G05

1. Introduction

Natural disasters, including floods, earthquakes, and landslides, are posing severe threats to the socio-economic resilience of communities, particularly in developing countries such as Indonesia. A strategic effort to mitigate the financial impacts of such disasters is the implementation of Community-Based Disaster Insurance (CBDI) schemes—insurance mechanisms specifically designed to protect the most vulnerable local community from disaster impacts. These schemes facilitate faster recovery processes and promote community participation in disaster risk management. In the context of CBDI, providing affordable premiums for communities becomes a main priority. A study by [1] showed that affordable premiums create a positive trend in disaster insurance participation, while still ensuring adequate protection against property damage, thus highlighting that premium affordability is a crucial factor for the implementation of CBDI schemes. Therefore, claim modeling—which is directly linked to premium determination—needs to be conducted in a realistic way to accurately represent claim risk. Affordable premiums can be achieved through nonstationary claim risk modeling, whereby the expected claims are updated to reflect time-varying risk dynamics. Such an approach prevents high premium setting and ensures that CBDI schemes remain sustainable while effectively addressing the needs of communities in disaster-prone areas. However, the risk of claims cannot be directly observed; it is inherently latent but closely related to observable variables such as claim frequency and claim severity. This motivates the use of Hidden Markov Model (HMM), which allows hidden risk states to be inferred from the observed claim data. HMM has been applied across several insurance domains, including vehicle claim estimation [2], aggregate claim dependence analyses [3], and optimal investment modeling for insurance firms under model uncertainty [4]. However, most HMM applications in insurance still rely on the assumption of stationarity, particularly in the transition probability matrix (t.p.m.) between hidden states and the initial probability distribution (i.p.d.). The assumption is often unsuitable for disaster-related claim data, which is typically highly volatile and shaped by changing risk factors over time, such as seasonal rainfall intensity and local geological conditions. For instance, the probability of low- or high-risk states may increase significantly during the rainy season or decrease after the construction of protective infrastructure. Such evolving dynamics are difficult to capture with a stationary HMM. To address this limitation, this study proposes a novel modeling method, namely the nonstationary transition Poisson-Lindley HMM (PL-HMM), designed specifically for community-based disaster insurance claims. The model incorporates three major components: (i) Poisson-Lindley distribution is used for the emission distribution, which provides a more flexible representation of the overdispersed claim frequency; (ii) an HMM framework to accommodate hidden risk claims; and (iii) a nonstationary transition through a sliding-window method, which allows the transition dynamics to evolve over time in response to the changing claim risk levels. Although overdispersion can be addressed using a Negative Binomial distribution, as previously shown by [3] in the context of insurance claims, such methods have generally been modeled under a stationary assumption. Moreover, a Negative Binomial distribution faces important limitations. The probability mass function (pmf) of Negative Binomial distributions rapidly declines as claim frequencies increase. For claim frequencies that are very large (e.g., exceeding 100), the corresponding

emission probabilities in an HMM become extremely small, often leading to numerical instability during the estimation and an increased algorithm complexity. These challenges make the Poisson-Lindley distribution a more robust and practical choice in nonstationary HMM frameworks for highly fluctuating disaster claim data. The choice of a nonstationary assumption is intended to produce affordable CBDI premiums by making them adaptive to dynamic risk patterns. Moreover, this assumption is grounded in the empirical reality that claim risk is observed through the claim frequency. Claim frequency data from CBDI schemes across regions (see Figure 1) exhibit clear shifts in variance, thus indicating nonstationarity [5]. Accordingly, the t.p.m. that evolves over time through a sliding-window approach becomes essential in the modeling process. In line with [4,6], which emphasize the need for flexibility to capture unobserved risk dynamics, the main contribution of this study lies in presenting a more realistic modeling framework to demonstrate that a disaster-related claim risk is not stationary. In other words, claim risk is dynamic and influenced by various external factors, thus requiring a nonstationary approach to be represented more accurately.

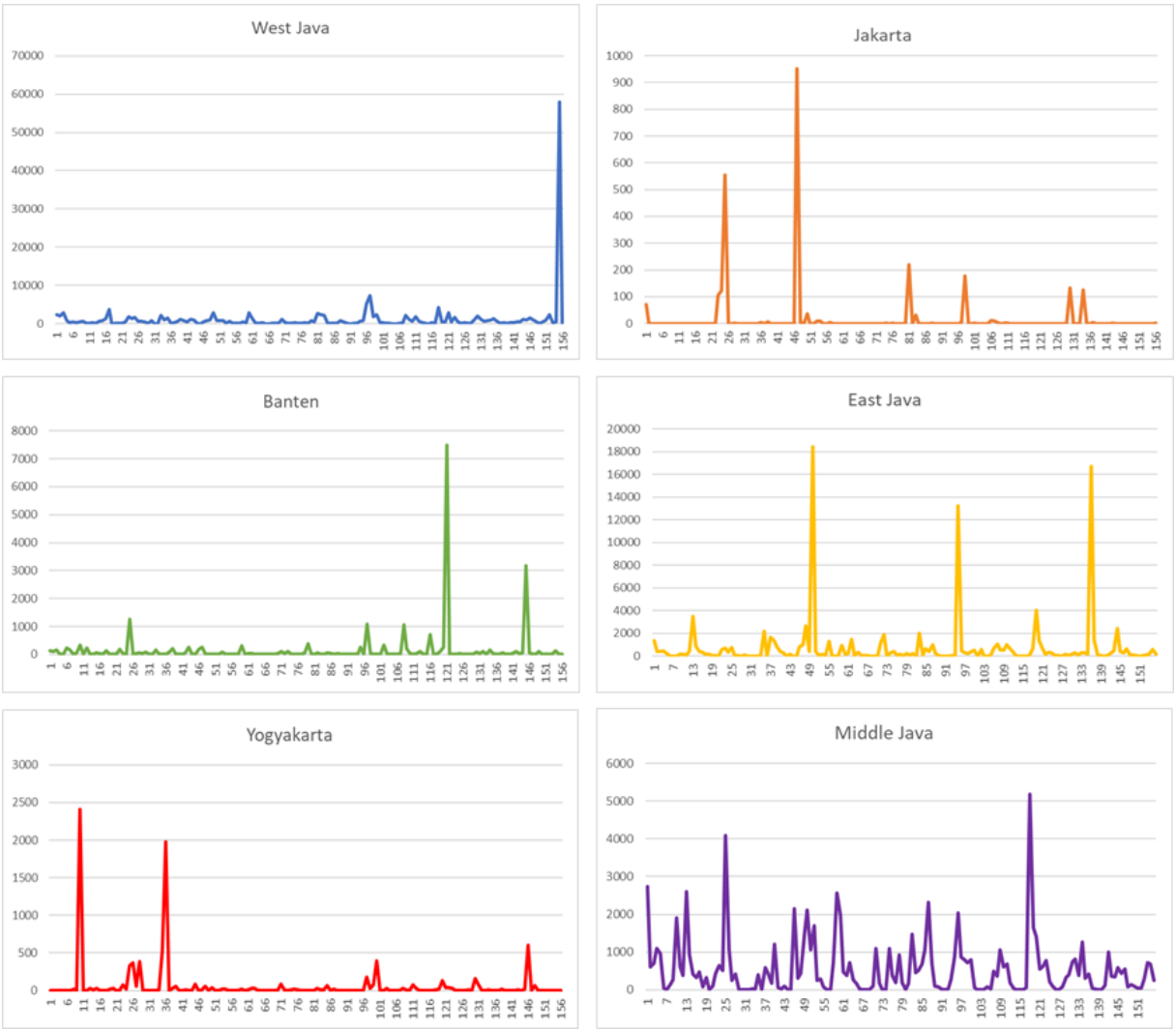

Figure 1. Nonstationary diagnostics of monthly claim frequencies.

2. Materials and methods

2.1. Poisson-Lindley distribution

The Poisson-Lindley distribution is a discrete probability distribution developed to address the issue of overdispersion, a condition where the variance of the data exceeds its mean, which is a common occurrence in claim frequency and disaster-related data [2,7]. This distribution arose from a mixture of the Poisson and Lindley distributions, which makes the model more flexible in representing data with high variability. According to [8], the pmf of a Poisson-Lindley distribution with parameter $\theta > 0$ is given by Eq (1):

$$P(Y = y; \theta) = \frac{\theta^2(\theta + y + 2)}{(1+\theta)^{y+3}}, y = 0,1,2, \ldots \tag{1}$$

The expectation of a Poisson-Lindley distribution is formulated in Eq (2):

$$E(Y) = \frac{\theta+2}{\theta(1+\theta)}. \tag{2}$$

2.2. HMM

According to [7], an HMM is a discrete-time stochastic process consisting of the pair $\{Y_t, X_t; t \in \mathbb{N}\}$, where $\{X_t, t \in \mathbb{N}\}$ forms a Markov Chain that represents the underlying causes of events that could not be directly observed. Meanwhile, $\{Y_t, t \in \mathbb{N}\}$ is a sequence of observations that depended on $\{X_t, t \in \mathbb{N}\}$. An HMM has several major parameters, as shown by [9,10].

The initial state distribution $\boldsymbol{\pi}$, denoted by $\pi_i = P(X_1 = i)$, is the probability of being in hidden state i at time 1 for $i = 1,2, \ldots, m$, where m is the total number of hidden states. In addition, the t.p.m.$A = [a_{ij}]$, where a_{ij} is represented by $P(X_{t+1} = j | X_t = i)$ as the time-independent transition probability of hidden state from i to j between time $t+1$ and t. The emission matrix $B = [b_i(y)]$, where $b_i(y)$ is denoted by $P(Y_t = y | X_t = i)$, is the probability of making an observation Y_t at a given state X_t for $i = 1,2, \ldots, m$ and $y = 1,2, \ldots, n$, where n is the number of observations.

2.3. Poisson HMM and the parameters estimation

A Poisson HMM is a special case of HMM, in which the observable variables $\{Y_t, t \in \mathbb{N}\}$ are assumed to follow a Poisson distribution with the parameter λ_i, depending on the hidden state $X_t = i$. The unobservable process $\{X_t, t \in \mathbb{N}\}$ is modeled as a discrete Markov chain with a state space $S = \{1, 2, \ldots, m\}$. Accordingly, three sets of parameters must be estimated, namely the state transition matrix A, the initial state distribution $\boldsymbol{\pi}$, and the emission distribution $b_i(y) = \frac{\lambda_i^y \cdot \exp(-\lambda_i)}{y!}$, where the vector $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m]$ represents the mean claim frequency for each hidden state.

Parameter estimation is generally performed using the Maximum Likelihood method, which is solved through either the Generalized Expectation-Maximization (GEM) algorithm or the Baum-Welch procedure. In the M-step, the parameter λ_i of the Poisson HMM is updated as follows:

$$\lambda_i^{new} = \frac{\sum_{t=1}^{T} \gamma_t(i) \cdot y_t}{\sum_{t=1}^{T} \gamma_t(i)}. \tag{3}$$

Equation (3) was first introduced by [11], who demonstrated that the Poisson HMM is capable of capturing variations in risk levels in insurance claim data that cannot be explained by a simple Poisson model. Furthermore, subsequent studies by [3,7] have shown that Eq (3) indeed maximizes the expectation of the complete ln-likelihood.

2.4. Data

The data used in this study is secondary data obtained from the official website of the National Disaster Management Authority (BNPB) of Indonesia through https://gis.bnpb.go.id/. The data is a monthly time series of the number of houses damaged by disaster from 2010 to 2022. Specifically, the dataset consists of the total number of damaged houses per month, which are classified according to the severity of the damage, namely minor, moderate, and severe. Additionally, this section provides an exploratory view across multiple regions to demonstrate that nonstationary is pervasive. The cross-region display is not to tailor the model by region, but rather to show that the model must be adaptive over time (time-varying risk). In every region we examine, we observe a variance shift—consistent with the introduction and supporting nonstationary transitions without merely "chasing data fit". Figure 1 presents six representative regions.

While Figure 1 illustrates that nonstationary is pervasive across regions, the simulation in this study focuses on West Java as a simulation. During the simulation, the total number of damaged houses per month was assumed to represent the frequency of claim under the CBDI scheme. This assumed that each instance of house damage due to disaster was associated with one claim in the CBDI framework.

The frequency of claim in this fluctuated significantly from one month to another, with certain periods showing very low claim, and others marked by sharp spikes in houses damage. This pattern signifies the existence of distinct phases of claim risk, which statistically assumes the presence of two or more hidden states governing the claim process. This observation supports the adoption of an HMM to capture the dynamics of transition between the underlying risk conditions. The descriptive statistics in Table 1 show that the average (mean) number of damaged houses per month was 1147.75, while the variance reached as high as 21,980,033.891935. The gap between the mean and variance strongly signifies the presence of overdispersion in the data [11]. The outcome shows that the variability exceeds the expectation under a standard Poisson model. Therefore, the Poisson-Lindley distribution, which is theoretically well-suited to model over dispersed count data, is considered appropriate for this study. The implementation of the nonstationary transition PL-HMM in this study is justified by the characteristics of the data, which demands a modeling method capable of accommodating both the changing risk patterns over time and the highly variable nature of disaster-related claim frequencies.

Table 1. Descriptive statistics of claim frequency in CBDI.

Mean	Median	Standard Deviation	Variance	Kurtosis	Skewness
1147.75	375.363362	4688.287	21,980,033.891935	141.028105	11.615884

3. Results

3.1. Nonstationary PL-HMM

The model assumes that $\{(X_t; Y_t); t \in \mathbb{N}_0\}$ is a stochastic process, where X_t represents the claim risk level following a nonhomogeneous Markov chain with a state space $S = \{1, 2, \dots, m\}$. The sequence of random variables $\{X_t, t \in \mathbb{N}_0\}$ shows claim risk levels that are not directly observable and constitute hidden states at time t. Meanwhile, $\{Y_t, t \in \mathbb{N}_0\}$ denote the observable sequence corresponding to the claim frequencies, where Y_t is a Poisson-Lindley distributed random variable that depends on the claim risk level X_t at time t. The relationship between the claim frequency and risk level is defined by Eq (4):

$$Y_t|X_t = i \sim PL(\theta_i), \text{ for } i = 1,2,\dots,m \text{ and } t \in \mathbb{N}_0. \tag{4}$$

Equation (4) shows that $\{Y_t, t \in \mathbb{N}_0\}$ is the observation sequence in the HMM with a Poisson-Lindley emission distribution, where θ_i represents the overdispersion parameter of the claim frequency influenced by hidden states $X_t = i$ (i.e., claim risk level).

The process introduces a sliding window method to capture the nonstationary dynamics of claim risk, where the observation period is divided into overlapping time segments denoted as w_k, the k^{th} window. Following the model, each window $w_k = \{1,2,3, \dots, T_k\}$, where T_k is the window length. The nonstationary transition probability from the risk level i to the risk level j in window w_k is mathematically expressed in Eq (5):

$$P(X_t = j|X_t = i\ ) = a_{ij}^{(w_k)}, t \in w_k, \tag{5}$$

where $a_{ij}^{(w_k)}$ represents the element of t.p.m. $A^{(w_k)}$ represents the window w_k with a dimension $m \times m$ and satisfies the condition $\sum_{j=1}^{m} a_{ij}^{(w_k)} = 1$ for all $i \in S$. Furthermore, the i.p.d. of the claim risk level at window w_1 is denoted by vector $\boldsymbol{\pi}^{(w_1)}$ and a row vector dimension $1 \times m$ that satisfies the condition $\sum_{i=1}^{m} \pi_i^{(w_1)} = 1$, as shown in Eq (6):

$$\boldsymbol{\pi}^{(w_1)} = \left[\pi_1^{(w_1)} \quad \pi_2^{(w_1)} \quad \cdots \quad \pi_m^{(w_1)}\right] = [P(x_1 = 1) \quad P(x_1 = 2) \quad \cdots \quad P(x_1 = m)]. \tag{6}$$

When the t.p.m. between claim risk levels is nonstationary, then the i.p.d for $k > 1$ also becomes nonstationary. Consequently, the i.p.d. of the claim risk level for $k > 1$ is given in Proposition 1.

Proposition 1. *Let $X = \{x_1, x_2, \dots, x_T\}$ be the sequence of classified claim risk levels, considered as hidden states divided into W segments based on sliding windows w_k. The t.p.m. with window w_k is shown by $A^{(w_k)} = \left[a_{ij}^{(w_k)}\right]$. When the i.d.p. of the claim risk level at window w_1 is denoted by vector $\boldsymbol{\pi}^{(w_1)}$, and the i.p.d. of claim risk levels at window w_k for $k > 1$ is given by $\boldsymbol{\pi}^{(w_k)} = \boldsymbol{\pi}^{(w_1)} \prod_{s=1}^{k-1} A^{(w_s)}$.*

Proof.

(i) Basis case: $\boldsymbol{\pi}^{(2)}$ for $k = 2$, the model has the following:

$$\boldsymbol{\pi}^{(w_2)} = P(x_2 = j) = \sum_{i=1}^{m} P(x_1 = i)P(x_2 = j|x_1 = i) = \sum_{i=1}^{m} \pi_i^{(w_1)} a_{ij}^{(w_1)}$$
$$= \pi_1^{(w_1)} a_{1j}^{(w_1)} + \pi_2^{(w_1)} a_{2j}^{(w_1)} + \cdots + \pi_m^{(w_1)} a_{mj}^{(w_1)} = \boldsymbol{\pi}^{(w_1)} A^{(w_1)}.$$

(ii) Inductive step: Assume the following for $k = n$:

$$\boldsymbol{\pi}^{(w_n)} = \boldsymbol{\pi}^{(w_1)} A^{(w_1)} A^{(w_2)} \dots A^{(w_{n-1})} = \boldsymbol{\pi}^{(w_1)} \prod_{s=1}^{n-1} A^{(w_s)}.$$

The process needs to show that the step holds for $k = n + 1$:

$$\boldsymbol{\pi}^{(w_{n+1})} = \boldsymbol{\pi}^{(w_1)}\, A^{(w_1)} A^{(w_2)} \dots A^{(w_{n-1})} A^{(w_n)} = \left(\boldsymbol{\pi}^{(w_1)} \prod_{s=1}^{n-1} A^{(w_s)}\right) A^{(w_n)} = \boldsymbol{\pi}^{(w_1)} \prod_{s=1}^{n} A^{(w_s)}.$$

Using the principle of mathematical induction, the i.p.d. of claim risk levels for any window w_k, where $k > 1$, is $\boldsymbol{\pi}^{(w_k)} = \boldsymbol{\pi}^{(w_1)} \prod_{s=1}^{k-1} A^{(s)}$.

□

Since hidden states are assumed to represent nonstationary claim risk levels, both t.p.m. and i.p.d. are considered to be nonstationary. However, the emission distribution $b_i(y)$ is not required to be nonstationary. This is based on a fundamental assumption of an HMM, as the probability of an observation at time t solely depends on a hidden state at the same time. Mathematically, the process used during the modeling is expressed as follows:

$$P\left(Y_t = y_t^{(w_k)} | X_t = i\right), t \in w_k. \tag{7}$$

Following this assumption of emission distribution $b_i\left(y_t^{(w_k)}\right)$ is constant across all windows w_k. This is in line with the assumed characteristics of the CBDI claim data, where the claim frequencies were driven by a specific risk level in each frequency class, irrespective of when those conditions occurred.

Based on Eqs (4) and (7), each row i of the emission distribution $b_i\left(y_t^{(w_k)}\right)$ represents the pmf of the claim frequency, which follows a Poisson-Lindley distribution with θ_i. Relating to the modeling, the pmf is mathematically defined in Eq (8):

$$b_i\left(y_t^{(w_k)}\right) = P\left(Y = y_t^{(w_k)} | X_t = i\right) = \frac{\theta_i^2\left(\theta_i + y_t^{(w_k)} + 2\right)}{(1+\theta_i)^{y_t^{(w_k)}+3}}, \tag{8}$$

where $b_i\left(y_t^{(w_k)}\right)$ describes how the i^{th} claim risk level generated claim frequency of value $y_t^{(w_k)}$.

According to the assumptions and Eqs (4)–(8), the parameters of the PL-HMM with nonstationary transition are $A^{(w_k)}$, $\boldsymbol{\pi}^{(w_k)}$, and $\boldsymbol{\theta}$. These parameters are estimated using the Baum-Welch algorithm. Before executing the algorithm, an initialization step is required for the parameters $A^{(w_k)}$, $\boldsymbol{\pi}^{(w_k)}$, and $\boldsymbol{\theta}$. The initial values of these parameters are determined using a percentile-based method, depending on the number of assumed claim risk levels. During this study, the number of claim risk levels is assumed to be $m = 2, 3, 4,$ and 5.

3.2. Parameters estimation

Next, the parameters of PL-HMM $\boldsymbol{\pi}^{(w_k)}, A^{(w_k)},$ and $\boldsymbol{\theta}$ are determined, thereby maximizing the likelihood of a given sequence of observations such that $P\left(\boldsymbol{y}^{(w_k)} \middle| \widehat{\boldsymbol{\pi}}^{(w_k)}, \hat{A}^{(w_k)}, \widehat{\boldsymbol{\theta}}\right) \geq P\left(\boldsymbol{y}^{(w_k)} \middle| \boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right)$, thus leading to a new PL-HMM $\widehat{\boldsymbol{\pi}}^{(w_k)}, \hat{A}^{(w_k)}, \widehat{\boldsymbol{\theta}}$. For the window w_k, the observed (window) ln-likelihood is as follows:

$$\mathcal{L}^{(w_k)}\left(\boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right) = \ln P\left(\boldsymbol{y}^{(w_k)} \middle| \boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right) = \ln \sum_{\boldsymbol{x}} P\left(\boldsymbol{x}, \boldsymbol{y}^{(w_k)} \middle| \boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right).$$

The observed likelihood is generally inconvenient to maximize directly [12]. Therefore, through the complete ln-likelihood formulation, the Generalized Expectation-Maximization (GEM) algorithm is used to estimate the parameters of the HMM as follows:

$$\mathcal{L}_C^{(w_k)}\left(\boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right) = \ln P\left(\boldsymbol{x}, \boldsymbol{y}^{(w_k)} \middle| \boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right)$$
$$= \sum_{i=1}^{m} I_1^{(w_k)}(i) \ln \pi_i^{(w_k)} + \sum_{t=1}^{T_k-1} \sum_{i=1}^{m} \sum_{j=1}^{m} I_t^{(w_k)}(i,j) \ln a_{ij}^{(w_k)} + \sum_{t=1}^{T_k} \sum_{i=1}^{m} I_t^{(w_k)}(i) \ln b_i(y_t^{(w_k)}),$$

where

$$I_t^{(w_k)}(i) = 1 \leftrightarrow X_t = i, t = 1,2, \dots, T_k,$$

$$I_t^{(w_k)}(i,j) = 1 \leftrightarrow X_t = i \text{ and } X_{t+1} = j, \ t = 1,2, \dots, T_k - 1.$$

E-Step. In the E-Step, given the observations of frequency claim and the current parameter estimates, the missing information about the claim risk level (hidden state) sequence is replaced by its conditional expectation. In other words, the unobserved indicator variables in the complete ln-likelihood are substituted with their posterior expectations, which are represented by $\gamma_t^{(w_k)}(i)$ and $\xi_t^{(w_k)}(i,j)$. This yields the following auxiliary function (expected complete ln-likelihood):

$$\sum_{i=1}^{m} \gamma_1^{(w_k)}(i) \ln \pi_i^{(w_k)} + \sum_{t=1}^{T_k-1} \sum_{i=1}^{m} \sum_{j=1}^{m} \xi_t^{(w_k)}(i,j) \ln a_{ij}^{(w_k)} + \sum_{t=1}^{T_k} \sum_{i=1}^{m} \gamma_t^{(w_k)}(i) \ln b_i(y_t^{(w_k)}). \tag{9}$$

Both $\gamma_t^{(w_k)}(i)$ and $\xi_t^{(w_k)}(i,j)$ are efficiently computed using the forward-backward algorithm. Let the forward and backward probabilities for window w_k be $\alpha_t^{(w_k)}(i)$ and $\beta_t^{(w_k)}(i)$, respectively:

$$\alpha_1^{(w_k)}(i) = \pi_i^{(w_k)} b_i\left(y_1^{(w_k)}\right),$$

$$\alpha_t^{(w_k)}(j) = \left(\sum_{i=1}^{m} \alpha_{t-1}^{(w_k)}(i) \cdot a_{ij}^{(w_k)}\right) b_j\left(y_t^{(w_k)}\right), \qquad t = 2,3, \dots, T_k.$$

$$\beta_{T_k}^{(w_k)}(i) = 1,$$

$$\beta_t^{(w_k)}(i) = \sum_{j=1}^{m} a_{ij}^{(w_k)} \cdot b_j\left(y_{t+1}^{(w_k)}\right) \cdot \beta_{t+1}^{(w_k)}(j), \qquad t = T_k - 1, T_k - 2, \dots, 1.$$

Here, the transition probability is computed from the claim risk levels i to j at time t as follows:

$$\xi_t^{(w_k)}(i,j) = \frac{\alpha_t^{(w_k)}(i) \cdot a_{ij}^{(w_k)} \cdot b_j\left(y_{t+1}^{(w_k)}\right) \cdot \beta_{t+1}^{(w_k)}(j)}{\sum_{i\prime}^{m} \sum_{j\prime}^{m} \alpha_t^{(w_k)}(i\prime) \cdot a_{i\prime j\prime}^{(w_k)} \cdot b_{j\prime}\left(y_{t+1}^{(w_k)}\right) \cdot \beta_{t+1}^{(w_k)}(j\prime)}. \tag{10}$$

The probability of being in the claim risk level i at time t is as follows:

$$\gamma_t^{(w_k)}(i) = \frac{\alpha_t^{(w_k)}(i)\cdot\beta_t^{(w_k)}(i)}{\sum_{k=1}^{m}\alpha_t^{(w_k)}(k)\cdot\beta_t^{(w_k)}(k)}. \tag{11}$$

M-Step. Subsequently, the GEM M-Step increases Eq (9) with respects to the parameters $\boldsymbol{\pi}^{(w_k)}$, $A^{(w_k)}$, and dan $\boldsymbol{\theta}$, thus yielding closed-form updates for the i.p.d. for the first window $\boldsymbol{\pi}^{(w_1)}$ as follows:

$$\pi_i^{(w_1)new} = \gamma_1^{(w_1)}(i), \tag{12}$$

For $k > 1$, the i.p.d. $\boldsymbol{\pi}^{(w_k)}$ is not updated independently, but instead is propagated from $\boldsymbol{\pi}^{(w_1)}$ through the sequence of the estimated t.p.m. according to Proposition 1. The t.p.m. $A^{(w_k)new} = \left[a_{ij}^{(w_k)new}\right]$ is updated as follows:

$$a_{ij}^{(w_k)new} = \frac{\sum_{t=1}^{T_k-1}\xi_t^{(w_k)}(i,j)}{\sum_{t=1}^{T_k-1}\sum_{j'=1}^{m}\xi_t^{(w_k)}(i,j')}. \tag{13}$$

For the emission distribution $b_i(y)$ is assumed stationary, thus signifying θ_i is the same across all windows, though it is re-estimated at each iteration. In the GEM framework, the update of θ_i need not maximize the auxiliary function [12]; it is sufficient to choose $\theta_i^{new} > 0$ such that each GEM iteration monotonically increases the observed (window) ln-likelihood. The formula used to update θ_i

$$\theta_i^{new} = \frac{-(\bar{y}_i-1)\pm\sqrt{(\bar{y}_i-1)^2+8\bar{y}_i}}{2\bar{y}_i}, \tag{14}$$

with $\bar{y}_i = \frac{\sum_k\sum_{t=1}^{T_k}\gamma_t^{(w_k)}(i)\cdot y_t^{(w_k)}}{\sum_k\sum_{t=1}^{T_k}\gamma_t^{(w_k)}(i)}$.

The total ln-likelihood is used to determine the convergence of the GEM algorithm. Specifically, the total ln-likelihood at iteration r is defined as follows:

$$L_r = \sum_{k=1}^{W}\mathcal{L}^{(w_k)}\left(\boldsymbol{\pi}^{(w_k)}, A^{(w_k)}, \boldsymbol{\theta}\right). \tag{15}$$

Equation (15) be evaluated using either the forward or the backward probabilities:

$$L_r = \sum_{k=1}^{W}\sum_{i=1}^{m}\alpha_{T_k}^{(w_k)}(i). \tag{16}$$

$$L_r = \sum_{k=1}^{W}\sum_{i=1}^{m}\pi_i^{(w_k)}\cdot b_i(y_1^{(w_k)})\cdot\beta_1^{(w_k)}(i), \text{ respectively.} \tag{17}$$

During the simulation, the iteration stops when $\varepsilon_r = |L_{r+1} - L_r| < \varepsilon$ (e.g., 10^{-6} to 10^{-4}).

3.3. Simulation study on claims frequency of community-based disaster insurance

In this study, the model is first applied for $m = 2$ in the simulation. The claim frequency data is arranged from smallest to largest; then, the risk boundaries between claim risks are determined using the percentile rule. In addition, the boundary percentiles for $m = 2$ and $T = 156$ are used. A snippet of the classification results for $m = 2$ are shown in Table 2.

Table 2. Claim frequency classification for 2 risk levels.

Month-Year	Claim Frequency	Risk Level	Claim Frequency at Risk Level 1	Claim Frequency at Risk Level 2
Jan-2010	2296	2	-	2296
Feb-2010	2001	2	-	2001
Mar-2010	2486	2	-	2486
Apr-2010	770	2	-	770
May-2010	374	1	374	-
⋮	⋮	⋮	⋮	⋮
Agst-2022	2305	2		2305
Sept-2022	265	1	265	-
Oct-2022	388	1	388	-
Nov-2022	57906	2	-	57906
Dec-2022	218	1	218	-

The 36-month horizon (w) is chosen to ensure a credible basis to estimate claim frequency risk. The three-year window balances stability and responsiveness, thereby covering at least three seasonal cycles and providing sufficient data volume for credible risk measurement. The 6-month step (s) reflects a feasible premium review period, consistent with the semi-annual adjustment practice in CBDI. The number of windows W formed is calculated as follows:

$$W = \frac{T-w}{s} + 1 = \frac{156-36}{6} + 1 = 21.$$

The outcome is 21 windows in total, thus leading to a set of nonstationary t.p.m. $A^{(w_1)}, A^{(w_2)}, \ldots, A^{(w_{21})}$. The constructed model is a nonstationary HMM formulated using the sliding window method. Moreover, a mathematical simulation is only shown for the first window (January 2010–December 2012). Simulations for the remaining windows follow the same procedure but are computed using Google Colab for efficiency.

The simulation starts with parameter initialization based on the classification in Table 2. The first parameter initialized during the simulation is the t.p.m.; a sample of the calculated t.p.m. for $m = 2$ is shown in Table 3.

Table 3. Transition probability matrix of claim risk levels for window w_k.

k	$A^{(w_k)}$	k	$A^{(w_k)}$
1	$\begin{bmatrix} 0.6 & 0.4 \\ 0.35 & 0.65 \end{bmatrix}$	11	$\begin{bmatrix} 0.833 & 0.167 \\ 0.364 & 0.636 \end{bmatrix}$
2	$\begin{bmatrix} 0.625 & 0.375 \\ 0.263 & 0.737 \end{bmatrix}$	12	$\begin{bmatrix} 0.81 & 0.19 \\ 0.286 & 0.714 \end{bmatrix}$
3	$\begin{bmatrix} 0.538 & 0.462 \\ 0.227 & 0.773 \end{bmatrix}$	13	$\begin{bmatrix} 0.818 & 0.182 \\ 0.231 & 0.769 \end{bmatrix}$
⋮	⋮	⋮	⋮
08	$\begin{bmatrix} 0.773 & 0.227 \\ 0.462 & 0.538 \end{bmatrix}$	19	$\begin{bmatrix} 0.765 & 0.235 \\ 0.222 & 0.778 \end{bmatrix}$
09	$\begin{bmatrix} 0.762 & 0.238 \\ 0.429 & 0.571 \end{bmatrix}$	20	$\begin{bmatrix} 0.706 & 0.294 \\ 0.222 & 0.778 \end{bmatrix}$
10	$\begin{bmatrix} 0.792 & 0.208 \\ 0.455 & 0.545 \end{bmatrix}$	21	$\begin{bmatrix} 0.688 & 0.312 \\ 0.316 & 0.684 \end{bmatrix}$

After the t.p.m. is determined for each window $w_k; k = 1,2, \dots, 21$, the next step is to calculate the i.p.d. for each risk level $i = 1,2$ in window 1.

$$\boldsymbol{\pi}^{(w_1)} = [0{,}444444 \quad 0{,}555556].$$

The calculations for windows $w_k; k = 2, 3, \dots, 21$, are performed using Proposition 1. Following, the simulation, a sample of the results are shown in Table 4.

Table 4. Initial probability distribution of risk levels for window w_k.

k	$\boldsymbol{\pi}^{(w_k)}$		k	$\boldsymbol{\pi}^{(w_k)}$	
1	[0.444444	0.555556]	11	[0.667447	0.332553]
2	[0.461111	0.538889]	12	[0.677032	0.322968]
3	[0.429922	0.570078]	13	[0.640765	0.359235]
⋮	⋮		⋮	⋮	
8	[0.459233	0.540767]	19	[0.518077	0.481923]
9	[0.604822	0.395178]	20	[0.503316	0.496684]
10	[0.630406	0.369594]	21	[0.465605	0.534395]

The last parameter initialization is the emission distribution for the parameter $\boldsymbol{\theta}$ at each risk level, which is estimated using the moment method. Based on the estimates θ_1 and θ_2, the emission distribution for risk levels 1 and 2 are given by the following:

$$b_1\left(y_t^{(w_k)}\right) = \frac{(\theta_1)^2\left(\theta_1 + y_t^{(w_k)} + 2\right)}{(1+\theta_1)^{y_t^{(w_k)}+3}} = \frac{(0.012133406865)^2\left(0.012133406865 + y_t^{(w_k)} + 2\right)}{(1+0.012133406865)^{y_t^{(w_k)}+3}},$$

$$b_2\left(y_t^{(w_k)}\right) = \frac{(\theta_2)^2\left(\theta_2 + y_t^{(w_k)} + 2\right)}{(1+\theta_2)^{y_t^{(w_k)}+3}}$$

$$= \frac{(0.00093779911)^2\left(0.00093779911 + y_t^{(w_k)} + 2\right)}{(1+0.00093779911)^{y_t^{(w_k)}+3}}, \text{respectively.}$$

After the initializing parameters $\boldsymbol{\pi}^{(w_k)}$, $A^{(w_k)}$, and $\boldsymbol{\theta}$ are established, the GEM algorithm is used to re-estimate the model parameters. Each iteration begins with the E-Step, where the expected complete ln-likelihood is computed using Eqs (10) and (11). This is followed by the M-Step, where the parameters are updated using Eqs (12)–(14). The algorithm is repeated until the total observed (window) ln-likelihood reaches convergence. Here, a convergence table is provided to display the evolution of the ln-likelihood and updated parameter from the last iteration.

The GEM algorithm is stopped at the 29th iteration when it reaches the convergence criterion $\varepsilon_r < \varepsilon$, with $\varepsilon = 10^{-5}$. The results in Table 5 demonstrate that Eqs (12)–(14) is numerically verified, thereby showing an increase in the observed (window) log-likelihood. The parameter estimates obtained at this final iteration are taken as the optimal values for the 2-state PL-HMM, including the t.p.m. and the i.p.d. These results are summarized in Tables 6 and 7, respectively.

Table 5. Iteration of GEM algorithm for PL-HMM with $m = 2$.

r	L_r	ε_r
0	−5832.423986	−
1	−5795.356346	37.067640
2	−5790.496406	4.859940
3	−5787.864127	2.632279
4	−5786.448840	1.415287
5	−5785.740907	0.707933
⋮	⋮	⋮
25	−5785.167434	0.000031
26	−5785.167456	0.000021
27	−5785.167471	0.000015
28	−5785.167481	0.000010
29	−5785.167487	0.000007

Table 6. Final transition probability matrix of claim risk levels for window w_k (after GEM algorithm).

k	$A^{(w_k)}$	k	$A^{(w_k)}$
1	$\begin{bmatrix}0.621740214534 & 0.378259785466\\ 0.223031476706 & 0.776968523294\end{bmatrix}$	11	$\begin{bmatrix}0.826721172052 & 0.173278827948\\ 0.367741957541 & 0.632258042459\end{bmatrix}$
2	$\begin{bmatrix}0.614795414826 & 0.385204585174\\ 0.195001154917 & 0.804998845083\end{bmatrix}$	12	$\begin{bmatrix}0.795937485753 & 0.204062514247\\ 0.275248630599 & 0.724751369401\end{bmatrix}$
3	$\begin{bmatrix}0.494574273985 & 0.505425726015\\ 0.186188621772 & 0.813811378228\end{bmatrix}$	13	$\begin{bmatrix}0.803601341262 & 0.196398658738\\ 0.215313193823 & 0.784686806177\end{bmatrix}$
⋮	⋮	⋮	⋮
08	$\begin{bmatrix}0.804031345116 & 0.195968654884\\ 0.370439742136 & 0.629560257864\end{bmatrix}$	19	$\begin{bmatrix}0.740153332663 & 0.259846667337\\ 0.145228512893 & 0.854771487107\end{bmatrix}$
09	$\begin{bmatrix}0.794788605377 & 0.205211394623\\ 0.330304403479 & 0.669695596521\end{bmatrix}$	20	$\begin{bmatrix}0.663504887086 & 0.336495112914\\ 0.109592247714 & 0.890407752286\end{bmatrix}$
10	$\begin{bmatrix}0.813530043168 & 0.186469956832\\ 0.369282720553 & 0.630717279447\end{bmatrix}$	21	$\begin{bmatrix}0.652212317886 & 0.347787682114\\ 0.123253902294 & 0.876746097706\end{bmatrix}$

Table 7. Final initial probability distribution of risk levels for window w_k (after GEM algorithm).

k	$\boldsymbol{\pi}^{(w_k)}$	k	$\boldsymbol{\pi}^{(w_k)}$
1	[0 1]	11	[0.58399483136 0.41600516864]
2	[0 1]	12	[0.628720860805 0.371279139195]
3	[0.223031476706 0.776968523294]	13	[0.65631176438 0.34368823562]
⋮	⋮	⋮	⋮
8	[0.332404376576 0.667595623424]	19	[0.456693808544 0.543306191456]
9	[0.405305158267 0.594694841733]	20	[0.412744879627 0.587255120373]
10	[0.546176655405 0.453823344595]	21	[0.390780686016 0.609219313984]

Additionally, a visualization of both the initial t.p.m. and the re-estimated t.p.m. at the 29th iteration of the GEM algorithm are shown in Figure 2.

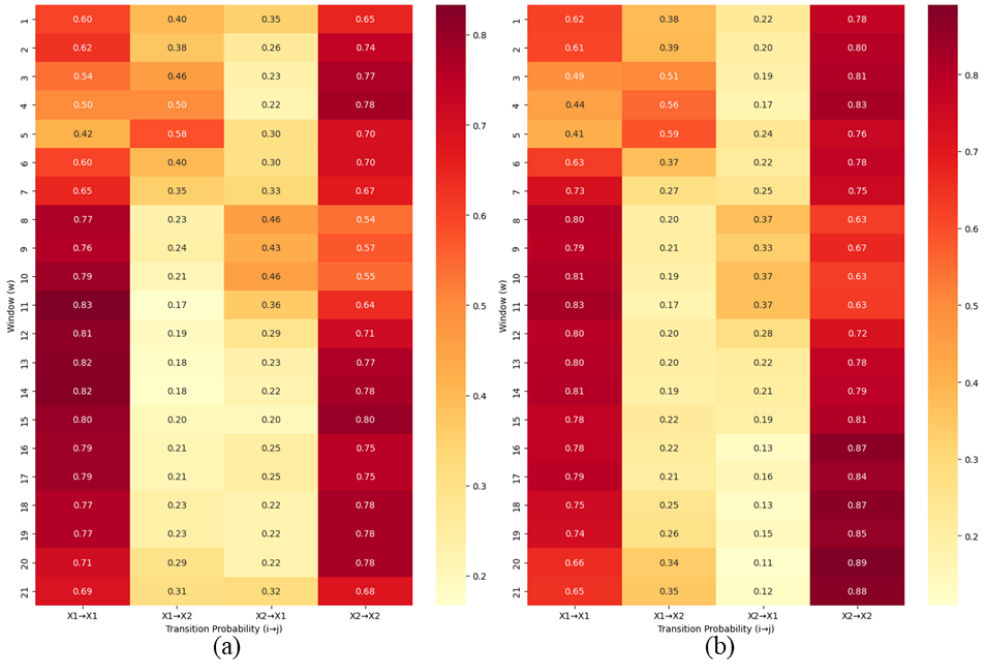

Figure 2. Heatmap transition probability matrices: (a) initial estimates and (b) after 29th iteration of GEM algorithm.

During the simulation, the initialization of t.p.m. in window 1 reflects the initial assumption as follows. When the system is in risk level 1, the probability of remaining in the level is relatively high at 0.6, while transitioning to risk level 2 is 0.4. The probability of moving from risk level 2 to 1 is lower at 0.35, compared to staying in risk level 2, which is 0.65. Based on the re-estimation results, the transition pattern between the risk levels shows sharpened probabilities. Following the discussion, the probability of remaining in risk level 1 increases from 0.6 to 0.62. This signifies that once a period of low claim frequency occurs, the condition tends to persist. The probability of staying in risk level 2 also increases to 0.78, thus suggesting a strong persistence of high-risk periods once the outcome occurred. Meanwhile, the probabilities of transitioning between the level 1 and level 2 states decreases, thus implying that the transition between states is not frequent. The t.p.m. obtained from the GEM algorithm offers an interpretation that the risk associated with the claim frequency shows either strong memory or persistence. Relating to the simulation, claims tend to remain in the same pattern or state for a certain period, which follows the concept of a nonstationary HMM, where transition probabilities reflect real-world disaster risk cycles, such as calm periods or periods of high disaster intensity.

Furthermore, the re-estimation initial probability distribution shows changes (see Tables 4 and 7) in the starting probabilities for each risk level across all windows. In particular, the initial probability for risk level 2 increases in several periods, thus indicating a higher probability of beginning a period with a high claim frequency, which reflects an initial condition already vulnerable to disasters. This result is consistent with the *West Java Provincial Disaster Risk Assessment 2022–2026* [13], which identified the province as a high-risk area for various hazards, particularly floods and landslides, that often cause large-scale housing damage (directly related to the CBDI claim frequency). Additionally, this change indicates that the GEM algorithm provides parameters with a maximum fit to historical claim data, thereby offering a more realistic representation of the initial risk of claims in each window.

To the last parameter, namely the emission distribution from the Poisson-Lindley parameter $\boldsymbol{\theta}$, the re-estimation results from the GEM algorithm indicate that the value of $\boldsymbol{\theta}$ increase but remain

relatively small (approaching zero). This value $\boldsymbol{\theta}$ implies a higher expected claim frequency while also reflecting substantial over-dispersion in CBDI-related claims. This finding suggests the formulated model successfully captures the over-dispersion phenomenon of claim frequencies, which are consistent with the reality of disasters in West Java that frequently cause large-scale housing damage [13], thereby generating high claim frequencies. The re-estimated emission distributions are presented below:

$$b_1\left(y_t^{(w_k)}\right) = \frac{(\theta_1^{new})^2\left(\theta_1^{new}+y_t^{(w_k)}+2\right)}{(1+\theta_1^{new})^{y_t^{(w_k)}+3}} = \frac{(0.015863225074)^2\left(0.015863225074+y_t^{(w_k)}+2\right)}{(1+0.015863225074)^{y_t^{(w_k)}+3}},$$

$$b_2\left(y_t^{(w_k)}\right) = \frac{(\theta_2^{new})^2\left(\theta_2^{new}+y_t^{(w_k)}+2\right)}{(1+\theta_2^{new})^{y_t^{(w_k)}+3}}$$
$$= \frac{(0.001419878595)^2\left(0.001419878595+y_t^{(w_k)}+2\right)}{(1+0.001419878595)^{y_t^{(w_k)}+3}}.$$

4. Discussion

The results of this study are in line with the findings by [14], which highlighted the trend in insurance claim modeling literature that positions HMM as a valid framework to capture the hidden dynamics influencing claims. Our findings show that a nonstationary PL-HMM can fix the issue of overdispersion, which cannot be adequately explained by a nonstationary Poisson-HMM. This is shown through a performance comparison between the nonstationary PL-HMM and the nonstationary Poisson-HMM across different numbers of hidden states (m), where the evaluation was based on the Bayesian Information Criterion (BIC).

Table 8 shows that the BIC values substantially decrease (in nonstationary Poisson HMM) as m increases, but remain at an extremely large scale (hundreds of thousands), making them less efficient in representing highly dispersed claim data. In contrast, nonstationary PL-HMM yields much smaller BIC values (on the order of tens of thousands), indicating that this model is more suitable for disaster-related claim data characterized by high variability. Moreover, a nonstationary PL-HMM with $m = 2$ produces the lowest BIC value, which indicates that this model is the most appropriate choice to represent CBDI claims.

Table 8. Comparison of nonstationary PL-HMM and Poisson-HMM.

Model	m	ln-likelihood	k	BIC
Nonstationary	2	−263,789.589706	45	527,806.42
Poisson HMM	3	−181,420.919543	131	363,503.37
	4	−123,643.314824	259	248,594.54
	5	−93,646.017518	429	189,458.42
Nonstationary PL-	2	−5785.167487	45	11,797.58
HMM	3	−5667.926737	131	11,997.38
	4	−5578.651092	259	12,465.21
	5	−5526.678839	429	13,219.75

After establishing the PL-HMM with $m = 2$ as the main model, the model's sensitivity to the choice of sliding window length and step size is evaluated. This analysis is important to ensure that

the model performance remains robust despite variations in the data horizon and the shifting interval of the window. The focus of this sensitivity analysis is placed on the t.p.m. Specifically, the sliding window is altered to 12 months and 24 months while keeping the step size fixed at 6 months. The results are presented in the form of heatmaps of the initial t.p.m. and the t.p.m. obtained from the last iteration of the GEM algorithm (see Figures 3 and 4).

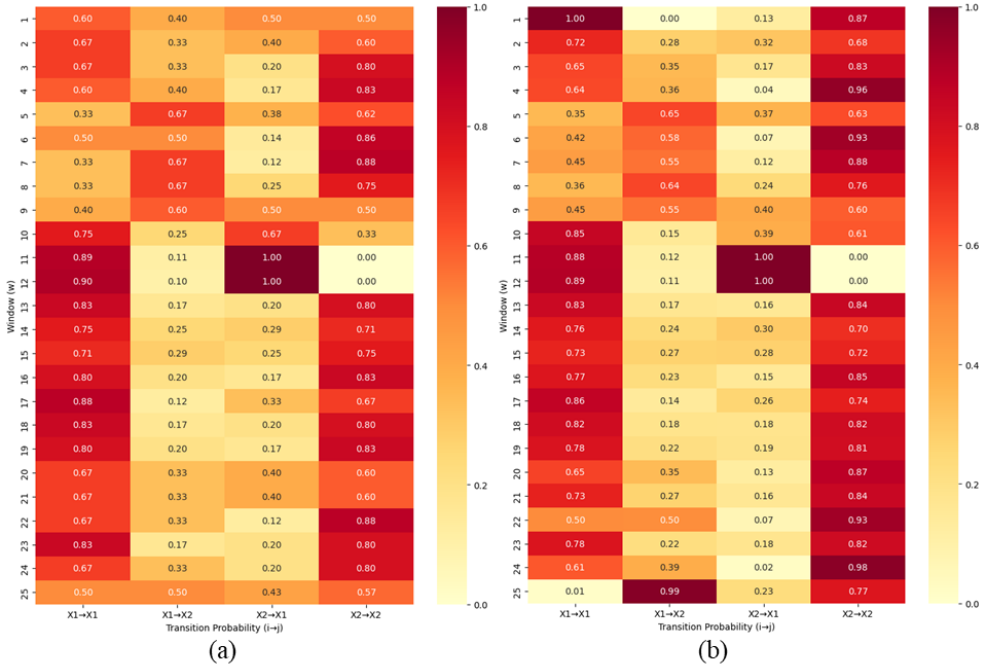

Figure 3. Heatmap transition probability matrices: (a) initial estimates and (b) after last iteration of GEM algorithm (sliding window $w = 12$).

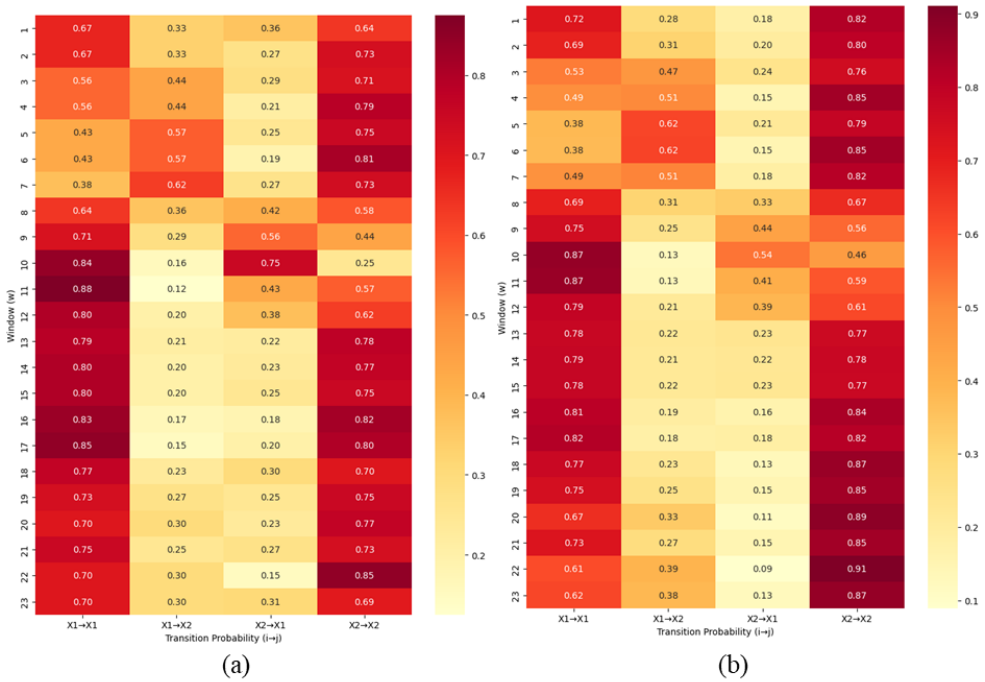

Figure 4. Heatmap transition probability matrices: (a) initial estimates and (b) after last iteration of GEM algorithm (sliding window $w = 24$).

The results obtained from altering the sliding window length to 12 or 24 months still demonstrate the same transition patterns as the 36-month specification, with the only difference of the number of resulting windows. This indicates that the choice of window length is more related to the risk measurement horizon (e.g., whether risk is evaluated over 12, 24, or 36 months), while the overall findings remain consistent. In all cases, the transition pattern between risk levels show sharpened probabilities. Specifically, the probability of remaining in risk level 1 increases, thus signifying that once a period of low claim frequency occurs, the condition tends to persist. Similarly, the probability of staying in risk level 2 also increases, thus suggesting a strong persistence of high-risk periods once they occur. Meanwhile, the probabilities of transitioning between level 1 and level 2 decreases, thus implying that such transitions are relatively infrequent. The findings consistently indicate persistence in claim-risk states, as reflected by high staying probabilities.

This study demonstrates the advantages of nonstationary Poisson HMM and PL-HMM to model CBDI claim frequencies, particularly in addressing overdispersion and capturing hidden risk dynamics. However, further research directions remain open to strengthen and extend these findings. First, future work could focus on exploring the implications of expected claim estimation under both nonstationary Poisson HMM and PL-HMM frameworks. A comparative calculation of expected values across these models would provide deeper insights into how the model choice influences the premium setting and risk evaluation in CBDI schemes. Second, the sensitivity analysis conducted in this study was limited to variations in sliding window length, while the step size was held constant. In a future study, we can extend the analysis to different step sizes in combination with varying window lengths. Such an approach would allow for a more comprehensive understanding of how temporal aggregation and updating frequency affect the parameter stability, transition patterns, and overall model performance.

5. Conclusions

We proposed a nonstationary PL-HMM as a novel approach to model the frequency of CBDI claims. The model allows for the frequency claim to exhibit both serial dependence and overdispersion through an underlying hidden risk state. We modified the Poisson HMM framework by incorporating the Poisson-Lindley distribution for emissions and introduced nonstationary transition probabilities using a sliding-window mechanism. Parameter estimation of the model was carried out using the GEM algorithm. To perform the algorithm, we proved one proposition (to ensure the nonstationary assumption) and one theorem to guarantee the increase of complete ln-likelihood of proposed model. To evaluate the performance of the proposed model, we conducted an empirical simulation on monthly claim frequency data in West Java Province. Nonstationary Poisson HMM and nonstationary PL-HMM with varying numbers of $m = 2, 3, 4,$ and 5 were compared using BIC. Among the competing models, the two-state nonstationary PL-HMM achieved the lowest BIC, thereby being selected as the most suitable representation of the CBDI claim frequency. A sensitivity analysis was further performed by altering the sliding-window horizon (12, 24, and 36 months), and confirmed that the estimated transition probability matrices preserved persistence patterns of the claim risk states, while differences across window lengths primarily reflected alternative horizons of risk measurement. The main advantage of the nonstationary PL-HMM lies in its ability to capture the time-varying risk of claim occurrences rather than assuming a uniform claim risk across an entire period. This is in line with the characteristics of disaster-related claims that typically alternate between periods of low and high frequencies. Overall, the model is still able to realistically represent the claim risk, thereby reflecting

the empirical fact that high-claim periods occur more frequently than low-claim periods. In addition, the model is flexible and can be applied to other regions that also exhibit overdispersion—whether mild or severe—in claim frequencies. Thus, this model provides an adaptive and applicable framework for various contexts of CBDI in disaster-prone areas.

Author contributions

Hilda Azkiyah Surya: Conceptualization, methodology, original draft preparation, writing, review and editing, visualization; Sukono: Conceptualization, validation, formal analysis, supervision, project administration, funding acquisition; Herlina Napitupulu: Software, validation, writing, review and editing, supervision; Noriszura Ismail: Validation, writing, review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

OpenAI's ChatGPT (version GPT-4) was used to assist in the language polishing of the manuscript, generation of the English translation of the abstract and discussion sections, and paraphrasing for academic tone.

Acknowledgments

The authors are grateful to Universitas Padjadjaran, who provided this Study fund by giving Padjadjaran Doctoral Program Scloarship (BPDP) with grant number 1792/UN6.3.1/PT.00/2024 and the Article Processing Charge (APC).

Conflict of interest

The authors declare no conflict of interest.

References

1. D. Lefutso, A. Ogundeji, G. Danso-Abbeam, Y. Nyam, Low-income households' willingness to pay for flood risk insurance in South Africa, *Progress in Disaster Science*, **25** (2025), 100403. https://doi.org/10.1016/j.pdisas.2024.100403
2. I. Azis, B. Setiawaty, I. Gusti Putu Purnaba, Modeling of vehicle insurance claim using exponential hidden Markov, *International Journal of Pure and Applied Mathematics*, **118** (2018), 309–320. https://doi.org/10.12732/ijpam.v118i2.15
3. Z. Oflaz, C. Yozgatligil, A. Selcuk-Kestel, Aggregate claim estimation using bivariate hidden markov model, *ASTIN Bulletin: The Journal of the IAA*, **49** (2019), 189–215. https://doi.org/10.1017/asb.2018.29
4. R. Elliott, T. Siu, Hedging options in a hidden Markov-switching local-volatility model via stochastic flows and a Monte-Carlo method, *J. Futures Markets*, **43** (2023), 925–950. https://doi.org/10.1002/fut.22422

5. L. Slater, B. Anderson, M. Buechel, S. Dadson, S. Han, S. Harrigan, et al., Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management, *Hydrol. Earth Syst. Sci.*, **25** (2021), 3897–3935. https://doi.org/10.5194/hess-25-3897-2021
6. A. Tsoi, S. Zhang, M. Hagenbuchner, Pattern discovery on Australian medical claim data-a systematic approach, *IEEE Trans. Knowl. Data Eng.*, **17** (2005), 1420–1435. https://doi.org/10.1109/TKDE.2005.168
7. V. Koerniawan, N. Sunusi, R. Raupong, Estimasi parameter model Poisson hidden Markov pada data Banyaknya Kedatangan Klaim Asuransi Jiwa, *ESTIMASI: Journal of Statistics and Its Application*, **1** (2020), 65–73. https://doi.org/10.20956/ejsa.v1i2.9302
8. W. Al-Nuaami, A. Heydari, H. Khamnei, The Poisson-Lindley distribution: some characteristics, with its application to SPC, *Mathematics*, **11** (2023), 2428. https://doi.org/10.3390/math11112428
9. N. Azizah, S. Astutik, Nurjannah, Two-state Poisson hidden Markov models for analysis of seismicity activity rates in west Nusa Tenggara, *IOP Conf. Ser.: Mater. Sci. Eng.*, **546** (2019), 052015. https://doi.org/10.1088/1757-899X/546/5/052015
10. K. Orfanogiannaki, D. Karlis, G. Papadopoulos, Identification of temporal patterns in the seismicity of Sumatra using Poisson Hidden Markov models, *Research in Geophysics*, **4** (2014), 4969. https://doi.org/10.4081/rg.2014.4969
11. R. Paroli, G. Redaelli, L. Spezia, Poisson hidden Markov models for time series of overdispersed insurance counts, *Proceedings of XXXI Astin Colloqium*, 2000, 461–474.
12. S. Chatterjee, O. Romero, S. Pequito, Analysis of a generalised expectation-maximisation algorithm for Gaussian mixture models: a control systems perspective, *Int. J. Control*, **95** (2022), 2734–2742. https://doi.org/10.1080/00207179.2021.1931964
13. *Kedeputian bidang sistem dan strategidirektorat pemetaan dan evaluasi risiko bencana, Kajian rsiko bncana nsional povinsi Jawa Barat 2022–2026*, Badan Nasional Penanggulangan Bencana, 2021. Available from: https://inarisk.bnpb.go.id/pdf/Jawa%20Barat/Dokumen%20KRB%20Prov.%20Jawa%20Barat_final%20draft.pdf.
14. H. Surya, Sukono, H. Napitupulu, N. Ismail, A systematic literature review of insurance claims risk measurement using the hidden Markov model, *Risks*, **12** (2024), 196. https://doi.org/10.3390/risks12110169