Mathematics and Computers in Simulation 223 (2024) 543–564

Contents lists available at ScienceDirect

Mathematics and Computers in Simulation

journal homepage: www.elsevier.com/locate/matcom

Original articles

Time series clustering based on latent volatility mixture modeling with applications in finance

F. Setoudehtazangi [a], T. Manouchehri [b], A.R. Nematollahi [b], M. Caporin [a,*]

[a] Dipartimento di Science Statistiche, Universit'a di Padova, Italy
[b] Department of Statistics, Shiraz University, Shiraz, Iran

ARTICLE INFO

Keywords:
EM algorithm
MoARCH models
Normal mean–variance mixture distributions
Penalized likelihood

ABSTRACT

Modeling financial time series data poses a significant challenge in the realm of time series analysis. The Autoregressive Conditional Heteroskedasticity (ARCH) model stands out as a potent tool for capturing time-varying volatility and heteroskedasticity in financial data. However, conventional ARCH models display sensitivity to departures from normality, leading to the development of extensions employing more flexible distributions. In this context, we propose a robust enhancement to the mixture of ARCH (MoARCH) model by integrating normal mean–variance mixture (NMVM) distributions to model component errors. The stochastic representation of the proposed model allows for a straightforward implementation of an Expectation Conditional Maximization Either (ECME) algorithm for obtaining Maximum Penalized Likelihood estimates (MPL). To thoroughly evaluate the model, we conduct four simulation studies to explore finite-sample properties, assess MPL estimators, scrutinize model robustness, and evaluate the accuracy of our proposal in fitting, clustering, and forecasting. Practical applications further highlight the effectiveness of our methodology, showcasing successful implementations across diverse real datasets.

1. Introduction

In the last decades, there has been increasing interest in using conditional heteroskedasticity models to analyze financial time series data characterized by extreme and/or skewed observations. Among these models, the Autoregressive Conditional Heteroskedastic (ARCH) process, introduced by Engle [15], has left an indelible mark. Its purpose is to address the conditional heteroskedasticity in time series data, issues of great significance in the realms of finance and economics. The evolution of the ARCH model has led to the emergence of various derived models, notably including the Generalized Autoregressive Conditional Heteroskedastic (GARCH) process by Bollerslev [8] and the Glosten–Jagannathan–Runkle GARCH (GJR-GARCH) model by Glosten et al. [19]. To gain a thorough understanding of the various GARCH models developed in financial econometrics over the last forty years, researchers may consult the comprehensive review by Bollerslev et al. [9]. This reference serves as a valuable resource, offering detailed insights into the diverse range of GARCH models that have influenced the field of financial econometrics.

While the GARCH model stands as one of the foremost contenders in this model class, it exhibits sensitivity to deviations from normality, in particular increased skewness and potential heavy tails. This has prompted the development of various extensions that take advantage of more flexible distributions. For further details, refer to [21,39]. In the context of our study, we tackle this sensitivity issue by employing a diverse class of distributions known as normal mean–variance mixture (NMVM) distributions. These

* Corresponding author.
E-mail addresses: fariborz.setoudehtazangi@studenti.unipd.it (F. Setoudehtazangi), t.manouchehri@shirazu.ac.ir (T. Manouchehri), ar.nematollahi@shirazu.ac.ir (A.R. Nematollahi), massimiliano.caporin@unipd.it (M. Caporin).

https://doi.org/10.1016/j.matcom.2024.04.031
Received 18 January 2024; Received in revised form 2 April 2024; Accepted 26 April 2024
Available online 6 May 2024

NMVM distributions encompass a wide range of lightly and heavily-tailed symmetric and asymmetric distributions, affording the flexibility needed to manage outliers and complex data.

Furthermore, a long-standing challenge in statistics and data science is the development of robust approaches to model and group data into economically meaningful clusters. A highly effective statistical approach in this regard involves the use of mixture models. In particular, this area boasts a substantial body of literature, as exemplified by McLachlan et al. [28] and McLachlan and Basford [27].

The exploration of time series clustering has yielded a diverse array of methodologies. Among these, finite mixtures of Markov chains have received increased attention from researchers, as evidenced by the works of Ridgeway [36] and Cadez et al. [12]. Another notable approach involves Bayesian clustering by dynamics (BCD), which has been explored by Sebastiani et al. [38] and Ramoni et al. [34], particularly in the context of time series modeled by Markov chains. Hidden Markov Models (HMMs) have also emerged as a powerful tool for clustering time series data. Rabiner et al. [33] seminal work in 1989 laid the groundwork for subsequent research, with notable contributions like [18,22] in the application of HMMs to time series clustering, and [2] especially in the realm of sequence data. Moreover, regression models, mixtures of regression models, and their extensions offer versatile frameworks for both modeling and clustering time series. Gaffney and Smyth [17] research exemplifies the potential of these approaches. Furthermore, the integration of autoregressive moving average (ARMA) and autoregressive (AR) models into mixture frameworks has been extensively explored for time series clustering, with notable studies by Xiong and Yeung [45] and Ren and Barnett [35]. Lastly, mixtures of linear Gaussian state space models have demonstrated efficacy in time series clustering, as evidenced by the work of Umatani et al. [41].

To contribute to the field of time series clustering, we introduce an innovative ARCH model, termed the mixture of ARCH (MoARCH) model which harnesses the normal mean variance mixture (NMVM) distribution. The NMVM distribution, first introduced by McNeil et al. [29], offers several key features, such as the ability to handle skewness and heavy tails. The class of NMVM distributions encompasses several significant variants, including the generalized hyperbolic (GH) distribution [4], the Birnbaum–Saunders normal mean–variance mixture (NMVBS) distribution [32], and the Lindley normal mean–variance mixture (NMVL) distribution [31] as special cases. Consequently, our mixture of Autoregressive Conditional Heteroskedasticity model based on the normal mean–variance mixture (MoARCH-NMVM) is highly adaptable and robust, effectively addressing skewness and heavy-tailedness within the MoARCH framework and we incorporate a penalized likelihood function to determine the optimal number of components and we leverage the stochastic representation of our proposed model to employ two extensions of the EM algorithm, including the ECM algorithm [30] and the ECME algorithm [25].

The structure of this paper unfolds as follows: In Section 2, we provide a concise overview of the NMVM distribution and highlight its key features. Section 3 presents the proposed NMVM-MoARCH model. Section 4 details the maximum penalized estimates (MPL) of the NMVM-ARCH(q) model using an EM-type algorithm, as well as the methods for determining the tuning parameter and the number of components. Section 5 deals with the practical aspects of estimating standard errors and selecting appropriate initial values. In Section 6, extensive simulation studies are carried out to evaluate the efficiency of the proposed approach. Subsequently, we apply the results to real-world data examples, demonstrating the practicality of our methodology. Finally, the paper concludes with a summary of the key findings in Section 7.

2. Normal mean–variance mixture distributions

The NMVM distribution belongs to a family of flexible distributions formed by scaling the mean and variance of a normal random variable with a positive scale-valued mixing random variable. In particular, a random variable Y is said to follow the normal mean–variance mixture distribution class, indicated by $Y \sim NMVM(\mu, \lambda, \sigma^2, \nu)$, if it is represented by a hierarchical mixing structure as follows:

$$Y|(W = w) \sim N(\mu + w\lambda, w\sigma^2), \tag{2.1}$$

$$W \sim h(w; \nu),$$

where $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R}$, and W is a positive mixing random variable, characterized by a probability density function (pdf) denoted as $h(\cdot; \nu)$, indexed by a possibly parameter vector ν.

Furthermore, it can be shown that Y has the following stochastic representation:

$$Y = \mu + \lambda W + \sqrt{W} Z, \tag{2.2}$$

where Z is a univariate normal random variable with mean 0 and variance σ^2. The random variable W is statistically independent of Z, see [29] for additional details.

The NMVM distribution has a PDF given by:

$$f_{NMVM}(y; \mu, \lambda, \sigma^2, \nu) = \int_0^\infty \phi(y; \mu + w\lambda, w\sigma^2) h(w; \nu) dw. \tag{2.3}$$

In this paper, we make use of the NMVM distribution family, which provides a suitable tool for modeling data with asymmetric and heavier tails compared to Gaussian distributions. In fact, by acting on W, a researcher might generate data with time varying mean and variance, easily introducing deviations from Normality.

One of the special cases of the NMVM family is the GH distribution, as specified in [29]. This subclass of NMVM is obtained by using the generalized inverse Gaussian (GIG) distribution proposed by Good [20] as the PDF of W. In this paper, we mainly use the GH distribution to model our data. For more details on the GH distribution, please refer to Appendix A.

3. NMVM-MoARCH model

The NMVM-MoARCH(q) model introduces an innovative framework for the comprehensive analysis of time series data, encompassing modeling, clustering, and forecasting. The primary goal is to achieve efficient and robust clustering of N multiple time series. This novel approach aims to categorize the N time series into an optimal number, denoted as g, of clusters represented by $(C_1, \dots, C_g)$, as determined by the components of the mixture model.

Let $y_{jt}^{(i)}$ denote the random variable representing the de-meaned asset return at time t $(t = 1, \dots, n_j)$. We consider a panel of multiple time series, denoted as $\{y_{jt}^{(i)}\}$, where $t = 1, \dots, n_j$, observed across N units $j = 1, \dots, N$. Additionally, we introduce the component-indicator random variable $Z_j^{(i)}$, which assigns time series j to one of g components or clusters C_i with a probability $P(Z_j^{(i)} = 1) = \pi^{(i)}$. Here, $\sum_{i=1}^{g} \pi^{(i)} = 1$ and $\pi^{(i)} > 0$ for $i = 1, \dots, g$ and $j = 1, \dots, N$. This integrated approach, combining modeling and clustering, offers a comprehensive framework for analyzing time series data.

Let $\boldsymbol{y}_j^{2(i)} = \left(1, y_{j(t-1)}^{2(i)}, \dots, y_{j(t-q)}^{2(i)}\right)$, $j = 1, \dots, N$ be a $(q+1)\times 1$ vector of variables and $\boldsymbol{\alpha}_i^{(i)} = \left(\alpha_{i0}^{(i)}, \alpha_{i1}^{(i)}, \dots, \alpha_{iq}^{(i)}\right)'$ be the coefficients for the conditional variance. Given $Z_j^{(i)} = 1$, in each single time series, we model the observations as follows:

Let $y_{jt}^{(i)} \equiv y_{jt} | Z_j^{(i)}$. Consequently, we have the following relationship:

$$y_{jt}^{(i)} = \sqrt{h_{jt}^{(i)}}\varepsilon_{jt}^{(i)}; \quad \varepsilon_{jt}^{(i)} \sim \text{NMVM}(0, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu}^{(i)}), \quad \text{with probability} \quad \pi^{(i)}; \quad i = 1, \dots, g, \tag{3.1}$$

where the volatility process $h_{jt}^{(i)}$ represents the conditional variance of $y_{jt}^{(i)}$ given $\mathcal{F}_{j,t-1}^{(i)}$ (where $\mathcal{F}_{j,t-1}^{(i)}$ is the information set available until time $t-1$). This conditional variance equals: $h_{jt}^{(i)} = \alpha_0^{(i)} + \alpha_1^{(i)} y_{j(t-1)}^{2(i)} + \cdots + \alpha_q^{(i)} y_{j(t-q)}^{2(i)} = \boldsymbol{\alpha}^{(i)\prime} \boldsymbol{y}_{jt}^{2(i)}$. Here, $\boldsymbol{y}_j^{2(i)}$ is a $(q+1)\times 1$ vector of variables for time series j, and $\boldsymbol{\alpha}^{(i)}$ denotes the coefficients for the conditional variance. To ensure proper identification, it is essential to set $E[h_{jt}^{(i)}] = 1$, where $\sigma^{2(i)}$ represents the unconditional variance of the innovations. This condition can be achieved by setting $\alpha_0^{(i)} = 1 - \sum_{j=1}^{q} \alpha_j^{(i)}$ coefficients.

Let $\boldsymbol{y}_j^{(i)} = \left(y_{j1}^{(i)}, \dots, y_{jn_j}^{(i)}\right)'$ come from a g-component mixture ARCH model, if the conditional density function with probability $\pi^{(i)}$ is given by:

$$f\left(y_{jt}^{(i)} | \mathcal{F}_{j,t-1}^{(i)}, \boldsymbol{\theta}^{(i)}\right) = f_{NMVM}\left(y_{jt}^{(i)}; 0, \lambda^{(i)}, \sigma^{2(i)} h_{jt}^{(i)}, \boldsymbol{\nu}^{(i)}\right), \quad i = 1, \dots, g; \; j = 1, \dots, N; \; t = 1, \dots, n_j, \tag{3.2}$$

where $\boldsymbol{\theta}^{(i)} = \left(\boldsymbol{\alpha}^{(i)\prime}, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu}^{(i)}\right)$.

The joint density function for the g-component mixture ARCH model can be defined through Eq. (3.2) as:

$$f_{MAR}\left(\boldsymbol{y}_{jt} | \boldsymbol{\mathcal{F}}_{j,t-1}, \boldsymbol{\Theta}\right) = \sum_{i=1}^{g} \pi^{(i)} \prod_{t=1}^{n_j} f\left(y_{jt}^{(i)} | \mathcal{F}_{j,t-1}^{(i)}, \boldsymbol{\theta}^{(i)}\right), \tag{3.3}$$

where $\boldsymbol{\Theta} = \left(\pi^{(1)}, \dots, \pi^{(g-1)}, \boldsymbol{\theta}^{(1)\prime}, \dots, \boldsymbol{\theta}^{(g)\prime}\right)$ represents the complete set of parameters for the NMVM-MoARCH(q), $\boldsymbol{\theta}^{(i)} = \left(\boldsymbol{\alpha}^{(i)\prime}, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu}^{(i)}\right)$, $\boldsymbol{y}_{jt} = \left(\boldsymbol{y}_{jt}^{(1)}, \dots, \boldsymbol{y}_{jt}^{(g)}\right)$, and $\boldsymbol{\mathcal{F}}_{j,t-1} = \left(\boldsymbol{\mathcal{F}}_{j,t-1}^{(1)}, \dots, \boldsymbol{\mathcal{F}}_{j,t-1}^{(g)}\right)$. The value of $\boldsymbol{\nu}^{(i)}$ is different for different models: it is equal to $\nu^{(i)}$ for NMVL, NIG, NMVBS, and GHST, equal to $\left(\kappa^{(i)}, \psi^{(i)}\right)'$ for VG, and equal to 0 for SL.

The log-likelihood of parameters for an iid sample $\left(\boldsymbol{y}_1, \dots, \boldsymbol{y}_N\right)$ can be calculated using the characterization (3.3) as:

$$\ell(\boldsymbol{\Theta}) = \sum_{j=1}^{N} \log f_{MAR}\left(\boldsymbol{y}_{jt} | \boldsymbol{\mathcal{F}}_{j,t-1}, \boldsymbol{\Theta}\right) = \sum_{j=1}^{N} \log\left(\sum_{i=1}^{g} \pi^{(i)} \prod_{t=1}^{n_j} f_{NMVM}\left(y_{jt}^{(i)}; 0, \lambda^{(i)}, \sigma^{2(i)} h_{jt}^{(i)}, \boldsymbol{\nu}^{(i)}\right)\right). \tag{3.4}$$

The NMVM-MoARCH(q) model can be expressed as a two-level hierarchical representation:

$$\frac{y_{jt}^{(i)}}{\sqrt{h_{jt}^{(i)}}} | \mathcal{F}_{j,t-1}^{(i)}, Z_j^{(i)} = 1 \sim \text{NMVM}\left(0, \lambda^{(i)}, \sigma^{2(i)}, \boldsymbol{\nu}^{(i)}\right),$$

$$\boldsymbol{Z}_j \sim Multinomial\left(1, \pi^{(1)}, \dots, \pi^{(g)}\right).$$

By utilizing the stochastic representation of the NMVM, the hierarchical representation of the MoARCH(q) model is expressed as follows:

$$\frac{y_{jt}^{(i)}}{\sqrt{h_{jt}^{(i)}}} | \mathcal{F}_{j,t-1}^{(i)}, W_{jt}^{(i)} = w_{jt}^{(i)}, Z_j^{(i)} = 1 \sim N\left(w_{jt}^{(i)} \lambda^{(i)}, w_{jt}^{(i)} \sigma^{2(i)}\right),$$

$$W_{jt}^{(i)} | Z_j^{(i)} = 1 \sim h\left(w_{jt}^{(i)}, \boldsymbol{\nu}^{(i)}\right),$$

$$\boldsymbol{Z}_j \sim Multinomial\left(1, \pi^{(1)}, \dots, \pi^{(g)}\right).$$

The alternative form of the above hierarchical representation is:

$$y_{jt}^{(i)} | \mathcal{F}_{j,t-1}^{(i)}, W_{jt}^{(i)} = w_{jt}^{(i)}, Z_j^{(i)} = 1 \sim N\left(\sqrt{h_{jt}^{(i)}}(w_{jt}^{(i)} \lambda^{(i)}), w_{jt}^{(i)} \sigma^{2(i)} h_{jt}^{(i)}\right), \tag{3.5}$$

$$W_{jt}^{(i)}|Z_j^{(i)} = 1 \sim h\left(w_{jt}, \boldsymbol{\nu}^{(i)}\right),$$

$$\boldsymbol{Z_j} \sim Multinomial\left(1, \pi^{(1)}, \ldots, \pi^{(g)}\right).$$

for $i = 1, \ldots, g,\ j = 1, \ldots, N,\ t = 1, \ldots, n_j$.

In this research, we apply the maximum likelihood inferential paradigm, where label switching emerges as a theoretical identifiability concern rather than a practical issue. Typically, this challenge can be addressed by establishing a specific order among the mixing proportions, represented as $\pi^{(1)} > \cdots > \pi^{(g)}$. In real-world applications, having too many components in mixture models can lead to overfitting, making our interpretations less reliable. On the flip side, if we have too few components, mixture models might not be flexible enough to capture the true structure of the data. Therefore, figuring out the right number of components is crucial. To tackle this, we use penalized log likelihood

$$\ell_{\text{pen}}(\boldsymbol{\Theta}) = \ell(\boldsymbol{\Theta}) - \sum_{j=1}^{N} n_j \gamma D_{f.\text{MAR}} \sum_{i=1}^{g} \left(\log(\varepsilon + \pi^{(i)}) - \log(\varepsilon)\right), \tag{3.6}$$

where $\ell(\Theta)$ represents the log-likelihood function, γ is a tuning parameter, ε is a small positive constant with a value of 10^{-6}, and $\mathrm{D_{f.MAR}}$ stands for the number of free parameters for each component. The free parameters in GHST − MoARCH(q), NIG − MoARCH(q), NMVBS − MoARCH(q), and NMVL − MoARCH(q) is $\mathrm{D_{f.MAR}} = q+3$, while in VG − MoARCH(q), it is $\mathrm{D_{f.MAR}} = q+4$, and in SL − MoARCH(q), it is $\mathrm{D_{f.MAR}} = q+2$. So, the penalization affects only the probabilities that series (j) belongs to component (i), thereby resulting in the penalization of Ng parameters.

Acquiring the suggested maximizer using the penalized log-likelihood (3.6) proves challenging due to the absence of an explicit solution. Consequently, we turn to an EM-type algorithm as our approach to address this computational complexity.

4. Maximum penalized estimation of the model parameters

In this section, an efficient EM-type algorithm is developed to obtain the Penalized Maximum Likelihood estimation of the parameters of NMVMN − MoARCH(q) using an incomplete-data framework, where missing data points, or unobserved (hidden) latent variables are considered in the estimation process. To obtain the penalized log-likelihood function based on the complete data, we start by calculating the log-likelihood function based on the complete data. Let $\boldsymbol{y} = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_N)'$, $\boldsymbol{W} = (\boldsymbol{W}'_1, \ldots, \boldsymbol{W}'_N)'$, and $\boldsymbol{Z} = (\boldsymbol{Z}'_1, \ldots, \boldsymbol{Z}'_N)'$, such that $\boldsymbol{W_j} = (W_{j1}, \ldots, W_{jN})'$ and $\boldsymbol{Z_j} = (Z_j^{(1)}, \ldots, Z_j^{(g)})'$; $j = 1, \ldots, N$. Using Eq. (3.5), the complete likelihood based on the complete-data, $\boldsymbol{y_c} = (\boldsymbol{y}', \boldsymbol{W}', \boldsymbol{Z}')'$ can be written as follows:

$$L(\boldsymbol{\Theta}) = \prod_{j=1}^{N} \prod_{t=1}^{n_j} \prod_{i=1}^{g} \left[\pi^{(i)} f_{y_{jt}^{(i)}|\mathcal{F}_{jt-1}^{(i)}, W_{jt}^{(i)}, Z_j^{(i)}}\left(y_{jt}^{(i)}; \boldsymbol{\theta}^{(i)}\right) f_{W_{jt}^{(i)}|Z_j^{(i)}}\left(w_{jt}^{(i)}, \boldsymbol{\nu}^{(i)}\right)\right]^{Z_j^{(i)}}, \tag{4.1}$$

where $f_{y_{jt}^{(i)}|\mathcal{F}_{jt-1}^{(i)}, W_{jt}^{(i)}, Z_j^{(i)}}\left(y_{jt}^{(i)}; \boldsymbol{\theta}^{(i)}\right) = N\left(y_{jt}^{(i)}; \sqrt{h_{jt}^{(i)}}(w_{jt}^{(i)}\lambda^{(i)}), w_{jt}^{(i)}\sigma^{2(i)}h_{jt}^{(i)}\right)$ with $h_{jt}^{(i)} = \boldsymbol{\alpha}^{(i)\prime}\boldsymbol{y}_{jt}^{2(i)}$ and $f_{W_{jt}^{(i)}|Z_j^{(i)}}\left(w_{jt}^{(i)}, \boldsymbol{\nu}^{(i)}\right) = h\left(w_{jt}^{(i)}; \boldsymbol{\nu}^{(i)}\right)$ representing a density function induced by the mixing distribution $H\left(w_{jt}^{(i)}; \boldsymbol{\nu}^{(i)}\right)$. As a result, the respective log likelihood-function based on the complete data is obtained by

$$\begin{aligned}
\ell(\boldsymbol{\Theta}) &= \sum_{j=1}^{N} \sum_{t=1}^{n_j} \sum_{i=1}^{g} Z_j^{(i)} \left[\log \pi^{(i)} + \log f(y_{jt}^{(i)}|\mathcal{F}_{jt-1}^{(i)}, w_{jt}^{(i)}, z_j^{(i)}, \boldsymbol{\theta}^{(i)}) + \log f(w_{jt}^{(i)}|z_j^{(i)} = 1)\right] \\
&= \sum_{j=1}^{N} \sum_{i=1}^{g} n_j Z_j^{(i)} \log(\pi^{(i)}) + \sum_{j=1}^{N} \sum_{t=1}^{n_j} \sum_{i=1}^{g} Z_j^{(i)} \log\left[\frac{1}{\sqrt{2\pi w_{jt}^{(i)} h_{jt}^{(i)} \sigma^{2(i)}}} exp\left\{-\frac{1}{2w_{jt}^{(i)}\sigma^{2(i)}h_{jt}^{(i)}}\left(y_{jt}^{(i)} - \sqrt{h_{jt}^{(i)}}(w_{jt}^{(i)}\lambda^{(i)})\right)^2\right\}\right], \\
&+ \sum_{j=1}^{N} \sum_{t=1}^{n_j} \sum_{i=} Z_j^{(i)} \log\left[f_{W_{jt}^{(i)}|Z_j^{(i)}}(w_{jt}^{(i)}; \boldsymbol{\nu}^{(i)})\right].
\end{aligned} \tag{4.2}$$

So, the final expression for the penalized log-likelihood function (3.6) becomes:

$$\begin{aligned}
\ell_p(\boldsymbol{\Theta}) &= \sum_{j=1}^{N} \sum_{i=1}^{g} n_j Z_j^{(i)} \log(\pi^{(i)}) - \frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{g} n_j Z_j^{(i)} \log \sigma^{2(i)} - \frac{1}{2} \sum_{j=1}^{N} \sum_{t=1}^{n_j} \sum_{i=1}^{g} Z_j^{(i)} \log(h_{jt}^{(i)}) \\
&- \frac{1}{2} \sum_{j=1}^{N} \sum_{t=1}^{n_j} \sum_{i=1}^{g} \frac{Z_j^{(i)}}{w_{jt}^{(i)}\sigma^{2(i)}h_{jt}^{(i)}} \left(y_{jt}^{(i)} - \sqrt{h_{jt}^{(i)}}(w_{jt}^{(i)}\lambda^{(i)})\right)^2 + \sum_{j=1}^{N} \sum_{t=1}^{n_j} \sum_{i=1}^{g} Z_j^{(i)} \log(h(w_{jt}^{(i)}; \boldsymbol{\nu}^{(i)})) \\
&- \sum_{j=1}^{N} n_j \gamma D_{f,MAR} \sum_{i=1}^{g} \left\{\log(\varepsilon + \pi^{(i)}) - \log(\varepsilon)\right\}.
\end{aligned} \tag{4.3}$$

The estimated values of $\Theta^{(k)}$ at the kth iteration are represented by $\hat{\Theta}^{(k)}$. Ignoring constant values, the conditional expectation of the complete log-likelihood function is expressed as follows:

$$\begin{aligned} Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}) = & \sum_{j=1}^{N}\sum_{i=1}^{g}\hat{Z}_j^{(i)(k)}\left[n_j\log(\pi^{(i)}) - \frac{1}{2}n_j\log(\sigma^{2(i)})\right] \\ & - \frac{1}{2}\sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{m}\hat{Z}_j^{(i)(k)}\left[\log(h_{jt}^{(i)}) + \frac{y_{jt}^{2(i)}}{\sigma^{2(i)}h_{jt}^{(i)}}\hat{U}_{jt}^{(i)(k)} + \frac{\lambda^{2(i)}}{\sigma^{2(i)}}\widehat{W}_{jt}^{(i)(k)} - 2\frac{y_{jt}^{2(i)}\lambda^{(i)}}{\sqrt{h_{jt}^{(i)}}\sigma^{2(i)}}\right] \\ & + \sum_{j=1}^{N}\sum_{t=1}^{n_j}\sum_{i=1}^{g}\hat{Z}_j^{(i)(k)}\hat{\zeta}_{jt}^{(i)(k)} - \sum_{j=1}^{N}n_j\gamma D_{f,MAR}\sum_{i=1}^{g}\left\{\log(\varepsilon+\pi^{(i)}) - \log(\varepsilon)\right\}, \end{aligned} \tag{4.4}$$

where

$$\hat{U}_{jt}^{(i)(k)} = E\left(\frac{1}{W_{jt}^{(i)}}\middle|y_{jt}^{(i)}, Z_j^{(i)} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\right),$$

$$\widehat{W}_{jt}^{(i)(k)} = E\left(W_{jt}^{(i)}\middle|y_{jt}^{(i)}, Z_j^{(i)} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\right),$$

$$\hat{\zeta}_{jt}^{(i)(k)} = E\left(\log h(w_{jt}^{(i)}; \boldsymbol{\nu}^{(i)})\middle|y_{jt}^{(i)}, Z_j^{(i)} = 1, \hat{\boldsymbol{\Theta}}^{(k)}\right).$$

Details are in Appendix B. $\hat{Z}_j^{(i)(k)}$ is the posterior probability that y_{jt} belongs to the ith cluster and is given by

$$\hat{Z}_j^{(i)(k)} = E\left(Z_j^{(i)}\middle|\boldsymbol{y}_j^{(i)}, \hat{\boldsymbol{\Theta}}^{(k)}\right) = \frac{\pi^{(i)(k)}\prod_{t=1}^{n_j} f_{NMVM}\left(y_{jt}^{(i)}; 0, \hat{\lambda}^{(i)(k)}, \hat{\sigma}^{2(i)(k)}h_{jt}^{(i)}, \boldsymbol{\nu}^{(i)(k)}\right)}{\sum_{s=1}^{g}\hat{\pi}^{(s)(k)}\prod_{t=1}^{n_j} f_{NMVM}\left(y_{jt}^{(s)}; 0, \hat{\lambda}^{(s)(k)}, \hat{\sigma}^{2(s)(k)}h_{jt}^{(s)}, \boldsymbol{\nu}^{(s)(k)}\right)}, \tag{4.5}$$

Additionally, updating $\boldsymbol{\nu}^{(i)(k)}$ can be simplified by using the CML-step of the ECME algorithm, thus eliminating the need to perform the complex calculation of $\hat{\zeta}_{jt}^{(i)(k)}$. **CM- steps**

Updating the parameters in the CM-steps is accomplished through the following sub-steps:

- (i) Update $\pi^{(i)}$ by maximizing (4.4) with respect to $\pi^{(i)}$ which gives

$$\hat{\pi}^{(i)(k+1)} = Max\left\{0, \frac{1}{1-g\gamma D_{f.GA}}\left[\frac{\sum_{j=1}^{N}\hat{Z}_j^{(i)(k)}n_j}{\sum_{j=1}^{N}n_j} - \gamma D_{f.GA}\right]\right\},$$

- (ii) Update $\lambda^{(i)}$ by maximizing (4.4) with respect to λ_i which gives

$$\hat{\lambda}^{(i)(k+1)} = \frac{\sum_{j=1}^{N}\sum_{t=1}^{n_j}\hat{Z}_j^{(i)(k)}y_{jt}^{(i)}/(\sqrt{\hat{h}_{jt}^{(i)}})}{\sum_{j=1}^{N}\sum_{t=1}^{n_j}\hat{Z}_j^{(i)(k)}\widehat{W}_{jt}^{(k)}},$$

- (iii) Update $\sigma^{2(i)}$ by maximizing (4.4) with respect to ${\sigma_i}^2$ which gives

$$\hat{\sigma}^{2(i)(k+1)} = \frac{\sum_{j=1}^{N}\sum_{t=1}^{n_j}\hat{Z}_j^{(i)(k)}\left(\frac{y_{jt}^{2(i)}\hat{U}_{jt}^{(i)(k)}}{2\hat{h}_{jt}^{(i)}} - \frac{2y_{jt}^{(i)}\hat{\lambda}^{(i)(k)}}{\sqrt{\hat{h}_{jt}^{(i)}}} + \widehat{W}_{jt}^{(i)(k)}(\hat{\lambda}^{(i)(k)})^2\right)}{\sum_{j=1}^{N}n_j\hat{Z}_j^{(i)(k)}},$$

- (iv) In the last step (CML-step), update $\boldsymbol{\nu}^{(i)}$ by maximizing (4.4) with respect to $\boldsymbol{\nu}^{(i)}$ which gives

$$\hat{\boldsymbol{\nu}}^{(i)(k+1)} = \arg\max_{\boldsymbol{\nu}^{(i)}}\left\{\sum_{j=1}^{N}\log\sum_{i=1}^{g}\hat{\pi}^{(i)(k)}\prod_{t=1}^{n_j} f_{NMVM}\left(y_{jt}^{(i)}; 0, \hat{\lambda}^{(i)(k+1)}, \hat{\sigma}^{2(i)(k+1)}\hat{h}_{jt}^{(i)(k+1)}, \boldsymbol{\nu}^{(i)}\right)\right\},$$

To achieve identification, we utilized a reparameterization of the model's parameters $(\alpha_0^{(i)}, \alpha_1^{(i)}, \ldots, \alpha_q^{(i)})$. These parameters must be greater than zero and satisfy $\sum_{p=1}^{q}\alpha_p^{(i)} < 1$. Therefore, we expressed them in the following way:

$$\alpha_0^{(i)} = 1 - \sum_{j=1}^{q}\alpha_j^{(i)},$$

and

$$\alpha_j^{(i)} = \frac{1 - \sum_{l=1}^{j-1} exp(-\eta_l^{(i)})}{1 + exp(-\eta_l^{(i)})},\ j = 1, \ldots, q.$$

Therefore, the parameters η, taking values in $\mathbb{R}$, will be estimated in place of the α parameters.

In this way, the unconditional value of the ARCH sequence is 1; moreover, the constraints on the alpha parameters ensure that the intercept will be positive.

In the ECME algorithm, the E-, CM-, and CML-steps are cyclically executed until a predetermined convergence criteria is satisfied. To ensure that the algorithm is progressing correctly, we use the Aitken acceleration method, proposed by Aitken [1]. At iteration (k), the observed log-likelihood is represented by $\ell_{ob}^{(k)}$ and the Aitken's acceleration factor, which is the ratio of successive increments, is calculated as $a^{(k)} = \frac{\ell_{ob}^{(k+1)} - \ell_{ob}^{(k)}}{\ell_{ob}^{(k)} - \ell_{ob}^{(k-1)}}$. For iteration $(k+1)$, the asymptotic log-likelihood is estimated as $\ell_{\infty}^{(k+1)} = \frac{\ell_{ob}^{(k)} + (\ell_{ob}^{(k+1)} - \ell_{ob}^{(k)})}{1 - a^{(k)}}$. When either $\ell_{\infty}^{(k+1)} - \ell_{ob}^{(k)} < tolerance$ or $k_m = 1000$ iterations are reached, the algorithm is considered to have converged. The tolerance value can be adjusted, and in this paper, it is set to 10^{-5}. The pseudocode for the proposed algorithm can be found in Algorithm 1.

Algorithm 1 The ECME Algorithm for Fitting the NMVM-MoARCH(q)

procedure *NMVM-MoARCH(q)*
- **inputs:** $\{y_{jt}^{(i)}\}, j = 1, ..., N; t = 1, ..., n_j$- the set of input data; **g** - the number of components;
- $\epsilon = 10^{-5}$ - the prespecified tolerance; $k_m = 1000$- the maximum number of iteration.
- **initialize:** Obtain the starting value $\widehat{\boldsymbol{\theta}}^{(0)}$.
- Compute the initial log-likelihood as $\ell_{ob}^{(0)}$.
- Set "Convergence = False".
- **for** $k = 0$ *to* k_m **do**
  - **for** *j=1 to N* **do**
    - **for** *i=1 to g* **do**
      - **for** $t = 1$ *to* n_j **do**
        - Obtain latent and missing information $\widehat{Z}_j^{(i)(k)}$, $\widehat{W}_{tj}^{(i)(k)}$, $\widehat{U}_{tj}^{(i)(k)}$ and $\widehat{\zeta}_{tj}^{(i)(k)}$.
      - **end for**
    - **end for**
  - **end for**
  - **for** *i=1 to g* **do**
    - Update the parameters $\widehat{\pi}^{(i)(k+1)}$, $\widehat{\sigma}^{2(i)(k+1)}$, $\widehat{\lambda}^{(i)(k+1)}$, $\widehat{\boldsymbol{\nu}}^{(i)(k+1)}$ and $\widehat{\boldsymbol{\alpha}}^{(i)(k)} = (\widehat{\alpha}_0^{(i)(k)}, \widehat{\alpha}_1^{(i)(k)}, ..., \widehat{\alpha}_q^{(i)(k)})$.
    - Set $\widehat{\boldsymbol{\Theta}}^{(k+1)} = (\widehat{\boldsymbol{\pi}}^{(k+1)}, \widehat{\boldsymbol{\theta}}^{(1)(k+1)}, ..., \widehat{\boldsymbol{\theta}}^{(g)(k+1)})$.
  - **end for**
  - Compute $\ell_{ob}^{(k+1)}$ and Compute the Aitken's acceleration factor.
  - set "Convergence = True" if $\ell_{\infty}^{(k+1)} - \ell_{ob}^{(k)} < tolerance$.
  - **if** *Convergence = True* **then**
    - break
  - **end if**
- **end for**
- **return** $\ell_{ob}^{(k+1)}$ and $\widehat{\boldsymbol{\theta}}^{(k+1)}$.

end procedure

4.1. Selection of tuning parameter

We outline the procedure for determining the tuning parameter γ in the penalized log-likelihood function (3.6). There are various methods in the literature for computing the tuning parameter, such as those used in standard the LASSO given by Tibshirani [40] and SCAD given by Fan and Li [16].

Our approach adheres to the methodology outlined in the work by Wang et al. [44], wherein we utilize the Bayesian Information Criterion (BIC). This is defined by:

$$BIC(\gamma) = \sum_{j=1}^{N} \log \sum_{i=1}^{\widehat{g}} \widehat{\pi}^{(i)(k)} \prod_{t=1}^{n_j} f_{NMVM}\left(y_{jt}^{(i)}; 0, \widehat{\lambda}^{(i)}, \widehat{\sigma}^{(i)} \widehat{h} j t^{(i)}, \widehat{\boldsymbol{\nu}}^{(i)}\right) - \frac{1}{2} \widehat{g} D_{f.MAR} \log \sum_{j=1}^{N} n_j,$$

and γ is estimated as:

$$\widehat{\gamma} = \arg\max_{\boldsymbol{\gamma}} BIC(\gamma),$$

where $\widehat{g}$ represents the number of NMVM – MoARCH(q) components.

5. Computational aspects

In this section, we addressed the computational aspects of estimating standard errors for the parameters in the NMVM – MoARCH(q) model. We employed the empirical information matrix to derive the asymptotic covariance matrix of the MPL estimates. The

individual score, involving intricate partial derivatives, was formulated to calculate the elements of this matrix. To approximate standard errors, we utilized the inverse of this matrix. Additionally, a systematic approach for determining effective initial values is detailed, emphasizing their critical role in ensuring the convergence and accuracy of the optimization algorithm.

5.1. Estimation of standard errors

In this section, we aim to determine the standard errors of the parameter estimates obtained from the NMVM – MoARCH(q) model. Under regularity conditions, the asymptotic covariance matrix of the MPL estimates, represented by $\widehat{\boldsymbol{\Theta}}$, can be estimated by taking the inverse of the empirical information matrix, which is expressed as:

$$\boldsymbol{I}(\boldsymbol{\Theta}) = \frac{\partial^2 \ell_{pen}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'},$$

where $\ell_{pen}(\boldsymbol{\Theta}) = \sum_{j=1}^{N} \ell_{pj}(\boldsymbol{\Theta})$, and $\ell_{pj}(\Theta) = \log \sum_{i=1}^{g} \pi^{(i)} \prod_{t=1}^{n_j} f(y_{jt}^{(i)} | \mathcal{F}_{j,t-1}^{(i)}, \boldsymbol{\theta}^{(i)}) - \sum_{j=1}^{N} n_j \gamma D_{f.MAR} \sum_{i=1}^{g} \Big(log(\varepsilon + \pi^{(i)}) - log(\varepsilon) \Big)$.

In real applications, the standard deviation of the MPL estimation is taken as the square root of the diagonal elements of the asymptotic covariance matrix. According to [5], the empirical information matrix is given by:

$$\boldsymbol{I}(\widehat{\boldsymbol{\Theta}}) = \sum_{j=1}^{N} s(\boldsymbol{y}_j^{(i)} | \widehat{\boldsymbol{\Theta}}) s'(\boldsymbol{y}_j^{(i)} | \widehat{\boldsymbol{\Theta}}),$$

The individual score $s(\boldsymbol{y}_j^{(i)} | \widehat{\boldsymbol{\Theta}}) = \partial \prod_{t=1}^{n_j} f(y_{jt}^{(i)} | \boldsymbol{\Theta}) / \partial \boldsymbol{\Theta}$ does not have an explicit form due to its partial derivative. To ascertain this, following the approach outlined in [26], we determine the expectation of the partial derivative of the individual penalized log-likelihood function in (3.6) as follows:

$$s(\boldsymbol{y}_j^{(i)} | \widehat{\boldsymbol{\Theta}}) = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} | y_{jt}^{(i)}, \widehat{\boldsymbol{\Theta}} \right], \tag{5.1}$$

The above score can then be divided into multiple components:

$$s(\boldsymbol{y}_j^{(i)} | \widehat{\boldsymbol{\Theta}}) = \Big(\widehat{s}_{jt,\pi^{(1)}}, \dots, \widehat{s}_{jt,\pi^{(g-1)}}, \widehat{s}'_{jt,\boldsymbol{\alpha}^{(1)}}, \dots, \widehat{s}'_{jt,\boldsymbol{\alpha}^{(g)}}, s_{jt,\lambda^{(1)}}, \dots, s_{jt,\lambda^{(g)}}, s_{jt,\sigma^{2(1)}}, \dots, s_{jt,\sigma^{2(g)}}, \widehat{s}'_{jt,\boldsymbol{\nu}^{(1)}}, \dots, \widehat{s}'_{jt,\boldsymbol{\nu}^{(g)}} \Big), \tag{5.2}$$

where its coordinate elements for $i = 1, \dots, g$ are given by

$$\widehat{s}_{jt,\pi^{(i)}} = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \pi^{(i)}} | y_{jt}^{(i)}, \widehat{\boldsymbol{\Theta}} \right] = \sum_{j=1}^{N} \left(\frac{\widehat{Z}_j^{(i)}}{\pi^{(i)}} - \frac{\widehat{Z}_j^{(g)}}{\pi^{(g)}} \right) - \sum_{j=1}^{N} n_j \gamma D_{f.MAR} \left(\frac{1}{\varepsilon + \pi^{(i)}} - \frac{1}{\varepsilon + \pi^{(g)}} \right),$$

$$\widehat{s}_{jt,\boldsymbol{\alpha}^{(i)}} = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \boldsymbol{\alpha}^{(i)}} | y_{jt}^{(i)}, \widehat{\boldsymbol{\Theta}} \right] = -\frac{1}{2} \sum_{j=1}^{N} \sum_{t=1}^{n_j} \widehat{Z}_j^{(i)} \left(\frac{1 + \sum_{l=1}^{q} y_{j(t-l)}^{2(i)}}{\widehat{h_{jt}^{(i)}}} \right) + \frac{1}{2} \sum_{j=1}^{N} \sum_{t=1}^{n_j} \frac{\widehat{Z}_j^{(i)} \widehat{U}_{jt}^{(i)}}{\widehat{\sigma}^{2(i)}} \left(\frac{1 + \sum_{l=1}^{q} y_{j(t-l)}^{2(i)}}{\widehat{h_{jt}^{(i)}}} \right),$$

$$\widehat{s}_{jt,\lambda^{(i)}} = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \lambda^{(i)}} \Big| y_{jt}^{(i)}, \widehat{\boldsymbol{\Theta}} \right] = \sum_{j=1}^{N} \sum_{t=1}^{n_j} \frac{\widehat{Z}_j^{(i)}}{\widehat{\sigma}^{2(i)}} \left(\frac{y_{jt}^{(i)}}{\sqrt{\widehat{h_{jt}^{(i)}}}} - \lambda^{(i)} \widehat{W}_{jt}^{(i)} \right),$$

$$\widehat{s}_{jt,\sigma^{2(i)}} = E\left[\frac{\partial \ell_p(\boldsymbol{\Theta})}{\partial \sigma^{2(i)}} | y_{jt}^{(i)}, \widehat{\boldsymbol{\Theta}} \right] = -\frac{1}{2} \sum_{j=1}^{N} \frac{\widehat{Z}_j^{(i)} n_j}{\sigma^{2(i)}} - \sum_{j=1}^{N} \sum_{t=1}^{n_j} \frac{\widehat{Z}_j^{(i)} \widehat{U}_{jt}^{(i)}}{2(\widehat{\sigma}^{2(i)})^2} \left(y_{jt}^{(i)} - \sqrt{\widehat{h}_{jt}^{(i)}} \left(\widehat{W}_{jt}^{(i)} \widehat{\lambda}^{(i)} \right) \right)^2.$$

As a result, the approximation of standard errors is:

$$se(\widehat{\theta}_k) \approx \sqrt{\left[\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\Theta}}) \right]_{kk}}, \tag{5.3}$$

where the (k,k)th element of the inverse of 5.1 is represented by $\left[\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\Theta}}) \right]_{kk}$.

5.2. Initial values

The initial values play an important role in the convergence and accuracy of the estimated parameters in the NMVM – MoARCH(q) model. The initial values serve as a starting point for the optimization algorithm used to estimate the parameters. If the initial values are not set appropriately, the optimization algorithm may converge to a local optimum, resulting in biased or inefficient estimates of the parameters. In addition, the initial values can affect the speed of convergence of the optimization algorithm, so it is important to choose initial values that are close to the true values to ensure fast convergence. The steps we have outlined for finding appropriate initial values aim to provide a good starting point for the optimization algorithm by taking into account the clustering of the data and by making use of OLS estimation and residuals. The steps are as follows:

- **step1:** Partition the data $(y_{jt}, j = 1, \dots, N; t = 1, \dots, n_j)$ into g categories based on the standard deviation. This variance-based clustering approach, utilizing standard deviation, ensures that clusters are formed considering the variations in the data rather than relying solely on averages.

- **step2:** Use the proportion of data points assigned to each category as the starting point $\widehat{\pi}^{(i)(0)}$ for each $i = 1, \ldots, g$. Then use ordinary least-square (OLS) estimation to obtain $\widehat{\boldsymbol{\alpha}}^{(i)(0)}$ as the initial value for the coefficients of the MoARCH model. When applying the OLS technique to estimate $\widehat{\boldsymbol{\alpha}}^{(i)(0)}$, depending on the specific requirements of the model and the dataset being analyzed, two approaches can be used. One approach could involve using a single series within the ith category to estimate $\widehat{\boldsymbol{\alpha}}^{(i)(0)}$. This can be achieved by creating linear equations based on the data points in this particular series. A moving window technique may be employed to iteratively estimate the parameter using subsets of the series data, allowing for a more dynamic estimation process. Alternatively, another approach could involve using all series assigned to the ith category to create a comprehensive linear equation system. By pooling data from multiple series within the same category, a more robust estimation of the initial value $\widehat{\boldsymbol{\alpha}}^{(i)(0)}$ can be obtained, taking into account the collective information from all relevant series. Ultimately, the choice between using a single series or multiple series to estimate $\widehat{\boldsymbol{\alpha}}^{(i)(0)}$ may depend on the specific characteristics of the data, the complexity of the model, and the desired level of accuracy in the estimation process. It is essential to consider the trade-offs between computational efficiency and the representativeness of the estimated parameter when deciding on the approach to apply the OLS technique in determining the initial values for the ith category in the algorithm. Additionally, use the mean squared errors of the residuals to initialize $\widehat{\sigma}^{2(i)(0)}$.
- **step3:** Fit NMVM models to the obtained residuals from step 2 to obtain the initial values for $\widehat{\lambda}^{(i)(0)}$ and $\widehat{\nu}^{(i)(0)}$.

6. Numerical study

This section includes a comprehensive numerical study that serves multiple purposes. First, the finite sample properties of Penalized Maximum Likelihood estimates obtained through the ECME algorithm are meticulously assessed. The focus is on the NMVBS – MoARCH(q) and SL – MoARCH(q) models, with an emphasis on their robust consistency across varying sample sizes. The investigation employs illustrative Monte Carlo samples and key performance metrics, including Absolute Relative Bias (ARB), Root Mean Squared Error (RMSE), standard deviation (STD), and approximate standard errors (ASE). The second aspect of the numerical study delves into the convergence behavior of the Bayesian Information Criterion (BIC) during the Expectation–Conditional Maximization Either (ECME) algorithm. The effectiveness of the proposed estimates in judiciously selecting the optimal number of components is highlighted through visualizations and analyses of BIC values. The third component of the study explores the robustness, misspecification, clustering, and forecasting capabilities of the proposed model. Through simulations in strongly and weakly separated datasets, the research evaluates the model's performance in fitting, clustering accuracy, and forecasting accuracy. Various error term distributions and scenarios are considered, providing insights into the versatility and effectiveness of the proposed model. Finally, a comprehensive benchmarking against a competing methodology, the scale-mixture of skew-normal (SMSN) class, is conducted. This comparative analysis involves generating Monte Carlo samples, computing BIC values, and recording miss-classification error rates (MCR). The results of this benchmarking highlight the suitability and superiority of the proposed NMVM – MoARCH(q) model, particularly the NIG-MoARCH(q) variant, for the considered data generation compared to SMSN-MoARCHs. The overall numerical study contributes empirical evidence and insights into the performance and capabilities of the proposed model under various scenarios and benchmarks.

6.1. Finite sample properties

The initial simulation study aims to assess the performance of maximum penalized likelihood (MPL) estimates obtained through the ECME algorithm and their corresponding standard errors using the information-based method described in Sections 4 and 5, respectively. For illustrative purposes, 500 Monte Carlo (MC) samples of sizes 500, 1000, and 2000 are generated from the NMVBS – MoARCH(q) and SL – MoARCH(q) model with g = 2 and q = 2. The assumed parameters are $\pi_1 = 0.4$, $\pi_2 = 0.6$, $\alpha_1 = (0.9, 0.05)$, $\alpha_2 = (0.85, 0.1)$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$, $\lambda_1 = 2$, $\lambda_2 = 3$, $\nu_1 = 2$, and $\nu_2 = 3$. To assess the accuracy of our estimates, we calculate the Absolute Relative Bias (ARB) and the Root Mean Squared Error (RMSE):

$$ARB = \frac{1}{R}\sum_{r=1}^{R}\left|\frac{\hat{\theta}_r - \theta_{true}}{\theta_{true}}\right| \quad and \quad RMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta_{true})^2},$$

where $\hat{\theta}_r$ represents the MPL estimate of a specific parameter in the rth replication, and θ_{true} denotes its true value. Furthermore, we calculate the standard deviation (STD) of MPL estimates across 500 replications, along with the corresponding average values of their approximate standard errors (ASE). These metrics are computed as follows:

$$ASE = \frac{1}{R}\sum_{r=1}^{R} se(\hat{\theta}_r) \quad and \quad STD = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\theta}_r - \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r\right)^2}.$$

The diminishing trends in both ARB and RMSE observed in Figs. 1 and 2 as the sample size (n) increases provide clear empirical evidence of the robust consistency in MPL estimators for NMVMBS – MoARCH(2) and SL – MoARCH(2), respectively. Additionally, the numerical findings in Table 1 highlight a convergence between the values of STD and ASE with increasing n. This convergence underscores the effectiveness of (5.3) in calculating the standard errors of MPL estimates for both NMVMBS – MoARCH(2) and SL – MoARCH(2). It is worth noting in passing that similar simulation studies were conducted for alternative NMVM – MoARCH(2) models, yielding comparable results that are not detailed here for brevity.

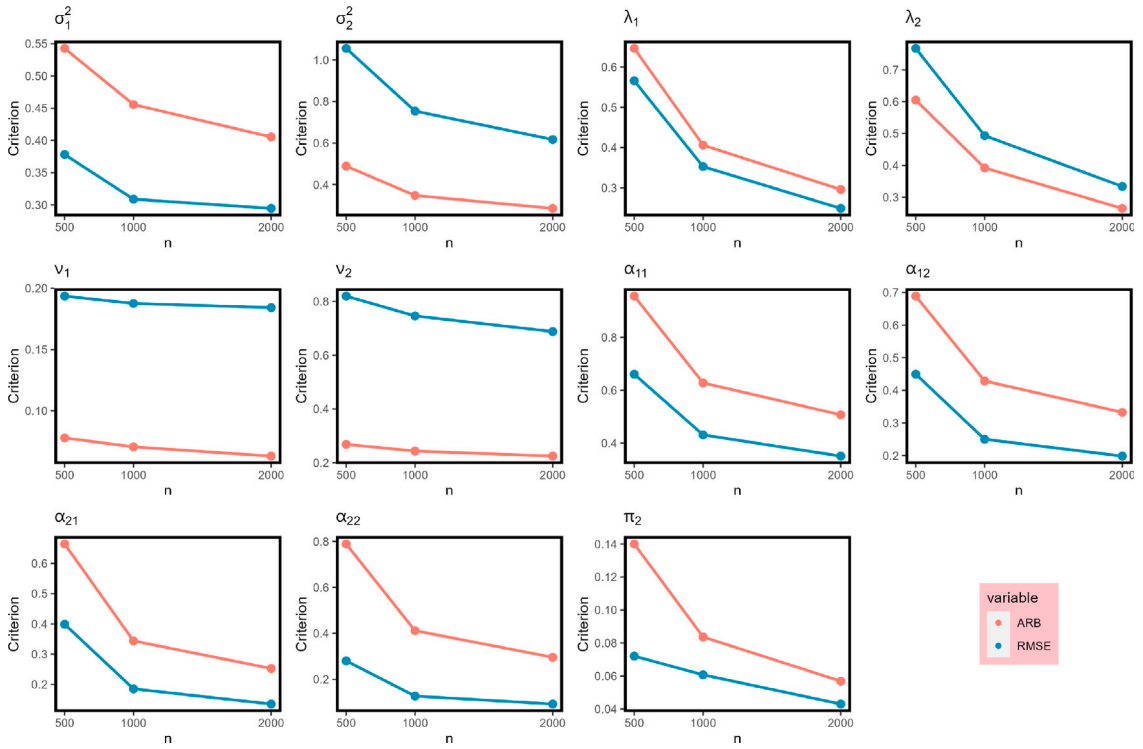


Fig. 1. Performance metrics of parameter estimates for NMVMBS-MoARCH(2) under varying sample sizes: Absolute ative Bias (ARB) and Root Mean Squared Error (RMSE).

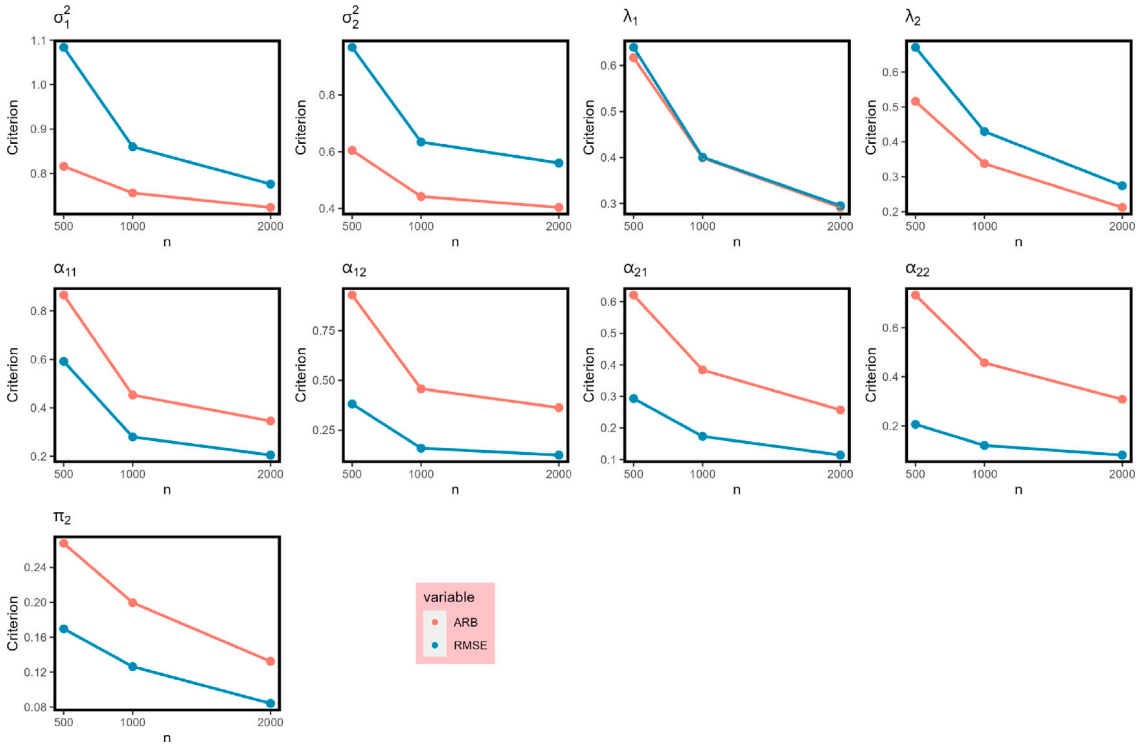


Fig. 2. Performance metrics of parameter estimates for SL – MoARCH(2) under varying sample sizes: Absolute Relative Bias (ARB) and Root Mean Squared Error (RMSE).

Table 1
Simulation results evaluating the asymptotic standard errors and empirical standard deviations of parameter estimates across different sample sizes.

Distribution	n	Measure	σ_1^2	σ_2^2	λ_1^2	λ_2^2	ν_1^2	ν_2^2	α_{11}	α_{12}	α_{21}	α_{22}	π_1
NMVBS	500	STD	0.0864	0.2359	0.3219	0.4595	0.0170	0.0305	0.2896	0.1890	0.2556	0.1277	0.0436
		ASE	0.3783	1.0556	0.5664	0.7678	0.1937	0.8197	0.3306	0.2250	0.3192	0.2243	0.1010
	1000	STD	0.0612	0.1358	0.1261	0.1914	0.0186	0.0255	0.1034	0.0517	0.0588	0.0268	0.0234
		ASE	0.3090	0.7539	0.3531	0.4941	0.1877	0.7459	0.2156	0.1251	0.1481	0.1016	0.0608
	2000	STD	0.0469	0.1025	0.0625	0.0883	0.0213	0.0255	0.0732	0.0322	0.0325	0.0142	0.0085
		ASE	0.2947	0.6170	0.2492	0.3343	0.1843	0.6878	0.1755	0.0993	0.1078	0.0737	0.0430
SL	500	STD	0.5375	0.2919	0.1640	0.1410	–	–	0.3453	0.2868	0.2071	0.1052	0.0473
		ASE	1.0843	0.5809	0.3200	0.2517	–	–	0.5921	0.3820	0.2934	0.2069	0.1697
	1000	STD	0.1704	0.0705	0.0643	0.0575	–	–	0.0821	0.0527	0.0731	0.0358	0.0226
		ASE	0.8602	0.3804	0.2005	0.1613	–	–	0.2799	0.1594	0.1734	0.1208	0.1264
	2000	STD	0.0810	0.0458	0.0350	0.0235	–	–	0.0336	0.0344	0.0169	0.0097	0.0336
		ASE	0.7762	0.3360	0.1478	0.1029	–	–	0.2048	0.1246	0.1139	0.0814	0.0842

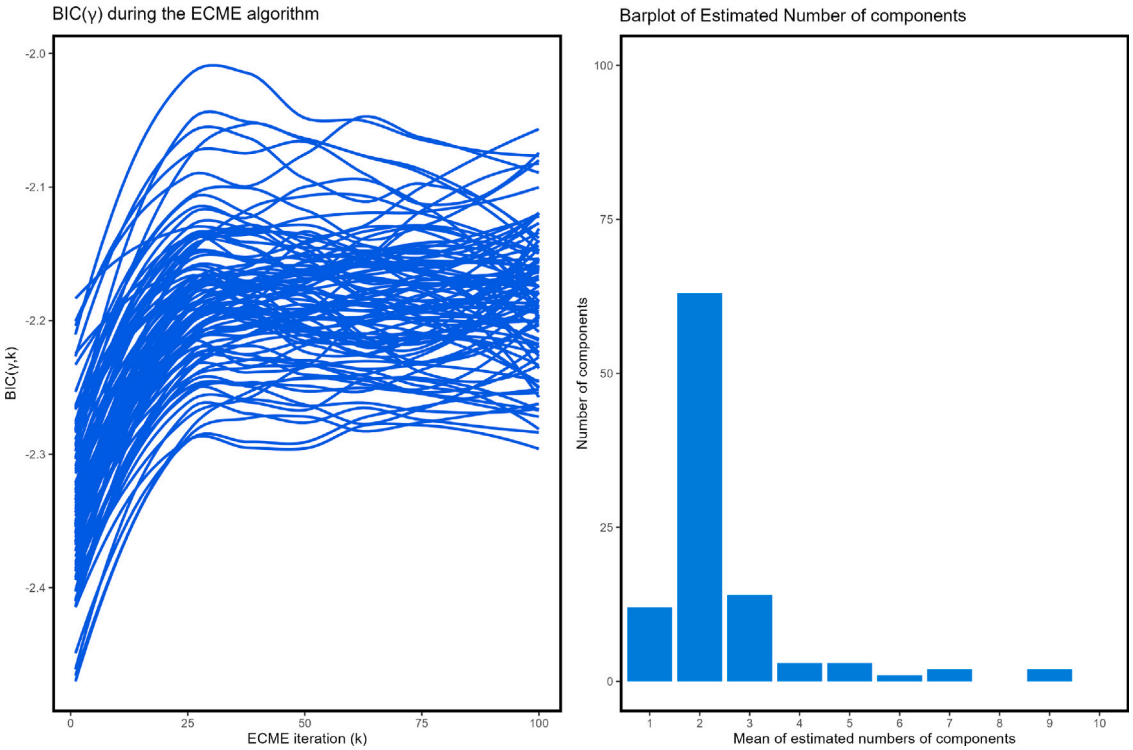


Fig. 3. Bayesian Information Criterion (BIC) values for 1000 samples obtained during the ECME algorithm (left), accompanied by a Barplot presenting the mean estimates of the number of components based on 1000 samples (right).

6.2. Convergence of BIC

In the context of the second simulation study, 100 Monte Carlo (MC) samples of size 1000 are generated from the NMVBS – MoARCH(q) model with g = 2 and q = 2 . Specified parameters include $\pi_1 = 0.4$, $\pi_2 = 0.6$, $\alpha_1 = (0.9, 0.05)$, $\alpha_2 = (0.85, 0.1)$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 2$, $\lambda_1 = 2$, $\lambda_2 = 3$, $\nu_1 = 2$, and $\nu_2 = 3$. Fig. 3 (left) illustrates the Bayesian Information Criterion (BIC(λ)) for each sample throughout the Expectation–Conditional Maximization Either (ECME) algorithm, showcasing their monotonic behavior and convergence. Additionally, Fig. 3 (right) presents a Barplot portraying the mean estimates of the number of components based on 100 samples. The Barplot emphasizes that the true number of components (two components) has the highest frequency, indicating convergence during the ECME algorithm. These results collectively highlight the effectiveness of the proposed estimates in judiciously selecting the optimal number of components.

6.3. Robustness, misspecification, clustering, and forecasting

This simulation examines the efficacy of our proposed model in modeling, clustering, and forecasting, particularly focusing on the conditional variance in diverse data scenarios. We simulate 100 samples of size $n = 1000$ from MoARCH(2) with $g = 3$, considering error terms from extended skew-normal (ESN) [13] and skew-t (ST) [3] distributions. Two distinct scenarios, representing strongly and weakly separated models, are as follows:

- **Strongly Separated Model:** Data is generated from ESN − MoARCH(2) and ST − MoARCH(2) models with $\pi_1 = 0.3$ and $\pi_2 = 0.3$. The remaining parameters are defined as: $\alpha_1 = (0.9, 0.05)$, $\alpha_2 = (0.85, 0.1)$, $\alpha_3 = (0.8, 0.15)$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 1$, $\sigma_3^2 = 2$, $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 4$, $\nu_1 = 3$, $\nu_2 = 5$, and $\nu_3 = 7$.
- **Weakly Separated Model:** Data is generated from ESN − MoARCH(2) and ST − MoARCH(2) models with $\pi_1 = 0.3$ and $\pi_2 = 0.3$. The remaining parameters are defined as: $\alpha_1 = (0.9, 0.05)$, $\alpha_2 = (0.85, 0.1)$, $\alpha_3 = (0.8, 0.15)$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.6$, $\sigma_3^2 = 1$, $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 4$, $\nu_1 = 3$, $\nu_2 = 5$, and $\nu_3 = 7$.

For each of the 100 replication trials, we fitted six NMVM − MoARCH(2) models with $g = 3$. To evaluate the suitability of the fitted models, we employ two penalized-likelihood information criteria: the corrected akaike information criterion (CAIC) [6] and the adjusted Bayesian information criterion (ABIC) [37]. These criteria are defined as $c(n, m) - 2l_p$, where m represents the number of free parameters in the candidate model, and l_p is the maximum penalized log-likelihood value. The penalty term $c(n, m)$ is $2nm/(n-m-1)$ for CAIC and $m \log((n+2)/24)$ for ABIC, respectively. Models exhibiting smaller values of these criteria are considered to offer a superior fit to the data. The assessment of clustering accuracy employed the adjusted rand index (ARI) [23] and the adjusted mutual information (AMI) [43]. Additionally, the average CPU time (in seconds) was recorded for each scenario. Importantly, we refocused our forecasting evaluation to emphasize the models' capability in predicting the conditional variance. This was quantified using mean squared prediction error (MSPE), absolute prediction error (APE), and mean absolute prediction error (MAPE), now redefined in the context of conditional variance prediction:

$$\text{MSPE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(h_i - \hat{h}_i)^2}, \quad \text{APE} = \frac{1}{n}\sum_{i=1}^{n}\left|h_i - \hat{h}_i\right|, \quad \text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{h_i - \hat{h}_i}{h_i}\right|.$$

Table 2 provides an overview of the performance of both strongly and weakly separated models in the context of ESN (light-tailed) and ST (heavy-tailed) data generation. In general, under the strong scenario, models exhibit slightly superior fitting, clustering accuracy, and forecasting accuracy for h_i compared to the weak scenario.

In the strong scenario of ESN-MoARCH(2) data generation, NIG-MoARCH(2), GHST-MoARCH(2), and NMVBS-MoARCH(2) demonstrate slightly better fitting performance compared to their counterparts. Although model selection and cluster validation indices for these MoARCHs are closely aligned, NIG-MoARCH(2), SL-MoARCH(2) and NMVBS-MoARCH(2) are generally favored due to their faster implementation, resulting in less consumed CPU time. Regarding forecasting accuracy, NIG-MoARCH(2) and GHST-MoARCH(2) outperform their counterparts. The results for weakly separated models under ESN (light-tailed) data generation are somewhat similar.

In the strong scenario of ST-MoARCH(2) data generation, NIG-MoARCH(2), NMVL-MoARCH(2), and NMVBS-MoARCH(2) exhibit slightly superior fitting performance compared to their counterparts. Similar to the ESN case, cluster validation indices for these MoARCHs are closely aligned, with NIG-MoARCH(2), SL-MoARCH(2), and NMVBS-MoARCH(2) being generally favored due to their faster implementation in terms of less consumed CPU time. In forecasting accuracy, NIG-MoARCH(2), GHST-MoARCH(2), and SL-MoARCH(2) show better performance than their counterparts. Findings for models with weak separation under ESN (light-tailed) data generation exhibit some resemblance.

When interpreting the resemblance in results between ESN (light-tailed) and ST (heavy-tailed) data generation, it becomes evident that models, especially NIG-MoARCH(2) and NMVBS-MoARCH(2), consistently exhibit superior fitting, forecasting, and cluster accuracy. Their preference is further underscored by faster implementation, resulting in less consumed CPU time. This suggests a robust performance, reinforcing the effectiveness of these models in capturing underlying patterns and enhancing computational efficiency.

6.4. Benchmarking against a competing methodology

In this simulation experiment, we conduct a comprehensive performance comparison between our proposed model and the scale-mixture of skew-normal (SMSN) class introduced by Branco and Dey [10]. Serving as an alternative skew extension to the classical MoARCH(q), the SMSN-MoARCH(q) is defined as:

$$\frac{y_{jt}^{(i)}}{\sqrt{h_{jt}^{(i)}}} \mid \mathcal{F}_{jt-1}, Z_j^{(i)} = 1 \sim \text{SMSN}\left(0, \lambda^{(i)}, \sigma^{2(i)}, \nu^{(i)}\right), \quad \boldsymbol{Z}_j \sim \text{Multinomial}(1, \pi^{(1)}, \ldots, \pi^{(g)}).$$

This model encompasses various special cases, such as ST-MoARCH(q), SSL-MoARCH(q), and SCN-MoARCH(q). While there exists limited empirical evidence comparing the practical performance of NMVM and SMSN models, it is prudent for users to carefully consider both in real-world analyses.

Table 2
Simulation results evaluating model selection, clustering, and prediction performance for the six sub-models within NMVM-MoARCH(q). Data generation assumes ESN and ST distributions for the error terms.

Error	Measure	SL		VG		NIG	
Generator		Strong	Weak	Strong	Weak	Strong	Weak
ESN	CAIC	18.1110	17.5268	18.9308	18.2618	14.4338	14.7715
	ABIC	125.0957	124.5115	131.7493	131.0803	110.3316	110.0583
	ARI	0.8989	0.8188	0.9168	0.8118	0.9007	0.8302
	AMI	0.8680	0.7893	0.8956	0.7823	0.8714	0.8029
	MPE	2.6514	2.9871	3.1748	3.8293	2.4465	1.0691
	MAPE	48.3181	57.3431	52.6887	62.9947	13.1262	13.5432
	MSPE	5.0264	5.7723	7.6857	10.5441	1.3325	1.3869
	CPU time	1.9712	1.7391	13.8734	15.8214	4.1284	4.7574
ST	CAIC	18.0912	17.6077	18.5730	18.0509	14.9564	14.8797
	ABIC	125.0759	124.5924	131.3915	130.8694	110.8541	110.1665
	ARI	0.8655	0.8323	0.8728	0.8356	0.8680	0.8404
	AMI	0.8425	0.8114	0.8473	0.8184	0.8408	0.8239
	MPE	2.1199	2.1562	3.6233	3.6958	1.5141	1.8395
	MAPE	38.4715	38.7878	43.4029	46.2569	10.5770	11.4133
	MSPE	5.5108	6.8829	16.9512	7.0150	2.7398	3.6008
	CPU time	2.5281	2.6061	16.4654	16.7620	3.1194	3.9374
Error	Measure	GHST		NMVBS		NMVL	
Generator		Strong	Weak	Strong	Weak	Strong	Weak
ESN	CAIC	14.5024	15.2389	15.0264	15.7343	15.2353	15.1234
	ABIC	110.4001	111.1366	110.3132	111.0211	110.5221	111.0212
	ARI	0.9190	0.8185	0.8962	0.8122	0.8762	0.8169
	AMI	0.8996	0.7878	0.8673	0.7811	0.8419	0.7884
	MPE	1.3843	2.4896	2.3569	3.0058	1.3436	1.3371
	MAPE	14.4872	37.9822	24.5157	44.3901	6.8333	9.0345
	MSPE	1.5528	10.2812	1.7545	9.6289	1.3224	1.3022
	CPU time	10.9784	11.6940	4.2964	4.8024	4.1084	4.8334
ST	CAIC	15.3250	15.8929	15.5515	16.0879	15.2279	15.5060
	ABIC	111.2227	111.7907	110.8383	111.3747	110.5147	111.4038
	ARI	0.8663	0.8349	0.8782	0.8366	0.8651	0.8365
	AMI	0.8385	0.8172	0.8617	0.8185	0.8364	0.8161
	MPE	1.4829	3.7143	1.9634	4.1914	1.7645	2.4200
	MAPE	12.2927	39.0286	28.5866	51.0278	7.1988	8.1507
	MSPE	2.7941	19.0459	3.1064	11.6276	4.5049	2.6998
	CPU time	5.1294	16.0754	2.7551	4.0970	3.9435	4.3100

To facilitate a comparative assessment, we generate 100 Monte Carlo samples, each of size 1000, from a 3-component MoARCH(2) model with $\pi_1 = 0.3$, $\pi_2 = 0.5$, and $\pi_3 = 0.2$. Ensuring fairness in comparisons, errors are generated from the extended skew-normal (ESN) distributions as proposed by Capitanio et al. [13]. The specified parameters include $\alpha_1 = (0.9, 0.05)$, $\alpha_2 = (0.85, 0.1)$, $\alpha_3 = (0.8, 0.15)$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.6$, $\sigma_3^2 = 1$, $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 4$, $\nu_1 = 0$, $\nu_2 = 0$, $\nu_3 = 0$. These simulated data effectively capture essential characteristics inherent in both competitive models. In each replication, three NMVM-MoARCH(q) (GHST, NIG, and NMVBS) and three SMSN-MoARCH(q) (ST, SSL, SCN) models are considered to fit the simulate dataset with ESN errors. BIC values and miss-classification error rates (MCR) are recorded, where smaller values indicate better model calibration. The heat-map pairwise comparison in Fig. 4 illustrates the model performance, showcasing the number of times each model outperforms others. The accompanying bar diagrams display cumulative pairwise comparison frequencies for each row. From the heat-map, we observe that NIG-MoARCH(q) offers the best overall fit to the data, while GHST-MoARCH(q) excels in gross clustering performance. Analyzing the overall frequencies for the two classes of MoARCHs, our proposed NMVM-MoARCHs demonstrates superior suitability for this data generation compared to SMSN-MoARCHs.

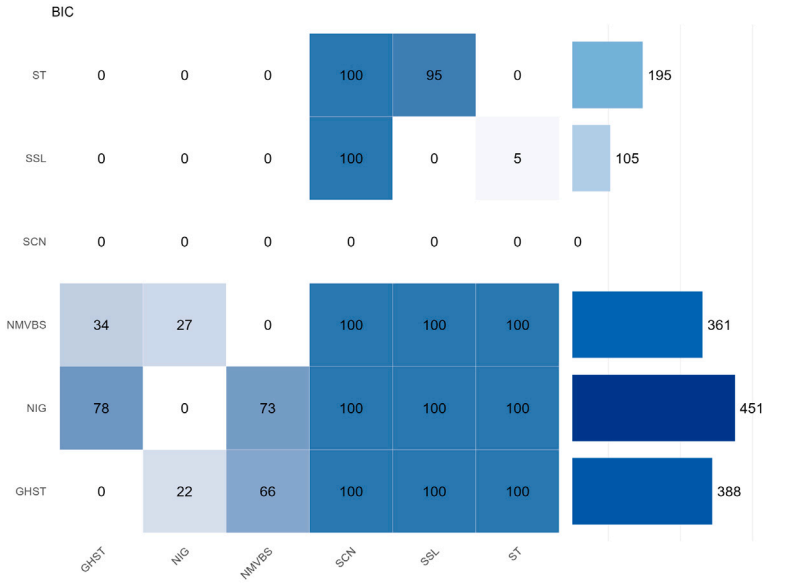
6.5. Practical applications

In this section, we present the robustness and applicability of the proposed MVMN-MoARCH(q) model through its implementation in two real-world datasets.

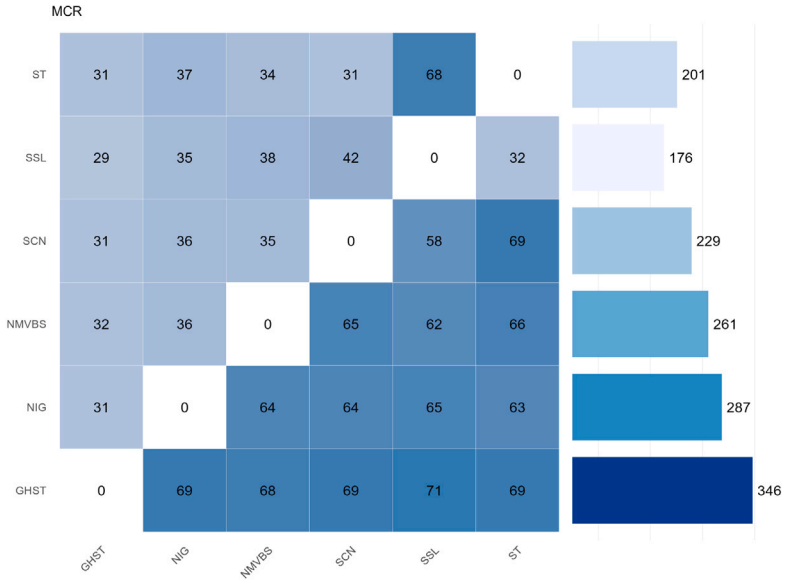
6.5.1. Example 1: S&P500 health care sector

In this first example, we use the 54 companies of the Health Care sector that belong to the S&P 500 equity market index from February 2018 to August 2023.

This dataset comprises data from 54 companies, totaling 1437 observations. To enhance interpretability, we applied the transformation $\log(\frac{x_t}{x_{t-1}})\times 10^2$. Subsequently, the dataset was divided into two segments: one for modeling and the other for validating



(a)



(b)

Fig. 4. Pairwise comparisons of NMVM-MoARCH(q) and SMSN-MoARCH(q) in terms of the BIC and MCR scores.

Table 3
Comparative analysis of 2-component NMVM-MoARCH models and their counterparts fitted to real S&P data. The lowest score for each criterion is highlighted in bold.

	m	LK	AIC	BIC	EDC	CAIC	ABIC
N	7	−1.1735	16.3470	79.6025	351.2252	16.3488	57.3563
SN	9	6.1295	5.7409	87.0694	436.2986	5.7438	58.4672
SSL	13	24.4494	−22.8989	94.5756	599.0178	−22.8930	53.2614
SCN	11	21.7556	−21.5113	77.8902	504.7259	−21.5070	42.9320
ST	11	28.3698	−34.7395	64.6620	491.4976	−34.7353	29.7037
VG	13	28.6352	−31.2704	86.2041	590.6463	−31.2645	44.8898
SL	9	9.3768	−0.7535	80.5750	429.8042	−0.7506	51.9728
NMVL	11	28.3801	−34.7603	64.6412	491.4769	−34.7560	29.6830
NMVBS	**11**	**29.9910**	**−37.9819**	**61.4196**	**488.2553**	**−37.9777**	**26.4614**
GHST	**11**	**29.2817**	**−36.5634**	**62.8381**	**489.6738**	**−36.5591**	**27.8799**
NIG	**11**	**30.4378**	**−38.8756**	**60.5259**	**487.3615**	**−38.8714**	**25.5676**

Table 4
Parameter estimation results and standard errors for a 2-component NIG – MoARCH model fitted to real S&P data.

Distribution	Measure	σ_1^2	σ_2^2	λ_1	λ_2	ν_1	ν_2	α_{11}	α_{12}	α_{21}	α_{22}	π_1
NIG (g = 2)	estimate	0.03452	0.04068	4.1620	4.9846	3.7068	4.9514	0.9361	0.0567	0.9404	0.0423	0.5185
	se	0.00207	0.00188	0.2831	0.3891	0.2603	0.3840	0.0405	0.0193	0.0424	0.0203	0.01087

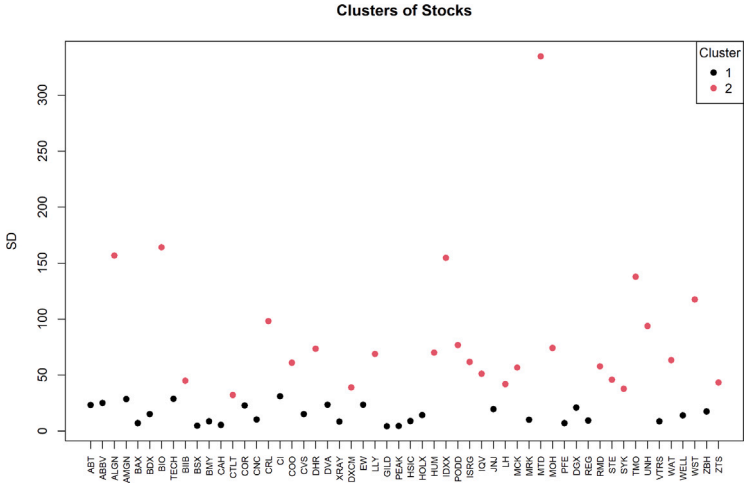


Fig. 5. Clustering obtained from real S&P data using a 2-component NIG – MoARCH model.

results. For modeling purposes, 80% of the data (1150 observations) was utilized, while the remaining 20% (287 observations) was reserved for testing and validation.

Fig. 9 in Appendix C, consistently indicates that the optimal number of clusters is 2 across all models for the given data.

Table 3 showcases the results of Maximum Penalized Likelihood (MPL) fitting obtained from a comprehensive analysis. This analysis comprises a 2-component N – MoARCH, four members of 2-component SMSN – MoARCHs, and six candidate models within 2-component NMVM – MoARCHs class, specifically SL, VG, NIG, GHST, NMVBS, and NMVL. The findings unequivocally underscore the superior performance of all six NMVM – MoARCH models compared to both the N – MoARCH and SMSN – MoARCHs. Notably, within the NMVM – MoARCH class, the NIG – MoARCH, NMVBS – MoARCH, and GHST – MoARCH exhibit exceptional performance. Remarkably, irrespective of the specific evaluation criterion employed, the NIG – MoARCH consistently emerges as the optimal fit among all the candidate models. Furthermore, in Table 4, we conducted MPL estimates of the 2-component NIG – MoARCH model obtained through the ECME algorithm. The corresponding standard errors were determined using the information-based method described in Sections 4 and 5.

Utilizing the optimal model, the 2-component NIG – MoARCH, chosen for its excellence in both clustering and forecasting, we present a comprehensive analysis. Fig. 5 visually illustrates the clustering outcomes for all stocks, employing the standard deviation(SD) of each company as a representative measure for the identified clusters.

Table 5 in our study presents an in-depth analysis of the prediction accuracy for the conditional variance h using various NMVM – MoARCH(2) models applied to real S&P data. This table specifically sheds light on the performance of NMVBS – MoARCH, GHST – MoARCH, and NIG – MoARCH models, with a particular focus on the exceptional prediction accuracy of the NIG – MoARCH

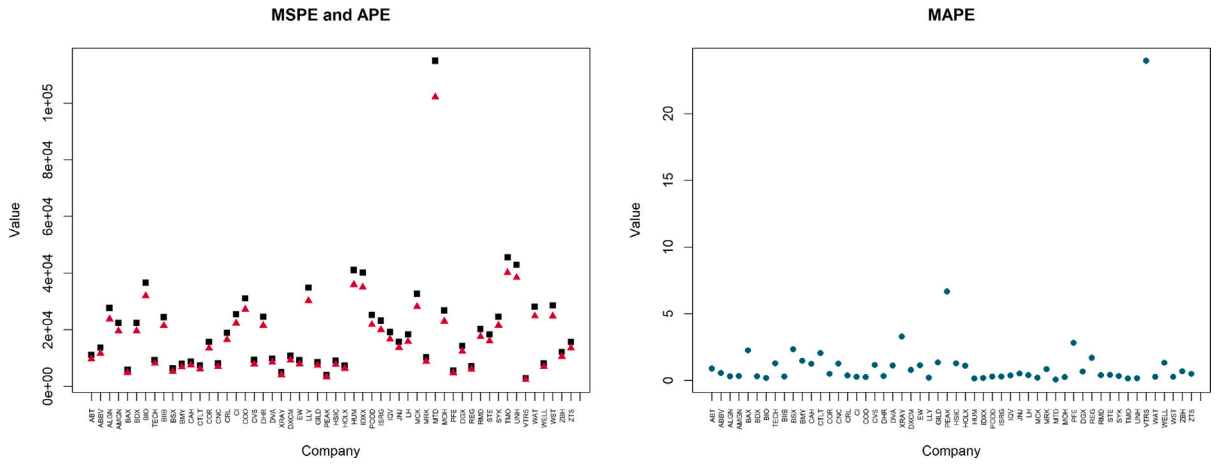


Fig. 6. Comparative analysis of MSPE, APE, and MAPE metrics for various companies based on a 2-component NIG – MoARCH model applied to S&P data.

Table 5
Prediction accuracy of 2-component NMVM – MoARCHs models and their counterparts fitted to real S&P data.

	MSPE	APE	MAPE
N	56 166.2159	46 468.3502	4.8784
SN	42 407.4780	35 655.0805	3.5448
SSL	64 889.2157	52 891.6782	6.0874
SCN	72 101.4259	58 022.8470	6.8029
ST	28 589.2688	24 062.2590	2.1286
VG	47 004.6449	38 889.4215	4.0021
SL	35 618.8327	29 958.4060	2.7863
NMVL	32 376.3890	27 375.8016	2.4551
NMVBS	**21918.9978**	**19021.4997**	**1.4991**
GHST	**25067.4900**	**21509.9051**	**1.8056**
NIG	**19933.7635**	**17339.5988**	**1.3283**

model. The table utilizes metrics such as the mean squared prediction error (MSPE), absolute prediction error (APE), and mean absolute percentage error (MAPE) to evaluate the models' forecasting accuracy for h. Among these, the NIG – MoARCH model distinguishes itself by consistently registering the lowest values across all three metrics, indicating its heightened accuracy in forecasting the conditional variance. The MSPE, APE, and MAPE values for the NIG – MoARCH model are significantly lower than those for the other models, implying a more precise and reliable prediction of the conditional variance h. This is particularly notable in the complex and volatile context of financial time series data represented by the S&P index. Complementing this analysis, Fig. 6 visually demonstrates the application of the best-performing NIG – MoARCH model across various stocks. This figure provides a comprehensive visual representation of the model's comparative performance in terms of prediction accuracy of h across different financial instruments. The visual comparison clearly illustrates the NIG – MoARCH model's superior ability in accurately predicting the conditional variance across a wide range of stocks, reinforcing the model's robustness and applicability in diverse financial scenarios.

6.5.2. *Example 2: life expectancy at birth*

Life expectancy at birth is a crucial demographic indicator that provides valuable insights into the overall health and well-being of populations across the globe. This metric represents the average number of years a newborn is expected to live under current mortality conditions. For this analysis, we leverage real-world data sourced from data.worldbank.org. Our study encompasses 176 countries, providing a comprehensive perspective on life expectancy trends worldwide.

To determine the optimal number of clusters, we employ BAIC metric. Fig. 10 in Appendix C illustrates that the true number of clusters for this dataset is three, consistent across all models. Table 6 presents a comparative analysis of different models within the NMVM-MoARCH framework, revealing that the GHST-MoARCH model exhibits superior performance across various criteria. The clustering of life expectancy among different countries is presented in Fig. 7. For individual country assignments within each cluster, please refer to Appendix D.

The robust clustering of time series data for 176 nations has revealed three distinct groups based on divergent longitudinal trends in life expectancy over the past six decades, as visualized in Fig. 8. In cluster 1, representing nations with world-leading health systems, life expectancy has consistently remained high, averaging around 70 years. The minimal variability in this cluster

Table 6
Comparative analysis of 3-component NMVM-MoARCH models fitted to life expectancy at birth data. The lowest score for each criterion is highlighted in bold.

	m	LK	AIC	BIC	EDC	CAIC	ABIC
VG	20	59.044747	−78.089495	67.862873	299.75258	−78.012367	4.305461
SL	14	41.719411	−55.438823	46.727835	209.05063	−55.400279	2.237645
NMVL	17	50.637143	−67.274286	56.785226	253.89148	−67.218108	2.761427
NMVBS	17	50.375499	−66.750999	57.308514	254.41477	−66.694821	3.284714
GHST	**17**	**50.974832**	**−67.949663**	**56.109849**	**253.2161**	**−67.893485**	**2.086050**
NIG	17	50.59689	−67.193784	56.865728	253.972	−67.137607	2.841928

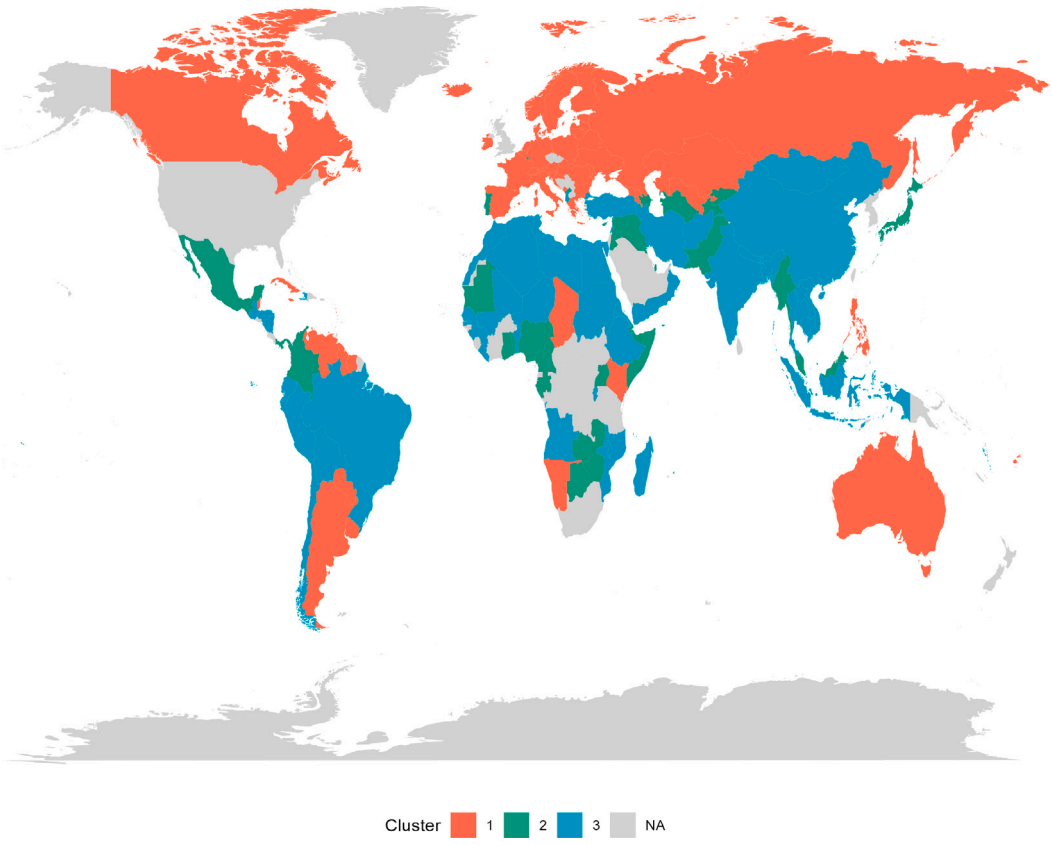


Fig. 7. Clustering obtained from real life expectancy data using a 3-component GHST-MoARCH model on a world map.

suggests resilience, likely stemming from robust infrastructure and access to healthcare. Cluster 2 exhibits a moderating but unstable lifespan, centered near 63 years, with increasing variability reflecting unequal advancements across member states. Notably, mean life expectancy in Cluster 3 showed convergence towards the moderate levels observed in Cluster 2 after 2010, as some nations within this cluster made progress in raising average lifespans from an initial low of 58 years. However, variability remained pronounced, indicating high cross-national inequality. The recent impact of the coronavirus pandemic has led to declining trajectories in mean life expectancy after 2021 across all clusters. This trend has particularly affected Cluster 3 countries, which were already experiencing systemic challenges, emphasizing the pandemic's asymmetrical impact and highlighting healthcare gaps globally. The persistence of variability in cluster 3 underscores the pandemic's disproportionate effects, further accentuating healthcare disparities on a global scale. While the unsupervised learning approach successfully identified heterogeneity in longitudinal patterns, trend analysis has revealed cluster-specific stresses and inequities. This signifies the potential for targeted interventions to support at-risk country groups grappling with lifespan deficits and inequality amid ongoing global disruptions.

7. Conclusions and discussions

This paper introduces a novel robust mixture of ARCH model designed for the modeling, clustering, and forecasting of complex time series data, leveraging the normal mean variance mixture (NMVM) distributions. In the pursuit of identifying the optimal

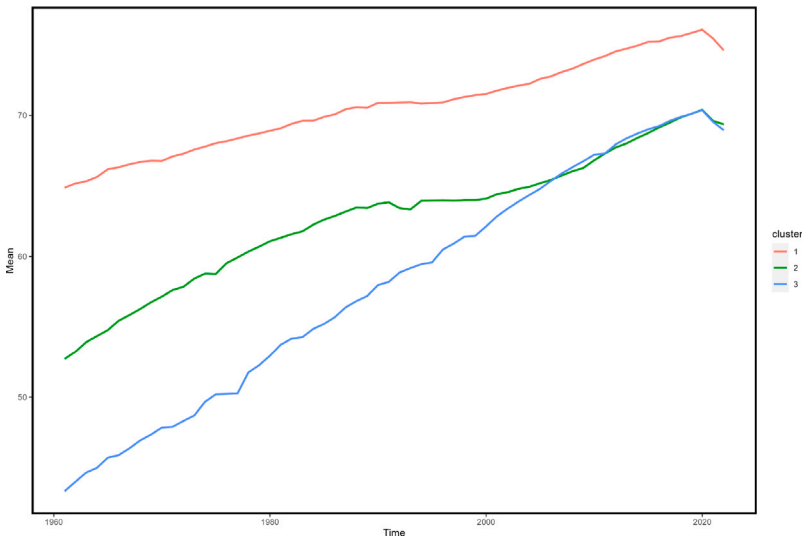


Fig. 8. Quantifying health outcomes through distinct cluster trajectories.

number of components, we adopt a penalized likelihood approach. Our main contribution lies in developing an efficient ECME algorithm structured within a convenient hierarchical representation of the model, streamlining the computation of maximum penalized likelihood (MPL) estimates. Addressing the challenge of approximating asymptotic standard errors, we employ an information matrix-based approach. To assess the model's efficacy, we conduct a thorough analysis using real data, showcasing its ability to capture essential data characteristics and facilitate precise inference. Through a series of simulation experiments, we demonstrate the superior performance of our computational techniques in achieving the desired asymptotic properties for MPL estimates. For those interested in implementing our methods, we provide R-coded computer programs, available upon request.

Looking ahead, several promising avenues for future research emerge. Recommended by the Associate Editor and the reviewers, extending our proposed approach to multivariate variants presents an intriguing avenue for future research. This extension could offer valuable tools for relevant financial areas, such as risk and portfolio management, thereby providing important results for academics, practitioners, and regulators. Additionally, exploring Bayesian techniques using Markov Chain Monte Carlo (MCMC) algorithms holds potential for obtaining even more accurate and precise parameter estimates. Furthermore, to enhance the model's flexibility in handling non-normally distributed data with increased skewness and the potential presence of heavy tails, we contemplate extending the approach to incorporate the family of skew-normal generalized hyperbolic distributions [42].

CRediT authorship contribution statement

F. Setoudehtazangi: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft. **T. Manouchehri:** Conceptualization, Methodology, Writing – original draft. **A.R. Nematollahi:** Conceptualization, Methodology, Supervision. **M. Caporin:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

F. Setoudehtazangi and M. Caporin acknowledge financial support from project PRIN2022 "PRICE: A New Paradigm for High-Frequency Finance", 2022C799SX.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

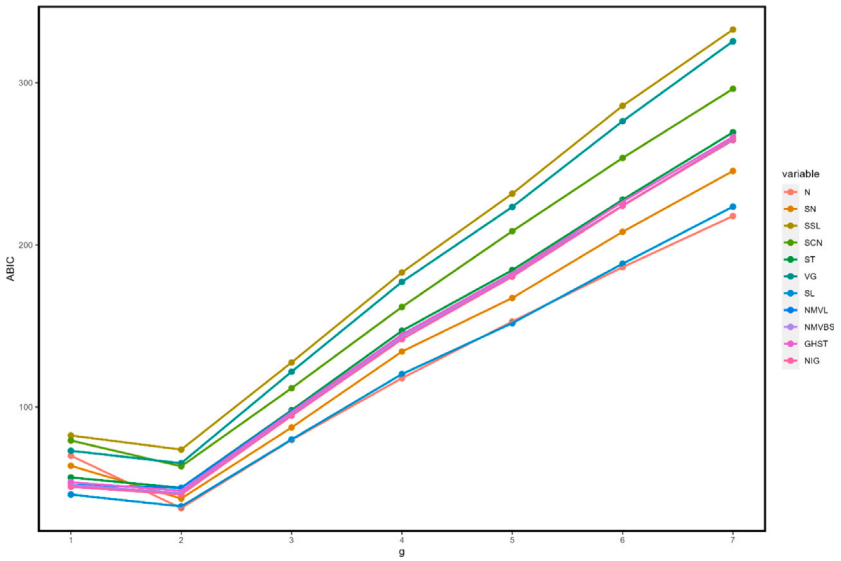


Fig. 9. Optimal cluster selection via ABIC criteria for NMVM-MoARCH models and their counterparts fitted to S&P data.

Appendix A. Model formulation

The GH distribution belongs to the NMVM family and is specified when the random mixing variable follows the GIG distribution ($W \sim GIG(\kappa, \chi, \psi)$). The GIG distribution has the following probability density function (PDF) for $w > 0$:

$$f_{GIG}(w;\kappa,\chi,\psi) = \frac{1}{2}\left(\frac{\psi}{\chi}\right)^{\frac{\kappa}{2}} \frac{w^{\kappa-1}}{K_\kappa(\sqrt{\psi\chi})} \exp\left\{-\frac{1}{2}\left(w^{-1}\chi + w\psi\right)\right\}. \tag{A.1}$$

The modified Bessel function of the third kind, $K_\kappa(y)$, is defined as:

$$K_\kappa(y) = \frac{1}{2}\int_0^\infty x^{\kappa-1} exp\big(-\frac{y}{2}(x+\frac{1}{x})\big)dx,$$

where $y > 0$ and $\kappa \in \mathbb{R}$. The parameters χ and ψ must satisfy the following conditions: (i) $\chi \geq 0$, $\psi > 0$ if $\kappa > 0$; (ii) $\chi > 0$, $\psi \geq 0$ if $\kappa < 0$; and (iii) $\chi > 0$, $\psi > 0$ if $\kappa = 0$. The expectations $E(W^r)$ and $E(\log W)$ are given by the following expressions for $r = \pm 1$:

$$E(W^r) = \big(\frac{\chi}{\psi}\big)^{\frac{r}{2}} R_{(\kappa,r)}\big(\sqrt{\chi\psi}\big) \tag{A.2}$$

and

$$E(\log W) = \log\big(\sqrt{\frac{\chi}{\psi}}\big) + \frac{1}{K_\kappa(\chi\psi)}\frac{\partial K_\kappa(\chi\psi)}{\partial\kappa}$$

where $R_{(\kappa,a)}(c) = \frac{K_{\kappa+a}(c)}{K_\kappa(c)}$. A random variable Y follows the GH distribution with the following PDF:

$$f_{GH}(y;\mu,\lambda,\sigma^2,\kappa,\chi,\psi) = \frac{\left(\frac{\psi}{\chi}\right)^{\kappa/2}\left(\psi+\frac{\lambda^2}{\sigma^2}\right)^{0.5-\kappa}}{\sqrt{2\pi\sigma^2}K_\kappa(\sqrt{\chi\psi})} \frac{K_{\kappa-0.5}\Big(\sqrt{\big(\psi+\lambda^2/\sigma^2\big)\big(\chi+\delta(y;\mu,\sigma^2)\big)}\Big)}{\Big(\sqrt{\big(\psi+\lambda^2/\sigma^2\big)\big(\chi+\delta(y;\mu,\sigma^2)\big)}\Big)^{0.5-\kappa}}$$
$$\times \exp\Big\{\frac{(y-\mu)\lambda}{\sigma^2}\Big\},$$

where $\delta(y;\mu,\sigma^2) = (y-\mu)^2/\sigma^2$.

Appendix B. Special cases of the NMVM distributions

We present the distinct members of the NMVM distributions and the necessary conditional expectations for our suggested ECME algorithm.

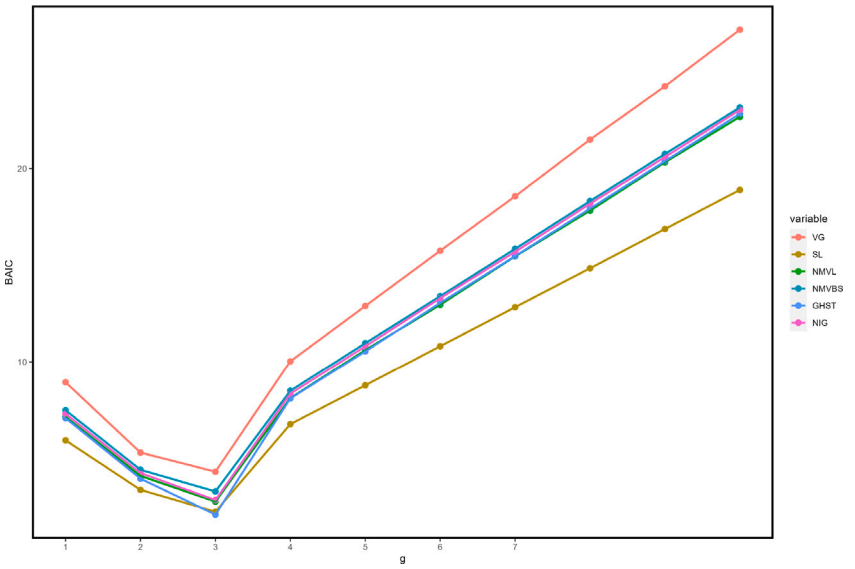


Fig. 10. Optimal cluster selection via BAIC criteria for NMVM-MoARCH models fitted to real life expectancy data.

- **The generalized hyperbolic skew-t distribution:** we present the Generalized Hyperbolic Skew-t (GHST) Distribution, a particular case of the GH Distribution. The random variable Y is said to follow the GHST distribution if its mixing random variable, W, follows an inverse gamma distribution represented by $W \sim Invgamma(\frac{\nu}{2}, \frac{\nu}{2})$. The mean and variance of this distribution are $E(W) = \frac{\nu}{\nu-2}$ and $Var(W) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}$, respectively. The GHST distribution is derived as a limiting case when $\kappa = -\frac{\nu}{2}$, $\chi = \nu$, and $\psi = 0$. The Probability Density Function (PDF) of the GHST distribution is given by:

$$f_{GHST}(y;\mu,\lambda,\sigma^2,\nu) = \Big(\frac{\lambda^2/\sigma^2}{\nu+\delta(y,\mu,\sigma^2)}\Big)^{(\nu+1)/4} \frac{\nu^{\nu/2}K_{\frac{\nu+1}{2}}\Big(\sqrt{\lambda^2(\nu+\delta(y,\mu,\sigma^2))/\sigma^2}\Big)}{\sqrt{2\pi\sigma^2}\Gamma(\nu/2)2^{\nu/2-1}exp\Big\{(\mu-y)\lambda/\sigma^2\Big\}},$$

  where $\delta(y,\mu,\sigma^2) = (y-\mu)^2/\sigma^2$. Using Bayes' rule, it can be deduced that $W|(Y=y) \sim GIG(\kappa_{ST}, \chi_{ST}, \psi_{ST})$, where $\kappa_{ST} = -\frac{(\nu+1)}{2}$, $\chi_{ST} = \delta(y,\mu,\sigma^2) + \nu$, and $\psi_{ST} = \frac{\lambda^2}{\sigma^2}$. Hence, the kth moment of W given $Y = y$ can easily be obtained from (A.2).
- **The normal inverse Gaussian distribution:** The normal inverse Gaussian (NIG) distribution can be derived from the generalized hyperbolic (GH) distribution by setting κ to -0.5, χ to 1, and ψ to ν^2. The probability density function (PDF) of the NIG distribution can be expressed as:

$$f_{NIG}(y;\mu,\lambda,\sigma^2,\nu) = \sqrt{\frac{\psi_{NIG}}{\pi^2\sigma^2\chi_{NIG}}}K_1(\sqrt{\chi_{NIG}\psi_{NIG}})exp\Big\{\frac{(y-\mu)\lambda}{\sigma^2}+\nu\Big\},$$

  where $\chi_{NIG} = 1+\delta(y,\mu,\sigma^2)$, $\psi_{NIG} = \nu^2+(\lambda/\sigma)^2$ and $K_{0.5}(y) = K_{-0.5}(y) = \sqrt{\pi/2}y^{-0.5}\exp\{-y\}$. Using Bayes' rule, we can conclude that $W|(Y=y) \sim GIG(-1, \chi_{NIG}, \psi_{NIG})$, and thus, the kth moment of W given $Y = y$ can be easily obtained using Eq. (A.2).
- **The variance-gamma distribution:** The VG distribution can be derived from the GH distribution by setting $\chi = 0$ and $\kappa > 0$. The PDF of the VG distribution is expressed as:

$$f_{VG}(y;\mu,\lambda,\sigma^2,\kappa,\psi) = \frac{\psi^\kappa K_{\kappa-0.5}\big(\sqrt{\chi_{VG}\psi_{VG}}\big)}{2^{\kappa-1}\sqrt{2\pi\sigma^2}\Gamma(\kappa)\big(\sqrt{\chi_{VG}/\psi_{VG}}\big)^{0.5-\kappa}}exp\Big\{\frac{(y-\mu)\lambda}{\sigma^2}\Big\},$$

  where $\chi_{VG} = \delta(y,\mu,\sigma^2)$ and $\psi_{VG} = \psi^2 + (\lambda/\sigma)^2$. As a result, it can be established that the distribution of $W|(Y=y)$ follows a GIG distribution with parameters $(\kappa-0.5, \chi_{VG}, \psi_{VG})$. Additionally, from Eq. (A.2), the kth moment of W given $Y = y$ for the VG distribution can be determined.

- **The skew Laplace distribution:** The Skew Laplace (SL) distribution can be obtained when the mixing random variable, denoted by W, in Eq. (2.1) follows an exponential distribution with mean 0.5, i.e., $W \sim Exp(0.5)$. The probability density function (PDF) of the SL distribution is given by:

$$f_{SL}(y;\mu,\sigma^2,\lambda) = \frac{1}{2\tau\sigma} exp\Big\{-\tau\frac{|y-\mu|}{\sigma} + \frac{\lambda(y-\mu)}{\sigma^2}\Big\}$$

where $\tau = \sqrt{1+(\frac{\lambda}{\sigma})^2}$. By applying Bayes' rule, it can be concluded that $W|(Y=y) \sim GIG(0.5, \chi_{ST}, \psi_{ST})$, where $\chi_{ST} = \delta(y,\mu,\sigma^2)$ and $\psi_{ST} = k^2$. Hence, the kth moment of W given $Y = y$ can be obtained using Eq. (A.2).

- **The normal mean–variance Birnbaum–Saunders distribution:** This distribution is obtained by assuming that the mixing random variable W in (2.1) has a Birnbaum–Saunders (BS) distribution [7], with shape and scale parameters α and β denoted by $W \sim \text{BS}(\alpha,\beta)$. The PDF of the BS distribution can be expressed as a mixture of two GIG distributions, as given in [14]:

$$f_{BS}(w;\alpha,\beta) = \frac{1}{2} f_{GIG}\big(w;\frac{1}{2},\frac{\beta}{\alpha^2},\frac{1}{\alpha\beta^2}\big) + \frac{1}{2} f_{GIG}\big(w;-\frac{1}{2},\frac{\beta}{\alpha^2},\frac{1}{\alpha\beta^2}\big).$$

If we assume that $W \sim BS(\alpha,1)$, we obtain the normal mean–variance Birnbaum–Saunders distribution [32], denoted by $f_{NMVBS}(y;\mu,\lambda,\sigma^2,\alpha)$, with the following PDF:

$$f_{NMVBS}(y;\mu,\lambda,\sigma^2,\alpha) = \frac{1}{2} f_{GH}\big(y;\mu,\lambda,\sigma^2,\frac{1}{2},\frac{1}{\alpha^2},\frac{1}{\alpha^2}\big) + f_{GH}\big(y;\mu,\lambda,\sigma^2,-\frac{1}{2},\frac{1}{\alpha^2},\frac{1}{\alpha^2}\big).$$

For any integer $r = \pm1, \pm2, \ldots$, the rth moment of W given $Y = y$ can be calculated as follows:

$$E(W^r|Y=y) = \Big(\frac{\chi_{BS}}{\psi_{BS}}\Big)^{k/2}\Big\{p(y)R(0,r)\Big(\sqrt{\chi_{BS}\psi_{BS}}\Big) + (1-p(y))R(-1,r)\Big(\sqrt{\chi_{BS}\psi^*_{BS}}\Big)\Big\},$$

where $\chi_{BS} = \delta(y,\mu,\sigma^2) + \alpha^{-2}$, $\psi_{BS} = (\lambda/\sigma)^2 + \alpha^{-2}$, and

$$p(y) = \frac{f_{GH}\big(y;\mu,\lambda,\sigma^2,\frac{1}{2},\frac{1}{\alpha^2},\frac{1}{\alpha^2}\big)}{f_{GH}\big(y;\mu,\lambda,\sigma^2,\frac{1}{2},\frac{1}{\alpha^2},\frac{1}{\alpha^2}\big) + f_{GH}\big(y;\mu,\lambda,\sigma^2,-\frac{1}{2},\frac{1}{\alpha^2},\frac{1}{\alpha^2}\big)}.$$

- **The normal mean–variance Lindley distribution:** The Normal Mean-Variance Lindley (NMVL) distribution, introduced in [31], describes a random variable Y where the mixing random variable W follows a Lindley distribution [24]. The probability density function (PDF) of the Lindley distribution is proportional to $\alpha^2(1+w^2)e^{-\alpha w}$, where α is a positive parameter and w is positive. The PDF of Y is expressed as the mixture of two Generalized Hyperbolic (GH) distributions:

$$f_{NMVL}(y;\mu,\sigma^2,\lambda,\alpha) = \frac{\alpha}{\alpha+1} f_{GH}\big(y;\mu,\lambda,\sigma^2,1,0,2\alpha\big) + \frac{1}{\alpha+1} f_{GH}\big(y;\mu,\lambda,\sigma^2,2,0,2\alpha\big)$$

The expected value of W^r conditioned on $Y = y$ can be calculated for $r = \pm1, \pm2, \ldots$ as follows:

$$E(W^r|Y=y) = \Big(\frac{\chi_{LI}}{\psi_{LI}}\Big)^{k/2}\Big\{p(y)R(0.5,r)\Big(\sqrt{\chi_{LI}\psi_{LI}}\Big) + (1-p(y))R(1.5,r)\Big(\sqrt{\chi_{LI}\zeta_{LI}}\Big)\Big\},$$

where $\chi_{LI} = \delta(y,\mu,\sigma^2)$, $\psi_{LI} = (\lambda/\sigma)^2 + 2\alpha$, and

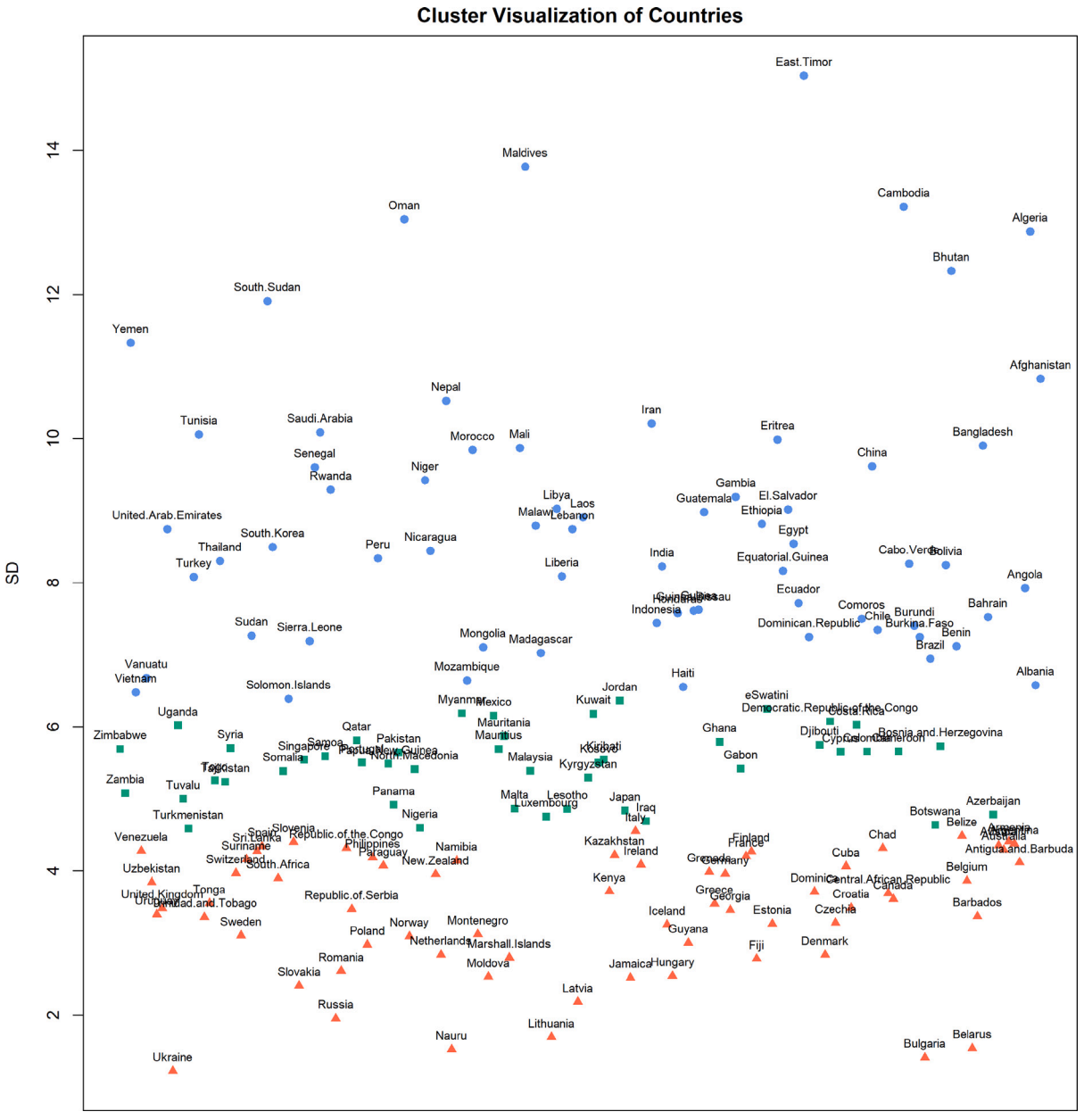
$$p(y) = \frac{\alpha f_{GH}\big(y;\mu,\lambda,\sigma^2,1,0,2\alpha\big)}{\alpha f_{GH}\big(y;\mu,\lambda,\sigma^2,1,0,2\alpha\big) + f_{GH}\big(y;\mu,\lambda,\sigma^2,1,0,2\alpha\big)}.$$

In the paper by Browne and McNicholas [11], a new parameterization was proposed for the Generalized Inverse Gaussian (GIG) distribution due to an identifiability issue. The issue is that $f_{GH}\big(y;\mu,\lambda,\sigma^2,\kappa,\chi,\psi\big)$ is equal to $f_{GH}\big(y;\mu,a\lambda,a\sigma^2,\kappa,\chi/a,a\psi\big)$ for any positive constant a, which leads to an identifiability problem. To resolve this, [11] set $\chi = \omega\eta$ and $\psi = \omega/\eta$ and showed that the finite mixture of GH distributions is identifiable with this new parameterization. This reparameterization is also consistent with the notation used in the paper by McNeil et al. [29].

Appendix C. Optimal number of clusters

See Figs. 9 and 10.

Appendix D. The clustering of life expectancy



References

[1] A.C. Aitken, Xxv.—on bernoulli's numerical solution of algebraic equations, Proc. Roy. Soc. Edinburgh 46 (1927) 289–305.
[2] A. Asilkalkan, X. Zhu, S. Sarkar, Finite mixture of hidden Markov models for tensor-variate time series data, Adv. Data Anal. Classif. (2023) 1–18.
[3] A. Azzalini, Distributions generated by pertubation of symmetry with emphasis on a multivariate skew t distribution, J. R. Stat. Soc. B 65 (2003) 579–602.
[4] O. Barndorff-Nielsen, C. Halgreen, Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 38 (4) (1977) 309–311.
[5] K. Basford, D. Greenway, G. McLachlan, D. Peel, Standard errors of fitted component means of normal mixtures, Comput. Statist. 12 (1) (1997) 1–18.
[6] E.J. Bedrick, C.-L. Tsai, Model selection for multivariate regression in small samples, Biometrics (1994) 226–231.
[7] Z.W. Birnbaum, S.C. Saunders, A new family of life distributions, J. Appl. Probab. 6 (2) (1969) 319–327.
[8] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (3) (1986) 307–327.
[9] T. Bollerslev, J. Russell, M. Watson, Volatility and Time Series Econometrics: Essays in Honor of Robert Engle, Oxford University Press, 2010.

[10] M.D. Branco, D.K. Dey, A general class of multivariate skew-elliptical distributions, J. Multivariate Anal. 79 (1) (2001) 99–113.
[11] R.P. Browne, P.D. McNicholas, A mixture of generalized hyperbolic distributions, Canad. J. Statist. 43 (2) (2015) 176–198.
[12] I. Cadez, D. Heckerman, C. Meek, P. Smyth, S. White, Visualization of navigation patterns on a web site using model-based clustering, in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp. 280–284.
[13] A. Capitanio, A. Azzalini, E. Stanghellini, Graphical models for skew-normal variates, Scand. J. Stat. 30 (1) (2003) 129–144.
[14] A.F. Desmond, On the relationship between two fatigue-life models, IEEE Trans. Reliab. 35 (2) (1986) 167–169.
[15] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica (1982) 987–1007.
[16] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc. 96 (456) (2001) 1348–1360.
[17] S. Gaffney, P. Smyth, Trajectory clustering with mixtures of regression models, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999, pp. 63–72.
[18] S. Ghassempour, F. Girosi, A. Maeder, Clustering multivariate time series using hidden Markov models, Int. J. Environ. Res. Public Health 11 (3) (2014) 2741–2763.
[19] L.R. Glosten, R. Jagannathan, D.E. Runkle, On the relation between the expected value and the volatility of the nominal excess return on stocks, J. Finance 48 (5) (1993) 1779–1801.
[20] I.J. Good, The population frequencies of species and the estimation of population parameters, Biometrika 40 (3–4) (1953) 237–264.
[21] P. Hall, Q. Yao, Inference in ARCH and GARCH models with heavy–tailed errors, Econometrica 71 (1) (2003) 285–317.
[22] S. Helske, J. Helske, Mixture hidden Markov models for sequence data: The seqHMM package in R, J. Stat. Softw. 88 (2019).
[23] L. Hubert, P. Arabie, Comparing partitions, J. Classification 2 (1985) 193–218.
[24] D.V. Lindley, Fiducial distributions and Bayes' theorem, J. R. Stat. Soc. Ser. B Stat. Methodol. (1958) 102–107.
[25] C. Liu, D.B. Rubin, The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence, Biometrika 81 (4) (1994) 633–648.
[26] T.A. Louis, Finding the observed information matrix when using the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (2) (1982) 226–233.
[27] G.J. McLachlan, K.E. Basford, Mixture Models: Inference and Applications to Clustering, vol. 38, M. Dekker New York, 1988.
[28] G.J. McLachlan, S.X. Lee, S.I. Rathnayake, Finite mixture models, Annu. Rev. Stat. Appl. 6 (2019) 355–378.
[29] A.J. McNeil, R. Frey, P. Embrechts, Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition, Princeton University Press, 2015.
[30] X.-L. Meng, D.B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework, Biometrika 80 (2) (1993) 267–278.
[31] M. Naderi, A. Arabpour, A. Jamalizadeh, Multivariate normal mean-variance mixture distribution based on lindley distribution, Comm. Statist. Simulation Comput. 47 (4) (2018) 1179–1192.
[32] R. Pourmousa, A. Jamalizadeh, M. Rezapour, Multivariate normal mean–variance mixture distribution based on Birnbaum–Saunders distribution, J. Stat. Comput. Simul. 85 (13) (2015) 2736–2749.
[33] L.R. Rabiner, C.-H. Lee, B.-H. Juang, J.G. Wilpon, HMM clustering for connected word recognition, in: International Conference on Acoustics, Speech, and Signal Processing, IEEE, 1989, pp. 405–408.
[34] M. Ramoni, P. Sebastiani, P. Cohen, Bayesian clustering by dynamics, Mach. Learn. 47 (2002) 91–121.
[35] B. Ren, I. Barnett, Autoregressive mixture models for clustering time series, J. Time Series Anal. 43 (6) (2022) 918–937.
[36] G. Ridgeway, Finite Discrete Markov Process Clustering, Technical Report, Technical Report TR 97-24, Microsoft Research, Redmond, WA, 1997.
[37] S.L. Sclove, Application of model-selection criteria to some problems in multivariate analysis, Psychometrika 52 (1987) 333–343.
[38] P. Sebastiani, M. Ramoni, P. Cohen, J. Warwick, J. Davis, Discovering dynamics using Bayesian clustering, in: Advances in Intelligent Data Analysis: Third International Symposium, IDA-99 Amsterdam, the Netherlands, August 9–11, 1999 Proceedings 3, Springer, 1999, pp. 199–209.
[39] F. Setoudehtazangi, et al., Robusti modelli di miscele per la modellazione e il clustering di dati di serie storiche complessi, Università degli studi di Padova, 2023.
[40] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[41] R. Umatani, T. Imai, K. Kawamoto, S. Kunimasa, Time series clustering with an EM algorithm for mixtures of linear Gaussian state space models, Pattern Recognit. 138 (2023) 109375.
[42] F. Vilca, N. Balakrishnan, C.B. Zeller, Multivariate skew-normal generalized hyperbolic distribution and its properties, J. Multivariate Anal. 128 (2014) 73–85.
[43] N.X. Vinh, J. Epps, J. Bailey, Information theoretic measures for clusterings comparison: is a correction for chance necessary? in: Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp. 1073–1080.
[44] H. Wang, R. Li, C.-L. Tsai, Tuning parameter selectors for the smoothly clipped absolute deviation method, Biometrika 94 (3) (2007) 553–568.
[45] Y. Xiong, D.-Y. Yeung, Time series clustering with ARMA mixtures, Pattern Recognit. 37 (8) (2004) 1675–1689.