



HAL
open science

A mixture integer-valued GARCH model

Mamadou Lamine Diop, Aliou Diop, Abdou Ka Diongue

► **To cite this version:**

Mamadou Lamine Diop, Aliou Diop, Abdou Ka Diongue. A mixture integer-valued GARCH model. 2013. hal-00831983

HAL Id: hal-00831983

<https://hal.science/hal-00831983>

Preprint submitted on 9 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A mixture integer-valued GARCH model

Mamadou Lamine DIOP

LERSTAD, Université Gaston Berger, Saint Louis, Sénégal.

Email: diopml@yahoo.fr

Aliou DIOP†

LERSTAD, Université Gaston Berger, Saint Louis, Sénégal.

Email: aliou.diop@ugb.edu.sn

Abdou Ka DIONGUE

LERSTAD, Université Gaston Berger, Saint Louis, Sénégal.

Email: abdou.diongue@ugb.edu.sn

Abstract. In this paper, we generalize the mixture integer-valued ARCH model (MINARCH) introduced by Zhu et al. (2010) to a mixture integer-valued GARCH (MINGARCH) for modeling time series of counts. This model include the ability to take into account the moving average (MA) components of the series. We give the necessary and sufficient first and second order stationarity conditions. The estimation is done via the EM algorithm. The model selection problem is studied by using three information criterions. We also study the performance of the method via simulations and include a real data application.

Keywords: Integer-valued · Mixture models · GARCH · EM algorithm.

1. Introduction

Time series count data are widely observed in real-world applications (epidemiology, econometrics, insurance ...). Many different approaches have been proposed to model time series count data, which are able to describe different types of marginal distribution. Zeger (1988) discusses a model for regression analysis with a time series of counts by illustrating the technique with an analysis of trends in U.S. polio incidence, Ferland et al. (2006) proposed an integer-valued autoregressive conditional heteroscedastic (INARCH) model to deal with integer-valued time series with overdispersion. Zhu (2011) propose a negative binomial INGARCH (NBINGARCH) model that can deal with both overdispersion and potential extreme observations simultaneously. Zhu (2012) introduce a generalized Poisson INGARCH model, which can account for both overdispersion and underdispersion, among others. Some extensions of the classical mixture models

†Corresponding author

to time series have been studied by many authors. For example Le et al. (1996) introduced the Gaussian mixture transition distribution (GMTD) models to capture the flat stretches.

In the literature, time series are often assumed to be driven by a unimodal innovation series. However, many time series may exhibit multimodality either in the marginal or the conditional distribution. For example, Martin (1992) proposed to model multimodal jump phenomena by a multipredictor autoregressive time series (MATS) model, Wong and Li (2000) generalized the GMTD model to the full mixture autoregressive (MAR) model whose predictive distribution could also be multimodal. Muller and Sawitzki (1991) propose and study a method for analyzing the modality of a distribution.

Recently, Zhu et al. (2010) generalize the INARCH model to the mixture (MINARCH) model, which has the advantages over the INARCH model because of its ability to handle multimodality and non-stationary components. But in their framework, they did not take into account the MA part of the model. Our main objective is to include the Moving Average (MA) part for the previous model, which leads to a generalized MINARCH model, namely the mixture MINGARCH model.

The paper is organized as follows. In Section 2 we describe the MINGARCH model and the stationarity conditions. The estimation procedures with an EM algorithm are discussed in Section 3. A simulation study is presented in Section 4. We illustrate the usefulness of the model in Section 5 by an empirical example. A brief discussion and concluding remarks are given in Section 6.

2. The mixture integer-valued GARCH model

The MINGARCH($K; p_1, \dots, p_K; q_1, \dots, q_K$) model is defined by :

$$\begin{cases} X_t = \sum_{k=1}^K \mathbf{1}(\eta_t = k) Y_{kt}, \\ Y_{kt} | \mathcal{F}_{t-1} : \mathcal{P}(\lambda_{kt}), \\ \lambda_{kt} = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki} X_{t-i} + \sum_{j=1}^{q_k} \beta_{kj} \lambda_{k(t-j)}, \end{cases} \quad (2.1)$$

with $\alpha_{k0} > 0$, $\alpha_{ki} \geq 0$, $\beta_{kj} \geq 0$, ($i = 1, \dots, p_k$, $j = 1, \dots, q_k$, $k = 1, \dots, K$).

where $\mathbf{1}(\cdot)$ denotes the indicator function, p_k and q_k are respectively the order of AR and MA for the k -th component, \mathcal{F}_{t-1} indicates the information given up to time $t-1$, η_t is a sequence of independent and identically distributed random variables with $\mathbb{P}(\eta_t = k) = \alpha_k$, $k = 1, \dots, K$. It is assumed that X_{t-j}

and η_t are independent for all t and $j > 0$, the variables Y_{kt} and η_t are conditionally independent given \mathcal{F}_{t-1} , $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_K$ for identifiability (see Titterton et al. (1985)) and $\sum_{k=1}^K \alpha_k = 1$.

The MINGARCH model is able to handle the conditional overdispersion in integer-valued time series. In fact, the conditional mean and variance are given by

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = \sum_{k=1}^K \alpha_k \lambda_{kt},$$

and

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = \mathbb{E}(X_t | \mathcal{F}_{t-1}) + \sum_{k=1}^K \alpha_k \lambda_{kt}^2 - \left(\sum_{k=1}^K \alpha_k \lambda_{kt} \right)^2.$$

This shows that we can have a strict inequality between the conditional mean and variance. Furthermore

$$\begin{aligned} \text{Var}(X_t) &= \mathbb{E}(\text{Var}(X_t | \mathcal{F}_{t-1})) + \text{Var}(\mathbb{E}(X_t | \mathcal{F}_{t-1})) \\ &= \mathbb{E} \left(\sum_{k=1}^K \alpha_k \lambda_{kt} + \sum_{k=1}^K \alpha_k \lambda_{kt}^2 - \left(\sum_{k=1}^K \alpha_k \lambda_{kt} \right)^2 \right) + \text{Var} \left(\sum_{k=1}^K \alpha_k \lambda_{kt} \right) \\ &= \mathbb{E}(X_t) + \sum_{k=1}^K \alpha_k \mathbb{E}(\lambda_{kt}^2) - \left(\mathbb{E}(X_t) \right)^2. \end{aligned}$$

Using the convexity inequality concerning the expectation, we can easily see that in general the variance is larger than the mean, which indicates that the MINGARCH model is also able to describe the time series count with overdispersion.

Let us now introduce the polynomials $D(B) = 1 - \beta_{k1}B - \beta_{k2}B^2 - \dots - \beta_{kq}B^q$, where B is the backshift operator. Suppose, from now on, that the roots of $D(z) = 0$ lie outside the unit circle which, for non-negative β_{kj} , is equivalent to saying that $\sum_{j=1}^q \beta_{kj} < 1$, for any fixed k . In the following, we consider : $p = \max(p_1, \dots, p_K)$; $q = \max(q_1, \dots, q_K)$; $\alpha_{ki} = 0$, for $i > p_k$; $\beta_{kj} = 0$, for $j > q_k$ and $L = \max(p, q)$.

The first- and second-order stationarity conditions for the MINGARCH model (2.1) are given in Theorem 2.1 and 2.3. The proof of the theorems is similar to that of Zhu et al. (2010).

THEOREM 2.1.

The necessary and sufficient condition for the existence of stationary solution is

that all roots of the equation :

$$1 - \sum_{i=1}^L \sum_{k=1}^K \alpha_k \alpha_{ki} Z^i - \sum_{l=1}^{\infty} \sum_{j_1 \dots j_{l+1}=1}^L \sum_{k=1}^K \alpha_k \alpha_{k j_{l+1}} \beta_{k j_1} \dots \beta_{k j_l} Z^{(j_1 + j_2 + \dots + j_{l+1})} = 0 \quad (2.2)$$

lie outside the unit circle.

Proof:

Let $\mu_t = \mathbb{E}(X_t) = \sum_{k=1}^K \alpha_k \mathbb{E}(\lambda_{kt})$, for all $t \in \mathbb{Z}$.

If the process is first-order stationary, we have $\mu_t = \mu$.

Since $\lambda_{kt} = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki} X_{t-i} + \sum_{j=1}^{q_k} \beta_{kj} \lambda_{k(t-j)}$

The recursion equation give, for all $m > 1$,

$$\begin{aligned} \lambda_{kt} &= \alpha_{k0} + \sum_{i=1}^L \alpha_{ki} X_{t-i} + \sum_{l=1}^m \sum_{j_1, \dots, j_l=1}^L \alpha_{k0} \beta_{k j_1} \dots \beta_{k j_l} \\ &\quad + \sum_{l=1}^m \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{k j_{l+1}} \beta_{k j_1} \dots \beta_{k j_l} X_{t-j_1 - \dots - j_{l+1}} \\ &\quad + \sum_{j_1, \dots, j_{m+1}=1}^L \beta_{k j_1} \dots \beta_{k j_{m+1}} \lambda_{k(t-j_1 - \dots - j_{m+1})} \end{aligned}$$

Let $C_{k0} = \alpha_{k0} + \sum_{l=1}^{\infty} \sum_{j_1, \dots, j_l=1}^L \alpha_{k0} \beta_{k j_1} \dots \beta_{k j_l}$. We define

$$\lambda'_{kt} = C_{k0} + \sum_{i=1}^L \alpha_{ki} X_{t-i} + \sum_{l=1}^{\infty} \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{k j_{l+1}} \beta_{k j_1} \dots \beta_{k j_l} X_{t-j_1 - j_2 - \dots - j_{l+1}}.$$

Since $\sum_{j=1}^L \beta_{kj} < 1$ it is easy to see that $0 \leq \lambda'_{kt} < \infty$ a.s. for any fixed t and k .

We will show below that $\lambda_{kt} = \lambda'_{kt}$ almost surely as $m \rightarrow \infty$ for any fixed t and k . In what follows, C will denote any positive constants whose value is unimportant and may vary from line to line. Let t and k be fixed now. It follows that for any $m \geq 1$

$$\begin{aligned} |\lambda_{kt} - \lambda'_{kt}| &\leq \sum_{l=m+1}^{\infty} \sum_{j_1 \dots j_l=1}^L \alpha_{k0} \beta_{k j_1} \dots \beta_{k j_l} + \sum_{l=m+1}^{\infty} \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{k j_{l+1}} \beta_{k j_1} \dots \beta_{k j_l} X_{t-j_1 - j_2 - \dots - j_{l+1}} \\ &\quad + \sum_{j_1, \dots, j_{m+1}=1}^L \beta_{k j_1} \dots \beta_{k j_{m+1}} \lambda_{k(t-j_1 - \dots - j_{m+1})}. \end{aligned}$$

First

$$E \left\{ \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{k j_{l+1}} \beta_{k j_1} \dots \beta_{k j_l} X_{t-j_1 - j_2 - \dots - j_{l+1}} \right\} \leq C \left(\sum_{j=1}^L \beta_{kj} \right)^l$$

and

$$E \left\{ \sum_{j_1, \dots, j_{m+1}=1}^L \beta_{kj_1} \cdots \beta_{kj_{m+1}} \lambda_{k(t-j_1-\dots-j_{m+1})} \right\} \leq C \left(\sum_{j=1}^L \beta_{kj} \right)^{m+1}$$

The expectation of the right-hand side of the above is bounded by

$$\left(C_{k0} + C_1 \left(1 - \sum_{j=1}^L \beta_{kj} \right)^{-1} \right) \left(\sum_{j=1}^L \beta_{kj} \right)^{m+1}$$

Let $A_m = \{ |\lambda_{kt} - \lambda'_{kt}| > \frac{1}{m} \}$. Then

$$\mathbb{P}(A_m) \leq m \left(C_{k0} + C_1 \left(1 - \sum_{j=1}^L \beta_{kj} \right)^{-1} \right) \left(\sum_{j=1}^L \beta_{kj} \right)^{m+1}.$$

Then, using Borel-Cantelli lemma and the fact that $A_m \subset A_{m+1}$, we can show that $\lambda_{kt} = \lambda'_{kt}$ a.s. Therefore,

$$\mu_t = \sum_{k=1}^K \alpha_k C_{k0} + \sum_{i=1}^L \sum_{k=1}^K \alpha_k \alpha_{ki} \mu_{t-i} + \sum_{l=1}^{\infty} \sum_{j_1 \dots j_{l+1}=1}^L \sum_{k=1}^K \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} \mu_{t-j_1-j_2-\dots-j_{l+1}}$$

The equation can be rewritten as :

$$\left(1 - \sum_{i=1}^L \sum_{k=1}^K \alpha_k \alpha_{ki} B^i - \sum_{l=1}^{\infty} \sum_{j_1 \dots j_{l+1}=1}^L \sum_{k=1}^K \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} B^{(j_1+j_2+\dots+j_{l+1})} \right) \mu_t = \sum_{k=1}^K \alpha_k C_{k0}$$

where B is the backward shift operator.

The necessary and sufficient condition for the existence of stationary solution is that all roots of the equation :

$$1 - \sum_{i=1}^L \sum_{k=1}^K \alpha_k \alpha_{ki} Z^i - \sum_{l=1}^{\infty} \sum_{j_1 \dots j_{l+1}=1}^L \sum_{k=1}^K \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} Z^{(j_1+j_2+\dots+j_{l+1})} = 0$$

lie outside the unit circle (Goldberg (1958)). □

COROLLARY 2.2. *A necessary and sufficient condition for the MINGARCH($K; 1, \dots, 1; 1, \dots, 1$) model to be stationary in the mean is that the roots of the equation*

$$1 + C_1 Z + C_2 Z^2 + \dots + C_K Z^K = 0$$

lie outside the unit circle where

$$C_1 = - \sum_{k=1}^K \left(\delta_k + \alpha_k \gamma_k \right)$$

and

$$C_j = (-1)^j \left[\sum_{k_1 > k_2 > \dots > k_j}^K \delta_{k_1} \delta_{k_2} \dots \delta_{k_j} + \sum_{k=1}^K \alpha_k \gamma_k \left(\sum_{\substack{k_1 > k_2 > \dots > k_{j-1} \\ k_1 \neq k, k_2 \neq k, \dots, k_{j-1} \neq k}}^K \delta_{k_1} \delta_{k_2} \dots \delta_{k_{j-1}} \right) \right]$$

for $j = 2, \dots, K$, with $\gamma_k = \alpha_{k1}$ and $\delta_k = \beta_{k1}$.

Proof :

The equation (2.2) becomes

$$1 - \sum_{k=1}^K \sum_{l=1}^{\infty} \alpha_k \gamma_k \delta_k^{l-1} Z^l = 0. \quad (2.3)$$

Assuming that $\|\delta_k Z\| < 1$, we have :

$$\sum_{k=1}^K \sum_{l=1}^{\infty} \alpha_k \gamma_k \delta_k^{l-1} Z^l = \sum_{k=1}^K \frac{\alpha_k \gamma_k Z}{1 - \delta_k Z}.$$

The equation (2.3) gives : $1 - \sum_{k=1}^K \frac{\alpha_k \gamma_k Z}{1 - \delta_k Z} = 0$.

For $k = 1, \dots, K$, the preceding equation is equivalent to:

$$\prod_{k=1}^K (1 - \delta_k Z) - \sum_{k=1}^K \alpha_k \gamma_k Z \prod_{\substack{k'=1 \\ k' \neq k}}^K (1 - \delta_{k'}) Z = 1 + C_1 Z + C_2 Z^2 + \dots + C_L Z^K = 0. \quad \square$$

THEOREM 2.3.

Suppose that the process X_t following a MINGARCH($K; p_1, \dots, p_K; q_1, \dots, q_K$) model is first-order stationary. A necessary and sufficient condition for the process to be second-order stationary is that all roots of $1 - c_1 Z - c_2 Z^2 - \dots - c_L Z^L =$

0 lie outside the unit circle, where

$$c_u = \sum_{k=1}^K \alpha_k \left(\Delta_{k,u} - \sum_{v=1}^{L-1} \Lambda_{kv} b_{vu} \omega_{u0} \right), \quad u = 1, \dots, L-1 \text{ and } c_L = \sum_{k=1}^K \alpha_k \Delta_{k,L}$$

with

$$\Delta_{k,i} = \sum_{l=0}^{\infty} \sum_{j_1+\dots+j_l=i} \alpha_{kj_{l+1}}^2 \beta_{kj_1}^2 \dots \beta_{kj_l}^2,$$

$$\Lambda_{kv} = \sum_{l \neq l'=0}^{\infty} \sum_{|j_1+j_2+\dots+j_{l+1}-j'_1-j'_2-\dots-j'_{l'+1}|=v} \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} \alpha_{kj'_{l'+1}} \beta_{kj'_1} \dots \beta_{kj'_{l'}},$$

and $B = (\omega_{ij})_{i,j=1}^{L-1}$, $B^{-1} = (b_{ij})_{i,j=1}^{L-1}$, two matrices such that

$$\omega_{i0} = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{i0kl}, \quad \omega_{iu} = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{iukl} \text{ for } u \neq i, \quad \omega_{ii} = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{iikl} - 1,$$

$$\delta_{iukl} = \sum_{|i-j_1-\dots-j_{l+1}|=u} \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l}.$$

Proof :

Let $\gamma_{it} = E(X_t X_{t-i})$ for $i = 0, 1, \dots, L$,

$$\begin{aligned} \gamma_{it} &= \sum_{i=1}^K \alpha_k E(\lambda_{kt} X_{t-i}) \\ &= \sum_{k=1}^K \alpha_{k_0} \alpha_k E(X_{t-i}) + \sum_{l=1}^m \sum_{k=1}^K \sum_{j_1, \dots, j_l=1}^L \alpha_{k_0} \alpha_k \beta_{kj_1} \dots \beta_{kj_l} E(X_{t-i}) \\ &+ \sum_{l=1}^m \sum_{k=1}^K \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_k \alpha_{k_{j+1}} \beta_{kj_1} \dots \beta_{kj_l} E(X_{t-j_1-\dots-j_{l+1}} X_{t-i}) \\ &+ \sum_{k=1}^K \sum_{j_1, \dots, j_{m+1}=1}^L \alpha_k \beta_{kj_1} \dots \beta_{kj_{m+1}} E(\lambda_{k(t-j_1-\dots-j_{m+1})} X_{t-i}). \end{aligned}$$

Using the same arguments as in the proof of Theorem 2.1, we can show that

almost surely

$$\begin{aligned}
\gamma_{it} &= \sum_{k=1}^K \alpha_{k_0} \alpha_k E(X_{t-i}) + \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{j_1, \dots, j_l=1}^L \alpha_{k_0} \alpha_k \beta_{k_{j_1}} \dots \beta_{k_{j_l}} E(X_{t-i}) \\
&+ \sum_{k=1}^K \sum_{j=1}^L \alpha_{k_j} \alpha_k E(X_{t-j} X_{t-i}) \\
&+ \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_k \alpha_{k_{j_{l+1}}} \beta_{k_{j_1}} \dots \beta_{k_{j_l}} E(X_{t-j_1-\dots-j_{l+1}} X_{t-i}) \\
&= I + II + III + IV
\end{aligned}$$

with

$$\begin{aligned}
III &= \sum_{k=1}^K \sum_{j=1}^L \alpha_{k_j} \alpha_k E(X_{t-j} X_{t-i}) \\
&= \sum_{k=1}^K \alpha_{ki} \alpha_k \gamma_{0,t-i} + \sum_{k=1}^K \sum_{j=1, i \neq j}^L \alpha_{kj} \alpha_k \gamma_{|j-i|,t} \\
&= \sum_{k=1}^K \alpha_{ki} \alpha_k \gamma_{0,t-i} + \sum_{k=1}^K \alpha_k \left(\sum_{|j-i|=1} \beta_{ki} \gamma_{1,t} + \dots + \sum_{|j-i|=i} \alpha_{kj} \gamma_{i,t} + \dots + \sum_{|j-i|=L-1} \alpha_{kj} \gamma_{L-1,t} \right) \\
&= \sum_{k=1}^K \alpha_k \delta_{i0k0} \gamma_{0,t-i} + \sum_{k=1}^K \sum_{u=1}^{L-1} \alpha_k \delta_{iuk0} \gamma_{u,t}
\end{aligned}$$

and

$$\begin{aligned}
IV &= \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_k \alpha_{k_{j_{l+1}}} \beta_{k_{j_1}} \dots \beta_{k_{j_l}} \gamma_{|i-j_1-\dots-j_{l+1}|,t} \\
&= \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{j_1+\dots+j_{l+1}=i}^L \alpha_k \alpha_{k_{j_{l+1}}} \beta_{k_{j_1}} \dots \beta_{k_{j_l}} \gamma_{0,t-i} \\
&+ \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{j_1+\dots+j_{l+1} \neq i}^L \alpha_k \alpha_{k_{j_{l+1}}} \beta_{k_{j_1}} \dots \beta_{k_{j_l}} \gamma_{|i-j_1-\dots-j_{l+1}|,t} \\
&= \sum_{l=1}^{\infty} \sum_{k=1}^K \alpha_k \delta_{i0kl} \gamma_{0,t-i} + \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{u=1}^{L-1} \alpha_k \delta_{iukl} \gamma_{u,t}
\end{aligned}$$

where

$$\delta_{iukl} = \sum_{|i-j_1-\dots-j_{l+1}|=u} \alpha_{k_{j_{l+1}}} \beta_{k_{j_1}} \dots \beta_{k_{j_l}}.$$

Then

$$III + IV = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{i0kl} \gamma_{0,t-i} + \sum_{l=0}^{\infty} \sum_{k=1}^K \sum_{u=1}^{L-1} \alpha_k \delta_{iukl} \gamma_{u,t}$$

where the first term of this summation ($l = 0$) is III .

Moreover, using the same notation, we get

$$\begin{aligned} I + II &= \left(\sum_{k=1}^K \alpha_{k_0} \alpha_k + \sum_{l=1}^{\infty} \sum_{k=1}^K \sum_{j_1, \dots, j_l=1}^L \alpha_{k_0} \alpha_k \beta_{kj_1} \dots \beta_{kj_l} \right) \mu \\ &= \left(\sum_{l=0}^{\infty} \sum_{k=1}^K \sum_{j_1, \dots, j_l=1}^L \alpha_{k_0} \alpha_k \beta_{kj_1} \dots \beta_{kj_l} \right) \mu =: K_1 \end{aligned}$$

Finally, for $i = 1, \dots, L$

$$K_1 + \omega_{i0} \gamma_{0,t-i} + \sum_{u=1}^{L-1} \omega_{iu} \gamma_{u,t} = 0$$

where

$$\omega_{i0} = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{i0kl}, \quad \omega_{iu} = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{iukl} \text{ for } u \neq i \text{ and } \omega_{ii} = \sum_{l=0}^{\infty} \sum_{k=1}^K \alpha_k \delta_{iikl} - 1.$$

Let $B = (\omega_{ij})_{i,j=1}^{L-1}$ and $B^{-1} = (b_{ij})_{i,j=1}^{L-1}$

Then

$$B(\gamma_{1,t}, \dots, \gamma_{L-1,t})^T = -(K_1 + \omega_{10} \gamma_{0,t-1}, \dots, K_1 + \omega_{(L-1)0} \gamma_{0,t-(L-1)})$$

which is equivalent to

$$(\gamma_{1,t}, \dots, \gamma_{L-1,t})^T = -B^{-1}(K_1 + \omega_{10} \gamma_{0,t-1}, \dots, K_1 + \omega_{(L-1)0} \gamma_{0,t-(L-1)}).$$

We can show that

$$\gamma_{i,t} = -K_1 \sum_{u=1}^{L-1} b_{iu} - \sum_{u=1}^{L-1} b_{iu} \omega_{u0} \gamma_{0,t-u}.$$

The conditional second moment is given by :

$$\begin{aligned}
\gamma_{0,t} &= \mathbb{E}(X_t) + \sum_{k=1}^K \alpha_k \mathbb{E}(\lambda_{kt}^2) \\
&= \mu + \sum_{k=1}^K \alpha_k \mathbb{E} \left(C_{k0} + \sum_{i=1}^L \alpha_{ki} X_{t-i} + \sum_{l=1}^{\infty} \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} X_{t-j_1-\dots-j_{l+1}} \right)^2 \\
&= C_0 + \sum_{k=1}^K \alpha_k \left(\sum_{i=1}^L \alpha_{ki}^2 \gamma_{0,t-i} + 2 \sum_{l=1}^{\infty} \sum_{i, j_1, \dots, j_{l+1}=1}^L \alpha_{ki} \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} \gamma_{|i-j_1-\dots-j_{l+1}|, t} \right) \\
&\quad + \sum_{k=1}^K \alpha_k \left(\sum_{l=1}^{\infty} \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{kj_{l+1}}^2 \beta_{kj_1}^2 \dots \beta_{kj_l}^2 \gamma_{0,t-j_1-j_2-\dots-j_{l+1}} \right) \\
&\quad + \sum_{k=1}^K \alpha_k \left(\sum_{l \neq l'=1}^{\infty} \sum_{\substack{j_1, \dots, j_l=1 \\ j'_1, \dots, j'_{l'}=1}}^L \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} \alpha_{kj'_{l'+1}} \beta_{kj'_1} \dots \beta_{kj'_{l'}} \gamma_{|j_1+\dots+j_{l+1}-j'_1-\dots-j'_{l'+1}|, t} \right) \\
&= C_0 + \sum_{k=1}^K \alpha_k \left(\sum_{l=0}^{\infty} \sum_{j_1, \dots, j_l=1}^L \alpha_{kj_{l+1}}^2 \beta_{kj_1}^2 \dots \beta_{kj_l}^2 \gamma_{0,t-j_1-j_2-\dots-j_{l+1}} \right) \\
&\quad + \sum_{k=1}^K \alpha_k \left(\sum_{l \neq l'=0}^{\infty} \sum_{\substack{j_1, \dots, j_l=1 \\ j'_1, \dots, j'_{l'}=1}}^L \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} \alpha_{kj'_{l'+1}} \beta_{kj'_1} \dots \beta_{kj'_{l'}} \gamma_{|j_1+\dots+j_{l+1}-j'_1-\dots-j'_{l'+1}|, t} \right) \\
&= C_0 + \sum_{k=1}^K \alpha_k \left(\sum_{i=1}^L \Delta_{k,i} \gamma_{0,t-i} + \sum_{v=1}^{L-1} \Lambda_{k,v} \gamma_{v,t} \right)
\end{aligned}$$

where

$$\begin{aligned}
C_0 &= \mu + \sum_{k=1}^K \alpha_k \left(C_{k0}^2 + 2C_{k0}\mu \left(\sum_{i=1}^L \alpha_{ki} + \sum_{l=1}^{\infty} \sum_{j_1, \dots, j_{l+1}=1}^L \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} \right) \right), \\
\Delta_{k,i} &= \sum_{l=0}^{\infty} \sum_{j_1+\dots+j_l=i} \alpha_{kj_{l+1}}^2 \beta_{kj_1}^2 \dots \beta_{kj_l}^2, \\
\Lambda_{k,v} &= \sum_{l \neq l'=0}^{\infty} \sum_{|j_1+j_2+\dots+j_{l+1}-j'_1-j_2-\dots-j'_{l'+1}|=v} \alpha_{kj_{l+1}} \beta_{kj_1} \dots \beta_{kj_l} \alpha_{kj'_{l'+1}} \beta_{kj'_1} \dots \beta_{kj'_{l'}}.
\end{aligned}$$

Then

$$\begin{aligned}
\gamma_{0,t} &= C_0 + \sum_{k=1}^K \alpha_k \left(\sum_{i=1}^L \Delta_{k,i} \gamma_{0,t-i} + \sum_{v=1}^{L-1} \Lambda_{kv} \gamma_{v,t} \right) \\
&= C_0 + \sum_{k=1}^K \alpha_k \left[\sum_{u=1}^L \Delta_{k,u} \gamma_{0,t-u} + \sum_{v=1}^{L-1} \Lambda_{kv} \left(-K_1 \sum_{u=1}^{L-1} b_{vu} - \sum_{u=1}^{L-1} b_{vu} \omega_{u0} \gamma_{0,t-u} \right) \right] \\
&= c_0 + \sum_{k=1}^K \alpha_k \left[\sum_{u=1}^L \Delta_{k,u} \gamma_{0,t-u} - \sum_{u=1}^{L-1} \left(\sum_{v=1}^{L-1} \Lambda_{kv} b_{vu} \omega_{u0} \right) \gamma_{0,t-u} \right] \\
&= c_0 + \sum_{k=1}^K \alpha_k \left[\sum_{u=1}^{L-1} \left(\Delta_{k,u} - \sum_{v=1}^{L-1} \Lambda_{kv} b_{vu} \omega_{u0} \right) \gamma_{0,t-u} + \Delta_{k,L} \gamma_{0,t-L} \right]
\end{aligned} \tag{2.4}$$

where

$$c_0 = C_0 - K_1 \sum_{v=1}^{L-1} \Lambda_{kv} \sum_{u=1}^{L-1} b_{vu}.$$

Let

$$c_u = \sum_{k=1}^K \alpha_k \left(\Delta_{k,u} - \sum_{v=1}^{L-1} \Lambda_{kv} b_{vu} \omega_{u0} \right), \quad u = 1, \dots, L-1 \text{ and } c_L = \sum_{k=1}^K \alpha_k \Delta_{k,L}.$$

Then the equation 2.4 is equivalent to :

$$\gamma_{0,t} = c_0 + \sum_{u=1}^L c_u \gamma_{0,t-u}.$$

A necessary and sufficient condition for the process to be second-order stationary is that all roots of $1 - c_1 Z - c_2 Z^2 - \dots - c_L Z^L = 0$ lie outside the unit circle. \square

3. Estimation procedure

In this section, we discuss the estimation of the parameters of a MINGARCH model by using the expectation-maximization (EM) algorithm (see Dempster et al. (1977)).

Suppose that the observation $X = (X_1, \dots, X_n)$ is generated from the MIN-GARCH model.

Let $Z = (Z_1, \dots, Z_n)$ be the unobserved random variable, where $Z_t = (Z_{1,t}, \dots, Z_{K,t})^T$ is a K -dimensional vector where

$$Z_{i,t} = \begin{cases} 1 & \text{if } X_t \text{ comes from the } i\text{-th component; } 1 \leq i \leq K, \\ 0 & \text{otherwise.} \end{cases}$$

The distribution of Z_t is

$$\mathbb{P}(Z_t = (1, 0, \dots, 0)^T) = \alpha_1, \dots, \mathbb{P}(Z_t = (0, 0, \dots, 0, 1)^T) = \alpha_K.$$

Let $\alpha = (\alpha_1, \dots, \alpha_{K-1})^T$, $\alpha_{(k)} = (\alpha_{k0}, \alpha_{k1}, \dots, \alpha_{kp_k})^T$, $\beta_{(k)} = (\beta_{k1}, \dots, \beta_{kq_k})^T$, $\theta_{(k)} = (\alpha_{(k)}^T, \beta_{(k)}^T)$ and $\theta = (\alpha, \theta_{(1)}, \dots, \theta_{(K)})^T \in \Theta$ (The parameters space).

The conditional distribution of the complete data $\underline{X}_t = (X_t, Z_t)$ is

$$\prod_{k=1}^K \left(\alpha_k \frac{\lambda_{kt}^{X_t} \exp(-\lambda_{kt})}{X_t!} \right)^{Z_{kt}}$$

and the conditional log-likelihood function at time t is given by

$$l_t = \sum_{k=1}^K Z_{kt} \log(\alpha_k) + X_t \sum_{k=1}^K Z_{kt} \log(\lambda_{kt}) - \sum_{k=1}^K Z_{kt} \lambda_{kt} - \log(X_t!).$$

The conditional log-likelihood is given by $l(\theta) = \sum_{t=1}^n l_t$.

$\sum_{t=1}^L l_t$ is the joint log-likelihood function of the first L random variables of the series and $l^*(\theta) = \sum_{t=L+1}^n l_t$ is called the conditional log-likelihood function.

When the sample size n is large, the influence of $\sum_{t=1}^L l_t$ will be negligible. In this study, the parameters will be estimated by maximizing the conditional log-likelihood function l^* . The conditional log-likelihood is then given by

$$l^*(\theta) = \sum_{t=L+1}^n \left\{ \sum_{k=1}^K Z_{kt} \log(\alpha_k) + X_t \sum_{k=1}^K Z_{kt} \log(\lambda_{kt}) - \sum_{k=1}^K Z_{kt} \lambda_{kt} - \log(X_t!) \right\}. \quad (3.5)$$

The first derivatives of the conditional log-likelihood with respect to θ are :

$$\frac{\partial l^*}{\partial \alpha_k} = \sum_{t=L+1}^n \left(\frac{Z_{kt}}{\alpha_k} - \frac{Z_{Kt}}{\alpha_K} \right), \quad k = 1, \dots, K-1, \quad (3.6)$$

$$\frac{\partial l^*}{\partial \alpha_{ki}} = \sum_{t=L+1}^n Z_{kt} \frac{X_t - \lambda_{kt}}{\lambda_{kt}} U(X_t, i), \quad k = 1, \dots, K, \quad i = 0, \dots, p_k \quad (3.7)$$

$$\frac{\partial l^*}{\partial \beta_{kj}} = \sum_{t=L+1}^n Z_{kt} \frac{X_t - \lambda_{kt}}{\lambda_{kt}} \lambda_{k,t-j}, \quad k = 1, \dots, K, \quad j = 1, \dots, q_k, \quad (3.8)$$

where

$$U(X_t, i) = \begin{cases} 1 & \text{if } i = 0, \\ X_{t-i} & \text{if } i \neq 0. \end{cases}$$

Given that the process $\{Z_t\}$ is not observed, the data that we have do not allow the estimation of the parameter θ . An iterative process EM procedure is proposed for estimating the parameters by maximizing the conditional log-likelihood function $l^*(\theta)$ consists of an (E-step) and an (M-step).

These steps are described in the following:

- (a) E-step: suppose that θ is known. The missing data \mathbf{Z} are then replaced by their conditional expectations, conditional on the parameters and on the observed data X . In this case the conditional expectation of the $k - th$ component of Z_t is just the conditional probability that the observation X_t comes from the $k - th$ component of the mixture distribution conditional on θ and X . Let $\tau_{k,t}$ be the conditional expectation of Z_{kt} . Then the E-step equation is given by :

$$\tau_{k,t}^{(s)} = \frac{\alpha_k^{(s-1)} \lambda_{kt}^{(s-1) X_t} \exp(-\lambda_{kt}^{(s-1)})}{\sum_{i=1}^K \alpha_i^{(s-1)} \lambda_{it}^{(s-1) X_t} \exp(-\lambda_{it}^{(s-1)})}$$

where $k = 1, 2, \dots, K$ and $t = L + 1, \dots, n$. $s = 1, 2, \dots$ represents the iteration number.

In practice, the $Z_{kt}^{(s)}$'s are set to the $\tau_{k,t}^{(s)}$'s from the previous E-step of the EM procedure.

- (b) M-step: The missing data Z are replaced by their conditional expectations on the parameters θ and on the observed data X_1, \dots, X_n . The estimates of the parameters θ can then be obtained by maximizing the conditional log-likelihood function $l^*(\theta)$ by equating expressions (3.7) - (3.8) to 0. The M-step equations become

$$\hat{\alpha}_k^{(s)} = \frac{1}{(n - L)} \sum_{t=L+1}^n \tau_{k,t}^{(s)}, \quad k = 1, \dots, K$$

From the equation (3.7), we have :

$$\sum_{t=L+1}^n \frac{\tau_{t,k}^{(s)} X_t}{\hat{\lambda}_{kt}} U(X_t, i) = \sum_{t=L+1}^n \tau_{k,t}^{(s)} U(X_t, i).$$

Then

$$\sum_{t=L+1}^n \left\{ \frac{\tau_{k,t}^{(s)} X_t}{\sum_{j=0}^{p_k} \hat{\alpha}_{kj}^{(s)} U(X_t, j) + \sum_{j=1}^{q_k} \hat{\beta}_{kj}^{(s)} \hat{\lambda}_{k(t-j)}^{(s)}} U(X_t, i) \right\} = \sum_{t=L+1}^n \tau_{k,t}^{(s)} U(X_t, i),$$

for $k = 1, \dots, K$, $i = 0, \dots, p_k$.

Similarly equation (3.8) gives :

$$\sum_{t=L+1}^n \frac{\tau_{k,t}^{(s)} X_t}{\hat{\lambda}_{kt}^{(s)}} \hat{\lambda}_{k,t-j}^{(s)} = \sum_{t=L+1}^n \tau_{k,t}^{(s)} \hat{\lambda}_{k,t-j}^{(s)}.$$

Then

$$\sum_{t=L+1}^n \left\{ \frac{\tau_{k,t}^{(s)} X_t}{\sum_{i=0}^{p_k} \hat{\alpha}_{ki}^{(s)} U(X_t, i) + \sum_{t=L+1}^{q_k} \hat{\beta}_{ki}^{(s)} \hat{\lambda}_{k,t-i}^{(s)}} \hat{\lambda}_{k,t-j}^{(s)} \right\} = \sum_{t=L+1}^n \tau_{k,t}^{(s)} \hat{\lambda}_{k,t-j}^{(s)},$$

for $k = 1, \dots, K$, $j = 1, \dots, q_k$.

The estimates of the parameters are then obtained by iterating these two steps until convergence.

Let $\theta_i^{(s)}$ be the i th component of $\theta^{(s)}$, then the criterion used for checking convergence of the EM procedure is

$$\max \left\{ \left| \frac{\theta_i^{(s+1)} - \theta_i^{(s)}}{\theta_i^{(s)}} \right|, s, i \geq 1 \right\} \leq 10^{-5}.$$

Among different strategies for choosing starting initial values for the EM algorithm (see Karlis and Xekalaki (2003) and Melnykova and Melnykovb (2012)), the random initialization method is employed in this paper (the initial values for $\theta^{(k)}$ are chosen randomly from a uniform distribution and the mixing proportions are generated from a Dirichlet distribution).

The asymptotic properties are not treated in this paper but they have been studied by many authors. For example, Nityasuddhia and Bohning (2003) have studied the asymptotic properties of the EM algorithm estimate for normal mixture models. They show that the EM algorithm gives reasonable solutions of the score equations in an asymptotic unbiased sense.

The performance of the EM algorithm is assessed by some simulation experiments.

4. Simulation studies

Monte Carlo experiment was conducted to investigate the performances of the EM estimation method. In all these simulation experiments, we use ($R = 100$) independent realizations of the MINGARCH (2.1) model with sizes $n = 100$, $n = 200$ and $n = 500$. The following two models were used in the experiment. The first denoted Model (I), is a MINGARCH(2; 1,1; 1,1) model with parameter values

$$\begin{pmatrix} \alpha_1 & \alpha_{10} & \alpha_{11} & \beta_{11} \\ \alpha_2 & \alpha_{20} & \alpha_{21} & \beta_{21} \end{pmatrix} = \begin{pmatrix} 0.75 & 1.00 & 0.20 & 0.30 \\ 0.25 & 5.00 & 0.50 & 0.30 \end{pmatrix}.$$

Table 1. Results of the simulation study with model (I).

Sample size	k		α_k	α_{k0}	α_{k1}	β_{k1}
100	1	True values	0.7500	1.0000	0.2000	0.3000
		Mean estimated	0.7410	1.1883	0.1833	0.2446
		RMSE	0.0523	0.5789	0.0623	0.2137
		MAE	0.0405	0.4726	0.0506	0.1801
	2	True values	0.2500	5.0000	0.5000	0.3000
		Mean estimated	0.2590	5.1660	0.4619	0.2901
		RMSE	0.0523	2.6410	0.2823	0.2588
		MAE	0.0405	2.2060	0.2103	0.2274
200	1	True values	0.7500	1.0000	0.2000	0.3000
		Mean estimated	0.7463	1.0093	0.1909	0.3054
		RMSE	0.0359	0.4429	0.0468	0.1773
		MAE	0.0291	0.3641	0.0381	0.1460
	2	True values	0.2500	5.0000	0.5000	0.3000
		Mean estimated	0.2537	5.2571	0.4612	0.2928
		RMSE	0.0359	2.2616	0.1728	0.2380
		MAE	0.0291	1.8728	0.1314	0.1976
500	1	True values	0.7500	1.0000	0.2000	0.3000
		Mean estimated	0.7510	1.0646	0.1959	0.2817
		RMSE	0.0259	0.2525	0.0272	0.1035
		MAE	0.0212	0.1867	0.0214	0.0783
	2	True values	0.2500	5.0000	0.5000	0.3000
		Mean estimated	0.2490	5.3064	0.5026	0.2688
		RMSE	0.0259	1.6316	0.0982	0.1702
		MAE	0.0212	1.3483	0.0774	0.1443

The second denoted Model (II), is a MINGARCH(3; 1,1,1; 1,1,1) model with parameter values

$$\begin{pmatrix} \alpha_1 & \alpha_{10} & \alpha_{11} & \beta_{11} \\ \alpha_2 & \alpha_{20} & \alpha_{21} & \beta_{21} \\ \alpha_3 & \alpha_{30} & \alpha_{31} & \beta_{31} \end{pmatrix} = \begin{pmatrix} 0.55 & 0.80 & 0.40 & 0.30 \\ 0.25 & 1.00 & 0.50 & 0.25 \\ 0.20 & 0.50 & 0.60 & 0.20 \end{pmatrix}$$

The performances of the estimators are evaluated by the root mean square error (RMSE) and the mean absolute error (MAE).

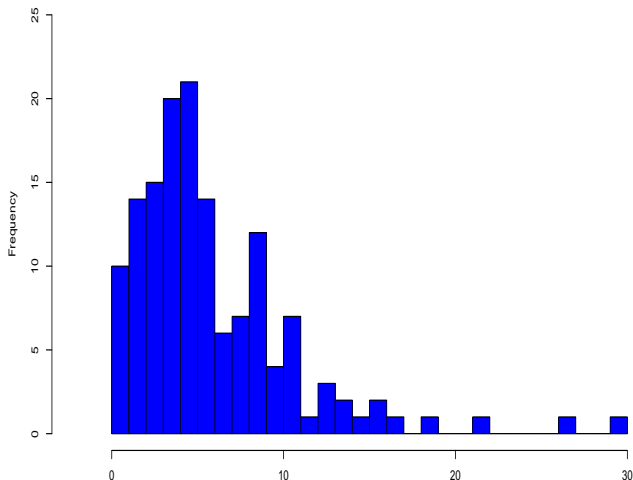
Based on the results in Tables 4 and 4, we can see that as the sample size increases, the estimates seem to converge to the true parameter values. The performance of the estimate improves when the sample size increases. But this performance varies depending on the parameters. Indeed the parameter estimate α_k seems to give good results for all sample sizes considered. For the parameter α_{k0} , the RMSE and the MAE are slightly higher.

Table 2. Results of the simulation study with model (II).

Sample size	k		α_k	α_{k0}	α_{k1}	β_{k1}
100	1	True values	0.5500	0.8000	0.4000	0.3000
		Mean estimated	0.5435	0.7671	0.4429	0.2163
		RMSE	0.1063	0.4997	0.1898	0.2339
		MAE	0.0828	0.4054	0.1482	0.1977
	2	True values	0.2500	1.0000	0.5000	0.2500
		Mean estimated	0.2240	1.0888	0.5344	0.2532
		RMSE	0.0802	0.7182	0.3804	0.2563
		MAE	0.0607	0.5504	0.2420	0.2113
	3	True values	0.2000	0.5000	0.6000	0.2000
		Mean estimated	0.2323	0.9516	0.4475	0.2714
		RMSE	0.0600	0.7127	0.2413	0.2263
		MAE	0.0429	0.5490	0.1895	0.1850
200	1	True values	0.5500	0.8000	0.4000	0.3000
		Mean estimated	0.5286	0.7471	0.4113	0.2552
		RMSE	0.1117	0.4363	0.1563	0.1942
		MAE	0.0838	0.3566	0.1190	0.1545
	2	True values	0.2500	1.0000	0.5000	0.2500
		Mean estimated	0.2316	1.0570	0.5340	0.2433
		RMSE	0.0785	0.6025	0.2584	0.1928
		MAE	0.0602	0.4787	0.1751	0.1506
	3	True values	0.2000	0.5000	0.6000	0.2000
		Mean estimated	0.2397	0.8867	0.4450	0.3042
		RMSE	0.0652	0.6088	0.2306	0.2439
		MAE	0.0452	0.4959	0.1806	0.1825
500	1	True values	0.5500	0.8000	0.4000	0.3000
		Mean estimated	0.5556	0.7040	0.4248	0.2725
		RMSE	0.0825	0.3246	0.1171	0.1797
		MAE	0.0614	0.2595	0.0934	0.1407
	2	True values	0.2500	1.0000	0.5000	0.2500
		Mean estimated	0.2182	0.9508	0.5223	0.2656
		RMSE	0.0620	0.4723	0.2059	0.2132
		MAE	0.0487	0.3853	0.1569	0.1576
	3	True values	0.2000	0.5000	0.6000	0.2000
		Mean estimated	0.2261	0.8985	0.4690	0.2815
		RMSE	0.0536	0.5883	0.1963	0.1988
		MAE	0.0298	0.4780	0.1586	0.1506

Table 3. Summary statistics of the crime counts series.

Sample size	Minimum	Maximum	Median	Mean	Variance	Skewness	Kurtosis
144	0	30	5	6.347	22.7317	1.9810	8.7646

**Figure 1.** Histogram of the crime counts series.

5. Real data examples

In this section we shall investigate the time series represents a count of CAD drug calls reported in the 22nd police car beat in Pittsburgh, during one month. It started in January 1990 and ended in December 2001. The data are available online at the forecasting principles site (<http://www.forecastingprinciples.com>), in the section about crime data. In the framework of Zhu et al. (2010), the bimodality index of Der and Everitt (2002) is used to show that the series is bimodal. In their results, they showed that the MINARCH model is more appropriate for this dataset than the INARCH model. But they are not taken account the **MA** components in the model.

In the following we fit a MINGARCH model (2.1) to the series with $K = 1, 2, 3$. The summary statistics are given in Table 5 . Mean and variance are estimated as 6.3470 and 22.7317, respectively. Hence the data seem to be overdispersed. The histogram of the series in Figure 1 show that the series is more or less bimodal. The autocorrelation function in Figure 2 implies that the third-order model can be considered. We consider the MINGARCH model with $1 \leq p \leq 3$ and $0 \leq q \leq 3$.

The model selection criteria considered here are the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the mixture regression

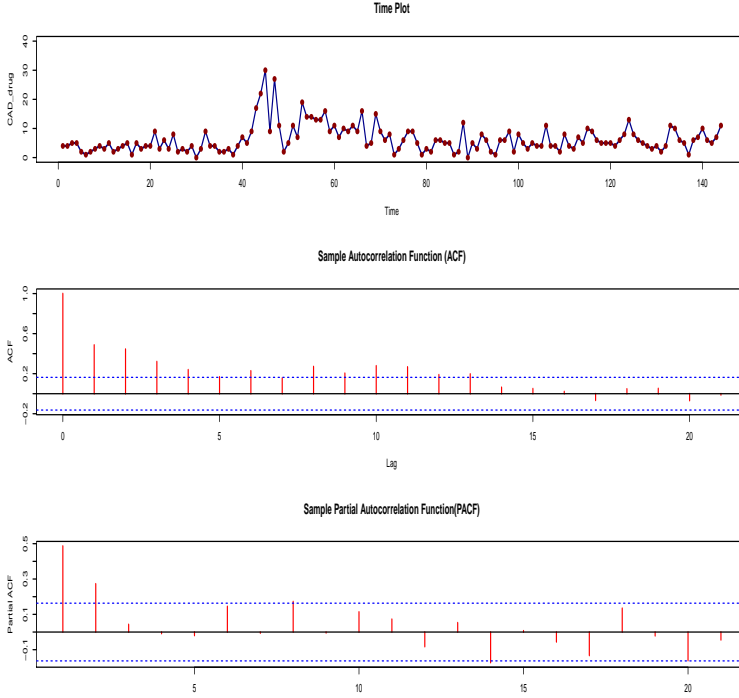


Figure 2. Crime counts series: the time plot, the sample autocorrelation and partial autocorrelation function.

criterion (MRC) proposed by Naik et al. (2007). These two first criteria are both defined as minus twice the maximized log-likelihood plus a penalty term. The first choice is the log-likelihood given by equation (3.5). The maximized (observed) log-likelihood is automatically generated by the EM estimation but it includes the information of the unobserved random variable \mathbf{Z} . The second choice is computed from the (conditional) probability density function of the MINGARCH model. It is defined as

$$l' = \sum_{t=L+1}^n \log \left\{ \sum_{k=1}^K \alpha_k \frac{\lambda_{kt}^{X_t} \exp(-\lambda_{kt})}{X_t!} \right\}.$$

We use l' because it may have better performance in finite samples (see Wong and Li (2000)). We use the following definition of AIC and BIC :

$$AIC = -2l' + 2 \left(2K - 1 + \sum_{k=1}^K p_k + \sum_{k=1}^K q_k \right),$$

$$BIC = -2l' + \log \left(n - \max(p_{\max}, q_{\max}) \right) \left(2K - 1 + \sum_{k=1}^K p_k + \sum_{k=1}^K q_k \right).$$

The third criterion consists of three terms: the first measures the lack of fit, the second imposes a penalty for regression parameters, and the third is the clustering penalty function. It is an extension of the AIC to mixture regression models. For the MINGARCH model, let $U = (X_{L+1}, \dots, X_n)^T$, $V = (V_{L+1}, \dots, V_n)^T$, $V_j = (1, X_{j-1}, \dots, X_{j-p}, \lambda_{k_j(j-1)}, \dots, \lambda_{k_j(j-q)})^T$, $k_j \mid \tau_{k_j, j} = \max\{\tau_{1, j}, \dots, \tau_{K, j}\}$, $j = L+1, \dots, n$, and $\theta_k^* = \left(\alpha_{(k)}^T, \mathbf{0}^T, \beta_{(k)}^T, \mathbf{0}^T \right)_{(p+q+1) \times 1}^T$, $\widehat{W}_k = \text{diag}((\widehat{\tau}_{k, L+1}, \dots, \widehat{\tau}_{k, n})^T)$, $\widehat{V}_k = \widehat{W}_k^{1/2} V$, $\widehat{H}_k = \widehat{V}_k \left(\widehat{V}_k^T \widehat{V}_k \right)^{-1} \widehat{V}_k^T$, $k = 1, \dots, K$.

Then the MRC is defined as

$$MRC = \sum_{k=1}^K \widehat{n}_k \log(\widehat{\sigma}_k^2) + \sum_{k=1}^K \frac{\widehat{n}_k(\widehat{n}_k + \widehat{h}_k)}{\widehat{n}_k - \widehat{h}_k - 2} - 2 \sum_{k=1}^K \widehat{n}_k \log(\widehat{\alpha}_k),$$

where $\widehat{n}_k = \text{tr}(\widehat{W}_k)$, $\widehat{h}_k = \text{tr}(\widehat{H}_k)$ and $\widehat{\sigma}_k^2 = (U - V\theta_k^*)^T \widehat{W}_k^{1/2} (I - \widehat{H}_k) (U - V\theta_k^*) / \widehat{n}_k$.

There are three aspects of model selection in the MINGARCH model. First, we must select the number of components K . Second, we must select the order of each **AR** component, i.e. p_k . Thirdly, we must select the order of each **MA** component, i.e. q_k .

The selection problem for the number of components is not discussed, we concentrate on the order selection of each component. The order of the components is chosen to be that minimizing the values of the three criterions.

The AIC, BIC and MRC values for the crime counts series are given in Tables 5, 5, 5 and 5.

The results in Table 5, the AIC, the BIC and the MRC retain respectively the two-component mixture model with $p = 3$, the two-component mixture model with $p = 2$ and the single-component model with $p = 3$ (as selected in the framework of Zhu et al. (2010)).

Based on the results in these tables (5, 5, 5 and 5), the AIC and the BIC retain the two-component mixture model with respectively $(p, q) = (1, 3)$ and $(p, q) = (1, 1)$. It confirms the result of the histogram and lends substantial

Table 4. AIC, BIC and MRC values for the crime counts series with $q = 0$.

K	AIC			BIC			MRC		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
1	825.08	806.21	803.77	831.00	815.08	815.56	554.38	544.71	543.85
2	762.40	755.24	751.92	777.22	775.93	778.46	600.39	590.37	588.23
3	763.73	757.93	756.49	787.43	790.45	797.77	632.82	622.69	686.01

Table 5. AIC, BIC and MRC values for the crime counts series with $K = 1$.

Order	AIC			BIC			MRC		
	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
$p = 1$	806.61	804.18	798.75	822.50	816.00	813.50	548.34	546.08	544.29
$p = 2$	804.98	806.89	800.75	816.81	821.67	818.44	546.41	548.58	546.36
$p = 3$	804.63	805.40	802.97	819.37	823.09	823.61	544.88	547.43	545.62

Table 6. AIC, BIC and MRC values for the crime counts series with $K = 2$.

Order	AIC			BIC			MRC		
	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
$p = 1$	754.24	749.83	745.47	754.24	776.43	797.69	546.38	531.45	533.94
$p = 2$	752.44	758.64	749.32	779.04	791.15	787.65	538.78	588.95	538.96
$p = 3$	753.48	751.45	752.96	785.92	789.79	797.19	590.52	540.92	542.03

Table 7. AIC, BIC and MRC values for the crime counts series with $K = 3$.

Order	AIC			BIC			MRC		
	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
$p = 1$	756.41	753.64	746.07	789.01	795.03	796.20	573.78	593.44	601.43
$p = 2$	756.05	759.68	762.29	797.44	809.93	821.26	632.30	888.62	707.24
$p = 3$	765.62	760.06	760.80	815.74	819.04	828.62	669.78	430.52	465.50

support to the two-component model. In practice, it is observed that the BIC criterion selects the model of dimension smaller than the AIC criterion, which is not surprising since the BIC penalizes more than the AIC (when $n > 7$). The MRC suggests the three-component model, but the third smallest MRC values (531.45) is obtained in the two-component model with $(p, q) = (1, 2)$. The values of the criteria AIC, BIC and MRC obtained in our model are better than those of model MINARCH. We can conclude that the MINGARCH model is more appropriate for this dataset than the MINARCH model.

6. Concluding remarks

The selection of K is more important as it will affect our interpretation of the MINGARCH model and the correct selection of the orders is dependent on the selected number of components, but this is difficult to handle and using AIC or BIC to choose K is somewhat non-standard of particular importance.

References

- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society B* 39, 1–38.
- Der, G. and B. Everitt (2002). *A Handbook of Statistical Analyses using SAS, second ed.* Chapman & Hall/CRC, London.
- Ferland, R., A. Latour, and D. Oraichi (2006). Integer-valued garch process. *Journal of Time Series Analysis* 27, 923–42.
- Goldberg, S. (1958). *Introduction to Difference Equations.* New York:Wiley.
- Karlis, D. and E. Xekalaki (2003). Choosing initial values for the em algorithm for finite mixtures. *Computational Statistics & Data Analysis* 41, 577 – 590.
- Le, N. D., R. D. Martin, and A. Z. Raftery (1996). Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association* 91, 1504–1515.
- Martin, V. L. (1992). Threshold time series models as multimodal distribution jump processes. *Journal of Time Series Analysis* 13, 79–94.
- Melnykova, V. and I. Melnykovb (2012). Initializing the em algorithm in gaussian mixture models with an unknown number of components. *Computational Statistics and Data Analysis* 56, 1381–1395.
- Muller, D. W. and G. Sawitzki (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association* 86, No. 415, 738–746.

- Naik, P. A., P. Shi, and C. L. Tsai (2007). Extending the akaike information criterion to mixture regression models. *Journal of the American Statistical Association* 102, 244–254.
- Nityasuddhia, D. and D. Bohning (2003). Asymptotic properties of the em algorithm estimate for normal mixture models with component specific variances. *Computational Statistics & Data Analysis* 41, 591– 601.
- Titterington, D. M., A. F. M. Smith, and U. E. Makov (1985). Statistical analysis of finite mixture distributions. In *Wiley, New York*.
- Wong, C. and W. Li (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society Series B* 62, 95–115.
- Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika* 75, 621–9.
- Zhu, F. (2011). A negative binomial integer-valued garch model. *Journal of Times Series Analysis* 32, 54–67.
- Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications* 389, 58–71.
- Zhu, F., Q. Li, and D. Wang (2010). A mixture integer-valued arch model. *Journal of Statistical Planning and Inference* 140, 2025–2036.