Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model

Orlando Yesid Esparza Albarracin, Airlane Pereira Alencar & Linda Lee Ho


To link to this article: https://doi.org/10.1080/00949655.2019.1599892

Published online: 09 Apr 2019.
Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model

Orlando Yesid Esparza Albarracin a, Airlane Pereira Alencar a, and Linda Lee Ho b

aIME-USP, São Paulo, Brazil; bEP-USP, São Paulo, Brazil

ABSTRACT
In this paper, we call attention of two observed features in practical applications of the Generalized Autoregressive Moving Average (GARMA) model due to the structure of its linear predictor. One is the multicollinearity which may lead to a non-convergence of the maximum likelihood, using iteratively reweighted least squares, and the inflation of the estimator’s variance. The second is that the inclusion of the same lagged observations into the autoregressive and moving average components confounds the interpretation of the parameters. A modified model, GAR-M, is presented to reduce the multicollinearity and to improve the interpretation of the parameters. The expectation and variance under stationarity conditions are presented for the identity and logarithm link function. In a general sense, simulation studies show that the maximum likelihood estimators based on the GARMA and GAR-M models are equivalent but the GAR-M estimators presented a little lower standard errors and some restrictions in the parametric space are imposed to guarantee the stationarity of the process. Also, a real data analysis illustrates the GAR-M fit for daily hospitalization rates of elderly people due to respiratory diseases from October 2012 to April 2015 in São Paulo city, Brazil.

1. Introduction

In many practical problems, the primary objective is to develop models that take into account the seasonality, changes in the population and other numerous covariates to observations registered over time as the daily number of asthma admissions in a hospital. This allows measuring the impact of covariates on time series and capturing a suitably dependence among observations. For data with normal distribution, the most common strategy is to fit a linear regression where the errors have a general autoregressive moving average (ARMA) structure [1]. For count time series, the generalized autoregressive moving average (GARMA) [2] and the generalized linear autoregressive moving average (GLARMA) [3] models extend the univariate Gaussian ARMA model to a flexible observation-driven model for non-Gaussian time series allowing to model discrete and continuous time series. These classes of models are widely used in areas of surveillance [4,5] where it is necessary to model discrete response time series in terms of covariates.
In this paper, we focus on the GARMA model, where the conditional mean of the response variable may depend on a set of covariates, as in the generalized linear models (GLM) [6]. However, the GARMA model also includes lagged values of the dependent variable in the set of regressors to model the serial dependence. Thus, it is possible to make inference about the effects of covariates on the mean of the response variable and take into account the serial dependence. In the GARMA model, the distribution of each observation conditioned on the past information belongs to the exponential family (as the Gaussian, Poisson, Gamma and Binomial distributions), allowing to model discrete and continuous time series.

The parameters are estimated by the maximum likelihood method and the optimization is carried out using an iteratively reweighted least squares (IRLS) algorithm, as for the GLM. However, the convergence of this algorithm (IRLS) may not be reached, mainly in the GARMA \((p,q)\) models, with \(p\) and \(q\) simultaneously different from zero. This occurs due to multicollinearity problems caused by the structure of the linear predictor. The existence of substantial correlations creates difficulties, namely, numerical instabilities and problems of identification and interpretation of the effects of the auto-regressive and the moving average terms included in the linear predictor to model the serial correlations among observations measured along time. Perfect multicollinearity leads to a non-identifiable model and an unfeasible estimation as it is not possible to invert a matrix in the IRLS method.

The main goal of the current paper is to deconstruct the GARMA \((p,q)\) models (with \(p\) and \(q\) simultaneously different from zero) to identify multicollinearity problems and understand the interpretation of the autoregressive and moving average parameters for each lag. The multicollinearity problems in the GARMA models are independent of the inclusion of covariates. Additionally, a modified GARMA model, namely GAR-M, is proposed to reduce the problem of multicollinearity in the GARMA model and to have a better interpretation of the parameters. The GARMA and the GAR-M models are evaluated using simulated data. Finally, a negative binomial GAR-M model is fitted to model the time series of daily number of hospital admissions due to respiratory diseases in the city São Paulo – Brazil from October 2012 to April 2015.

This paper is organized as follows: the GARMA model is defined in Section 2; the deconstruction of the GARMA model in Section 3; the modified GARMA model is proposed in Section 4; simulations studies are presented in Section 5; a real data analysis on daily hospital admission due to respiratory diseases is presented in Section 6 and discussions and conclusions are outlined in Section 7.

### 2. Generalized autoregressive moving average model

In the GARMA model, introduced by Benjamin et al. [2], the distribution of each observation \(y_t\), for \(t = 1, \ldots, n\), conditioned on the past information \(H_t = \{x_1, \ldots, x_t, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\}\), belongs to the exponential family. The conditional density function is expressed as

\[
  f(y_t|H_t) = \exp \left\{ \frac{y_t \nu_t - b(\nu_t)}{\varphi} + d(y_t, \varphi) \right\},
\]

where \(\nu_t\) is the canonical parameter; \(\varphi\) the scale parameter, \(b(.)\) and \(d(.)\) are specific functions that define a particular distribution from the exponential family, as Normal, Gamma and Poisson distributions.
The terms \( \mu_t = E(y_t|H_t) = b'(\nu_t) \) represent the conditional mean of \( y_t \) given \( H_t \) and \( \text{var}(y_t|H_t) = \varphi b''(\nu_t) = \varphi v(\mu_t) \) the conditional variance, where \( v(\mu_t) \) is called the variance function. For example, the Poisson distribution belongs to the exponential family distribution, with: \( \varphi = 1, \nu_t = \ln(\mu_t), b'(\nu_t) = \exp(\nu_t), d(y_t, \varphi) = -\ln(y_t!) \) and \( v(\mu_t) = \mu_t. \)

As in the Generalized Linear Models (GLM) [7], \( \mu_t \) is related to a linear predictor, \( \eta_t \), by a twice-differentiable one-to-one monotonic link function \( g(\cdot) \). However, in the GARMA models, the autoregressive and the moving average terms are included to model possible serial correlations present in the observations \( y_t \) that are observed over time. The linear predictor for the GARMA model [2] is expressed as

\[
g(\mu_t) = \eta_t = \mathbf{x}_t' \mathbf{\beta} + \sum_{j=1}^{p} \phi_j A(y_{t-j}, \mathbf{x}_{t-j}') \mathbf{\beta} + \sum_{j=1}^{q} \theta_j M(y_{t-j}, \mu_{t-j}),
\]

where \( \mathbf{x}_t \) is a vector of \( r \) explanatory variables and \( \mathbf{\beta}' = (\beta_1, \beta_2, \ldots, \beta_r) \). The autoregressive parameters are \( \phi' = (\phi_1, \ldots, \phi_p) \); the moving average parameters are \( \theta' = (\theta_1, \ldots, \theta_q) \); and \( A \) and \( M \) are functions of the autoregressive and moving average terms, respectively.

The moving average error terms, \( M \), can be different types of residuals, for example, the deviance residuals, Pearson residuals, the residuals measured on the original scale or on the predictor scale [i.e. \( g(\mu_t) - \eta_t \)]. In practice, the linear predictor (2) can be rewritten in a simpler form as

\[
g(\mu_t) = \eta_t = \mathbf{x}_t' \mathbf{\beta} + \sum_{j=1}^{p} \phi_j \left[ g(y_{t-j}) - \mathbf{x}_{t-j}' \mathbf{\beta} \right] + \sum_{j=1}^{q} \theta_j \left[ g(y_{t-j}) - \eta_{t-j} \right].
\]

A particular GARMA \((p, q)\) model is defined by Equations (1) and (3). For certain link functions, it may be necessary to replace \( y_{t-j} \) by \( y_{t-j}^c \) in (3) to avoid the non-existence of \( g(y_{t-j}) \) for certain values of \( y_{t-j} \). For example, if the logarithm is the link function, any zero values of \( y_t \) must be replaced by a threshold parameter \( c \), satisfying \( 0 < c < 1 \) (in Benjamin et al. [2] \( c = 0.1 \) is used).

The model parameters denoted by \( \gamma = (\mathbf{\beta}', \phi', \theta')' \) are estimated by the conditional maximum likelihood. The linear predictor in (3), for the observations \( y_{n+1}, \ldots, y_n \), can be written in the matrix form as \( \eta_t = B' \gamma \), where the matrix \( B \) is the matrix of explanatory variables and the autoregressive and moving average terms (more details in Appendix 1). The conditional maximum likelihood estimators (MLE) can be obtained using an iterative weighted least squares process. At the iteration \( (k + 1) \), the estimation is updated as

\[
\hat{\gamma}^{(k+1)} = \left( B' W^{(k)} B \right)^{-1} B' W^{(k)} z^{(k)}
\]

The convergence of (4) occurs generally in a finite number of steps. For more details, see Appendix 1.

3. Deconstructing the GARMA model

In this section, the GARMA \((p,q)\) models defined by Equations (1) and (3), with \( p \) and \( q \) simultaneously different from zero, are deconstructed. First the structure of the linear
predictor is studied, writing each linear predictor \((\eta_t)\) as function of lagged linear predictors and lagged observations in order to understand their contributions to the predictor. Then, the multicollinearity of GARMA models may be deeply understood, motivating the proposal of the modified GAR-M model.

### 3.1. The structure of the linear predictor

For ease the development, firstly, only the GARMA(1,1) model is considered. The linear predictor, as defined by (3), for the GARMA(1,1) model with a \(g(\cdot)\) link function, is expressed as

\[
\eta_t = g(\mu_t) = x_t' \beta + \phi_1 \left\{ g(y_{t-1}) - x_{t-1}' \beta \right\} + \theta_1 \left\{ g(y_{t-1}) - \eta_{t-1} \right\}.
\]  

(5)

To initialize the process, let \(\eta_1 = g(\mu_1) = x_1' \beta\). Then

\[
\eta_2 = g(\mu_2) = x_2' \beta + \phi_1 \left\{ g(y_1) - x_1' \beta \right\} + \theta_1 \left\{ g(y_1) - \eta_1 \right\}
\]

\[
= x_2' \beta + \phi_1 \left\{ g(y_1) - x_1' \beta \right\} + \theta_1 \left\{ g(y_1) - x_1' \beta \right\}.
\]

Note that, for \(t = 2\), the effect of the previous instant is considered at the same time by the parameters \(\phi_1\) and \(\theta_1\). Then, for \(t = 3\) the linear predictor is

\[
\eta_3 = g(\mu_3) = x_3' \beta + \phi_1 \left\{ g(y_2) - x_2' \beta \right\} + \theta_1 \left\{ g(y_2) - \eta_2 \right\}
\]

\[
= x_3' \beta + \phi_1 \left\{ g(y_2) - x_2' \beta \right\} + \theta_1 \left\{ g(y_2) - x_2' \beta \right\}
\]

\[
- (\phi_1 + \theta_1) \left\{ g(y_1) - x_1' \beta \right\}.
\]

Observe that the effect of the previous instant \((t = 2)\) is associated simultaneously with the parameters \(\phi_1\) and \(\theta_1\). But in this case, the parameter \(\theta_1\) additionally takes into account the effect of older instants. Making successive substitutions, \(\eta_t\) can be written as

\[
\eta_t = x_t' \beta + \phi_1 \left\{ g(y_{t-1}) - x_{t-1}' \beta \right\} + \theta_1 \left\{ g(y_{t-1}) - x_{t-1}' \beta + c_t \right\},
\]

with \(c_t = \sum_{j=1}^{t-2} (-\theta_1)^{j-1}(\theta_1 + \phi_1) \left\{ g(y_{t-1-j}) - x_{t-1-j}' \beta \right\}. \)

(6)

Writing the linear predictor as in (6), it is easy to observe that the effect of the previous term, \(g(y_{t-1}) - x_{t-1}' \beta\), is taken into account simultaneously by the autoregressive \(\phi_1\) and the moving average \(\theta_1\) parameters. This fact leads to multicollinearity in the GARMA model causing estimation problems, which are detailed in the next section, and confounds the interpretation of the parameters. For instance, for daily time series, the parameter \(\phi_1\) takes into account only the effect of the previous day and the parameter \(\theta_1\) considers not only the direct effect of the previous day, but also all the older observations attributing decreasing weights with \((-\theta_1)^{j-1}(\theta_1 + \phi_1))\). The expressions \(\{g(y_{t-1}) - x_{t-1}' \beta\}\) and
\{g(y_{t-1}) - \dot{x}_{t-1}^j \beta + \epsilon_t \} are called hereafter as the autoregressive and moving average terms, respectively.

On the other hand, the performance of the GARMA(1,0) and GARMA(0,1) models is similar when \( \theta_1 = \phi_1 \). This occurs because, in this case, both models give the same weight for the common term \{g(y_{t-1}) - \dot{x}_{t-1}^j \beta \} (see (6)). For the GARMA(0,1) the older terms are multiplied by decreasing weights \( \phi_j^t \), for \( j = 2, 3, \ldots \). For example, for \( \theta_1 = \phi_1 = 0.2 \) the highest weight attributed to older instants in the GARMA(0,1) model is 0.04.

For the GARMA(2,1) model, the linear predictor, as defined by (3), considering \( g(\cdot) \) as a link function, is expressed as

\[
\eta_t = \dot{x}_t^j \beta + \phi_1 \left\{ g(y_{t-1}) - \dot{x}_{t-1}^j \beta \right\} + \phi_2 \left\{ g(y_{t-2}) - \dot{x}_{t-2}^j \beta \right\} + \theta_1 \left\{ g(y_{t-1}) - \dot{x}_{t-1}^j \beta - (\theta_1 + \phi_1) \left\{ g(y_{t-2}) - \dot{x}_{t-2}^j \beta \right\} + \epsilon_{1,t} \right\}
\]

with \( \epsilon_{1,t} = \sum_{j=1}^{(t-3)} (-\theta_1)^{j-1} \left[ \theta_1 (\theta_1 + \phi_1) - \phi_2 \right] \left\{ g(y_{t-2-j}) - \dot{x}_{t-2-j}^j \beta \right\} \)

It is worth noting that the effects of two previous terms \( g(y_{t-1}) - \dot{x}_{t-1}^j \beta \) and \( g(y_{t-2}) - \dot{x}_{t-2}^j \beta \) are measured simultaneously by the autoregressive parameters \( \phi_1, \phi_2 \) and the moving average parameter \( \theta_1 \). In general, in the GARMA(\( p,1 \)) models with \( p > 1 \), the effect of the \( p \) previous terms \( \{g(y_{t-j}) - \dot{x}_{t-j}^j \beta \} \) for \( j = 1, 2, \ldots p \) are taken into account simultaneously by the autoregressive parameters and the moving average parameter leading to problems of multicollinearity and hindering the interpretation of these parameters.

On the other hand, in the GARMA(1,\( q \)) models with \( q > 1 \), many previous terms are measured simultaneously, with different weights, by the \( q \) moving average parameters. For example, in the GARMA(1,2), the moving average parameters \( \theta_1 \) and \( \theta_2 \) take into account, simultaneously, all terms \( \{g(y_j) - \dot{x}_{t-j}^j \beta \} \), for \( j < t-2 \).

### 3.2. Multicollinearity problems

In the GARMA(\( p,q \)) models, with \( p \) and \( q \) simultaneously different from zero, problems of multicollinearity exist due to the structure of the linear predictor defined in (3), since the autoregressive and the moving average terms are highly correlated for some regions of parameter space. Then, the convergence of the maximum likelihood, using IRLS, to find the estimators of the parameters may not be achieved. In this section, the regions of the parameter space that induce more multicollinearity in the GARMA(1,1) models are studied. In cases where there are covariates \( x_t \) in the model, it is assumed that these are not correlated with each other.

To illustrate the relationship between the autoregressive terms of the linear predictor in the GARMA(1,1) model, four different scenarios are considered assuming that the response variables consist of a time series of counts under the Poisson distribution with the natural logarithm as a link function and assuming that \( \dot{x}_t \beta = \beta_0 \) for all \( t \). Figure 1 presents scatter plots of the autoregressive terms (that is, \( \{g(y_{t-1}) - \beta_0 \} \)) versus the moving average terms (that is, \( \{g(y_{t-1}) - \eta_{t-1} = \{g(y_{t-1}) - \beta_0 + \epsilon_t \} \). In all cases, there is a linear relationship and their Pearson sample correlations are greater than 0.79.
Figure 1. Scatter plots for some \((\phi_1, \theta_1) - \text{GARMA}(1,1)\). (a) \(\phi_1 = 0.4\) and \(\theta_1 = 0.3\). (b) \(\phi_1 = -0.2\) and \(\theta_1 = 0.5\). (c) \(\phi_1 = 0.1\) and \(\theta_1 = -0.5\). (d) \(\phi_1 = -0.2\) and \(\theta_1 = -0.4\).

Figure 2 shows a correlation heatmap of the autoregressive term and the moving average term with values of \(\theta_1\) and \(\phi_1\) in the interval \((-1, 1)\). For 50% of the parametric space \((-1, 1)^2\), the sample Pearson correlation between the autoregressive term and the moving average term is greater than 0.65.

To illustrate the convergence problems, the matrix \(B'WB\) in (4) is examined for a \(\text{GARMA}(1,1)\) with \(x_t'\beta = \beta_0\) for all \(t\) (which is a stationarity condition for the GARMA process) and link function \(g(\cdot)\). In this case, the matrix \(B'WB\) is expressed as

\[
\begin{bmatrix}
\sum_{t=m+1}^{n} w_t \\
\sum_{t=m+1}^{n} w_t h_t \\
\sum_{t=m+1}^{n} w_t h_t^2 \\
\sum_{t=m+1}^{n} w_t (g(y_{t-1}) - \eta_{t-1}) \\
\sum_{t=m+1}^{n} w_t h_t (g(y_{t-1}) - \eta_{t-1}) \\
\sum_{t=m+1}^{n} w_t (g(y_{t-1}) - \eta_{t-1})^2
\end{bmatrix}
\]
where \( w_t = \nu_t^{-1}(d g(\mu_t)/d\mu_t)^{-2} \) and \( h_t = g(y_{t-1}) - \beta_0 \). Considering that \( g(y_t) - \eta_t \) may be written as \( g(y_t) - \beta_0 + c_t \), where \( c_t \) is defined in (6), the matrix \( B'WB \) is rewritten as

\[
\begin{bmatrix}
\sum_{t=m+1}^{n} w_t \\
\sum_{t=m+1}^{n} w_t h_t \\
\sum_{t=m+1}^{n} w_t (h_t + c_{t-1}) \\
\sum_{t=m+1}^{n} w_t h_t \\
\sum_{t=m+1}^{n} w_t h_t^2 \\
\sum_{t=m+1}^{n} w_t (h_t^2 + c_{t-1}h_t) \\
\sum_{t=m+1}^{n} w_t (h_t^2 + 2c_{t-1}h_t + c_{t-1}^2) \\
\end{bmatrix}
\]

Note that each element of the column 3 is equal to the respective element of the column 2 plus a term that depends on \( c_t \). A very high correlation (> 0.8) between these two columns is observed for several pairs of values of \( \theta_1 \) and \( \phi_1 \) causing problems of invertibility of the matrix \( B'WB \). In many cases this matrix is singular.

4. **A modified GARMA**

In this section, a modified GARMA model, namely GAR-M\((p,q)\) with \( p \geq 1 \) and \( q \geq 1 \), is presented as a proposal to reduce the problem of multicollinearity in the GARMA models, and to improve the interpretation of the parameters. The distribution of each observation \( y_t \) for \( t = 1, \ldots, n \), conditional on the past information \( \{H_t = x_t, \ldots, x_1, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\} \) belongs to the exponential family as in (1). However, in the GAR-M model, the possible serial correlation in the observations \( y_t \) is modelled in terms of the linear predictor for \( \mu_t \), by a twice-differentiable one-to-one monotonic function \( g(\cdot) \), as follows:

\[
\eta_t = g(\mu_t) = \beta x_t + \tau_t 
\]
where

\[ \tau_t = \sum_{j=1}^{p} \phi_j \{ g(y_{t-j}) - x'_{t-j} \beta \} + \theta_t \sum_{j=q}^{t-1} \theta_1^{t-j} \{ g(y_{t-p-j}) - x'_{t-p-j} \beta \}. \]  

(8)

In the GAR-M model, the linear predictor defined in (8) is written such that the effect of the previous \( p \)-instants are only weighted by the autoregressive parameters \( \phi_j, j = 1, \ldots, p \) and the effects of the older observations are taken into account by a single parameter \( \theta_1 \) with decreasing weights excluding the previous \( p+1-q \) instants. For \( q = 1 \), the previous \( p \)-instants that have already been considered by the autoregressive parameters \( \phi_j, j = 1, \ldots, p \) are not considered, and for \( q = 2 \) is excluded one more previous instant just considering the older observations from \( t - p - 2 \). However, in practice \( q = 1 \) is more usual. Note that, when \( \theta_1 = 0 \), the conditional mean depends only on the first \( p \) lagged terms \( g(y_{t-p}) - x'_{t-p} \beta \). The parameters of the GAR-M model can be estimated by the conditional maximum likelihood method using IRLS as in the GARMA models.

### 4.1. Properties of the GAR-M model

In this subsection, theoretical expressions are derived for the marginal mean and variance of \( y_t \) following a GAR-M model with the link functions: identity and natural logarithm. Additionally, restrictions on the parametric space for stationarity conditions for the GAR-M(1,1) model are determined. Recalling, the terms \( \mu_t = E(y_t | H_t) \) and \( \varphi \nu(\mu_t) = \text{Var}(y_t | H_t) \) are respectively the mean and variance of \( y_t \) conditioned on the past information and \( E(y_t | x_t) \) and \( \text{Var}(y_t | x_t) \) represent the marginal (unconditional on the previous information) mean and variance of \( y_t \), where \( \varphi \) and \( \nu(\mu_t) \) are respectively the scale parameter and variance function of the exponential family distribution.

**Theorem 4.1:** The marginal mean and variance of \( y_t \) for the GAR-M model defined by (1) and (8) with identity link function are

(a) \( E(y_t | x_t) = x'_t \beta \),

(b) \( \text{Var}(y_t | x_t) = \varphi \text{E}[\Psi'(B) \nu(\mu_t)] \),

provided that \( \Psi'(B) = 1 + \psi_1^2 B + \psi_2^2 B^2 + \cdots \) (as in (A5) in the Appendix 2) exists. The marginal mean is stationary if \( x'_t \beta = \beta_0 \) for all \( t \). The proofs of Theorems 1a and 1b are respectively presented in Appendices A.1 and A.2.

**Theorem 4.2:** The marginal mean and variance of \( y_t \) for the GAR-M model with the natural logarithm link function are approximated as

(a) \( E(y_t | x_t) \approx e^{x'_t \beta} \left[ 1 - \frac{\psi}{2 \sigma^2} \sum_{j=1}^{t-1} \alpha_j \Psi(B) E \left( \nu(\mu_{t-j}) \right) \right] \)

(b) \( \text{Var}(y_t | x_t) \approx \left[ E(y_t) \right]^2 \Psi^2(B) \left\{ \varphi E \left( \frac{\nu(\mu_t)}{\mu_t^2} \right) + \text{E}(a_t^2) \right\} \)

provided that \( \Psi'(B) = 1 + \psi_1^2 B + \psi_2^2 B^2 + \cdots \) (as in (A5)) exists. As in the Theorem 4.1, the marginal mean is stationary if \( x'_t \beta = \beta_0 \) for all \( t \). The expressions for \( \alpha's, \)
at and $c_x$ are defined respectively in (9), (A7) and (A10). The proofs are presented in Appendices A.3 and A.4. For the Poisson distribution, the stationary marginal mean and variance of $y_t$ of the GAR-M model with identity link and natural logarithm link functions are given respectively by

$$E(y_t | x_t) = \beta_0,$$

$$\text{Var}(y_t | x_t) = \Psi^{(2)}(1)\beta_0,$$

and

$$E(y_t | x_t) \approx \exp(\beta_0) \left[ 1 - \frac{\Psi^{(2)}(1)\beta_0}{2c_x^2} \sum_{j=1}^{t-1} \alpha_j \right],$$

$$\text{Var}(y_t | x_t) \approx [E(y_t | x_t)]^2 \Psi^2(B) \left\{ E \left( \frac{1}{\mu_t} \right) + E(a_t^2) \right\}$$

where $\Psi^{(2)}(1) = 1 + \psi_1^2 + \psi_2^2 + \cdots$ and the $\alpha_j$’s are the parameters of the GARMA($p,q$) model, written as

$$\alpha_j = \begin{cases} 
\phi_j & \text{if } 1 \leq j \leq p \\
\theta_1^{j-p} & \text{if } j \geq p + q.
\end{cases}$$

4.2. Restrictions on the parametric space of the GAR-M

The marginal mean of $y_t$ in the GAR-M($p,q$) models with link functions identity and natural logarithm is constant if $x'_t \beta = \beta_0$ for all $t$ and the operator $\Psi(B)$ exists (that is, the roots of the characteristic autoregressive polynomial are outside the unit circle). The relationship between the roots and coefficients may be used to show that the following two conditions are necessary for stationarity [8]. In order to get roots greater than one, both conditions

$$\sum \alpha_j < 1$$

$$|\alpha_j| < 1$$

are necessary but not sufficient. For the GAR-M($1,q$) model, the first condition ($\sum \alpha_j > 1$) leads to an additional parametric restriction,

$$\phi_1 < \frac{\theta_1^{n-q} + 1 - 2\theta_1}{1 - \theta_1},$$

where the parameters $\phi_1$ and $\theta_1$ assume values in the interval $(-1,1)$. Figure 3 shows the restricted parameter space of the GAR-M($1,1$) model for stationary processes.

4.3. Residual analysis

The residual analysis can be performed as in the GARMA models. To verify if the residuals are uncorrelated and normally distributed, it is recommended to analyse respectively their autocorrelation function and their normal Q–Q plot. Benjamin et al. [2] advocate to use the
Figure 3. The parameter restriction space for GAR-M(1,1) model.

normalized conditional (randomized) quantile residuals of Dunn and Smyth [9] for discrete GARMA models, as the distribution of the deviance and Pearson residuals are highly nonnormally distributed for count data with low fitted means. The normalized randomized quantile residuals are given by $q_t = G^{-1}(u_t)$, where $G^{-1}$ is the inverse cumulative distribution function of a standard normal variable and $u_t$ is a random value simulated from the uniform distribution in the interval $[F(y_t - 1, \hat{\mu}_t), F(y_t, \hat{\mu}_t)]$, where $F(y_t - 1, \hat{\mu}_t)$ is the fitted conditional cumulative distribution function.

5. Simulation study

Two simulation studies were performed to measure the consistency of the parameter estimators of the GAR-M model in terms of the mean square errors and to compare the performance of the GARMA($p,q$) and GAR-M($p,q$) models with similar serial correlation function. Only moderate correlations ($< 0.7$) between the autoregressive (AR) and moving average (MA) term were considered, because for higher correlation values it was not possible to obtain the maximum likelihood estimates of the parameters of GARMA models using IRLS due to its non-convergence.

5.1. Consistency of the GAR-M parameters estimators

To measure the consistency of the parameter estimators of the GAR-M model, time series $y_t$ with different lengths $n$ were simulated 10,000 times according to a Poisson-GAR-M(1,1) model with a logarithm link function and a Gamma-GAR-M(1,1) model with an identity link function, then the parameters of the GAR-M model are estimated using the IRLS method. The average, the standard deviation (S.D.) and the mean square error (MSE) of estimators are obtained. Asymptotic confidence intervals also were calculated assuming that the maximum likelihood (ML) estimators are asymptotically Gaussian, unbiased with variance given by the inverse of Fisher information matrix, to determine the coverage percentage. In simulations studies, it was confirmed empirically that the normal distribution
can be assumed as distribution for the parameters of the GARMA and GAR-M models for \( n \geq 100 \).

Time series were simulated with linear predictor given by

\[
g(\mu_t) = \beta_0 + \beta_1 \cos(2\pi t/52.25) + \tau_t,
\]

where \( \tau_t \) is defined in (8).

**Scenario 1a.** In the first scenario, time series were simulated according to a Poisson-GAR-M(1,1) model with linear predictor given in (11), natural logarithm as link function and parameters \( \beta_0 = \ln(19) \approx 2.944; \beta_1 = \ln(1.28) \approx 0.246; \phi_1 = 0.22 \) and \( \theta_1 = 0.1 \). Table 1 shows descriptive measures of the estimates considering time series with different lengths (\( n \)). The estimated values, for \( n = 100 \), are close to the true values, the coverage percentages are higher than 89\% and the BIAS and MSE values are low.

**Scenario 1b.** In this scenario, time series \( y_t \) were simulated according to a Gamma-GAR-M(1,1) model with linear predictor given in (11), identity as link function and parameters \( \beta_0 = 19; \beta_1 = 1.28; \phi_1 = 0.22; \theta_1 = 0.1 \). Descriptive measures of the estimates for the scenario 1b are shown in Table 2. In the case of \( n = 100 \), the estimated values are close to the true values and the coverage percentages for all parameters are higher than 90\%. It is worth to note that, in this case, the MSE values are higher for \( \beta_0 \) and \( \beta_1 \) in relation to \( \phi_1 \) and \( \theta_1 \).

---

**Table 1.** Descriptive measures of estimates for Scenario 1a.

<table>
<thead>
<tr>
<th>( n )</th>
<th>Parameter</th>
<th>True value</th>
<th>Mean</th>
<th>S.D.</th>
<th>% coverage</th>
<th>BIAS</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>( \beta_0 )</td>
<td>2.944</td>
<td>2.940</td>
<td>0.029</td>
<td>90.9</td>
<td>0.004</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>0.246</td>
<td>0.247</td>
<td>0.040</td>
<td>90.4</td>
<td>-0.001</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>( \phi_1 )</td>
<td>0.220</td>
<td>0.194</td>
<td>0.100</td>
<td>94.0</td>
<td>0.026</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.064</td>
<td>0.098</td>
<td>95.4</td>
<td>0.036</td>
<td>0.010</td>
</tr>
<tr>
<td>500</td>
<td>( \beta_0 )</td>
<td>2.944</td>
<td>2.943</td>
<td>0.016</td>
<td>93.0</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>0.246</td>
<td>0.247</td>
<td>0.022</td>
<td>92.9</td>
<td>-0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( \phi_1 )</td>
<td>0.220</td>
<td>0.215</td>
<td>0.043</td>
<td>94.8</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.093</td>
<td>0.041</td>
<td>95.7</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>1000</td>
<td>( \beta_0 )</td>
<td>2.944</td>
<td>2.944</td>
<td>0.012</td>
<td>94.9</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>0.246</td>
<td>0.247</td>
<td>0.016</td>
<td>94.1</td>
<td>-0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>( \phi_1 )</td>
<td>0.220</td>
<td>0.219</td>
<td>0.035</td>
<td>95.9</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.096</td>
<td>0.033</td>
<td>96.5</td>
<td>0.004</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Table 2.** Descriptive measures of estimates for Scenario 1b.

<table>
<thead>
<tr>
<th>( n )</th>
<th>Parameter</th>
<th>True value</th>
<th>Mean</th>
<th>S.D.</th>
<th>% coverage</th>
<th>BIAS</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>( \beta_0 )</td>
<td>19.01</td>
<td>19.00</td>
<td>1.431</td>
<td>90.5</td>
<td>-0.010</td>
<td>1.782</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>1.280</td>
<td>1.275</td>
<td>1.090</td>
<td>89.9</td>
<td>0.005</td>
<td>1.329</td>
</tr>
<tr>
<td></td>
<td>( \phi_1 )</td>
<td>0.220</td>
<td>0.196</td>
<td>0.097</td>
<td>92.9</td>
<td>0.024</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.062</td>
<td>0.086</td>
<td>93.9</td>
<td>0.038</td>
<td>0.011</td>
</tr>
<tr>
<td>500</td>
<td>( \beta_0 )</td>
<td>19.00</td>
<td>19.00</td>
<td>0.947</td>
<td>94.1</td>
<td>-0.001</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>1.280</td>
<td>1.284</td>
<td>0.601</td>
<td>93.5</td>
<td>-0.004</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td>( \phi_1 )</td>
<td>0.220</td>
<td>0.215</td>
<td>0.046</td>
<td>93.9</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.093</td>
<td>0.042</td>
<td>94.1</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>1000</td>
<td>( \beta_0 )</td>
<td>19.00</td>
<td>18.99</td>
<td>0.325</td>
<td>94.1</td>
<td>&lt;0.001</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>1.280</td>
<td>1.281</td>
<td>0.455</td>
<td>95.0</td>
<td>-0.001</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>( \phi_1 )</td>
<td>0.220</td>
<td>0.218</td>
<td>0.037</td>
<td>94.7</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.096</td>
<td>0.034</td>
<td>95.7</td>
<td>0.004</td>
<td>0.001</td>
</tr>
</tbody>
</table>
5.2. Performance of the GARMA\((p,q)\) and GAR-M\((p,q)\) models

In order to compare the performance of the GARMA\((1,1)\) and GAR-M\((1,1)\) models, a second simulation study was performed. Stationary time series \(y_t\) were simulated according to a Poisson distribution with a logarithm link function and a Gamma distribution with an identity link function. To avoid multicollinearity problems in the Poisson-GARMA\((1,1)\) model, the values of the autoregressive and moving average parameters considered were \(\phi_1 = 0.7\) and \(\theta_1 = 0.1\), respectively. Additionally, no covariate was included and the parameter \(\beta_0 = \ln(19) \approx 2.944\). For this model, the correlation between the autoregressive term \(\{g(y_{t-1}) - \beta_1\}\) and moving average term \(\{g(y_{t-1}) - \eta_{t-1}\}\) is around 0.65. An analogous model, in the sense to have a similar serial correlation function, is a Poisson-GAR-M\((1,1)\) model with a logarithm link function and parameters \(\beta_0 = \ln(19) \approx 2.944\), \(\phi_1 = 0.64\), \(\theta_1 = 0.05\). Also, it was considered GAR-M with the same parameters considered for the GARMA\((1,1)\) model. The similarity of the autocorrelation functions may be visualized in Figure 4.

Table 3 shows the estimates of the parameters of the Poisson-GARMA\((1,1)\) and Poisson-GAR-M\((1,1)\) models, respectively, for \(n = 500\). Note that, for both models the average of the parameter estimates is close to the true values and the percentage coverages confidence intervals for both models are similar. For the second scenario, a Gamma-GARMA model

![Figure 4. Sample ACF of \(y_t\) for Poisson-GARMA\((1,1)\) with \(\phi_1 = 0.7, \theta_1 = 0.1\) and GAR-M\((1,1)\) with \(\phi_1 = 0.64, \theta_1 = 0.05\) models.](image)

**Table 3.** Mean, standard deviation, C.I. coverage, Bias and MSE for the GARMA and GAR-M models – Scenario 2a – Poisson.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True value</th>
<th>Mean</th>
<th>S.D.</th>
<th>% Coverage</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GARMA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>2.944</td>
<td>2.941</td>
<td>0.036</td>
<td>90.8</td>
<td>0.003</td>
<td>0.009</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>0.700</td>
<td>0.688</td>
<td>0.045</td>
<td>92.3</td>
<td>0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>0.100</td>
<td>0.104</td>
<td>0.042</td>
<td>95.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GAR-M (similar ACF)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>2.944</td>
<td>2.940</td>
<td>0.032</td>
<td>91.0</td>
<td>0.004</td>
<td>0.007</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>0.640</td>
<td>0.635</td>
<td>0.043</td>
<td>94.3</td>
<td>0.005</td>
<td>0.010</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>0.050</td>
<td>0.047</td>
<td>0.042</td>
<td>95.7</td>
<td></td>
<td>0.007</td>
</tr>
<tr>
<td><strong>GAR-M (same parameters)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>2.944</td>
<td>2.941</td>
<td>0.035</td>
<td>89.9</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>0.700</td>
<td>0.689</td>
<td>0.043</td>
<td>95.3</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>0.100</td>
<td>0.102</td>
<td>0.041</td>
<td>94.7</td>
<td></td>
<td>0.002</td>
</tr>
</tbody>
</table>
Figure 5. ACF of \( y_t \) for Gamma-GARMA(1,1) and GAR-M(1) models.

Table 4. Mean, standard deviation, C.I. coverage, Bias and MSE for the GARMA and GAR-M models – Scenario 2b – Gamma.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True Value</th>
<th>Mean</th>
<th>S.D.</th>
<th>% Coverage</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>19.00</td>
<td>18.969</td>
<td>0.800</td>
<td>89.9</td>
<td>0.031</td>
<td>0.203</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>0.700</td>
<td>0.688</td>
<td>0.059</td>
<td>94.3</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.103</td>
<td>0.043</td>
<td>95.1</td>
<td>-0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>GAR-M (similar ACF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>19.00</td>
<td>18.97</td>
<td>0.666</td>
<td>90.1</td>
<td>0.028</td>
<td>0.102</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>0.640</td>
<td>0.626</td>
<td>0.045</td>
<td>97.3</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.045</td>
<td>0.040</td>
<td>0.034</td>
<td>94.7</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>GAR-M (same parameters)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>19.00</td>
<td>18.97</td>
<td>0.791</td>
<td>89.7</td>
<td>0.030</td>
<td>0.181</td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>0.700</td>
<td>0.689</td>
<td>0.052</td>
<td>94.5</td>
<td>0.010</td>
<td>0.002</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.100</td>
<td>0.096</td>
<td>0.036</td>
<td>96.7</td>
<td>0.006</td>
<td>0.001</td>
</tr>
</tbody>
</table>

was considered with parameters \( \beta_0 = 19 \), \( \phi_1 = 0.7 \) and \( \theta_1 = 0.1 \) and a similar model, as done for the Poisson simulation, a Gamma-GAR-M(1,1) model with parameters \( \beta_0 = 19 \), \( \phi_1 = 0.64 \) and \( \theta_1 = 0.045 \). In this case, it was considered an identify link function. For this GARMA model, the correlation between the autoregressive and moving average terms is around 0.66. The similarity of two autocorrelation functions may be visualized in Figure 5. The results are similar to the Scenario 2a as presented in Table 4. The standard errors are a little lower for the GAR-M model than the values obtained for the GARMA models.

6. Real data analysis

In this section, a GAR-M(1,1) model is fitted to analyse the count time series of the daily number of hospital admissions due to respiratory diseases for people aged over 60 years in the city of São Paulo – Brazil from October 2012 to April 2015 including explanatory variables. The daily hospitalization time series is obtained from the Hospital Information System at the Health Secretary of São Paulo (PRO-AIM). The logarithm of the expected number of hospitalizations is

\[
\eta_t = \ln(\mu_t) = x_t' \beta = \beta_0 + \beta_1 \sin(2\pi t / 365) + \beta_2 \text{Mon}_t + \beta_3 \text{Fri}_t + \beta_4 \text{Sat}_t + \beta_5 \text{Sun}_t + \tau_t
\] (12)
where Mon\(_t\), Fri\(_t\), Sat\(_t\) and Sun\(_t\) are indicator variables equal to 1 for each weekday and zero otherwise to consider weekly seasonality \[10\] and

\[
\tau_t = \phi_1 \left\{ g(y_{t-1}) - x_{t-1}' \beta \right\} + \sum_{j=1}^{t-2} \theta_j^j \left\{ g(y_{t-1-j}) - x_{t-1-j}' \beta \right\}.
\] (13)

as in (8).

Firstly, a GLM model was fitted to the data but their first residual autocorrelations were significant. Then a GAR-M(1,0) and a GAR-M(0,1) models are considered but the residuals were still autocorrelated. Thus, the GAR-M(1,1) was fitted and finally their residual autocorrelations were all non-significant as presented in Figure 7. A GAR-M(2,0) model was also considered but the parameter \(\phi_2\) was not significant (\(p > .05\)). Figure 6 shows the observed daily number of admissions, its predicted values using the Negative Binomial GAR-M(1,1) model with the linear predictor as in (13).

Figure 7 presents the quantile-quantile plot of the Gaussian residuals and they seem to be normally distributed. Also, the normality hypothesis was accepted using the Shapiro–Wilk test (\(p = .110\)), confirming that the assumption of a Negative Binomial distribution for the counts is appropriate. The estimates of all coefficients for the NB-GAR-M(1,1) are in Table 5. The autoregressive and moving average parameters are significant (\(p < .05\)). The daily residuals (\(\ln(y_t) - x_t' \beta\)) vary from \(-0.5\) to \(0.5\) and the estimated autoregressive parameter is 0.222. This means that an increase of 0.1 in the previous daily residual leads to

\[\text{Figure 6. Number of admissions due to respiratory diseases for people aged over 60 years in São Paulo from October 2012 to April 2015 and predicted counts.}\]

\[\text{Figure 7. Correlogram and Q–Q plot of residuals for GAR-M (1,1) model. (a) ACF plot and (b) Q–Q plot.}\]
Table 5. Estimates, standard errors, and p-values GAR-M(1,1) model.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>S.E.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (\beta_0)</td>
<td>3.717</td>
<td>0.015</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Sine (\beta_1)</td>
<td>-0.132</td>
<td>0.008</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Sunday (\beta_2)</td>
<td>-0.298</td>
<td>0.022</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Monday (\beta_3)</td>
<td>0.047</td>
<td>0.021</td>
<td>.023</td>
</tr>
<tr>
<td>Friday (\beta_4)</td>
<td>-0.084</td>
<td>0.021</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Saturday (\beta_5)</td>
<td>-0.233</td>
<td>0.022</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>0.222</td>
<td>0.031</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>0.062</td>
<td>0.031</td>
<td>.043</td>
</tr>
<tr>
<td>Dispersion ((k))</td>
<td>0.005</td>
<td>.001</td>
<td></td>
</tr>
</tbody>
</table>

an average increase of 2.2\%\((\exp(\hat{\phi}_1 \times 0.1))\) in the average daily number of hospitalizations of elderly people due to respiratory diseases in São Paulo city.

On the other hand, it was not possible to get the maximum likelihood estimates of the parameters of Negative Binomial-GARMA\((1,1)\) model using IRLS due to its non-convergence. However, the parameters of this GARMA\((1,1)\) model may nowadays be estimated using the `optim` function in R (as recommended in the new update package `garmaFit` [11] available in 2017). Even though, the interpretation of parameters is unclear once the estimates of autorregressive and moving average parameters, that take into account simultaneously the effect of the previous day, have different signals \(\hat{\phi}_1 = 0.362\ (0.052)\) and \(\hat{\theta}_1 = -0.130\ (0.045)\). The negative term of the moving average (MA) coefficient estimate seems to cancel part of the over estimated autoregressive (AR) estimate. In addition, the standard errors of the AR and MA estimators (even for \(\beta\)'s) for the Negative Binomial-GARMA\((1,1)\) model seems inflated due to the multicollinearity, since they are larger than the S.E. of the GAR-M estimates (see Table 5). Finally, for this real data, the sample Pearson correlation between the autoregressive term \(g(y_{t-1}) - x'_{t-1}\beta\) and moving average term \(g(y_{t-1}) - \eta_{t-1}\) for the fitted GARMA\((1,1)\) model was 0.898.

7. Discussions and conclusions

The main goal of this paper is to analyse the structure of the GARMA\((p,q)\) models, proposed by Benjamin et al. [2], with \(p\) and \(q\) simultaneously different from zero. In the GARMA models, the conditional mean is related, as in the GLM, to a linear predictor by a twice-differentiable one-to-one monotonic link function. However, the GARMA model also includes lagged values of the dependent variable and covariates through the autoregressive terms and lagged residuals, for example, the Pearson residuals, through the moving average terms. These both terms are included to model possible serial correlations present among observations measured along time. However, the simultaneous inclusion of autoregressive and moving average terms in the linear predictor, independent of the choice of the link function and the distribution from the exponential family, induces multicollinearity in the GARMA\((p,q)\) models.

For ease the development in this paper, a GARMA\((1,1)\) model was deconstructed writing each linear predictor \(\eta_t\) as function of lagged linear predictors and lagged observations in order to understand their contributions to the predictor. It was observed that the effect of the previous terms, \(g(y_{t-1}) - x'_{t-1}\beta\), is taken into account simultaneously by the autoregressive \(\phi_1\) and the moving average \(\theta_1\) parameters inducing multicollinearity,
turning unclear the interpretation of these parameters and leading, in many times, to the non-convergence of the algorithm (IRLS). For 50% of the parametric space \((-1, 1)^2\), the sample Pearson correlation between the autoregressive term \(g(y_{t-1}) - x'_{t-1} \beta\) and the moving average term \(g(y_{t-1}) - \eta_{t-1} = g(y_{t-1}) - x'_{t-1} \beta + c_t\) is greater than 0.65.

In general, the structure of linear predictor of the GARMA\((p, q)\) models with \(p\) and \(q\) simultaneously different from zero induces multicollinearity because the previous terms are measured simultaneously by autoregressive and moving average parameters.

To deal with these problems, a modified model, namely GAR-M\((p, q)\) model, is proposed (which also takes into account the long range dependence). For the proposed model, the linear predictor is written such that the effect of the previous \(p\)-instants are only weighted by the autoregressive parameters \(\phi_j, j = 1, \ldots, p\), and the effect of older observations are taken into account by a single parameter \(\theta_1\), with decreasing weights. In simulation studies, it was shown that the estimates are close to the true values with high confidence interval coverage for the parameters, as presented by the GARMA models. In a general sense, the alternative model leads better to the multicollinearity problem and improves the interpretation of the parameters of the model. However, the parameter space is more restricted to achieve stationarity.

An interesting feature of the GAR-M\((1, 1)\) models is that, in several cases, for small values of the parameter \(\theta_1\), it is possible to reproduce a similar autocorrelation function of the GARMA\((p, q)\) models. In addition, the asymptotic normality of the parameters of the GAR-M model, analysed by the simulations studies, indicated that the normal distribution can be assumed for time series with lengths \(n\) greater than 100.

Finally, a GAR-M\((1, 1)\) model with Negative Binomial distribution with logarithm function as link function was used to model the count time series of the daily hospital admission due to respiratory diseases for people aged over 60 years in the city of São Paulo – Brazil from October 2012 to April 2015. This model includes seasonal components (day-of-the-week and the sine function) in the expected number of hospitalizations. The autoregressive and moving average parameters were significant \((p < .05)\). The analysis of residuals confirmed that no assumption was violated. Additionally, GARMA\((p, 0)\) models were fitted, but only for \(p = 1, \phi_1\) was significant. For \(p = 2\) and \(3\), only \(\phi_1\) was significant. For the GARMA\((0, q)\), the same occurred, only \(\theta_1\) was significant for \(q = 1, 2, 3\). Considering only GARMA\((1, 0)\) and GARMA\((0, 1)\), the residuals were still correlated as in the paper. Only our GAR-M\((1, 1)\) was able to remove all the residual autocorrelation.

On the other hand, it is not possible to get the maximum likelihood estimates of the parameters of a Negative Binomial-GARMA\((1, 1)\) model using IRLS due to its non-convergence. However, the parameters of GARMA\((1, 1)\) model may nowadays be estimated using the optim function in R (new version of the package garmaFit). Even though, in our real data analysis, the interpretation of parameters is unclear once the estimates of the autoregressive and moving average parameters have different signals \((\hat{\phi}_1 = 0.362 (0.052)\) and \(\hat{\theta}_1 = -0.130 (0.045)\)). The negative term of the moving average coefficient estimate seems to cancel part of the over estimated autorregressive estimate and the standard error for this GARMA\((1, 1)\) model is inflated due to the multicollinearity in comparison to the S.E. of the estimators using the GAR-M\((1, 1)\) model.

Making successive substitutions, the linear predictor for the GARMA\((1, 1)\) model can be written as \(\eta_t = x'_{t} \beta + (\phi_1 + \theta_1)\{g(y_{t-1}) - x'_{t-1} \beta\} + \theta_1 c_t\), where \(c_t\) is a small value in
relation to the lagged residual \( \{g(y_{t-1}) - X'_{t-1} \beta \} \). This representation shows that there is parameter redundancy and this fact may lead to problems of identifiability in the GARMA models. In Shumway and Stoffer [1] these problems of parameter redundancy in the linear ARMA models are discussed. The authors comment that if we were unaware of parameter redundancy, we might claim the data are correlated when in fact they are not.

In a general sense, the modified GARMA models, proposed in this paper, have a similar performance to the GARMA models, reduce the multicollinearity and improve the interpretation of the autoregressive and moving average parameters. However some restriction in the parametric space are imposed to guarantee the stationarity of the process. On the other hand, the GARMA model may be used including only autoregressive or moving average parameters. It may be sufficient to model the serial correlation. For instance, in Benjamin et al. [2] a GARMA(0,2) model with a negative binomial conditional distribution was fitted to a well-known time series dataset of poliomyelitis counts and in Dugas et al. [5] a GARMA(3,0) model was used to model the number of emergency department (ED) visits and hospitalizations in the United States integrating Google Flu Trends information.

**Funding**

The authors would like to acknowledge CNPq, CAPES and FAPESP (Grant number 2018/04654-9) for the financial support.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Orlando Albarracin http://orcid.org/0000-0001-6037-6518

Airlane Pereira Alencar http://orcid.org/0000-0002-0779-0426

Linda Lee Ho http://orcid.org/0000-0001-9984-8711

**References**


Appendices

Appendix 1. Estimation of the parameters

The model parameters denoted by \( \gamma = (\beta', \phi', \theta')' \) are estimated by the conditional maximum likelihood. The likelihood function is the product of conditional densities \( f(y|H_t) \). Then, the log-likelihood of the data \( \{y_{m+1}, \ldots, y_n\} \) conditioned on the first \( m \) observations and on \( \eta_t = g(y_t^*) \) for \( t = 1, 2, \ldots, i, (i = \max(p, q)) \) and \( m \geq i \) is given by

\[
l = \sum_{i=m+1}^{n} \ln f(y_t|H_t).
\]

(A1)

An iterative Fisher’s score method can be used to maximize the conditional log-likelihood \( l \) function leading to an iterative reweighted least squares (IRLS). This procedure is similar to the optimization procedure used for GLM.

Let \( A_1 \) and \( A_2 \) matrices defined as follows

\[
A_1 = \begin{bmatrix}
g(y_m) - x_m^\prime \beta & \ldots & g(y_{m+1-p}) - x_{m+1-p}^\prime \beta \\
g(y_{m+1}) - x_{m+1}^\prime \beta & \ldots & g(y_{m+2-p}) - x_{m+2-p}^\prime \beta \\
\vdots & & \vdots \\
g(y_{n-1}) - x_{n-1}^\prime \beta & \ldots & g(y_{n-p}) - x_{n-p}^\prime \beta 
\end{bmatrix},
\]

\[
A_2 = \begin{bmatrix}
g(y_m) - \eta_m & \ldots & g(y_{m+1-q}) - \eta_{m+1-q} \\
g(y_{m+1}) - \eta_{m+1} & \ldots & g(y_{m+2-q}) - \eta_{m+2-q} \\
\vdots & & \vdots \\
g(y_{n-1}) - \eta_{n-1} & \ldots & g(y_{n-q}) - \eta_{n-q}
\end{bmatrix}.
\]

The linear predictor in (3), for the observations \( y_{m+1}, \ldots, y_n \), can be written in the matrix form as

\[
\eta_t = B'y_t,
\]

where \( B = [X \quad A_1 \quad A_2] \), and \( X \) is the matrix of explanatory variables. The conditional maximum likelihood estimators (MLE) can be obtained by an iterative weighted least squares process. At the \((k+1)\) iteration, the estimation is updated \( \gamma^{(k+1)} \) as

\[
\gamma^{(k+1)} = (B'W^{(k)}B)^{-1}B'W^{(k)}z^{(k)},
\]

(A2)

where \( z^{(k)} \) is the dependent variable and \( W^{(k)} \) is a diagonal matrix of weights, that change at each \( k \)-th iteration, given by the expressions

\[
z = \eta + W^{-1/2}V^{-1/2}(y - \mu)
\]

and

\[
W = \text{diag}(w_{m+1}, \ldots, w_n)
\]

where \( y = (y_{m+1}, \ldots, y_n)' \), \( \mu = (\mu_{m+1}, \ldots, \mu_n) \), \( V = \text{diag}(v(\mu_{m+1}), \ldots, v(\mu_n)) \) is a diagonal matrix of variance functions and \( w_j = v(\mu_j)^{-1}(d^2g(\mu_j)/d\mu_j)^{-2} \).
The convergence of (A2) occurs generally in a finite number of steps. Evaluating \( \gamma \) using IRLS allows to obtain the approximate large-sample conditional variance of \( \hat{\gamma} \) as a by product of the iterative process (Green [12]; Kaufmann [13]). It can be shown that asymptotically \( \sqrt{(n-m)}(\hat{\gamma} - \gamma) \sim N(0, I(\gamma)^{-1}) \), where

\[
I(\gamma) = \lim_{n \to \infty} \frac{n}{n-m} \left\{ \sum_{t=m}^{n} \left( \frac{\partial \eta_t}{\partial \gamma} \right) \left( \frac{\partial \eta_t}{\partial \gamma} \right)' \right\}.
\]

**Appendix 2. Proofs of Theorems 4.1 and 4.2**

**A.1 Proof of Theorem 4.1(a)**

Considering the identity link function, the linear predictor in (3) can be written as

\[
\mu_t = x_t' \beta + \sum_{j=1}^{t-1} \alpha_j \left\{ y_{t-j} - x_{t-j} \beta \right\},
\]

where,

\[
\alpha_j = \begin{cases} 
\phi_j & \text{if } 1 \leq j \leq p \\
(\theta)^{-p} & \text{if } j \geq p + q.
\end{cases}
\]

Let \( y_t = \mu_t + \epsilon_t \). Then the \( \epsilon_t \) are uncorrelated errors with mean 0. Replacing \( \mu_t \) by (A3) and defining \( w_t = y_t - x_t \beta \) gives

\[
w_t = \sum_{j=1}^{t-1} \alpha_j w_{t-j} + \epsilon_t
\]

\[
(1 - \alpha_1 - \cdots - \alpha_{t-1})w_t = \epsilon_t
\]

\[
\Phi(B)w_t = \epsilon_t
\]

\[
w_t = \Psi(B)\epsilon_t
\]

where

\[
\Psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots
\]

with \( \sum_{j=0}^{\infty} |\psi_j| < \infty \), provided that the roots of \( \Phi(B) \) are outside the unit circle. Then

\[
E(y_t|x_t) = E(x_t' \beta + w_t) = x_t' \beta,
\]

as \( E(w_t) = 0, \forall t \).

**A.2 Proof of Theorem 4.1(b)**

The marginal variance of \( y_t \) is given by

\[
\text{Var}(y_t|x_t) = \text{Var}(w_t + x_t' \beta | x_t) = E(w_t^2)
\]

\[
= E\left( \sum_{j=0}^{\infty} \psi_j B^j \epsilon_t \sum_{i=0}^{\infty} \psi_i B^i \epsilon_t \right)
\]

\[
= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_i E(\epsilon_{t-j} \epsilon_{t-i})
\]

\[
= \sum_{j=0}^{\infty} \psi_j^2 E(\epsilon_{t-j}^2)
\]
where $\Psi^{(2)}(B) = 1 + \psi^2B + \psi^2B^2 + \cdots$. The variance of the martingale errors is, $\text{Var}(\epsilon_t) = E(\epsilon_t^2) = E[E(\epsilon_t^2|H_t)]$, where, $\text{Var}(\epsilon_t|H_t) = \text{Var}((y_t - \mu_t)|H_t) = \text{Var}(y_t|H_t) = \varphi\nu(\mu_t)$. Hence

$$\text{Var}(y_t|x_t) = \varphi E[\Psi^{(2)}(B)\nu(\mu_t)].$$

### A.3 Proof of Theorem 4.2(a)

Considering the ln link function, $g(\mu_t) = \ln(\mu_t)$, the linear predictor in (3) can be written as

$$\ln(\mu_t) = x_t'\beta + \sum_{j=1}^{t-1} \alpha_j \left\{ \ln(y_{t-j}) - x_{t-j}'\beta \right\}, \quad (A6)$$

where $\alpha_j$ is defined in (A4). Applying the Taylor series expansion for $\ln(y_t)$ at the point $\mu_t$ gives

$$\ln(y_t) = \ln(\mu_t) + \frac{1}{\mu_t}(y_t - \mu_t) + a_t, \quad (A7)$$

where, $a_t$ is the approximation error. Substituting $\ln(\mu_t)$ from (A6) in (A7) and let $w'_t = \ln(y_t) - x_t'\beta$, gives

$$w'_t = \sum_{j=1}^{t-1} \alpha_j w'_{t-j} + \frac{1}{\mu_t}(y_t - \mu_t) + a_t$$

$$\Phi(B)w'_t = \frac{1}{\mu_t}(y_t - \mu_t) + a_t$$

$$w'_t = \Psi(B) \left[ \frac{1}{\mu_t}(y_t - \mu_t) + a_t \right]$$

where $\Psi(B) = \Phi^{-1}(B)$. Hence

$$E(w'_t) = \Psi(B)E(a_t), \quad (A8)$$

because $E(y_t - \mu_t/\mu_t) = 0$. Then

$$E[\ln(y_t)] = x_t'\beta + \Psi(B)E(a_t).$$

Taking the exponential function of both sides in (A6) and applying the Taylor series for the function $e^x$ at the point $x = 0$ yields

$$\mu_t = \exp(x_t'\beta) \left[ 1 + \sum_{j=1}^{t-1} \alpha_j \left\{ \ln(y_{t-j}) - x_{t-j}'\beta \right\} + b_t \right], \quad (A9)$$

where $b_t$ is the approximation error. Hence

$$E(\mu_t|x_t) = \exp(x_t'\beta) + \exp(x_t'\beta) \left[ \sum_{j=1}^{t-1} \alpha_j E\left\{ \ln(y_{t-j}) - x_{t-j}'\beta \right\} + E(b_t) \right]$$

$$= \exp(x_t'\beta) + \exp(x_t'\beta) \left[ \sum_{j=1}^{t-1} \alpha_j E\left( w'_{t-j} \right) + E(b_t) \right]$$

$$= \exp(x_t'\beta) + \exp(x_t'\beta) \left[ \sum_{j=1}^{t-1} \alpha_j \Psi(B)E(a_{t-j}) + E(b_t) \right].$$
The error \( a_t \) in (A7) can be rewritten as

\[
a_t = -\frac{(y_t - \mu_t)^2}{2c_x^2} = -\frac{\epsilon_t^2}{2c_x^2}, \text{ for some } c_x \in (\mu_t, y_t). \tag{A10}
\]

Applying the law of iterated expectations, the \( E[\epsilon_t^2] \) is expressed by

\[
E[\epsilon_t^2] = E[E(\epsilon_t^2|H_t)] = E(\text{Var}(\epsilon_t|H_t)) = E[\text{Var}(y_t|H_t)] = E[\varphi(\mu_t)].
\]

Thus, \( E(a_t) \) is given by

\[
E(a_t) = \frac{-E[\varphi(\mu_t)]}{2c_x^2}.
\]

Hence

\[
E(y_t|x_t) = E[E(y_t|H_t)] = e^{\phi \beta} \left[ 1 - \frac{\varphi}{2c_x^2} \sum_{j=1}^{n} \alpha_j \Psi(B)E[\psi(\mu_{t-j})] + b_t \right]
\]

where \( b_t \), the remainder term in (A9), can be written as \( b_t = e^{b_x}/2[\log(\mu_t) - x'_t\beta]^2 \) for some \( b_x \in (0, \log(\mu_t) - x'_t) \).

### A.4 Proof of Theorem 4.2(b)

In this proof, we will use the delta method for deriving the variance of \( \ln(y_t) \). To apply the delta method, the first order Taylor expansion about \( \ln(y_t) \) at the point \( \mu_t \) is calculated

\[
\ln(y_t) = \ln(\mu_t) + \frac{1}{\mu_t}(y_t - \mu_t) + \delta_t,
\]

where \( \delta_t = o_p(y_t - \mu_t) \) [14]. Hence the variance of the \( \ln(y_t) \) is approximately

\[
\text{Var}[\ln(y_t)] \approx \frac{1}{[E(y_t)]^2} \text{Var}(y_t).
\]

The variance of \( \ln(y_t) \) is given by

\[
\text{Var}[\ln(y_t)] = \text{Var}(\omega'_t) = E(\omega'_t^2) - (E(\omega'_t))^2.
\]

where \( \omega'_t \) is defined in the proof of the Theorem 4.2(a). Thus,

\[
E(\omega'_t^2) = E(\Psi(B)(s_t + a_t)\Psi(B)(s_t + a_t)),
\]

denoting \( s_t = y_t - \mu_t/\mu_t \), we have

\[
E(\omega'_t^2) = E \left( \sum_{i=0}^{\infty} \psi_i(s_{t-i} + a_{t-i}) \sum_{j=0}^{\infty} \psi_j(z_{t-j} + a_{t-j}) \right)
\]

\[
= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j E \left\{ (s_{t-i} + a_{t-i})(s_{t-j} + a_{t-j}) \right\}
\]

Assuming that \( s_t \) is uncorrelated and furthermore also that \( a_t \) and \( s_t \) are uncorrelated, it yields

\[
E(\omega'_t^2) = \Psi^{(2)}(B) \{ E(s_t^2) + E(a_t^2) \}.
\]

with the variance of \( s_t \) expressed as

\[
\text{Var}(s_t) = E(s_t^2)
\]

\[
= E \left\{ E \left[ \left( \frac{y_t - \mu_t}{\mu_t} \right)^2 \, | H_t \right] \right\}
\]
\begin{align*}
&= E \left\{ \frac{1}{\mu_t^2} \left[ \text{Var}(y_t|H_t) + (E(y_t|H_t))^2 \right] - \frac{2}{\mu_t} E(y_t|H_t) + 1 \right\} \\
&= \varphi E \left\{ \frac{v(\mu_t)}{\mu_t^2} \right\}.
\end{align*}

Hence

$$\text{Var}(y_t) \approx [E(y_t)]^2 \psi^2(B) \left\{ \varphi E \left( \frac{v(\mu_t)}{\mu_t^2} \right) + E(\sigma_t^2) \right\},$$

where $E(y_t)$ is given in (A11).