Modelos generalizados auto-regressivos e de médias móveis: gráficos de controle, multicolinearidade e novo modelo modificado

Orlando Yesid Esparza Albarracin

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Doutor em Ciências

Programa: Estatística

Orientador: Prof. Dra. Airlane Pereira Alencar

Coorientador: Prof. Dra. Linda Lee Ho

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES/CNPq/FAPESP

São Paulo, Outubro de 2017

Modelos generalizados auto-regressivos e de médias móveis: gráficos de controle, multicolinearidade e novo modelo modificado

Esta é a versão original da tese elaborada pelo candidato Orlando Yesid Esparza Albarracín, tal como submetida à Comissão Julgadora.

Generalized autoregressive and moving average models: control charts, multicollinearity, and a new modified model.

Esta versão da dissertação/tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 24/10/2017. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Profª. Drª. Airlane Pereira Alencar - IME-USP
- Profª. Drª. Denise Aparecida Botter - IME-USP
- Prof. Dr. Roberto da Costa Quinino - UFMG
- Prof. Dr. Marcela A. Guerreiro Machado UNESP - Guaratinguetá
- Prof. Drª. Thelma Sáfadi - UFLA

,

Agradecimentos

Con esta tesis le pongo punto final a esa historia que comencé a escribir en diciembre de 2012 en Bucaramanga, en mi país natal Colombia. En São Paulo descubrí que la función seno era perfecta para modelar mis días en el país de la samba. En los períodos $[0,\pi/2]$ descubrí que los límites son mentales y que es posible volver realidad los sueños. En los períodos $[\pi/2, 3\pi/2]$ entendí que el talento no lo era todo y que el estudio solo era una parte de mí existir. Finalmente, en los períodos $[3\pi/2,2\pi]$ encontré personas que me dieron su mano en señal de apoyo y pude colocarle, uno que otro, color a esta ciudad gris.

Hoy mi nombre me recuerda a quien nunca le podré decir adiós. Papá, como no estoy seguro si después de la muerte se puede seguir leyendo quiero que todos sepan que tienes un hijo doctor gracias a los esfuerzos tuyos y de nuestra Esperanza.

Un agradecimiento especial a mi Esperanza, espero que sea lo último que pierda. A mis dos hermanos: Nancy y Sneider, gracias por todo.

Doutor Alberto, gracias por todo. Después de tres o cuatro o cinco reencarnaciones tendré como retribuirle todo el apoyo que me dio.

Angeliquita, mujer, gracias por aguantarme, ayudarme y contar hasta veintidós después de escucharme.

Gracias familia y amigos !

Quiero agradecer, en destaque, en mi idioma, a mis dos orientadoras: Airlane e Linda sin ellas esta historia no tendría punto final.

Aos meus amigos brasileiros, não sei se existe outro povo mais legal do que vocês, obrigado por tudo!

Resumo

Albarracín, O. Y. E. **Generalized autoregressive and moving average models: gráficos de controle, multicolinearidade e um modelo modificado.** 2017. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

Recentemente, no campo da saúde, gráficos de controle têm sido propostos para monitorar a morbidade ou a mortalidade decorrentes de doenças. Este trabalho está composto por três artigos. Nos dois primeiros artigos, gráficos de controle CUSUM e EWMA foram propostos para monitorar séries temporais de contagens com efeitos sazonais e de tendência usando os modelos Generalized autoregressive and moving average models (GARMA), em vez dos modelos lineares generalizados (GLM), como usualmente são utilizados na prática. Diferentes estatísticas baseadas em transformações, para variávies que seguem uma distribuição Binomial Negativa, foram usadas nestes gráficos de controle. No segundo artigo foram propostas duas novas estatísticas baseadas na razão da função de log-verossimilhança. Diferentes cenários que descrevem perfis de doenças foram considerados para avaliar o efeito da omissão da correlação serial nesses gráficos de controle. Este impacto foi medido em termos do Average Run Lenght (ARL). Notou-se que a negligência da correlação serial induz um aumento de falsos alarmes. Em geral, todas as estatísticas monitoradas apresentaram menores valores de ARL_0 para maiores valores de autocorrelação. No entanto, nenhuma estatística entre as consideradas mostrou ser mais robusta, no sentido de produzir o menor aumento de falsos alarmes nos cenários considerados.

No último artigo, foram estudados os modelos GARMA (p, q) com p e q simultaneamente diferentes de zero, uma vez que duas características foram observadas na prática. A primeira é a presença de multicolinearidade, que induz à não-convergência do método de máxima verossimilhança usando mínimos quadrados ponderados reiterados. A segunda é a inclusão dos mesmos termos defasados nos componentes autorregressivos e de médias móveis. Um modelo modificado, GARMA-M, foi apresentado para lidar com a multicolinearidade e melhorar a interpretação dos parâmetros. Em sentido geral, estudos de simulação mostraram que o modelo modificado fornece estimativas mais próximas dos parâmetros e intervalos de confiança com uma cobertura percentual maior do que a obtida nos modelos GARMA. No entanto, algumas restrições no espaço paramétrico são impostas para garantir a estacionariedade do processo. Por último, uma análise de dados reais ilustra o ajuste do modelo GARMA-M para o número de internações diárias de idosos devido a doenças respiratórias de outubro de 2012 a abril de 2015 na cidade de São Paulo, Brasil.

Palavras-chave: Gráficos de controle, Series temporais, CUSUM, GARMA, Multicolinearidade.

Abstract

Albarracín, O. Y. E. **Generalized autoregressive and moving average models: gráficos de controle, multicolinearidade e um modelo modificado.** 2017. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

Recently, in the health surveillance area, control charts have been proposed to decide if the morbidity or mortality of a specific disease reached an epidemic level. This thesis is composed by 3 papers. In the first two papers, CUSUM and EWMA control charts were proposed to monitor count time series with seasonal and trend effects using the Generalized Autoregressive and Moving Average models (GARMA), instead of the independent Generalized Linear Model (GLM) as it is usually used in practice. Different statistics based on transformations, for variables that follow a Negative Binomial distribution, were used in these control charts. In the second paper, two new statistics were proposed based on the ratio of log-likelihood function. Different scenarios describing disease profiles were considered to evaluate the effect of omission of serial correlation in EWMA and CUSUM control charts. The performance of CUSUM and EWMA charts when the serial correlation is neglected in the regression model was measure in terms of average run length (ARL). In summary, when the autocorrelation is neglected, fitting a pure GLM instead of a GARMA model will lead to an increase of false alarms. However, no statistics among the tested ones seem to be robust, in a sense to produce the smallest increase of false alarms in all scenarios. In general, all monitored statistics presented a smaller ARL_0 for higher values of autocorrelation.

In the last paper, the GARMA models (p, q) with p and q simultaneously different from zero were studied since that two features were observed in practice. One is the multicollinearity, which may lead to a non-convergence of the maximum likelihood, using iteratively reweighted least squares. The second is the inclusion of the same lagged observations into the autoregressive and moving average components confounding the interpretation of the parameters. In a general sense, simulation studies show that the modified model provide estimators closer to the parameters and offer confidence intervals with higher coverage percentage than obtained with the GARMA model, but some restrictions in the parametric space are imposed to guarantee the stationarity of the process. Also, a real data analysis illustrate the GARMA-M fit for daily hospilatization rates of elderly people due to respiratory diseases from October 2012 to April 2015 in São Paulo city, Brazil.

Keywords: Control charts, surveillance, CUSUM, EWMA, GARMA, Multicollinearity.

Sumário

Capítulo 1

Introduction

In the manufacturing area the concept of quality is related to ensuring which the final product only presents random fluctuations, that in some cases are uncontrollable variations due to humidity, tear of machinery or other conditions. Process monitoring allows to decide when the variation is natural and when corrections are necessary avoiding major losses when problems are detected as soon as they occur. In the surveillance area, the control charts have been adopted, due to its easy implementation, to monitor the morbidity or mortality of diseases to detect epidemics, as well as to take urgent decisions and to plan health services.

A control chart consists basically of points which representing a statistic calculated from sample observations and three reference lines: an upper control limit (UCL), a lower control limit (LCL), and a center line (LC) that represents the target value of the characteristic to be monitored. If some point is above the UCL or below the LCL, the process is called out-of-control. On the other hand, if all points are within the control limits, the process is considered in-control. The performance of the control chart is generally based on the distribution of the number of points required until a first point exceeds the control limits. One of these parameters and the most used in the literature is the Average Run Lenght (ARL) at a given quality level is the average number of samples taken before an action signal is given.

Recently, in the public health context, control charts have been built based on generalized linear models (GLM) [McCullagh e Nelder, 1989]. In these models, important covariate information may be included as the seasonal pattern (since more hospitalizations due to several diseases are expected during the winter), adjustment for at-risk population or other explanatory variables. Some control charts based on GLMs are found for example in Hohle e Mazick [2009], Rogerson e Yamada [2004] and Unkel *et al.* [2012]. A key assumption usually made on count process monitoring is the independence of observations. However, autocorrelation may be present in time series data and may have a strong impact on the properties of control charts as discussed in Chao-Wen e Reynolds Jr [2001].

Another proposal to monitor count time series consists in building charts based on diverse models for count time series designed for different types of marginal distribution and autocorrelation structure. Heinen [2003] introduced the INGARCH models that can be understood as a GLM with Poisson distribution and the identity link as a systematic component. Weiß [2009] showed that the simple two-parametric INARCH(1) model is able to describe both the observed serial dependence and overdispersion satisfactorily. It can be considered as a counterpart to the very popular Poisson INAR(1) model, but being able to describe overdispersion.

In the surveillance context, some control charts for count data are found for example in Weiß e Testik [2011]. They proposed the Poisson INAR(1) CUSUM chart and investigated the effects of assuming a Poisson model when there is overdispersion and also showed an extension, the Winsorized Poisson INAR(1) CUSUM chart, for achieving robustness. Other proposals for surveillance of Negative Binomial daily disease counts are found in Sparks *et al.* [2010, 2011]. In the first paper, an adaptive cumulative sum (CUSUM) plan is developed for signaling unusually high incidence in counts with a non-homogeneous mean and in the second an optimal exponentially weighted moving average

(EWMA) is proposed for detecting epidemics.

The present thesis is composed by 3 paper. In the first two papers, CUSUM and EWMA control charts are proposed to monitor count time series using the generalized autoregressive and moving average (GARMA) model [Benjamin *et al.*, 2003] instead of the independent GLM. The impact, in terms of average run length (ARL), is measured on the performance of CUSUM and EWMA charts when the serial correlation is neglected in the regression model by simulation studies considering scenarios that describe the profiles of time series of infectious diseases. The GARMA model extends the univariate Gaussian ARMA time series model to a flexible observation-driven model for non-Gaussian time series data, combining the GLM flexibility to model a function of the conditional expected value and the inclusion of lagged terms to model the autocorrelation. Conditioned on the past history of the process, the distribution of the dependent variable belongs to the exponential family of distributions (as the Gaussian, Poisson, Gamma and Binomial distributions).

In the first paper, different statistics based on normalizing transformations, the deviance residual and the likelihood ratio assuming that observations follow a Negative Binomial distribution are used to build CUSUM control charts to monitor counts with time varying means, including trend and seasonal effects. The monitoring of the weekly number of hospital admissions due to respiratory diseases for people aged over 65 years in the city São Paulo-Brazil is considered as an illustration of the current method.

In the second paper, three like-EWMA control charts are built in order to detect faster upward changes since, in the public health area, large observations are of concern since they may be associated to epidemics and two alternative statistics were proposed to monitor autocorrelated count time series assuming the Negative Binomial distribution.

The last paper, was motivated by two observed features in practical applications of GARMA(p,q) models (with p and q simultaneously different from zero) 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, to find the estimates of the parameters. The second is the inclusion of the same lagged observations into the autoregressive and moving average components confounding the interpretation of the parameters. A modified model, GARMA-M, is presented to deal with the multicollinearity and improves the interpretation of the parameters. In a general sense, simulation studies show that the modified model provide estimators closer to the parameters and offer confidence intervals with higher coverage percentage than obtained with the GARMA model, but some restrictions in the parametric space are imposed to guarantee the stationarity of the process. Also, A real data analysis illustrate the GARMA-M fit for daily hospitalization rates of elderly people due to respiratory diseases from October 2012 to April 2015 in São Paulo city, Brazil.

1.1 Goals

The main goal of the first two papers is to evaluate the importance of taking into account the serial correlation when modelling time series of counts measuring, by simulation, the performance of CUSUM e EWMA control charts to monitor autocorrelated count time series when a GLM model with trend and seasonal effects is fitted instead of the true GARMA model, that generated the data, with the same covariates. This performance is measured in terms of the Average Run Lengths (ARLs). In the last paper, the main aim 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. Additionally, to propose a modified GARMA model, namely GARMA-M to reduce the problem of multicollinearity in the GARMA model and to have a better interpretation of the parameters.

1.2 Contributions

The main contributions of this work are the following:

- To propose control charts that take into account the serial correlation when time series of count are monitored.
- To measure in terms of the Average Run Lengths (ARLs) the performance of CUSUM and EWMA charts when the serial correlation is neglected fitting a GLM instead of the true GARMA model.
- To propose two alternative statistics to monitor autocorrelated count time series assuming the Negative Binomial distribution.
- To propose a modified GARMA model, namely GARMA-M to reduce the problem of multicollinearity in the GARMA model and to have a better interpretation of the parameters.

1.3 Thesis organization

This thesis is organized as follows: The first paper, CUSUM chart to monitor autocorrelated counts using Negative Binomial GARMA model is presented in Section 2, the second paper, EWMA chart for count time series using GARMA model is preseted in Section 3 and the last paper, Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model in Section 4.

CUSUM chart to monitor autocorrelated counts using Negative Binomial GARMA model

1 Orlando Yesid Esparza Albarracn, IME-USP ∗ E-mail: orlando@ime.usp.br
2 Airlane Pereira Alencar, IME-USP
3 Linda Lee Ho, EP-USP, Brasil

Abstract

Cumulative sum (CUSUM) control charts have been used for health surveillance due to its efficiency to detect soon small shifts in the monitored series. However, these charts may fail when data are autocorrelated. An alternative procedure is to build a control chart based on the residuals after fitting autoregressive moving average (ARMA) models but these models usually assume Gaussian distribution for the residuals.
In practical health surveillance, count series can be modelled by Poisson or Negative Binomial regression, this last to control overdispersion. To include serial correlations, generalized autoregressive moving average (GARMA) models are proposed. The main contribution of the current paper is to measure the impact, in terms of average run length (ARL) on the performance of CUSUM charts when the serial correlation is neglected in the regression model. Different statistics based on transformations, the deviance residual and the likelihood ratio are used to build CUSUM control charts to monitor counts with time varying means, including trend and seasonal effects. The monitoring of the weekly number of hospital admissions due to respiratory diseases for people aged over 65 years in the city São Paulo-Brazil is considered as an illustration of the current method.

Keywords: Health Surveillance; Average Run Length; Autocorrelation; Time series; Control charts.

1 Introduction

Control charts have been often used to monitor production processes. Due to their easy implementation, their usage has been adopted in other sciences as to monitor service quality, public health surveillance and so on. An example is the application of CUSUM control charts to monitor count time series, like the

weekly number of hospitalizations, to decide if the morbidity or mortality of a specific disease reached an epidemic level. CUSUM control charts are frequently used in health surveillance since they are more efficient than the Shewhart chart for detection of small shifts. Recently, in the surveillance context [1;2;3;4], control charts based on generalized linear models (GLM) [5] have been built. In these control charts, important covariate information may be included as the seasonal pattern (since more hospitalizations due to several diseases are expected during the winter), adjustment for at-risk population or other explanatory variables. Some control charts based on GLMs are found for example in Hohle and Mazick [6], Rogerson and Yamada [7] and Rossi et al [8].

A key assumption usually made on count process monitoring is the independence of observations. However, autocorrelation may be present in time series data and may have a strong impact on the properties of control charts as discussed in Chao-Wen and Reynolds Jr [9].

In this paper, CUSUM control charts are used for monitoring the expected value of count time series where the observations are generated by a generalized autoregressive and moving average (GARMA) model [10] instead of the independent generalized linear model (GLM). The GARMA model extends the univariate Gaussian ARMA time series model to a flexible observation-driven model for non-Gaussian time series data, combining the GLM flexibility to model a function of the conditional expected value and the inclusion of lagged terms to model the autocorrelation. Conditioned on the past history of the process, the distribution of the dependent variable belongs to the exponential family of distributions. This model is used to describe a variety of time-dependent response variables which also have time-dependent covariates. In the surveillance area, the GARMA model with a negative binomial distribution were proposed by Benjamin et al. [10] and used to model the well-known time series of poliomyelitis counts and Dugas et al. [11] evaluated the effects of Google Flu Trends, meteorological and temporal information on the number of influenza-related patient visits.

In the literature, diverse models for count time series have been proposed, designed for different types of marginal distribution and autocorrelation structure. Heinen [12] introduced the INGARCH models that can be understood as a GLM with Poisson distribution and the identity link as a systematic component. Weiß [13] showed that the simple two-parametric INARCH(1) model is able

to describe both the observed serial dependence and overdispersion satisfactorily. It can be considered as a counterpart to the very popular Poisson INAR(1) model, but being able to describe overdispersion. Recently Weiß[14] showed how to approximate the marginal process distribution with the help of the Poisson-Charlier expansion and the goodness of this approximation is investigated, also in view of approximating the average run lengths (ARLs).

In the surveillance context, some control charts for count data are found for example in Weiß & Testik[15]. They proposed the Poisson INAR(1) CUSUM chart and investigated the effects of assuming a Poisson model when there is overdispersion and also showed an extension, the Winsorized Poisson INAR(1) CUSUM chart, for achieving robustness. Other proposals for surveillance of Negative Binomial daily disease counts are found in Sparks et al.[16;17]. In the first paper, an adaptive cumulative sum (CUSUM) plan is developed for signaling unusually high incidence in counts with a non-homogeneous mean and in the second an optimal exponentially weighted moving average (EWMA) is proposed for detecting epidemic.

The weekly number of hospital admissions due to respiratory diseases for people aged over 65 years in the city São Paulo-Brazil from 2006 to 2010 is considered as an illustration of the current proposal. The residual analysis indicated that the GARMA(2,0) model is well fitted to these data and the CUSUM chart based on this fit is built to detect epidemics in 2011.

The main goal of the current paper is to evaluate, by simulation, the performance of CUSUM control charts to monitor autocorrelated count time series when a GLM model with trend and seasonal effects is fitted instead of the true GARMA model, that generated the data, with the same covariates. This performance is measured in terms of the Average Run Lengths (ARLs). The CUSUM charts considered in this paper are based on transformed variables, the deviance residual(DR), as well as the CUSUM chart proposed by Rogerson and Yamada[7](RY), extended here to include the Negative Binomial distribution, and the CUSUM chart proposed by Hohle and Paul[18] based on the likelihood ratio (LR) statistic for the Negative Binomial distribution.

This paper is organized as follows: the statistics used to build one-sided CUSUM control charts for GLM and GARMA models are presented in Section 2; the GARMA model fitted and the corresponding CUSUM control chart for the daily number of hospital admissions due to respiratory diseases for people

aged over 65 years is described in Section 3; the evaluation of CUSUM control charts for autocorrelated data based on simulations are commented in Section 4 and conclusions and discussions are outlined in Section 5.

2 Building control chart for GARMA models to monitor count data series

In general, control charts are built to identify shifts in the expected value and these shifts may be described as an increase or a decrease of this mean response. In surveillance studies only larger counts are of concern since they may be associated to epidemics. Therefore, in the present study, all control charts are built only with an upper control limit to detect significant increases in the expectations.

In order to build control charts to monitor count data series, it is assumed that the expected value $\mu_t = E(y_t)$ may vary over time as a function of past observed values $y_{t-1}, ..., y_1$; their past expected values, $\mu_{t-1}, ..., \mu_1$, and explanatory variables $\mathbf{x_t}, ..., \mathbf{x_1}$. Also, only a single observation Y_t is available at each time t and Y_t assumes non negative integer values in $\{0, 1, \dots, \}$.

The GARMA model[10] is a combination of the GLM and the ARMA model, allowing the inclusion of the autoregressive and moving average elements in a generalized linear model. In the GARMA model, the distribution of each observation y_t, for $t = 1, ..., n$, conditioned on the past information $\mathbf{H}_t = (\mathbf{x_t}, ..., \mathbf{x_1}, \mathbf{y_{t-1}}, ..., \mathbf{y_1}, \mu_{\mathbf{t-1}}, ..., \mu_{\mathbf{1}})$, belongs to the exponential family, such as the Gaussian, Poisson, Gamma and Binomial distributions.

In this paper, it is assumed that $Y_t|\mathbf{H}_t$ follows a Negative Binomial (NB) distribution and the sequence $Y_1|\mathbf{H}_1, Y_2|\mathbf{H}_2, \dots, Y_n|\mathbf{H}_n$ consists of conditionally independent random variables. The conditional probability of Y_t for a NB distribution is

$$f(Y_t|\mathbf{H}_t, \gamma) = exp\left\{ Y_t \ln\left(\frac{\mu_t}{\mu_t + \gamma}\right) + \gamma \ln\left(\frac{\gamma}{\mu_t + \gamma}\right) + \ln\left(\frac{\Gamma(\gamma + Y_t)}{\Gamma(Y_t + 1)\Gamma(\gamma)}\right)\right\}, \tag{1}$$

with $E(Y_t|\mathbf{H}_t) = \mu_t$ and $Var(Y_t|\mathbf{H}_t) = \mu_t + \mu_t^2/\gamma$. The gamma function in (1) is an extension of the factorial function and for a positive integer n, $\Gamma(n) = (n-1)!$.

The parametrization in (1) is known as the NB-2 in the literature [19] and obtained as a Poisson model with a gamma random effect or as the distribution of the number of failures until the r-th success, implying a non-constant coefficient of variation ($1+\mu/\gamma$). It is worth noting that for a Poisson distribution $E(Y_t|\mathbf{H}_t) = Var(Y_t|\mathbf{H}_t) = \mu_t$, implying that the NB distribution induces a larger variance than the Poisson, what is called overdispersion.

As the counts Y_t are observed along time, they may present serial correlation and can be modelled in the GARMA model including autoregressive and moving average terms in the linear predictor function for μ_t as:

$$g(\mu_t) = \eta_t = \mathbf{x'_t}\boldsymbol{\beta} + \sum_{\mathbf{j=1}}^{\mathbf{p}} \phi_{\mathbf{j}} \mathbf{A}(\mathbf{y_{t-j}}, \mathbf{x_{t-j}}, \boldsymbol{\beta}) + \sum_{\mathbf{j=1}}^{\mathbf{q}} \theta_{\mathbf{j}} \mathbf{M}(\mathbf{y_{t-j}}, \mu_{\mathbf{t-j}}) \qquad (2)$$

with $\mu_t = E(y_t|\mathbf{H}_t)$. The autoregressive parameters are $\boldsymbol{\phi}'=(\phi_1, ..., \phi_p)$; the moving average parameters are $\boldsymbol{\theta}'=(\theta_1, ..., \theta_q)$; and A and M are functions of the autoregressive and moving average terms, respectively. The complete specification of the GARMA model includes that Y_t follows the NB distribution. The ARMA model with covariates is a particular case when $Y_t|\mathbf{H}_t$ follows a Gaussian distribution and A and M are identity functions.

Specifically, choosing the logarithm as the link function, the linear predictor of the GARMA model is expressed as

$$ln(\mu_t) = \eta_t = \mathbf{x'_t}\boldsymbol{\beta} + \sum_{\mathbf{j=1}}^{\mathbf{p}} \phi_{\mathbf{j}} \left\{\mathbf{ln}(\mathbf{y_{t-j}}) - \mathbf{x'_{t-j}}\boldsymbol{\beta}\right\} + \sum_{\mathbf{j=1}}^{\mathbf{q}} \theta_{\mathbf{j}} \left\{\mathbf{ln}(\mathbf{y_{t-j}}) - \eta_{\mathbf{t-j}}\right\}. \quad (3)$$

This model generalizes the model proposed by Sparks et al. [16] that already included the autoregressive terms to account for the autocorrelation.

The location parameters $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ and the scale parameter (as γ for the NB2 distribution) are estimated by the maximum likelihood method using the iterative weighted least squares, as in the GLM. The diagnostic analysis are based on the normalized conditional quantile residuals proposed by Dunn & Smyth [20] as recommended in Benjamin et al. [10]. The null correlation and Gaussian distribution of these residuals are evaluated respectively by a correlogram and a quantile-quantile plot. Also, the Shapiro-Wilk test is used to evaluate the normality of the residuals.

The expected value of Y_t is denoted as $\mu_{0,t}$ and $\mu_{1,t} = \delta\mu_{0,t}$, $\delta > 0$, when the process is respectively in-control and out of control, and as μ_t when it is not

specified the state of the process.

To monitor non-normally distributed data, a first approach traditionally used in statistical process control consists of applying some transformation on the original count variable in order to get an approximately normal distribution. Then, control charts are built with the transformed variable and the control limits are determined under a normal distribution.

In this paper, several monitored statistics are considered to build the CUSUM control chart. The first four statistics are transformations of the original count data to achieve a standardized normal distribution (one is based on the deviance residuals as proposed in Alencar et al.[21]); and the last two are derived from the ratio of likelihood functions. All statistics are summarized in Table 1 and detailed in Alencar et al.[21].

Table 1: Monitored statistics

Method	$Z_{i,t}$
Rossi et al.[8] (RS)	$\frac{Y_t - 3\mu_t + 2\sqrt{Y_t\mu_t}}{2\sqrt{\mu_t}}$
Jorgensen[22] (JG)	$\frac{Y_t - \mu_t}{\sqrt{\gamma\pi_t/(1-\pi_t)^2}};\ \pi_t = \frac{\mu_t}{(\mu_t+\gamma)}$
Guan[23] (GN)	$\sqrt{\gamma - 0.5}\left(\sqrt{\frac{Y_t + 0.385}{\gamma - 0.75}} - \sqrt{\frac{\mu_t + 0.385}{\gamma - 0.75}}\right),$ $\gamma > 0.75$
Deviance residual (DR)	$sign(Y_t - \mu_t)\sqrt{(d_t^2)}$
Rogerson and Yamada[7](RY)	$Y_t - \frac{-\gamma ln\{(\gamma + \mu_{0,t})/(\gamma + \mu^*_{1,t})\}}{\ln\{\mu^*_{1,t}(\gamma + \mu_{0,t})/\mu_{0,t}(\gamma + \mu^*_{1,t})\}}$
Hohle and Paul[18] (LR)	$\ln\left\{\frac{f_{\mu^*_{1,t}}(y_t)}{f_{\mu_{0,t}}(y_t)}\right\}$

Note: $\mu^*_{1,t} = \Delta\mu_{0,t}$

The d_t^2 in deviance residual is defined as

$$d_t^2 = \begin{cases} 2\phi \ln(1+\mu_t/\gamma), & if\ Y_t = 0 \\ 2Y_t \ln\left(\frac{Y_t}{\mu_t}\right) - 2\gamma(1+Y_t/\gamma)\ln\left(\frac{1+Y_t/\gamma}{1+\mu_t/\gamma}\right) & if\ Y_t > 0. \end{cases}$$

In order to calculate the statistics based on the likelihood ratio, RY and LR, it is necessary to define the out-of-control mean $\mu_1^*(t)$. In this context, it is usual assumed $\mu_1^*(t) = \Delta\mu_0(t)$, with $\Delta = 2$, as in Hohle and Paul[18] and also in our analysis.

It is worth noting that the RY statistics may be written as LR divided by $\ln\left(\frac{2(\mu_{0,t}+\gamma)}{2\mu_{0,t}+\gamma}\right)$ for the NB distribution. More details are presented in the Appendix.

For all monitored statistics, the CUSUM chart may be expressed as

$$C_{i,t} = max(0, C_{i,t-1} + Z_{i,t} - k_i), i = \{\text{RS, JG, GN, DR, RY, LR}\}, \quad (4)$$

where $C_{i,0} = 0$ and $Z_{i,t}$ is the i-th statistic presented in Table 1. Whenever $C_{i,t} > h_i$, the process is considered out-of-control. In general, the distribution of the transformed statistics are not Gaussian (and this was confirmed for all these statistics excepting the deviance residual in Alencar et al.[21]). Due to this fact, the parameters (k, h) of CUSUM charts were exhaustively searched by simulation when the process is in-control and out-of-control.

For the first four statistics of Table 1, a general framework to build the CUSUM charts could be divided in four steps:

- Step 1: Fit a GARMA model which includes explanatory variables to the time series data for a non-epidemic period. It is essential to evaluate whether all the model assumptions are valid. The predicted values of the in-control expectation, $\mu_{0,t}$, is calculated.

- Step 2: Fix the increase (δ) in the mean level for the epidemic period, such that $\mu_{1,t} = \delta\mu_{0,t}$.

- Step 3: Fix the k value and then choose the control limit, h. Simulate time series using the chosen GARMA model for the in-control and out-of-control periods and calculate the ARL_0 and ARL_1, as the mean time until the false and true alarms, respectively. Change the value of h, until yielding the desired ARL_0 and calculate ARL_1.

- Step 4: Change the k value and repeat the Step 3, until getting the pair (k, h) that achieves the smaller value of ARL_1.

In practice, it is usual to adopt $0 < k < 1$, but an exhaustive search must be executed. We considered $k > 1$, but the ARL_1 values have increased in our examples. For the two statistics based on likelihood ratio, the value of k may be negative. In particular, we obtained negative values for k in real data analysis (Table 3). In this analysis the desired ARL_0 is equal to 70 weeks, which is equivalent to the usual value of 500 days proposed in previous reviews about public health surveillance[18]. To estimate the ARLs, 10 thousand run lengths are simulated for the in-control and for out-of-control periods. All data analysis and simulations are implemented in the R software.

To measure the effect of neglecting the serial correlation in surveillance analysis using CUSUM control charts, data were simulated from a GARMA model and the GLM model (under independence and NB distribution) are fitted considering the same explanatory variables and the linear predictor $ln(\mu_t) = \eta_t = \mathbf{x_t'}\boldsymbol{\beta}$. Then the in-control Average Run length (ARL_0), its respective standard deviation (SD), the in-control median run length (MRL_0) and out-of-control average run length (ARL_1) for a shift in the mean of the process were calculated.

3 Real data analysis - GARMA model

In this section, a GARMA model is fitted to the count time series of the weekly number of hospital admissions due to respiratory diseases for people aged over 65 years in the city of São Paulo-Brazil. Weekly data from January, 2006 to December, 2010 are used to fit the model, which include explanatory variables, and the predicted values for 2011 are calculated. Only a Negative Binomial distribution is considered since the scale parameter γ is too small, indicating the presence of overdispersion, since $Var(y_t|H_t) = \mu_t + \mu_t^2/\gamma$. Additionally, the logarithmic function expressed in (3) is chosen as the link function with

$$\mathbf{x}_t'\boldsymbol{\beta} = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{52.25}\right) + \beta_2 \sin\left(\frac{2\pi t}{52.25}\right) + \beta_3 \times t, \tag{5}$$

where t is the number of each week, $t = 1, \ldots, 261$.

Figure 1 shows the observed weekly number of admissions, its predicted values using the NB-GARMA model and using a NB-GLM model with the linear

predictor in (5). Also, the forecasts for 2011 are presented for both models.

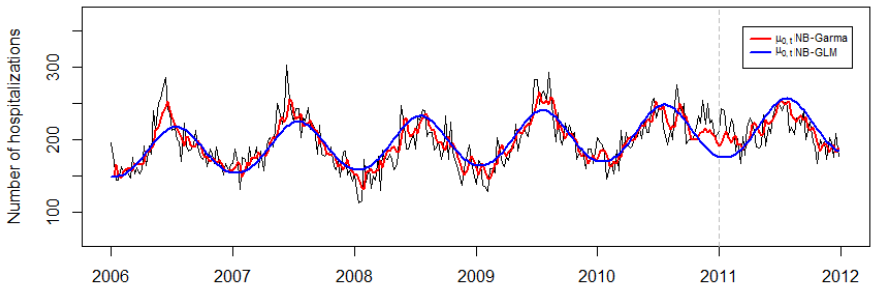


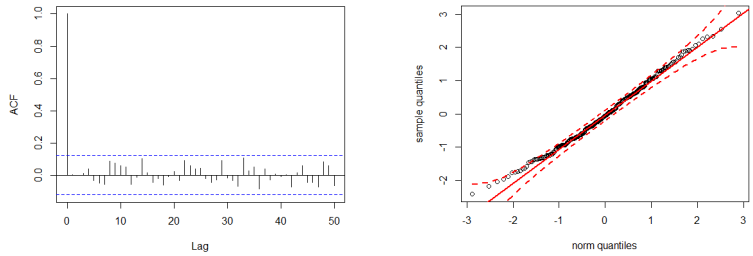
Figure 1: Weekly number of hospitalizations, predicted values until 2010 and forecasts for 2011 using NB-GARMA and NB-GLM models

Firstly, a GLM model was fitted to the data but their first deviance residuals autocorrelation were significant. Then a GARMA(1,0) model is considered, but their residuals were still autocorrelated. Thus the GARMA(2,0) is fitted and finally their residual autocorrelations were all non-significant as presented in Figure 2a. The GARMA(2,0) model was chosen as the best model since the residual analysis indicated that no model assumption seems to be violated. Figure 2b presents the quantile-quantile plot of the Gaussian residuals and they seem to be normally distributed. Also, the normality hypothesis is accepted using the Shapiro Wilk test ($p = 0.2141$), confirming that the assumption of a Negative Binomial distribution for the counts is appropriate. The estimates of all coefficients for the NB-GARMA(2,0) are in Table 2. Both the autoregressive parameters are significant ($p < 0.001$).

Table 2: Estimates, standard errors and p-values - GARMA(2,0) model

Coefficient	Estimate	S.E.	p-value
Intercept (β_0)	5.180	0.033	<0.001
Cosine (β_1)	-0.177	0.021	0.021
Sine (β_2)	-0.048	0.021	<0.001
t ($100\beta_3$)	0.063	0.00021	0.002
(ϕ_1)	0.364	0.061	<0.001
(ϕ_2)	0.219	0.060	<0.001
Dispersion (γ^{-1})	0.0052	0.001	<0.001

Hospitalization data are simulated according to a NB-GARMA (2,0) model ($p = 2, q = 0$) with all parameters equal to the estimates shown in Table 2 to



(a) ACF plot (b) Q-Q plot

Figure 2: Correlogram and Q-Q plot of residuals - GARMA (2,0) model

get 10 thousand simulated run lengths. With in-control simulated run lengths, the parameters h and k are computed to meet an $ARL_0 = 70$ weeks and to provide the lowest value of ARL_1 for shifts $\delta = \{1.25; 1.5\}$. The results are in Table 3. Analysing these results, all statistics present tiny differences in ARL_1 and MRL_1 (only values of ARL_1 are included in Table 3), but differences of one week are observed until detecting an increase of 25% in the mean number of hospitalizations. Lowest ARL_1s are achieved using the LR statistic. In general, it may take six weeks to detect an increase of 25% and almost three weeks to detect a larger increase.

Table 3: CUSUM results for the hospitalization data - GARMA(2,0)

						ARL_1	ARL_1
	k	h	ARL_0	SD	MRL_0	$\delta = 1.25$	$\delta = 1.50$
RS	0.6	4.02	70.05	0.633	50	6.31	2.86
GN	0.3	1.94	69.99	0.634	51	6.26	2.87
JG	0.5	2.62	70.02	0.669	49	6.06	2.75
DR	0.5	2.43	69.96	0.673	49	5.99	2.74
RY	-66.45	50.820	69.98	1.055	49	5.86	3.17
LR	-18.88	12.512	69.98	0.817	49	5.51	2.66

The CUSUM charts depicted in Figure 3 are built to detect increases of at least 25% in 2011, using the values of h and k presented in Table 3. In order to compare all CUSUM charts, each statistic was divided by its corresponding threshold h. So signals are triggered whenever these standardized statistics are larger than 1. All control charts indicate an epidemic scenario in the second week of January, in the beginning of February and April and the second week of December. Indeed, an increase of the number of hospitalizations (more than

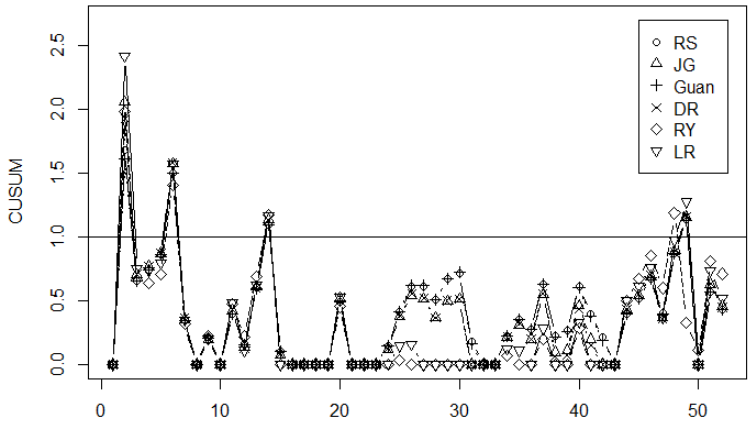


Figure 3: CUSUM control charts for a GARMA(2,0) model

50% higher than the average number of hospitalizations in the previous years) was observed in January 2011, in the second week of February and also in the first week of April.

4 Simulation Study

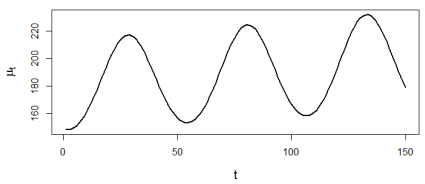
In this section, a simulation study is conducted to measure the impact on the CUSUM control chart performance in terms of ARL, if the correlated data are generated by a NB-GARMA model but the serial correlation is neglected, fitting only the usual NB-GLM.

To measure the effect of serial correlation, time series were simulated from a NB-GLM model with mean $\mu_{0,t}$

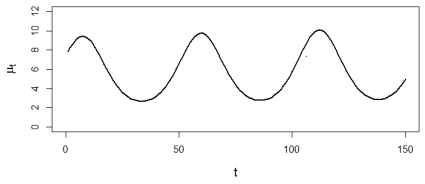
$$ln(\mu_{0,t}) = x_t^{'}\beta = \beta_0 + \beta_1 cos\left(\frac{2\pi t}{52.25}\right) + \beta_2 sin\left(\frac{2\pi t}{52.25}\right) + \beta_3 t, \qquad (6)$$

to get 10 thousand simulated run lengths and three scenarios to describe the different profiles of time-series of infectious diseases, as shown in Fig 4, are used in this study. All coefficients and the dispersion parameter for $\mu_{0,t}^a$ are replaced by the estimates in Table 2, and for $\mu_{0,t}^b$ and $\mu_{0,t}^c$ the coefficients are taken from Hohle and Paul[18].

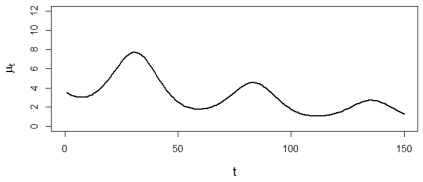
Then, the parameters k and h of CUSUM control charts based on the statis-



(a) $\mu^a_{0,t}$: $\beta = (5.18, -0.177, -0.048, 0.00063)$, $\gamma = 0.0052$.



(b) $\mu^b_{0,t}$: $\beta = (1.6, 0.4, 0.5, 0)$, $\gamma = 0.0052$.



(c) $\mu^c_{0,t}$: $\beta = (1.76, -0.45, -0.39, -0.01)$, $\gamma = 0.16$.

Figure 4: Examples of in-control parameter $\mu_{0,t}$

tics summarized in Table 1 are computed to reach an ARL_0 around 500 days (equivalently around 70 weeks) and to provide the lowest value of ARL_1 for an increase of 25% in the mean weekly number of admissions. In order to measure the effect of autocorrelation in the performance of CUSUM control charts, the NB-GARMA(1,0) and the NB-GARMA(0,1) models with parameters $\phi_1 = \theta_1 = \{0.05; 0.1; 0.2; 0.36\}$ were simulated and the performance of the CUSUM charts developed for the GLM were measured in terms of ARL_0 and ARL_1. The last value 0.36 was included because it is the estimate of ϕ_1 in Table 2. In all simulations, values of standard deviation (SD) for ARL_1 vary around 0.01 and therefore they did not put in Tables 4-6.

Due to the asymmetry of the waiting time until detecting an out-of-control regime, the median run length (MRL) is also presented. This procedure measures the impact of using a CUSUM chart for independent GLM data when the true model is a correlated GARMA model. All the results for GARMA(1,0) and GARMA(0,1) when $\mu_{0,t} = \mu^a_{0,t}$ are presented in Table 4. For the GARMA(1,0) model, the values of ARL_0 and MRL_0 plotted against the values of $\phi_1 \in [0\ ;1]$ for all statistics are respectively in Figures 5a and 5b.

Note that in the presence of serial correlation, i.e. $\phi_1 > 0$ or $\theta_1 > 0$, the lowest values of ARL_0 and MRL_0 were produced by the CUSUM charts built with LR and Rossi statistics, then they lead to a larger increase of false alarms,

Table 4: ARL_0, its SD, MRL_0, ARL1 for an increase of 25% in the mean of GARMA(1,0) and GARMA(0,1) processes when fitting a GLM with mean $\mu^a_{0,t}$.

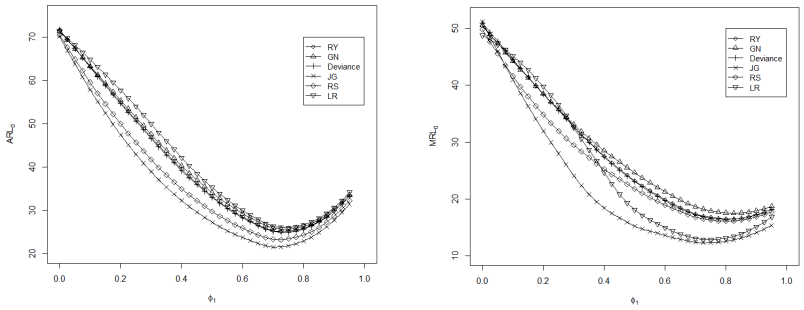
		Statistics					
		RS	GN	JG	DR	RY	LR
	ARL_0 (SD)	69.9 (0.64)	70.1 (0.64)	69.9 (0.69)	69.9 (0.69)	70.0(0.92)	70.0(0.85)
$\phi = 0$	MRL_0	49	49	49	49	47	48
$\theta = 0$	ARL_1	2.07	2.01	1.88	1.89	1.54	1.74
	ARL_0 (SD)	65.4 (0.61)	67.9 (0.63)	67.6(0.66)	67.6(0.66)	68.3 (0.89)	64.4 (0.76)
$\phi = 0.05$	MRL_0	45	47	47	47	47	46
	ARL_1	2.12	2.05	1.93	1.93	1.57	1.78
	ARL_0 (SD)	59.4 (0.54)	63.4 (0.58)	63.1 (0.62)	63.3 (0.62)	64.8 (0.84)	57.1 (0.67)
$\phi = 0.1$	MRL_0	41	45	44	44	47.00	42
	ARL_1	2.13	2.06	1.93	1.94	1.57	1.78
	ARL_0 (SD)	50.4 (0.46)	56.2 (0.52)	54.9 (0.53)	55.4 (0.53)	58.7 (0.75)	48.0(0.55)
$\phi = 0.2$	MRL_0	36	39	39	39	43	35
	ARL_1	2.18	2.13	2.00	2.00	1.61	1.82
	ARL_0 (SD)	37.4 (0.35)	42.7 (0.40)	41.7 (0.41)	42.3 (0.41)	45.2 (0.57)	35.0 (0.39)
$\phi = 0.36$	MRL_0	27	30	29	29	26	20
	ARL_1	2.32	2.29	2.14	2.15	1.71	1.91
	ARL_0 (SD)	65.3 (0.60)	67.6 (0.63)	67.5(0.66)	67.3(0.66)	68.3 (0.88)	64.4 (0.79)
$\theta = 0.05$	MRL_0	45	47	47	47	47	46
	ARL_1	2.10	2.03	1.91	1.89	1.54	1.77
	ARL_0 (SD)	59.5 (0.55)	63.8 (0.59)	63.3 (0.62)	63.2 (0.62)	68.4 (0.86)	59.1 (0.67)
$\theta = 0.1$	MRL_0	41	45	44	44	47.00	42
	ARL_1	2.14	2.08	1.95	1.93	1.56	1.80
	ARL_0 (SD)	50.5 (0.47)	55.7 (0.52)	55.4 (0.55)	55.7 (0.55)	60.0 (0.77)	48.2(0.56)
$\theta = 0.2$	MRL_0	35	38	38	39	43	34
	ARL_1	2.17	2.14	1.98	1.97	1.60	1.82
	ARL_0 (SD)	39.1 (0.35)	43.7 (0.40)	42.5 (0.41)	42.8 (0.41)	44.3 (0.56)	35.9 (0.40)
$\theta = 0.36$	MRL_0	28	31	30	30	25	21
	ARL_1	2.26	2.21	2.07	2.08	1.69	1.88
	k	1.0	0.7	1.0	1.0	-50.11	-18.16
	h	2.753	0.945	1.468	1.322	19.86	11.07

while for the statistic RY the false alarms increase more slowly. Also, the RY statistics presents the lowest ARL_1 values, indicating that in this case study, the RY seems to present the best performance. The same pattern is found for the GARMA(0,1) model. Another result is that the first residual autocorrelation is close to the value of ϕ for the GARMA(1,0) model and close to θ for the GARMA(0,1) model. For the first estimated autocorrelation close to 0.2, obtained simulating a GARMA(1,0) with $\phi_1 = 0.2$ or a GARMA(0,1) with $\theta_1 = 0.2$, false alarms may be detected 15 to 20 weeks earlier than expected, since the desired ARL_0 is 70 and it decays to 55 or 50 depending on the statistics.

For the first simulated scenario ($\mu^a_{0,t}$) and some values of ϕ, the values of k and h (in the Appendix) were searched in order to provide an $ARL_0 = 70$ and the best ARL_1 for an increase of 25%. These values are really close to the values found for the independent GLM model for small values of ϕ, so the values

for k and h for the GLM may be used as initial values for the GARMA (1,0). Results for GARMA (0,1) are not reported here because they are similar to the GARMA(1,0).

The results for GARMA(1,0) model for the other two scenarios are presented in Tables 5 and 6, respectively. When $\phi_1 = 0.0$, the statistics LR and RY lead the highest ARL_1 values, since they detect averagely 4 weeks later an increase of 25% in the mean than the other statistics. Although, these statistics produce to the smallest increase of false alarms when $\phi_1 > 0$, i.e, in presence of serial correlation. For $\phi_1 = 0.2$ the ARL_0 decays to 61 when the LR and RY statistics are implemented, while for other statistics the ARL_0 decays to around 50 weeks. Finally, for the scenario $\mu^c_{0,t}$ all statistics present tiny differences in ARL_0 and ARL_1. However, the JG statistic presents the highest ARL_0 values in presence of serial correlation, but also the highest ARL_1 values. Results of the scenarios $\mu^b_{0,t}$ and for $\mu^c_{0,t}$ for the GARMA(0,1) are not reported here because they are similar to the GARMA(1,0), (as observed for the first scenario (see Table 4)).



(a) ARL_0 (b) MRL_0

Figure 5: Plots of ARL_0 and MRL_0 versus ϕ_1 for observations following NB-GARMA$(1, 0)$ for $\mu_{0,t} = \mu^a_{0,t}$

5 Conclusions and discussions

The main goal of this paper is to evaluate the importance of taking into account the serial correlation when modelling time series of counts. It is very well known that neglecting serial correlation or heteroscedaticity in regression models leads to biased variance estimators, but our goal is to measure the impacts on the

Table 5: ARL_0, its SD, MRL_0, ARL1 for an increase of 25% in the mean of GARMA(1,0) processe when fitting a GLM with mean $\mu^b_{0,t}$.

		Statistics					
		RS	GN	JG	DR	RY	LR
	ARL_0 (SD)	69.9 (0.67)	69.9 (0.67)	69.9 (0.66)	70.0 (0.67)	69.9 (0.69)	69.9 (0.69)
$\phi = 0$	MRL_0	50	52	49	50	48	48
$\theta = 0$	ARL_1	8.52	8.22	8.76	8.58	12.60	12.57
	ARL_0 (SD)	64.4 (0.63)	64.2 (0.60)	66.4 (0.64)	64.4 (0.62)	69.0 (0.67)	68.7 (0.67)
$\phi = 0.05$	MRL_0	45	48	46	45	48	47
	ARL_1	8.79	8.46	9.21	8.84	12.88	12.93
	ARL_0 (SD)	58.9 (0.57)	58.5 (0.56)	60.5 (0.58)	58.5 0.56)	66.7 (0.65)	67.0 (0.60)
$\phi = 0.1$	MRL_0	42	43	42	42	46	46
	ARL_1	8.94	8.68	9.39	8.97	13.06	13.17
	ARL_0 (SD)	50.3 (0.49)	50.4 (0.49)	52.4 (0.50)	50.0 (0.49)	61.2 (0.66)	61.2 (0.60)
$\phi = 0.2$	MRL_0	35	35	36	35	42	41
	ARL_1	9.21	8.99	9.64	9.21	13.35	13.45
	ARL_0 (SD)	40.2 (0.39)	40.8 (0.40)	41.4 (0.40)	39.8 (0.39)	50.4 (0.49)	50.5 (0.49)
$\phi = 0.36$	MRL_0	28	27	29	27	35	35
	ARL_1	9.95	9.81	10.38	9.90	13.22	13.30
	k	0.40	0.10	0.40	0.40	0.10	0.10
	h	2.883	1.704	3.197	2.650	3.617	2.346

Table 6: ARL_0, its SD, MRL_0, ARL1 for an increase of 25% in the mean of GARMA(1,0) processe when fitting a GLM with mean $\mu^c_{0,t}$.

		Statistics					
		RS	GN	JG	DR	RY	LR
	$\text{ARL}_0(SD)$	70.0 (0.85)	69.9 (1.10)	69.9 (0.63)	69.9 (0.77)	69.9 (1.07)	70.0 (0.84)
$\phi = 0$	MRL_0	39	32	51	45	30	42
$\theta = 0$	ARL_1	18.57	17.42	19.94	19.11	18.11	19.77
	$\text{ARL}_0(SD)$	63.9 (0.73)	61.5 (0.95)	65.8 (0.57)	63.7 (0.67)	62.5 (0.98)	63.8 (0.82)
$\phi = 0.05$	MRL_0	38	31	49	42	30	38
	ARL_1	18.80	17.51	20.37	19.28	18.19	19.83
	$\text{ARL}_0(SD)$	60.2 (0.68)	56.5 (0.86)	63.3 (0.56)	59.4 (0.62)	60.1 (0.94)	59.5 (0.75)
$\phi = 0.1$	MRL	36	29	48	40	30	36
	ARL_1	18.89	17.43	20.65	19.29	18.42	20.03
	$\text{ARL}_0(SD)$	51.0 (0.51)	45.8 (0.62)	55.0 (0.47)	50.7 (0.48)	49.4 (0.68)	51.1 (0.57)
$\phi = 0.2$	MRL_0	34	28	43	36	29	34
	ARL_1	19.27	17.59	21.20	19.59	18.66	20.44
	$\text{ARL}_0(SD)$	40.6 (0.37)	35.3 (0.39)	44.8 (0.38)	40.2 (0.36)	37.9 (0.41)	40.4 (0.39)
$\phi = 0.36$	MRL_0	29	24	34	29	26	28
	ARL_1	19.66	17.29	21.23	19.46	18.26	19.61
	k	0.1	0.2	0.2	0.1	0.1	0.1
	h	5.002	1.228	4.617	3.647	4.335	1.921

average run length to detect outbreaks if the CUSUM chart is based on the independent GLM model, as it is usual in the surveillance area.

A model that allies the inclusion of autocorrelation and allows to consider the Poisson or Negative Binomial distribution is the GARMA model proposed by Benjamin et al. [10]. Higher values of autoregressive or moving average parameters implies higher autocorrelations for the time series generated by this model.

Our real data analysis fitted an GARMA(2,0) model for the weekly number of hospital admissions due to respiratory diseases for people over 65 years old in São Paulo city from January 2006 to December 2010. This model includes a seasonal component and a linear trend in the expected number of hospitalizations. The estimated trend coefficient is 0.00063, implying that after each year it is expected an annual increase of 3.3% ($exp(\beta_3 * 52) = 1.033$) in the mean number of hospitalizations of elderly people due to respiratory diseases in São Paulo city. The CUSUM control charts for 2011 considering the statistics RS, GN, JG, DR, RY and LR indicated alarms in January, February, April and December. These weeks really presented higher number of hospitalization (an increase higher than 35%), indicating that all methods are able to detect deviations in the mean number of hospitalizations.

The impact on the performance of the CUSUM control charts to monitor count time series is measured in terms of the Average Run Lengths (ARLs) when the autocorrelation is ignored and a GLM model under independence is fitted, as it is supposed to occur in practice, instead of a GARMA model, that is used to simulate the time series.

The CUSUM performances were compared in a simulation study considering 3 scenarios that include linear trend and seasonal components, this latter consists of cosine and sine functions with cycles of 52 weeks. The first scenario is based on the estimated model for the hospitalizations, the second one is also simulated in Hohle and Paul[18], and the third scenario is based on the analysis of real data presented in Hohle and Paul[18].

In general, all monitored statistics presented a smaller ARL_0 for higher values of autocorrelation. The RY statistics presented a better performance in the first scenario; the RY and LR statistics, both based on the likelihood, presented higher ARLs in the second case; and the JG statistics was the best in the third scenario. There is no evidence that one of these statistics is more robust than the other in case of autocorrelated data.

In summary, when the autocorrelation is neglected fitting a pure GLM instead of a GARMA model will lead to an increase of false alarms. However no statistics among the tested ones seem to be robust, in a sense to produce the smallest increase of false alarms in the evaluated models. About ARL_1, it looks that the misspecification of the autocorrelation does not provoke a serious delay. Since in Tables 5-6 for each statistic $Z_{i,t}$ the ARL_1 for correlated data

($\phi \neq 0$ or $\theta \neq 0$) is similar to the ARL_1 for uncorrelated data ($\phi = 0$ and $\theta = 0$).

The effect of autocorrelation in CUSUM charts was also analysed by Chao Wen and Reynolds Jr[9]. They consider an autoregressive process with an additional random error, which may be written as an ARMA(1,1) model. Based on simulations, CUSUM charts of the original observations and residuals are built and the ARL values are very similar for both approaches. Our proposal is different since we consider that, in practice, several surveillance studies build control charts after fitting a GLM to estimate the time-varying mean of the process, and then we measure the effects of autocorrelated GARMA processes using the CUSUM charts designed for the GLM under independence.

Other extensions of the GLM in the literature may be considered to model correlated data and build control charts as the generalized seasonal autoregressive integrated moving average for counts in Briet et al.[24] and models that include conditional heteroscedasticity as integer-valued ARCH and GARCH models proposed by Weiß[25] and Zhu[26]. To monitor time series of attributes, alternative models are the binomial ARMA model as proposed in Startz[27] and the beta ARMA model in Cribari-Neto[28].

Finally, after fitting a model to time series data, it is essential to evaluate whether all the model assumptions are valid. If any assumption is violated, ideally another model must be proposed and fitted to encompass all the features of the data. Based on the best model the control parameters must be exhaustively searched based on simulation data to achieve the desired ARL_0 with the lowest ARL_1.

Acknowledgements

The authors would like to acknowledge CNPq,(Process 304670/2014-6) Capes and Fapesp (Grant number 13/00506-1) for the financial support. We would like to thank both reviewers for their insightful comments on the paper.

References

1 Unkel S, Farrington C, Garthwaite P, Robertson C, NickBox A. Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society A. 2012;175(1):40–82.

2 Fricker R. Introduction to Statistical Methods for Biosurveillance - with emphasis on syndromic surveillance. Cambridge University Press; 2013.

3 Richards SC, Woodall WH, Purdyc G. Surveillance of Nonhomogeneous Poisson Processes. Technometrics. 2015;57(3):388–394.

4 Rakitzis AC, Castagliola P, Maravelakis PE. Cumulative sum control charts for monitoring geometrically inflated Poisson processes: An application to infectious disease counts data. Statistical methods in medical research. 2016;p. 0962280216641985.

5 McCullagh P, Nelder JA. Generalized Linear Models. 2nd ed. London: Chapman & Hall/CRC; 1989.

6 Hohle M, Mazick A. Aberration detection in R illustrated by Danish mortality monitoring. Biosurveillance: A Health Protection Priority. 2009;.

7 Rogerson PA, Yamada I. Approaches to Syndromic Surveillance When Data Consist of Small Regional Counts. Morbidity and Mortality Weekly Report. 2004;53/Supplement:79–85.

8 Rossi G, Lampugnani l, Marchi M. An aproximate CUSUM procedure for surveillance of health events. Statistics in Medicine. 1999;18:2111–2122.

9 Chao-Wen L, Reynolds Jr MR. CUSUM charts for monitoring an autocorrelated process. Journal of Quality Technology. 2001;33(3):316.

10 Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. Journal of the American Statistical Association. 2003;98(461):214–223.

11 Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google flu trends. PloS one. 2013;8(2):e56176.

12 Heinen A. Modelling time series count data: an autoregressive conditional Poisson model. Available at SSRN 1117187. 2003;.

13 Weiß CH. Modelling time series of counts with overdispersion. Statistical Methods and Applications. 2009;18(4):507–519.

14 Weiß CH. The INARCH (1) model for overdispersed time series of counts. Communications in Statistics-Simulation and Computation. 2010;39(6):1269–1291.

15 Weiß CH, Testik MC. The Poisson INAR (1) CUSUM chart under overdispersion and estimation error. IIE Transactions. 2011;43(11):805–818.

16 Sparks RS, Keighley T, Muscatello D. Early warning CUSUM plans for surveillance of negative binomial daily disease counts. Journal of Applied Statistics. 2010;37(11):1911–1929.

17 Sparks R, Keighley T, Muscatello D. Optimal exponentially weighted moving average (EWMA) plans for detecting seasonal epidemics when faced with non-homogeneous negative binomial counts. Journal of Applied Statistics. 2011;38(10):2165–2181.

18 Hohle M, Paul M. Count data regression chart for the monitoring of surveillance time series. Computational Statistics and Data Analyisis. 2008;52:4357–4368.

19 Hardin JW, Hilbe JM, Hilbe J. Generalized Linear Models and Extensions. 2nd ed. Stata Press; 2007.

20 Dunn PK, Smyth GK. Randomized Quantile Residuals. Journal of Computational and Graphical Statistics,. 1996;5:236–244.

21 Alencar AP, Ho LL, Albarracin OYE. CUSUM control charts to monitor series of Negative Binomial count data. Statistical Methods in Medical Research. 2015;p. DOI: 10.1177/0962280215592427.

22 Jorgensen B. The Theory of Dispersion Models. 2nd ed. London: Chapman and Hall; 1996.

23 Guan Y. Variance stabilizing transformations of Poisson, binomial and negative binomial distributions. Statistical and Probability Letters. 2009;79:1621–1629.

24 Briët OJ, Amerasinghe PH, Vounatsou P. Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. PloS one. 2013;8(6):e65761.

25 Weiß C. A Poisson INAR(1) model with serially dependent innovations. Metrika. 2015;78:829–851.

26 Zhu F. A negative binomial integer-valued GARCH model. Journal of Time Series Analysis. 2011;32(1):54–67.

27 Startz R. Binomial autoregressive moving average models with an application to US recessions. Journal of business & economic statistics. 2012;.

28 Rocha AV, Cribari-Neto F. Beta autoregressive moving average models. Test. 2009;18(3):529–545.

Appendix

In the GARMA model, the distribution of each observation y_t, for $t = 1, ..., n$, conditioned on the past information $\mathbf{H}_t = (\mathbf{x_t}, ..., \mathbf{x_1}, \mathbf{y_{t-1}}, ..., \mathbf{y_1}, \mu_{\mathbf{t-1}}, ..., \mu_{\mathbf{1}})$, can be expressed as distribution belonging to the exponential family, that is, the conditional density function is:

$$f(y_t|\mathbf{H}_t) = \exp\left\{\frac{y_t\upsilon_t - b(\upsilon_t)}{\varphi} + d(y_t,\varphi)\right\}, \tag{7}$$

where υ_t is the canonical parameter; φ the scale parameter, $b(.)$ and $d(.)$ are specific functions that identify a particular distribution from the exponential family, $\mathbf{x}$ is a vector of r explanatory variables and $\mu_t = b'(\upsilon_t) = E(y_t|\mathbf{H}_t)$ represent the conditional mean of y_t given $\mathbf{H}_t$. For each time point, the likelihood ratio is given as follows

$$\begin{aligned}\frac{f(y_t|\mathbf{H}_{1,t})}{f(y_t|\mathbf{H}_{0,t})} &= \exp\left\{\frac{1}{\varphi}\left[y_t\upsilon_{t,1} - b(\upsilon_{t,1}) - y_t\upsilon_{t,0} + b(\upsilon_{t,0})\right]\right\}\\ &= \exp\left\{\frac{1}{\varphi}\left[y_t(\upsilon_{t,1}-\upsilon_{t,0}) - (b(\upsilon_{t,1}) - b(\upsilon_{t,0}))\right]\right\}\end{aligned}$$

where, $f(y_t|\mathbf{H}_{1,t})$ and $f(y_t|\mathbf{H}_{0,t})$ are the conditional distribution for process with means $\mu_{1,t}$ and $\mu_{0,t}$ respectively. Thus, the statistic LR used in this paper is given by

$$Z_{6,t} = \log\left(\frac{f(y_t|\mathbf{H}_{t,1})}{f(y_t|\mathbf{H}_{t,0})}\right) = \frac{1}{\varphi}\left[y_t(\upsilon_{t,1}-\upsilon_{t,0}) - (b(\upsilon_{t,1}) - b(\upsilon_{t,0}))\right]$$

On the other hand, Hawkins(1998) proposed for any member of the exponential family, a statistics to test whether process that has gone from an in-control parameter value υ_0 to an out-of-control value υ_1 and defined the parameter k as follows

$$k = -\frac{b(\upsilon_1) - b(\upsilon_0)}{\upsilon_1 - \upsilon_0}. \tag{8}$$

Rogerson and Yamada (2004) proposed a Poisson CUSUM chart with Time-Varying Expectations including the previous parameter k varying over time. Alencar et al. (2015) proposed a correction of k_t for the Negative Binomial distribution. The statistic RY used in this paper is given by

$$\begin{aligned}Z_{5,t} &= y_t - k_t\\ &= y_t - \frac{b(\upsilon_{t,1}) - b(\upsilon_{t,0})}{\upsilon_{t,1} - \upsilon_{t,0}}\\ &= \frac{y_t(\upsilon_{t,1}-\upsilon_{t,0}) - (b(\upsilon_{t,1}) - b(\upsilon_{t,0}))}{\upsilon_{t,1} - \upsilon_{t,0}}\\ &= \frac{\varphi LR}{\upsilon_{t,1} - \upsilon_{t,0}}.\end{aligned}$$

In this paper, it is assumed that $Y_t|\mathbf{H}_t$ follows a Negative Binomial distribution, for a fixed value of γ, the probability distribution of $Y_t|\mathbf{H}_t$ written as an element

of the exponential family is given by

$$f(Y_t|\mathbf{H}_t,\gamma) = exp\left\{Y_t \ln\left(\frac{\mu_t}{\mu_t+\gamma}\right) + \gamma \ln\left(\frac{\gamma}{\mu_t+\gamma}\right) + \ln\left(\frac{\Gamma(\gamma+Y_t)}{\Gamma(Y_t+1)\Gamma(\gamma)}\right)\right\}$$

with $\upsilon_t = \ln\left(\frac{\mu_t}{\mu_t+\gamma}\right)$, $b(\upsilon_t) = -\gamma \ln\left(1-e^{\upsilon_t}\right)$ and $\varphi = 1$.

The likelihood-ratio assuming Negative Binomial distribution is given as

$$\begin{aligned} Z_{5,t} &= y_t - k_t \\ &= \frac{Z_{6,t}}{\ln\left(\frac{2(\mu_{0,t}+\gamma)}{2\mu_{0,t}+\gamma}\right)}. \end{aligned}$$

Table 7: The values k and h which provide an $ARL_0 = 70$ and achieves the smaller value of ARL_1 for $\delta = 1.25$ for a GARMA(1,0) with mean $\mu_{0,t}^a$.

	RS		GN		JG		DR		LR		RY	
ϕ	k	h	k	h	k	h	k	h	k	h	k	h
0.05	1.0	2.730	0.7	0.927	1.0	1.444	1.0	1.300	-19.04	13.350	-50.40	20.23
0.10	1.0	2.735	0.6	1.100	0.9	1.615	0.9	1.465	-19.04	13.300	-50.40	20.09
0.20	1.0	2.737	0.6	1.096	0.9	1.605	0.9	1.464	-17.05	8.674	-47.22	16.01
0.36	0.9	2.976	0.5	1.312	0.8	1.799	0.8	1.645	-15.04	5.895	-44.80	13.74

EWMA chart for count time series using GARMA model

1 Orlando Yesid Esparza Albarracn, IME-USP ∗ E-mail: orlando@ime.usp.br
2 Airlane Pereira Alencar, IME-USP
3 Linda Lee Ho, EP-USP, Brasil

Abstract

Exponentially weighted moving average (EWMA) charts and Cumulative sum (CUSUM) control charts have been used for health surveillance due to its efficiency to detect soon small shifts in the monitored count series.
In practical health surveillance, count series can be modelled by Poisson or Negative Binomial regression, this last to control overdispersion. However, in these proposals the serial correlation is usually omitted. An alternative procedure is to build a control chart based on the residuals after fitting autoregressive moving average (ARMA) models but these models usually assume Gaussian distribution for the residuals. The main contribution of the current paper is to measure the impact, in terms of average run length (ARL) on the performance of EWMA charts when the serial correlation is neglected in the regression model.

Keywords: Health Surveillance; Average Run Length; Autocorrelation; Time series; Control charts.

1 Introduction

Recently the cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts have been widely used to detect an increase in unusual events rate in public health. They are viable alternatives to Shewhart control charts due their efficiency to detect quickly the occurrence of natural outbreaks and to issue an emergency alarm as soon as possible (See Shmueli and Burkom[1]). A feature of these control charts, compared to Shewhart charts, is that the information from the past samples are accumulated up to the current sample and then the decision about the process is taken. Authors as Lucas[2] and Hawkins[3] affirm that these charts are more efficient than the Shewhart charts in detecting small changes in the parameters of interest.

In the literature, several control charts based on fitting a Generalized linear model (GLM), with Poisson or Negative binomial distributions, have been proposed to monitor count data as morbidity or mortality rates. The data are registered over time and may depend on numerous covariates as changes in the population and seasonal pattern (since more hospitalizations due to several diseases are expected during the winter). Some CUSUM charts based on GLMs are found for example in Hohle and Mazick[4], Rogerson and Yamada[5] and Rossi et al[6].

Another proposal to monitor count time series are charts based on normalizing transformations[7;8]. A key assumption usually made on count process monitoring is the independence of observations. However, the autocorrelation may be present in the observations and may have a strong impact on the properties of control charts as discussed in Chao-Wen and Reynolds Jr[9] and Albarracin et al[10].

In this paper, EWMA control charts are proposed for monitoring the expected value of autocorrelated observations using a generalized autoregressive and moving average (GARMA) model (proposed by Benjamin et al.[11]) instead of the independent generalized linear model (GLM). This model extends the univariate Gaussian ARMA time series model to a flexible observation-driven model for non-Gaussian time series data and the conditional expected value related to a linear predictor as in the GLM. However, the GARMA model includes autoregressive and moving average terms to model a possible serial correlation. In the surveillance area, a GARMA model with a Negative Binomial distribution was fitted in Dugas et al.[12] to evaluate the effects of Google Flu Trends, meteorological and temporal information on the number of influenza-related patient visits and in Benjamin et al.[11] to monitor the well-known time series of poliomyelitis counts. Other proposals to monitor count data are based on INAR models. Weiß & Testik[13] proposed the Poisson INAR(1) CUSUM chart and investigated the effects of assuming a Poisson model when there is overdispersion. In the surveillance context, Sparks et al.[14] proposed an optimal exponentially weighted moving average (EWMA) for detecting seasonal epidemic in non-homogeneous Negative Binomial counts.

The main goal of the current paper is to evaluate, by simulation, the performance of three like-EWMA control charts to monitor autocorrelated count time series when a GLM model with trend and seasonal effects is fitted instead of the

true GARMA model, that generated the data, with the same covariates. The performance is measured in terms of the Average Run Lengths (ARLs). The EWMA charts considered in this article are based on transformed variables to achieve the normal distribution as the deviance residual (DR) and the conditional quantile residual (QR), as well as two statistics proposed here based on the likelihood ratio (LR) for the Negative Binomial distribution .

This paper is organized as follows: the GARMA model 2; the statistics used to build one-sided EWMA control charts are presented in Section 2.2; the three type of EWMA control charts in Section 2.3; the evaluation of EWMA control charts for autocorrelated data based on simulations are commented in Section 3 and conclusions and discussions are outlined in Section 4.

2 EWMA charts for GARMA models

2.1 GARMA models

In the GARMA model, the distribution of each observation y_t, for $t = 1, ..., n$, conditioned on the past information $\mathbf{H}_t = \left\{\underline{\mathbf{x}}_t, ..., \underline{\mathbf{x}}_1, y_{t-1}, ..., y_1, \mu_{t-1}, ..., \mu_1\right\}$, belongs to the exponential family. The terms $\mu_t = E(y_t|\mathbf{H}_t)$ represents the conditional mean of y_t given $\mathbf{H}_t$ and $Var(y_t|H_t) = \varphi\nu(\mu_t)$ the conditional variance, where $\nu(\mu_t)$ is a specific functions for each particular distribution from the exponential family, called the variance function. In this paper, it is assumed that the observations $Y_t|\mathbf{H}_t$ follows a Negative Binomial (NB) distribution and the sequence $Y_1|\mathbf{H}_1, Y_2|\mathbf{H}_2, \ldots, Y_n|\mathbf{H}_n$ consists of conditionally independent random variables. The conditional probability of Y_t for a NB distribution is

$$f(Y_t|\mathbf{H}_t, \gamma) = exp\left\{Y_t \ln\left(\frac{\mu_t}{\mu_t + \gamma}\right) + \gamma \ln\left(\frac{\gamma}{\mu_t + \gamma}\right) + \ln\left(\frac{\Gamma(\gamma + Y_t)}{\Gamma(Y_t + 1)\Gamma(\gamma)}\right)\right\}, \tag{1}$$

with conditional variance given by $Var(y_t|H_t) = \mu_t + \mu_t^2/\gamma$, where γ is called dispersion or scale parameter. The parametrization in (1) is known as the NB-2 in the literature[15] and obtained as a Poisson model with a gamma random effect or as the distribution of the number of failures until the r-th success,

implying a non-constant coefficient of variation $(1 + \mu/\gamma)$. It is worth noting that for a Poisson distribution $E(Y_t|\mathbf{H}_t) = Var(Y_t|\mathbf{H}_t) = \mu_t$, implying that the NB distribution induces a larger variance than the Poisson, what is called overdispersion.

As in the Generalized Linear Models (GLM), μ_t is related to a linear predictor. However, in the GARMA models, the autoregressive and the moving average terms are included to model the serial correlations present in the observations y_t that are observed over time. In practice, the linear predictor for the GARMA model is expressed as

$$g(\mu_t) = \eta_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{p} \phi_j \left\{ g(y_{t-j}) - \underline{\mathbf{x}}_{t-j}'\underline{\boldsymbol{\beta}} \right\} + \sum_{j=1}^{q} \theta_j \left\{ g(y_{t-j}) - \eta_{t-j} \right\}, \quad (2)$$

where $\underline{\mathbf{x}}_t$ is a vector of r explanatory variables, $\underline{\boldsymbol{\beta}}' = (\beta_1, \beta_2, ..., \beta_r)$ is the vector of the coefficients related to the explanatory variables, $\boldsymbol{\phi}'=(\phi_1, ..., \phi_p)$ are the autoregressive parameters and $\boldsymbol{\theta}'=(\theta_1, ..., \theta_q)$ are the moving average parameters. In this paper, the logarithm was choose as the link function $g(\cdot)$. These parameters and the scale parameter (as γ for the NB2 distribution) are estimated by the maximum likelihood method using the iterative weighted least squares procedure, as in the GLM. The diagnostic analysis are based on the normalized conditional quantile residuals proposed by Dunn & Smyth [16] and as recommended in Benjamin et al. [11]. The null correlation and Gaussian distribution of these residuals are evaluated respectively by a correlogram and a quantile-quantile plot.

2.2 EWMA charts

In the present study, we measure, in terms of the average run length (ARL), the impact on the performance of three like-EWMA charts when the serial correlation is neglected fitting a generalized linear model. The expected value of Y_t when the process is in-control and out of control is denoted as $\mu_{0,t}$ and $\mu_{1,t} = \mu_{0,t} \cdot \exp(k)$, $k > 0$, respectively, and as μ_t when it is not specified the state of the process. In this paper, we considered two transformations to the original count data to achieve a standardized normal distribution and proposed two statistics based on the likelihood ratio (LR) to monitor count time series assuming that the observations follow a Negative Binomial distribution. The deviance residual,

DR,[17] was proposed as an alternative statistic to build CUSUM chart for count data in Alencar et al[8] since that for non-normal distributed data, it follows a standard normal distribution and presents good properties as the variance stability . In the case of Y_t following a Negative Binomial distribution, the DR is defined as

$$DR_t = z_{1,t} = sign(Y_t - \mu_t)\sqrt{d_t^2}, \tag{3}$$

where the d_t^2 is defined as

$$d_t^2 = \begin{cases} 2\gamma \ln(1 + \mu_t/\gamma), & if\ Y_t = 0 \\ 2Y_t \ln\left(\frac{Y_t}{\mu_t}\right) - 2\gamma(1 + Y_t/\gamma) \ln\left(\frac{1+Y_t/\gamma}{1+\mu_t/\gamma}\right) & if\ Y_t > 0. \end{cases}$$

On the other hand, Benjamin et al.[11] suggest to use the randomized Quantile Residuals (QR), proposed in Dunn and Smyth[16], for discrete GARMA models since the distribution of the deviance and Pearson residuals are highly non-normally distributed for count data with low fitted means. The QR is defined as

$$QR_t = z_{2,t} = \Phi^{-1}(u_t) \tag{4}$$

where Φ^{-1} is the inverse of the cumulative standard normal distribution function and u_t is a random value simulated from the uniform distribution in the interval $[F(y_t - 1, \mu_t), F(y_t, \mu_t)]$ and $F(\cdot)$ is the cumulative distribution function of a Negative Binomial distribution.

Finally, two monitored statistics based on the ratio of log-likelihood functions (LR) for Negative Binomial are presented. In Lorden[18], for the first time, a sequence of hypotheses test based on the LR is proposed to detect a change point. Recently, Xu et al.[19] proposed control chart to monitor the process mean subject to linear drifts using a generalized likelihood ratio and Hohle and Paul[7], CUSUM control charts for Poisson and Negative Binomial distributions for monitoring time series of counts using GLM. In order to detect a change point using a sequence of n observations, the LR for counts following a Negative Binomial, fixed the value of γ, is expressed as

$$LR_t = ln\left\{\frac{f(y_t|\mathbf{H}_{1,t})}{f(y_t|\mathbf{H}_{0,t})}\right\} = y_t \cdot k + (y_t+\gamma)\ ln\left(\frac{\mu_{0,t}+\gamma}{\mu_{0,t}\cdot \exp(k)+\gamma}\right) \qquad (5)$$

where, $f(y_t|\mathbf{H}_{1,t})$ and $f(y_t|\mathbf{H}_{0,t})$ are the conditional distribution for process with means $\mu_{1,t}$ and $\mu_{0,t}$ respectively. Note that values of LR_t yielded by (5) may be negative (the fraction in the logarithm operator is < 1) mainly calculated with high values of y_t (as in real counts time series analyzed in Albarracin et al[10]). Additionally, the values of LR_t are not independent (as they are function of y_t). These two features turn difficulty the interpretation and implementation of EWMA charts once after finding the upper control limit, the EWMA chart does not sign a shift in the mean when an out-of-control process is considered. To deal with these problems, an alternative statistic is proposed, namely, LRc_t, defined after centered LR_t in $\mu_{0,t}$ by

$$LRc_t = z_{3,t} = (y_t-\mu_{0,t})\cdot k + (y_t-\mu_{0,t})\ ln\left(\frac{\mu_{0,t}+\gamma}{\mu_{0,t}\cdot \exp(k)+\gamma}\right) \qquad (6)$$

. The second statistic K_t proposed here consists of the determination of the maximum likelihood estimator (MLE) of the shift size k ($\mu_{1,t} = \mu_{0,t}\exp(k)$) based on the observations y_1,...,y_n. The maximization of the log-likelihood $l_n = \sum_{t=1}^{n} lnf(y_t|\mathbf{H}_t)$ as function of k is optimized by following the lines of Lawless[20] to determine the first and second derivatives

$$\frac{\partial l_n}{\partial k} = \sum_{t=1}^{n} \frac{y_t - \mu_{0,t}\exp(k)}{1+\mu_{0,t}\exp(k)/\gamma}; \quad \frac{\partial^2 l_n}{\partial k^2} = \sum_{t=1}^{n} \frac{\gamma\mu_{0,t}\exp(k)(y_t+\gamma)}{(\gamma+\mu_{0,t}\exp(k))^2}$$

In general, numerical methods are used to solve these expressions, as Newton-Raphson iterations. The convergence usually occurs in just few update-steps. Then, the last statistic proposed here, namely, k_t, is given by solution of $\frac{\partial l_n}{\partial k} = 0$ for each t.

$$k_t = z_{4,t} = \exp\left(\widehat{k}_t\right) \qquad (7)$$

When the process is in-control $E(z_{4,t}|\mu_{0,t}) = 1$. The independence of K_t and LRc_t statistics will be confirmed by simulation.

2.3 One-sided EWMA charts

In order to detect increases in the expected values, three like-EWMA control charts are presented and built only with upper control limits since for the detection of shifts on a specific direction, one-sided charts are more effective than two-side ones (see Shu et al[21]). In addition, in public health area large observations are of concern since they may be associated to epidemics. The first chart considered in this paper is the traditional EWMA chart proposed by Roberts[22], where the weight for each older data point decreases exponentially, giving much more importance to the recent observations. Its theoretical properties and the performance was investigated by Lucas and Saccuine[2]. It is expressed as

$$C_{i,t,1} = \lambda Z_{i,t} + (1-\lambda)C_{i,t-1,1},\ t \geq 1, \tag{8}$$

$i = 1,\ 2,\ 3,\ 4$, where λ is a smoothing constant such that $0 < \lambda \leq 1$. For $\lambda = 1$, (8) correspond to the Shewhart control chart.

The second EWMA control chart consists to reset the chart to the target whenever it is less than the target in order to detect faster the upward changes. Some applications are found in Champ et al[23], Gan[24] and Shu et al[21]. Thus, resetting the one-sided EWMA chart in (8) results

$$C_{t,i,2} = \max\left\{\mu_{0,t},\ \lambda Z_{t,i} + (1-\lambda)C_{t-1,i,2}\right\}. \tag{9}$$

Finally, Shu and Jiang[25] suggested resetting the current observation or normalized observation to the target but not the one-sided EWMA statistic. This process consists in first Winsorizes the statistic Z_t and then applies the conventional EWMA chart. That is,

$$C_{t,i,3} = \lambda Z^{+}_{t,i} + (1-\lambda)C_{t-1,i,3} \tag{10}$$

where, $Z^{+}_{t,i} = \max\{\mu_{0,t}, Z_{t,i}\}$. Note that if $Z_t \sim N(0,1)$, the mean and variance of the Winsorized normal variable, Z^{+}_{t}, are given by:

$$E(Z^{+}_{t}) = \frac{1}{\sqrt{2\pi}};\ Var(Z^{+}_{t}) = \frac{1}{2} - \frac{1}{2\pi}.$$

To shift the mean of C_t to zero, it is convenient to rewrite the EWMA recursion

in (10) as

$$C^*_{t,i,3} = \lambda \left(Z^+_{t,i} - E(Z^+_{t,i}) \right) + (1-\lambda) C_{t-1,i,3} \tag{11}$$

All three one-sided EWMA control charts will give an alarm at

$$h_{ij} = min \left\{ t : C_{t,i,j} > L_{ij} \sigma_{z_i} \right\},$$

where $L_{ij} > 0$ is a constant and σ_{z_i} is the asymptotic standard deviation of the statistic $Z_{i,t}$, $i = 1, \ldots, 4$ and $j = 1, 2, 3$. The starting value is often set to the target, that is, $C_{0,i,j} = \mu_{0,t}$. In this paper, the value of the control parameter L_{ij} is determined to get in-control average run length, ARL, equal to 70 weeks, which is equivalent to the usual value of 500 days proposed in the previous contributions about public health surveillance [7;26].

3 Simulation Study

In this section, a simulation study is conducted to measure the effect on the performance of the three like-EWMA control chart (namely, EWMA_1, EWMA_2 and EWMA_3), previously presented, in terms of ARL_0 and ARL_1, when the autocorrelated data are generated by a NB-GARMA model but the serial correlation is neglected, fitting only the usual NB-GLM as usually is assumed in the surveillance literature. Time series are simulated from a NB-GLM model, to get 10 thousand simulated run lengths, with mean $\mu_{0,t}$

$$ln(\mu_{0,t}) = x_t^{'} \beta = \beta_0 + \beta_1 cos\,(wt) + \beta_2 sin\,(wt) + \beta_3 t, \tag{12}$$

where $w = \frac{2\pi}{52.25}$ for weekly data. Three scenarios, that describe the profiles of time-series of infectious diseases, are considered in this study with mean $\mu^a_{0,t}$, $\mu^b_{0,t}$ and $\mu^c_{0,t}$ and parameters $\underset{\sim}{\beta}^{'}_a = (5.18, -0.048, -0.176, 0.00063)$, $\underset{\sim}{\beta}^{'}_b = (3.019, 0.250, 0.261, 0.05)$ and $\underset{\sim}{\beta}^{'}_c = (4.16, 0.5, 0.4, 0.01)$ respectively. The first and second scenarios are based on the estimated model for the real data present in Alencar et al [27] and Grabowska et al [28]. In the third scenario the coefficients of *sine* and *cosine* are the same used in a simulated study in Hohle and Paul [7]. In the first two papers, the time series of weekly number of hospital admissions due to respiratory diseases in São Paulo-Brazil were monitored using CUSUM chart

and the second, the invasive pneumococcal diseases in Germany were analyzed.

Then, the threshold L and the parameter λ based on simulations are computed to reach an ARL_0 around 70 weeks and to provide the lowest value of ARL_1 for an increase of 25% ($\exp(k) = 1.25$). In order to measure the impact of the autocorrelation in the performance of three like-EWMA control charts developed for independents observations, time series from a NB-GARMA(1,0) with parameters $\phi_1 = \{0.05; 0.1; 0.2; 0.36\}$ were simulated and monitored by the EWMA charts developed for the GLM. Values of ARL_0, its standard deviation (SD) and ARL_1 for GARMA(1,0) when $\mu_{0,t} = \mu^a_{0,t}$ are presented in Table 1. Due to the asymmetry of the run length distribution, the median run length (MRL) is also presented in the same table.

Table 1: Scenario 1 - Values of ARL_0, SD, MRL_0 and ARL_1 for an increase of 25% in the mean of GARMA(1.0) processes

.

		ARL_0 (SD)			MRL_0			ARL_1		
ϕ	Statistics	$EWMA_1$	$EWMA_2$	$EWMA_3$	$EWMA_1$	$EWMA_2$	$EWMA_3$	$EWMA_1$	$EWMA_2$	$EWMA_3$
0.00	DR	69.94 (0.68)	69.84 (0.66)	70.02 (0.77)	49	49	43	2.08	2.35	2.70
	LRc	70.06 (0.64)	70.00 (0.65)	70.03 (0.72)	50	49	44	2.13	2.40	1.90
	QR	69.91 (0.67)	69.96 (0.65)	70.04 (0.69)	49	50	48	2.22	2.41	3.18
	kt	70.17 (0.71)	70.45 (0.70)	70.53 (0.74)	48	49	49	1.99	2.20	**1.80**
0.05	DR	62.04 (0.60)	61.56 (0.58)	66.67 (0.72)	44	44	42	2.11	2.36	1.80
	LRc	61.72 (0.58)	62.26 (0.56)	67.90 (0.70)	43	44	43	2.15	2.41	1.91
	QR	62.85 (0.61)	62.71 (0.60)	64.80 (0.64)	44	44	45	2.23	2.44	1.95
	kt	61.57 (0.61)	61.66 (0.61)	65.06 (0.67)	44	45	46	1.99	2.21	1.80
0.10	DR	57.75 (0.55)	56.16 (0.52)	64.73 (0.70)	42	41	41	2.13	2.38	1.81
	LRc	57.62 (0.54)	57.04 (0.51)	65.37 (0.66)	41	41	43	2.16	2.43	1.92
	QR	58.21 (0.57)	56.04 (0.52)	60.59 (0.59)	41	41	44	2.27	2.48	2.00
	kt	56.96 (0.56)	56.53 (0.55)	61.01 (0.62)	42	42	44	2.01	2.22	1.82
0.20	DR	47.84 (0.47)	45.69 (0.43)	56.80 (0.64)	33	33	34	2.18	2.44	1.86
	LRc	48.12 (0.46)	46.94 (0.43)	57.38 (0.61)	33	33	35	2.22	2.50	1.98
	QR	48.65 (0.48)	46.19 (0.44)	51.34 (0.51)	33	33	35	2.29	2.50	2.02
	kt	47.78 (0.48)	45.39 (0.44)	50.45 (0.52)	33	33	35	2.06	2.29	1.88
0.36	DR	37.10 (0.37)	34.17 (0.32)	44.72 (0.50)	26	25	26	2.27	2.56	1.96
	LRc	37.36 (0.36)	35.18 (0.32)	45.61 (0.49)	26	26	28	2.32	2.62	2.08
	QR	37.14 (0.36)	34.10 (0.32)	38.01 (0.38)	26	24	26	2.40	2.63	2.14
	kt	36.69 (0.37)	33.92 (0.33)	38.40 (0.39)	25	24	26	2.15	2.39	1.99

Note that for uncorrelated data, i.e, $\phi_1 = 0$, the K_t statistic, that is proposed in this paper, presents the lowest ARL_1 values for the three EWMA control charts, indicating that in this case study, it seems to present the best performance to detect a shift of 25% in the mean. In the presence of serial correlation, i.e. $\phi_1 > 0$ all monitored statistics produced similar increases of false alarms, i.e, $ARL_0 < 70$, for each type of EWMA. However, the performance is different for the three types of EWMA charts. For the $EWMA_1$ and $EWMA_2$ control charts, all statistics produce lower ARL_0 values than in the $EWMA_3$, indicating that, in this scenario, $EWMA_3$ is less affected by the negligence of the serial correlation. For example, when $\phi_1 = 0.2$ the ARL_0 decayed around to

46 for EWMA_1 and EWMA_2 for all statistics, while for the EWMA_3 the false alarms increase more slowly, since the ARL_0 decayed to 57 or 50 depending on the statistics.

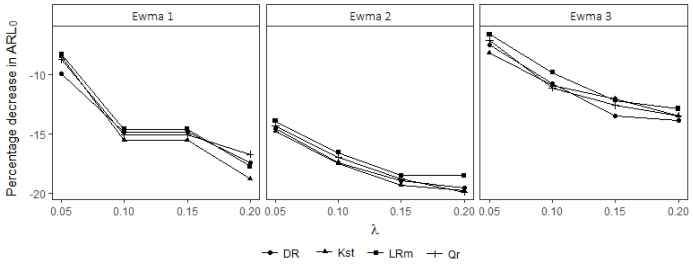


Figure 1: Percentage decrease in ARL_0 relative to the parameter λ - Scenario 1

Figure 1 shows the percentage decrease in ARL_0 relative to the parameter λ in all EWMA control charts when $\phi_1 = 0.1$. Note that, in this case study, all EWMA charts signal false alarms earlier than expected for high λ values, independent of the statistics considered. However, the decrease of the ARL_0 is lower for the EWMA_3. For $\lambda = 0.2$, the percentage decrease in ARL_0 is around 13.5% for the last EWMA while for the EWMA_1 and EWMA_2 is around to 18.2% and 19.5%, respectively.

Table 2: Scenario 2 - Values of ARL_0, its SD, MRL_0 and ARL_1 for an increase of 25% in the mean of GARMA(1.0) processes when fitting a GLM with mean $\mu_{0,t}^b$

		ARL_0 (SD)			MRL_0			ARL_1		
ϕ	Statistics	EWMA_1	EWMA_2	EWMA_3	EWMA_1	EWMA_2	EWMA_3	EWMA_1	EWMA_2	EWMA_3
0	DR	70.11 (0.82)	70.03 (0.66)	69.98 (0.80)	39	50	41	**4.73**	7.26	**4.86**
	LRc	70.02 (0.73)	69.96 (0.68)	70.01 (0.75)	47	51	47	5.61	6.82	5.49
	QR	69.94 (0.65)	69.95 (0.65)	70.00 (0.67)	48	48	48	7.32	7.84	7.11
	Kt	69.89 (0.71)	69.97 (0.64)	69.86 (0.69)	44	47	44	6.18	8.21	6.30
0.05	DR	69.01 (0.82)	63.71 (0.60)	68.66 (0.80)	38	46	39	4.88	7.32	5.04
	LRc	69.04 (0.74)	63.69 (0.62)	69.54 (0.76)	45	47	46	5.75	6.88	5.65
	QR	65.69 (0.62)	63.85 (0.60)	65.75 (0.63)	45	44	45	7.50	7.97	7.28
	Kt	68.72 (0.72)	64.67 (0.59)	68.70 (0.70)	42	44	43	6.34	8.24	6.50
0.10	DR	66.74 (0.81)	57.63 (0.55)	68.83 (0.81)	35	41	39	5.05	7.54	5.31
	LRc	67.20 (0.73)	58.76 (0.58)	68.76 (0.76)	43	43	45	5.95	7.15	5.88
	QR	61.79 (0.58)	59.19 (0.54)	62.64 (0.60)	42	41	43	7.60	8.09	7.44
	Kt	67.24 (0.71)	59.23 (0.54)	69.14 (0.71)	41	40	43	6.54	8.43	6.76
0.20	DR	64.28 (0.79)	47.73 (0.44)	63.16 (0.75)	33	35	35	5.46	7.84	5.75
	LRc	64.63 (0.72)	48.69 (0.47)	63.34 (0.71)	41	35	40	6.39	7.46	6.37
	QR	53.20 (0.50)	49.79 (0.45)	54.16 (0.52)	37	35	37	7.93	8.38	7.79
	Kt	64.88 (0.70)	50.34 (0.45)	64.02 (0.67)	39	36	40	6.97	8.73	7.23
0.36	DR	61.87 (0.79)	38.23 (0.37)	54.91 (0.66)	30	27	30	6.25	8.40	6.54
	LRc	61.26 (0.71)	39.11 (0.39)	55.28 (0.63)	36	27	33	7.21	8.11	7.19
	QR	44.46 (0.42)	39.39 (0.36)	43.46 (0.42)	32	29	31	8.57	8.90	8.47
	Kt	61.37 (0.69)	40.45 (0.37)	56.55 (0.62)	35	29	35	7.78	9.33	8.11

The values of L and λ of the scenario one are in the Appendix. Note that,

although $\lambda = 0.2$ leads to the lowest ARL_0 values in EWMA charts developed for the GLM when uncorrelated observations are monitored, it was the best value of λ to detect earlier a shift of 25% in the mean in the first scenario.

In Table 2, all the results for GARMA(1,0) for the second simulated scenario $\mu^b_{0,t}$ are presented. Note that, when the EMWA$_2$ is used to monitor the uncorrelated data, all statistics produce the highest ARL_1 values. In the other two EWMA control charts, the DR statistic seems to present the best results since it produces the lowest ARL_1 values. However, this statistic presents also the lowest MRL_0 values, signaling false alarms faster than the expected. Thus, the LRc statistic seems to be better once it produces high MRL_0 values and using the EMWA$_3$, it detects averagely an increase of 25% in the mean, 0.6 weeks later than the DR statistic. Note also that, the performance of EWMA$_1$ and EWMA$_3$ are similar. However, when the traditional EWMA is used, all monitored statistics, except the QR statistic, are less affected by the negligence of the high correlation serial ($\phi_1 = 0.36$).

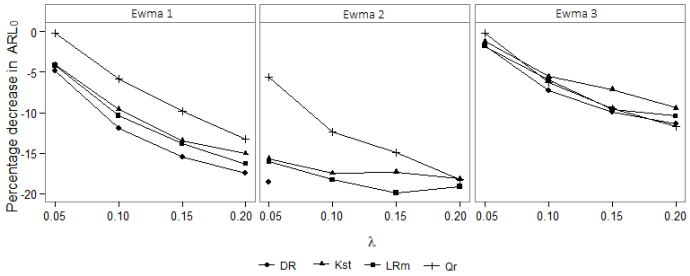


Figure 2: Percentage decrease in ARL_0 relative to the parameter λ - Scenario 2

The percentage decrease in ARL_0 relative to the parameter λ for all EWMA control charts when $\phi_1 = 0.1$ is apresented in Figure 2. Note that, when the EWMA$_3$ is used the percentage decrease in ARL_0 is lower. As in the first scenario all EWMA charts signal false alarms earlier than expected for high λ values, independent of the statistics considered. Finally, it is worth noting that, the QR statistic is the most robust in a sense to produce the smallest increase of false alarms for the EWMA$_1$ and EWMA$_2$.

In Table 3, all the results for GARMA(1,0) for the third simulated scenario $\mu^c_{0,t}$ are presented. Note that, when $\phi_1 = 0.0$, for all EWMA control charts, all monitored statistics produce similar ARL_1 values for an increase of 25% in

Table 3: Scenario 3 -Values of ARL_0. its SD. MRL_0. ARL_1 for an increase of 25% in the mean of GARMA(1.0) processes when fitting a GLM with mean $\mu^c_{0.t}$

		ARL_0 (SD)			MRL_0			ARL_1		
ϕ	Statistics	$EWMA_1$	$EWMA_2$	$EWMA_3$	$EWMA_1$	$EWMA_2$	$EWMA_3$	$EWMA_1$	$EWMA_2$	$EWMA_3$
0	DR	69.95 (0.77)	69.97 (0.67)	69.94 (0.78)	43	49	43	2.59	2.96	2.24
	LRc	69.98 (0.74)	69.98 (0.69)	69.96 (0.78)	48	52	48	2.61	2.69	2.27
	QR	70.00 (0.65)	70.05 (0.65)	70.00 (0.67)	48	48	47	2.96	3.19	2.66
	Kt	70.06 (0.71)	70.01 (0.63)	70.15 (0.65)	41	45	43	3.14	4.21	3.37
0.05	DR	68.57 (0.76)	62.29 (0.59)	67.48 (0.75)	41	44	41	2.61	2.99	2.26
	LRc	68.57 (0.73)	63.32 (0.62)	67.94 (0.75)	46	49	46	2.63	2.73	2.29
	QR	64.56 (0.60)	63.45 (0.58)	65.59 (0.62)	45	45	45	3.02	3.24	2.71
	Kt	67.86 (0.69)	64.65 (0.57)	69.30 (0.65)	40	42	42	3.16	4.24	3.41
0.10	DR	64.34 (0.72)	56.49 (0.55)	64.25 (0.72)	38	40	39	2.64	3.03	2.30
	LRc	64.61 (0.70)	57.37 (0.57)	65.66 (0.73)	44	45	45	2.66	2.75	2.33
	QR	57.77 (0.54)	56.47 (0.53)	60.46 (0.59)	40	39	41	3.05	3.27	2.76
	Kt	64.95 (0.67)	60.05 (0.53)	67.48 (0.63)	38	39	41	3.19	4.29	3.46
0.20	DR	60.19 (0.69)	46.59 (0.44)	58.74 (0.67)	34	33	34	2.69	3.10	2.37
	LRc	60.60 (0.67)	47.48 (0.47)	59.66 (0.68)	39	35	36	2.71	2.81	2.39
	QR	49.85 (0.48)	46.71 (0.43)	51.20 (0.49)	35	33	36	3.12	3.36	2.84
	Kt	60.79 (0.65)	51.84 (0.44)	62.79 (0.60)	35	36	39	3.25	4.35	3.54
0.36	DR	51.53 (0.62)	34.84 (0.33)	46.33 (0.54)	27	25	26	2.88	3.33	2.58
	LRc	51.84 (0.61)	35.11 (0.35)	47.56 (0.55)	29	22	24	2.90	3.05	2.60
	QR	39.67 (0.38)	35.53 (0.32)	39.37 (0.38)	29	26	28	3.34	3.57	3.09
	Kt	53.04 (0.59)	40.86 (0.34)	51.79 (0.50)	30	30	34	3.47	4.57	3.82

the mean. However, it's worth noting that using the $EWMA_3$ chart the LRc statistic leads to better results since produce the lowest ARL_1 value and the highest MRL_0 value. On the other hand, when $\phi_1 > 0$, i.e, in a presence of serial correlation the Kt statistic produces the smallest increases of false alarms. For $\phi_1 = 0.2$ the percentage decrease in ARL_0 is around 26.0% while for the other statistics is on average 36.0%. However, in this case, the DR and LRc statistics detect about two weeks before an increase of 25 % in the mean.

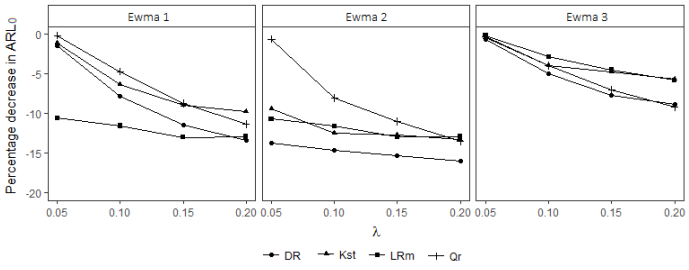


Figure 3: Percentage decrease in ARL_0 relative to the parameter λ - Scenario 3

Figure 3 shows the percentage decrease in ARL_0 relative to the parameter λ in all EWMA control charts when $\phi_1 = 0.1$. In this third case study, again high λ values lead to detect false alarms earlier than expected. Although, in the $EWMA_3$ this decrease is lower. Also note that the LMc statistic presents

the highest percentage decrease in ARL_0 using EMWA_1. The values of L and λ of the scenario three are in the Appendix. Note that, in this case the λ values that lead the lowest ARL_1 values are 0.05 and 0.2.

4 Conclusions and discussions

The main goal of this paper is to measure the impacts on the average and median run length (ARL and MRL) to detect outbreaks in autocorrelated count time series when EWMA charts based on the independent GLM model are used, as it is usual in the surveillance area. Then, three like-EWMA control charts were built only with upper control limits since that in public health area, large observations may be associated to epidemics. The first chart is the traditional EWMA proposed by Roberts[22]. The second consists in resetting the traditional EWMA to the target whenever it is less than the target in order to detect faster the upward changes. Finally, in the last chart, first the monitored statistic is winsorized and then applies the conventional EWMA chart.

All charts were based on four statistics, two transformations to achieve a standardized normal distribution: the deviance residual, DR, proposed as an alternative transformation to build CUSUM chart for count data in Alencar et al[8] and the randomized Quantile Residuals (QR), proposed in Dunn and Smyth[16] as an alternative to the deviance and Pearson residuals for count data with low fitted means. On the other hand, two alternative statistics were proposed in this paper to monitor autocorrelated count time series assuming the Negative Binomial distribution. The first one consists in centralizing the ratio of log-likelihood functions (LR) widely used in the literature assuming independent observations. However, it presents some undesirable features for observations with high values and, when uncorrelated counts are considered, the LR values are also uncorrelated. The second is based on the maximum likelihood estimator (MLE) of the shift size k ($\mu_{1,t} = \mu_{0,t}\exp(k)$) since that it is easy to interpret.

The threshold L and parameter λ are computed, by simulations, to reach an ARL_0 around 70 weeks and to provide the lowest value of ARL_1 for an increase of 25% ($\exp(k) = 1.25$).

The EWMA performances when the autocorrelation is ignored and a GLM model under independence is fitted, instead of a true GARMA model (that is, the time series simulated as GARMA but analyzed as GLM), were compared

in a simulation study considering 3 scenarios that describe the profiles of time-series of infectious diseases. The first and second are based on the estimated parameters of models fitted for real data presented, respectively, in Alencar et al[27] and in[?] . In the third scenario, the coefficients of *sine* and *cosine* components are the same used in Hohle and Paul[7].

The GARMA models[11] allow to consider the Poisson or Negative Binomial distribution and to relate the expected value to a linear predictor as in the GLM model. However, in the GARMA model includes autoregressive and moving average terms to model serial correlation. Higher values of autoregressive or moving average parameters implies higher autocorrelations for the time series generated by this model.

For uncorrelated data ($\phi_1 = 0$), in the 3 scenarios, all monitored statistics produced similar performance to detect a shift of 25% in the mean in the EWMA charts However, the EWMA_2 seemed to present the worst performance since that when it was used all statistics produced the highest ARL_1 values. It is worth noting that, in the first scenario, the k_t statistic, proposed here, seems to have the best performance since that presented the lowest ARL_1 values and high MRL_0 values for the three EWMA control charts. In the second and third scenario, the DR statistic produced the lowest ARL_1 values. However, the DR statistic presented also the lowest MRL_0 values signaling false alarms faster than the expected. Thus, the LRc statistic seemed to be better once it produced the highest MRL_0 values and detected averagely an increase of 25% in the mean, only 0.6 weeks later using the EWMA_3 chart.

On the other hand, when a pure GLM is fitted instead of a GARMA model that generated the count series, all EWMA charts were affected by the negligence of the correlation producing all statistics ARL_0 values lower than the expected. It is noteworthy that, no statistics seem to be robust, in a sense to produce the smallest increase of false alarms in the analyzed scenarios. Similar results are found in Albarracin et al[10] where the authors studied the impact of omitted the autocorrelation on the CUSUM chart.

.

Acknowledgements

The authors would like to acknowledge CNPq,(Process 304670/2014-6) Capes and Fapesp (Grant number 13/00506-1) for the financial support. We would like to thank both reviewers for their insightful comments on the paper.

References

1 Shmueli G, Burkom H. Statistical challenges facing early outbreak detection in biosurveillance. Technometrics. 2010;52(1):39–51.

2 Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. Technometrics. 1990;32(1):1–12.

3 Hawkins DM, Olwell DH. Cumulative sum charts and charting for quality improvement. Springer Science & Business Media; 2012.

4 Hohle M, Mazick A. Aberration detection in R illustrated by Danish mortality monitoring. Biosurveillance: A Health Protection Priority. 2009;.

5 Rogerson PA, Yamada I. Approaches to Syndromic Surveillance When Data Consist of Small Regional Counts. Morbidity and Mortality Weekly Report. 2004;53/Supplement:79–85.

6 Rossi G, Lampugnani l, Marchi M. An aproximate CUSUM procedure for surveillance of health events. Statistics in Medicine. 1999;18:2111–2122.

7 Hohle M, Paul M. Count data regression chart for the monitoring of surveillance time series. Computational Statistics and Data Analyisis. 2008;52:4357–4368.

8 Alencar AP, Ho LL, Albarracin OYE. CUSUM control charts to monitor series of Negative Binomial count data. Statistical Methods in Medical Research. 2015;p. DOI: 10.1177/0962280215592427.

9 Chao-Wen L, Reynolds Jr MR. CUSUM charts for monitoring an autocorrelated process. Journal of Quality Technology. 2001;33(3):316.

10 Albarracin O, Alencar A, Lee HL. CUSUM chart to monitor autocorrelated counts using Negative Binomial GARMA model. Statistical methods in medical research. 2017;p. 962280216686627–962280216686627.

11 Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. Journal of the American Statistical Association. 2003;98(461):214–223.

12 Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google flu trends. PloS one. 2013;8(2):e56176.

13 Weiß CH, Testik MC. The Poisson INAR (1) CUSUM chart under overdispersion and estimation error. IIE Transactions. 2011;43(11):805–818.

14 Sparks R, Keighley T, Muscatello D. Optimal exponentially weighted moving average (EWMA) plans for detecting seasonal epidemics when faced with non-homogeneous negative binomial counts. Journal of Applied Statistics. 2011;38(10):2165–2181.

15 Hardin JW, Hilbe JM, Hilbe J. Generalized Linear Models and Extensions. 2nd ed. Stata Press; 2007.

16 Dunn PK, Smyth GK. Randomized Quantile Residuals. Journal of Computational and Graphical Statistics,. 1996;5:236–244.

17 Pierce DA, Schafer DW. Residuals in generalized linear models. Journal of the American Statistical Association. 1986;81(396):977–986.

18 Lorden G. Likelihood ratio tests for sequential k-decision problems. The Annals of Mathematical Statistics. 1972;p. 1412–1427.

19 Xu L, Wang S, Reynolds MR. A generalized likelihood ratio control chart for monitoring the process mean subject to linear drifts. Quality and Reliability Engineering International. 2013;29(4):545–553.

20 Lawless JF. Negative binomial and mixed Poisson regression. Canadian Journal of Statistics. 1987;15(3):209–225.

21 Shu L, Jiang W, Wu S. A one-sided EWMA control chart for monitoring process means. Communications in StatisticsSimulation and Computation®. 2007;36(4):901–920.

22 Roberts S. Control chart tests based on geometric moving averages. Technometrics. 1959;1(3):239–250.

23 Champ CW, Woodall WH, Mohsen HA. A generalized quality control procedure. Statistics & probability letters. 1991;11(3):211–218.

24 Gan F. Monitoring Poisson observations using modified exponentially weighted moving average control charts. Communications in Statistics-Simulation and Computation. 1990;19(1):103–124.

25 Shu L, Jiang W. A new EWMA chart for monitoring process dispersion. Journal of Quality Technology. 2008;40(3):319.

26 Rossi G, Sarto SD, Marchi M. A new risk-adjusted Bernoulli cumulative sum chart for monitoring binary health data. Statistical methods in medical research. 2016;25(6):2704–2713.

27 Alencar AP, Lee Ho L, Albarracin OYE. CUSUM control charts to monitor series of Negative Binomial count data. Statistical methods in medical research. 2017;26(4):1925–1935.

28 Grabowska K, Högberg L, Penttinen P, Svensson Å, Ekdahl K. Occurrence of invasive pneumococcal disease and number of excess cases due to influenza. BMC infectious diseases. 2006;6(1):58.

Appendix

Table 4: λ and L parameters for the first scenario

	$EWMA_1$		$EWMA_2$		$EWMA_3$	
	λ	L	λ	L	λ	L
DR	0.2	1.762	0.2	2.023	0.05	1.076
LRc	0.2	10.432	0.2	11.904	0.05	6.618
QR	0.2	1.978	0.2	2.181	0.2	2.338
Kst	0.2	0.191	0.2	0.217	0.2	0.236

Table 5: λ and L parameters for the third scenario

	Ewma 1		Ewma 2		Ewma 3	
	λ	L	λ	L	λ	L
DR	0.05	0.856	0.20	1.952	0.05	1.952
LRm	0.05	2.363	0.20	4.448	0.05	4.448
QR	0.20	2.712	0.20	2.768	0.20	2.768
Kst	0.05	0.422	0.05	0.596	0.05	0.596

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

Orlando Yesid Esparza Albarracin
IME-USP, São Paulo, Brazil
E-mail: orlando@ime.usp.br
Airlane Pereira Alencar
IME-USP, São Paulo, Brazil
Linda Lee Ho
EP-USP, São Paulo, Brazil

Summary. 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, to find the estimates of the parameters. The second is the inclusion of the same lagged observations into the autoregressive and moving average components confounding the interpretation of the parameters.
A modified model, GARMA-M, is presented to deal with the multicollinearity and improves the interpretation of the parameters. In a general sense, simulation studies show that the modified model provide estimators closer to the parameters and offer confidence intervals with higher coverage percentage than obtained with the GARMA model, but some restrictions in the parametric space are imposed to guarantee the stationarity of the process. Also, A real data analysis illustrate the GARMA-M fit for daily hospilatization rates of elderly people due to respiratory diseases from October 2012 to April 2015 in So Paulo city, Brazil.

Keywords: hospilatization rates; count time series; exponential family distribution; ILRS algorithm

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 (Shumway and Stoffer, 2010). For count time series, the generalized autoregressive moving average (GARMA) (Benjamin et al., 2003) and the generalized linear autoregressive moving average (GLARMA) (Dunsmuir et al., 2015) models extend the univariate Gaussian ARMA time series 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 (Dugas et al., 2013; Albarracin et al., 2017) where it is necessary to model discrete response time series in terms of covariates.

In this paper, we focus on the GARMA model. A model where the conditional mean of the response variable may depend on a set of covariates and includes the serial correlations among observations measured along time. Thus, it is possible to make inference on 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 values time series. A function of the conditional mean (called link function) is a linear combination of the covariates (linear predictor), as in the generalized linear models (GLM) (McCullagh and Nelder, 1989). But the GARMA model includes lagged values of the dependent variable in the set of regressors to model the serial dependence.

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 among the explanatory variables creates difficulties, namely, numerical instabilities and problems of identification and interpretation of the effects of each explanatory variables on the mean response variable. Perfect multicollinearity leads to non-identifiable model and an unfeasible estimation as it is not possible to invert a matrix in the Least Squares (LS) method. In addition, the interpretation of the parameters in the GARMA models with p and q, simultaneously, different from zero is unclear as a consequence of the structure of the linear predictor .

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 GARMA-M is proposed to reduce the problem of multicollinearity in the GARMA model and to have a better interpretation of the parameters.

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. (2003), the distribution of each observation y_t, for $t = 1, ..., n$, conditioned on the past information $\mathbf{H}_t = \{\underline{\mathbf{x}}_t, ..., \underline{\mathbf{x}}_1, y_{t-1}, ..., y_1, \mu_{t-1}, ..., \mu_1\}$, belongs to the exponential family. The conditional density function

is expressed as

$$f(y_t|\mathbf{H}_t) = \exp\left\{\frac{y_t\upsilon_t - b(\upsilon_t)}{\varphi} + d(y_t,\varphi)\right\}, \tag{1}$$

where υ_t is the canonical parameter; φ the scale parameter, $b(.)$ and $d(.)$ are specific functions that define a particular distribution from the exponential family, as Normal, Gamma, Poisson distributions.

The terms $\mu_t = E(y_t|\mathbf{H}_t) = b'(\upsilon_t)$ represents the conditional mean of y_t given $\mathbf{H}_t$ and $var(y_t|H_t) = \varphi b''(\upsilon_t) = \varphi\nu(\mu_t)$ the conditional variance, where $\nu(\mu_t)$ is called the variance function. For example, the Poisson distribution belongs to the exponential family distribution, with: $\varphi = 1$, $\upsilon_t = \ln(\mu_t)$, $b(\upsilon_t) = exp(\upsilon_t)$, $d(y_t,\varphi) = -\ln(y_t!)$ and $\nu(\mu_t) = \mu_t$.

As in the Generalized Linear Models (GLM) (McCulloch and Neuhaus, 2001), μ_t is related to a linear predictor, η_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 is expressed as,

$$g(\mu_t) = \eta_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{p}\phi_j A(y_{t-j},\underline{x}_{t-j},\underline{\beta}) + \sum_{j=1}^{q}\theta_j M(y_{t-j},\mu_{t-j}), \tag{2}$$

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

$$g(\mu_t) = \eta_t = \underline{\mathbf{x}}_t'\underline{\beta} + \sum_{j=1}^{p}\phi_j\left\{g(y_{t-j}) - \underline{\mathbf{x}}_{t-j}'\underline{\beta}\right\} + \sum_{j=1}^{q}\theta_j\left\{g(y_{t-j}) - \eta_{t-j}\right\}. \tag{3}$$

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} 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. (2003) $c = 0.1$ is used).

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

is updated $\underline{\gamma}^{(k+1)}$ as

$$\hat{\underline{\gamma}}^{(k+1)} = \left(B^{'}W^{(k)}B\right)^{-1} B^{'}W^{(k)}z^{(k)}, \tag{4}$$

where $W^{(k)}$ is a diagonal matrix of weights w_t, that change at each k-th iteration and $z^{(k)}$ is the dependent variable given by the expression $z = \eta + W^{-1/2}V^{-1/2}(y-\mu)$, where V is a diagonal matrix of variance functions $\nu(\mu_t)$ and $w_t^{-1} = \nu(\mu_t)\left(\frac{dg(\mu_t)}{d\mu_t}\right)^2$, for $t = m+1, ..., n$, with $m \geq max\{p, q\}$.

The convergence of (4) occurs generally in a finite number of steps. Estimating $\underline{\gamma}$ using IRLS allows to obtain the approximate large-sample conditional variance of $\widehat{\gamma}$ as a by product of the iterative process (Green (1984); Kaufmann (1987)). It can be shown that asymptotically $\sqrt{(n-m)}(\widehat{\underline{\gamma}} - \underline{\gamma}) \sim N(0, I(\underline{\gamma})^{-1})$, where

$$I(\underline{\gamma}) = \lim_{n\to\infty} \frac{\widehat{\varphi}}{n-m}\left\{\sum_{t=m}^{n} w_t \left(\frac{\partial \eta_t}{\partial \underline{\gamma}}\right)\left(\frac{\partial \eta_t}{\partial \underline{\gamma}}\right)^{'}\right\}. \tag{5}$$

The GARMA model extends previous works of Li (1994); Zeger and Qaqish (1988). They had already included the autoregressive and moving average terms, respectively, in exponential family models. If $\theta_j = 0$ for $j = 1, .., q$ in (3), then this corresponds to an autorregressive model for counts presented in Zeger and Qaqish (1988). If $\phi_j = 0$ for $j = 1, .., p$ in (3), it is a pure moving average model for counts considered by Li (1994).

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 (η_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 GARMA-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

$$\eta_t = g(\mu_t) = \underline{x}_t^{'}\underline{\beta} + \phi_1\left\{g(y_{t-1}) - \underline{x}_{t-1}^{'}\underline{\beta}\right\} + \theta_1\left\{g(y_{t-1}) - \eta_{t-1}\right\}. \tag{6}$$

To initialize the process, let $\eta_1 = g(\mu_1) = \underline{x}_1^{'}\underline{\beta}$. Then,

$$\begin{aligned}\eta_2 = g(\mu_2) &= \underline{x}_2^{'}\underline{\beta} + \phi_1\left\{g(y_1) - \underline{x}_1^{'}\underline{\beta}\right\} + \theta_1\left\{g(y_1) - \eta_1\right\} \\ &= \underline{x}_2^{'}\underline{\beta} + \phi_1\left\{\boldsymbol{g(y_1) - \underline{x}_1^{'}\underline{\beta}}\right\} + \theta_1\left\{\boldsymbol{g(y_1) - \underline{x}_1^{'}\underline{\beta}}\right\}.\end{aligned}$$

Note that, for $t = 2$, the effect of the previous instant is considered at the same time by the parameters ϕ_1 and θ_1. Then, for $t = 3$ the linear predictor is

$$\begin{aligned}\eta_3 = g(\mu_3) &= \underline{x}_3'\underline{\beta} + \phi_1\left\{g(y_2) - \underline{x}_2'\underline{\beta}\right\} + \theta_1\left\{g(y_2) - \eta_2\right\} \\ &= \underline{x}_3'\underline{\beta} + \phi_1\left\{\boldsymbol{g(y_2)} - \underline{\boldsymbol{x}}_{\boldsymbol{2}}'\underline{\boldsymbol{\beta}}\right\} + \theta_1\Big\{\boldsymbol{g(y_2)} - \underline{\boldsymbol{x}}_{\boldsymbol{2}}'\underline{\boldsymbol{\beta}} - \\ &\quad (\phi_1 + \theta_1)\left\{g(y_1) - \underline{x}_1'\underline{\beta}\right\}\Big\}.\end{aligned}$$

Observe that the effect of the previous instant ($t = 2$) is associated simultaneously with the parameters ϕ_1 and θ_1. But in this case, the parameter θ_1 additionally takes into account the effect of the older instants. Making successive substitutions, η_t can be written as

$$\eta_t = \underline{x}_t'\underline{\beta} + \phi_1\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta}\right\} + \theta_1\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta} + c_t\right\}, \tag{7}$$

$$\text{with } c_t = \sum_{j=1}^{t-2}(-\theta_1)^{j-1}(\theta_1 + \phi_1)\left[g(y_{t-1-j}) - \underline{x}_{t-1-j}'\underline{\beta}\right].$$

Writing the linear predictor as in (7), it is easy to observe that the effect of the previous term, $g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta}$, is taken into account simultaneously by the autoregressive ϕ_1 and the moving average θ_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 ϕ_1 takes into account only the effect of the previous day and the parameter θ_1 considers not only the direct effect of the previous day, but also all the older observations attributing decreasing weights with alternating signs $((-\theta_1)^{j-1}(\theta_1 + \phi_1))$. The expressions $\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta}\right\}$ and $\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta} + c_t\right\}$ are called hereafter as the autoregressive and moving average terms, respectively.

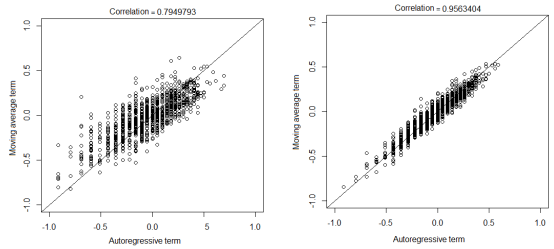
Another interesting feature in the GARMA models is the similar performance of GARMA(1,0) and GARMA(0,1) models when $\theta_1 = \phi_1$. This occurs because both models give the same weight for the common term $\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta}\right\}$. (the GARMA(0,1) includes decreasing weights θ_1^j, for $j = 2, 3, ...$ for the older terms). For example, for $\theta_1 = \phi_1 = 0.2$ the highest weight attributed to older instances in the moving average term in the GARMA(0,1) 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

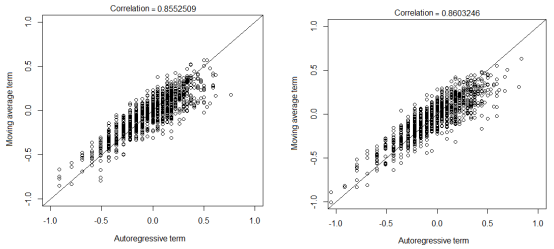
$$\begin{aligned}\eta_t = \underline{x}_t'\beta + \phi_1\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta}\right\} + \phi_2\left\{g(y_{t-2}) - \underline{x}_{t-2}'\underline{\beta}\right\} + \\ \theta_1\left\{g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta} - (\theta_1 + \phi_1)\left\{g(y_{t-2}) - \underline{x}_{t-2}'\underline{\beta}\right\} + c_{1,t}\right\}\end{aligned}$$

$$\text{with } c_{1,t} = \sum_{j=1}^{(t-3)}(-\theta_1)^{j-1}\left[\theta_1(\theta_1 + \phi_1) - \phi_2\right]\left\{g(y_{t-2-j}) - \underline{x}_{t-2-j}'\underline{\beta}\right\}$$

It is worth noting that the effects of two previous terms $g(y_{t-1}) - \underline{x}_{t-1}^{'}\underline{\beta}$ and $g(y_{t-2}) - \underline{x}_{t-2}^{'}\underline{\beta}$ are measured simultaneously by the autoregressive parameters ϕ_1, ϕ_2 and the moving average parameter θ_1. In general, in the GARMA(p,1) models with $p > 1$, the effect of the p previous terms $\left\{g(y_{t-j}) - \underline{x}_{t-j}^{'}\underline{\beta}\right\}$ for $j = 1, 2, ...p$ are taken into account simultaneously by the autorregressive parameters and the moving average parameter leading to problems of multicollinearity and hindering the interpretation of these parameters.



$\phi_1 = 0.4$ and $\theta_1 = 0.3$ $\quad$ $\phi_1 = -0.2$ and $\theta_1 = 0.5$



$\phi_1 = 0.1$ and $\theta_1 = -0.5$ $\quad$ $\phi_1 = -0.2$ and $\theta_1 = -0.4$

Fig. 1. Scatter plots for some (ϕ_1, θ_1) - GARMA(1,1).

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 θ_1 and θ_2 taken into account, simultaneously, all terms $\left\{g(y_j) - \underline{x}_j^{'}\underline{\beta}\right\}$, 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 parameter space that induce more multicollinearity in the GARMA(1,1) models are studied. In cases where there are covariates $\underline{x}_t$ 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 count under the Poisson distribution with the natural logarithm as a link function and assuming that $x_t\beta = \beta_1$ for all t. Figure 1 presents scatter plots of the autoregressive terms (that is, $\{g(y_{t-1}) - \beta_1\}$) versus the moving average terms (that is, $\{g(y_{t-1}) - \eta_{t-1}\} = \{g(y_{t-1}) - \beta_1 + c_t\}$). It is worth noting that, in all cases, there is a linear relationship and their Pearson sample correlations are greater than 0.79.
Figure 2 shows a correlation heatmap of the autoregressive term and the moving averages term with values of θ_1 and ϕ_1 in the interval $(-1, 1)$. In about 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.

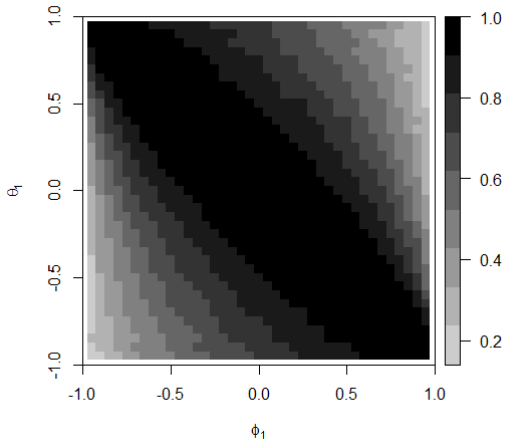


Fig. 2. Correlation between the AR and MA terms of the GARMA (1,1)

To illustrate the convergence problems, the matrix $B^{'}WB$ in (4) is examined for a GARMA(1,1) with $x_t^{'}\beta = \beta_1$ for all t (which is a condition of stationarity for the mean of a GARMA process) and link function $g(\cdot)$. In this case, the matrix $B^{'}WB$ is expressed as

$$\begin{bmatrix} \sum_{t=m+1}^{n} w_t & \sum_{t=m+1}^{n} w_t h_t & \sum_{t=m+1}^{n} w_t(g(y_{t-1}) - \eta_{t-1}) \\ \sum_{t=m+1}^{n} w_t h_t & \sum_{t=m+1}^{n} w_t h_t^2 & \sum_{t=m+1}^{n} w_t h_t(g(y_{t-1}) - \eta_{t-1}) \\ \sum_{t=m+1}^{n} w_t(g(y_{t-1}) - \eta_{t-1}) & \sum_{t=m+1}^{n} w_t h_t(g(y_{t-1}) - \eta_{t-1}) & \sum_{t=m+1}^{n} w_t(g(y_{t-1}) - \eta_{t-1})^2 \end{bmatrix}$$

where $w_t = \nu_t^{-1}\left(\frac{dg(\mu_t)}{d\mu_t}\right)^{-2}$ and $h_t = g(y_{t-1}) - \beta_1$. Considering that $g(y_t) - \eta_t$ be writ-

ten as $g(y_t) - \beta_1 + c_t$, where c_t is defined in (7), the matrix $B^{'}WB$ is rewritten as

$$\begin{bmatrix} \sum_{t=m+1}^{n} w_t & \sum_{t=m+1}^{n} w_t h_t & \sum_{t=m+1}^{n} w_t h_t + \sum_{t=m+1}^{n} c_{t-1} w_t \\ \sum_{t=m+1}^{n} w_t h_t & \sum_{t=m+1}^{n} w_t h_t^2 & \sum_{t=m+1}^{n} w_t h_t^2 + \sum_{t=m+1}^{n} c_{t-1} w_t h_t \\ \sum_{t=m+1}^{n} w_t (h_{t+1} + c_{t-1}) & \sum_{t=m+1}^{n} w_t (h_t^2 + c_{t-1} h_t) & \sum_{t=m+1}^{n} w_t \left\{ h_t^2 + 2c_{t-1} h_t + c_{t-1}^2 \right\} \end{bmatrix}$$

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 θ_1 and ϕ_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 GARMA-M(p, q) with $p \geq 1$ and $q \geq 1$ is presented as a proposal to reduce the problem of multicolinearity in the GARMA models, and to improve the interpretation of the parameters. The distribution of each observation y_t for $t = 1, ..., n$, conditional on the past information $\{\mathbf{H}_t = x_t, ..., x_1, y_{t-1}, ..., y_1, \mu_{t-1}, ..., \mu_1\}$ belongs to the exponential family as in (1). However, in the GARMA-M model, the possible serial correlation in the observations y_t is modelled in terms of the linear predictor for μ_t, by a twice-differentiable one-to-one monotonic function $g(\cdot)$, as follows:

$$\eta_t = g(\mu_t) = \underline{x}_t' \underline{\beta} + \tau_t \tag{8}$$

where,

$$\tau_t = \sum_{j=1}^{p} \phi_j \left\{ g(y_{t-j}) - \underline{x}_{t-j}' \underline{\beta} \right\} + \theta_1 \sum_{j=q}^{t-p-1} \theta_1^{j-q} \left\{ g(y_{t-p-j}) - \underline{x}_{t-p-j}' \underline{\beta} \right\}. \tag{9}$$

The parameters of the GARMA-M model can be estimated by the conditional maximum likelihood method using IRLS as in the GARMA models. In the GARMA-M model, the linear predictor defined in (9) is written such that the effect of the previous p-instants are only weighted by the autoregressive parameters ϕ_j, $j = 1, \ldots, p$ and the effects of the older observations are taken into account by a single parameter θ_1 with decreasing weights, excluding, for $q = 1$, the previous instants that have already been considered by the autoregressive parameters ϕ_j, $j = 1, \ldots, p$, and for $q = 2$ is excluded one more previous instant just considering the older observations from $t - p - 2$. However, in practice is more usual $q = 1$. Note that, when $\theta_1 = 0$, the conditional mean depends only on the first p lagged terms $g(y_{t-p}) - \underline{x}_{t-p}' \underline{\beta}$.

4.1. Properties of GARMA-M

In this subsection, theoretical expressions are derived for the marginal mean and variance of y_t following a GARMA-M model with the link functions: identity and natural

logarithm. Additionally, restrictions on the parametric space for stationary conditions for the GARMA-M(1,1) model are determined. Recalling, the terms $\mu_t = E(y_t|H_t)$ and $\varphi\nu(\mu_t) = Var(y_t|H_t)$ are respectively the mean and variance of y_t conditioned on the past information and $E(y_t|x_t)$ and $Var(y_t|x_t)$ represent the marginal (unconditional on the previous information) mean and variance of y_t, where φ and $\nu(\mu_t)$ are respectively the scale parameter and variance function of the exponential family distribution.

Theorem 1.
The marginal mean and variance of y_t of a GARMA-M model defined by (1) and (9) with identity link function are

$$\begin{aligned} a) \quad & E(y_t|x_t) \quad = \underline{x}_t'\underline{\beta}, \\ b) \quad & Var(y_t|x_t) = \varphi E[\Psi^{(2)}(B)\nu(\mu_t)], \end{aligned}$$

provided that $\Psi^{(2)}(B) = 1 + \psi_1^2 B + \psi_2^2 B^2 + ...$ (as in (18)) exists. The marginal mean is stationary if $\underline{x}_t'\underline{\beta} = \beta_0$ for all t. The proofs of Theorem 1a and 1b are respectively presented in Appendices B.1 and B.2.

Theorem 2.
The marginal mean and variance of y_t of a GARMA-M model with the logarithm natural link function are

$$\begin{aligned} a) \quad & E(y_t|x_t) \quad \approx e^{x_t'\beta}\left[1 - \frac{\varphi}{2c_x^2}\sum_{j=1}^{t-1}\alpha_j\Psi(B)E\left[\nu(\mu_{t-j})\right]\right] \\ b) \quad & Var(y_t|x_t) \approx [E(y_t)]^2\Psi^2(B)\left\{\varphi E\left(\frac{\nu(\mu_t)}{\mu_t^2}\right) + E(a_t^2)\right\} \end{aligned}$$

provided that $\Psi^{(2)}(B) = 1 + \psi_1^2 B + \psi_2^2 B^2 + ...$ (as in (18)) exists. As in the Theorem 1, the marginal mean is stationary if $x_t'\beta = \beta_0$ for all t. The expressions for $\alpha's$, a_t and c_x are defined respectively in (10), (20) and (22). The proofs are presented in Appendices B.3 and B.4. For the Poisson distribution, the stationary marginal mean and variance of y_t of the GARMA-M model with identity link and natural logarithm link function are given by

$$\begin{aligned} E(y_t|x_t) &= \beta_0 \\ Var(y_t|x_t) &= \Psi^{(2)}(1)\beta_0, \end{aligned}$$

and

$$\begin{aligned} E(y_t|x_t) &\approx \exp(\beta_0)\left[1 - \frac{\Psi^{(2)}(1)\beta_0}{2c_x^2}\sum_{j=1}^{t-1}\alpha_j\right], \\ Var(y_t|x_t) &\approx [E(y_t|x_t)]^2\Psi^2(B)\left\{E\left(\frac{1}{\mu_t}\right) + E(a_t^2)\right\} \end{aligned}$$

where $\Psi^{(2)}(1) = 1 + \psi_1^2 + \psi_2^2 + ...$ and the $\alpha' s_j$ are the parameters of the GARMA(p,q) model, written as

$$\alpha_j = \begin{cases} \phi_j & \text{if } 1 \leq j \leq p \\ \theta_1^{j-p} & \text{if } j \geq p+q. \end{cases} \tag{10}$$

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

The marginal mean of y_t in the GARMA-M(p,q) models with link functions identity and logarithm natural is stationary if $\underline{x}_t'\beta = \beta_0$ 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 (Cryer and Kellet, 1986). That is, to get roots greater than one, both conditions

$$\begin{cases} \sum \alpha_j & < 1 \\ |\alpha_j| & < 1 \end{cases}$$

are necessary but not sufficient. For the GARMA-M(1,q) model, the first condition, i.e, the sum of the $\alpha' s_j$ must be less than 1 leads to an additional parametric restriction,

$$\phi_1 < \frac{\theta_1^{n-q} + 1 - 2\theta_1}{1 - \theta_1}, \tag{11}$$

where the parameters ϕ_1 and θ_1 assume values in the interval $(-1, 1)$. Figure 3 shows the parameter restriction space of the GARMA-M(1,1) model.

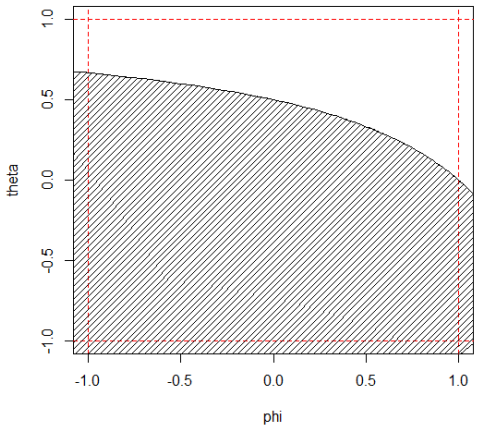


Fig. 3. The parameter restriction space for GARMA-M(1,1) model

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 analyze respectively

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

n	Parameter	True value	Mean	S.D.	% coverage	BIAS	MSE
100	β_0	2.944	2.940	0.024	80.9	0.004	0.001
	β_1	0.246	0.247	0.034	81.4	-0.001	0.003
	ϕ_1	0.220	0.194	0.100	94.0	0.026	0.011
	θ_1	0.100	0.064	0.098	95.4	0.036	0.010
500	β_0	2.944	2.943	0.016	82.0	0.001	<0.001
	β_1	0.246	0.247	0.022	81.9	-0.001	<0.001
	ϕ_1	0.220	0.215	0.043	94.8	0.005	0.002
	θ_1	0.100	0.093	0.041	95.7	0.007	0.002
1000	β_0	2.944	2.944	0.012	86.9	<0.001	<0.001
	β_1	0.246	0.247	0.016	90.1	-0.001	<0.001
	ϕ_1	0.220	0.219	0.035	95.9	0.001	0.001
	θ_1	0.100	0.096	0,033	96.5	0.004	0.001

their autocorrelation function and their normal Q-Q plot. Benjamin et al. (2003) advocate to use the normalized conditional (randomized) quantile residuals of Dunn and Smyth (1996) 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 variate 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 conditional fitted cumulative distribution function.

5. Simulation study

Two simulations studies are performed to measure the consistency of the parameter estimators of the GARMA-M model in terms of the mean square errors and to compare the performance of the GARMA(p,q) and GARMA-M(p, q) models with similar serial correlation function.

5.1. Consistency of the GARMA-M parameters estimators

To measure the consistency of the parameter estimators of the GARMA-M model, time series y_t with different lengths n are simulated 10.000 times according to Poisson-GARMA-M(1,1) with a logarithm link function and Gamma-GARMA-M(1,1) with an identity link function, then the parameters of GARMA-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 ML estimators are asymptotically Gaussian, unbiased with variance given by the inverse of Fisher information matrix, to determine the coverage percentage.

Scenario 1a. In the first scenario, time series are simulated according to Poisson-GARMA-M(1,1) with linear predictor given by

$$ln(\mu_t) = \beta_0 + cos(2\pi t/52.25)\beta_1 + \tau_t$$

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

n	Parameter	True value	Mean	S.D.	% coverage	BIAS	MSE
100	β_0	19.01	19.00	0.548	84.5	-0.010	0.782
	β_1	1.280	1.275	0.790	87.5	0.005	1.329
	ϕ_1	0.220	0.196	0.097	92.9	0.024	0.012
	θ_1	0.100	0.062	0.086	93.9	0.038	0.011
500	β_0	19.00	19.00	0.387	90.1	-0.001	0.150
	β_1	1.280	1.284	0.543	89.5	-0.004	0.295
	ϕ_1	0.220	0.215	0.046	93.9	0.005	0.002
	θ_1	0.100	0.093	0.042	94.1	0.007	0.002
1000	β_0	19.00	18.99	0.325	90.1	<0.001	0.106
	β_1	1.280	1.281	0.455	89.9	-0.001	0.211
	ϕ_1	0.220	0.218	0.037	94.7	0.002	0.002
	θ_1	0.100	0.096	0.034	95.7	0.004	0.001

with parameters: $\beta_0 = ln(19) \approx 2.944$; $\beta_1 = ln(1.28) \approx 0.246$; $\phi_1 = 0.22$; $\theta_1 = 0.1$. 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 and the coverage percentages for ϕ_1 and θ_1 are higher than $> 94\%$, but it was 81% for β_0 and β_1, while for $n = 1000$ these coverage percentages are at least 87%.

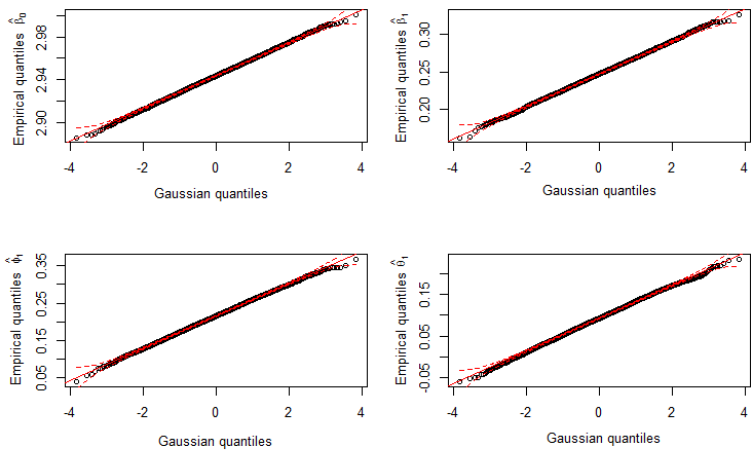


Fig. 4. QQ plots (simulated estimates parameters) for Scenario 1a with $n = 500$

Scenario 1b. In this scenario time series y_t are simulated according to Gamma-GARMA-M(1,1) with linear predictor given by

$$\mu_t = \beta_0 + cos(2\pi t/52.25)\beta_1 + \tau_t$$

with 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 $n = 1000$ 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 values of MSE are higher for β_0 and β_1. Figures 4 and 5 show a normal quantile-quantile plot of the estimated parameters, for scenario 1a and 1b for $n = 500$, they seem to be symmetric and normally distributed.

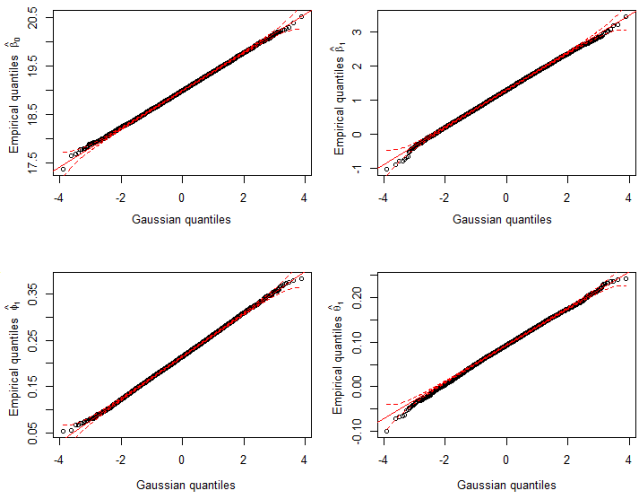


Fig. 5. QQ plots (simulated estimates parameters) for Scenario 1b with $n = 500$

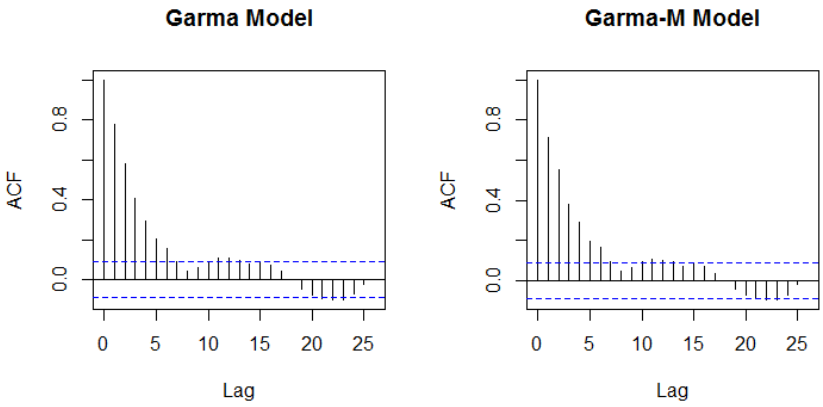


Fig. 6. ACF of y_t for Poisson-GARMA(1,1) and GARMA-M(1,1) models

5.2. Performance of the GARMA(p, q) and GARMA-M(p, q) models

In order to compare the performance of the GARMA(1,1) and GARMA-M(1,1) models, a second simulation study is performed. Stationary time series y_t are simulated according to Poisson distribution with a logarithm link function and Gamma distribution with an

Table 3. Estimates of parameters of the two models - Scenario 2a - Poisson

	Estimator	True Value	Mean	S.D.	% Coverage	Bias	MSE
GARMA	β_0	2.944	2.941	0.015	40.8%	0.003	0.006
	ϕ_1	0.700	0.688	0.036	92.3%	0.012	0.011
	θ_1	0.100	0.104	0.055	95.1%	-0.004	0.001
GARMA-M	β_0	2.944	2.940	0.016	47.9%	0.004	0.007
	ϕ_1	0.640	0.635	0.043	94.3%	0.005	0.010
	θ_1	0.050	0.047	0.013	95.7%	0.007	0.001
GARMA-M	β_0	2.944	2.941	0.016	42.1%	0.003	0.006
	ϕ_1	0.700	0.689	0.043	95.3%	0.011	0.012
	θ_1	0.100	0.102	0.045	94.7%	-0.002	0.001

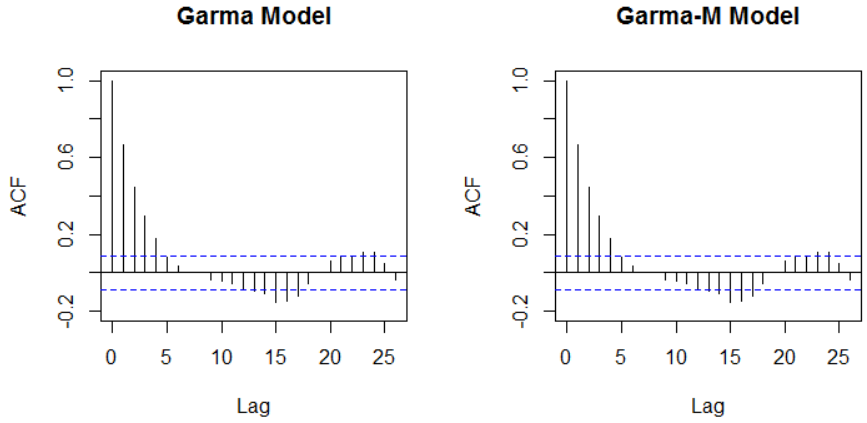


Fig. 7. ACF of y_t for Gamma-GARMA(1,1) and GARMA-M(1) models

identity link function. To avoid multicollinearity problems in the Poisson-GARMA(1,1) model the values of parameters : $\beta_0 = \ln(19) \approx 2.944$, $\phi_1 = 0.7$ and $\theta_1 = 0.1$ were considered. For this model, the correlation between the autoregressive and moving average terms is around 0.65. An analogous model Poisson- GARMA-M(1,1) with a logarithm link function, in the sense to have a similar serial correlation function, is considered with parameters $\beta_0 = \ln(19) \approx 2.944$, $\phi_1 = 0.64$ and $\theta_1 = 0.05$. The similarity of the autocorrelation functions may be visualized in Figure 6.

Table 3 shows the estimates of the parameters of the Poisson- GARMA(1,1) and Poisson- GARMA-M(1,1) models, respectively, for $n = 500$. Note that, for both models the average of the parameters estimates are close to the true values, however the percentage of coverage of β_0 parameter for both models are low, but for the GARMA-M model it is higher than in the GARMA model. By simulation, it was verified that for high values of ϕ_1 or θ_1 the distribution of the estimated parameter β_0 is a little asymmetric which leads to low coverage percentage.

For the second scenario, time series according to Gamma-GARMA model, with an

Table 4. Estimates of parameters of the two models- Scenario 2b - Gamma

	Estimator	True Value	Mean	S.D.	% Coverage	Bias	MSE
GARMA	β_0	19.00	18.969	0.275	44.2	0.031	0.203
	ϕ_1	0.700	0.688	0.036	94.3	0.012	0.002
	θ_1	0.100	0.103	0.055	95.1	-0.003	0.004
GARMA-M	β_0	19.00	18.97	0.216	50.1	0.028	0.102
	ϕ_1	0.640	0.626	0.043	97.3	0.014	0.002
	θ_1	0.045	0.040	0.012	94.7	0.005	0.001
GARMA-M	β_0	19.00	18.97	0.266	46.1	0.030	0.181
	ϕ_1	0.700	0.689	0.039	94.5	0.010	0.002
	θ_1	0.100	0.096	0.035	96.7	0.006	0.001

identify link function, are simulated with parameters: $\beta_0 = 19$, $\phi_1 = 0.7$ and $\theta_1 = 0.1$. For this model, the correlation between the autoregressive and moving average terms is around 0.66. An analogous model Gamma- GARMA-M(1,1), in the sense to have a similar serial correlation function, is considered with parameters $\beta_0 = 19$, $\phi_1 = 0.64$ and $\theta_1 = 0.04$ with an identify link function. The similarity of two autocorrelation functions may be visualized in Figure 7. The results are similar to the scenario 2a as presented in Table 4. However, the values of the MSE in this case are higher.

6. Real data analysis

In this section, a GARMA-M model is fitted to analyze 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 including explanatory variables. The daily admission time series is obtained from Hospital Information System at Health Secretary of São Paulo (PRO-AIM). Only a Negative Binomial distribution was considered once the scale parameter, k, is too small, indicating the presence of overdispersion, since $Var(y_t|H_t) = \mu_t + \mu_t^2/k$. Additionally, the logarithm function is the chosen link function $g(\mu_t) = ln(\mu_t)$ with

$$x_t'\beta = \beta_0 + \beta_1 sin(2\pi t/365) + \beta_2 Mon_t + \beta_3 Fri_t + \beta_4 Sat_t + \beta_5 Sun_t + \tau_t \qquad (12)$$

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 (Draper and Smith, 2014) and

$$\tau_t = g(y_{t-1}) - \underline{x}_{t-1}'\underline{\beta} + \sum_{j=1}^{t-2} \theta_1^j \left\{ g(y_{t-1-j}) - \underline{x}_{t-1-j}'\underline{\beta} \right\}. \qquad (13)$$

as in (9). Firstly, a GLM model was fitted to the data but their first residuals autocorrelations were signficant. Then a GARMA-M(1,0) and a GARMA-M(0,1) models are considered but the residuals were still autocorrelated. Thus the GARMA-M(1,1) is fitted and finally their residual autocorrelations were all non-significant as presented in Figure 9. A GARMA-M(2,0) model was also considered but the parameter ϕ_2 was not significant (p>0.05). Figure 8 shows the observed daily number of admissions, its predicted values using the negative binomial GARMA-M(1,1) model with the linear predictor as

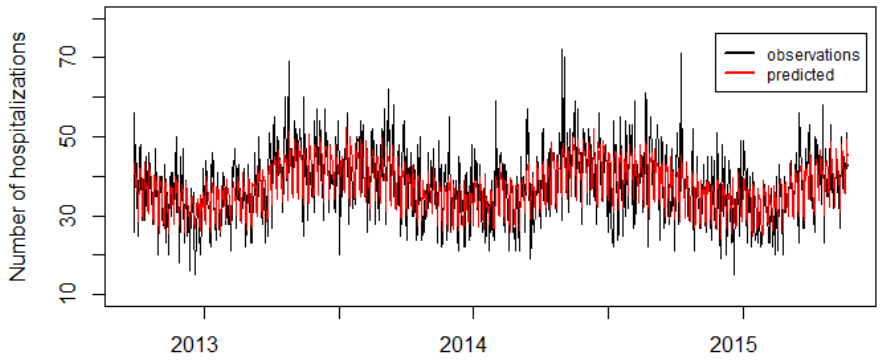


Fig. 8. Number of admissions due to respiratory diseases for people aged over 60 years in So Paulo from October 2012 to April 2015 and predicted counts.

in (13).

Figure 9 presents the quantile-quantile plot of the Gaussian residuals and they seem to be normally distributed. Also, the normality hypothesis is accepted using the ShapiroWilk test (p= 0.110), confirming that the assumption of a Negative Binomial distribution for the counts is appropriate. The estimates of all coefficients for the NB-GARMA-M(1,1) are in Table 5. Both the autoregressive parameters are significant (p<0.05). The daily residuals $(ln(y_t) - x_t\beta)$ vary from -0.5 to 0.5 thus the estimated autorregressive parameter is 0.222. This means that an increase of 0.1 unit of previous daily residual leads an average increase of 2.2% $(exp(\phi_1 * 0.1))$ percent in the average number of hospitalizations of elderly people due to respiratory diseases in São Paulo city.

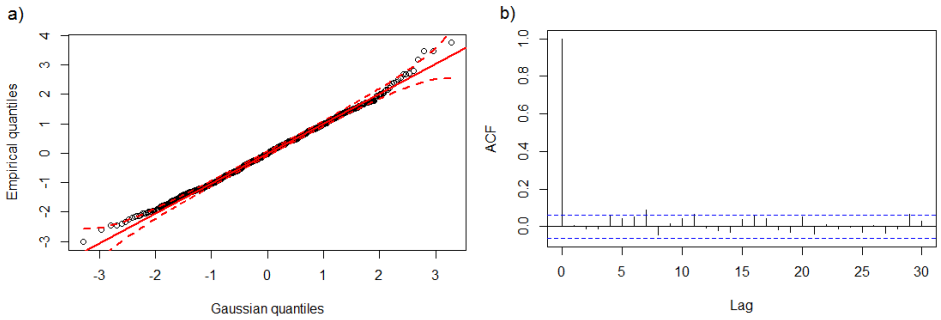


Fig. 9. Correlogram and Q-Q plot of residuals GARMA-M (1,1) model. (a) ACF plot and (b) Q-Q plot

On the other hand, it is not possible to get the maximum likelihood estimates of the parameters of GARMA(1,1) using IRLS due a non converge. The sample Pearson corre-

Table 5. Estimates, standard errors, and p-values
GARMA-M(1,1) model

Coefficient	Estimate	S.E.	p-value
Intercept (β_0)	3.717	0.015	<0.001
Sine (β_1)	-0.132	0.008	<0.001
Sunday (β_2)	-0.298	0.022	<0.001
Monday (β_3)	0.047	0.021	0.023
Friday (β_4)	-0.084	0.021	<0.001
Saturday (β_5)	-0.233	0.022	<0.001
ϕ_1	0.222	0.031	<0.001
θ_1	0.062	0.031	0.043
Dispersion (k)	0.005	0.001	

lation between the autoregressive and moving average terms for this real data was 0.898. However, the parameters of GARMA(1,1) may be estimated using the *optim* function in R (as recommended in the new package update *garmaFit*) but 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 signal ($\widehat{\phi}_1 = 0.362$ (0.052) and $\widehat{\theta}_1 = -0.130$ (0.045)). The negative term of the moving average coefficient estimate seems to cancel part of the over estimated autorregressive estimate, the standard error for the GARMA(1,1) is inflated due to the multicollinearity.

7. Conclusions and discussions

The main goal of this paper is to analyze the performance of the GARMA model proposed by Benjamin et al. (2003). In this model, the mean as in the GLM is related to a linear predictor by a twice- differentiable one-to-one monotonic link function. However in the GARMA model, autoregressive and the moving average terms are included to model possible serial correlations present in the observations. The moving average error terms included in the linear predictor can be different type of residuals, for example, the deviance residuals, Pearson residuals measured on the original scale or on the predictor scale (i.e, $g(y_t) - g(\mu_t)$). The simultaneous inclusion of this type of residual and autoregressive terms in the linear predictor, independent of choice of the link function and the distribution from the exponential family, induces multicollinearity in the GARMA(p,q) models with p and q simultaneously different from zero. In these cases, the same lagged observations are taken into account simultaneously in the autoregressive and the moving average parameters leading to the non-convergence of the algorithm IRLS used to find the estimates of the parameters turning unclear the interpretation of the parameters.

Another interesting feature in the GARMA models is the similar performance of GARMA(1,0) and GARMA(0,1) models when $\theta_1 = \phi_1$. This occurs due to, in this case, both models give the same weight for the common term $\left\{g(y_{t-1}) - x_{t-1}^{'}\beta\right\}$ and for the older instants, also considered in the moving average term, are given decreasing weights θ_1^j for j=2,3,... In addition, in the GARMA(1,q) models with $q > 1$, many previous terms are measured simultaneously, with different weights, by the q moving average parameters turning again confuse the interpretation of the parameters.

To deal with these problems, a modified model namely GARMA-M(p,q) model is proposed (which also takes into account the long range dependence or the long memory). 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 ϕ_j, $j = 1, \ldots, p$ and the effect of the older observations are taken into account by a single parameter θ_1, with decreasing weights. In simulation studies, it was shown that the estimates are close to the true values, but like the GARMA models, in GARMA-M with high values for the autoregressive parameters, the distribution of the coefficient estimates of the explanatory variable is slightly asymmetric which leads to a low coverage percentage for these parameters.

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 get stationarity. For example, for GARMA-M(1,1) with $\phi_1 > 0$ the value of θ_1 has to be lower than 0.5, but θ_1 is usually smaller in practice. An interesting feature in the GARMA-M(1,1) models is that, in several cases, for small values of the parameter θ_1 is possible to reproduce a similar autocorrelation function of the GARMA models (p,q). In our simulation studies, the estimates of the parameters of GARMA-M model are close to the true values and a little higher coverage percentage than the original GARMA model is obtained. However, these results can not be generalized. In addition, the asymptotic normality of the parameters of the GARMA-M model, analyzed by simulations studies, indicated that the normal distribution can be assumed. However, it should be noted that there is a little bias in the estimation of the parameter θ_1.

Finally, it should be noted that, the performance of the GARMA and GARMA-M models with inflated zeros should be analyzed in more detail. In addition, an offset term can be added to the linear predictor of the GARMA-M model.

Acknowledgements

The authors would like to acknowledge CNPq,(Process 304670/2014-6) Capes and Fapesp (Grant number 13/00506-1) for the financial support.

References

Albarracin, O. Y. E., A. P. Alencar, and L. Lee Ho (2017). Cusum chart to monitor autocorrelated counts using negative binomial garma model. *Statistical methods in medical research*, 0962280216686627.

Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association 98*(461), 214–223.

Cryer, J. D. and N. Kellet (1986). *Time series analysis*, Volume 101. Springer.

Draper, N. R. and H. Smith (2014). *Applied regression analysis*. John Wiley & Sons.

Dugas, A. F., M. Jalalpour, Y. Gel, S. Levin, F. Torcaso, T. Igusa, and R. E. Rothman (2013). Influenza forecasting with google flu trends. *PloS one 8*(2), e56176.

Dunn, P. K. and G. K. Smyth (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics, 5*, 236–244.

Dunsmuir, W. T., D. J. Scott, et al. (2015). The glarma package for observation driven time series regression of counts. *Journal of Statistical Software 67*(7), 1–36.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, 149–192.

Kaufmann, H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory. *The Annals of Statistics*, 79–98.

Li, W. (1994). Time series models based on generalized linear models: some further results. *Biometrics*, 506–511.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2th ed.). London: Chapman & Hall/CRC.

McCulloch, C. E. and J. M. Neuhaus (2001). *Generalized linear mixed models*. Wiley Online Library.

Shumway, R. H. and D. S. Stoffer (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, 1019–1031.

Appendix

A. Estimation of the parameters

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

$$l = \sum_{t=m+1}^{n} \ln\ f(y_t|H_t). \tag{14}$$

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

$$A_1 = \begin{bmatrix} g(y_m) - x_m^{'}\beta & ... & g(y_{m+1-p}) - x_{m+1-p}^{'}\beta \\ g(y_{m+1}) - x_{m+1}^{'}\beta & ... & g(y_{m+2-p}) - x_{m+2-p}^{'}\beta \\ \vdots & & \vdots \\ g(y_{n-1}) - x_{n-1}^{'}\beta & ... & g(y_{n-p}) - x_{n-p}^{'}\beta \end{bmatrix},$$

$$\mathrm{A}_2 = \begin{bmatrix} g(y_m) - \eta_m & ... & g(y_{m+1-q}) - \eta_{m+1-q} \\ g(y_{m+1}) - \eta_{m+1} & ... & g(y_{m+2-q}) - \eta_{m+2-q} \\ \vdots & & \vdots \\ g(y_{n-1}) - \eta_{n-1} & ... & g(y_{n-q}) - \eta_{n-q} \end{bmatrix}.$$

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

$$\eta_t = B'\underline{\gamma},$$

where $B = \begin{bmatrix} X & A_1 & A_2 \end{bmatrix}$, 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

$$\hat{\gamma}^{(k+1)} = \left(B^{'}W^{(k)}B\right)^{-1} B^{'}W^{(k)}z^{(k)}, \tag{15}$$

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

$$W = diag\left\{w_{(m+1)}, ..., w_n\right\}$$

where, $y = (y_{(m+1)}, ..., y_n)^{'}$, $\mu = (\mu_{(m+1)}, ..., \mu_n^{'})$, $V = diag\left\{\nu(\mu_{(m+1)}), ..., \nu(\mu_n)\right\}$ is a diagonal matrix of variance functions and $w_j = \nu(\mu_j)^{-1}\left(\dfrac{dg(\mu_j)}{d\mu_j}\right)^{-2}$.

The convergence of (15) occurs generally in a finite number of steps. Evaluating γ using IRLS allows to obtain the approximate large-sample conditional variance of $\widehat{\gamma}$ as a by product of the iterative process (Green (Green, 1984) ; Kaufmann (Kaufmann, 1987)). It can be shown that asymptotically $\sqrt{(n-m)}(\widehat{\gamma} - \gamma) \sim N(0, I(\gamma)^{-1})$, where

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

B. Proofs of Theorems 1 and 2

B.1. Proof of Theorem 1a.

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

$$\mu_t = x_t'\beta + \sum_{j=1}^{t-1} \alpha_j \left\{ y_{t-j} - x_{t-j}'\beta \right\}, \tag{16}$$

where,

$$\alpha_j = \begin{cases} \phi_j & \text{if } 1 \leq j \leq p \\ \theta^{j-p} & \text{if } j > p. \end{cases} \tag{17}$$

Let $y_t = \mu_t + \epsilon_t$. Then the ϵ_t are uncorrelated errors with mean 0. Replacing μ_t by (16) and defining $w_t = y_t - x_t'\beta$ gives,

$$\begin{aligned} w_t &= \sum_{j=1}^{t-1} \alpha_j w_{t-j} + \epsilon_t \\ w_t &= \Psi(B)\epsilon_t \end{aligned}$$

where,

$$\Psi(B) = \Phi^{-1}(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + ... \tag{18}$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$, provided that $\Phi(B)$ to be invertible. Then,

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

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

B.2. Proof of Theorem 1b.

The marginal variance of y_t is given by,

$$\begin{aligned} Var(y_t|x_t) &= Var(w_t + x_t'\beta|x_t) = E(w_t^2) = \\ &= E\left(\sum_{j=0}^{\infty} \psi_j B^j \epsilon_t \sum_{i=0}^{\infty} \psi_i B^i \epsilon_t \right) \\ &= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_i E(\epsilon_{t-j}\epsilon_{t-i}) \\ &= \sum_{j=0}^{\infty} \psi_j^2 E(\epsilon_{t-j}^2) \\ &= \Psi^{(2)}(B) E(\epsilon_t^2), \end{aligned}$$

where, $\Psi^{(2)}(B) = 1+\psi_1^2 B+\psi_2^2 B^2+...$ The variance of the martingale errors is, $Var(\epsilon_t) = E(\epsilon_t^2) = E[E(\epsilon_t^2|H_t)]$, where, $Var(\epsilon_t|H_t) = Var((y_t - \mu_t)|H_t) = Var(y_t|H_t) = \varphi\nu(\mu_t)$. Hence,

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

B.3. Proof of Theorem 2a.

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

$$\ln(\mu_t) \quad = \quad x_t'\beta + \sum_{j=1}^{t-1} \alpha_j \left\{\ln(y_{t-j}) - x_{t-j}'\beta\right\}, \tag{19}$$

where α_j is defined in (17). Applying the Taylor series expansion for $\ln(y_t)$ at the point μ_t gives

$$\ln(y_t) = \ln(\mu_t) + \frac{1}{\mu_t}(y_t - \mu_t) + a_t, \tag{20}$$

where, a_t is the approximation error. Substituting $\ln(\mu_t)$ from (19) in (20) and let $w_t' = \ln(y_t) - x_t'\beta$, gives

$$\begin{aligned} w_t' &= \sum_{j=1}^{t-1} \alpha_j w_{t-j}' + \frac{1}{\mu_t}(y_t - \mu_t) + a_t \\ \Phi(B)w_t' &= \frac{1}{\mu_t}(y_t - \mu_t) + a_t \\ w_t' &= \Psi(B)\left[\frac{1}{\mu_t}(y_t - \mu_t) + a_t\right] \end{aligned}$$

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

$$E(w_t') = \Psi(B)E(a_t), \tag{21}$$

because $E\left(\dfrac{y_t - \mu_t}{\mu_t}\right) = 0$. Then,

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

Taking the exponential function of both sides in (19) and applying the Taylor series for the function e^x at the point $x = 0$ yields

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

where b_t is the approximation error. Hence,

$$\begin{aligned} E(\mu_t|x_t) &= \exp\left(x_t'\beta\right) + \exp\left(x_t'\beta\right)\left[\sum_{j=1}^{t-1}\alpha_j E\left\{\ln(y_{t-j}) - x_{t-j}'\beta\right\} + E(b_t)\right] \\ &= \exp\left(x_t'\beta\right) + \exp\left(x_t'\beta\right)\left[\sum_{j=1}^{t-1}\alpha_j E\left(w_{t-j}'\right) + E(b_t)\right] \\ &= \exp\left(x_t'\beta\right) + \exp\left(x_t'\beta\right)\left[\sum_{j=1}^{t-1}\alpha_j \Psi(B)E(a_{t-j}) + E(b_t)\right]. \end{aligned}$$

The error a_t in (20) can be rewritten as

$$a_t = -\frac{(y_t-\mu_t)^2}{2c_x^2} = -\frac{\epsilon_t^2}{2c_x^2}, \text{ for some } c_x \in (\mu_t, y_t). \tag{22}$$

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

$$E[\epsilon_t^2] = E[E(\epsilon_t^2|H_t)] = E(Var(\epsilon_t+\mu_t|H_t)) = E[Var(y_t|H_t)] = E[\varphi\nu(\mu_t)].$$

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

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

Hence,

$$E(y_t|x_t) = E[E(y_t|H_t)] = e^{x_t'\beta}\left[1 - \frac{\varphi}{2c_x^2}\sum_{j=1}^{t-1}\alpha_j\Psi(B)E\left[\nu(\mu_{t-j})\right]\right] \tag{23}$$

assuming that $E(b_t) = 0$.

B.4. Proof of Theorem 2b.

In this proof, the Taylor expansions are applied to calculate the variance of $f(x)$. Hence,

$$Var\left[\ln(y_t)\right] = \frac{1}{\left[E(y_t)\right]^2}Var(y_t) + d_t,$$

where d_t is the approximation error. The variance of $\ln(y_t)$, is given by

$$Var\left[\ln(y_t)\right] = Var(w_t') = E(w_t'^2) - (E(w_t'))^2.$$

where w_t' is defined in the proof of the Theorem 2a. Thus,

$$E(w_t'^2) = E(\Psi(B)(s_t+a_t)\Psi(B)(s_t+a_t)),$$

denoting $s_t = \dfrac{y_t - \mu_t}{\mu_t}$, we have

$$\begin{aligned} E(w_t'^2) &= E\left(\sum_{i=0}^{\infty}\psi_i(s_{t-i}+a_{t-i})\sum_{j=0}^{\infty}\psi_j(z_{t-j}+a_{t-j})\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j E\left\{(s_{t-i}+a_{t-i})(s_{t-j}+a_{t-j})\right\} \end{aligned}$$

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

$$E(w_t'^2) = \Psi^{(2)}(B)\left\{E(s_t^2)+E(a_t^2)\right\}.$$

with the variance of s_t expressed as

$$\begin{aligned} Var(s_t) &= E(s_t^2) \\ &= E\left\{E\left[\left(\frac{y_t-\mu_t}{\mu_t}\right)^2|H_t\right]\right\} \\ &= E\left\{\frac{1}{\mu_t^2}\left[Var(y_t|H_t)+(E(y_t|H_t))^2\right]-\frac{2}{\mu_t}E(y_t|H_t)+1\right\} \\ &= \varphi E\left\{\frac{\nu(\mu_t)}{\mu_t^2}\right\}. \end{aligned}$$

Hence,

$$Var(y_t) = [E(y_t)]^2\Psi^2(B)\left\{\varphi E\left(\frac{\nu(\mu_t)}{\mu_t^2}\right)+E(a_t^2)\right\},$$

assuming that $E(d_t)=0$ and $E(y_t)$ is given in (23).

Capítulo 5

Conclusions

The main goal of the two first papers in this thesis was to evaluate the importance of taking into account the serial correlation when time series of counts are monitored using CUSUM and EWMA charts. The GARMA model proposed by Benjamin *et al.* [2003] was considered in this study due to it allows to consider the Poisson or Negative Binomial distribution and to include of autoregressive and the moving average terms to model the autocorrelation.

The impacts on the performance of CUSUM and EWMA control charts when the serial correlation is omitted and a GLM model under independence is fitted, as it is supposed to occur in practice, instead of a GARMA model, that is used to simulate the time series was measured in terms of average run length (ARL) in a simulation study where different scenarios that describe the profiles of time series of infectious diseases was considered. Different statistics based on transformations, the deviance residual and the likelihood ratio were used to build CUSUM control charts. On the other hand, two alternative statistics were proposed to build EWMA charts assuming the Negative Binomial distribution. The first one consists in centralizing the ratio of log-likelihood functions (LR) widely used in the literature and the second is based on the maximum likelihood estimator (MLE) of the shift size k ($\mu_{1,t} = \mu_{0,t}exp(k)$) since that it is easy to interpret.

In general, when the autocorrelation is neglected adjusting a "pure"GLM instead of a GARMA model will lead to an increase of false alarms. However no statistics among the tested ones seem to be robust, in a sense to produce the smallest increase of false alarms in the evaluated models. About ARL_1, it looks that the misspecification of the autocorrelation does not provoke a serious delay. The monitoring of the weekly number of hospital admissions due to respiratory diseases for people aged over 65 years in the city São Paulo-Brazil was considered as an illustration of the current method. A GARMA(2,0) was fitted from January 2006 to December 2010 including a seasonal component and a linear trend in the expected number of hospitalizations. Then, the CUSUM control charts for 2011 considering all statistics indicated alarms in January, February, April and December. These weeks really presented higher number of hospitalization (an increase higher than 35%), indicating that all methods are able to detect deviations in the mean number of hospitalizations.

Finally, in third paper the main goal was to analyze the performance of the GARMA(p, q) model with p and q simultaneously different from zero. In these cases, the same lagged observations are taken into account simultaneously in the autoregressive and the moving average parameters leading to the non-convergence of the algorithm IRLS used to find the estimates of the parameters turning unclear the interpretation of the parameters. To deal with these problems, a modified model namely GARMA-M(p,q) model was proposed (which also takes into account the long range dependence or the long memory). In simulation studies, it was shown that the estimates are close to the true values, but like the GARMA models, in GARMA-M with high values for the autoregressive parameters, the distribution of the coefficient estimates of the explanatory variable is slightly asymmetric which leads to a low coverage percentage for these parameters. 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 get stationarity. Finally, a GARMA-M model was fitted to analyze 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 including explanatory variables, being the GARMA-M a good option model to fitted count time series.

Referências Bibliográficas

Benjamin *et al.***(2003)** Michael A Benjamin, Robert A Rigby e D Mikis Stasinopoulos. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98(461): 214–223. Citado na pág. 2, 67

Chao-Wen e Reynolds Jr(2001) Lu Chao-Wen e Marion R Reynolds Jr. Cusum charts for monitoring an autocorrelated process. *Journal of Quality Technology*, 33(3):316. Citado na pág. 1

Heinen(2003) Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*. Citado na pág. 1

Hohle e Mazick(2009) M. Hohle e A. Mazick. Aberration detection in r illustrated by danish mortality monitoring. *Biosurveillance: A Health Protection Priority*. Citado na pág. 1

McCullagh e Nelder(1989) P. McCullagh e J. A. Nelder. *Generalized Linear Models*. Chapman & Hall/CRC, London, 2th edição. Citado na pág. 1

Rogerson e Yamada(2004) P. A. Rogerson e I. Yamada. Approaches to syndromic surveillance when data consist of small regional counts. *Morbidity and Mortality Weekly Report*, 53/Supplement:79–85. Citado na pág. 1

Sparks *et al.***(2010)** Ross S Sparks, Tim Keighley e David Muscatello. Early warning cusum plans for surveillance of negative binomial daily disease counts. *Journal of Applied Statistics*, 37(11): 1911–1929. Citado na pág. 1

Sparks *et al.***(2011)** RS Sparks, T Keighley e D Muscatello. Optimal exponentially weighted moving average (ewma) plans for detecting seasonal epidemics when faced with non-homogeneous negative binomial counts. *Journal of Applied Statistics*, 38(10):2165–2181. Citado na pág. 1

Unkel *et al.***(2012)** S Unkel, C Farrington, PH Garthwaite, C Robertson e A NickBox. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society A*, 175(1):40–82. Citado na pág. 1

Weiß(2009) Christian H Weiß. Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, 18(4):507–519. Citado na pág. 1

Weiß e Testik(2011) Christian H Weiß e Murat Caner Testik. The poisson inar (1) cusum chart under overdispersion and estimation error. *IIE Transactions*, 43(11):805–818. Citado na pág. 1