
GARMA models, a new perspective using
Bayesian methods and transformations

Breno Silveira de Andrade

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Breno Silveira de Andrade

GARMA models, a new perspective using Bayesian methods and
transformations

Doctoral dissertation submitted to the Departamento de Estatística - DEs - UFSCar and to the Instituto de Ciências Matemáticas e de Computação - ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Joint Graduate Program in Statistics DEs-UFSCar/ICMC-USP. FINAL VERSION

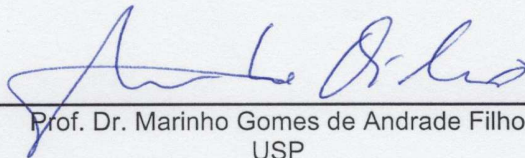
Advisor: Prof. Dr. Marinho Gomes de Andrade Filho

São Carlos

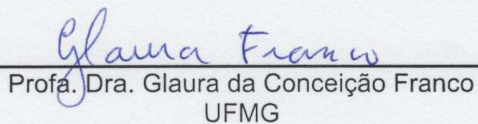
February 2017

Folha de Aprovação

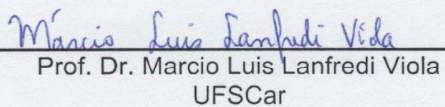
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Breno Silveira de Andrade, realizada em 16/12/2016:



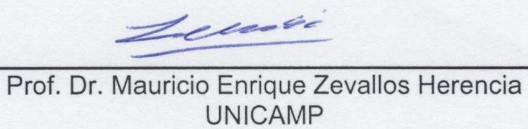
Prof. Dr. Marinho Gomes de Andrade Filho
USP



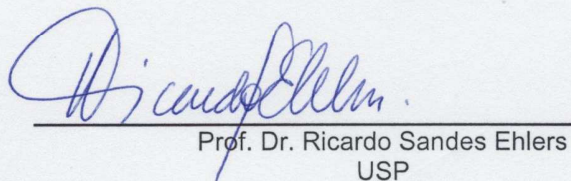
Profa. Dra. Glaucia da Conceição Franco
UFMG



Prof. Dr. Marcio Luis Lanfredi Viola
UFSCar



Prof. Dr. Mauricio Enrique Zevallos Herencia
UNICAMP



Prof. Dr. Ricardo Sandes Ehlers
USP

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Breno Silveira de Andrade

Modelos GARMA, uma nova perspectiva usando métodos
Bayesianos e transformações.

Tese apresentada ao Departamento de Estatística -
Des/UFSCar e ao Instituto de Ciências Matemáticas
e de Computação - ICMC-USP, como parte dos
requisitos para obtenção do título de Mestre ou
Doutor em Estatística - Programa Interinstitucional
de Pós-Graduação em Estatística UFSCar-USP.
VERSÃO FINAL

Orientador: Prof. Dr. Marinho Gomes de Andrade
Filho

São Carlos
Fevereiro de 2017

Agradecimentos

Agradeço, primeiramente a Deus e Nossa Senhora.

Agradeço também:

A meus pais e demais familiares.

A Paula, pelo amor, amizade e total apoio.

A todos os amigos que sempre estiveram presentes, contribuindo com discussões, críticas e sugestões.

Ao professor Marinho Gomes de Andrade Filho, pela orientação segura, e pelo incentivo durante todo o curso de pós-graduação.

Ao professor Jacek Leśkow, pela orientação durante meu doutorado sanduíche.

E a todos que contribuíram direta ou indiretamente na obtenção deste título.

Abstract

Generalized autoregressive moving average (GARMA) models are a class of models that was developed for extending the univariate Gaussian ARMA time series model to a flexible observation-driven model for non-Gaussian time series data. This work presents the GARMA model with discrete distributions and application of resampling techniques to this class of models. We also proposed the Bayesian approach on GARMA models. The TGARMA (Transformed Generalized Autoregressive Moving Average) models were proposed, using the Box-Cox power transformation. Last but not least we proposed the Bayesian approach for the TGARMA (Transformed Generalized Autoregressive Moving Average).

Keywords: Transformed Generalized ARMA model, Generalized ARMA model, Bayesian approach, discrete distributions, Continuous distributions.

Resumo

Modelos Autoregressivos e de médias móveis generalizados (GARMA) são uma classe de modelos que foi desenvolvida para estender os conhecidos modelos ARMA com distribuição Gaussiana para um cenário de séries temporais não Gaussianas. Este trabalho apresenta os modelos GARMA aplicados a distribuições discretas, e alguns métodos de reamostragem aplicados neste contexto. É proposto neste trabalho uma abordagem Bayesiana para os modelos GARMA. O trabalho da continuidade apresentando os modelos GARMA transformados, utilizando a transformação de Box-Cox. E por último porém não menos importante uma abordagem Bayesiana para os modelos GARMA transformados.

Palavras-chave: ARMA Transformado Generalizado, ARMA Generalizado, abordagem Bayesiana, distribuições discretas, distribuições contínuas.

Contents

List of Figures	xiii
List of Tables	xv
1 Introduction	1
2 GARMA models and moving block bootstrap	5
2.1 Generalized Autoregressive Moving Average (GARMA) model	5
2.1.1 Model Definition	5
2.1.2 Poisson GARMA model	7
2.1.3 Binomial GARMA model	7
2.1.4 Negative Binomial	8
2.1.5 Seasonal Component	8
2.1.6 Maximum Likelihood Estimation and Inference	8
2.1.7 Predictions with GARMA models	9
2.2 Resampling methods	12
2.2.1 Moving Block Bootstrap	12
2.3 Simulation Study	13
2.4 Application to Real Data Sets	16
2.4.1 Dengue Fever Real Data Analysis	17
2.4.2 Monthly Morbidity in São Paulo	21

3	Bayesian GARMA Models for Count Data	25
3.1	Generalized Autoregressive Moving Average Model	25
3.1.1	Poisson GARMA model	26
3.1.2	Binomial GARMA model	26
3.1.3	Negative Binomial	27
3.2	Bayesian Approach on GARMA Models	27
3.2.1	Defining the Prior Densities	27
3.2.2	Bayesian prediction for GARMA models	28
3.2.3	Algorithm used to calculate the $CI_{(1-\delta)}$ for predictions	31
3.3	Simulation Study	32
3.4	Bayesian Real Data Analysis	33
3.4.1	Automobile data set	35
3.4.2	Epidemiology data set	37
3.4.3	Mortality data set	40
4	Transformed GARMA model	43
4.1	TGARMA model	43
4.1.1	Model definition	44
4.1.2	Examples	45
4.1.3	Model Fitting	46
4.2	Moving Block Bootstrap on TGARMA models	48
4.2.1	Forecasting for TGARMA	48
4.3	Simulation Study	49
4.3.1	Bootstrap Simulation Study	54
4.4	Real data analysis	55
5	Bayesian Transformed GARMA Models	61
5.1	Transformed Generalized Autoregressive Moving Average (TGARMA) Model	61
5.1.1	Model definition	62

5.1.2	Examples	63
5.2	Bayesian Approach on TGARMA Models	64
5.2.1	Defining the Prior densities	65
5.2.2	Bayesian prediction for GARMA models	66
5.3	Simulation Study	69
5.4	Real data analysis	72
6	Discussion	77
6.1	Acknowledgments	78
	Bibliography	79

List of Figures

2.1	Estimated densities with Monte Carlo and bootstrap negative binomial	14
2.2	Estimated densities with Monte Carlo and bootstrap binomial	15
2.3	Estimated densities with Monte Carlo and bootstrap Poisson	15
2.4	Graph of Number of Hospitalizations caused by dengue	17
2.5	ACF and PACF for number of hospitalizations caused by dengue . . .	17
2.6	Graph of bootstrap densities of each parameter	19
2.7	Adjusted values versus real values of Number of Hospitalizations caused by dengue	19
2.8	Residual Analysis of Hospitalizations caused by dengue	20
2.9	Predictions with GARMA(1,0) Negative Binomial model with Hospitalizations caused by dengue series	20
2.10	Graph of Morbidity in São Paulo	21
2.11	ACF and PACF for Morbidity in São Paulo	21
2.12	Graph of bootstrap densities of each parameter	23
2.13	Adjusted values versus real values of Monthly Morbidity in São Paulo	23
2.14	Residual Analysis of Monthly Morbidity in São Paulo	24
2.15	Predictions with GARMA(2,0) negative binomial model with Monthly Morbidity in São Paulo	24
3.1	Graph of number of automobile production in Brazil.	35
3.2	Residual Analysis for the number of automobile production in Brazil under a GARMA(1,1) negative binomial model.	37

3.3	Predictions for the number of automobile production in Brazil with a GARMA(1,1) Negative Binomial model.	38
3.4	Number of hospitalizations caused by Dengue Fever.	38
3.5	Residual Analysis of Hospitalizations caused by Dengue.	39
3.6	Predictions with GARMA(1,2) Negative Binomial model with Hospitalizations caused by Dengue series.	40
3.7	Number of deaths in Brazil.	40
3.8	Residual analysis of the number of deaths in Brazil.	41
3.9	Predictions with GARMA(1,0) Binomial model with Number of death in Brazil series.	42
4.1	Densities of parameters with $\lambda = 0.3$	51
4.2	Densities of parameters with $\lambda = 0.5$	52
4.3	Bootstrap densities for TGARMA models	54
4.4	Graph of Annual Swedish fertility rates	56
4.5	Auto Correlation Function and Partial Auto Correlation Function of Annual Swedish fertility rates	56
4.6	λ empirical bootstrapped density Annual Swedish fertility rates	57
4.7	Graph of true values versus the estimated values of the residuals of Annual Swedish fertility rates series	58
4.8	Autocorrelation function and partial autocorrelation function of the residuals of rate of Annual Swedish fertility rates series	59
4.9	Original Predictions with GARMA(1,0) Gamma model with Rate of Annual Swedish fertility rates series	59
5.1	Graph of Annual Swedish fertility rates	73
5.2	Auto Correlation Function and Partial Auto Correlation Function of Annual Swedish fertility rates	73
5.3	Autocorrelation function and partial autocorrelation function of the residuals of rate of Annual Swedish fertility rates series	75
5.4	Original Predictions with GARMA(1,0) Gamma model with Rate of Annual Swedish fertility rates series	75

List of Tables

2.1	Negative Binomial GARMA(1,1) Confidence intervals n=100	14
2.2	Binomial GARMA(1,1) Confidence intervals n=100	15
2.3	Poisson GARMA(1,1) Confidence intervals n=100	16
2.4	Model Selection Criteria using Number of Hospitalizations caused by dengue	18
2.5	Estimates of Hospitalizations caused by dengue series with GARMA(1,0) Negative Binomial	18
2.6	Model Selection Criteria using Monthly Morbidity in São Paulo	22
2.7	Estimates of Monthly Morbidity in São Paulo series with GARMA(2,0) negative binomial	22
3.1	Parameters values to simulate from Negative Binomial GARMA(p,q).	32
3.2	Monte Carlo experiments. Corrected bias, corrected errors and mean acceptance rates for the Bayesian estimation of Negative Binomial GARMA(p,q) model.	34
3.3	Proportions of correct model chosen via Bayesian criteria with Negative Binomial GARMA(p,q) models.	34
3.4	Bayesian selection criteria for the number of automobile production in Brazil.	36
3.5	Estimation results. GARMA(1,1) Negative Binomial model for number of automobile production in Brazil.	36
3.6	Bayesian selection criteria for the number of hospitalizations caused by Dengue Fever.	39

3.7	Estimation results. GARMA(1,2) negative binomial model for the number of hospitalizations caused by Dengue Fever.	39
3.8	Bayesian selection criteria using the number of deaths in Brazil. . . .	41
3.9	Estimates of the number of deaths in Brazil series with GARMA(1,0) Binomial.	41
4.1	Gamma GARMA(p,q) real values	50
4.2	TGARMA(1,1) with gamma distribution and Box-Cox power transformation	51
4.3	Proportions of correct model using BIC with gamma TGARMA(p,q) model	52
4.4	TGARMA(2,2) with Gamma distribution and Box-Cox Power transformation	53
4.5	Bootstrap TGARMA(1,1) Gamma	54
4.6	Coverage Bootstrap TGARMA(1,1) Gamma	55
4.7	Criterion selection using Annual Swedish fertility rates	57
4.8	Estimates of Annual Swedish fertility rates series with TGARMA(1,0) Gamma	58
5.1	TGARMA(1,1) with gamma distribution and Box-Cox power transformation	70
5.2	Proportions of correct model chosen using Bayesian criterion with Gamma GARMA(p,q) model	71
5.3	TGARMA(2,2) with gamma distribution and Box-Cox power transformation	71
5.4	Criterion selection using Annual Swedish fertility rates	74
5.5	Estimates of Annual Swedish fertility rates series with TGARMA(1,0) Gamma	74

Introduction

Counts as observations have attracted significant attention from time series specialists recently. Different contexts present such kind of data: the number of incidences of a certain disease (Poliomyelitis in U.S. [Zeger \(1988\)](#) and Asthma cases in a hospital in Campbelltown, [Davis et al. \(1999\)](#)). Integer financial data sets, like the number of transaction price movements were also discussed (see for example, [Liesenfeld et al. \(2006\)](#) and [Rydberg and Shephard \(2003\)](#)). There are important data sets coming from Brazil like the number of automobile production in Brazil, the number of hospitalizations caused by Dengue Fever and number of death in Brazil that will be analyzed.

Parameter and observation driven models provide a flexible framework for modelling time series of counts. So far, a wide variety of models for count time series have been discussed in literature usually embedded in the framework of integer valued ARMA type models (see for example [Biswas and Song \(2009\)](#)). An overview of these kind of models can be found in [Davis et al. \(1999\)](#) while [Zeger \(1988\)](#) and [Chan and Ledolter \(1995\)](#) explicitly discuss and develop estimation techniques for Poisson generalized linear models with an autoregressive latent process in the mean.

Significant progress has been made in modelling such data. Parameter and observation driven models described by [Cox \(1981\)](#) allow a discrete value approach. Well known extensions of ARMA models to count data are: DARMA and INARMA models of [Jacobs and Lewis \(1978\)](#) and techniques for Poisson

GLM with an autoregressive latent process in the mean in [Davis et al. \(1999\)](#) and [Zeger \(1988\)](#). [Jung et al. \(2006\)](#) compares various models for time series of counts which can account for discreteness, overdispersion and serial correlation. [Zhu \(2010\)](#) proposed a negative binomial INGARCH model applied to polio data. Modelling polio data was also discussed in [Zeger \(1988\)](#).

The focus of this research are GARMA models, Bayesian methods, transformations and applications. The motivation for our work comes from the observation-driven model developed by [Benjamin et al. \(2003\)](#). The real advantage of GARMA models is in providing an exact parametrization of conditional distributions with the help of exponential family. Therefore, in GARMA models, the likelihood can be explicitly expressed for any fixed set of parameter values. Such models can also be used with a variety of time-dependent response variables.

Important aspect of the Chapter 2 is the application of resampling techniques to GARMA models. One of the advantages of the resampling procedures is to improve the evaluation of the confidence intervals and coverage probabilities. Resampling methods in time series context have significant advantage over the classical methods based on asymptotic distributions. In resampling approach confidence intervals are created directly from the sampling distributions of the estimators. The role of resampling is to approximate such sampling distributions. GARMA models are no exception to this rule and we provide both theoretical and real data evidence for that.

The idea of resampling is based on approximating the sampling distribution of an estimator $\hat{\theta}$ via an empirical counterpart based on replicates. In the case of dependent data, one has to carefully introduce resampling techniques while approximating a sampling distribution of $\hat{\theta}$. For a more elaborate study of resampling techniques for nonstationary time series the reader is referred e.g. to [Lešek and Synowiecki \(2010\)](#).

There is an extensive research dedicated to resampling for independent data structures. [Efron \(1979\)](#) introduced the method and presented fundamental results using resampling techniques. Resampling methods for times series have been discussed in the monograph of in [Lahiri \(2003\)](#). In this context one has to mention pioneering work of [Bühlmann \(2002\)](#), [Politis \(2003\)](#) and [Härdle et al. \(2003\)](#) who have published significant results on resampling for dependent data. More recently, results on seasonal and periodic time series resampling

has been developed by [Leškow and Synowiecki \(2010\)](#), [Dehay et al. \(2014\)](#) and [Dudek et al. \(2014a\)](#).

The Chapter 3 extends the work of [Benjamin et al. \(2003\)](#), giving rise to the Bayesian approach on the generalized autoregressive moving average (GARMA) model. This approach presents some gain in terms of estimation, that could be more adequate using different loss functions. The use of Bayesian selection criteria is also an important contribution from this article. Last but not least the application of discrete models on important Brazilian real data providing a new perspective on this field.

In the Chapter 4 we propose using Box-Cox transformations in order to be able to describe our data with GARMA model. Therefore, [de Andrade et al. \(2016b\)](#) proposed model is TGARMA (Transformed Generalized Autoregressive Moving Average). The use of transformations have been shown a good alternative to reduce these kind of problems. [Hamasaki and Kim \(2007\)](#) described a Box and Cox power-transformation to confined and censored nonnormal responses in regression. [da Silva et al. \(2011\)](#) proposed the use of Box-Cox transformation and regression models to deal with fecal egg count data. [Gillard \(2012\)](#) presented a study using Box-Cox family of transformations, the paper comments about problems with asymmetry in the transformed data. [Castillo and F.G. \(2013\)](#) commented about many fields where the Box-Cox transformation can be used, and also proposed a method to improve the forecasting models. [Ahmad et al. \(2015\)](#) combined Box-Cox and bootstrapping idea in one algorithm, the Box-Cox is to ensure the data is normally distributed and bootstrap to deal with small and limited sample size data.

The transformation parameter λ is estimated using the profile likelihood (PL). Moreover, we assess the correctness of the choice of λ using bootstrap method. This method provides confidence intervals for λ . The partial likelihood method was introduced by [Cox \(1975\)](#) and is based entirely on the conditional distribution of the current response, given past responses, and past covariates information and functions thereof can be used for inference. [Zhu and Ghodsi \(2006\)](#) presented a procedure to dimensionality selection maximizing a profile likelihood function. [Huang et al. \(2013\)](#) proposed an efficient equation for estimating the index parameter and unknown link function using adaptive profile-empirical-likelihood inferences. [Cole et al. \(2014\)](#) provide a primer on maximum likelihood, Profile Likelihood and Penalized Likelihood which have proven useful in epidemiologic research.

Bayesian computational methods based on Markov chain Monte Carlo (MCMC) can be utilized to address the complexity of the profile likelihood. Thus, our main contribution concerns inference under the Bayesian framework providing MCMC procedures to evaluate the joint posterior distribution of model parameters. [de Andrade et al. \(2016a\)](#) presented a Bayesian approach for GARMA models, indicating advantages of using the Bayesian methods.

The Chapter 5 extends the TGARMA models, giving rise to the Bayesian approach on the transformed generalized autoregressive moving average (TGARMA) model. This approach presents some gain on the estimate, that could be more adequate using different loss functions. A prior distribution on the parameter λ add information of this parameter, also the posterior density guarantee properties on the transformation parameter. The using of Bayesian selection criterion is also an import gain on this article. Properties of MCMC were used to improve the predictions and construct confidence intervals. Last but certainly not least the application of Swedish fertility rates.

The remainder of this work is organized as follows. Chapter 2 defines the GARMA model with discrete distributions and application of resampling techniques to this class of models. The Bayesian approach on GARMA models are presented in Chapter 3. The Chapter 4 the TGARMA (Transformed Generalized Autoregressive Moving Average) is proposed. The Chapter 5 defines the Bayesian approach for the TGARMA (Transformed Generalized Autoregressive Moving Average). Finally, Chapter 6 gives some concluding remarks.

GARMA models and moving block bootstrap

Abstract

Generalized autoregressive moving average (GARMA) models were developed to extend the univariate Gaussian ARMA time series. In particular, such models can be applied to describe discrete time series. Our paper discusses the Moving Block Bootstrap (MBB) to improve the inference for the parameters of the GARMA model. We provide consistency theorem for MBB applied to GARMA models. Real data related to morbidity caused by external causes considering children younger than 1 year in São Paulo and hospitalizations caused by dengue disease in Ribeirão Preto state of São Paulo in Brazil are analyzed using our approach.

2.1 Generalized Autoregressive Moving Average (GARMA) model

2.1.1 Model Definition

The GARMA model, introduced by [Benjamin et al. \(2003\)](#), specifies the conditional distribution of each observation y_t , for $t = 1, \dots, n$ given the previous

information set $F_{t-1} = (y_1, \dots, y_{t-1}, \mu_1, \dots, \mu_{t-1})$. The conditional density belongs to exponential family and is given by

$$f(y_t|F_{t-1}) = \exp\left(\frac{y_t\alpha_t - b(\alpha_t)}{\varphi} + d(y_t, \varphi)\right), \quad (2.1)$$

where α_t and φ are canonical and scale parameters, respectively. Moreover $b(\cdot)$ and $d(\cdot)$ are specific functions that define the particular exponential family. The conditional mean and conditional variance of y_t given F_{t-1} are represented as:

$$\begin{aligned} \mu_t = b'(\alpha_t) &= E(y_t|F_{t-1}) \\ \text{Var}(y_t|F_{t-1}) &= \varphi b''(\alpha_t), \end{aligned} \quad (2.2)$$

with $t = 1, \dots, n$ and y_1, \dots, y_n are observed.

Following the Generalized Linear Models (GLM) approach the parameter μ_t is related to the predictor η_t by a twice differentiable one-to-one monotonic function g , called *link function*. In general, we can also include the set of covariates \mathbf{x} into our model. Moreover, we can we add an additional component allowing autoregressive moving average terms to be included. In such a case our model will have a form:

$$g(\mu_t) = \eta_t = x_t'\beta + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - x_{t-j}'\beta\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\}. \quad (2.3)$$

The parameters p and q are identified using the classical BIC or AIC criteria. For more information the reader is referred to [Kedem and Fokianos \(2002\)](#) and [Li \(1994\)](#).

The GARMA(p,q) model is defined by the equations (2.1) and (2.3). For certain functions g , it may be necessary to replace y_t with y_t^{new} in (2.3) to avoid the non-existence of $g(y_t)$ for certain values of y_t . The form y_t^{new} depends on the particular function g and is defined for specific cases later.

In our approach we will not include covariates. We will consider three important discrete GARMA models: Poisson, binomial and negative binomial. We will present each one with a density and a respective predictor. The equations (2.2)

are important to relate the interest exponential family distribution to the predictor.

2.1.2 Poisson GARMA model

Suppose that $y_t|F_{t-1}$ follow Poisson distribution with mean μ_t , thus

$$f(y_t|F_{t-1}) = \exp \{y_t \log(\mu_t) - \mu_t - \log(y_t!)\}. \quad (2.4)$$

Here, $Y_t|F_{t-1}$ has a distribution in the exponential family with $\varphi = 1$, $\alpha_t = \log(\mu_t)$, $b(\alpha_t) = \exp(\alpha_t)$, $c(y_t, \varphi) = -\log(y_t!)$ and $\nu(\mu_t) = \mu_t$.

The canonical link function for this model is the logarithmic function, thus the linear predictor is given by

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}^{new}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^{new}) - \log(\mu_{t-j})\}. \quad (2.5)$$

In the equation above $y_{t-j}^{new} = \max(y_{t-j}, c)$, $0 < c < 1$. In sequel we will drop the superscript "new" understanding that we truncate y_t from below we needed. The Poisson GARMA model is defined by the equations (2.4) and (2.5).

2.1.3 Binomial GARMA model

Suppose that $y_t|F_{t-1}$ follows a binomial distribution with the mean μ_t , thus

$$f(y_t|F_{t-1}) = \exp \left\{ y_t \log \left(\frac{\mu_t}{m - \mu_t} \right) + m \log \left(\frac{m - \mu_t}{m} \right) + \log \left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} \right) \right\}. \quad (2.6)$$

The canonical link function for this model is the logarithmic function. The linear predictor is given by

$$\log \left(\frac{\mu_t}{m - \mu_t} \right) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}) - \log(\mu_{t-j})\}, \quad (2.7)$$

where m known.

2.1.4 Negative Binomial

Let y_t a time series such that, $y_t | \mathbf{F}_{t-1} \sim NB(k, \mu_t)$, thus

$$f(y_t | \mathbf{F}_{t-1}) = \exp \left(y_t \log \left\{ \frac{\mu_t}{\mu_t + k} \right\} + k \log \left\{ \frac{k}{\mu_t + k} \right\} + \log \left\{ \frac{\Gamma(k + y_t)}{\Gamma(y_t + 1)\Gamma(k)} \right\} \right), \quad (2.8)$$

that belongs to exponential family with k known

The link function for this model is the logarithmic function

$$\log \left(\frac{\mu_t}{\mu_t + k} \right) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}) - \log(\mu_{t-j})\}, \quad (2.9)$$

2.1.5 Seasonal Component

The seasonal components of the model will be represented as β_{S_1} and β_{S_2} , using functions \cos and \sin respectively. These two terms will be included in the predictor, thus

$$\begin{aligned} \log(\mu_t) = & \beta_0 + \beta_{S_1} \cos \left(\frac{2\pi t}{12} \right) + \beta_{S_2} \sin \left(\frac{2\pi t}{12} \right) + \\ & + \sum_{j=1}^p \phi_j \{\log y_{t-j}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}) - \log(\mu_{t-j})\}. \end{aligned} \quad (2.10)$$

The next subsection contains the Maximum Likelihood Estimation MLE for the GARMA models.

2.1.6 Maximum Likelihood Estimation and Inference

The GARMA model-fitting procedure described herein performs the maximum likelihood estimation (see [Benjamin et al. \(2003\)](#)). The estimation method is based on the standard GLM.

Let $\{y_t\}$ be a time series where the equations (2.1) and (2.3) are satisfied. The parameter vector is $\gamma' = (\beta', \phi', \theta')$, where $\beta = (\beta_0 \dots \beta_m)'$, $\phi = (\phi_1 \dots \phi_p)'$ and also $\theta = (\theta_1 \dots \theta_q)'$. For the estimation procedure the approximated likelihood function in the r first observations is used. $F_r = \{y_1 \dots y_r\}$, where $r = \max(p, q)$. The partial likelihood function can be constructed considering that y_{t-1} and y_t are conditionally independent. Thus

$$\begin{aligned} L(\beta, \phi, \theta | F_n) &\propto \prod_{t=r+1}^n f(y_t | F_t) \\ &\propto \prod_{t=r+1}^n \exp \left\{ \frac{y_t g(\mu_t) - b(g^{-1}(\mu_t))}{\varphi} + c(y_t, \varphi) \right\}, \end{aligned} \quad (2.11)$$

where $g(\mu_t)$ is the link function given by

$$g(\mu_t) = x'_t \beta + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - x'_{t-j} \beta\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - g(\mu_{t-j})\}. \quad (2.12)$$

In the above equations $t = r + 1, \dots, n$. The equations (2.11) and (2.12) do not have a closed form solution therefore a numerical optimization routine will be used.

The GARMA models present interesting asymptotic characteristics, see [Benjamin et al. \(2003\)](#). This theory works well to big data sets, however in discrete real data is common deal with small data sets. The resampling methods can be a solution to asymptotic problems.

2.1.7 Predictions with GARMA models

Prediction plays a key role in time series analysis. The estimate $\hat{\eta}_t$, for $t = r + 1, \dots, n$ are obtained by

$$\hat{\eta}_t = x_t \hat{\beta} + \sum_{j=1}^p \hat{\phi}_j \{g(y_{t-j}) - x_{t-j} \hat{\beta}\} + \sum_{j=1}^q \hat{\theta}_j \{g(y_{t-j}) - \hat{\eta}_{t-j}\} \quad (2.13)$$

Using this equation (2.13), the mean of the process is given by $\hat{\mu}_t = g^{-1}(\hat{\eta}_t)$, for $t = r + 1, \dots, n$.

The prediction of futures values y_{t+h} , for $h > 0$ is given by $\hat{y}_{n+h} = E(Y_{n+h} | F_{n+1})$,

where the information until n were known. And also

$$F_{n+1} = \{x_{n+1}, x_n, \dots, x_1, y_n, y_{n-1}, \dots, y_1, \mu_n, \mu_{n-1}, \dots, \mu_1\}.$$

\hat{y}_{t+h} is called prediction with origin n and horizon h . The prediction for GARMA models is made recursively by the linear predictor of each model. Using the MLE estimate can be obtained

$$\hat{\eta}_{n+1} = x_{n+1}\hat{\beta} + \sum_{j=1}^p \hat{\phi}_j \{g(\hat{y}_{n-j+1}) - x_{n-j+1}\hat{\beta}\} + \sum_{j=1}^q \hat{\theta}_j \{g(\hat{y}_{n-j+1}) - \hat{\eta}_{n-j+1}\}. \quad (2.14)$$

Thus the prediction 1 step ahead ($h = 1$) is calculated by $\hat{y}_{n+1} = g^{-1}(\hat{\eta}_{n+1}) = \mu_{n+1}$, where

$$\begin{cases} \hat{y}_{t-j+h} = y_{t-j+h}, h \leq j \\ \hat{y}_{t-j+h} = g^{-1}(\hat{\eta}_{t-j+h}), h > j \end{cases}$$

For $h > 1$

$$\hat{\eta}_{n+h} = x_{n+h}\hat{\beta} + \sum_{j=1}^p \hat{\phi}_j \{g(\hat{y}_{n-j+h}) - x_{n-j+h}\hat{\beta}\} + \sum_{j=1}^q \hat{\theta}_j \{g(\hat{y}_{n-j+h}) - \hat{\eta}_{n-j+h}\}, \quad (2.15)$$

Confidence intervals for predictions present the idea of more information about the forecast values. In GARMA models, the predictive distribution does not follow a Gaussian density, thus, the usual method based on the asymptotic confidence intervals, evaluated with the unconditional variance, are not the most appropriate. Therefore, the confidence intervals for predictions will be evaluated using quantiles of the estimated distribution. The Algorithm used to calculate the $CI_{(1-\delta)}$ for predictions is the follow:

Confidence Interval Algorithm

STEP 1 Let a sequence of forecast values \hat{y}_{t+h} for $h = 1, \dots, H$.

STEP 2 Take $h = 1, k = 0, y_{t+h}^{(0)} = 0, S_{t+h}^{(0)} = 0$ and also initiate $LB = 0, UB = 0$.

STEP 3 Using the initial values evaluate the equation:

$$f(y_{t+h}^{(k)} | \beta^{(j)}, \Phi^{(j)}, \Theta^{(j)}, \widehat{F}_{t+h}) = \exp \left(\frac{y_{t+h}^{(k)} \alpha_{t+h}^{(j)} - b(\alpha_{t+h}^{(j)})}{\varphi} + d(y_{t+h}^{(k)}, \varphi) \right), \quad (2.16)$$

and also,

$$\widehat{p}(y_{t+h}^{(k)} | \widehat{F}_{t+h}) = \frac{1}{Q} \sum_{j=1}^Q f(y_{t+h}^{(k)} | \beta^{(j)}, \Phi^{(j)}, \Theta^{(j)}, \widehat{F}_{t+h}). \quad (2.17)$$

STEP 4 Using $\widehat{p}(y_{t+h}^{(k)} | \widehat{F}_{t+h})$ compute $S_{t+h}^{(k+1)}$ with

$$S_{t+h}^{(k+1)} = S_{t+h}^{(k)} + \widehat{p}(y_{t+h}^{(k)} | \widehat{F}_{t+h}) \quad (2.18)$$

STEP 5 If $LB = 0$ and $S_{t+h}^{(k+1)} \geq \delta$, $\rightarrow y_{t+h,\delta} = y_{t+h}^{(k)}$ and $LB = 1$.

STEP 6 If $UB = 0$ and $S_{t+h}^{(k+1)} \leq (1 - \delta)$, $\rightarrow y_{t+h,(1-\delta)} = y_{t+h}^{(k)}$ and $UB = 1$.

STEP 7 If $LB = 0$ or $UB = 0$, take $k = k + 1$ and $y_{t+h}^{(k)} = y_{t+h}^{(k-1)} + 1$, repeat steps 3 and 4 until $LB = 1$ and $UB = 1$.

The percentiles $100\delta\%$ and $100(1 - \delta)\%$ are represented by $y_{t+h,\delta}$ and $y_{t+h,(1-\delta)}$ respectively, and given by:

$$y_{t+h,\delta} = \max \left\{ y_{t+h}^{(r)} \mid \sum_{k=1}^r \widehat{p}(y_{t+h}^{(k)} | \widehat{F}_{t+h}) \leq \delta \right\}. \quad (2.19)$$

$$y_{t+h,(1-\delta)} = \min \left\{ y_{t+h}^{(r)} \mid \sum_{k=1}^r \widehat{p}(y_{t+h}^{(k)} | \widehat{F}_{t+h}) \geq (1 - \delta) \right\}. \quad (2.20)$$

The confidence interval $100(1 - \delta)\%$ for the predictions is given by:

$$CI_{(1-\delta)} = [y_{t+h,\delta}; y_{t+h,(1-\delta)}]$$

Due to the results of Benjamin et al. (2003) the asymptotic distribution of the partial likelihood estimator is known. However, the information matrix of the corresponding asymptotic normal law is quite difficult to estimate. This gives us a strong motivation to consider resampling techniques in the GARMA models. As it also known, resampling techniques provide a better alternative to construct

confidence intervals in the time series field. The next section introduces the concept of bootstrap applied in GARMA models.

2.2 Resampling methods

Resampling methods are based on recalculating the value of the estimator on samples that are drawn from the initial sample in a special way. In such a way we obtain an approximation of the distribution of the investigated estimator. This technique in the time series context is computationally intense as one has to many times recalculate complicated algorithms. Therefore, fast computing is a key.

Since [Efron \(1979\)](#) seminal paper, bootstrap algorithms have generated considerable attention. Two main characteristics explain their popularity. Due to bootstrap approach we are able to approximate the sampling distribution of an estimator more effectively than using asymptotic counterparts. This is especially true in time series context, where frequently the asymptotic distribution may have parameters that are not estimable in practice. Resampling methods also allow computing valid asymptotic quantiles of the limiting distribution in different situations, even when the limiting distribution is unknown. One has to bear in mind, however, that before using a given resampling routine we have to prove its consistency. Consistency of resampling means that for increasing sample size the quantiles produced by resampling converge to quantiles of the corresponding asymptotic distribution.

2.2.1 Moving Block Bootstrap

The MBB algorithm was intensively studied, the algorithm and many examples can be found in the monograph of [Lahiri \(2003\)](#). We will present the assumptions to guarantee the consistency of the MBB applied on the GARMA models.

Theorem 4.1 Let the following assumptions be fulfilled:

- A1** $\{y_t\}$ time series that follows GARMA model (2.1) and (2.3) with parameters $\gamma = (\beta, \phi, \theta)$
- A2** $\{y_t\}$ fulfills the α -mixing condition with $\sum_{\tau=1}^{\infty} \alpha(\tau) < \infty$. Moreover the following moment continuity condition is fulfilled
($2 + \delta$ moment condition): There exist $\delta > 0$, $r \in [0, 1 + \delta)$ and nonnegative constants d_1 and d_2 such that considering (2.3):

$$E \left[|Y_t - \mu_t|^{2+\delta} \middle| \mu_t \right] \leq d_1 |\mu_t|^r + d_2$$

Under assumptions **(A1)** and **(A2)** the MBB resampling method is consistent. This means that

$$\sup_x |Pr^* \{ \sqrt{n}(\hat{\gamma}_n^* - \hat{\gamma}_n) \leq x \} - Pr \{ \sqrt{n}(\hat{\gamma}_n - \gamma_n) \leq x \}| \xrightarrow{P} 0 \quad (2.21)$$

In the formula above the inequalities are understood coordinatewise.

Remark: Due to the assumptions of GARMA models we have the existence of all moments. Therefore, the requirement on the rate of convergence of mixing function is weaker than in a general stationary case. The conditional second moment continuity condition is fulfilled e.g. for Poisson GARMA models and binomial GARMA models with fixed m . For general discussion of consistency of conditions for time series the reader is referred to [Dudek et al. \(2014b\)](#) or to monograph of [Lahiri \(2003\)](#).

Proof of the Theorem: We start our proof by noting that any time series $\{y_t\}$ fulfilling the conditions **(A1)** and **(A2)** of GARMA model is stationary in the strict sense. See [Woodard et al. \(2011\)](#) Theorem 14, page 815. Strict stationarity of $\{y_t\}$ together with mixing condition $\sum_{\tau=1}^{\infty} \alpha(\tau) < \infty$ provides for consistency of MBQR procedure: see [Lahiri \(2003\)](#) and also [Dudek et al. \(2014b\)](#) for detailed arguments.

2.3 Simulation Study

The aim of this section is to validate the bootstrap confidence intervals using simulation methods. We present one figure and one table for each model, the figure presents the estimated densities with a Monte Carlo simulation and bootstrap replication.

The objective of this comparison is to show that the bootstrap replicates, present similar densities than the Monte Carlo study. In other words, the Figures [2.1](#), [2.2](#) and [2.3](#) show that in our results the bootstrap densities clearly replicate the original density of the estimates. This result motivates the use of the bootstrap method to withdraw more information about the parameter of interest.

Our study also presents the Tables 2.1, 2.2 and 2.3 where we get the results about the performance of the coverage using Moving Block Bootstrap (MBB) and asymptotic methods. We used 1000 bootstraps replications, 1000 different series.

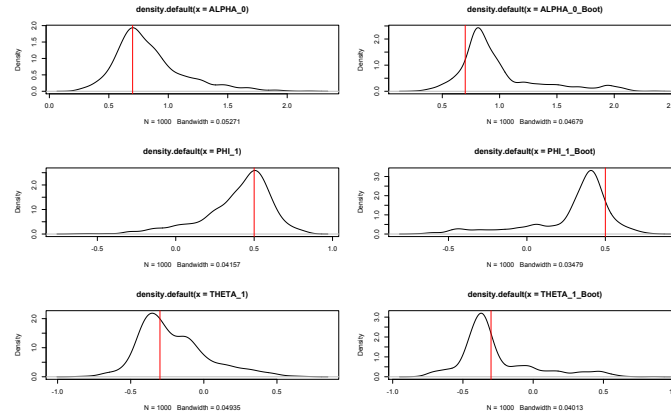


Figure 2.1: Estimated densities with Monte Carlo and bootstrap negative binomial

Table 2.1: Negative Binomial GARMA(1,1) Confidence intervals n=100

Parameter	Bootstrap Inferior Rejection	Bootstrap Coverage	Bootstrap Superior Rejection
β_0	01.20%	98.80%	00.00%
ϕ_1	00.00%	99.20%	00.80%
θ_1	00.40%	99.60%	00.00%
Parameter	Asymptotic Inferior Rejection	Asymptotic Coverage	Asymptotic Superior Rejection
β_0	06.40%	89.70%	03.90%
ϕ_1	03.80%	89.80%	06.40%
θ_1	06.70%	88.10%	05.20%

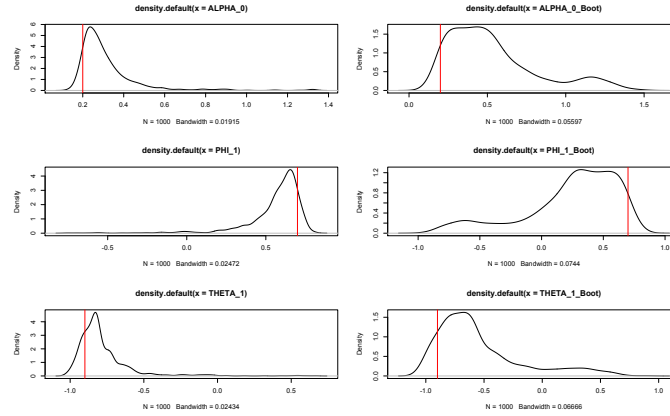


Figure 2.2: Estimated densities with Monte Carlo and bootstrap binomial

Table 2.2: Binomial GARMA(1,1) Confidence intervals n=100

Parameter	Bootstrap Inferior Rejection	Bootstrap Coverage	Bootstrap Superior Rejection
β_0	05.30%	94.70%	00.00%
ϕ_1	00.00%	95.40%	04.60%
θ_1	00.10%	99.90%	00.00%
Parameter	Asymptotic Inferior Rejection	Asymptotic Coverage	Asymptotic Superior Rejection
β_0	09.10%	90.00%	00.90%
ϕ_1	00.90%	90.10%	09.00%
θ_1	04.40%	90.60%	05.00%

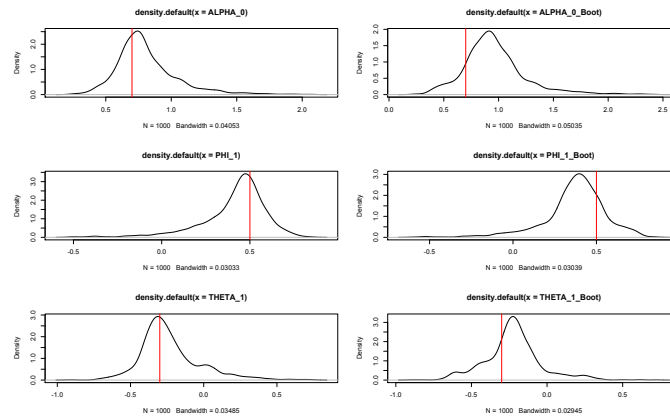


Figure 2.3: Estimated densities with Monte Carlo and bootstrap Poisson

The Tables 2.1, 2.2 and 2.3 present the 95% coverage obtained using two methods. The bootstrap coverage represents the 2.5% and 97.5% values of the bootstrap replicates considering each parameter. The asymptotic coverage

present the values of 95% coverage using the Gaussian confidence intervals, we used the inverse of Fisher matrix to obtain the variances of the estimates and multiplied by -1.96 and 1.96 that represent the respective Gaussian value for 95%.

The results of Tables 2.1, 2.2 and 2.3 clearly show the coverage obtained using the bootstrap confidence interval is higher than the coverage obtained using asymptotic confidence intervals. The results were verified for the binomial, negative binomial and Poisson GARMA models, which motivate the using of bootstrap methods to improve the confidence intervals.

We also present the Figures 2.1, 2.2 and 2.3 showing that the bootstrap replicates densities, in our results, are close than the Monte Carlo densities. In other words, the bootstrap method replicate the origina density of the parameters in the GARMA models, thus the bootstrap confidence intervals should represent a better option than the asymptotic ones.

Table 2.3: Poisson GARMA(1,1) Confidence intervals n=100

Parameter	Bootstrap Inferior Rejection	Bootstrap Coverage	Bootstrap Superior Rejection
β_0	00.10%	99.30%	00.60%
ϕ_1	00.60%	99.40%	00.00%
θ_1	00.00%	99.60%	00.40%
Parameter	Asymptotic Inferior Rejection	Asymptotic Coverage	Asymptotic Superior Rejection
β_0	05.00%	87.40%	07.60%
ϕ_1	09.70%	82.00%	08.30%
θ_1	10.20%	80.80%	09.00%

We can observe the clear improvement of the coverage using the MBB method. The strong asymmetry present on GARMA parameters impairs the quality of the asymptotic coverage. However, the MBB calibrates the confidence intervals to the asymmetric behavior, providing better estimates. There is a extensively discussion about the length of the block. We carried on different sizes of block, and also the random size using the geometric function. In our results we obtained better results following the Lahiri (2003) assumptions about the block length to be between $n^{\frac{1}{4}} \leq b \leq n^{\frac{1}{2}}$.

2.4 Application to Real Data Sets

We analyze 2 different real data sets. One considering the number of hospitalizations and the second considering the Morbidity.

2.4.1 Dengue Fever Real Data Analysis

We analyze the number of hospitalizations caused by dengue in Ribeirão Preto (São Paulo state) between January 2005 to December 2015.

Dengue disease is transmitted by several species of mosquito within the *genus Aedes*, especially *A.aegypti*. The *Aedes* mosquito is easily identifiable by the distinctive black and white stripes on its body. It prefers to breed in a clean and stagnant water. The summer months have higher volumes of rain, thus more clean and stagnant water is available. For Brazil this disease poses a significant public health problem. Therefore, analyzing this data helps healthcare institutions to prepare for possible outbreaks of dengue fever.

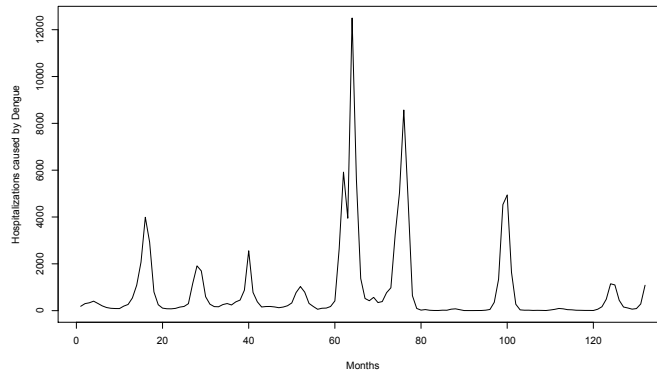


Figure 2.4: Graph of Number of Hospitalizations caused by dengue

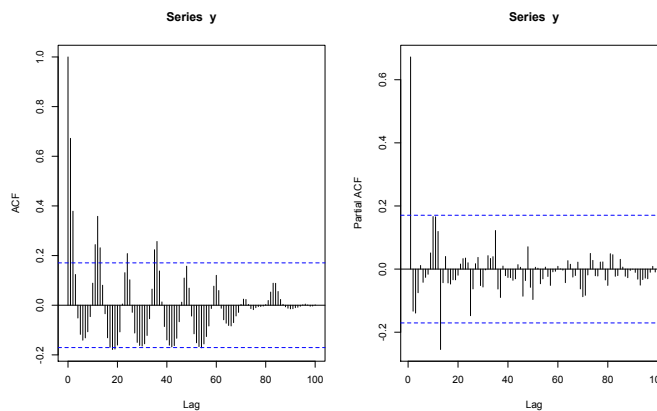


Figure 2.5: ACF and PACF for number of hospitalizations caused by dengue

From Figure 2.4 and Figure 2.5 one clearly sees the seasonal component in the data. Therefore, we will add to the model two seasonal components, considering the period of 12 months. These components improve the model, linking the seasonal behavior to the number of hospitalizations caused by dengue.

The Table 2.4 presented the values of BIC and AIC from different models and orders.

Table 2.4: Model Selection Criteria using Number of Hospitalizations caused by dengue

Poisson	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
BIC	65263.46	68860.81	65784.61	55854.89
AIC	65254.81	68849.27	65773.08	55840.47
Binomial	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
BIC	94143.41	96430.65	96460.43	80731.63
AIC	94134.76	96419.12	96448.87	80717.21
Negative Binomial	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
BIC	3127.27	4260.31	4107.65	3135.64
AIC	3118.62	4248.78	4096.12	3121.22

The extra parameter k was selected trying different values and analyzing the likelihood value, 15 was the chosen value. The Moving Average parameter θ_1 present the 0 inside the confidence intervals, thus we go forward with the GARMA(1,0) model with negative binomial. The Table 2.5 present the MLE estimate and the bootstrap confidence intervals in the GARMA model with the seasonality correction.

Table 2.5: Estimates of Hospitalizations caused by dengue series with GARMA(1,0) Negative Binomial

Parameter	MLE estimate	Lower Bootstrap Bound	Upper Bootstrap Bound
β_0	0.5517	0.5071	2.7321
β_{S_1}	0.6369	0.4288	0.9836
β_{S_2}	0.5763	0.4952	1.2413
ϕ_1	0.9269	0.5787	0.9423

We evaluated the MLE results: $CI_{\beta_0} = (0.3492, 2.7543)$, $CI_{\beta_{S_1}} = (0.3547, 0.9192)$, $CI_{\beta_{S_2}} = (0.5055, 1.6471)$ and $CI_{\phi_1} = (0.4911, 0.9628)$. Comparing with Table 2.5 we can clearly observe that the proposed methodology improved considerably the length of the confidence intervals. The Figure 2.6 presents the graph of estimated densities of the bootstrap samples. All the normality test reject the normality assumption of the samples.

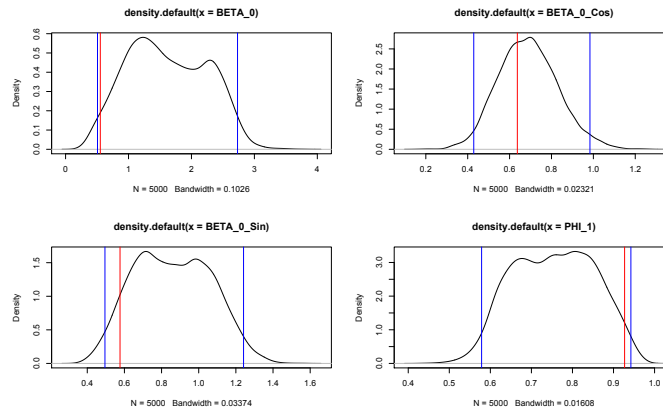


Figure 2.6: Graph of bootstrap densities of each parameter

The Figure 2.7 presents a quantile plot with the true values on x axis and the estimated values on y axis. The line represents the perfect model with real values on axis x and y.

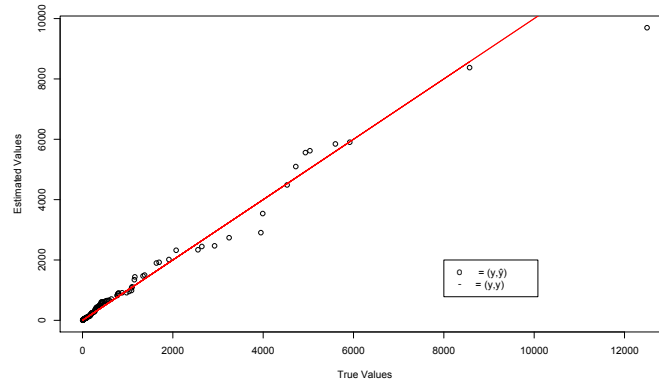


Figure 2.7: Adjusted values versus real values of Number of Hospitalizations caused by dengue

The residual analysis is presented on Figure 2.8 and indicates that the residuals are Non-correlated and Gaussian distributed with Mean=-0.1718 and Sd=1.8947. The Shapiro-Francia test present a p-value of 0.2325 and the Anderson-Darling test present a value 0.3424 which confirm the Gaussian supposition, see [Conover \(1971\)](#).

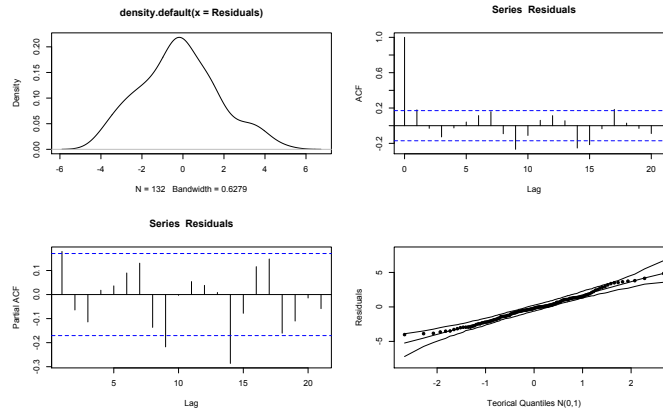


Figure 2.8: Residual Analysis of Hospitalizations caused by dengue

The 8 last values of the series were removed and fitted the model without them. Prediction one step ahead for 8 values was evaluated, thus the predicted value be compared with the true value.

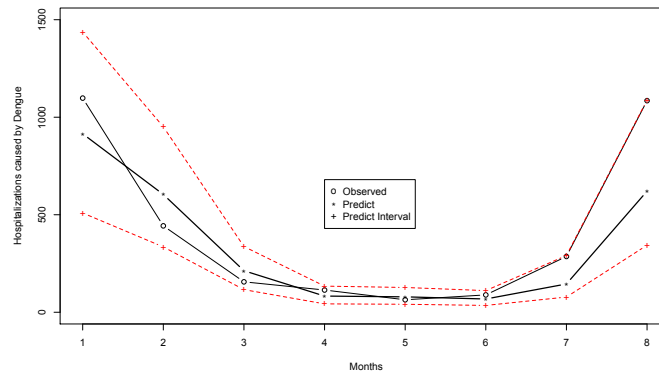


Figure 2.9: Predictions with GARMA(1,0) Negative Binomial model with Hospitalizations caused by dengue series

The predictions are close to the real values, in some cases bigger than the real ones, in other lower than the real one, which indicates good predictions as we can see on Figure 2.9. The confidence intervals for the predictions contain all the real values. The criterion MAPE, see [Chen and Yang \(2004\)](#), was calculated using the predictions given on Figure 2.9. The value was 30.17% which confirmed good predictions.

2.4.2 Monthly Morbidity in São Paulo

We analyze the morbidity caused by external causes considering children younger than 1 year in São Paulo - Brazil. Morbidity represents the incidence or prevalence of a disease or of all diseases, we took observations between January of 1998 to December of 2015, containing 216 observations as we can see in Figure 2.10.

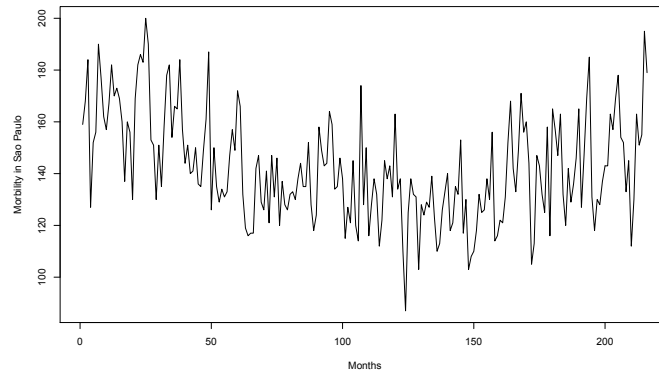


Figure 2.10: Graph of Morbidity in São Paulo

Figure 2.11 presents the autocorrelation function and the partial autocorrelation function of Morbidity in São Paulo.

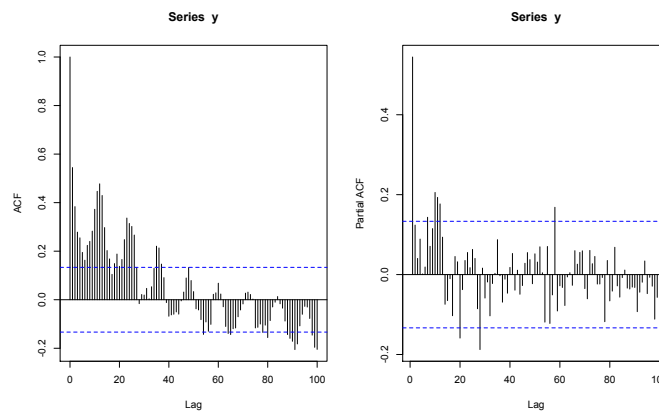


Figure 2.11: ACF and PACF for Morbidity in São Paulo

The morbidity represents considerable part of public spending in Brazil. The main causes of morbidity are aggressions and transportation accidents, considering children younger than 1 year in São Paulo we can analyze this issue

focussed in one of the most important states in Brazil.

We presented in this section three real data sets that will be adjusted in this work. We selected the first and the third one without seasonal component, that allows the using of the traditional moving block bootstrap. While, the second real data about dengue fever, presented seasonal component, requiring the use of periodic moving block bootstrap [Leškow and Synowiecki \(2010\)](#).

The Table 2.6 presented the values of BIC and AIC from different models and orders.

Table 2.6: Model Selection Criteria using Monthly Morbidity in São Paulo

Poisson	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
BIC	3953.91	4425.47	4391.69	4384.94
AIC	3943.79	4411.97	4378.19	4368.07
Binomial	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
BIC	4997.76	5095.27	5212.33	5075.65
AIC	4987.64	5081.77	5198.83	5058.77
Negative Binomial	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
BIC	3822.07	3790.88	3784.90	3791.30
AIC	3811.95	3777.38	3771.40	3777.91

The extra parameter k was selected trying different values and analyzing the likelihood value, 50 was the chosen value. The Moving Average parameter θ_1 also present the 0 inside the confidence intervals, thus we go forward with the GARMA(2,0) model with negative binomial. Table 2.7 present the MLE estimates of the model GARMA(2,0) negative binomial with the respective bootstrap confidence intervals for each parameter.

Table 2.7: Estimates of Monthly Morbidity in São Paulo series with GARMA(2,0) negative binomial

Parameter	MLE estimate	Lower Bootstrap Bound	Upper Bootstrap Bound
β_0	0.8150	0.6958	2.5216
ϕ_1	0.5620	0.3489	0.6764
ϕ_2	0.2743	0.0271	0.4110

We evaluated the MLE confidence intervals: $CI_{\beta_0} = (-0.0561, 1.6863)$, $CI_{\phi_1} = (0.3803, 0.7437)$ and $CI_{\phi_2} = (0.0909, 0.4578)$ Comparing with Table 2.7 we can clearly observe that the proposed methodology improved considerably the length of the confidence intervals. The Figure 2.12 presents the graph of estimated densities of the bootstrap samples. All the normality test reject the normality assumption of the samples.

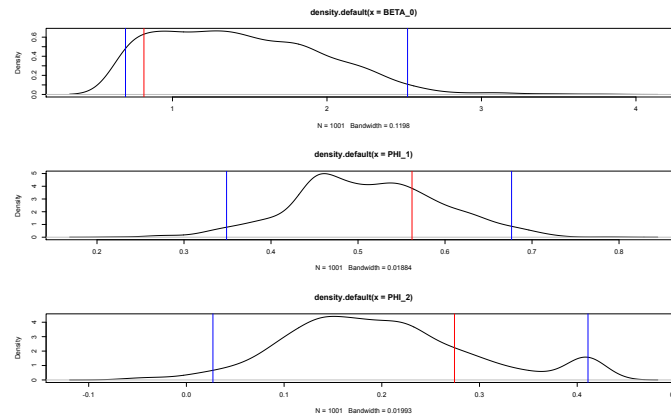


Figure 2.12: Graph of bootstrap densities of each parameter

The Figure 2.13 presents a quantile plot with the true values on x axis and the estimated values on y axis. The line represents the perfect model with real values on axis x and y.

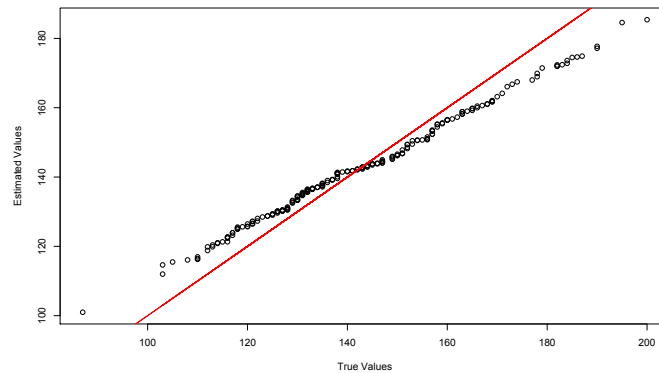


Figure 2.13: Adjusted values versus real values of Monthly Morbidity in São Paulo

The residual analysis presented on Figure 2.14 indicates that the residuals are Non-correlated and Gaussian distributed with Mean=0.0480 and Sd=0.7473. The Kolmogorov-Smirnov test present a p-value of 0.7329 which confirm the Gaussian supposition, see [Conover \(1971\)](#).

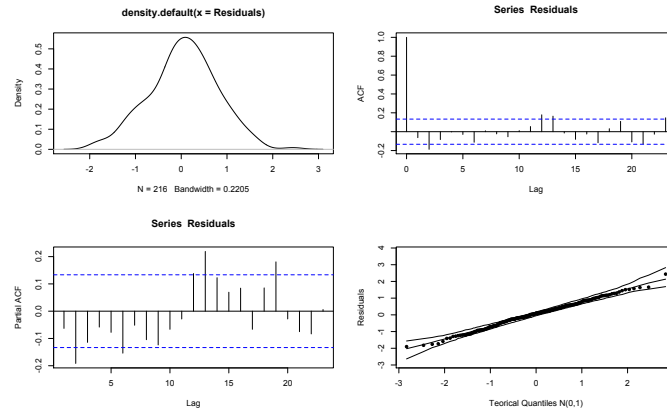


Figure 2.14: Residual Analysis of Monthly Morbidity in São Paulo

The 8 last values of the series were removed and fitted the model without them. Prediction one step ahead for 8 values was evaluated, thus the predicted value be compared with the true value.

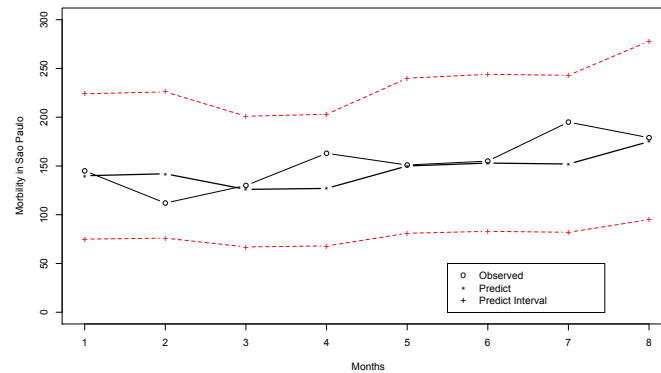


Figure 2.15: Predictions with GARMA(2,0) negative binomial model with Monthly Morbidity in São Paulo

The predictions are close to the real values, in some cases bigger than the real ones, in other lower than the real one, which indicates good predictions as we can see on Figure 2.15. The confidence intervals for the predictions contain all the real values. The criterion MAPE, see [Chen and Yang \(2004\)](#), was calculated using the predictions given on Figure 2.15. The value was 10.20% which confirmed good predictions.

Bayesian GARMA Models for Count Data

Abstract

Generalized autoregressive moving average (GARMA) models are a class of models that were developed for extending the univariate Gaussian ARMA time series model to a flexible observation-driven model for non-Gaussian time series data. This work presents Bayesian approach for GARMA models with Poisson, binomial and negative binomial distributions. A simulation study was carried out to investigate the performance of Bayesian estimation and Bayesian model selection criteria. Also three real datasets were analysed using the Bayesian approach on GARMA models.

3.1 Generalized Autoregressive Moving Average Model

The GARMA model, introduced by Benjamin et al. (2003), assumes that the conditional distribution of each observation y_t , for $t = 1, \dots, n$ given the previous information set $F_{t-1} = (x_1, \dots, x_{t-1}, y_1, \dots, y_{t-1}, \mu_1, \dots, \mu_{t-1})$ belongs to the exponential family. The conditional density is given by,

$$f(y_t|F_{t-1}) = \exp\left(\frac{y_t\alpha_t - b(\alpha_t)}{\varphi} + d(y_t, \varphi)\right), \quad (3.1)$$

where α_t e φ are conical and scale parameter respectively, with $b(\cdot)$ e $d(\cdot)$ being specific functions that define the particular exponential family. The conditional mean and conditional variance of y_t given F_{t-1} is represented by the terms $\mu_t = E(y_t|F_{t-1}) = b'(\alpha_t)$ and $Var(y_t|F_{t-1}) = \varphi b''(\alpha_t)$, with $t = 1, \dots, n$.

Just as in Generalized Linear Models (GLM, [McCullagh and Nelder \(1989\)](#)), μ_t , is related to the linear predictor, η_t , by a twice-differentiable one-to-one monotonic link function $g(\cdot)$. The linear predictor for the GARMA model is given by,

$$g(\mu_t) = \eta_t = x'_t \beta + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - x'_{t-j} \beta\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\}. \quad (3.2)$$

The GARMA (p,q) model is defined by equations (3.1) and (3.2). For certain functions g , it may be necessary to replace y_t with y_t^* in (3.2) to avoid the non-existence of $g(y_t)$ for certain values of y_t . The form y_t^* depends on the particular function $g(\cdot)$ and is defined for specific cases later.

The definition of GARMA model allows to consider the adjust of exogenous variables x'_t however in this work the term $x'_t \beta$ will be considered as a constant β_0 . For count data time series we will consider the following distributions.

3.1.1 Poisson GARMA model

Suppose that $y_t|F_{t-1}$ follows a Poisson distribution with mean μ_t . Then,

$$f(y_t|F_{t-1}) = \exp \{y_t \log(\mu_t) - \mu_t - \log(y_t!)\}. \quad (3.3)$$

and $Y_t|F_{t-1}$ has distribution in the exponential family with $\varphi = 1$, $\alpha_t = \log(\mu_t)$, $b(\alpha_t) = \exp(\alpha_t)$, $c(y_t, \varphi) = -\log(y_t!)$ and $\nu(\mu_t) = \mu_t$. The canonical link function for this model is the logarithmic function, so that the linear predictor is given by,

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}^*\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^*) - \log(\mu_{t-j})\}, \quad (3.4)$$

Where $y_{t-j}^* = \max(y_{t-j}, c)$, $0 < c < 1$. The Poisson GARMA model is defined by equations (3.3) and (3.5).

3.1.2 Binomial GARMA model

Suppose that $y_t|F_{t-1}$ follows a binomial distribution with mean μ_t . Then,

$$f(y_t|F_{t-1}) = \exp \left\{ y_t \log \left(\frac{\mu_t}{m - \mu_t} \right) + m \log \left(\frac{m - \mu_t}{m} \right) + \log \left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} \right) \right\}.$$

The canonical link function for this model is the logarithmic function. The linear predictor is given by,

$$\log\left(\frac{\mu_t}{m - \mu_t}\right) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}^*\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^*) - \log(\mu_{t-j})\}, \quad (3.5)$$

with $y_{t-j}^* = \max(y_{t-j}, c)$, $0 < c < 1$, and m is known.

3.1.3 Negative Binomial

Let y_t a time series such that $y_t | \mathbf{F}_{t-1} \sim NB(k, \mu_t)$. Then,

$$f(y_t | \mathbf{F}_{t-1}) = \exp\left(k \log\left\{\frac{k}{\mu_t + k}\right\} + y_t \log\left\{\frac{\mu_t}{\mu_t + k}\right\} + \log\left\{\frac{\Gamma(k + y_t)}{\Gamma(y_t + 1)\Gamma(k)}\right\}\right),$$

which belongs to the exponential family with k known. The link function for this model is the logarithmic function

$$\log\left(\frac{k}{\mu_t + k}\right) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}^*\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^*) - \log(\mu_{t-j})\},$$

with $y_{t-j}^* = \max(y_{t-j}, c)$, $0 < c < 1$.

3.2 Bayesian Approach on GARMA Models

3.2.1 Defining the Prior Densities

Using the logarithmic in link function to guarantee positive values for any values of the vectors $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$, $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_p)$ and $\boldsymbol{\Theta} = (\theta_1, \dots, \theta_q)$. β , ϕ_i . Thus, a multivariate Gaussian prior will be proposed for each parameter.

$$\begin{aligned} \boldsymbol{\beta} &\sim N(\boldsymbol{\mu}_0, \sigma_0^2 \mathbf{I}_0), \\ \boldsymbol{\Phi} &\sim N(\boldsymbol{\mu}_1, \sigma_1^2 \mathbf{I}_1) \\ \boldsymbol{\Theta} &\sim N(\boldsymbol{\mu}_2, \sigma_2^2 \mathbf{I}_2) \end{aligned}$$

where $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ are vectors with length m , p and q respectively, σ_0^2 , σ_1^2 and σ_2^2 represent the prior variance and \mathbf{I}_0 , \mathbf{I}_1 and \mathbf{I}_2 are $m \times m$, $p \times p$ and $q \times q$ identity matrices respectively. The construction of the multivariate Gaussian depends on

hyper parameters, when there is no prior knowledge on these parameters it can be considered a vary large variance making the prior densities flats. The partial likelihood function for GARMA models can be constructed as follows

$$\begin{aligned} L(\boldsymbol{\beta}, \Phi, \Theta|Y) &\propto \prod_{t=r+1}^n f(y_t|F_{t-1}) \\ &\propto \prod_{t=r+1}^n \exp\left(\frac{y_t\alpha_t - b(\alpha_t)}{\varphi} + d(y_t, \varphi)\right), \end{aligned}$$

where $\alpha_t = g(\mu_t)$, which represent the link function given by

$$g(\mu_t) = x'_t\boldsymbol{\beta} + \sum_{j=1}^p \phi_j\{g(y_{t-j}^*) - x'_{t-j}\} + \sum_{j=1}^q \theta_j\{g(y_{t-j}^*) - g(\mu_{t-j})\},$$

for all $t = r + 1, \dots, n$.

The posterior density is obtained combining the likelihood function with the prior densities. Let the vector $\mathbf{Y} = (y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1, \dots)$ represent the necessary information to construct the likelihood function. The posterior density is then given by,

$$\pi(\boldsymbol{\beta}, \Phi, \Theta|Y) \propto L(\boldsymbol{\beta}, \Phi, \Theta|Y)\pi_0(\boldsymbol{\beta}, \Phi, \Theta). \quad (3.6)$$

However, the joint posterior density of parameters in the GARMA models can not be obtained in closed form. Therefore, Markov chain Monte Carlo (MCMC) sampling strategies will be employed for obtaining samples from this joint posterior distribution. In particular, we use a Metropolis-Hastings algorithm to yield the required realisations. We adopt a sampling scheme where the parameters are updated as o single block and at each iteration we generate new values from a multivariate normal distribution centred around the maximum likelihood estimates with a variance-covariance proposal matrix given by the inverse Hessian evaluated at the posterior mode.

3.2.2 Bayesian prediction for GARMA models

Important aspect of our work is the Bayesian approach on GARMA models, furthermore forecasting futures values. The Bayesian model defined by the equation (3.6) present posterior information obtained via combination of the likelihood function and the prior densities. The goal of forecasting is predict futures

values of the information y_{t+h} , $h \geq 1$, when all the information available is until time t . To evaluate this forecasting is necessary to find the predictive density function $p(y_{t+h}|Y)$.

Denoting the information set $\widehat{F}_{t+h} = (\widehat{x}_{t+h}, \dots, x_t, x_{t-1}, \dots, \widehat{y}_{t+h-1}, \dots, y_t, y_{t-1}, \dots, \widehat{\mu}_{t+h-1}, \dots, \mu_t, \mu_{t-1}, \dots)$, where $\widehat{y}_{t+h-i} = y_{t+h-i}$, if $h \leq i$, else $\widehat{y}_{t+h-i} = E\{y_{t+h-i}|\widehat{F}_{t+h-i}\}$, $i = 1, 2, \dots, h+1$. The general idea is that \widehat{F}_{t+h} contains all the data observed until the time t , for the future time $t+h$, $h \geq 1$, the set \widehat{F}_{t+h} is completed with forecasts of necessary information to estimate y_{t+h} . Starting with,

$$f(y_{t+h}|\beta, \Phi, \Theta, \widehat{F}_{t+h}) = \exp\left(\frac{y_{t+h}\alpha_{t+h} - b(\alpha_{t+h})}{\varphi} + d(y_{t+h}, \varphi)\right), \quad (3.7)$$

The conditional mean and variance of y_{t+h} given \widehat{F}_{t+h} is represented by the terms $\widehat{\mu}_{t+h} = E(y_{t+h}|\widehat{F}_{t+h}) = b'(\alpha_{t+h})$ and $Var(y_{t+h}|\widehat{F}_{t+h}) = \varphi b''(\alpha_{t+h})$. The μ_{t+h} , is related to the predictor, η_{t+h} , by a twice-differentiable one-to-one monotonic link function $g(\cdot)$. The linear predictor for the GARMA model is given by,

$$g(\mu_{t+h}) = \eta_{t+h} = \widehat{x}'_{t+h}\beta + \sum_{j=1}^p \phi_j \{g(\widehat{y}_{t+h-j}) - \widehat{x}'_{t+h-j}\beta\} + \sum_{j=1}^q \theta_j \{g(\widehat{y}_{t+h-j}) - \widehat{\eta}_{t+h-j}\}. \quad (3.8)$$

With the equation (3.7) and posterior density (3.6), the predictive density for y_{t+h} can be written as,

$$p(y_{t+h}|\widehat{F}_{t+h}) = \int_{\{\beta, \Phi, \Theta\} \in \Omega} f(y_{t+h}|\beta, \Phi, \Theta, \widehat{F}_{t+h}) \pi(\beta, \Phi, \Theta|Y) d\beta d\Phi d\Theta.$$

The aim is determine the predictive density using the MCMC algorithm, thus

$$\widehat{p}(y_{t+h}|\widehat{F}_{t+h}) = \frac{1}{Q} \sum_{j=1}^Q f(y_{t+h}|\beta^{(j)}, \Phi^{(j)}, \Theta^{(j)}, \widehat{F}_{t+h}). \quad (3.9)$$

Given the predictive density, the next step is to evaluate the prediction $E(y_{t+h}|\widehat{F}_{t+h}) = \widehat{y}_{t+h}$.

$$E(y_{t+h}|\widehat{F}_{t+h}) = \int_{y_{t+h} \in R} y_{t+h} p(y_{t+h}|\widehat{F}_{t+h}) dy_{t+h}. \quad (3.10)$$

Substituting the equation (3.9) the equation (3.10) can be rewritten by,

$$E(y_{t+h}|\widehat{\mathbf{F}}_{t+h}) = \int_{y_{t+h} \in R} y_{t+h} \left[\int_{\{\beta, \Phi, \Theta\} \in \Omega} f(y_{t+h}|\beta, \Phi, \Theta, \widehat{\mathbf{F}}_{t+h}) \pi(\beta, \Phi, \Theta|Y) d\beta d\Phi d\Theta \right] dy_{t+h}.$$

Using properties of integer, we can rewrite (3.11) as,

$$E(y_{t+h}|\widehat{\mathbf{F}}_{t+h}) = \int_{\{\beta, \Phi, \Theta\} \in \Omega} \left[\int_{y_{t+h} \in R} y_{t+h} f(y_{t+h}|\beta, \Phi, \Theta, \widehat{\mathbf{F}}_{t+h}) dy_{t+h} \right] \pi(\beta, \Phi, \Theta|Y) d\beta d\Phi d\Theta.$$

The equation (3.11) represent

$$E(y_{t+h}|\widehat{\mathbf{F}}_{t+h}) = \int_{\{\beta, \Phi, \Theta\} \in \Omega} \left[E(y_{t+h}|\beta, \Phi, \Theta, \widehat{\mathbf{F}}_{t+h}) \right] \pi(\beta, \Phi, \Theta|Y) d\beta d\Phi d\Theta. \quad (3.11)$$

Denoting by $\mu_{t+h}(\beta, \Phi, \Theta, \widehat{\mathbf{F}}_{t+h}) = E(y_{t+h}|\beta, \Phi, \Theta, \widehat{\mathbf{F}}_{t+h})$. Hence, using the MCMC vector $(\beta^{(j)}, \Phi^{(j)}, \Theta^{(j)})$, $j = 1, 2, \dots, Q$, the $E(y_{t+h}|\widehat{\mathbf{F}}_{t+h})$ can be estimated by

$$\widehat{y}_{t+h} = \frac{1}{Q} \sum_{k=1}^Q \mu_{t+h}(\beta^{(k)}, \Phi^{(k)}, \Theta^{(k)}, \widehat{\mathbf{F}}_{t+h}), \quad (3.12)$$

where

$$g(\mu_{t+h}^{(k)}) = \widehat{x}'_{t+h} \beta^{(k)} + \sum_{j=1}^p \phi_j^{(k)} \{g(\widehat{y}_{t+h-j}) - \widehat{x}'_{t+h-j} \beta^{(k)}\} + \sum_{j=1}^q \theta_j^{(k)} \{g(\widehat{y}_{t+h-j}) - \widehat{\eta}_{t+h-j}^{(k)}\}. \quad (3.13)$$

Credible intervals for \widehat{y}_{t+h} can be calculated using the $100\alpha\%$, and $100(1 - \alpha)\%$ quantiles of the MCMC sample $\mu_{t+h}^{(k)}$, with $k = 1, \dots, Q$. An approach to estimate the credible interval of \widehat{y}_{t+h} is the Highest Posterior Density (HPD), see [Chen and Shao \(1998\)](#).

A $100(1 - \alpha)\%$ HPD region for \hat{y}_{t+h} are a subset $C \in R$ defined by $C = \{y_{t+h} : p(y_{t+h} | \hat{F}_{t+h}) \geq \kappa\}$, where κ is the largest number such that

$$\int_{y_{t+h} \geq \kappa} p(y_{t+h} | \hat{F}_{t+h}) dy_{t+h} = 1 - \alpha. \quad (3.14)$$

We can use the $\hat{p}(y_{t+h} | \hat{F}_{t+h})$ MCMC estimates, given by the equation (3.9), to estimate the $100(1 - \alpha)\%$ HPD region. The next section contains all the Bayesian simulation study. Metrics were used to verify the quality and performance of the adjust.

3.2.3 Algorithm used to calculate the $CI_{(1-\delta)}$ for predictions

1. Let a sequence of forecast values \hat{y}_{t+h} for $h = 1, \dots, H$.
2. Take $h = 1, k = 0, y_{t+h}^{(0)} = 0, S_{t+h}^{(0)} = 0$ and also initiate $LB = 0, UB = 0$.
3. Using the initial values evaluate the equation:

$$f(y_{t+h}^{(k)} | \beta^{(j)}, \Phi^{(j)}, \Theta^{(j)}, \hat{F}_{t+h}) = \exp \left(\frac{y_{t+h}^{(k)} \alpha_{t+h}^{(j)} - b(\alpha_{t+h}^{(j)})}{\varphi} + d(y_{t+h}^{(k)}, \varphi) \right),$$

and also,

$$\hat{p}(y_{t+h}^{(k)} | \hat{F}_{t+h}) = \frac{1}{Q} \sum_{j=1}^Q f(y_{t+h}^{(k)} | \beta^{(j)}, \Phi^{(j)}, \Theta^{(j)}, \hat{F}_{t+h}).$$

4. Using $\hat{p}(y_{t+h}^{(k)} | \hat{F}_{t+h})$ compute $S_{t+h}^{(k+1)}$ with

$$S_{t+h}^{(k+1)} = S_{t+h}^{(k)} + \hat{p}(y_{t+h}^{(k)} | \hat{F}_{t+h})$$

5. If $LB = 0$ and $S_{t+h}^{(k+1)} \geq \delta$, $\rightarrow y_{t+h,\delta} = y_{t+h}^{(k)}$ and $LB = 1$.
6. If $UB = 0$ and $S_{t+h}^{(k+1)} \leq (1 - \delta)$, $\rightarrow y_{t+h,(1-\delta)} = y_{t+h}^{(k)}$ and $UB = 1$.
7. If $LB = 0$ or $UB = 0$, take $k = k + 1$ and $y_{t+h}^{(k)} = y_{t+h}^{(k-1)} + 1$, repeat steps 3 and 4 until $LB = 1$ and $UB = 1$.

The percentiles $100\delta\%$ and $100(1 - \delta)\%$ are represented by $y_{t+h,\delta}$ and $y_{t+h,(1-\delta)}$ respectively, and given by:

$$y_{t+h,\delta} = \max \left\{ y_{t+h}^{(r)} \mid \sum_{k=1}^r \widehat{p}(y_{t+h}^{(k)} \mid \widehat{F}_{t+h}) \leq \delta \right\}.$$

$$y_{t+h,(1-\delta)} = \min \left\{ y_{t+h}^{(r)} \mid \sum_{k=1}^r \widehat{p}(y_{t+h}^{(k)} \mid \widehat{F}_{t+h}) \geq (1 - \delta) \right\}.$$

The confidence interval $100(1 - \delta)\%$ for the predictions is given by:

$$CI_{(1-\delta)} = [y_{t+h,\delta}; y_{t+h,(1-\delta)}]$$

The next section contains all the Bayesian simulation study. Metrics were used to verify the quality and performance of the adjust.

3.3 Simulation Study

In this section we conduct a simulation study for negative binomial GARMA(p, q) models with different orders p and q . The actual parameter values used to simulate the artificial series are shown in Table 3.1 and the parameter k of the negative binomial was fixed at $k = 15$. These values were chosen taking into account that a GARMA model can be nonstationary since they are in the exponential family and the variance function depends on the mean. So, we opted to chose parameter values that would generate moderate values for the time series. The experiment was replicated $m = 1000$ times for each model. For each dataset we used the prior distributions as described in Section 5.2 with mean zero and variance 200. We then drew samples from the posterior distribution discarding the first 1000 draws as burn-in and keeping every 3rd sampled value resulting in a final sample of 5000 values. All the computations were implemented using the open-source statistical software language and environment R [R Development Core Team \(2010\)](#).

Table 3.1: Parameters values to simulate from Negative Binomial GARMA(p, q).

Order	β_0	ϕ_1	ϕ_2	θ_1	θ_2
(1,1)	0.80	0.50	-	0.30	-
(1,2)	1.00	0.30	-	0.40	0.25
(2,1)	0.55	0.30	0.40	0.20	-
(2,2)	0.65	0.30	0.40	0.25	0.35

The performance of the Bayesian estimation was evaluated using three metrics: the corrected bias (CB), the corrected error (CE) and the mean

acceptance rates in the MCMC algorithm called Acceptance Probabilities (AP). These metrics are defined as,

$$\begin{aligned}
 CB &= \frac{1}{m} \sum_{i=1}^m \left| \frac{\theta - \hat{\theta}^{(i)}}{\theta} \right|, \\
 CE^2 &= \frac{1}{Var} \frac{1}{m} \sum_{i=1}^m (\hat{\theta}^{(i)} - \theta)^2 \\
 AP &= \frac{1}{m} \sum_{i=1}^m \hat{r}^{(i)},
 \end{aligned}$$

where $\hat{\theta}^{(i)}$ and $\hat{r}^{(i)}$ are the estimate of parameter θ and the computed acceptance rate respectively for the i -th replication, $i = 1, \dots, m$. In this paper we take the posterior means of θ as point estimates. Also, the variance term (Var) that appears in the definition of CE is the sample variance of $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(m)}$.

The estimation results appear in Table 3.2 where the posterior mean and variance (in brackets) as well as the aforementioned metrics are shown for each model and parameter. These results indicate good properties with relatively small values of the corrected bias (CB), values of the corrected error (CE) around 1 and acceptance probabilities between 0.20 and 0.70.

We also include Table 3.3 with the proportions of correct model choice using three popular Bayesian model selection criteria. Specifically, we adopt the expected Bayesian information criterion (EBIC, [Carlin and Louis \(2001\)](#)), the Deviance information criterion (DIC, [Spiegelhalter et al. \(2002\)](#)) and the conditional predictive ordinate (CPO, [Gelfand et al. \(1992\)](#)) to select the order of the GARMA models. Each column in this table contains the model order and the associated proportions of correct model choice according to EBIC, DIC and CPO criteria. Higher proportions of correct model choices are observed as the sample sizes increase for all models and criteria. Also, EBIC and CPO tend to perform better for GARMA(1,1) and GARMA(1,2) models but none performed particularly well with GARMA(2,2) models.

Finally, this simulation study was carried out also for the Poisson and binomial distributions with results similar to the ones shown. These results are not included to save space.

3.4 Bayesian Real Data Analysis

In this section, we apply the methodology described so far to three real time series of count data. For each series we estimated GARMA(p, q) models with

Table 3.2: Monte Carlo experiments. Corrected bias, corrected errors and mean acceptance rates for the Bayesian estimation of Negative Binomial GARMA(p,q) model.

Parameter	Mean(Var)(1,1)	CB(1,1)	CE(1,1)	AP(1,1)	Mean(Var)(1,2)	CB(1,2)	CE(1,2)	AP(1,2)
β_0	0.8571(0.0065)	0.0984	1.2247	0.3746	1.0823(0.0196)	0.1276	1.1592	0.3182
ϕ_1	0.4695(0.0026)	0.0947	1.1637	0.3511	0.2554(0.0097)	0.2820	1.0965	0.2702
ϕ_2	-	-	-	-	-	-	-	-
θ_1	0.2927(0.0033)	0.1531	1.0071	0.6480	0.4099(0.0091)	0.1900	1.0048	0.4327
θ_2	-	-	-	-	0.2478(0.0037)	0.1929	1.0001	0.5882
Parameter	Mean(Var)(2,1)	CB(2,1)	CE(2,1)	AP(2,1)	Mean(Var)(2,2)	CB(2,2)	CE(2,2)	AP(2,2)
β_0	0.6198(0.0097)	0.1740	1.2240	0.2786	0.7344(0.0079)	0.1497	1.3171	0.3397
ϕ_1	0.2798(0.0152)	0.3295	1.0127	0.1422	0.2887(0.0054)	0.1959	1.0111	0.2282
ϕ_2	0.3794(0.0066)	0.1661	1.0307	0.2091	0.3414(0.0049)	0.1485	1.0787	0.2348
θ_1	0.2012(0.0182)	0.5334	0.9995	0.3214	0.2430(0.0052)	0.2307	1.0040	0.5237
θ_2	-	-	-	-	0.3464(0.0027)	0.1193	1.0017	0.6614

Table 3.3: Proportions of correct model chosen via Bayesian criteria with Negative Binomial GARMA(p,q) models.

EBIC				
Size	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
200	0.9379	0.3042	0.5626	0.4450
500	0.9799	0.6156	0.8048	0.5825
1000	0.9852	0.9039	0.8471	0.6772
DIC				
Size	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
200	0.6316	0.4804	0.5445	0.4437
500	0.6876	0.6476	0.6221	0.4925
1000	0.7155	0.7364	0.6469	0.7154
CPO				
Size	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
200	0.8078	0.3493	0.5575	0.4112
500	0.8188	0.5925	0.5993	0.4625
1000	0.8325	0.7266	0.6152	0.7317

varying orders and computed the Bayesian selection criteria EBIC, DIC and CPO for model comparison. In all cases we used the diagnostic proposed by Geweke (1992) to assess convergence of the chains. This is based on a test for equality of the means of the first and last part of the chain (by default the first 10% and the last 50%). If the samples are drawn from the stationary distribution, the two means are equal and the statistic has an asymptotically standard normal distribution. The calculated values of Geweke statistics were all between -2 and 2 which is an indication of convergence of the Markov chains.

3.4.1 Automobile data set

The first real data set analysed is the number of automobile production in Brazil between January 1993 and December 2013. The data is available from <http://www.anfavea.com.br/tabelas.html>. The original observations were divided by 1000 to reduce the magnitude of the data.

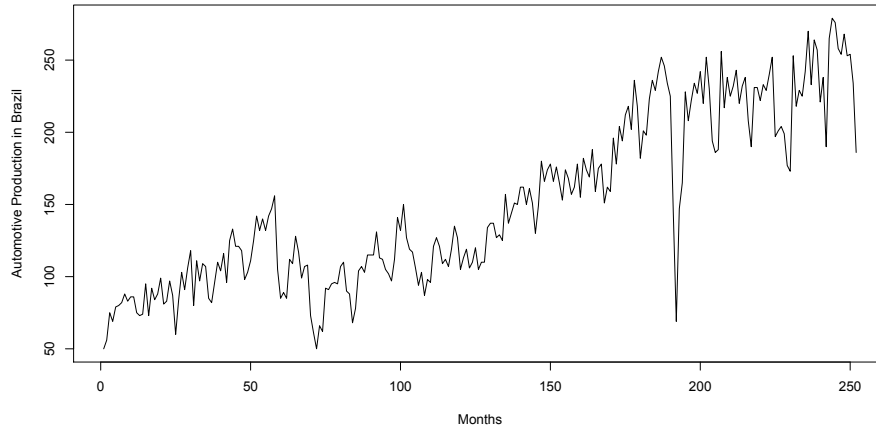


Figure 3.1: Graph of number of automobile production in Brazil.

The automotive industry is extremely important as it can influence other industries activities. For example, 50% of the world rubber production, 25% of the world glass production and 15% of the world iron production are destined to the automotive industry. The behaviour of the data along time depicted in Figure 3.1 seems to indicate that an extra term should be included to take into account a (possibly nonlinear) trend. The term $\beta_{\text{exp}} = \log(t)$ was then included in the model equation to account for this long-term increase.

The results regarding selection criteria are summarized in Table 3.4. We note that the three criteria indicate that the most appropriate model was the GARMA(1,1) Negative Binomial. Also, Table 3.5 presents the estimation results for the selected GARMA(1,1) Negative Binomial model with the extra parameter fixed at $k = 150$.

We also performed a residual analysis based on the so called quantile residuals which are the common choice for generalized linear models. In fact, quantile residuals are the only useful residuals for binomial, negative binomial or Poisson data when the response takes on only a small number of distinct values (Dunn and Smyth (1996)). These are given by $r_t = \Phi^{-1}(\mathbf{F}_{y_t}(y_t|F_{t-1}))$ where \mathbf{F}_{y_t} represent the cumulative distribution function of the associated discrete

Table 3.4: Bayesian selection criteria for the number of automobile production in Brazil.

Poisson	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	3046.61	3074.24	3045.70	3074.45	3071.21	3067.97
DIC	3032.06	3064.97	3030.38	3064.55	3046.02	3065.89
CPO	-1519.88	-1536.12	-1519.65	-1535.15	-1536.76	-1540.29
Binomial	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	3559.79	3814.33	3559.11	3813.91	3759.32	3738.38
DIC	3545.12	3736.57	3544.19	3794.01	3738.90	3713.19
CPO	-1782.36	-1930.30	-1780.67	-1949.77	-1929.13	-1909.67
Negative Binomial	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	2547.67	2792.38	2546.76	2799.21	2787.56	2785.10
DIC	2537.71	2777.16	2531.85	2779.28	2767.32	2760.13
CPO	-1269.48	-1427.72	-1267.34	-1430.66	-1426.47	-1423.09

Table 3.5: Estimation results. GARMA(1,1) Negative Binomial model for number of automobile production in Brazil.

Parameter	Mean	Variance	HPD Credible Interval	AP
β_0	0.3834	0.0006	(0.3543; 0.4159)	0.3710
β_{exp}	0.0850	0.0002	(0.0814; 0.0884)	0.3163
ϕ_1	0.8447	0.0005	(0.8379; 0.8521)	0.3038
θ_1	0.1149	0.0005	(0.1064; 0.1244)	0.6323

distribution. In practice, when dealing with discrete distributions we need to introduce some randomization to produce continuous normal residuals. The residual analysis summarized in Figure 3.2 which indicates that the residuals are non-correlated and Gaussian distributed with mean 0.0767 and standard deviation 1.2295. Kolmogorov-Smirnov and Lilliefors normality tests returned p -values of 0.4502 and 0.0743 respectively which provides evidence for Gaussian assumption (Conover (1999)).

Finally, we performed a prediction exercise using the last 9 observations of the original series as follows. For each $k = 1, \dots, 9$ the GARMA(1,1) negative binomial model was fitted to the series y_1, \dots, y_{n-k} and an out-of-sample one-step ahead prediction \hat{y}_{n-k+1} was produced. These predictions can then be compared with the true values. The results are illustrated in Figure 3.3 from which we can see that the prediction errors are overall small. A formal comparison was made by calculating the mean absolute percentage error (MAPE, Hyndman (2006)) and we obtained the value 6.07%.

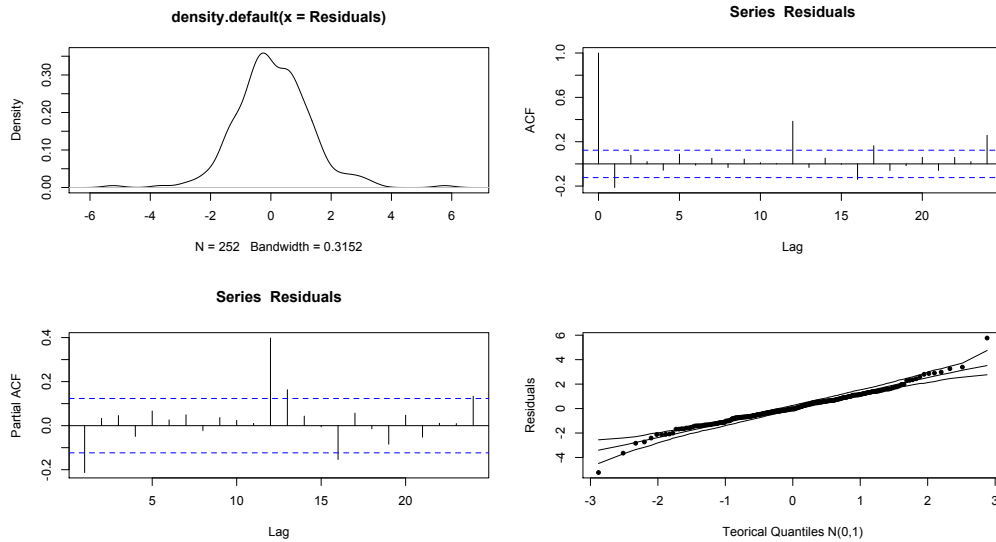


Figure 3.2: Residual Analysis for the number of automobile production in Brazil under a GARMA(1,1) negative binomial model.

3.4.2 Epidemiology data set

This real data set comprises the number of hospitalizations caused by Dengue Fever in Campina Grande city (Brazil) between January 1998 and October 2003. Dengue Fever is transmitted by several species of mosquito within the *genus Aedes*, principally *A. aegypti*. The *Aedes* mosquito is easily identifiable by the distinctive black and white stripes on its body. It prefers to lay eggs on clean and stagnant water. Analysing the autocorrelation function of this data, a seasonal behaviour is characterised. This is because the Summer months in this region present higher volume of rain, thus leading to more clean and stagnant water. Figure 3.4 clearly sees the seasonal component in the data, therefore we included two seasonal components in the model, β_{S_1} and β_{S_2} , using *cosine* and *sine* functions respectively, and also considering the period of 12 months. These components are expected to improve model estimation.

The results regarding the selection criteria are summarized in Table 3.6 from which we can conclude that the most appropriate model was the GARMA(1,2) Negative Binomial. Note that the three criteria gave the same indication. Table 3.7 shows the estimation results for the selected GARMA(1,2) Negative Binomial model with the extra parameter fixed at $k = 30$.

Again we performed a residual analysis based on quantile residuals. This is summarized in Figure 3.5 which indicates that the residuals are non-correlated and Gaussian distributed with mean 0.0258 and standard deviation 1.5571.

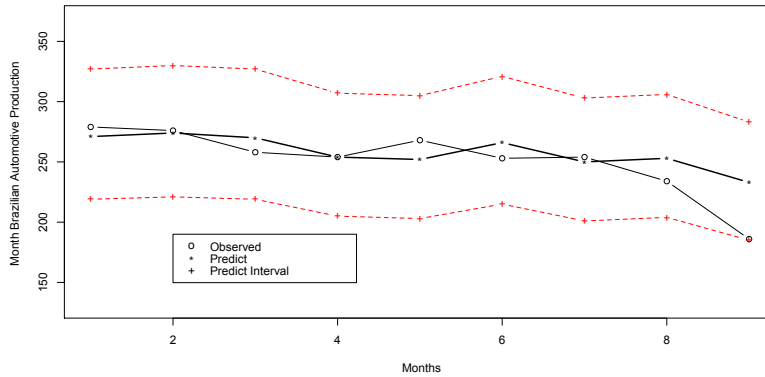


Figure 3.3: Predictions for the number of automobile production in Brazil with a GARMA(1,1) Negative Binomial model.

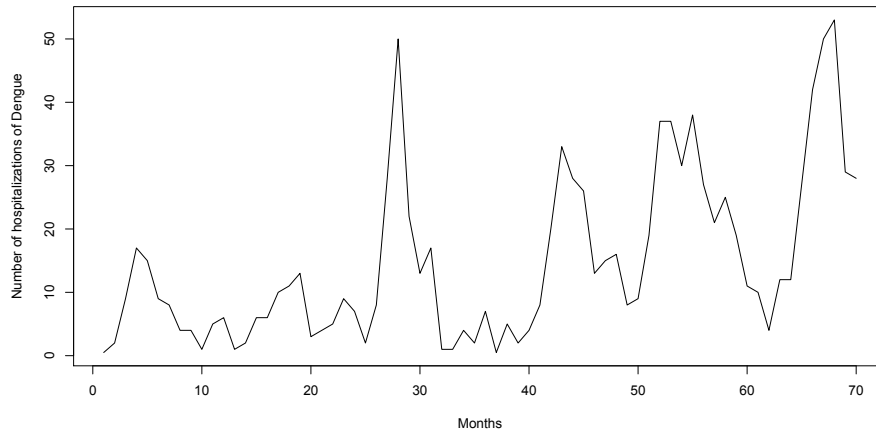


Figure 3.4: Number of hospitalizations caused by Dengue Fever.

The Kolmogorov-Smirnov and Shapiro-Wilk normality tests returned p -values of 0.4856 and 0.1176 respectively thus giving evidence for the Gaussian assumption.

A similar prediction exercise was performed for this data. So, we fitted a GARMA(1,2) negative binomial model to y_1, \dots, y_{n-k} and computed an out-of-sample one-step ahead prediction \hat{y}_{n-k+1} for $k = 1, \dots, 9$. Figure 3.6 shows the predictions, prediction intervals and the real observations for comparison. It can be seen that, although relatively close to the actual values, predictions for May, June, July and August 2003 are consistently below the observations. The MAPE criterion was calculated as 47.81%.

Table 3.6: Bayesian selection criteria for the number of hospitalizations caused by Dengue Fever.

Poisson	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	632.82	633.13	633.73	632.48	632.65	628.20
DIC	580.66	581.04	581.86	581.25	580.32	578.31
CPO	-794.03	-794.87	-794.69	-794.11	-793.83	-792.34
Binomial	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	690.62	689.28	690.34	656.56	688.82	655.30
DIC	679.14	679.92	679.42	642.12	674.83	637.19
CPO	-345.89	-346.16	-345.76	-327.13	-348.04	-324.83
Negative Binomial	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	507.89	508.97	509.36	504.12	509.09	505.89
DIC	519.66	520.93	520.22	518.30	523.11	523.19
CPO	-256.35	-255.88	-256.10	-254.24	-257.64	-256.26

Table 3.7: Estimation results. GARMA(1,2) negative binomial model for the number of hospitalizations caused by Dengue Fever.

Parameter	Mean	Variance	HPD Credible Interval	AP
β_0	1.1916	0.0566	(0.7443; 1.6068)	0.1090
β_{S_1}	-0.2571	0.0035	(-0.3753;-0.1407)	0.6196
β_{S_2}	0.1424	0.0040	(0.0156; 0.2649)	0.5858
ϕ_1	0.5796	0.0078	(0.4230; 0.7456)	0.0968
θ_1	0.1214	0.0112	(-0.0853; 0.3273)	0.3391
θ_2	0.0987	0.0053	(-0.0470; 0.2358)	0.3978

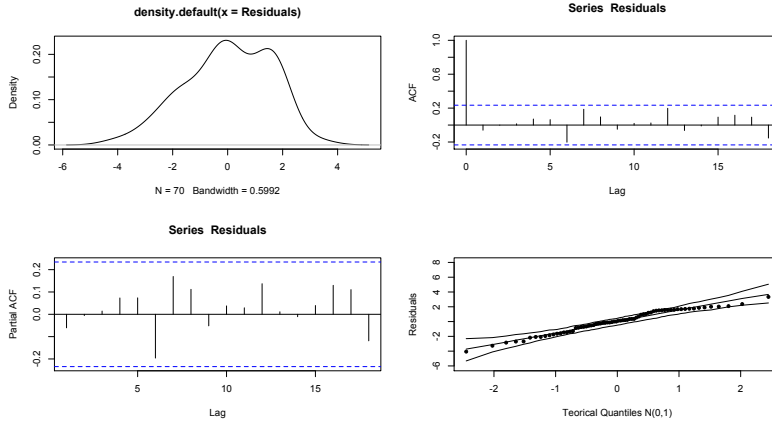


Figure 3.5: Residual Analysis of Hospitalizations caused by Dengue.

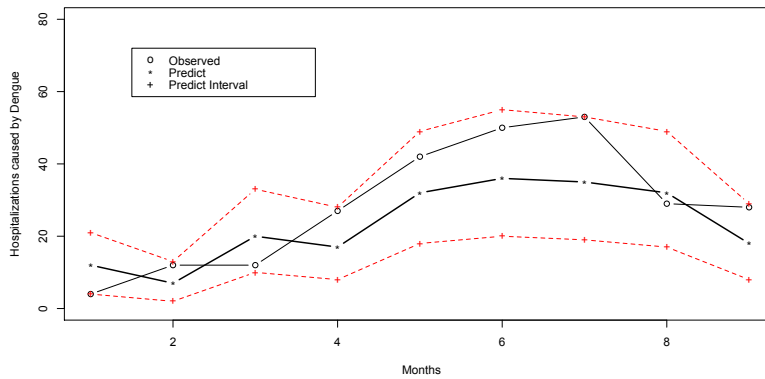


Figure 3.6: Predictions with GARMA(1,2) Negative Binomial model with Hospitalizations caused by Dengue series.

3.4.3 Mortality data set

Our last real data set is the number of deaths in Brazil between January 1984 and December 2007. This data is available from the Brazilian Health Ministry at <http://www2.datasus.gov.br/DATASUS> and is depicted in Figure 3.7. Likewise the first example, the original series was divided by 1000 to reduce the magnitude of the data. As in the first example, we think there is a point for the inclusion of an extra term here too since the series exhibits a long-term (possibly nonlinear) increase. So, a new component $\beta_{\text{exp}} = \log(t)$ was added to the model equation as this is expected to improve model estimation.

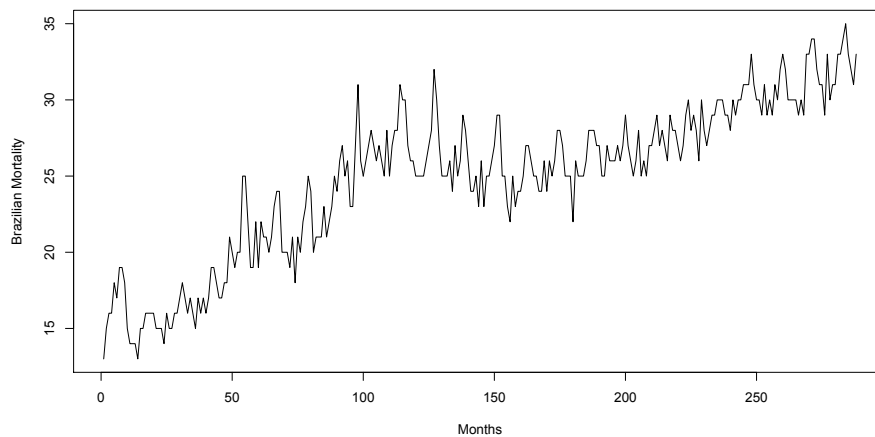


Figure 3.7: Number of deaths in Brazil.

Table 3.8: Bayesian selection criteria using the number of deaths in Brazil.

Poisson	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	1549.55	1560.41	1546.60	1566.37	1565.79	1566.80
DIC	1531.53	1566.68	1531.10	1570.34	1571.11	1573.77
CPO	-766.42	-773.35	-765.49	-784.20	-784.99	-785.48
Binomial	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	1351.42	1412.95	1357.52	1391.79	1399.13	1404.81
DIC	1341.42	1391.56	1342.28	1371.54	1378.10	1379.88
CPO	-670.73	-705.64	-671.10	-695.29	-716.72	-708.52
Negative Binomial	GARMA(1,0)	GARMA(2,0)	GARMA(1,1)	GARMA(1,2)	GARMA(2,1)	GARMA(2,2)
EBIC	1705.33	1709.12	1700.61	1735.23	1734.39	1738.30
DIC	1693.59	1696.61	1685.18	1714.76	1713.51	1712.47
CPO	-851.35	-855.13	-842.01	-866.51	-866.45	-866.47

Looking at the Bayesian selection criteria given in Table 3.8 we can conclude that the best model for this particular data is the GARMA(1,0) Binomial model. There are only three parameters in this model and the estimation results are shown in Table 3.9. Here the extra parameter was fixed at $m = 45$.

Table 3.9: Estimates of the number of deaths in Brazil series with GARMA(1,0) Binomial.

Parameter	Mean	Variance	HPD Credible Interval	AP
β_0	0.4154	0.0006	(0.3739; 0.4724)	0.2272
β_{exp}	0.0713	0.0004	(0.0651; 0.0774)	0.3503
ϕ_1	0.7637	0.0007	(0.7462; 0.7788)	0.1885

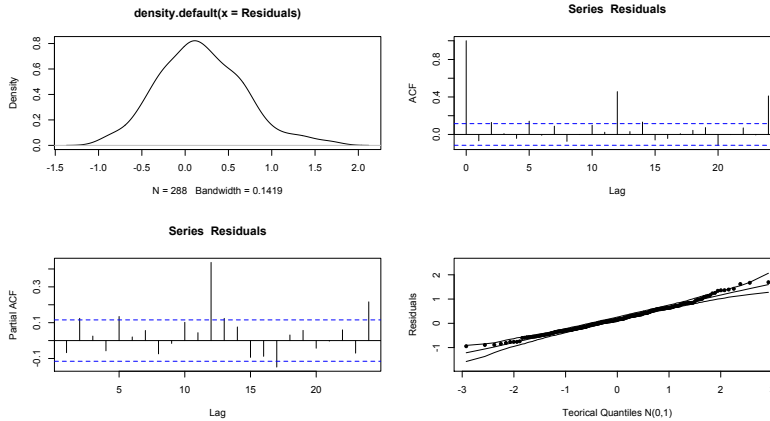


Figure 3.8: Residual analysis of the number of deaths in Brazil.

The residual analysis summarized in Figure 3.8 indicates that the residuals are non-correlated and Gaussian distributed with mean 0.1850 and standard

deviation 0.4894. The Kolmogorov-Smirnov and Anderson-Darling normality tests returned p -values of 0.6736 and 0.1304 respectively thus indicating evidence for the Gaussian assumption.

Likewise the previous examples we repeated the prediction exercise here. This time we used the 10 last observations as the series is longer. So, the GARMA(1,0) binomial model was fitted to the series y_1, \dots, y_{n-k} and a one-step ahead prediction \hat{y}_{n-k+1} was produced for $k = 1, \dots, 10$. The results are illustrated in Figure ?? from which we can see that the prediction errors are again overall small. Using these prediction errors the calculated value for the MAPE criterion was 3.63%.

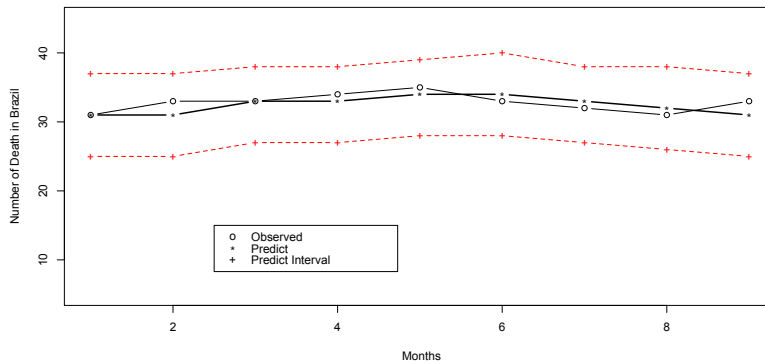


Figure 3.9: Predictions with GARMA(1,0) Binomial model with Number of death in Brazil series.

Transformed GARMA model

Abstract

Real time series can present anomalies, like non-additivity, non-normality and heteroscedasticity which makes using GARMA models impossible. Our paper introduces a new class of models called Transformed Generalized Autoregressive Moving Average (TGARMA) models, that allow using transformations to guarantee the GARMA assumptions. We present an extensive simulation study of the influence of the transformation on GARMA estimation. We also propose using bootstrap methods to get more information about the distribution of the transformation parameter. We apply the methodology to data related to Annual Swedish fertility rates.

4.1 TGARMA model

[Box and Cox \(1964\)](#) commented that many important results in statistical analysis follow from the assumption that the population being sampled or investigated is normally distributed with a common variance and additive error structure. For this reason, these authors presented a transformation called Box-Cox power transformation that has generated a great deal of interests, both in theoretical work and in practical applications.

This family has been modified [Cox \(1981\)](#) to take account of the discontinuity at $\lambda = 0$, such that

$$y_t^{(\lambda)} = \begin{cases} \frac{(Y_t^{\lambda}-1)}{\lambda}; \lambda \neq 0 \\ \log(Y_t); \lambda = 0 \end{cases}$$

Sakia (1992), Manly (1976) and Draper and Cox (1969) discuss others transformations which have the same goal: reduce anomalies in the data. The literature recommends the use of Box-Cox power transformation as a general transformation. The next section presents the TGARMA approach using the Box-Cox power transformation.

4.1.1 Model definition

The TGARMA model specifies the conditional distribution of each transformed observation $y_t^{(\lambda)}$, for $t = 1, \dots, n$ given the previous information set, defined by $F_{t-1}^{(\lambda)} = (y_1^{(\lambda)}, \dots, y_{t-1}^{(\lambda)}, \mu_1, \dots, \mu_{t-1})$. The conditional density belongs to exponential family and is given by

$$f(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \exp \left(\frac{y_t^{(\lambda)} \alpha_t - b(\alpha_t)}{\varphi} + d(y_t^{(\lambda)}, \varphi) \right), \quad (4.1)$$

where α_t and φ are canonical and scale parameters, respectively. Moreover $b(\cdot)$ and $d(\cdot)$ are specific functions that define the particular exponential family. The conditional mean and conditional variance of y_t given $F_{t-1}^{(\lambda)}$ are represented as:

$$\begin{aligned} \mu_t = b'(\alpha_t) &= E(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) \\ \text{Var}(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) &= \varphi b''(\alpha_t), \end{aligned} \quad (4.2)$$

with $t = 1, \dots, n$.

Following the Generalized Linear Models (GLM) approach the parameter μ_t is related to the predictor η_t by a twice differentiable one-to-one monotonic function g , called *link function*. In general, we can also include set of covariates \mathbf{x} into our model. Moreover, we can we add an additional component allowing autoregressive moving average terms to be included. In such a case our model will have a form:

$$g(\mu_t) = \eta_t = x'_t \beta + \sum_{j=1}^p \phi_j \{g(y_{t-j}^{(\lambda)}) - x'_{t-j} \beta\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}^{(\lambda)}) - \eta_{t-j}\}. \quad (4.3)$$

The parameters p and q are identified using the classical BIC or AIC criteria. For more information the reader is referred to [Kedem and Fokianos \(2002\)](#) and [Li \(1994\)](#).

The TGARMA(p,q) model is defined by the equations (4.1) and (4.3). For certain functions g , it may be necessary to replace $y_t^{(\lambda)}$ with $y_t^{(\lambda^{new})}$ in (4.3) to avoid the non-existence of $g(y_t^{(\lambda)})$ for certain values of y_t . The form $y_t^{(\lambda^{new})}$ depends on the particular function g and is defined for specific cases later.

We will consider two important continuous GARMA models: gamma and inverse Gaussian. We will present each one with a density and a respective predictor. The simulation study and real data analysis were done for each of the distributions.

4.1.2 Examples

Gamma TGARMA model

Suppose that $y_t^{(\lambda)} | F_{t-1}^{(\lambda)}$ follows gamma distribution with the mean μ_t , thus

$$f(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu_t} \right)^\nu y_t^{(\lambda)(\nu-1)} \exp \left(-\frac{y_t^{(\lambda)} \nu}{\mu_t} \right). \quad (4.4)$$

Here, $y_t^{(\lambda)} | F_{t-1}^{(\lambda)}$ has a distribution in the exponential family with $\alpha_t = -\frac{1}{\mu_t}$, $b(\alpha_t) = -\log \left(\frac{\nu}{\mu_t} \right)$, $c(y_t^{(\lambda)}, \varphi) = \frac{1}{\Gamma(\nu)} + \nu \log(\nu) + (\nu-1) \log(y_t^{(\lambda)})$ and $\varphi = \nu$. Moreover $y_t^{(\lambda)} > 0$, $\mu_t > 0$ and $\nu > 0$. And $E(Y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \mu_t$ where $Var(Y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \mu_t \nu$.

The canonical link function for this model is the logarithmic function, thus the linear predictor is given by

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}^{(\lambda^{new})}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^{(\lambda^{new})}) - \log(\mu_{t-j})\}. \quad (4.5)$$

In the equation above $y_{t-j}^{(\lambda^{new})} = \max(y_{t-j}, c)$, $0 < c < 1$. In sequel we will drop the superscript "new" understanding that we truncate $y_t^{(\lambda)}$ from below if needed. The gamma TGARMA model is defined by the equations (4.4) and (4.5).

Inverse Gaussian TGARMA model

Suppose that $y_t^{(\lambda)} | F_{t-1}^{(\lambda)}$ follows an inverse Gaussian distribution with the mean μ_t , thus

$$f(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \exp \left\{ \frac{1}{\sigma^2} \left[-\frac{2y_t^{(\lambda)}}{\mu_t^2} + \frac{1}{\mu_t} \right] - \frac{1}{2} \log(2\pi\sigma^2 y_t^{(\lambda)^3}) - \frac{1}{2\sigma^2 y_t^{(\lambda)}} \right\}. \quad (4.6)$$

The canonical link function for this model is the logarithmic function. The linear predictor is given by

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \left\{ \log(y_{t-j}^{(\lambda)}) \right\} + \sum_{j=1}^q \theta_j (\log(y_{t-j}^{(\lambda)}) - \log(\mu_{t-j})). \quad (4.7)$$

The inverse Gaussian TGARMA model is defined by the equations (4.6) and (4.7).

4.1.3 Model Fitting

The TGARMA model-fitting procedure described herein performs the maximum likelihood estimation (see Benjamin et al. (2003)). The estimation method is based on the standard GLM.

Let $\{y_t^{(\lambda)}\}$ be a time series where the equations (4.1) and (4.3) are satisfied. The parameter vector is $\gamma' = (\beta', \phi', \theta')$, where $\beta = (\beta_0 \dots \beta_m)'$, $\phi = (\phi_1 \dots \phi_p)'$ and also $\theta = (\theta_1 \dots \theta_q)'$. For the estimation procedure the conditional likelihood

function is used where the $F_r^{(\lambda)} = \{y_1^{(\lambda)} \dots y_r^{(\lambda)}\}$, for $r = \max(p, q)$. The partial likelihood function can be constructed considering that $y_{t-1}^{(\lambda)}$ and $y_t^{(\lambda)}$ are conditionally independent. Thus

$$\begin{aligned} L(\beta, \phi, \theta | F_n^{(\lambda)}) &\propto \prod_{t=r+1}^n f(y_t^{(\lambda)} | F_t^{(\lambda)}) \\ &\propto \prod_{t=r+1}^n \exp \left\{ \frac{y_t^{(\lambda)} g(\mu_t) - b(g^{-1}(\mu_t))}{\varphi} + d(y_t^{(\lambda)}, \varphi) \right\}, \end{aligned} \quad (4.8)$$

where $g(\mu_t)$ is the link function given by

$$g(\mu_t) = x_t' \beta + \sum_{j=1}^p \phi_j \{g(y_{t-j}^{(\lambda)}) - x_{t-j}' \beta\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}^{(\lambda)}) - g(\mu_{t-j})\}. \quad (4.9)$$

In the above equations $t = r + 1, \dots, n$. The equations (4.8) and (4.9) do not have a closed form solution therefore a numerical optimization routine will be used. Due to the results of Benjamin et al. (2003) the asymptotic distribution of the partial likelihood estimator is known. However, the information matrix of the corresponding asymptotic normal law is quite difficult to estimate. This gives us a strong motivation to consider resampling techniques in the GARMA models. As it also known, resampling techniques provide a better alternative to construct confidence intervals in the time series field de Andrade et al. (2016c).

The profile likelihood for the parameter vector γ and scalars ϕ and λ is expressed in terms of the transformed series $y^{(\lambda)} = (y_{r+1}^{(\lambda)}, \dots, y_n^{(\lambda)})^T$ conditioned on the first r transformed observations, where $r = \max\{p, q\}$, is

$$PL(\beta, \phi, \theta, \lambda) = \prod_{t=r+1}^n L(\beta, \phi, \theta | F_t^{(\lambda)}) J(\lambda, y_t), \quad (4.10)$$

where $J(\lambda, y_t)$ is the Jacobian of the transformation from y_t to $y_t^{(\lambda)}$. The inference method is made selecting a range $\lambda^{(1)}, \dots, \lambda^{(k)}$, and evaluating the profile likelihood on each $\lambda^{(i)}$, $i = 1, \dots, k$.

$$\hat{\lambda} = \max \left(PL(\hat{\gamma}^{(\lambda^i)}, \hat{\phi}^{(\lambda^i)}, \lambda^i) \right), i = 1, \dots, k \quad (4.11)$$

The profile likelihood depends on the Jacobian of the transformation, so

different transformation provide different profiles likelihood. The Box-Cox power transformation was used, thus

$$J(\lambda, y_t) = \begin{cases} \sum_{i=1}^n \{(\lambda - 1) \log(y_i)\}; & \lambda \neq 0 \\ \sum_{i=1}^n \{-\log(y_i)\}; & \lambda = 0 \end{cases}$$

4.2 Moving Block Bootstrap on TGARMA models

The estimation of λ , the transformation parameter, is done using the profile likelihood (PL). The PL provides a single value $\hat{\lambda}$ that corresponds to the maximum of the selected interval. The information about λ is very scarce, therefore, resampling methods can be used to approximate the empirical distribution of $\hat{\lambda}$ enabling the construction of confidence intervals for λ .

The Moving Block Bootstrap (MBB) will be used to generate B replicates time series, enabling the use of PL to select the best value of the transformation parameter. We use the following

Resampling Algorithm

- STEP 1** Apply the Moving Block Bootstrap approach to y_t that comes from TGARMA. Select randomly the block starting point. Select the size of the block that should be between $n^{\frac{1}{4}}$ and $n^{\frac{1}{2}}$, where n represent the length of the original time series.
- STEP 2** If the initial point plus the size of the block is smaller than n , take the block $Bl(y_t)$. The chosen block is now the first piece of the resampled time series y_t^* .
- STEP 3** Repeat [**STEP 1**] and [**STEP 2**] until the length of y_t^* equals n .

In the sequel we provide a complete simulation study, checking computationally the properties of the algorithm above. Firstly, however, we will briefly show the usefulness of our approach to the prediction problem.

4.2.1 Forecasting for TGARMA

Below we provide final formula for conditional expectation and conditional variance. The lengthy technical calculations are omitted due to the space constraints.

When $\lambda \neq 0$ the inverse Box-Cox gets

$$Y_t = (1 + \lambda \mu_t)^{1/\lambda} \{1 + \nu_t (Y_t^{(\lambda)} - \mu_t)\}^{1/\lambda}, \quad (4.12)$$

where $\nu_t = \lambda/(1 + \lambda \mu_t)$. Applying the binomial expansion, we obtain

$$E(Y_t|F_{t-1}) = (1 + \lambda \mu_t)^{1/\lambda} \left\{ 1 + \sum_{i=2}^{\infty} \frac{a_i \mu_{ti}}{i! (1 + \lambda \mu_t)^i} \right\}, \quad (4.13)$$

where $a_i = \prod_{j=0}^{i-1} (1 - j\lambda)$ and $\mu_{ti} = E((Y_t^{(\lambda)} - \mu_t)^i | F_{t-1})$ is the i th conditional central moment of $Y_t^{(\lambda)}$.

Equation (4.13) generalizes the expansion given by Pankratz and Dudley (1987).

We also obtain that conditional variance of Y_t can be approximated by

$$(1 + \lambda \mu_t)^{2/\lambda} \left\{ \frac{\mu_{t2}}{(1 + \lambda \mu_t)^2} + \frac{(1 - \lambda) \mu_{t3}}{(1 + \lambda \mu_t)^3} + \frac{(1 - \lambda) \{(7 - 11\lambda) \mu_{t4} - 3(1 - \lambda) \mu_{t2}^2\}}{12 (1 + \lambda \mu_t)^4} \right\}.$$

4.3 Simulation Study

The performance of the Maximum Likelihood estimation was evaluated using two measures: the corrected bias (CB), the corrected error (CE). These measures are defined as,

$$CB = \frac{1}{m} \sum_{i=1}^m \left| \frac{\theta - \hat{\theta}^{(i)}}{\theta} \right|,$$

$$CE^2 = \frac{1}{Var} \frac{1}{m} \sum_{i=1}^m (\hat{\theta}^{(i)} - \theta)^2$$

where $\hat{\theta}^{(i)}$ are the estimates of parameter θ for the i -th replication, $i = 1, \dots, m$. The variance term (Var) that appears in the definition of CE is the sample variance of $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(m)}$. We have also calculated the coverage representing the proportion of simulated cases when the confidence intervals contain the true simulated value.

The estimation results appear in Tables 4.2 and 4.4 where the aforementioned metrics are shown for each model and parameter. These results indicate good

properties with relatively small values of the corrected bias (CB), values of the corrected error (CE) around 1 and also coverage around 95%.

In Table 4.3 we present results of the simulation experiment that was checking proportion of times when the correct model was selected. The BIC criterion was used to select the best model, the criterion was used also with different sizes of series. Figures 4.1 and 4.2 represent results of a Monte Carlo study, the densities of the estimates are presented, showing approximately a Gaussian behavior.

The experiment was repeated 5000 times for each value of λ , with 500 observations in each of artificial time series. The simulation study for TGARMA models was conducted with Gamma with the orders (1,1),(2,1),(1,2) and (2,2). The results of the orders (1,2) and (2,1) was omitted, but is very similar of the results of orders (1,1) and (2,2) that follow.

Table 4.1: Gamma GARMA(p,q) real values

Order	β_0	ϕ_1	ϕ_2	θ_1	θ_2	ν
(1,1)	0.70	0.50	-	-0.30	-	2
(1,2)	-0.10	0.50	-	-0.30	0.30	2
(2,1)	0.10	0.50	0.30	-0.40	-	2
(2,2)	-0.10	0.30	0.20	0.25	-0.35	2

The real values used to simulated the artificial series are presented in Table 1, each line corresponds to the order of the model.

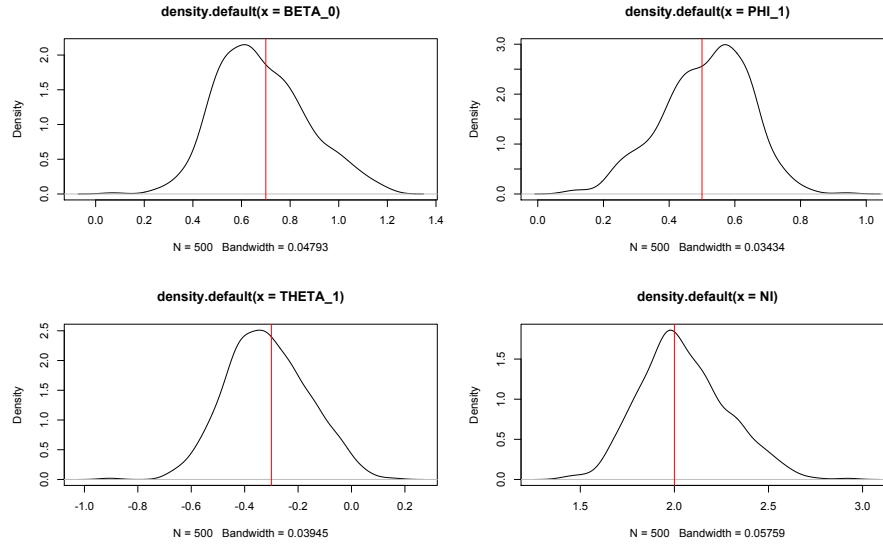
Figure 4.1: Densities of parameters with $\lambda = 0.3$

Table 4.2: TGARMA(1,1) with gamma distribution and Box-Cox power transformation

Parameter($\lambda = 0.3$)	CB	CE	Coverage
ν	0.0398	0.9724	0.9614
β_0	0.2217	1.0900	0.9252
ϕ_1	0.0488	1.0942	0.9218
θ_1	0.0593	0.9574	0.9624
Parameter($\lambda = 0.5$)	CB	CE	Coverage
ν	0.0410	1.0000	0.9578
β_0	0.2255	1.1034	0.9232
ϕ_1	0.0482	1.0746	0.9250
θ_1	0.0576	0.9398	0.9632
Parameter($\lambda = 0.7$)	CB	CE	Coverage
ν	0.0438	1.0613	0.9444
β_0	0.2207	1.0790	0.9286
ϕ_1	0.0480	1.0681	0.9370
θ_1	0.0553	0.9072	0.9714
Parameter($\lambda = 0.9$)	CB	CE	Coverage
ν	0.0458	1.1036	0.9366
β_0	0.2228	1.0896	0.9372
ϕ_1	0.0476	1.0436	0.9438
θ_1	0.0526	0.8629	0.9734

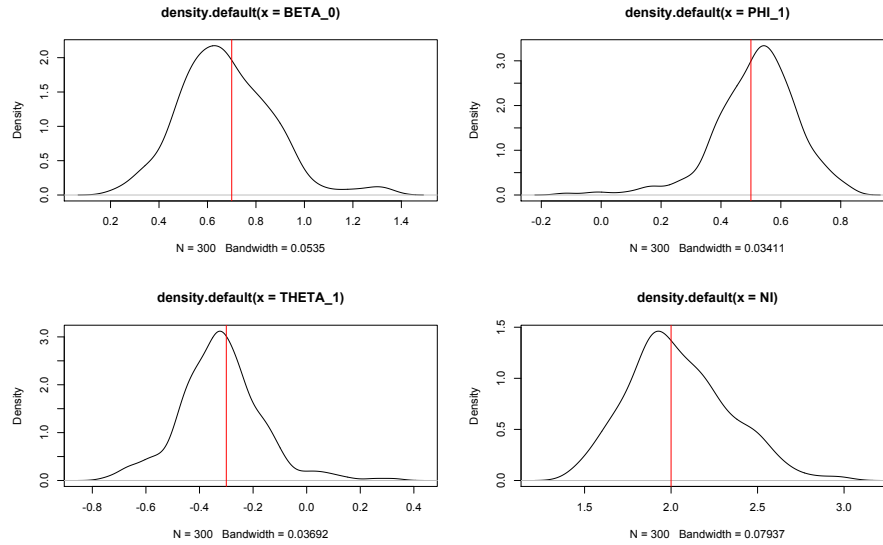
Figure 4.2: Densities of parameters with $\lambda = 0.5$

Table 4.3: Proportions of correct model using BIC with gamma TGARMA(p,q) model

$\lambda = 0.3$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.8284	0.4976	0.4866	0.2896
500	0.8980	0.5646	0.5075	0.3062
1000	0.9455	0.6102	0.5432	0.3445
$\lambda = 0.5$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.7980	0.4198	0.03410	0.2858
500	0.8965	0.4728	0.3625	0.3003
1000	0.9360	0.4910	0.4454	0.3252
$\lambda = 0.7$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.7855	0.4290	0.4859	0.2884
500	0.8693	0.4474	0.5422	0.3042
1000	0.9280	0.4763	0.5810	0.3312
$\lambda = 0.9$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.7645	0.5250	0.5434	0.3038
500	0.8630	0.5334	0.5542	0.3102
1000	0.9294	0.5503	0.6110	0.3365

Table 4.4: TGARMA(2,2) with Gamma distribution and Box-Cox Power transformation

Parameter($\lambda= 0.3$)	CB	CE	Coverage
ν	0.0104	1.0094	0.9542
β_0	1.3310	1.0992	0.9308
ϕ_1	0.2486	1.0264	0.9323
ϕ_2	0.2620	1.0183	0.9342
θ_1	0.2908	1.0160	0.9313
θ_2	0.1108	1.0366	0.9527
Parameter($\lambda= 0.5$)	CB	CE	Coverage
ν	0.0105	1.0166	0.9569
β_0	1.3510	1.0781	0.9398
ϕ_1	0.2490	1.0235	0.9398
ϕ_2	0.2645	1.0297	0.9379
θ_1	0.2932	1.0137	0.9365
θ_2	0.1080	1.0081	0.9469
Parameter($\lambda= 0.7$)	CB	CE	Coverage
ν	0.0114	1.0795	0.9419
β_0	1.3380	1.0649	0.9527
ϕ_1	0.2450	1.0068	0.9340
ϕ_2	0.2580	1.0113	0.9321
θ_1	0.2860	0.9930	0.9404
θ_2	0.1091	1.0119	0.9503
Parameter($\lambda= 0.9$)	CB	CE	Coverage
ν	0.0116	1.1176	0.9304
β_0	1.4260	1.1256	0.9533
ϕ_1	0.2553	1.0453	0.9297
ϕ_2	0.2625	1.0189	0.9373
θ_1	0.3012	1.0353	0.9304
θ_2	0.1148	1.0449	0.9505

The simulation study was carried out to verify the influence of selecting the value of λ on the TGARMA model estimates. Figure 4.1 presents the estimate densities for each parameter of the TGARMA(1,1) model with $\lambda = 0.3$, we observed a Gaussian behavior in the parameters. Table 4.2 presents results for the TGARMA model with one autoregressive and one moving average terms, the metrics proposed above showed that for different values of λ , the CB, CE and coverage still indicating good properties. Figure 2 presents the estimate densities for each parameter of the TGARMA(1,1) model with $\lambda = 0.5$, we also observed a Gaussian behavior in the parameters. Table 4.3 presents the proportions of correct model selected by using the BIC criterion, four different models were proposed, (1,1), (2,1), (1,2) and (2,2). The results presented on Table 4.2 shows that the higher proportions indicate the corrected model, unrelated to the λ value. The Table 4.4 presents results for the TGARMA model with two autoregressive

and two moving average terms, the results follow the same aspect of the Table 4.1. The next subsection contains all the results related to bootstrap on TGARMA models.

4.3.1 Bootstrap Simulation Study

We carried out two important analyses of the bootstrap simulation study. The Figure 4.3 presents the estimated density of the bootstrap replicates, the vertical lines represent true value, the PL estimate, the bootstrap mean and the confidence intervals.

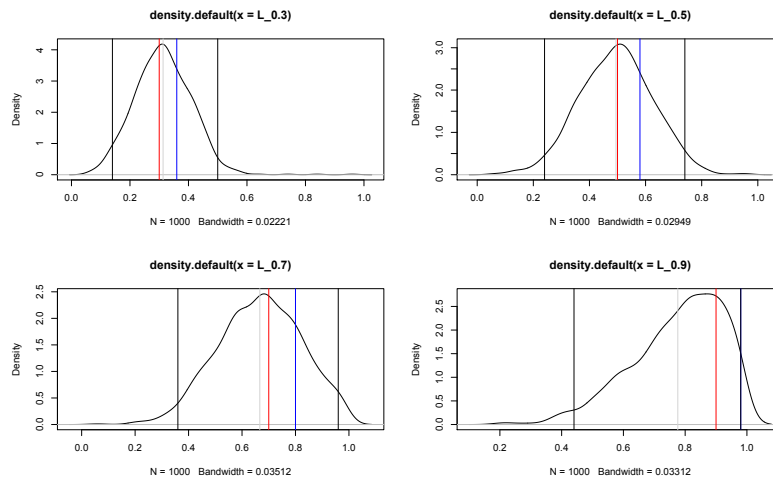


Figure 4.3: Bootstrap densities for TGARMA models

Table 4.5: Bootstrap TGARMA(1,1) Gamma

True Value	PL Estimate	Bootstrap Mean	Bootstrap Median
0.30	0.36	0.3132	0.32
0.50	0.58	0.4947	0.50
0.70	0.80	0.6670	0.68
0.90	0.98	0.7761	0.80

We also conducted a simulation study testing the coverage of the PL using the bootstrap replicates. We carried out a coverage simulation study with 500 bootstrapped replicates, 500 times, both studies were conducted with artificial series of the length 200.

Careful reading of Table 4.3 is leading us to conclude that the mean and the median of bootstrap distribution $\hat{\lambda}$ provide better estimates of the true value λ compared with the PL estimate. Table 4.6 presents the coverage of the transformation parameter via bootstrap quantiles, also compares maximum partial likelihood bias with the bootstrap mean bias. The results presented on Table 4.6 indicate that the coverage is close to 95% and the bias calculated via bootstrap mean lower than the MPLE bias for $\lambda < 0.5$

Table 4.6: Coverage Bootstrap TGARMA(1,1) Gamma

True Value	Coverage	MPLE Bias	Bootstrap Bias
0.30	95.80%	0.2417	0.2166
0.50	94.40%	0.2001	0.1978
0.70	95.20%	0.1717	0.1816
0.90	83.40%	0.1437	0.2691

4.4 Real data analysis

The demography of Sweden is monitored by Statistics Sweden (SCB). As of 31 December 2013, Sweden's population was estimated to be 9.64 million people, making it the 90th most populous country in the world. The three biggest cities are Stockholm, Gothenburg and Malmö. Approximately 85% of the country's population resides in urban areas.

The real data set represents the Annual Swedish fertility rates (1000's) 1750-1849. We obtained the data set on <https://datamarket.com/data/set/22s2/annual-swedish-fertility-rates-1000s-1750-1849-thomas-1940>.

Figure 4.4 present the graph of Annual Swedish fertility rates and Figure 4.5 the Auto Correlation Function and Partial Auto Correlation Function of Annual Swedish fertility rates.

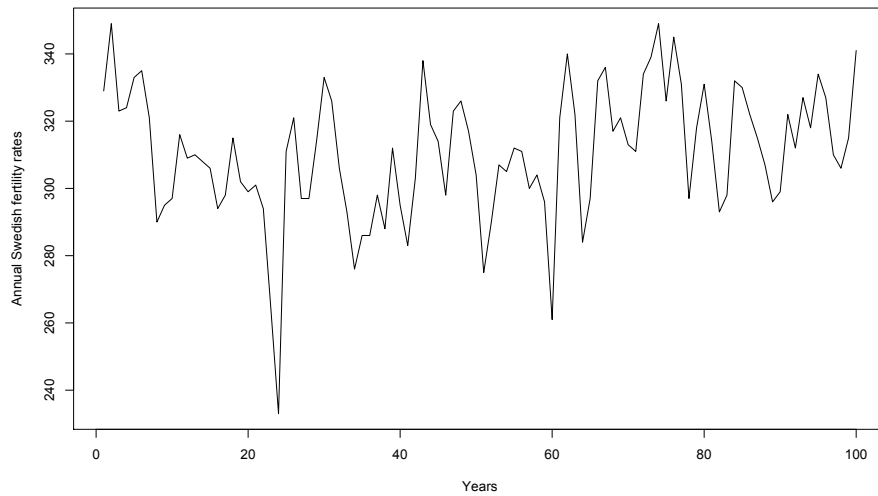


Figure 4.4: Graph of Annual Swedish fertility rates

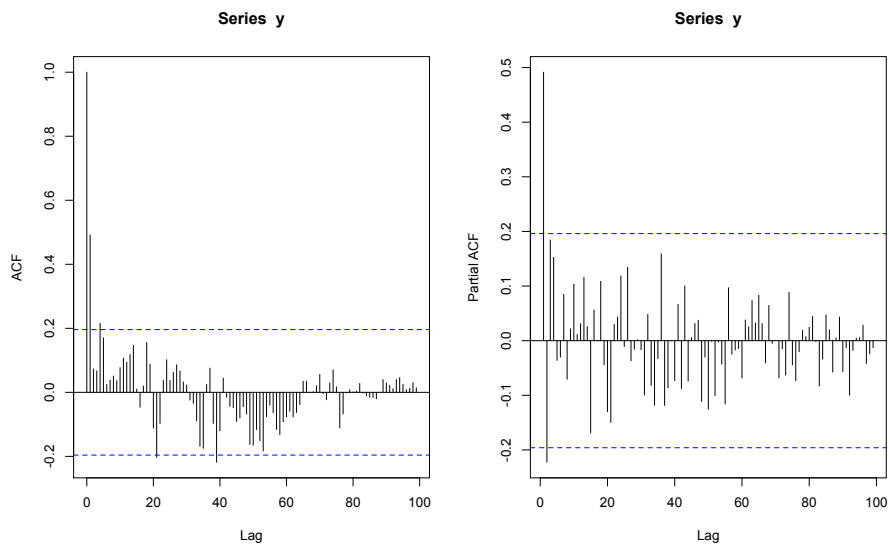


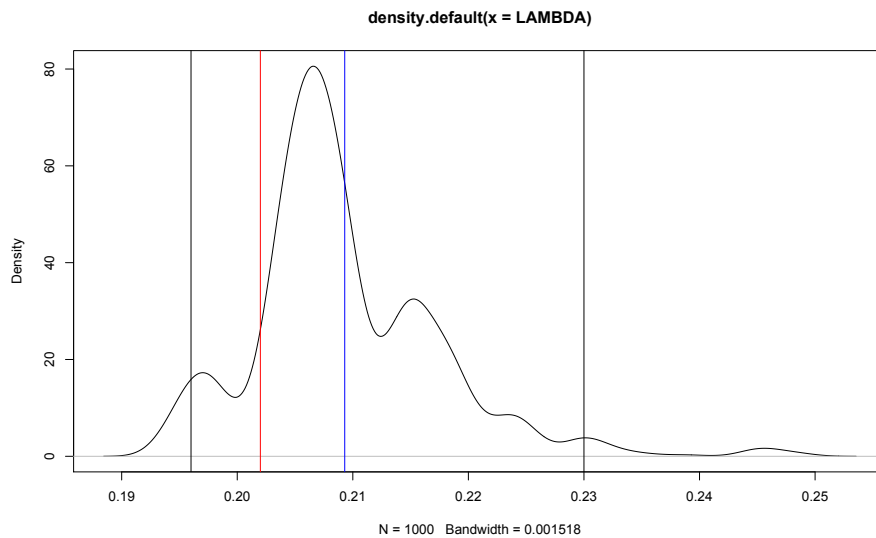
Figure 4.5: Auto Correlation Function and Partial Auto Correlation Function of Annual Swedish fertility rates

The selection criterion AIC and BIC was used to select between the gamma and inverse Gaussian and also to select the order of the model. Below we provide the Table 4.7 with the results.

Table 4.7: Criterion selection using Annual Swedish fertility rates

Gamma	TGARMA(1,0)	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
AIC	582.56	582.85	655.41	655.40	658.12
BIC	592.98	593.17	665.83	665.85	667.71
Inverse Gaussian	TGARMA(1,0)	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
AIC	765.87	766.59	785.46	785.45	786.22
BIC	775.12	777.01	795.88	795.87	796.14

The range of the transformation parameter λ was selected from $0 \leq \lambda < 1$, with step 0.001. The Partial Likelihood selected the value $\hat{\lambda} = 0.202$. With bootstrap replicates we obtained the confidence intervals for λ . The inferior bound was 0.196 and the superior bound 0.23. We also we evaluate the mean of the bootstrap replicates, which was 0.2093.

Figure 4.6: λ empirical bootstrapped density Annual Swedish fertility rates

Just like the simulation study showed, the mean of bootstrapped replicates represent a better estimate of λ . Therefore, we used the Box Cox Power transformation with this value and adjusted the model GARMA(1,0) with gamma distribution. Figure 4.6 present the empirical density of the transformation parameter, the lines represent the Partial Likelihood selected the value and the mean of the bootstrap replicates.

Below on Table 4.8 we provide the Moving Block Bootstrap to evaluate the confidence intervals of the parameters (de Andrade et al. (2016c)).

Table 4.8: Estimates of Annual Swedish fertility rates series with TGARMA(1,0) Gamma

Parameter	Estimate	Inferior Bootstrap Bound	Superior Bootstrap Bound
β_0	1.2234	1.2226	1.2287
ϕ_1	0.4916	0.4890	0.4923
ν	6.5041	6.4872	6.7019

The goodness of fit can be checked using the graph of true values versus the estimated values.

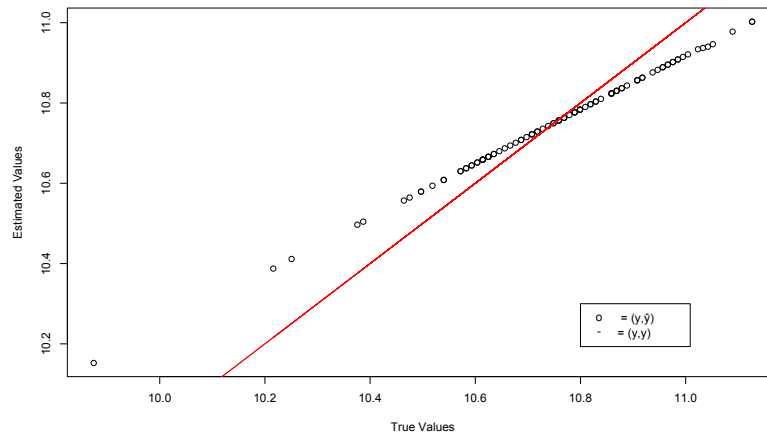


Figure 4.7: Graph of true values versus the estimated values of the residuals of Annual Swedish fertility rates series

The Figure 4.7 presents a quantile plot with the true values on x axis and the estimated values on y axis. The line represents the perfect model with real values on axis x and y.

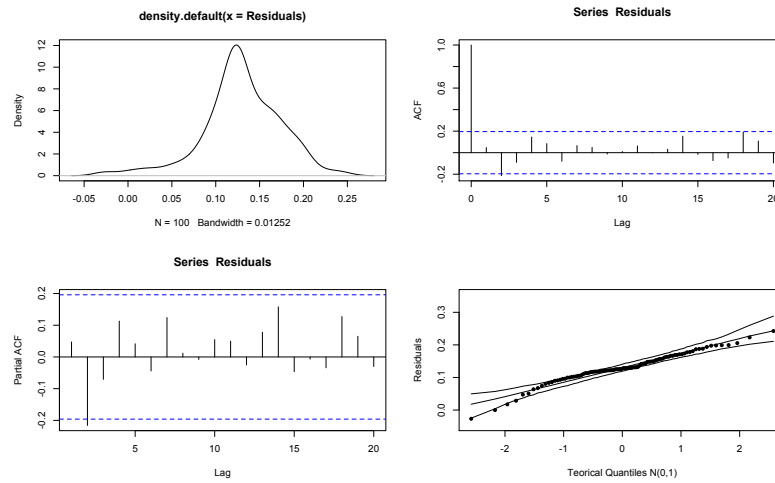


Figure 4.8: Autocorrelation function and partial autocorrelation function of the residuals of rate of Annual Swedish fertility rates series

Quantile residuals are based on the idea of inverting the estimated distribution function for each observation to obtain exactly standard normal residuals. In the case of discrete distributions, such as the binomial, negative binomial and Poisson, some randomization is introduced to produce continuous normal residuals. The residuals is given by $r_t = \Phi^{-1}(\mathbf{F}_{y_t}(y_t|F_{t-1}))$ where \mathbf{F}_{y_t} represent the cumulative distribution function for the respective density. Figure 4.8 present the residual analysis.

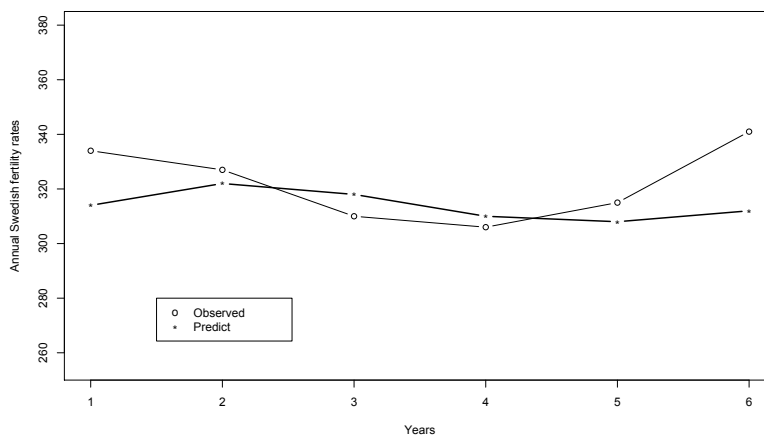


Figure 4.9: Original Predictions with GARMA(1,0) Gamma model with Rate of Annual Swedish fertility rates series

The prediction were made by the median. Only the first term of Taylor expansion was used. Using the estimate, predictions of 6 steps ahead of the original series can be made. The 6 last values of the series were removed and fitted the model without them. Prediction one step ahead for 6 years values, thus the predicted value be compared with the true value. Figure 4.9 present the predictions, the MAPE was calculated to quantify the quality of predictions.

The MAPE was calculated to assess the quality of predictions, the value was 03.70% which indicated good predictions.

Bayesian Transformed GARMA Models

Abstract

The Transformed Generalized Autoregressive Moving Average (TGARMA) models were proposed to deal with non-additivity, non-normality and heteroscedasticity in real data. This work presents a Bayesian approach for TGARMA models, extending the original model. The Bayesian methodology brings some advantages as prior densities and Bayesian model selection criteria. We conducted a simulation study to investigate the performance of Bayesian estimation and Bayesian model selection criteria. In addition, a real dataset was analysed using the proposed approach.

5.1 Transformed Generalized Autoregressive Moving Average (TGARMA) Model

[Box and Cox \(1964\)](#) commented that many important results in statistical analysis follow from the assumption that the population being sampled or investigated is normally distributed with a common variance and additive error structure. For this reason, these authors presented a transformation called Box-Cox power transformation that has generated a great deal of interests, both in theoretical work and in practical applications.

This family has been modified [Cox \(1981\)](#) to take account of the discontinuity at $\lambda = 0$, such that

$$y_t^{(\lambda)} = \begin{cases} \frac{(Y_t^\lambda - 1)}{\lambda}; & \lambda \neq 0 \\ \log(Y_t); & \lambda = 0 \end{cases}$$

[Sakia \(1992\)](#), [Manly \(1976\)](#) and [Draper and Cox \(1969\)](#) discuss others transformations which have the same goal: reduce anomalies in the data. The literature recommends the use of Box-Cox power transformation as a general transformation. The next section presents the TGARMA approach using the Box-Cox power transformation.

5.1.1 Model definition

The TGARMA model specifies the conditional distribution of each transformed observation $y_t^{(\lambda)}$, for $t = 1, \dots, n$ given the previous information set, defined by $F_{t-1}^{(\lambda)} = (y_1^{(\lambda)}, \dots, y_{t-1}^{(\lambda)}, \mu_1, \dots, \mu_{t-1})$. The conditional density belongs to exponential family and is given by

$$f(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \exp \left(\frac{y_t^{(\lambda)} \alpha_t - b(\alpha_t)}{\varphi} + d(y_t^{(\lambda)}, \varphi) \right), \quad (5.1)$$

where α_t and φ are canonical and scale parameters, respectively. Moreover $b(\cdot)$ and $d(\cdot)$ are specific functions that define the particular exponential family. The conditional mean and conditional variance of y_t given $F_{t-1}^{(\lambda)}$ are represented as:

$$\begin{aligned} \mu_t = b'(\alpha_t) &= E(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) \\ \text{Var}(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) &= \varphi b''(\alpha_t), \end{aligned} \quad (5.2)$$

with $t = 1, \dots, n$.

Following the Generalized Linear Models (GLM) approach the parameter μ_t is related to the predictor η_t by a twice differentiable one-to-one monotonic function g , called *link function*. In general, we can also include set of covariates \mathbf{x} into our model. Moreover, we can we add an additional component allowing autoregressive moving average terms to be included. In such a case our model will have a form:

$$g(\mu_t) = \eta_t = x_t' \beta + \sum_{j=1}^p \phi_j \{g(y_{t-j}^{(\lambda)}) - x_{t-j}' \beta\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}^{(\lambda)}) - \eta_{t-j}\}. \quad (5.3)$$

The parameters p and q are identified using the classical BIC or AIC criteria. For more information the reader is referred to [Kedem and Fokianos \(2002\)](#) and [Li \(1994\)](#).

The TGARMA(p,q) model is defined by the equations (5.1) and (5.3). For certain functions g , it may be necessary to replace $y_t^{(\lambda)}$ with $y_t^{(\lambda^{new})}$ in (5.3) to avoid the non-existence of $g(y_t^{(\lambda)})$ for certain values of y_t . The form $y_t^{(\lambda^{new})}$ depends on the particular function g and is defined for specific cases later.

We will consider two important continuous GARMA models: gamma and inverse Gaussian. We will present each one with a density and a respective predictor. The simulation study and real data analysis were done for each of the distributions.

5.1.2 Examples

Gamma TGARMA model

Suppose that $y_t^{(\lambda)} | F_{t-1}^{(\lambda)}$ follows gamma distribution with the mean μ_t , thus

$$f(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu_t} \right)^\nu y_t^{(\lambda)(\nu-1)} \exp \left(-\frac{y_t^{(\lambda)} \nu}{\mu_t} \right). \quad (5.4)$$

Here, $y_t^{(\lambda)} | F_{t-1}^{(\lambda)}$ has a distribution in the exponential family with $\alpha_t = -\frac{1}{\mu_t}$, $b(\alpha_t) = -\log \left(\frac{\nu}{\mu_t} \right)$, $c(y_t^{(\lambda)}, \varphi) = \frac{1}{\Gamma(\nu)} + \nu \log(\nu) + (\nu-1) \log(y_t^{(\lambda)})$ and $\varphi = \nu$. Moreover $y_t^{(\lambda)} > 0$, $\mu_t > 0$ and $\nu > 0$. And $E(Y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \mu_t$ where $Var(Y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \mu_t \nu$.

The canonical link function for this model given by inverse function, however we choose to use the logarithmic function, thus the linear predictor is given by

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \{\log y_{t-j}^{(\lambda^{new})}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^{(\lambda^{new})}) - \log(\mu_{t-j})\}. \quad (5.5)$$

In the equation above $y_{t-j}^{(\lambda^{new})} = \max(y_{t-j}, c)$, $0 < c < 1$. In sequel we will drop the superscript "new" understanding that we truncate $y_t^{(\lambda)}$ from below if needed. The gamma TGARMA model is defined by the equations (5.4) and (5.5).

Inverse Gaussian TGARMA model

Suppose that $y_t^{(\lambda)} | F_{t-1}^{(\lambda)}$ follows an inverse Gaussian distribution with the mean μ_t , thus

$$f(y_t^{(\lambda)} | F_{t-1}^{(\lambda)}) = \exp \left\{ \frac{1}{\sigma^2} \left[-\frac{2y_t^{(\lambda)}}{\mu_t^2} + \frac{1}{\mu_t} \right] - \frac{1}{2} \log(2\pi\sigma^2 y_t^{(\lambda)^3}) - \frac{1}{2\sigma^2 y_t^{(\lambda)}} \right\}. \quad (5.6)$$

The canonical link function for this model is the inverse of square function, however we choose to use the logarithmic function. The linear predictor is given by

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^p \phi_j \left\{ \log(y_{t-j}^{(\lambda)}) \right\} + \sum_{j=1}^q \theta_j (\log(y_{t-j}^{(\lambda)}) - \log(\mu_{t-j})). \quad (5.7)$$

The inverse Gaussian TGARMA model is defined by the equations (5.6) and (5.7).

5.2 Bayesian Approach on TGARMA Models

The MLE approach present the profile likelihood [Cox \(1975\)](#) to select the transformed value. Using a Bayesian approach, the λ value can be actualized, may providing better estimate than the MLE method.

5.2.1 Defining the Prior densities

Box and Cox (1964) shown that the transformation parameter present good properties when its belong to $[-1, 1]$ interval. Hence a prior density will respect this assumption.

$$\lambda \in [-1, 1]$$

The prior density for the transformation parameter λ is a uniform defined in $[-1, 1]$, given by

$$\pi(\lambda) = \frac{1}{2} \quad (5.8)$$

We added to the model prior information for the transformation parameter $\pi(\lambda)$. Using the logarithmic link function to guarantee positive values for any values of the vectors $\beta = (\beta_1, \dots, \beta_m)$, $\Phi = (\phi_1, \dots, \phi_p)$ and $\Theta = (\theta_1, \dots, \theta_q)$. Thus, a multivariate Gaussian prior will be proposed for each parameter, i.e.

$$\begin{aligned} \beta &\sim N(\boldsymbol{\mu}_0, \sigma_0^2 \mathbf{I}_0) \\ \Phi &\sim N(\boldsymbol{\mu}_1, \sigma_1^2 \mathbf{I}_1) \\ \Theta &\sim N(\boldsymbol{\mu}_2, \sigma_2^2 \mathbf{I}_2), \end{aligned}$$

where $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are vectors with lengths m , p and q respectively; σ_0^2 , σ_1^2 and σ_2^2 represent the prior variance, and \mathbf{I}_0 , \mathbf{I}_1 and \mathbf{I}_2 are $m \times m$, $p \times p$ and $q \times q$ identity matrices, respectively. The construction of the multivariate Gaussian depends on hyper-parameters, when there is no prior knowledge on these parameters, it can be considered a very large variance, making the prior densities flat. The parameters ν and σ^2 for the gamma and inverse Gaussian distribution respectively, must be positive. Thus we propose for these parameters a lognormal distribution, guaranteeing positive values as prior information.

$$\begin{aligned} \nu &\sim LN(\mu_3, \sigma_3^2) \\ \sigma^2 &\sim LN(\mu_4, \sigma_4^2) \end{aligned}$$

The hyper-parameters μ_3 and μ_4 represent the mean of the lognormal distribution and the σ_3^2 and σ_4^2 representing the variance, can also be considered

a very large variance, making the prior densities flat. The partial likelihood function for GARMA models can be constructed as follows:

$$\begin{aligned} L(\boldsymbol{\beta}, \Phi, \Theta, u|Y) &\propto \prod_{t=r+1}^n f(y_t|F_{t-1}) \\ &\propto \prod_{t=r+1}^n \exp\left(\frac{y_t\alpha_t - b(\alpha_t)}{\varphi} + d(y_t, \varphi)\right), \end{aligned}$$

where $u = \nu$ if we have the TGARMA with gamma distribution, and $u = \sigma^2$ if we have the TGARMA with Inverse Gaussian distribution and also $\alpha_t = g(\mu_t)$, which represents the link function given by

$$g(\mu_t) = x_t'\boldsymbol{\beta} + \sum_{j=1}^p \phi_j\{g(y_{t-j}^*) - x_{t-j}'\boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j\{g(y_{t-j}^*) - g(\mu_{t-j})\},$$

for all $t = r + 1, \dots, n$.

The posterior density is obtained combining the likelihood function with the prior densities. Let the vector $\mathbf{Y} = (y_t^{(\lambda)}, y_{t-1}^{(\lambda)}, \dots, y_1^{(\lambda)}, x_t, x_{t-1}, \dots, x_1, \dots)$ represent the necessary information to construct the likelihood function. The posterior density is then given by,

$$\pi(\boldsymbol{\beta}, \Phi, \Theta, u, \lambda|\mathbf{Y}) \propto L(\boldsymbol{\beta}, \Phi, \Theta, u|\mathbf{Y})\pi_0(\boldsymbol{\beta}, \Phi, \Theta, u, \lambda). \quad (5.9)$$

Where $\pi_0(\cdot)$ denotes a joint prior distribution. However, the joint posterior density of parameters in the GARMA models can not be obtained in closed form. Therefore, Markov chain Monte Carlo (MCMC) sampling strategies will be employed for obtaining samples from this joint posterior distribution. We used a Metropolis-Hastings algorithm to yield the required realizations and adopted a sampling scheme where the parameters are updated as a single block, and at each iteration we generated new values from a multivariate normal distribution centred around the maximum likelihood estimates with a variance-covariance proposal matrix given by the inverse Hessian evaluated at the posterior mode.

5.2.2 Bayesian prediction for GARMA models

The Bayesian model is defined by the equation (5.9). This equation present information of present values considering the likelihood function and the prior densities. However, the interest is in futures values which represent the model

for the information y_{t+h} .

$$f_{\Theta}(y_{t+h}|F_t) \quad (5.10)$$

The model of y_{t+h} provide the equation

$$\pi(y_{t+h}, \Theta|F_t) = f_{\Theta}(y_{t+h}|F_t)\pi(\Theta|F_t). \quad (5.11)$$

Thus, the predictive density is established and given by

$$P(y_{t+h}|F_t) = \int_{\theta \in \Theta} \pi(y_{t+h}, \Theta|F_t) d\theta. \quad (5.12)$$

The predictive density (5.12) is rewritten by

$$P(y_{t+h}|F_t) = \int_{\theta \in \Theta} f_{\Theta}(y_{t+h}|F_t)\pi(\Theta|F_t) d\theta. \quad (5.13)$$

The aim is determine the predictive density using the MCMC algorithm, thus

$$\hat{P}(y_{t+h}|F_t) = \frac{1}{Q} \sum_{j=1}^Q f(y_{t+h}|\theta^{(j)}, F_t). \quad (5.14)$$

The MCMC vector provide a vector of predictive densities of y_{t+h} , thus

$$\hat{P}(y_{t+h}|F_t) = \frac{1}{Q} \sum_{j=1}^Q f_{\Theta^{(j)}}(y_{t+h}|F_t). \quad (5.15)$$

Given the predictive density, the next step is evaluate the prediction $E(y_{t+h}|F_t) = \hat{y}_{t+h}$

$$E(y_{t+h}|F_t) = \int_{y_{t+h} \in Y_{t+h}} y_{t+h} P(y_{t+h}|F_t) dy_{t+h}. \quad (5.16)$$

The equation (5.16) can be rewritten by

$$E(y_{t+h}|F_t) = \int_{y_{t+h} \in Y_{t+h}} y_{t+h} \left[\int_{\theta \in \Theta} f_{\Theta}(y_{t+h}|F_t)\pi(\Theta|F_t) d\theta \right] dy_{t+h}. \quad (5.17)$$

Using properties of integer, the limits of the equation (5.17) can be replaced, thus

$$E(y_{t+h}|F_t) = \int_{\theta \in \Theta} \left[\int_{y_{t+h} \in Y_{t+h}} y_{t+h} f_{\Theta}(y_{t+h}|F_t) dy_{t+h} \right] \pi(\Theta|F_t) d\theta. \quad (5.18)$$

The equation (5.18) represent

$$\int_{\theta \in \Theta} [E(y_{t+h}|F_t, \Theta)] \pi(\Theta|F_t) d\theta = \mu_{t+h}(\Theta). \quad (5.19)$$

Hence, using the MCMC vector the $E(y_{t+h}|F_t)$ is evaluated by

$$\hat{E}(y_{t+h}|F_t) = \hat{y}_{t+h} = \frac{1}{Q} \sum_{j=1}^Q \mu_{t+h}(\theta^{(j)}), \quad (5.20)$$

where

$$g(\mu_{t+h}^{(k)}) = \beta_0^{(k)} + \sum_{i=1}^p \phi_i^{(k)} g(y_{t+h-i}) + \sum_{j=1}^q \theta_j^{(k)} [g(y_{t+h-j}) - g(\mu_{t+h-j})] \quad (5.21)$$

Note that,

$$E(y_{t+h-j}|F_t) = \begin{cases} y_{t+h-j}, & h \leq j \\ \hat{y}_{t+h-j}, & h > j \end{cases}$$

$$\hat{\mu}_{t+h-j} = \begin{cases} \hat{\mu}_{t+h-j}, & h \leq j \\ \hat{y}_{t+h-j}, & h > j \end{cases}$$

These predictions represent the transformed series, however the real interest is forecasting the original series.

$$E[y_t|F_t] = E[\lambda(y_t^{(\lambda)} + 1)^{\frac{1}{\lambda}}|F_t], \quad (5.22)$$

The properties of MCMC will be used to evaluate the original predictions. Using the estimate mean is possible apply the inverse transformation, thus obtaining the original prediction.

$$\begin{aligned}\mu_t^{(k)} &= [\lambda\mu_t^{(k)} + 1]^{\frac{1}{\lambda}} \\ \mu_{t+h}^{(k)} &= [\lambda\hat{\mu}_{t+h}^{(k)} + 1]^{\frac{1}{\lambda}}\end{aligned}\quad (5.23)$$

These predictions are an important contribution of this work. Note that the original predictions are obtained without any supposition or expansion, a crucial gain of the Bayesian approach.

The confidence intervals for $\hat{\mu}_{t+h}$ can be evaluated using the MCMC sample to calculate $\{\mu_{t+h}^{(j)} = g^{-1}(\eta_{t+h}(\gamma^{(j)}))\}$, with $j = 1, \dots, M$. Once calculated M values of the prediction $\hat{\mu}_{t+h}$, the percentile $100\sigma\%$ and $100(1 - \sigma)\%$ are selected, and using 5.23 we obtained the confidence interval for $\hat{\mu}_{t+h}$. The next section contains all the Bayesian simulation study. We use the Bayesian selection criterion to verify the quality of choice for each TGARMA model.

5.3 Simulation Study

In this section we present a simulation study for GARMA(1,1) with gamma and inverse Gaussian. Tables 5.1 and 5.3 show these results.

The performance of the Bayesian estimation was evaluated using three metrics: the corrected bias (CB), the corrected error (CE) and the mean acceptance rates in the MCMC algorithm called Acceptance Probabilities (AP). These metrics are defined as,

$$\begin{aligned}CB &= \frac{1}{m} \sum_{i=1}^m \left| \frac{\theta - \hat{\theta}^{(i)}}{\theta} \right|, \\ CE^2 &= \frac{1}{\tau^2} \frac{1}{m} \sum_{i=1}^m (\hat{\theta}^{(i)} - \theta)^2 \\ AP &= \frac{1}{m} \sum_{i=1}^m \hat{r}^{(i)},\end{aligned}$$

where $\hat{\theta}^{(i)}$ and $\hat{r}^{(i)}$ are the estimate of parameter θ and the computed acceptance rate, respectively, for the i -th replication, $i = 1, \dots, m$. In this paper we take the posterior means of θ as point estimates. Furthermore, the variance term (τ^2) that appears in the definition of CE is the sample variance of $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(m)}$.

Therefore, we opted to chose parameter values that would generate moderate values for the time series. The Table 5.2 present the proportions of correct model chosen using Bayesian criterion with Gamma GARMA(p,q) model. We conduced the study with different values of λ .

The experiment was replicated $m = 1000$ times for each model. For each dataset, we used the prior distributions as described in Section 3 with mean zero and variance 200. We then drew samples from the posterior distribution discarding the first 1000 draws as burn-in and keeping every 3rd sampled value, resulting in a final sample of 5000 values. All the computations were implemented using the open-source statistical software language and environment R (R Development Core Team (2010)).

Table 5.1: TGARMA(1,1) with gamma distribution and Box-Cox power transformation

Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.30	0.3025	0.0039	0.1504	0.9982	0.5939
ν	0.50	0.5032	0.0011	0.0528	1.0023	0.7414
α_0	0.70	0.6970	0.0277	0.1793	0.9976	0.6302
α_1	0.50	0.4970	0.0019	0.0718	0.9996	0.5553
ϕ_1	0.30	0.3008	0.0016	0.1036	0.9976	0.5863
Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.50	0.5052	0.0068	0.1266	1.0015	0.6771
ν	0.50	0.5030	0.0010	0.0511	1.0040	0.7455
α_0	0.70	0.7027	0.0281	0.1882	0.9996	0.6370
α_1	0.50	0.4997	0.0018	0.0689	0.9995	0.5574
ϕ_1	0.30	0.3008	0.0016	0.1069	1.0015	0.5854
Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.70	0.7095	0.0110	0.1208	1.0006	0.7276
ν	0.50	0.5076	0.0013	0.0598	1.0182	0.7450
α_0	0.70	0.7011	0.0316	0.2019	0.9964	0.6384
α_1	0.50	0.4978	0.0019	0.0722	0.9976	0.5611
ϕ_1	0.30	0.3008	0.0015	0.1096	0.9966	0.5873
Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.90	0.8783	0.0107	0.0969	1.0203	0.7634
ν	0.50	0.5114	0.0010	0.0550	1.0581	0.7401
α_0	0.70	0.6581	0.0217	0.1778	1.0382	0.6270
α_1	0.50	0.4925	0.0018	0.0685	1.0140	0.5553
ϕ_1	0.30	0.3020	0.0019	0.1147	0.9991	0.5831

Table 5.2: Proportions of correct model chosen using Bayesian criterion with Gamma GARMA(p,q) model

Model	$\lambda = 0.3$		$\lambda = 0.5$	
	GARMA(1,1)	GARMA(2,2)	GARMA(1,1)	GARMA(2,2)
EBIC	0.9820	0.4640	0.9920	0.4220
DIC	0.7900	0.7660	0.7940	0.7760
CPO	0.4260	0.7860	0.4300	0.8040
Model	$\lambda = 0.7$		$\lambda = 0.9$	
	GARMA(1,1)	GARMA(2,2)	GARMA(1,1)	GARMA(2,2)
EBIC	0.9880	0.4900	0.9890	0.4540
DIC	0.8080	0.7860	0.7800	0.7510
CPO	0.4800	0.7800	0.4860	0.7950

Table 5.3: TGARMA(2,2) with gamma distribution and Box-Cox power transformation

Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.30	0.3013	0.0010	0.0762	0.9971	0.3535
ν	0.50	0.5060	0.0008	0.0474	1.0170	0.7407
α_0	0.50	0.5332	0.0398	0.3074	1.0100	0.4085
α_1	0.30	0.2790	0.0178	0.3406	1.0084	0.2771
α_2	-0.20	-0.2064	0.0020	0.1810	1.0062	0.5685
ϕ_1	0.40	0.4184	0.0182	0.2566	1.0055	0.2478
ϕ_2	-0.30	-0.2788	0.0140	0.2956	1.0121	0.2421
Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.50	0.5031	0.0020	0.0654	0.9989	0.4252
ν	0.50	0.5043	0.0009	0.0498	1.0063	0.7521
α_0	0.50	0.5289	0.0380	0.3019	1.0076	0.4871
α_1	0.30	0.2841	0.0169	0.3319	1.0040	0.3257
α_2	-0.20	-0.2055	0.0020	0.1814	1.0042	0.5764
ϕ_1	0.40	0.4139	0.0171	0.2461	1.0023	0.3301
ϕ_2	-0.30	-0.2838	0.0133	0.2903	1.0064	0.3457
Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.70	0.7015	0.0036	0.0660	0.9961	0.5057
ν	0.50	0.5064	0.0009	0.0501	1.0174	0.7378
α_0	0.50	0.5321	0.0406	0.3102	1.0085	0.4414
α_1	0.30	0.2822	0.0175	0.3347	1.0048	0.3271
α_2	-0.20	-0.2076	0.0018	0.1732	1.0118	0.6342
ϕ_1	0.40	0.4153	0.0173	0.2483	1.0026	0.2621
ϕ_2	-0.30	-0.2813	0.0133	0.2905	1.0089	0.2928
Parameter	Real value	Mean	Variance	CB	CE	AP
λ	0.90	0.9025	0.0055	0.0642	0.9948	0.5932
ν	0.50	0.5068	0.0009	0.0520	1.0177	0.7285
α_0	0.50	0.5347	0.0442	0.3131	1.0079	0.5121
α_1	0.30	0.2793	0.0156	0.3093	1.0075	0.2318
α_2	-0.20	-0.2069	0.0016	0.1654	1.0087	0.5478
ϕ_1	0.40	0.4191	0.0157	0.2304	1.0060	0.2114
ϕ_2	-0.30	-0.2807	0.0126	0.2741	1.0089	0.2407

This section confirms the good performance of the Bayesian methods. The Tables 5.1 and 5.3 presented acceptable values for CB, CE and AP, that should be near 0, 1 and between 0.30 and 0.80, respectively. The Table 5.2 present

higher proportions indicating the corrected model, with different values of λ . We presented the results of GARMA(1,1) and GARMA(2,2) but similar results were obtained with the orders (1,2),(2,1). We also conducted the same study using the inverse Gaussian distribution, similar results were obtained.

In all cases we used the diagnostic proposed by Geweke (1992) to assess convergence of the chains. This is based on a test for equality of the means of the first and last parts of the chain (by default the first 10% and the last 50). If the samples are drawn from the stationary distribution, the two means are equal and the statistic has an asymptotically standard normal distribution. The calculated values of Geweke statistics were all between -2 and 2, which is an indication of convergence of the Markov chains.

5.4 Real data analysis

The demography of Sweden is monitored by Statistics Sweden (SCB). As of 31 December 2013, Sweden's population was estimated to be 9.64 million people, making it the 90th most populous country in the world. The three biggest cities are Stockholm, Gothenburg and Malmö. Approximately 85% of the country's population resides in urban areas.

The real data set represents the Annual Swedish fertility rates (1000's) 1750-1849. We obtained the data set on <https://datamarket.com/data/set/22s2>. Figures 5.1 and 5.2 present the graph and Auto Correlation Function and Partial Auto Correlation Function of Annual Swedish fertility rates respectively.

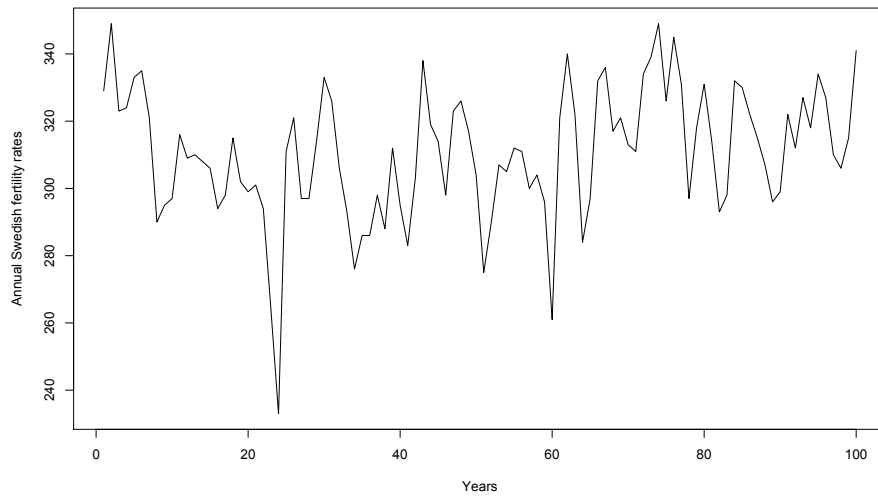


Figure 5.1: Graph of Annual Swedish fertility rates

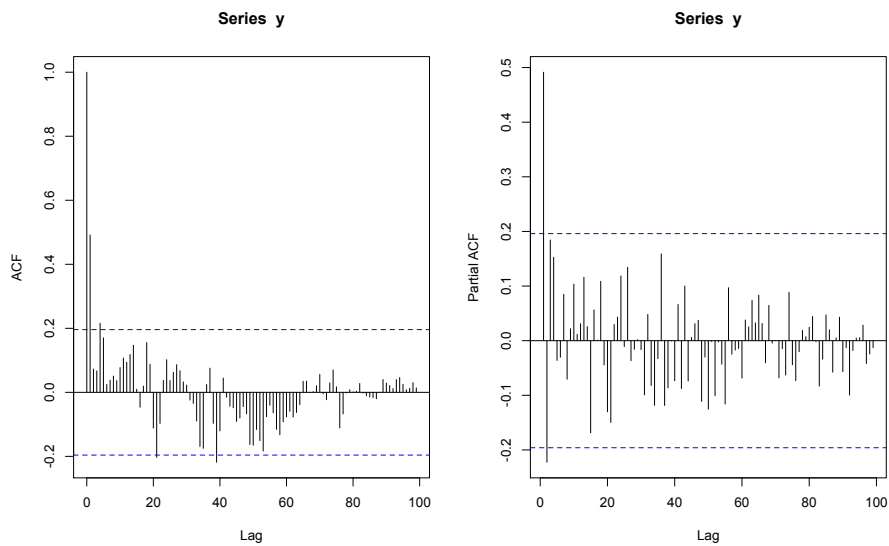


Figure 5.2: Auto Correlation Function and Partial Auto Correlation Function of Annual Swedish fertility rates

The Bayesian selection criterion were used to select between the gamma and inverse Gaussian and also to select the order of the model. Table 5.4 we present the results.

Table 5.4: Criterion selection using Annual Swedish fertility rates

Gamma	TGARMA(1,0)	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
EBIC	1506.72	1507.51	1579.31	1544.57	1606.73
DIC	1496.66	1497.32	1570.84	1530.09	1596.48
CPO	-299.91	-300.76	-409.56	-306.09	-435.69
Inverse Gaussian	TGARMA(1,0)	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
EBIC	1664.14	1664.89	1683.51	1681.97	1686.55
DIC	1651.29	1652.56	1667.54	1665.44	1665.82
CPO	-336.96	-337.35	-381.22	-374.79	-392.01

The selected model was GARMA(1,0) with gamma distribution as we can see in Table 5.4. Table 5.5 presents the Bayesian estimates of the model with posterior mean and standard deviation in brackets, the 95% HPD intervals and acceptance probability.

Table 5.5: Estimates of Annual Swedish fertility rates series with TGARMA(1,0) Gamma

Parameter	Posterior Mean(Sd)	Inferior HPD Bound	Superior HDP Bound	AP
β_0	0.7888(0.0537)	0.6497	0.9269	0.4928
ϕ_1	0.7157(0.0224)	0.6657	0.7650	0.4961
ν	3.4782(0.3516)	2.8195	4.1212	0.6701
λ	0.3145(0.0326)	0.2513	0.3754	0.7382

Quantile residuals are based on the idea of inverting the estimated distribution function for each observation to obtain exactly standard normal residuals. The residuals is given by $r_t = \Phi^{-1}(\mathbf{F}_{y_t}(y_t|F_{t-1}))$ where \mathbf{F}_{y_t} represent the cumulative distribution function for the respective density.

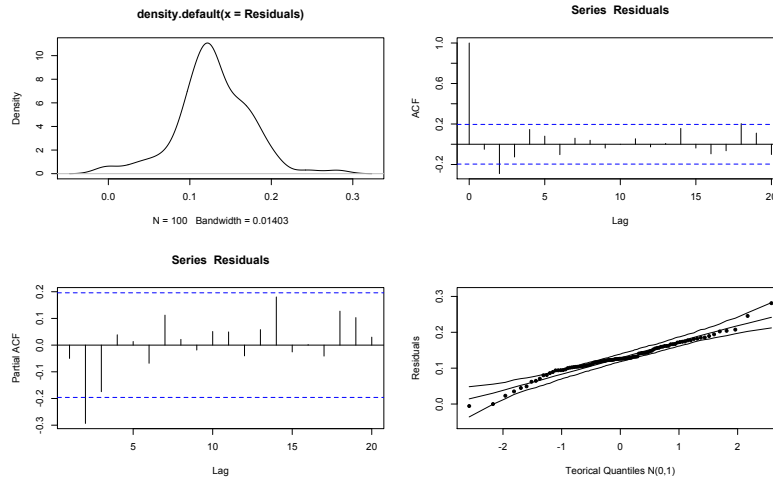


Figure 5.3: Autocorrelation function and partial autocorrelation function of the residuals of rate of Annual Swedish fertility rates series

The residual analysis is an important part to check the adequacy of the model. The Figure 5.3 confirms residuals following Gaussian distribution and non-correlated.

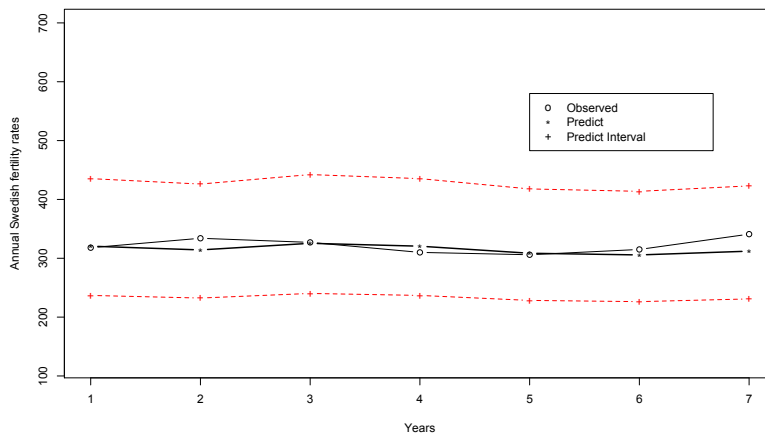


Figure 5.4: Original Predictions with GARMA(1,0) Gamma model with Rate of Annual Swedish fertility rates series

The prediction were made by the median. Only the first term of Taylor expansion was used. Using the estimate, predictions of 6 steps ahead of the original series can be made. The 6 last values of the series were removed and fitted the model without them. Figure 5.4 presents predictions one step ahead for

6 years values, thus the predicted value be compared with the true value.

The MAPE was calculated to assess the quality of predictions, the value was 03.70% which indicated good predictions. The predictions using the Gaussian distribution was evaluated, and presented higher values of MAPE.

Discussion

In the Second Chapter we have shown that the MBB algorithm is consistent, that is the quantiles derived by MBB method are equivalent with the quantiles derived from the asymptotic distribution. For GARMA models considered in this paper obtaining asymptotic distributions of the ML estimates is very complicated as it entails calculating complicated asymptotic Fisher information matrix. Thus, MBB method provides a clear advantage over the classical approach. Additionally, in Section 5 we have performed an extensive Monte Carlo study to check the coverage of MBB confidence intervals and compare it with the coverage of classical asymptotic confidence intervals. Our results clearly show the superiority of MBB method, the coverage rates are consistently better throughout our experiments.

In the Third Chapter we discuss a Bayesian approach for estimation, comparison and prediction of GARMA time series models. We analysed three different discrete models: Poisson, binomial and negative binomial. We implemented MCMC algorithms to carry out the simulation study and the methodology was also applied on three real discrete time series data.

Properties of the Bayesian estimation and the performance of Bayesian selection criteria were assessed with our simulation study. The analysis with real data also provided good estimates and predictions via parsimonious models. All in all our results suggest that, as indicated in the original GARMA paper, this class

of models have potential uses for modelling overdispersed time series count data.

In the Forth Chapter the model under study is a GARMA model that was recently introduced by Benjamin et al. (2003). There are many potential uses of GARMA model in time series analysis. For example, while analyzing discrete time series, GARMA structures allow embedding continuous type ARMA model in the parametrization of the discrete distribution (see de Andrade et al. (2016c)). Due to increased popularity of GARMA models, there is a need to find transformations letting the model correspond to data more flexibly. We propose Box-Cox transformation. Our research shows that the Partial Likelihood method can be successfully applied in finding optimal value of λ for the Box-Cox transformation. Such value, say $\hat{\lambda}$, can then be used to transform the data and use the GARMA model for statistical inference. Additionally, the moving block bootstrap technique is used to get confidence intervals for the true unknown value of λ in the Box-Cox transformation. Having confidence intervals for λ is an additional bonus as it indicates preferred subintervals of $[0,1]$ for λ .

In the Fifth Chapter we discussed a Bayesian approach for estimation, comparison and prediction of TGARMA time series models. We analyzed two different continuous models: gamma and inverse Gaussian. We implemented MCMC algorithms to carry out the simulation study and the methodology was also applied on a real time series dataset. Properties of the Bayesian estimation and the performance of Bayesian selection criteria also were assessed with our simulation study. The analysis with real data also provided good estimates and predictions via parsimonious models. Our results suggest that, as indicated in the original TGARMA paper, this class of models have potential uses for modelling non-additivity, non-normality and heteroscedasticity continuous time series.

6.1 Acknowledgments

The Author Breno Andrade gratefully acknowledges the financial support from Brazilian research agency CAPES, process number: 99999.003692/2015-09 that has allowed him to spend an academic year at CTU in Cracow. The technical assistance obtained from the Institute of Telecommunication of Cracow Technical University is also gratefully acknowledged here.

Bibliography

- Ahmad, W. M. A. W., Zakaria, S. B., Aleng, N. A., Halim, N. A., and Ali, Z. (2015). Box-cox transformation and bootstrapping approach to one sample t-test. *World Applied Sciences Journal*, 33(5):704–708. [3](#)
- Benjamin, M., Rigby, R., and Stasinopoulos, D. (2003). Generalized autoregressive moving average models. . *J. Amer. Statist. Assoc.*, 98:214–223. [2](#), [3](#), [5](#), [8](#), [9](#), [11](#), [25](#), [46](#), [47](#), [78](#)
- Biswas, A. and Song, P. X.-K. (2009). Discrete-valued ARMA processes. *Statistics and Probability Letters*, 79(17):1884–1889. [1](#)
- Box, G. and Cox, D. (1964). An analysis of transformation. . *J. R. Statist. Soc. B*, 26:211–252. [43](#), [61](#), [65](#)
- Bühlmann, P. (2002). Bootstrap for time series. *Statistical Science*, 17(1):52–72. [2](#)
- Carlin, B. P. and Louis, T. A. (2001). *Bayes and Empirical Bayes methods for data analysis*. Boca Raton: Chapman and Hall/CRC. [33](#)
- Castillo, J. and F.G., B. (2013). Improving trip forecasting models by means of the box-cox transformation. *Transportmetrica A: Transport Science*, 9:653–674. [3](#)
- Chan, K. and Ledolter, J. (1995). Monte Carlo E.M. estimation for time series models involving counts. . *Journal of the American Statistical Association*, 90:242–251. [1](#)
- Chen, M. and Shao, Q. (1998). Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, 8:69–92. [30](#)
- Chen, Z. and Yang, Y. (2004). Assessing forecast accuracy measures. *Preprint Series*, (2004-2010):2004–10. [20](#), [24](#)

-
- Cole, S. R., Chu, H., and Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *American journal of epidemiology*, 179(2):252–260. [3](#)
- Conover, W. J. (1971). *Practical nonparametric statistics*. Number 978-0471160687. John Wiley and Sons. [19](#), [23](#)
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. Wiley series in Probability and Statistics. Wiley, New York, 3rd edition. [36](#)
- Cox, D. (1975). Partial likelihood. . *Biometrika*, 62:69–76. [3](#), [64](#)
- Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. . *Scandinavian Journal of Statistics*, 8:93–115. [1](#), [43](#), [62](#)
- da Silva, M. V. G. B., Van Tassel, C. P., Sonstegard, T. S., Cobuci, J. A., and Gasbarre, L. C. (2011). Box–cox transformation and random regression models for fecal egg count data. *Frontiers in genetics*, 2. [3](#)
- Davis, R. A., Dunsmuir, W. T., and Wang, Y. (1999). Modelling time series of counts data. . *Asymptotic, Nonparametric, and Time Series*, Ed. S. Ghosh:63–114. [1](#), [2](#)
- de Andrade, B. S., Andrade, M. G., and Ehlers, R. S. (2016a). Bayesian gamma models for count data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1(4):192–205. [4](#)
- de Andrade, B. S., Leśkow, J., and Andrade, M. G. (2016b). Transformed gamma model: Properties and simulations. *Communications in Statistics - Simulation and Computation*. [3](#)
- de Andrade, S. B., Leśkow, J., and Andrade, M. G. (2016c). Moving block quantile residual bootstrap in gamma models: An application for the time series of dengue case count. *61th Annual Brazilian region meeting of the International Biometrics Society, Salvador, Bahia, Brazil*. [47](#), [57](#), [78](#)
- Dehay, D., Dudek, A., and Leśkow, J. (2014). Subsampling for continuous-time almost periodically correlated processes. *Journal of Statistical Planning and Inference*, 150:142–158. [3](#)
- Draper, N. and Cox, D. (1969). On distributions and their transformation to normality. . *J. R. Statist. Soc. B*, 31:472–476. [44](#), [62](#)

-
- Dudek, A., Leśkow, J., Paparoditis, E., and Politis, D. (2014a). A generalized block bootstrap for seasonal time series. *Journal of Time Series Analysis*, 35:89–114. [3](#)
- Dudek, A. E., Leśkow, J., Paparoditis, E., and Politis, D. N. (2014b). A generalized block bootstrap for seasonal time series. *Journal of Time Series Analysis*, 35:89–114. [13](#)
- Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5:236–244. [35](#)
- Efron, B. (1979). Bootstrap methods: Another look at the jack- knife. *Ann. Statist*, 7:1–26. [2](#), [12](#)
- Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. . *Bayesian Statistics*, 4:147–167. [33](#)
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In Bayesian Statistics*, 4:169–193. [34](#), [72](#)
- Gillard, J. (2012). A generalised box–cox transformation for the parametric estimation of clinical reference intervals. *Journal of Applied Statistics*, 39:2231–2245. [3](#)
- Hamasaki, T. and Kim, S. (2007). Box and cox power-transformation to confined and censored nonnormal responses in regression. *Computational Statistics and Data Analysis*, 51:3788 – 3799. [3](#)
- Härdle, W., Horowitz, J., and Kreiss, J. P. (2003). Bootstrap methods for time series. *Institute of Statistical Rev.*, 71:435–459. [2](#)
- Huang, Z., Pang, Z., and Zhang, R. (2013). Adaptive profile-empirical-likelihood inferences for generalized single-index models. *Computational Statistics and Data Analysis*, 62:70–82. [3](#)
- Hyndman, R. J. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688. [36](#)
- Jacobs, P. and Lewis, W. (1978). Discrete time series generated by mixtures I: Correlation and runs properties. . *Journal of Royal Statistical Society*, 40:94–105. [1](#)

-
- Jung, R. C., Kukuk, M., and Liesenfeld, R. (2006). Time series of count data: Modeling, estimation and diagnostics. . *Comput. Stat. Data Anal*, 51:2350–2364. [2](#)
- Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. John Wiley & Sons, New Jersey, USA. [6](#), [45](#), [63](#)
- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Number 978-1-4419-1848-2. Springer Series in Statistics, New York. [2](#), [12](#), [13](#), [16](#)
- Leśkow, J. and Synowiecki, R. (2010). On bootstrapping periodic random arrays with increasing period. *Metrika*, 71:253–279. [2](#), [3](#), [22](#)
- Li, W. K. L. (1994). Time series model based on generalized linear models: Some further results. . *Biometrics*, 50:506–511. [6](#), [45](#), [63](#)
- Liesenfeld, R., Nolte, I., and Pohlmeier, W. (2006). Modelling financial transaction price movements: a dynamic integer count data model. *Empirical Econom.*, 30:795–825. [1](#)
- Manly, B. (1976). Exponential data transformation. . *The Statistician*, 25:37–42. [44](#), [62](#)
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London, 2nd. [26](#)
- Pankratz, A. and Dudley, U. (1987). Forecasts of power-transformed series. . *J. of Forecasting*, 6:239–248. [49](#)
- Politis, D. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219–230. [2](#)
- R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. [32](#), [70](#)
- Rydberg, T. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Finan. Econom.*, 1:2–25. [1](#)
- Sakia, R. (1992). The Box-Cox transformation technique: a review. . *The Statistician*, 41:168–178. [44](#), [62](#)
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64:583–639. [33](#)

Woodard, D. B., Matteson, D. S., and Henderson, S. G. (2011). Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics*, 5:800–828. [13](#)

Zeger, S. L. (1988). A regression model for time series of counts. . *Biometrika*, 75(4):621–629. [1](#), [2](#)

Zhu, F. (2010). A negative binomial integer-valued GARCH model. . *Journal of Time Series Analysis*, 32:54–67. [2](#)

Zhu, M. and Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics and Data Analysis*, 51:918–930. [3](#)