Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Estatística

Seleção de ordem em modelos GARMA: uma perspectiva bayesiana

Katerine Zuniga Lastra

Porto Alegre, Fevereiro de 2024.

CIP - Catalogação na Publicação

Lastra, Katerine
Seleção de ordem em modelos GARMA: uma perspectiva bayesiana / Katerine Lastra. -- 2024.
57 f.
Orientador: Guilherme Pumi.

Coorientadora: Taiane Schaedler Prass.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Matemática e Estatística, Programa de Pós-Graduação em Estatística, Porto Alegre, BR-RS, 2024.

1. Séries temporais de contagem. 2. Modelos de regressão. 3. Análise Bayesiana. 4. Cadeia de Markov com salto reversível. I. Pumi, Guilherme, orient. II. Schaedler Prass, Taiane, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Dissertação submetida por Katerine Zuniga Lastra como requisito parcial para a obtenção do título de Mestre em Estatística pelo Programa de Pós-Graduação em Estatística da Universidade Federal do Rio Grande do Sul.

Orientador:

Prof. Dr. Guilherme Pumi

Co-orientadora:

Prof. Dra. Taiane Schaedler Prass

Comissão Examinadora:

Profa. Dra. Airlane Pereira Alencar (IME - USP)

Profa. Dra. Renata Rojas Guerra (CCNE - UFSM e PPGEst - UFRGS)

Profa. Dra. Silvana Schneider (PPGEst - UFRGS)

Data de Apresentação: 16 de fevereiro de 2024

"I suppose it is tempting,
if the only tool you have is a hammer,
to treat everything as if it were a nail."
(Abraham Maslow, 1966)

Agradecimentos

Quero expressar minha profunda gratidão a todas as pessoas que tornaram possível a conclusão desta jornada acadêmica. Em primeiro lugar, agradeço aos meus amados pais, Sergio e Nieves, cujo amor incondicional e apoio incessante foram as forças motrizes por trás de todas as minhas conquistas. Seu amor foi a luz que guiou meus passos e, por isso, expresso minha eterna gratidão.

Agradeço aos meus irmãos, pelo seu carinho e apoio que foram fundamentais ao longo desta jornada, e à minha amiga Aline, por ter contribuído para que minha experiência no Brasil fosse tão maravilhosa.

Ao meu orientador, Guilherme Pumi, quero expressar minha sincera gratidão pela orientação valiosa e inspiradora. Seu comprometimento e expertise foram fundamentais para o meu desenvolvimento acadêmico e profissional. Agradeço por compartilhar seu conhecimento e por ser um mentor exemplar. E à minha coorientadora, Taiane Schaedler Prass, agradeço por sua orientação cuidadosa e contribuições significativas para o meu trabalho de pesquisa. Sua sabedoria e apoio foram essenciais para o sucesso deste projeto.

Não posso deixar de reconhecer a contribuição do Programa de Pós-Graduação em Estatística da UFRGS, bem como o suporte financeiro da CAPES, que viabilizaram a realização deste estudo.

Resumo

A estimação de modelos autorregressivos de média móvel generalizados (GARMA) baseia-se predominantemente em métodos frequentistas, principalmente naqueles baseados em verossimilhança. Em contraste, abordagens Bayesianas têm sido menos exploradas e utilizadas na literatura. No que diz respeito aos modelos GARMA de contagem, a estimação Bayesiana vem ganhando reconhecimento por seus resultados promissores. No entanto, a seleção do modelo nesse contexto frequentemente se baseia na utilização de critérios de informação. Apesar de sua prevalência, é sabido que critérios de informação Bayesianos no contexto de modelos GARMA apresentam resultados desanimadores em simulações, o que desestimula o seu uso em aplicações. Isso é particularmente verdadeiro quando se trata da habilidade de identificar corretamente modelos, mesmo com amostras de tamanho grande. Neste trabalho abordamos o problema de seleção da ordem em modelos GARMA para séries temporais de contagem sob a perspectiva Bayesiana, através da utilização da abordagem conhecida como Reversible Jump Markov Chain Monte Carlo (RJMCMC). Além de discutir detalhes da abordagem RJMCMC no contexto de modelos GARMA de contagem, apresentamos um estudo de simulações para verificar seu desempenho em amostras finitas, considerando vários modelos GARMA, sob diversos cenários. Apresentamos ainda uma análise de sensibilidade com relação a escolha das prioris utilizadas, comparando com os critérios de informação. Para ilustrar a aplicação da abordagem proposta, são utilizados dados de produção automobilística no Brasil de janeiro de 1993 a dezembro de 2013.

Abstract

The estimation of generalized autoregressive moving average (GARMA) models is predominantly based on frequentist methods, especially those grounded in likelihood. In contrast, Bayesian approaches have been less explored and utilized in the literature. Regarding GARMA count models, Bayesian estimation has gained recognition for its promising results. However, model selection in this context often relies on the use of information criteria. Despite their prevalence, it is known that Bayesian information criteria in the context of GARMA models yield discouraging results in simulations, deterring their use in applications. This is particularly true when it comes to correctly identifying models, even with large sample sizes. In this work, we address the problem of order selection in GARMA models for count time series from a Bayesian perspective, using the approach known as Reversible Jump Markov Chain Monte Carlo (RJMCMC). In addition to discussing details of the RJMCMC approach in the context of GARMA count models, we present a simulation study to assess its performance in finite samples, considering various GARMA models under different scenarios. We also conduct a sensitivity analysis regarding the choice of priors used. To illustrate the application of the proposed approach, we use data from automotive production in Brazil between January 1993 and December 2013.

Índice

Capítulo 1

Introdução

É muito comum em aplicações que os dados de interesse sejam séries temporais discretas, tipicamente resultante da contagem de algum tipo de evento ou ocorrência de interesse. A análise de séries temporais de contagem é uma área de estudo importantíssima em estatística e ciência de dados, tendo aplicações em diversas áreas. Por exemplo, (Zeger and Qaqish, 1988; Davis et al., 2000; Christou and Fokianos, 2015; Franco et al., 2019; Triantafyllopoulos et al., 2019; Roy and Karmakar, 2021) consideram a modelagem do número de certas doenças. Em atuárias, Freeland and McCabe (2004) consideram a modelagem da contagem mensal de requerentes de prestações por perda de salário, a fim de estimar a duração esperada dos requerentes no sistema e (Weiß et al., 2020; Weiß and Feld, 2020) consideram a modelagem da contagem mensal de greves nos EUA com o fim de seleção de ordem dentro de uma dada família de modelos e diagnóstico. No mercado financeiro, (Liesenfeld et al., 2006; Cui and Zheng, 2017; Weiß and Zhu, 2024) consideram a modelagem e previsão da evolução do número de transações de duas ações em um determinado período de tempo. Na área comercial, também existem diversas aplicações. Para mais detalhes, consulte (Berry and West, 2020; Homburg et al., 2021; Liu and Nason, 2023). Finalmente, na segurança, Brännäs and Johansson (1994) consideram uma aplicação do número de acidente de tráfico em uma determinada localidade e Weiß et al. (2019) consideram as contagens mensais de relatórios de crimes relacionados a roubos, e Weiß et al. (2023) o número mensal de sinistros causados por queimaduras na indústria de manufatura pesada.

Ao contrário de outras séries temporais, que se concentram em medidas contínuas, tais como temperatura ou pressão, as séries de contagem são discretas e apresentam particularidades próprias, que tornam sua análise desafiante. Por exemplo, a distribuição das ocorrências tipicamente apresenta algum grau de assimetria, sendo obviamente não-gaussianas. A menos que severas restrições sejam feitos nos parâmetros, uma abordagem via modelos lineares tradicionais não é diretamente aplicável em séries temporais de contagem. Adicionalmente, em muitas aplicações de interesse prático existe a influência de fatores externos, como feriados, eventos sazonais ou outras condições que podem afetar a modelagem das séries. A compreensão das características dessas séries e a aplicação de técnicas de modelagem adequadas é crucial para a análise de séries temporais de contagem.

Na análise de séries temporais de contagem, os modelos estatísticos para dados dependentes podem ser classificados em modelos da classe parameter-driven e modelos da classe observation-driven, conforme a classificação proposta por Cox et al. (1981). Modelos da classe parameter-driven estendem os modelos lineares generalizados, integrando um processo autorregressivo latente no desfecho médio condicional do processo de contagem. Por outro lado, os modelos da classe observation-driven se baseiam diretamente na contagem observada em cada intervalo para discernir a dinâmica temporal, especificando um modelo para a distribuição da contagem a cada momento.

Diversas abordagens têm sido propostas para análises de séries temporais de contagem. Uma abordagem bastante difundida é dos modelos da família INAR(1), incluindo modelos INAR(1) de coeficiente aleatório (Yu et al., 2019), o modelo INAR(1) modificado (Khoo et al., 2017), com equi-dispersão, subdispersão e superdispersão (Bourguignon and Weiß, 2017; Pavlopoulos and Karlis, 2008) e modelos baseados em thinning, (Liu and Zhu, 2020). Por outro lado, uma das classes mais promissoras de modelos para séries temporais de contagem, e objeto de estudo deste trabalho, são os chamados modelos autorregressivos de média móvel generalizados - GARMA (Generalized Autorregressive Moving Average), sistematizado sob esse nome em Benjamin et al. (2003). Tais modelos são da classe observation-driven e compreendem basicamente uma combinação de modelos lineares generalizados (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) com modelos de séries temporais tradicionais. Nesse contexto, devemos especificar o que chamamos de componentes aleatória e sistemática do modelo. A componente aleatória modela a distribuição da resposta Y_t condicionada à informação observada até o tempo $t-1$. Já a componente sistemática é responsável pela dinâmica evolutiva de alguma medida de interesse, como por exemplo, a média, mediana, quantil, etc.

No contexto frequentista, a estimação dos parâmetros em modelos GARMA é tipicamente feita utilizando-se máxima verossimilhança condicional (Rocha and Cribari-Neto, 2009; Bayer et al., 2017; Melo and Alencar, 2020) ou parcial (Pumi et al., 2019, 2021, 2023), enquanto que seleção de modelos é feita usando-se critérios de informação ou ainda utilizando-se abordagem similar ao método de Box-Jenkings. No contexto bayesiano, Grande et al. (2023) estudam a estimação Bayesiana para os modelos βARMA (Rocha and Cribari-Neto, 2009), enquanto de Andrade et al. (2015) e Ehlers (2019) trabalham no contexto GARMA para séries temporais de contagem. Já a seleção de modelos é feita usando-se critérios de informações adaptados ao contexto Bayesiano ou utilizando-se uma abordagem baseada em Reversible Jump Markov Chain Monte Carlo (RJMCMC) introduzio por Green (1995). Esta abordagem, que será vista em detalhes na Seção 3.1, permite que a geração de amostras de uma distribuição alvo em um espaço de dimensão variável.

Considerando apenas a literatura de RJMCMC, Campbell (2004) propuseram uma metodologia para seleção de modelos bayesianos em modelos SETAR (Self-Exciting Threshold Autoregressive) usando RJMCMC. Em Troughton (1999) é introduzido um método RJMCMC com dois estágios para a seleção de modelos ARMA (autoregressive moving average), generalizado para o contexto de modelos com longa dependência por Eğrioğlu and Günay (2010). Por fim, Karakuş et al. (2016, 2017) contribuíram com técnicas de estimação bayesiana para algumas generalizações de modelos ARMA. Fan et al. (2009) introduziram uma estratégia para simplificar simulações em RJMCMC, enquanto Meyer-Gohde and Neuhoff (2015) empregaram o RJMCMC em um modelo de crescimento neoclássico (ver Solow (1956)), contestando suposições AR(1) convencionais. Porém, no contexto de modelos GARMA, o único trabalho que conhecemos nessa linha é o de Casarin et al. (2010), onde os autores apresentaram um algoritmo RJMCMC para inferência e seleção bayesianas em processos βAR específicos.

Objetivo

Existe muito criticismo quanto ao uso de critérios de informação para seleção de modelos, mas as vezes eles são úteis. Não é o caso para modelos GARMA: os resultados apresentados em de Andrade et al. (2015) são desapontadores com baixa poder de identificação dos modelos corretos mesmo com amostras de tamanho grande. Porem o presente estudo tem por objetivo propor e estudar a seleção de modelos GARMA de contagem no contexto Bayesiano utilizando uma abordagem de RJMCMC, incluindo uma análise de sensibilidade em relação à escolha dos hiper-parâmetros das prioris utilizadas.

Novidades do trabalho

Neste trabalho, introduzimos e estudamos o RJMCMC para seleção dos modelos GARMA de contagem, e a estimação dos parâmetros, realizando uma análise de sensibilidade em relação às priores utilizados. Além disso, fornecemos a implementação do código em Nimble, que pode ser desafiadora, mas destaca-se pela sua robustez e poder. Para avaliar a eficácia da abordagem proposta, realizamos simulações abrangentes e comparativos com critérios de informação estabelecidos. Também ilustramos a utilidade de nossa proposta por meio de sua aplicação em conjuntos de dados sobre a produção automobilística no Brasil, no período de janeiro de 1993 a dezembro de 2013, enfatizando sua relevância prática.

Suporte computacional

Na parte computacional, implementamos o código em R (R Core Team, 2020), utilizando o pacote `Nimble` (de Valpine et al., 2017), uma poderosa ferramenta para modelagem estatística e amostragem RJMCMC, que simplifica a especificação e a estimação de modelos estatísticos bayesianos e de verossimilhança. Nimble destaca-se como uma ferramenta flexível e poderosa para modelagem estatística em R, sua flexibilidade na especificação do modelo permite atender a necessidades específicas, enquanto a capacidade de combinar algoritmos de amostragem pode resultar em uma maior eficiência na estimativa de parâmetros. Além disso, sua integração perfeita com o R permite aproveitar as capacidades estatísticas e de manipulação de dados do R em conjunto com as funcionalidades de modelagem do Nimble. Com um desenvolvimento ativo, Nimble é adequado para modelar cenários estatísticos complexos, incluindo modelos hierárquicos e estruturas não convencionais. No entanto, é importante notar que, como qualquer ferramenta estatística, o Nimble pode ter uma curva de aprendizado, especialmente para aqueles que não estão familiarizados com a programação em R ou com a modelagem bayesiana. Embora o Nimble ofereça uma documentação sólida, alguns usuários podem sentir que informações específicas ou exemplos detalhados são limitados em comparação com ferramentas mais estabelecidas tornando desafiador a implementação do código. Além disso, em comparação com ferramentas mais populares como Stan, JAGS ou WinBUGS, o Nimble pode não ser tão amplamente reconhecido ou utilizado na comunidade estatística. No entanto, é importante ressaltar que a popularidade não necessariamente reflete a qualidade ou eficácia da ferramenta.

Organização do trabalho

No Capítulo 2, apresentamos e discutimos de forma concisa os modelos GARMA para séries temporais de contagem. No Capítulo 3, aprofundamos a discussão sobre o método Reversible Jump Markov Chain Monte Carlo. Para concluir, no Capítulo 4, destacamos as principais conclusões alcançadas neste trabalho, juntamente com propostas para pesquisas futuras. O artigo completo com os resultados da pesquisa pode ser encontrado no Apêndice A.

Capítulo 2

MODELOS GARMA

Modelos GARMA combinam a simplicidade e robustez dos modelos ARMA, já tradicionais em séries temporais, com a flexibilidade dos modelos lineares generalizados. Estas duas abordagens são combinadas em um modelo do tipo observation-driven, onde a componente aleatória modela as características distribucionais do modelo, condicionado à informação observada até o tempo t, enquanto a componente determinística modela a dependência serial do modelo a partir de alguma medida de interesse, como a média, a mediana ou quantil condicional, utilizando-se uma função de ligação ao estilo de modelos lineares generalizados.

2.1 Especificação do Modelo GARMA

Seja $\{Y_t\}_{t\in\mathbb{Z}}$ um processo estocástico de interesse e seja $\{\boldsymbol{x}_t\}_{t\in\mathbb{Z}}$ um conjunto de covariáveis exógenas r-dimensionais a serem incluídas no modelo. Seja $\mathscr{F}_t = \sigma\left\{\boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, \cdots, Y_{t-1}, Y_{t-2}, \cdots\right\}$ a informação disponível ao observador no tempo t. Em Benjamin et al. (2003), os modelos GARMA consideram que a distribuição de Y_t, dada a informação observada até o tempo t, pertence à família exponencial na forma canônica, ou seja, podemos escrever a densidade ou massa de probabilidade condicional como

$$f\big(y;\omega_t,\varphi|\mathscr{F}_{t-1}\big) = \exp\bigg\{\frac{y\omega_t - b(\omega_t)}{\varphi} + c(y,\varphi)\bigg\}, \tag{2.1}$$

onde, ω_t e φ são os parâmetros canônico e de escala, respectivamente, com $b(\cdot)$ e $c(\cdot)$ sendo funções específicas que definem a família exponencial particular. Nos modelos GARMA tradicionais, ω_t depende do tempo, enquanto φ não, o que se reflete na notação. A média condicional e a variância de Y_t, dado $\mathscr{F}_{t-1}$, são dados por $\mu_t = E\big(Y_t|\mathscr{F}_{t-1}\big) = b'\big(\omega_t\big)$ e $\mathrm{Var}\,\big(Y_t|\mathscr{F}_{t-1}\big) = \varphi b''\big(\omega_t\big) = \varphi V\big(\mu_t\big)$, para $t \in \{1,\cdots,n\}$. No componente sistemático do modelo, a média condicional μ_t está relacionada ao preditor linear através de uma função de ligação g monótona e duas vezes diferenciável. A estrutura mais comumente usada para o componente sistemático inclui covariáveis e uma estrutura ARMA da forma

$$\eta_t = g(\mu_t) = \alpha + \boldsymbol{x}'_t\boldsymbol{\beta} + \sum_{j=1}^{p}\phi_j\big[g\big(Y_{t-j}\big) - \boldsymbol{x}'_{t-j}\boldsymbol{\beta}\big] + \sum_{j=1}^{q}\theta_j\big[g\big(Y_{t-j}\big) - \eta_{t-j}\big], \tag{2.2}$$

em que η_t é o preditor linear, α é um intercepto, $\boldsymbol{\beta} = (\beta_1,\cdots,\beta_r)'$ é o vetor de parâmetros relacionado às covariáveis, $\boldsymbol{\phi} = (\phi_1,\cdots,\phi_p)'$ e $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_q)'$ são os coeficientes AR e MA, respectivamente. Um modelo GARMA(p,q) é definido por (2.1) e (2.2). A componente sistemática (2.1) pode ser contínua ou discreta. No caso de séries temporais de contagens, os modelos mais comumente aplicados são revistos na próxima seção.

2.2 Modelos GARMA para séries temporais de contagens

Nesta seção apresentamos os modelos GARMA para dados de contagem mais comumente estudados na literatura, a saber, Poisson GARMA, binomial GARMA e binomial negativa GARMA.

2.2.1 Modelo Poisson GARMA

Quando $Y_t|\mathscr{F}_{t-1}$ segue uma distribuição de Poisson com media μ_t temos

$$f\left(y|\mathscr{F}_{t-1}\right) = \exp\left\{y\log(\mu_t) - \mu_t - \log(y!)\right\} I(y \in \mathbb{N}). \tag{2.3}$$

Portanto, $Y_t|\mathscr{F}_t$ pertence à família exponencial canônica com

$$\varphi = 1, \quad \omega_t = \log(\mu_t), \quad b(\omega_t) = \exp\{\omega_t\}, \quad c(y_t, \varphi) = -\log(y_t!), \quad \mu_t = \exp(\omega_t),$$

e $V(\mu_t) = \mu_t$. O preditor linear é prescrito como

$$\log(\mu_t) = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{j=1}^{p}\phi_j\left[\log(Y_{t-j}^*) - \boldsymbol{x}_{t-j}'\beta\right] + \sum_{j=1}^{q}\theta_j\left[\log(Y_{t-j}^*) - \log(\mu_{t-j})\right], \tag{2.4}$$

onde $Y_t^* = \max\{Y_t, c\}$, e $0 < c < 1$ é uma constante definida pelo usuário cuja função é evitar problemas numéricos com o logaritmo quando y_t é 0. O modelo Poisson GARMA é definido por (2.3) e (2.4). A forma (2.4) possui problemas devido à introdução da constante c, não sendo uma unanimidade na literatura. Davis et al. (1999) apresentam uma investigação em relação à formulação do preditor, incorporando a utilização dos resíduos de Pearson para os termos de médias móveis. Por outro lado, Brumback et al. (2000) introduziu uma proposta semelhante ao GARMA-Poisson, denominada TGLM (Transformed Generalized Linear Models), que oferece quatro abordagens distintas para a construção do preditor. Nesse trabalho foi considerado a abordagem de Davis et al. (1999).

2.2.2 Modelo Binomial GARMA

Quando $Y_t|\mathscr{F}_{t-1} \sim B\left(m, p_t\right)$, com $m > 0$ conhecido, e $\mu_t = \mathbb{E}(Y_t|\mathscr{F}_{t-1}) = mp_t$ temos

$$\begin{aligned} f\left(y;\mu_t|\mathscr{F}_{t-1}\right) &= \exp\left\{y\log\left(\frac{p_t}{1-p_t}\right) + m\log(1-p_t) + \log\left(\frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\right)\right\} \\ &= \exp\left\{y\log\left(\frac{\mu_t}{m-\mu_t}\right) + m\log\left(\frac{m-\mu_t}{m}\right) + \log\left(\frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\right)\right\} \end{aligned} \tag{2.5}$$

que é um membro da família exponencial canônica com $\varphi = 1$,

$$\omega_t = \log\left(\frac{\mu_t}{m-\mu_t}\right), \; b(\omega_t) = m\log\left(\frac{m}{m-\mu_t}\right), \; \text{and} \;\; c(y_t, \varphi) = \log\left(\frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\right).$$

O modelo binomial GARMA é especificado por (2.4) e (2.5). Em particular,

$$\mu_t = \frac{m\exp(p_t)}{1+\exp(p_t)} \quad \text{e} \quad V(\mu_t) = \frac{\mu_t(m-\mu_t)}{m}.$$

2.2.3 Modelo Binomial Negativo GARMA

No modelo binomial negativo GARMA, consideramos a distribuição de probabilidade condicional como sendo a distribuição binomial negativa, ou seja, $Y_t|\mathscr{F}_{t-1} \sim \mathrm{NB}(k, p_t)$, com $k > 0$ conhecido. Neste caso $\mathbb{E}(Y_t|\mathscr{F}_{t-1}) = \frac{k(1-p_t)}{p_t}$, de modo que $p_t = \frac{k}{\mu_t + k}$ e daí

$$f(y; \mu_t|\mathscr{F}_{t-1}) = \exp\left\{k \log\left(\frac{k}{\mu_t + k}\right) + y \log\left(\frac{\mu_t}{\mu_t + k}\right) + \log\left(\frac{\Gamma(k+y)}{\Gamma(y+1)\Gamma(k)}\right)\right\}, \tag{2.6}$$

que pertence à família exponencial com

$$\varphi = 1, \quad \omega_t = \log\left(\frac{\mu_t}{\mu_t + k}\right), b(\omega_t) = -k \log\left(\frac{k}{\mu_t + k}\right), \quad c(y_t, \varphi) = \log\left(\frac{\Gamma(k+y)}{\Gamma(y+1)\Gamma(k)}\right).$$

As equações (2.4) e (2.6) especificam o modelo GARMA binomial negativo com

$$\mu_t = \frac{k \exp(p_t)}{1 + \exp(p_t)} \quad \text{e} \quad V(\mu_t) = \frac{\mu_t(k + \mu_t)}{k}.$$

2.3 Abordagem Bayesiana para modelagem GARMA

Nesta seção, tratamos da estimação Bayesiana no contexto de modelos GARMA estudados neste trabalho. Considerando a especificação geral (2.1), a função de verossimilhança para o modelo é dada por

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t) \propto \exp\left\{\sum_{t=s+1}^{n} \frac{y_t\theta_t - b(\omega_t)}{\varphi} + c(y_t, \varphi)\right\},$$

em que s é o ponto de partida da função de verossimilhança, geralmente assumido como $s = 0$, como em Benjamin et al. (2003) e Pumi et al. (2019), mas às vezes tomado como $s = \max\{p, q\}$, como em Rocha and Cribari-Neto (2009) e de Andrade et al. (2015). Neste trabalho, utilizaremos $s = 0$. Vamos assumir distribuições a priori normais com média zero e variância σ^2 para os parâmetros $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ e α_0, ou seja, tomamos $\phi_i \sim N(0, \sigma^2)$, $\theta_j \sim N(0, \sigma^2)$, $\beta_j \sim N(0, \sigma^2)$, e $\alpha_0 \sim N(0, \sigma^2)$, para $i \in 1, \cdots, p$ e $j = 1, \cdots, q$. Assumindo independência entre os parâmetros, a distribuição conjunta a priori é

$$\pi_0(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2\right)\right\}. \tag{2.7}$$

Com a especificação a priori acima, a distribuição a posteriori para o modelo se torna

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t) \propto \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t)\pi_0(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0).$$

A seguir, apresentamos fórmulas explícitas para as funções de verossimilhança de cada modelo estudado.

Modelo Poisson GARMA

Suponha que $\{Y_t\}_{t\in\mathbb{Z}}$ represente uma série temporal de dados de contagem, com distribuição condicional Poisson com média μ_t, ou seja, com massa de probabilidade condicional dada por (2.3). Neste caso, a função de verossimilhança para o modelo Poisson GARMA é dada por

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | \mathscr{F}_t) = \exp\left\{ \sum_{t=1}^{n} \Big[y_t \log(\mu_t) - \mu_t - \log(y_t!) \Big] \right\},$$

com

$$\mu_t = \exp\left\{ \alpha_0 + \sum_{i=1}^{p} \phi_i \log(y^*_{t-i}) + \sum_{j=1}^{q} \theta_j \log\left(\frac{y^*_{t-j}}{\mu_{t-j}} \right) \right\}.$$

De (2.7), a densidade a posteriori para o modelo Poisson GARMA é dada por

$$\begin{aligned}\pi\left(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | \mathscr{F}_t\right) \propto \exp\Bigg\{ & \sum_{t=1}^{n} \left[y_t \left(\alpha_0 + \sum_{i=1}^{p} \phi_i \log(y^*_{t-i}) + \sum_{j=1}^{q} \theta_j \log\left(\frac{y^*_{t-j}}{\mu_{t-j}} \right) \right) \right] \\ & - \sum_{t=1}^{n} \exp\left\{ \alpha_0 + \sum_{i=1}^{p} \phi_i \log(y^*_{t-i}) + \sum_{j=1}^{q} \theta_j \log\left(\frac{y^*_{t-j}}{\mu_{t-j}} \right) \right\} \\ & - \frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2 \right) \Bigg\}.\end{aligned}$$

Modelo Binomial GARMA

Quando $Y_t|\mathscr{F}_{t-1} \sim B(m, p_t)$, isto é, distribuição Binomial com parâmetro $m > 0$ conhecido e $p_t \in (0, 1)$ representando a probabilidade de sucesso no tempo t, a massa de probabilidade condicional é dada em (2.5), de forma que a função de verossimilhança condicional se torna

$$\begin{aligned}\mathcal{L}(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, |\mathscr{F}_t) &= \prod_{t=1}^{n} \exp\Bigg\{ y_t \log\left(\frac{\mu_t}{m - \mu_t} \right) + m \log\left(\frac{m - \mu_t}{m} \right) \\ &\qquad + \log\left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m - y_t + 1)} \right) \Bigg\} \\ &\propto \exp\left\{ \sum_{t=1}^{n} \left[y_t \log\left(\frac{\mu_t}{m - \mu_t} \right) + m \log\left(\frac{m - \mu_t}{m} \right) \right] \right\}.\end{aligned}$$

A densidade a posteriori para o modelo binomial GARMA é dada por

$$\begin{aligned}\pi\left(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | \mathscr{F}_t\right) \propto \exp\Bigg\{ & \sum_{t=1}^{n} \left[y_t \log\left(\frac{\mu_t}{m - \mu_t} \right) + m \log\left(\frac{m - \mu_t}{m} \right) \right] + \\ & - \frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2 \right) \Bigg\}.\end{aligned}$$

Modelo Binomial Negativa GARMA

Quando Y_t segue uma distribuição Binomial negativa com parâmetro $k > 0$ conhecido e p_t representando a probabilidade de ocorrência do evento em estudo no tempo t, a massa de probabilidade condicional é dada por (2.6). Portanto, a função de verossimilhança condicional para o modelo binomial negativo GARMA é dada por

$$\mathcal{L}(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, |\mathscr{F}_t) = \prod_{t=1}^{n} \exp\left\{ k \log\left(\frac{k}{\mu_t + k}\right) + y_{\mathrm{t}} \log\left(\frac{\mu_{\mathrm{t}}}{\mu_{\mathrm{t}} + k}\right) + \log\left(\frac{\Gamma(k + y_t)}{\Gamma(y_{\mathrm{t}} + 1)\Gamma(k)}\right) \right\}$$
$$\propto \exp\left\{ k \sum_{t=1}^{n} \log\left(\frac{k}{\mu_{\mathrm{t}} + k}\right) + \sum_{t=1}^{n} y_t \log\left(\frac{\mu_{\mathrm{t}}}{\mu_{\mathrm{t}} + k}\right) \right\},$$

De (2.7), a densidade a posteriori para o modelo binomial negativo GARMA é dada por

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t) \propto \exp\left\{ \sum_{t=1}^{n} \left[y_t \log\left(\frac{\mu_t}{m - \mu_t}\right) + m \log\left(\frac{m - \mu_t}{m}\right) \right] + \right.$$
$$\left. - \frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2 \right) \right\}.$$

Capítulo 3

Método Computacional

Neste capítulo, apresentaremos o método computacional Reversible Jump Markov Chain Monte Carlo (RJMCMC). Explicaremos concisamente como esse método é usado para simular amostras da distribuição a posteriori e nos aprofundaremos em como funciona seu algoritmo de amostragem. Ao longo do capítulo, também exploraremos em detalhes as diversas aplicações e vantagens desse método em séries temporais de contagem.

3.1 Reversible Jump Markov Chain Monte Carlo

O método conhecido como Reversible Jump Markov chain Monte Carlo, introduzido por Green (1995), é uma extensão do algoritmo Metropolis-Hastings que permite a geração de amostras de uma distribuição alvo em espaços de diferentes dimensões. A dimensão do espaço de parâmetros pode variar entre as iterações permitindo que este método seja utilizado como um método Bayesiano para seleção de modelos.

De acordo com Green (1995), em um contexto de modelagem Bayesiana, seja $\{M_k, k \in K\}$ uma coleção enumerável de modelos candidatos, onde o índice k atua como uma variável indicadora auxiliar do modelo, e K representa o escopo dos modelos considerados. Ao modelo M_k é tipicamente associado um vetor de n_k parâmetros desconhecidos, digamos $\boldsymbol{\xi}_k \in \mathbb{R}^{n_k}$, que podem assumir diferentes valores para modelos diferentes. Existe uma estrutura hierárquica natural expressa pela modelagem da distribuição conjunta de $(k, \boldsymbol{\xi}_k, y)$ como

$$p(k, \boldsymbol{\xi}_k, y) \propto p(k)p(\boldsymbol{\xi}_k|k)p(y|\boldsymbol{\xi}_k, k).$$

A inferência Bayesiana sobre k e $\boldsymbol{\xi}_k$ será baseada na distribuição a priori $p(k, \boldsymbol{\xi}_k|y)$, resultando em uma distribuição a posteriori dada por

$$p(\boldsymbol{\xi}_k|y, k) \propto p(y|\boldsymbol{\xi}_k, k)p(\boldsymbol{\xi}_k|k),$$

onde $p(y|\boldsymbol{\xi}_k, k)$ e $p(\boldsymbol{\xi}_k|k)$ representam o modelo de probabilidade e a distribuição a priori dos parâmetros do modelo M_k, respectivamente. Então a probabilidade a posteriori é dado como,

$$p(k, \boldsymbol{\xi}_k|y) \propto p(k)p(\boldsymbol{\xi}_k|k, y).$$

A densidade conjunta a posteriori é o alvo do amostrador RJMCMC sobre o espaço de estados $\Theta = \cup_{k \in K}\big(k, \mathbb{R}^{n_k}\big)$. Dentro de cada iteração, o algoritmo RJMCMC atualiza os parâmetros de acordo com a ordem do modelo e, em seguida, a ordem do modelo de acordo com os parâmetros. Se o estado atual da cadeia de Markov for $(k, \boldsymbol{\xi}_k)$, então um pseudoalgoritmo para o método RJMCMC é o seguinte:

Algoritmo RJMCMC geral

Passo 1. Uma transição para o modelo M'_k é proposta com probabilidade $J(k \to k')$.

Passo 2. Amostra-se ν a partir de uma densidade auxiliar $q(\nu|\boldsymbol{\xi}_k, k, k')$.

Passo 3. Define-se $\left(\boldsymbol{\xi}'_k, \nu'\right) := g_{k,k'}(\boldsymbol{\xi}_k, \nu)$, onde $g_{k,k'}(\cdot)$ é uma bijeção entre $(\boldsymbol{\xi}_k, \nu)$ e $\left(\boldsymbol{\xi}'_k, \nu'\right)$.

Passo 4. O novo modelo é aceito com probabilidade

$$\alpha_{k \to k'} = \min\left\{1, \frac{p(y|k', \boldsymbol{\xi}'_k)p(\boldsymbol{\xi}'_k)p(k')J(k' \to k)q(\nu'|\boldsymbol{\xi}_{k'}, k', k)}{p(y|k, \boldsymbol{\xi}_k)p(\boldsymbol{\xi}_k)p(k)J(k \to k')q(\nu|\boldsymbol{\xi}_{k'}, k', k)} \times \left|\frac{\partial g_{k,k'}(\boldsymbol{\xi}_k, \nu)}{\partial(\boldsymbol{\xi}_{k'}, \nu)}\right|\right\}.$$

Repetindo-se as etapas 1–4, gera-se uma amostra $k_l, l = 1, \cdots, L$ para os parâmetros do modelo, do qual obtemos

$$\hat{p}(k|y) = \frac{1}{L}\sum_{l=1}^{L} I(k_l = k),$$

onde $I(\cdot)$ é a função indicadora. Na prática, $J(k \to k')$ é geralmente considerado $N(0, \sigma^2)$, onde σ^2 é um hiperparâmetro de escala.

3.1.1 RJMCMC para modelos GARMA

Em modelos ARMA, o algoritmo acima pode ser aplicado a partir de um conjunto de índices especificando os modelos. Transições são feitas entre índices que representam uma bijeção entre as possíveis ordens (p, q) dos modelos desejados e os números inteiros positivos até um determinado valor máximo, essencialmente como dado nos Passos 1 e 3 do algoritmo geral acima, como proposto em Troughton (1999) e, mais geralmente, em Eğrioğlu and Günay (2010). Ideia semelhante aos modelos ARMA pode ser aplicada para modelos GARMA. Porém, neste contexto tipicamente transições apenas consideram modelos ARMA/GARMA completos, o que é bastante restritivo levando-se em conta a totalidade de submodelos possíveis.

Ao invés de adotar esta perspectiva, aproveitamos a flexibilidade do algoritmo RJMCMC para propor transições de forma mais direta com uma melhor cobertura do espaço de parâmetros. No caso particular de modelos GARMA, específica-se os valores de p e q para o maior modelo GARMA(p, q) a ser explorado pelo algoritmo. Sejam p_m e q_m estes valores máximos, com parâmetros associados $\boldsymbol{\phi}_m := (\phi_1, \cdots, \phi_{p_m})'$ e $\boldsymbol{\theta}_m := (\theta_1, \cdots, \theta_{q_m})'$. Em cada transição, o algoritmo determina se cada um dos parâmetros ϕ_i e θ_j, para $i \in \{1, \cdots, p_m\}$ e $j \in \{1, \cdots, q_m\}$, deve entrar no modelo ou não, com determinada probabilidade dada a priori, dado a atual configuração do sistema. Caso afirmativo, o parâmetro é amostrado normalmente. Caso negativo, o parâmetro é fixado em 0. Esta abordagem além de mais simples, é tipicamente adotada nos diversos softwares e pacotes gerais que implementam o RJMCMC genérico, facilitando inclusive sua utilização.

Uma vantagem adicional de propor transição a transição a inclusão ou não de cada parâmetro é que o algoritmo explora configurações de modelos GARMA que raramente são considerados na prática. Por exemplo, se $p_m = q_m = 3$, o algoritmo pode considerar uma configuração de GARMA$(3, 3)$ com defasagens específicas, no qual apenas ϕ_3 e θ_3 são diferentes de zero. Um modelo desse tipo, raramente é ajustado usando-se a abordagem tradicional de Box e Jenkins. Por outro lado, esta flexibilidade pode tornar o espaço de possíveis modelos amplo demais e dificultar a estabilização do algoritmo, o que na prática se traduz na necessidade de gerar longas cadeias para atingir convergência.

No caso específico de modelos GARMA, o Passo 4 do algoritmo permanece o mesmo, com as probabilidades de transição calculadas a partir das probabilidades de inclusão de cada entrada dos vetores $\boldsymbol{\phi}_m$ e $\boldsymbol{\theta}_m$. O parâmetro α, por sua vez, é sempre amostrado.

Capítulo 4

Conclusões e trabalhos futuros

Neste trabalho, estudamos o problema de seleção da ordem em modelos GARMA para séries temporais de contagem sob uma perspectiva Bayesiana, usando a abordagem Reversible Jump Markov Chain Monte Carlo. Os resultados encontram-se no artigo apresentado no Apêndice A. inicialmente, apresentamos os modelos GARMA para dados de contagem e a abordagem RJMCMC. O desempenho da abordagem proposta é então avaliada em extensas simulações de Monte Carlo, onde são avaliados estimativas pontual, intervalar e porcentagem de identificação do modelo correto e os efeitos da aplicação de burn-in e thinning nas cadeias. Analisamos ainda a sensibilidade da metodologia com relação à escolha dos hiperparâmetros e influência do tamanho das cadeias geradas.

De modo geral, as simulações mostraram que o método é bastante efetivo na identificação do verdadeiro modelo, desde que seja bem calibrado. A aplicação de um burn-in às cadeias geradas mostrou-se muito efetiva para a obtenção de bons resultados, mitigando inclusive a escolha de determinados hiperparâmetros, aos quais a metodologia se mostrou sensível. Por outro lado, muito embora o método gere cadeias bastante correlacionadas, observou-se que a aplicação de thinning não produz resultados significativamente melhores no geral.

Uma parte importante do trabalho, mas que pode passar despercebida, é a implementação computacional do método. Além de os modelos GARMA para dados de contagem serem tradicionalmente computacionalmente sensíveis, a implementação do RJMCMC foi feita utilizando o pacote `Nimble` do R que, embora seja um pacote poderoso, possui uma codificação própria com alguns detalhes que tornam a programação necessária dos modelos GARMA para a utilização com os amostradores RJMCMC altamente não-trivial, o que tomou uma quantidade de tempo grande para sua correta implementação. Em breve o código será disponibilizado livremente no Github.

Os resultados deste estudo indicam que o método Reversible-Jump Markov chain Monte Carlo é uma ferramenta muito útil na seleção de modelos dentro da abordagem Bayesiana aplicada a séries temporais de contagem. O método apresenta-se como promissor em aplicações do mundo real nas mais diversas áreas. Como perspectiva para futuras pesquisas, sugerimos investigar a aplicação do RJMCMC em modelos GARMA que incorporem distribuições contínuas, assim como em modelos INAR destinados a dados de contagem. Além disso, contemplamos a possibilidade de desenvolver um pacote ou script dedicado para compartilhar a implementação computacional desse método, com a finalidade de tornar a abordagem mais acessível tanto para pesquisadores quanto para profissionais interessados nesta área.

Referências Bibliográficas

de Andrade, B.S., Andrade, M.G., Ehlers, R.S., 2015. Bayesian GARMA models for count data. Communications in Statistics: Case Studies, Data Analysis and Applications 1, 192–205.

Bayer, F.M., Bayer, D.M., Pumi, G., 2017. Kumaraswamy autoregressive moving average models for double bounded environmental data. Journal of Hydrology 555, 385–396.

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

Berry, L.R., West, M., 2020. Bayesian forecasting of many count-valued time series. Journal of Business & Economic Statistics 38, 872–887.

Bourguignon, M., Weiß, C.H., 2017. An INAR (1) process for modeling count time series with equidispersion, underdispersion and overdispersion. Test 26, 847–868.

Brännäs, K., Johansson, P., 1994. Time series count data regression. Communications in Statistics-Theory and Methods 23, 2907–2925.

Brumback, B.A., Ryan, L.M., Schwartz, J.D., Neas, L.M., Stark, P.C., Burge, H.A., 2000. Transitional regression models, with application to environmental time series. Journal of the American Statistical Association 95, 16–27.

Campbell, E.P., 2004. Bayesian selection of threshold autoregressive models. Journal of Time Series Analysis 25, 467–482.

Casarin, R., Valle, L.D., Leisen, F., 2010. Bayesian model selection for beta autoregressive processes. Bayesian Analysis 7, 385–410.

Christou, V., Fokianos, K., 2015. On count time series prediction. Journal of Statistical Computation and Simulation 85, 357–373.

Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., Lauritzen, S.L., 1981. Statistical analysis of time series: Some recent developments [with discussion and reply]. Scandinavian Journal of Statistics , 93–115.

Cui, Y., Zheng, Q., 2017. Conditional maximum likelihood estimation for a class of observation-driven time series models for count data. Statistics & Probability Letters 123, 193–201.

Davis, R.A., Dunsmuir, W.T., Wang, Y., 1999. Modeling time series of count data. Statistics textbooks and monographs 158, 63–114.

Davis, R.A., Dunsmuir, W.T., Wang, Y., 2000. On autocorrelation in a Poisson regression model. Biometrika 87, 491–505.

Eğrioğlu, E., Günay, S., 2010. Bayesian model selection in ARFIMA models. Expert Systems with Applications 37, 8359–8364.

Ehlers, R.S., 2019. A Conway-Maxwell-Poisson GARMA Model for Count Data. arXiv:1901.07473.

Fan, Y., Peters, G.W., Sisson, S.A., 2009. Automating and evaluating reversible jump MCMC proposal distributions. Statistics and Computing 19, 409–421.

Franco, G.C., Migon, H.S., Prates, M.O., 2019. Time series of count data. Brazilian Journal of Probability and Statistics 33, 756–781.

Freeland, R.K., McCabe, B.P., 2004. Analysis of low count time series data by Poisson autoregression. Journal of time series analysis 25, 701–722.

Grande, A.F., Pumi, G., Cybis, G.B., 2023. Bayesian Analysis of Beta Autoregressive Moving Average Models. arXiv:2307.07042.

Green, P.J., 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika 82, 711–732.

Homburg, A., Weiß, C.H., Alwan, L.C., Frahm, G., Göb, R., 2021. A performance analysis of prediction intervals for count time series. Journal of Forecasting 40, 603–625.

Karakuş, O., Kuruoğlu, E.E., Altınkaya, M.A., 2016. Bayesian estimation of polynomial moving average models with unknown degree of nonlinearity, in: 2016 24th European Signal Processing Conference (EUSIPCO), IEEE. pp. 1543–1547.

Karakuş, O., Kuruoğlu, E.E., Altinkaya, M.A., 2017. Nonlinear model selection for PARMA processes using RJMCMC, in: 2017 25th European Signal Processing Conference (EUSIPCO), IEEE. pp. 2056–2060.

Khoo, W.C., Ong, S.H., Biswas, A., 2017. Modeling time series of counts with a new class of INAR (1) model. Statistical Papers 58, 393–416.

Liesenfeld, R., Nolte, I., Pohlmeier, W., 2006. Modelling financial transaction price movements: a dynamic integer count data model. Empirical Economics 30, 795–825.

Liu, H., Nason, G., 2023. New methods for network count time series. arXiv preprint arXiv:2312.01944 .

Liu, Z., Zhu, F., 2020. A new extension of thinning-based integer-valued autoregressive models for count data. Entropy 23, 62.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman & Hall / CRC.

Melo, M., Alencar, A., 2020. Conway–Maxwell–Poisson Autoregressive Moving Average Model for Equidispersed, Underdispersed, and Overdispersed Count Data. Journal of Time Series Analysis 41, 830–857. doi:https://doi.org/10.1111/jtsa.12550.

Meyer-Gohde, A., Neuhoff, D., 2015. Generalized exogenous processes in dsge: A bayesian approach. Technical Report. SFB 649 Discussion Paper.

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General) 135, 370–384.

Pavlopoulos, H., Karlis, D., 2008. INAR (1) modeling of overdispersed count series with an environmental application. Environmetrics: The official journal of the International Environmetrics Society 19, 369–393.

Pumi, G., Prass, T.S., Souza, R.R.a., 2021. A dynamic model for double-bounded time series with chaotic-driven conditional averages. Scandinavian Journal of Statistics 48, 68–86.

Pumi, G., Prass, T.S., Taufemback, C.G., 2023. Unit-weibull autoregressive moving average models. To appear.

Pumi, G., Valk, M., Bisognin, C., Bayer, F.M., Prass, T.S., 2019. Beta autoregressive fractionally integrated moving average models. Journal of Statistical Planning and Inference 200, 196 – 212.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. TEST 18, 529–545.

Roy, A., Karmakar, S., 2021. Time-varying auto-regressive models for count time-series. Electronic Journal of Statistics 15, 2905–2938.

Solow, R.M., 1956. A contribution to the theory of economic growth. The quarterly journal of economics 70, 65–94.

Triantafyllopoulos, K., Shakandli, M., Campbell, M., 2019. Count time series prediction using particle filters. Quality and Reliability Engineering International 35, 1445–1459.

Troughton, P.T., 1999. Simulation methods for linear and nonlinear time series models with application to distorted audio signals. Ph.D. thesis. University of Cambridge.

de Valpine, P., Turek, D., Paciorek, C.J., Anderson-Bergman, C., Lang, D.T., Bodik, R., 2017. Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics 26, 403–413.

Weiß, C., Scherer, L., Aleksandrov, B., Feld, M., 2020. Checking model adequacy for count time series by using pearson residuals. Journal of Time Series Econometrics 12, 20180018.

Weiß, C.H., Aleksandrov, B., Faymonville, M., Jentsch, C., 2023. Partial autocorrelation diagnostics for count time series. Entropy 25, 105.

Weiß, C.H., Feld, M.H.J., 2020. On the performance of information criteria for model identification of count time series. Studies in Nonlinear Dynamics & Econometrics 24, 20180012.

Weiß, C.H., Feld, M.H.J., Mamode Khan, N., Sunecher, Y., 2019. Inarma modeling of count time series. Stats 2, 284–320.

Weiß, C.H., Zhu, F., 2024. Conditional-mean multiplicative operator models for count time series. Computational Statistics & Data Analysis 191, 107885.

Yu, M., Wang, D., Yang, K., 2019. A class of observation-driven random coefficient INAR (1) processes based on negative binomial thinning. Journal of the Korean Statistical Society 48, 248–264.

Zeger, S.L., Qaqish, B., 1988. Markov Regression Models for Time Series: A Quasi-Likelihood Approach. Biometrics 44, 1019–1031.

Apêndice A

Artigo Lastra et al. (2024)

Autores: Katerine Zuniga Lastra, Guilherme Pumi e Taiane Schaedler Prass

Título: Order selection in GARMA models for count time series: a Bayesian perspective

Ano: 2024

Order selection in GARMA models for count time series: a Bayesian perspective

Katerine Zuniga Lastra[a], **Guilherme Pumi**[a,*] **and Taiane Schaedler Prass**[a]

Abstract

Estimation in GARMA models has traditionally been carried out under the frequentist approach. To date, Bayesian approaches for such estimation have been relatively limited. In the context of GARMA models for count time series, Bayesian estimation achieves satisfactory results in terms of point estimation. Model selection in this context often relies on the use of information criteria. Despite its prominence in the literature, the use of information criteria for model selection in GARMA models for count time series have been shown to present poor performance in simulations, especially in terms of their ability to correctly identify models, even under large sample sizes. In this study, we study the problem of order selection in GARMA models for count time series, adopting a Bayesian perspective through the application of the Reversible Jump Markov Chain Monte Carlo approach. Monte Carlo simulation studies are conducted to assess the finite sample performance of the developed ideas, including point and interval inference, sensitivity analysis, effects of burn-in and thinning, as well as the choice of related priors and hyperparameters. An application to the time series of automobile production in Brazil is presented, showcasing the method's capabilities and exploring further its flexibility.

Keywords: Count time series; Regression models; Bayesian analysis; Reversible Jump Markov Chain.

MSC: 62M10, 62F12, 62E20, 62G20, 60G15.

1 Introduction

Counting time series typically arise when the interest lies in the count of certain events happening during times intervals. They are ubiquitous to all fields of study and applications abundant. For instance, Zeger and Qaqish (1988) and Davis et al. (2000) analyzed the incidence of certain diseases. In the field of insurance, Freeland and McCabe (2004) presented an application to the monthly count data set of claimants for wage loss benefit, in order to estimate the expected duration of claimants in the system. Liesenfeld et al. (2006) studied fluctuations in the financial market whereas Weiß (2007) considered time series of count in the context of quality management strategies and Brännäs and Johansson (1994) modeled the number of traffic accidents in a given location.

Models for time series of count are mainly modeled under the frameworks of parameter and observation driven models, according to Cox's classification Cox et al. (1981). The former extends generalized linear models by incorporating a latent process into the conditional mean of the counting process, while the latter directly rely on the count observed in each interval to discern the temporal dynamics, specifying a model for the distribution of the count at each moment.

Among the modeling approaches, the class of GARMA (generalized autoregressive moving average) models, introduced under this name by Benjamin et al. (2003), have been extensively

[a]Mathematics and Statistics Institute and Programa de Pós-Graduação em Estatística - Federal University of Rio Grande do Sul.

[*]Corresponding author. This Version: April 12, 2024

E-mails: zunigalastrakaterine@gmail.com (K.Z. Lastra), guilherme.pumi@ufrgs.br (G. Pumi), taiane.prass@ufrgs.br (T.S. Prass)

studied in the recent literature, being considered one of the most promising approaches to non-Gaussian time series modeling. GARMA are observation driven models which merge the classical ARMA modeling approach within the flexibility of the Generalized Linear Model (GLM) framework. Inference in GARMA models are usually conducted under a frequentist framework, based on conditional or partial likelihood. The literature considering Bayesian inference in GARMA models is less abundant. For instance, in the context of continuously distributed GARMA models, Casarin et al. (2010) and Grande et al. (2023) consider Bayesian inference in the context of the βARMA of Rocha and Cribari-Neto (2009). de Andrade et al. (2015) considers Bayesian inference in the context of GARMA models for count time series in the classical framework, namely, when the conditional distribution is a member of the canonical exponential family.

One important matter is model selection for GARMA models. In the frequentist framework this is usually attained by using either a Box and Jenkings-like approach or by using information criteria. information criteria are also widely applied in the context of Bayesian model selection. One important alternative is the so-called Reversible Jump Markov Chain Monte Carlo (RJMCMC) approach, introduced by Green (1995). The RJMCMC is an extension of the Metropolis-Hastings algorithm allowing the generation of samples from a target distribution in spaces of different dimensions. To the best of our knowledge, the only work considering an RJMCMC approach in the context of GARMA models is Casarin et al. (2010)which considered model selection using an RJMCMC approach in the context of a subclass of βAR models.

In this paper, we propose and discuss model selection in GARMA models for count time series using an RJMCMC approach. The commonly applied general purpose RJMCMC can be adapted to be used in the context of GARMA models following the approach employed by Troughton (1999a) in the context of ARMA(p, q) models. The main idea is to enumerate the possible combinations of model orders and use this enumeration as index for model transition. We shall consider a different approach however, in which transitions are determine by inclusion/exclusion of each parameter, given the current state of the chain, according to a prior inclusion probability. The proposed approach allow for more flexibility in model configuration, widening the scope of possible models to be visited by the chain.

The paper is organized as follows. In Section 2, a review of the GARMA model class is conducted, addressing key concepts related to Bayesian inference for these models and the RJMCMC method. In Section 3, we carry out a Monte Carlo simulation to assess the finite sample performance of the proposed Bayesian approach, with emphasis on model selection. In section 4 we present a real data application of the proposed methodology to the time series of automobile production in Brazil. Lastly, we present conclusions and future works.

2 GARMA Models

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a stochastic process of interest and let $\{\boldsymbol{x}_t\}_{t\in\mathbb{Z}}$ be a set of r-dimensional exogenous covariates to be included in the model. Let $\mathscr{F}_t = \sigma\left\{\boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, \cdots, Y_{t-1}, Y_{t-2}, \cdots\right\}$ be the information available to the observer at time t. In Benjamin et al. (2003), GARMA models are considering that the distribution of Y_t given the information observed up to time t belongs to the exponential family in canonical form, that is

$$f\big(y; \omega_t, \varphi|\mathscr{F}_{t-1}\big) = \exp\bigg\{\frac{y\omega_t - b(\omega_t)}{\varphi} + c(y, \varphi)\bigg\}, \tag{1}$$

where, ω_t and φ are the canonical and scale parameters, respectively, with $b(\cdot)$ and $c(\cdot)$ being specific functions that define the particular exponential family. In traditional GARMA models, ω_t is time dependent while φ is not, which is reflected in the notation. The conditional mean and variance of Y_t, given $\mathscr{F}_{t-1}$, are given by $\mu_t = E\big(Y_t|\mathscr{F}_{t-1}\big) = b'\big(\omega_t\big)$ and $\mathrm{Var}\big(Y_t|\mathscr{F}_{t-1}\big) = \varphi b''\big(\omega_t\big) = \varphi V\big(\mu_t\big)$, with $t \in \{1, \cdots, n\}$. In the systematic component of the model, the conditional mean μ_t is related to the linear predictor possibly through a twice differentiable invertible link function g. The most commonly used structure for the systematic component includes covariates and an ARMA structure of the form

$$\eta_t = g(\mu_t) = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \big[g\big(Y_{t-j}\big) - \boldsymbol{x}_{t-j}'\boldsymbol{\beta}\big] + \sum_{j=1}^{q} \theta_j \big[g\big(Y_{t-j}\big) - \eta_{t-j}\big], \tag{2}$$

where η_t is the linear predictor, α is an intercept, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_r)'$ is the parameter vector related to the covariates, $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)'$ are the AR and MA coefficients, respectively. A GARMA(p, q) model is defined by (1) and (2).

The systematic component (1) can be continuous (Rocha and Cribari-Neto, 2009; Bayer et al., 2017; Benaduce and Pumi, 2023, among others), discrete (Benjamin et al., 2003; Melo and Alencar, 2020; Sales et al., 2022, among others), or even of the mixed type (Bayer et al., 2023). In the case of a time series of counts, the most commonly applied GARMA models are reviewed in the next section.

2.1 GARMA models for time series of counts

In this Section, we present a brief description of the three most applied GARMA models for counting data, namely, the Poisson GARMA, binomial GARMA, and negative binomial GARMA models.

2.1.1 Poisson GARMA model

When $Y_t|\mathscr{F}_{t-1}$ follows a Poisson distribution with mean μ_t we have

$$f\big(y; \mu_t|\mathscr{F}_{t-1}\big) = \exp\big\{y \log(\mu_t) - \mu_t - \log(y!)\big\} I(y \in \mathbb{N}). \tag{3}$$

Hence $Y_t|\mathscr{F}_t$ belongs to the canonical exponential family with

$$\varphi = 1, \quad \omega_t = \log(\mu_t), \quad b(\omega_t) = \exp\{\omega_t\}, \quad c(y_t, \varphi) = -\log(y_t!), \quad \mu_t = \exp(\omega_t),$$

and $V(\mu_t) = \mu_t$. The linear predictor is prescribed as

$$\log(\mu_t) = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \big[\log(Y_{t-j}^*) - \boldsymbol{x}_{t-j}'\boldsymbol{\beta}\big] + \sum_{j=1}^{q} \theta_j \big[\log(Y_{t-j}^*) - \log(\mu_{t-j})\big], \tag{4}$$

where $Y_t^* = \max(Y_t, c)$, for $0 < c < 1$, is a user-defined threshold applied to avoid numerical problems with the logarithm when y_t is too close to 0. The Poisson GARMA model is defined by (3) and (4).

2.1.2 Binomial GARMA model

When $Y_t|\mathscr{F}_{t-1} \sim B(m, p_t)$, with $m > 0$ known, and $\mu_t = \mathbb{E}(Y_t|\mathscr{F}_{t-1}) = mp_t$ we have

$$f(y;\mu_t|\mathscr{F}_{t-1}) = \exp\bigg\{y\log\bigg(\frac{p_t}{1-p_t}\bigg) + m\log(1-p_t) + \log\bigg(\frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\bigg)\bigg\}$$
$$= \exp\bigg\{y\log\bigg(\frac{\mu_t}{m-\mu_t}\bigg) + m\log\bigg(\frac{m-\mu_t}{m}\bigg) + \log\bigg(\frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\bigg)\bigg\}, \quad (5)$$

which is a member of the canonical exponential family with $\varphi = 1$,

$$\omega_t = \log\bigg(\frac{\mu_t}{m-\mu_t}\bigg),\ b(\omega_t) = m\log\bigg(\frac{m}{m-\mu_t}\bigg),\ \text{and}\ \ c(y_t,\varphi) = \log\bigg(\frac{\Gamma(m+1)}{\Gamma(y+1)\Gamma(m-y+1)}\bigg).$$

The binomial GARMA model is specified by (4) and (5). In particular,

$$\mu_t = \frac{m\exp(p_t)}{1+\exp(p_t)} \quad \text{and} \quad V(\mu_t) = \frac{\mu_t(m-\mu_t)}{m}.$$

2.1.3 Negative binomial GARMA model

When $Y_t|\mathscr{F}_{t-1} \sim \mathrm{NB}(k, p_t)$, with $k > 0$ known, we have $\mathbb{E}(Y_t|\mathscr{F}_{t-1}) = \frac{k(1-p_t)}{p_t}$, so that $p_t = \frac{k}{\mu_t+k}$ and hence

$$f(y;\mu_t|\mathscr{F}_{t-1}) = \exp\bigg\{k\log\bigg(\frac{k}{\mu_t+k}\bigg) + y\log\bigg(\frac{\mu_t}{\mu_t+k}\bigg) + \log\bigg(\frac{\Gamma(k+y)}{\Gamma(y+1)\Gamma(k)}\bigg)\bigg\}, \quad (6)$$

which belongs to the exponential family with

$$\varphi = 1, \quad \omega_t = \log\bigg(\frac{\mu_t}{\mu_t+k}\bigg), b(\omega_t) = -k\log\bigg(\frac{k}{\mu_t+k}\bigg), \quad c(y_t,\varphi) = \log\bigg(\frac{\Gamma(k+y)}{\Gamma(y+1)\Gamma(k)}\bigg).$$

Equations (4) and (6) specify the negative binomial GARMA(p, q) model with

$$\mu_t = \frac{k\exp(p_t)}{1+\exp(p_t)} \quad \text{and} \quad V(\mu_t) = \frac{\mu_t(k+\mu_t)}{k}.$$

2.1.4 Bayesian approach to GARMA modeling

The partial likelihood function for the model is given by

$$\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\theta},\alpha_0|\mathscr{F}_t) \propto \exp\bigg\{\sum_{t=s+1}^{n} \frac{y_t\theta_t - b(\omega_t)}{\varphi} + c(y_t,\varphi)\bigg\},$$

where ω_t is the canonical parameter of the model and s is the starting point of the likelihood function, most often taken as $s = 0$ as in Benjamin et al. (2003) and Pumi et al. (2019) but sometimes taken as $s = \max\{p, q\}$ as in Rocha and Cribari-Neto (2009) and de Andrade et al. (2015). In this work we shall employ $s = 0$.

For α_0, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, we shall assume normal prior distributions with zero mean and variance σ^2 for each component, that is $\phi_i \sim N(0,\sigma^2)$, $\theta_j \sim N(0,\sigma^2)$ and $\alpha_0 \sim N(0,\sigma^2)$, for $i \in$

$\{1, \cdots, p\}$ and $j = \{1, \cdots, q\}$. Assuming independence between the parameters, the joint prior distribution is

$$\pi_0\big(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0\big) \propto \exp\bigg\{ -\frac{1}{2\sigma^2}\bigg(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2\bigg)\bigg\}. \quad (7)$$

Therefore, the posterior conditional distribution for the model is written as

$$\pi\big(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t\big) \propto \mathcal{L}\big(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t\big)\pi_0\big(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0\big).$$

In what follows, we present explicit formulae for the likelihood functions of each proposed model.

Poisson GARMA Model

The partial likelihood function for the Poisson GARMA model is given by

$$\mathcal{L}\big(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t\big) = \exp\bigg\{ \sum_{t=1}^{n} \Big[y_t \log(\mu_t) - \mu_t - \log(y_t!)\Big]\bigg\},$$

with

$$\mu_t = \exp\bigg\{\alpha_0 + \sum_{i=1}^{p} \phi_i \log(y_{t-i}^*) + \sum_{j=1}^{q} \theta_j \log\bigg(\frac{y_{t-j}^*}{\mu_{t-j}}\bigg)\bigg\}.$$

The posterior density for the Poisson GARMA model is given by

$$\begin{aligned}\pi\left(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t\right) \propto \exp\bigg\{ &\sum_{t=1}^{n} \bigg[y_t\bigg(\alpha_0 + \sum_{i=1}^{p} \phi_i \log(y_{t-i}^*) + \sum_{j=1}^{q} \theta_j \log\bigg(\frac{y_{t-j}^*}{\mu_{t-j}}\bigg)\bigg)\bigg] \\ &- \sum_{t=1}^{n} \exp\bigg\{\alpha_0 + \sum_{i=1}^{p} \phi_i \log(y_{t-i}^*) + \sum_{j=1}^{q} \theta_j \log\bigg(\frac{y_{t-j}^*}{\mu_{t-j}}\bigg)\bigg\} \\ &- \frac{1}{2\sigma^2}\bigg(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2\bigg)\bigg\}.\end{aligned}$$

Binomial GARMA Model

The partial likelihood function for the GARMA binomial model is given by

$$\begin{aligned}\mathcal{L}\big(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, |\mathscr{F}_t\big) = \prod_{t=1}^{n} \exp\bigg\{&y_t \log\bigg(\frac{\mu_t}{m - \mu_t}\bigg) + m \log\bigg(\frac{m - \mu_t}{m}\bigg) \\ &+ \log\bigg(\frac{\Gamma(m+1)}{\Gamma\big(y_t+1\big)\Gamma\big(m - y_t + 1\big)}\bigg)\bigg\} \\ \propto \exp\bigg\{ &\sum_{t=1}^{n} \bigg[y_t \log\bigg(\frac{\mu_t}{m - \mu_t}\bigg) + m \log\bigg(\frac{m - \mu_t}{m}\bigg)\bigg]\bigg\}.\end{aligned}$$

The posterior density for the binomial GARMA model is given by

$$\begin{aligned}\pi\big(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t\big) \propto \exp\bigg\{ &\sum_{t=1}^{n} \bigg[y_t \log\bigg(\frac{\mu_t}{m - \mu_t}\bigg) + m \log\bigg(\frac{m - \mu_t}{m}\bigg)\bigg] + \\ &- \frac{1}{2\sigma^2}\bigg(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2\bigg)\bigg\}.\end{aligned}$$

Binomial Negativa GARMA Model

The approximate likelihood function for the negative binomial GARMA model is given by

$$\mathcal{L}(\alpha_0, \boldsymbol{\phi}, \boldsymbol{\theta}, |\mathscr{F}_t) = \prod_{t=1}^{n} \exp\left\{k \log\left(\frac{k}{\mu_t + k}\right) + y_{\mathrm{t}} \log\left(\frac{\mu_{\mathrm{t}}}{\mu_{\mathrm{t}} + k}\right) + \log\left(\frac{\Gamma(k + y_t)}{\Gamma(y_{\mathrm{t}} + 1)\Gamma(k)}\right)\right\}$$
$$\propto \exp\left\{k \sum_{t=1}^{n} \log\left(\frac{k}{\mu_{\mathrm{t}} + k}\right) + \sum_{t=1}^{n} y_t \log\left(\frac{\mu_{\mathrm{t}}}{\mu_{\mathrm{t}} + k}\right)\right\},$$

From (7), the posterior density for the negative binomial GARMA model is given by

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0|\mathscr{F}_t) \propto \exp\left\{\sum_{t=1}^{n}\left[y_t \log\left(\frac{\mu_t}{m - \mu_t}\right) + m \log\left(\frac{m - \mu_t}{m}\right)\right] + \right.$$
$$\left. - \frac{1}{2\sigma^2}\left(\alpha_0^2 + \sum_{i=1}^{p} \phi_i^2 + \sum_{j=1}^{q} \theta_j^2\right)\right\}.$$

2.1.5 Reversible-jump Markov chain Monte Carlo

As mentioned in the introduction, the method known as Reversible-jump Markov chain Monte Carlo (RJMCMC), introduced by Green (1995), is an extension of the Metropolis-Hastings algorithm allowing the generation of samples of a target distribution in spaces of different dimensions. The dimension of the parameter space is allowed to vary between iterations and is commonly used as a Bayesian method for model selection. According to Green (1995) in a Bayesian modeling context, one has a countable collection of candidate models $\{M_k, k \in K\}$, where the index k serves as an auxiliary indicator variable of the model and K represent the scope of the considered models. The model M_k has a vector of $k+1$ unknown parameters, say $\boldsymbol{\xi}_k \in R^{k+1}$, that can assume different values for different models. There is a natural hierarchical structure expressed by modeling the joint distribution of $(k, \boldsymbol{\xi}_k, y)$ as

$$p(k, \boldsymbol{\xi}_k, y) \propto p(k)p(\xi_k|k)p(y|\boldsymbol{\xi}_k, k).$$

The Bayesian inference about k and $\boldsymbol{\xi}_k$ will be based on the posterior distribution $p(k, \boldsymbol{\xi}_k|y)$, given by

$$p(\boldsymbol{\xi}_k|y, k) \propto p(y|\boldsymbol{\xi}_k, k)p(\boldsymbol{\xi}_k|k)$$

Where $p(y|\boldsymbol{\xi}_k, k)$ and $p(\boldsymbol{\xi}_k|k)$ represent the probability model and the prior distribution of the model parameters M_k, respectively. Thus, the posterior probability is given as,

$$p(k, \boldsymbol{\xi}_k|y) \propto p(k)p(\boldsymbol{\xi}_k|k, y)$$

According to Casarin et al. (2010), the posterior joint distribution is the target distribution of the RJMCMC sampler over the state space $\Theta = \cup_{k\in K}(k, \mathbb{R}^{n_k})$. Within each iteration, the RJMCMC algorithm updates the parameters given the model order and then the model order given the parameters. If the current state of the Markov chain is $(k, \boldsymbol{\xi}_k)$, then a possible version of the RJMCMC algorithm is as follows:

General RJMCMC algorithm

Step 1. Propose a visit to model $M_{k'}$ with probability $J(k \to k')$.

Step 2. ν is sampled from a proposal density $q(\nu|\boldsymbol{\xi}_k, k, k')$.

Step 3. Set $(\boldsymbol{\xi}_{k'}, \nu') = g_{k,k'}(\boldsymbol{\xi}_{k'}, \nu)$, where $g_{k,k'}(\cdot)$ is a bijection between $(\boldsymbol{\xi}_k, \nu)$ e $(\boldsymbol{\xi}_{k'}, \nu')$.

Step 4. The acceptance probability of the new model is

$$\alpha_{k\to k'} = \min\left\{1, \frac{p(y|k', \boldsymbol{\xi}_{k'})p(\boldsymbol{\xi}_{k'})p(k')J(k'\to k)q(\nu'|\boldsymbol{\xi}_{k'}, k', k)}{p(y|k, \boldsymbol{\xi}_k)p(\boldsymbol{\xi}_k)p(k)J(k\to k')q(\nu|\boldsymbol{\xi}_k, k', k)} \times \left|\frac{\partial g_{k,k'}(\boldsymbol{\xi}_k, \nu)}{\partial(\boldsymbol{\xi}_{k'}, \nu)}\right|\right\}.$$

Looping through steps 1–4 generates a sample $\{k_l, l = 1, \cdots, L\}$ for the model indicators and

$$\hat{p}(k|y) = \frac{1}{L}\sum_{l=1}^{L} I(k_l = k)$$

where $I(\cdot)$ is the indicator function. In practice, $J(k \to k')$ is usually taken as $N(0, \sigma^2)$, where σ^2 is a scale hyperparameter.

In the case of ARMA models, the general RJMCMC algorithm can be applied by indexing the model order (p, q) by means of a bijection between the scope of models of interest, say $\{(p, q) \in \mathbb{N}^2 : 1 \le p \le p_m, 1 \le q \le q_m\}$ for p_m and q_m the maximum values of p and q desired, and the positive integers. In this way, the RJMCMC algorithm for ARMA follow essentially steps 1 through 4 above, as presented in Troughton (1999b) and extended to ARFIMA models by Eğrioğlu and Günay (2010). The same approach can, in principle, be used in the context of GARMA models. One criticism to this approach is that transitioning between models via the indexing of (p, q) implies that only "complete" models are considered in each transition. This constraint may be somewhat limiting, especially when exploring the entire scope of possible ARMA submodels. Additionally, the implementation of this approach can be challenging and less generalizable due to the need for careful indexing.

Instead, we propose a more direct and simplified approach to the RJMCMC for GARMA models. This method not only facilitates a more thorough exploration of GARMA submodels but is also easier to implement using widely available general RJMCMC packages and software. We start by determining values p_m and q_m for which the the most complex model of interest is of order (p_m, q_m). Let $\boldsymbol{\phi}_m := (\phi_1, \cdots, \phi_{p_m})'$ and $\boldsymbol{\theta}_m := (\theta_1, \cdots, \theta_{q_m})'$ be the associated AR and MA parameters, respectively. Transitions from one model to another occur by determining whether each parameter ϕ_i, for $i \in \{1, \cdots, p_m\}$, is to be included in the model or not, according to a prior inclusion probability, given the current chain state. If a particular ϕ_i is to be included in the model, then it is sampled normally. Otherwise the parameter is set to 0. The procedure is repeated to cover parameters θ_j, $j \in \{1, \cdots, q_m\}$.

By proposing, transition by transition, which parameters to include in the model (given the current state), the algorithm explores model configurations that are rarely considered in practice. For instance, for $p_m = q_m = 3$, the algorithm might propose a model for which only ϕ_3 and θ_3 are different from 0. The selection of p_m and q_m are important in this context, given the potential for $2^{p_m+q_m}$ submodels that can be proposed using this approach. While larger values of p_m and q_m may increase the algorithm's flexibility, they also present a challenge as the resulting scope of possible models may be too extensive, requiring very large chains for the MCMC sampler to converge.

3 Simulations

In this section, we present a Monte Carlo simulation study aimed at evaluating the finite sample performance of the proposed model selection in the context of GARMA(p, q) count models. In the simulation, we consider the point and interval estimation of the parameters of interest and also the percentage of models correctly selected by the proposed approach.

As expected, given the characteristics of the RJMCMC, and widely reported in the literature, samples from the posterior distribution obtained via RJMCMC are typically sensible to the initial values and to the scale hyperparameter σ^2, associated with the transition probability and are highly correlated as well. Mitigating the influence of initial values in the posterior sample is usually attained through a burn-in, whereas, autocorrelation in the sample can be mitigated by using a thinning approach. With that in view, we also provide a sensitivity analysis with respect to the burn-in, thinning and the scale hyperparameter used.

The simulation was carried out using the software `R` (R Core Team, 2020), version 4.0.3. To perform the RJMCMC, we use `R` package `Nimble` (de Valpine et al., 2023).

3.1 Effects of Burn-in

In this section, we examine the finite sample performance of point and interval estimation of the proposed Bayesian approach for the GARMA Binomial model with different values of p and q, different values of the hyperparameter $\sigma \in \{0.5, 5, 10, 15\}$ and burn-in values $\{0, 1000, 3000, 5000\}$. Observe that the proposed RJMCMC approach perform model estimation and point estimation at the same time, hence, being different than the Bayesian approach presented in de Andrade et al. (2015).

3.1.1 GAR(p) models

The first set of experiments considers GAR(1) models with $(\alpha, \phi) \in \{(1, 0.5), (-0.5, -0.4)\}$ and GAR(2) models with $(\alpha, \phi_1, \phi_2) \in \{(-1, 0, -0.4), (1, 0, 0.5)\}$ and $m = 15$. To generate the time series, a burn-in of 100 points was considered and a constant of $c = 0.3$ for the binomial GARMA models was used, independently of the model considered. We generated time series of size $n = 1{,}000$ and a total of 1,000 replications of each scenario were performed.

In all scenarios, RJMCMC was performed considering maximum orders $p_m = 3$ and $q_m = 3$ with a non-informative prior probability of 0.5 for the inclusion of each parameter. We consider a $N(0, 0.3^2)$ prior for α and a $N(0, 0.2^2)$ for the AR parameters. These can be considered somewhat informative, but larger values of the hyperparameters were found to cause numerical instability when compiling the `Nimble` code, often making compilation impossible. We consider zero-mean normally distributed reversible jump proposals with standard deviation (scale) taking values on $\{0.5, 5, 8\}$. For each scenario, a single chain containing 30,000 iterations was generated.

Credible intervals (CIs) were obtained using two methods: the highest posterior density interval (HPD) and the empirical (emp) one, based on the sample from the posterior distribution obtained. To obtain HPD intervals, we use function `emp.hpd` from the `R` package `TeachingDemos` (Snow, 2020), while for empirical credible intervals, we apply the `R` function `quantile`. All credible intervals are presented considering a confidence level of 0.05. To calculate the effective sample size (ESS) for each parameter, we use function `effectiveSize` from the `R` package `coda` (Plummer et al., 2006).

The simulation results are shown in Table 1 and 2. For each scenario, we present point estimates obtained as the average (Mean), median (Median), and standard deviation (sd) of the posterior distribution, along with the average HPD intervals, denoted by HPD, obtained by averaging the limits of the credible intervals. For each parameter, we included the frequency for which the CIs correctly identify the parameter as significant or not, according to the data generating process, denoted by %cob. In this sense, a value of 99% for a parameter that is non-zero means that in 990 of the 1,000 replications, the parameter was correctly identified as significant based on the CI; on the other hand, the same percentage for a parameter that is 0, means that in 990 out of the 1,000 replications correctly identified the parameter as non-significant based on the CI.

When no burn-in is applied, we observe that as the value of σ increases, so does the bias in the estimates for the GAR(1) model. For the GAR(2), a pattern is not so easily identifiable and the effect of σ in the estimation is less noticeable. In this case, the smallest bias for the non-zero parameters was obtained for $\sigma = 10$. Very little difference is observed when we apply the mean or median to obtain point estimation, with a slight advantage for the median in both cases. Effective sample size is low for most parameters, especially for the non-zero ones, due to high correlation in the sample. For the GAR(1), effective sample size seems to increase as σ decreases. The percentage of correctly identified models is lower for the GAR(1) model than for the GAR(2) and this percentage seems to decrease as σ increases in both cases. As expected, when the HPD credible intervals are considered for model identification, a higher percentage of correctly identified models is obtained when compared to the quantile based CIs. The best scenario in this metric was when $\sigma = 0.5$ with an advantage of almost 30% to the worst case for the GAR(1) case for both CIs. For the GAR(2) model these numbers are about 10% considering HPD and about 20% for the quantile based CIs. For the GAR(2) models, the percentage of correctly identified models is fairly high, above 97% for both CIs, but for the GAR(1) it can be considered on the low end. As for standard deviations, these do not seem to be impacted by σ.

Applying a burn-in improves the results in all cases and in all metrics. The most interesting feature, however, is that the effect of σ is highly mitigated upon applying a burn-in, yielding more dependable results overall. This is especially observed in the percentage of correctly identified models, which in the case of GAR(1) increases from fairly low values to values around 90% in all cases. For the GAR(2) these values are around 99% in all cases. Effective sample size also generally increases upon applying a burn-in, but in most cases the improvement is marginal.

Regarding the size of the burn-in, for the GAR(1) the improvements obtained from applying a size 3,000 burn-in compared to 1,000 are very noticeable, while for the GAR(2), the effect is not as noticeable. In both cases, the improvement obtained by using a burn-in of size 5,000 compared to 3,000 is small under all metrics.

Table 1: RJMCMC Simulation Results for GAR(1) Models with burn-in $\{0, 1000, 3000, 5000\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Burn-in	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -0.5$			-0.509	-0.503	0.149	70			-0.524	-0.508	0.160	54
	$\phi_1 = -0.4$			-0.369	-0.385	0.093	81			-0.358	-0.382	0.104	55
	$\phi_2 = 0.0$			-0.012	-0.001	0.054	114			-0.010	-0.001	0.049	88
0	$\phi_3 = 0.0$	74.0%	66.4%	-0.011	-0.003	0.043	185	55.8%	48.2%	-0.012	-0.003	0.042	133
	$\theta_1 = 0.0$			-0.027	-0.002	0.090	71			-0.037	-0.003	0.100	45
	$\theta_2 = 0.0$			0.018	0.002	0.061	147			0.019	0.003	0.058	121
	$\theta_3 = 0.0$			0.006	0.001	0.044	274			0.007	0.001	0.043	227
	$\alpha = -0.5$			-0.508	-0.503	0.132	84			-0.517	-0.506	0.142	65
	$\phi_1 = -0.4$			-0.373	-0.386	0.081	102			-0.366	-0.384	0.091	72
	$\phi_2 = 0.0$			-0.010	-0.001	0.044	144			-0.010	-0.001	0.044	95
1000	$\phi_3 = 0.0$	86.2%	80.6%	-0.010	-0.003	0.037	230	71.8%	64.9%	-0.011	-0.003	0.037	148
	$\theta_1 = 0.0$			-0.023	-0.002	0.079	91			-0.030	-0.003	0.088	59
	$\theta_2 = 0.0$			0.015	0.002	0.051	204			0.017	0.002	0.052	144
	$\theta_3 = 0.0$			0.006	0.001	0.039	369			0.006	0.001	0.039	265
	$\alpha = -0.5$			-0.504	-0.502	0.125	86			-0.506	-0.502	0.129	72
	$\phi_1 = -0.4$			-0.377	-0.387	0.075	110			-0.374	-0.386	0.078	87
	$\phi_2 = 0.0$			-0.009	-0.001	0.043	147			-0.010	-0.001	0.043	96
3000	$\phi_3 = 0.0$	92.5%	89.2%	-0.010	-0.003	0.037	223	89.4%	84.2%	-0.010	-0.003	0.037	147
	$\theta_1 = 0.0$			-0.019	-0.002	0.072	99			-0.022	-0.002	0.076	71
	$\theta_2 = 0.0$			0.014	0.002	0.049	226			0.015	0.002	0.050	162
	$\theta_3 = 0.0$			0.006	0.001	0.039	365			0.006	0.001	0.039	262
	$\alpha = -0.5$			-0.503	-0.502	0.124	81			-0.503	-0.502	0.125	70
	$\phi_1 = -0.4$			-0.377	-0.387	0.074	105			-0.376	-0.386	0.075	88
	$\phi_2 = 0.0$			-0.009	-0.001	0.043	141			-0.010	-0.001	0.043	93
5000	$\phi_3 = 0.0$	92.8%	89.8%	-0.010	-0.003	0.036	214	93.1%	90.5%	-0.010	-0.003	0.036	143
	$\theta_1 = 0.0$			-0.019	-0.002	0.071	94			-0.020	-0.002	0.072	71
	$\theta_2 = 0.0$			0.014	0.002	0.048	225			0.014	0.002	0.049	164
	$\theta_3 = 0.0$			0.006	0.001	0.039	360			0.006	0.001	0.039	255
Burn-in	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -0.5$			-0.532	-0.510	0.167	47			-0.538	-0.513	0.167	46
	$\phi_1 = -0.4$			-0.353	-0.381	0.110	47			-0.350	-0.380	0.112	45
	$\phi_2 = 0.0$			-0.010	-0.001	0.047	84			-0.009	-0.001	0.045	81
0	$\phi_3 = 0.0$	48.0%	40.8%	-0.012	-0.003	0.040	127	44.3%	37.6%	-0.011	-0.002	0.039	119
	$\theta_1 = 0.0$			-0.042	-0.004	0.106	37			-0.045	-0.005	0.108	35
	$\theta_2 = 0.0$			0.020	0.004	0.057	102			0.020	0.003	0.056	93
	$\theta_3 = 0.0$			0.006	0.001	0.041	215			0.006	0.001	0.040	187
	$\alpha = -0.5$			-0.522	-0.507	0.149	59			-0.527	-0.510	0.150	57
	$\phi_1 = -0.4$			-0.361	-0.383	0.097	63			-0.359	-0.382	0.099	59
	$\phi_2 = 0.0$			-0.010	-0.002	0.044	87			-0.009	-0.001	0.044	82
1000	$\phi_3 = 0.0$	64.5%	58.8%	-0.011	-0.003	0.038	133	60.9%	55.6%	-0.011	-0.003	0.037	122
	$\theta_1 = 0.0$			-0.035	-0.004	0.094	49			-0.037	-0.005	0.096	44
	$\theta_2 = 0.0$			0.018	0.004	0.053	117			0.018	0.003	0.053	104
	$\theta_3 = 0.0$			0.006	0.001	0.039	230			0.006	0.001	0.039	193
	$\alpha = -0.5$			-0.509	-0.503	0.133	68			-0.512	-0.506	0.132	67
	$\phi_1 = -0.4$			-0.371	-0.385	0.082	81			-0.370	-0.384	0.083	77
	$\phi_2 = 0.0$			-0.010	-0.002	0.044	87			-0.010	-0.001	0.043	82
3000	$\phi_3 = 0.0$	83.8%	79.7%	-0.011	-0.003	0.037	133	80.9%	76.6%	-0.010	-0.003	0.037	119
	$\theta_1 = 0.0$			-0.025	-0.003	0.079	61			-0.026	-0.004	0.081	56
	$\theta_2 = 0.0$			0.016	0.003	0.050	138			0.016	0.002	0.050	119
	$\theta_3 = 0.0$			0.006	0.001	0.039	224			0.006	0.001	0.039	185
	$\alpha = -0.5$			-0.504	-0.501	0.127	69			-0.507	-0.504	0.126	69
	$\phi_1 = -0.4$			-0.375	-0.386	0.076	84			-0.374	-0.385	0.076	82
	$\phi_2 = 0.0$			-0.010	-0.002	0.043	85			-0.009	-0.001	0.042	82
5000	$\phi_3 = 0.0$	92.1%	89.9%	-0.011	-0.003	0.037	129	89.5%	87.0%	-0.010	-0.003	0.037	116
	$\theta_1 = 0.0$			-0.021	-0.003	0.073	63			-0.022	-0.003	0.074	59
	$\theta_2 = 0.0$			0.015	0.003	0.049	141			0.014	0.002	0.049	125
	$\theta_3 = 0.0$			0.006	0.001	0.039	215			0.006	0.001	0.039	177

Table 2: RJMCMC Simulation Results for GAR(2) Models considering burn-in $\{0, 1000, 3000, 5000\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Burn-in	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1.0$			-0.942	-0.966	0.135	85			-0.949	-0.969	0.127	76
	$\phi_1 = 0.0$			-0.036	-0.006	0.082	77			-0.033	-0.006	0.074	70
	$\phi_2 = -0.4$			-0.394	-0.396	0.063	234			-0.391	-0.395	0.068	164
0	$\phi_3 = 0.0$	98.7%	97.3%	-0.021	-0.006	0.049	199	95.9%	89.0%	-0.019	-0.006	0.045	153
	$\theta_1 = 0.0$			0.034	0.005	0.083	83			0.031	0.004	0.076	83
	$\theta_2 = 0.0$			-0.002	-0.000	0.062	210			-0.004	-0.000	0.067	135
	$\theta_3 = 0.0$			0.011	0.004	0.048	296			0.010	0.004	0.045	222
	$\alpha = -1.0$			-0.946	-0.967	0.116	109			-0.947	-0.968	0.116	88
	$\phi_1 = 0.0$			-0.033	-0.006	0.071	98			-0.032	-0.006	0.070	72
	$\phi_2 = -0.4$			-0.395	-0.396	0.056	290			-0.395	-0.396	0.058	225
1000	$\phi_3 = 0.0$	99.2%	99.1%	-0.019	-0.006	0.042	268	98.5%	97.5%	-0.019	-0.006	0.041	166
	$\theta_1 = 0.0$			0.031	0.005	0.072	106			0.030	0.005	0.071	85
	$\theta_2 = 0.0$			-0.001	-0.000	0.056	256			-0.001	-0.000	0.057	177
	$\theta_3 = 0.0$			0.010	0.004	0.043	399			0.009	0.004	0.042	245
	$\alpha = -1.0$			-0.946	-0.968	0.115	105			-0.946	-0.967	0.115	85
	$\phi_1 = 0.0$			-0.032	-0.006	0.070	98			-0.033	-0.006	0.070	69
	$\phi_2 = -0.4$			-0.395	-0.396	0.056	275			-0.395	-0.396	0.056	220
3000	$\phi_3 = 0.0$	99.3%	99.2%	-0.019	-0.006	0.042	256	99.3%	98.8%	-0.019	-0.006	0.041	159
	$\theta_1 = 0.0$			0.030	0.005	0.071	107			0.031	0.005	0.071	82
	$\theta_2 = 0.0$			-0.001	0.000	0.056	241			-0.000	0.000	0.056	172
	$\theta_3 = 0.0$			0.010	0.004	0.042	382			0.010	0.004	0.042	233
	$\alpha = -1.0$			-0.946	-0.968	0.115	98			-0.946	-0.967	0.115	80
	$\phi_1 = 0.0$			-0.032	-0.006	0.070	94			-0.033	-0.006	0.070	67
	$\phi_2 = -0.4$			-0.395	-0.396	0.056	256			-0.395	-0.396	0.056	207
5000	$\phi_3 = 0.0$	99.3%	99.3%	-0.019	-0.006	0.042	242	99.1%	98.8%	-0.019	-0.006	0.041	150
	$\theta_1 = 0.0$			0.030	0.005	0.071	104			0.031	0.005	0.071	81
	$\theta_2 = 0.0$			-0.001	-0.000	0.056	226			-0.000	0.000	0.056	160
	$\theta_3 = 0.0$			0.010	0.004	0.042	361			0.010	0.004	0.042	220
Burn-in	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1.0$			-0.956	-0.971	0.125	73			-0.957	-0.971	0.128	71
	$\phi_1 = 0.0$			-0.031	-0.005	0.069	72			-0.031	-0.006	0.069	68
	$\phi_2 = -0.4$			-0.389	-0.395	0.072	145			-0.387	-0.394	0.075	133
0	$\phi_3 = 0.0$	92.1%	81.5%	-0.018	-0.006	0.043	143	87.2%	76.8%	-0.018	-0.006	0.042	134
	$\theta_1 = 0.0$			0.029	0.004	0.070	87			0.029	0.005	0.070	80
	$\theta_2 = 0.0$			-0.007	-0.001	0.069	115			-0.008	-0.001	0.072	105
	$\theta_3 = 0.0$			0.009	0.004	0.043	194			0.009	0.004	0.043	180
	$\alpha = -1.0$			-0.951	-0.969	0.115	85			-0.950	-0.969	0.117	81
	$\phi_1 = 0.0$			-0.030	-0.005	0.067	72			-0.032	-0.006	0.068	67
	$\phi_2 = -0.4$			-0.393	-0.396	0.060	197			-0.393	-0.395	0.062	188
1000	$\phi_3 = 0.0$	97.8%	95.7%	-0.018	-0.006	0.041	146	95.3%	90.9%	-0.018	-0.006	0.041	133
	$\theta_1 = 0.0$			0.029	0.005	0.069	88			0.030	0.005	0.069	79
	$\theta_2 = 0.0$			-0.002	-0.000	0.059	150			-0.003	-0.000	0.061	140
	$\theta_3 = 0.0$			0.010	0.004	0.042	203			0.009	0.004	0.042	181
	$\alpha = -1.0$			-0.948	-0.968	0.113	84			-0.946	-0.967	0.113	81
	$\phi_1 = 0.0$			-0.031	-0.006	0.068	70			-0.032	-0.007	0.069	64
	$\phi_2 = -0.4$			-0.395	-0.396	0.056	203			-0.395	-0.396	0.056	200
3000	$\phi_3 = 0.0$	99.2%	98.7%	-0.018	-0.006	0.041	139	99.1%	98.5%	-0.019	-0.006	0.041	127
	$\theta_1 = 0.0$			0.029	0.005	0.069	88			0.031	0.006	0.070	75
	$\theta_2 = 0.0$			-0.001	-0.000	0.056	152			-0.001	0.000	0.056	147
	$\theta_3 = 0.0$			0.010	0.004	0.042	192			0.010	0.004	0.042	174
	$\alpha = -1.0$			-0.947	-0.967	0.113	79			-0.946	-0.966	0.113	77
	$\phi_1 = 0.0$			-0.032	-0.006	0.068	68			-0.033	-0.007	0.069	63
	$\phi_2 = -0.4$			-0.395	-0.396	0.056	191			-0.395	-0.396	0.056	190
5000	$\phi_3 = 0.0$	99.4%	99.2%	-0.019	-0.007	0.041	130	99.2%	99.0%	-0.019	-0.007	0.041	123
	$\theta_1 = 0.0$			0.030	0.005	0.069	86			0.031	0.006	0.070	74
	$\theta_2 = 0.0$			-0.001	0.000	0.056	142			-0.001	-0.000	0.055	138
	$\theta_3 = 0.0$			0.010	0.004	0.042	179			0.010	0.004	0.042	167

3.1.2 GMA(q) models

In this section we consider GMA(1) models with parameters $\alpha = -0.5$ and $\theta_1 = -0.5$ and GMA(2) with $(\alpha, \theta_1, \theta_2) \in \{(0.5, 0, 0.5), (-0.5, 0, -0.5), (-1, 0, 0.6)\}$. Hyperparameter m was set to 40 and we consider $c = 0.3$ for the binomial. To generate the required time series, a burn-in of 100 points was applied yielding a final sample size of $n = 1{,}000$. We generate 1,000 replicas of each proposed scenario.

Regarding the RJMCMC procedures, they are the same as in the previous analysis, namely, maximum orders were taken as $p_m = 3$ and $q_m = 3$, accompanied by a non-informative prior probability of 0.5 for the inclusion of each parameter. Priors for α and the MA parameters were $N(0, 0.3^2)$ and $N(0, 0.2^2)$ respectively, whereas $\sigma \in \{0.5, 5, 10, 15\}$.A single chain of 30,000 iterations was sampled for each scenario.

The results are presented in Table 3 and 4. Regarding the hyperparameter σ, in both scenarios $\sigma = 0.5$ yielded the worst results, whereas little difference in point estimation is observed for $\sigma \in \{5, 10, 15\}$. Overall, the effects of the burn-in in point estimation are considerably less noticeable than in the GAR case. Considering model identification, in most cases applying a burn-in is even slightly detrimental, especially in the GMA(1) case. The percentage of correctly identified models is lower for the GMA(1) model compared to the GMA(2) model and this percentage appears to increase as σ increases for GMA(1), although this pattern is not observed for the GMA(2). In the GMA situation, applying a burn-in does not seem to significantly improve the results.

Table 3: RJMCMC Simulation Results for GMA(1) Models considering burn-in $\{0, 1000, 3000, 5000\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Burn-in	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
0	$\alpha = -0.5$	89.2%	88.4%	-0.277	-0.301	0.181	11	94.0%	93.7%	-0.356	-0.389	0.153	11
	$\phi_1 = 0.0$			-0.075	-0.061	0.066	14			-0.044	-0.028	0.050	31
	$\phi_2 = 0.0$			-0.006	-0.002	0.033	28			-0.005	-0.002	0.025	36
	$\phi_3 = 0.0$			-0.001	0.001	0.029	35			-0.003	-0.002	0.022	43
	$\theta_1 = -0.5$			-0.402	-0.408	0.080	451			-0.430	-0.437	0.071	1037
	$\theta_2 = 0.0$			-0.011	-0.005	0.049	1288			-0.005	-0.002	0.041	1305
	$\theta_3 = 0.0$			-0.002	-0.001	0.040	3144			0.000	-0.000	0.036	1964
1000	$\alpha = -0.5$	89.1%	88.1%	-0.279	-0.302	0.174	11	93.5%	93.3%	-0.354	-0.386	0.150	11
	$\phi_1 = 0.0$			-0.076	-0.062	0.065	14			-0.045	-0.029	0.050	31
	$\phi_2 = 0.0$			-0.005	-0.002	0.031	29			-0.005	-0.002	0.024	36
	$\phi_3 = 0.0$			-0.000	0.001	0.027	37			-0.003	-0.002	0.021	43
	$\theta_1 = -0.5$			-0.401	-0.407	0.079	470			-0.430	-0.436	0.070	1229
	$\theta_2 = 0.0$			-0.012	-0.005	0.047	1391			-0.006	-0.002	0.040	1310
	$\theta_3 = 0.0$			-0.002	-0.001	0.038	3479			-0.000	-0.000	0.035	2026
3000	$\alpha = -0.5$	87.6%	86.9%	-0.280	-0.302	0.170	11	92.8%	92.6%	-0.348	-0.378	0.147	11
	$\phi_1 = 0.0$			-0.077	-0.064	0.064	14			-0.048	-0.033	0.050	31
	$\phi_2 = 0.0$			-0.004	-0.002	0.029	30			-0.005	-0.002	0.024	34
	$\phi_3 = 0.0$			0.000	0.001	0.025	37			-0.003	-0.002	0.021	41
	$\theta_1 = -0.5$			-0.400	-0.406	0.078	467			-0.427	-0.433	0.070	1143
	$\theta_2 = 0.0$			-0.012	-0.005	0.045	1574			-0.006	-0.003	0.040	1262
	$\theta_3 = 0.0$			-0.002	-0.001	0.036	3904			-0.000	-0.000	0.034	1963
5000	$\alpha = -0.5$	85.6%	85.5%	-0.277	-0.298	0.168	11	92.0%	91.8%	-0.342	-0.368	0.146	10
	$\phi_1 = 0.0$			-0.078	-0.066	0.063	13			-0.050	-0.036	0.049	29
	$\phi_2 = 0.0$			-0.004	-0.002	0.028	29			-0.005	-0.002	0.024	33
	$\phi_3 = 0.0$			0.001	0.001	0.025	36			-0.003	-0.002	0.021	39
	$\theta_1 = -0.5$			-0.399	-0.404	0.077	460			-0.425	-0.430	0.069	1055
	$\theta_2 = 0.0$			-0.013	-0.005	0.044	1609			-0.007	-0.003	0.040	1191
	$\theta_3 = 0.0$			-0.003	-0.001	0.036	3735			-0.000	-0.000	0.034	1839

Burn-in	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
0	$\alpha = -0.5$	94.9%	95.0%	-0.352	-0.384	0.159	15	95.7%	95.3%	-0.358	-0.389	0.154	13
	$\phi_1 = 0.0$			-0.049	-0.034	0.054	31			-0.046	-0.031	0.052	40
	$\phi_2 = 0.0$			-0.004	-0.001	0.024	47			-0.005	-0.001	0.023	49
	$\phi_3 = 0.0$			-0.001	-0.001	0.021	52			-0.001	-0.000	0.019	55
	$\theta_1 = -0.5$			-0.426	-0.432	0.075	899			-0.428	-0.435	0.074	851
	$\theta_2 = 0.0$			-0.008	-0.003	0.041	1069			-0.006	-0.002	0.041	863
	$\theta_3 = 0.0$			-0.001	-0.000	0.035	1756			-0.001	-0.001	0.034	1492
1000	$\alpha = -0.5$	94.6%	94.5%	-0.349	-0.380	0.157	14	95.0%	94.8%	-0.355	-0.385	0.153	13
	$\phi_1 = 0.0$			-0.050	-0.035	0.054	31			-0.047	-0.032	0.052	39
	$\phi_2 = 0.0$			-0.004	-0.001	0.023	46			-0.005	-0.002	0.023	48
	$\phi_3 = 0.0$			-0.001	-0.001	0.020	51			-0.001	-0.000	0.019	54
	$\theta_1 = -0.5$			-0.425	-0.431	0.073	1190			-0.428	-0.434	0.072	1371
	$\theta_2 = 0.0$			-0.008	-0.003	0.041	1058			-0.006	-0.002	0.041	854
	$\theta_3 = 0.0$			-0.001	-0.001	0.034	1758			-0.001	-0.001	0.034	1503
3000	$\alpha = -0.5$	92.9%	92.8%	-0.343	-0.372	0.155	14	94.3%	93.6%	-0.349	-0.377	0.152	13
	$\phi_1 = 0.0$			-0.052	-0.039	0.054	31			-0.049	-0.035	0.052	37
	$\phi_2 = 0.0$			-0.004	-0.001	0.023	45			-0.005	-0.002	0.023	47
	$\phi_3 = 0.0$			-0.001	-0.001	0.020	50			-0.001	-0.001	0.020	52
	$\theta_1 = -0.5$			-0.423	-0.429	0.073	1104			-0.426	-0.432	0.072	1274
	$\theta_2 = 0.0$			-0.008	-0.004	0.041	1014			-0.007	-0.002	0.041	819
	$\theta_3 = 0.0$			-0.001	-0.001	0.034	1681			-0.001	-0.001	0.034	1437
5000	$\alpha = -0.5$	91.4%	91.5%	-0.336	-0.363	0.154	13	93.4%	92.8%	-0.343	-0.369	0.151	12
	$\phi_1 = 0.0$			-0.055	-0.042	0.054	30			-0.051	-0.038	0.051	37
	$\phi_2 = 0.0$			-0.004	-0.001	0.024	43			-0.005	-0.002	0.023	45
	$\phi_3 = 0.0$			-0.001	-0.001	0.021	47			-0.001	-0.001	0.020	51
	$\theta_1 = -0.5$			-0.421	-0.426	0.072	1029			-0.424	-0.429	0.071	1191
	$\theta_2 = 0.0$			-0.009	-0.004	0.041	958			-0.007	-0.002	0.041	773
	$\theta_3 = 0.0$			-0.001	-0.001	0.034	1572			-0.001	-0.001	0.034	1350

Table 4: RJMCMC Simulation Results for GMA(2) Models considering burn-in $\{0, 1000, 3000, 5000\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Burn-in	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1$			-0.813	-0.843	0.191	32			-0.929	-0.951	0.121	125
	$\phi_1 = 0.0$			-0.050	-0.027	0.072	33			-0.029	-0.007	0.067	102
	$\phi_2 = 0.0$			-0.009	-0.003	0.047	48			-0.005	-0.001	0.048	212
0	$\phi_3 = 0.0$	98.9%	99.1%	-0.020	-0.010	0.044	61	99.4%	99.4%	-0.014	-0.006	0.038	281
	$\theta_1 = 0.0$			0.040	0.017	0.071	127			0.025	0.003	0.065	116
	$\theta_2 = 0.6$			0.586	0.586	0.053	290			0.587	0.588	0.052	518
	$\theta_3 = 0.0$			0.026	0.012	0.049	223			0.019	0.006	0.044	214
	$\alpha = -1$			-0.822	-0.846	0.167	37			-0.929	-0.950	0.116	132
	$\phi_1 = 0.0$			-0.048	-0.027	0.064	38			-0.029	-0.007	0.065	104
	$\phi_2 = 0.0$			-0.008	-0.003	0.041	54			-0.005	-0.001	0.045	221
1000	$\phi_3 = 0.0$	98.4%	98.6%	-0.019	-0.011	0.039	69	99.1%	99.3%	-0.013	-0.006	0.036	294
	$\theta_1 = 0.0$			0.038	0.018	0.064	157			0.025	0.003	0.063	119
	$\theta_2 = 0.6$			0.586	0.586	0.050	350			0.588	0.588	0.048	527
	$\theta_3 = 0.0$			0.025	0.012	0.044	343			0.019	0.006	0.042	231
	$\alpha = -1$			-0.822	-0.845	0.164	36			-0.929	-0.950	0.116	124
	$\phi_1 = 0.0$			-0.047	-0.027	0.062	39			-0.030	-0.007	0.065	100
	$\phi_2 = 0.0$			-0.008	-0.004	0.041	52			-0.005	-0.001	0.045	209
3000	$\phi_3 = 0.0$	97.9%	98.0%	-0.020	-0.011	0.039	66	99.1%	99.1%	-0.013	-0.006	0.036	278
	$\theta_1 = 0.0$			0.037	0.018	0.062	189			0.025	0.004	0.063	116
	$\theta_2 = 0.6$			0.586	0.586	0.050	332			0.588	0.588	0.049	497
	$\theta_3 = 0.0$			0.025	0.012	0.044	361			0.019	0.007	0.042	220
	$\alpha = -1$			-0.820	-0.843	0.164	34			-0.929	-0.950	0.116	117
	$\phi_1 = 0.0$			-0.048	-0.028	0.062	38			-0.030	-0.007	0.065	98
	$\phi_2 = 0.0$			-0.008	-0.004	0.041	49			-0.005	-0.001	0.045	194
5000	$\phi_3 = 0.0$	97.5%	97.8%	-0.020	-0.011	0.039	62	99.1%	99.3%	-0.014	-0.006	0.036	259
	$\theta_1 = 0.0$			0.038	0.019	0.062	213			0.025	0.003	0.063	118
	$\theta_2 = 0.6$			0.586	0.586	0.050	316			0.588	0.588	0.048	469
	$\theta_3 = 0.0$			0.025	0.013	0.044	362			0.019	0.007	0.042	209

Burn-in	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1$			-0.836	-0.863	0.169	33			-0.841	-0.868	0.165	34
	$\phi_1 = 0.0$			-0.042	-0.022	0.061	35			-0.040	-0.021	0.057	37
	$\phi_2 = 0.0$			-0.008	-0.003	0.041	45			-0.008	-0.004	0.040	46
0	$\phi_3 = 0.0$	99.2%	99.2%	-0.019	-0.010	0.039	55	99.0%	98.8%	-0.019	-0.009	0.038	54
	$\theta_1 = 0.0$			0.033	0.014	0.060	168			0.030	0.013	0.057	178
	$\theta_2 = 0.6$			0.585	0.586	0.053	438			0.586	0.586	0.053	454
	$\theta_3 = 0.0$			0.023	0.010	0.044	234			0.022	0.009	0.042	214
	$\alpha = -1$			-0.834	-0.860	0.162	34			-0.838	-0.864	0.159	34
	$\phi_1 = 0.0$			-0.043	-0.024	0.059	35			-0.041	-0.022	0.057	37
	$\phi_2 = 0.0$			-0.008	-0.004	0.040	45			-0.008	-0.005	0.039	46
1000	$\phi_3 = 0.0$	98.5%	98.7%	-0.019	-0.010	0.038	56	98.2%	98.2%	-0.019	-0.010	0.037	54
	$\theta_1 = 0.0$			0.034	0.015	0.059	166			0.031	0.014	0.057	173
	$\theta_2 = 0.6$			0.586	0.586	0.050	397			0.587	0.587	0.049	423
	$\theta_3 = 0.0$			0.023	0.011	0.043	243			0.022	0.010	0.042	215
	$\alpha = -1$			-0.828	-0.852	0.161	33			-0.832	-0.856	0.158	34
	$\phi_1 = 0.0$			-0.045	-0.026	0.060	34			-0.042	-0.024	0.057	36
	$\phi_2 = 0.0$			-0.008	-0.004	0.040	43			-0.009	-0.005	0.039	44
3000	$\phi_3 = 0.0$	97.2%	97.2%	-0.019	-0.011	0.038	54	96.5%	96.7%	-0.020	-0.010	0.038	52
	$\theta_1 = 0.0$			0.035	0.017	0.060	159			0.033	0.015	0.057	166
	$\theta_2 = 0.6$			0.586	0.586	0.050	369			0.587	0.587	0.049	391
	$\theta_3 = 0.0$			0.024	0.012	0.043	229			0.023	0.010	0.042	205
	$\alpha = -1$			-0.824	-0.847	0.161	31			-0.827	-0.851	0.158	32
	$\phi_1 = 0.0$			-0.046	-0.027	0.060	33			-0.044	-0.026	0.057	34
	$\phi_2 = 0.0$			-0.008	-0.004	0.040	41			-0.009	-0.005	0.039	41
5000	$\phi_3 = 0.0$	96.3%	96.4%	-0.020	-0.012	0.038	51	95.8%	96.0%	-0.020	-0.011	0.038	50
	$\theta_1 = 0.0$			0.036	0.018	0.060	152			0.034	0.017	0.057	158
	$\theta_2 = 0.6$			0.586	0.586	0.050	350			0.587	0.587	0.049	367
	$\theta_3 = 0.0$			0.025	0.012	0.043	219			0.024	0.011	0.042	194

3.2 Effects of Thinning

Thinning is a technique usually applied when a sample presents considerable autocorrelation. In this section we evaluate the effects of thinning in terms of point and interval estimation, as well as in the effective sample size of each parameter and model identification. The model parameters and other details are kept the same as in Section 3.1. Also notice that no burn-in was applied in this exercise, so that results when no thinning is applied correspond to the case of no burn-in in the previous section.

3.2.1 GAR(p) models

Considering the GAR(1) and GAR(2) models presented in Section 3.1, we now study the effects of applying thinning of lags $\{5, 10, 20\}$ in the posterior samples prior to inference. The case of no thinning corresponds to the case of no burn-in presented in the previous section. The results are presented in Table 5 and 6. Regarding point estimation, applying any thinning does not improve the results in any way. This is expected since both the sample mean and sample median are consistent estimator even under dependence in the data. Hence, even applying a thinning of 20, the final sample is of size 1,500, which is still sufficiently large to guarantee that the sample mean and sample median are very close to the ones obtained with no thinning. Similar reasoning apply to the construction of credibility intervals, which in turn imply that thinning is expected to have little impact on model selection. These results are all reasonable considering that thinning is mainly used to reduce the correlation in the sample improving effective sample size. So, does effective sample size values improve after application of the thinning? Well, not quite. The simulation results shown borderline improvements at best, and even some decline in a few cases, especially for the GAR(2) model.

Table 5: RJMCMC Simulation Results for GAR(1) Models considering thinning $\{5, 10, 20\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Thinning	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -0.5$			-0.509	-0.503	0.149	63			-0.524	-0.508	0.160	47
	$\phi_1 = -0.4$			-0.369	-0.385	0.093	68			-0.358	-0.382	0.104	45
	$\phi_2 = 0.0$			-0.012	-0.001	0.054	103			-0.010	-0.001	0.049	83
5	$\phi_3 = 0.0$	74.0%	66.5%	-0.011	-0.003	0.043	166	55.7%	48.2%	-0.012	-0.003	0.042	125
	$\theta_1 = 0.0$			-0.027	-0.002	0.090	65			-0.037	-0.003	0.100	42
	$\theta_2 = 0.0$			0.018	0.002	0.061	128			0.019	0.003	0.058	99
	$\theta_3 = 0.0$			0.006	0.001	0.044	244			0.007	0.001	0.043	200
	$\alpha = -0.5$			-0.509	-0.503	0.149	60			-0.524	-0.508	0.160	45
	$\phi_1 = -0.4$			-0.369	-0.385	0.093	64			-0.358	-0.382	0.104	43
	$\phi_2 = 0.0$			-0.012	-0.001	0.054	99			-0.010	-0.001	0.049	83
10	$\phi_3 = 0.0$	73.7%	66.6%	-0.011	-0.003	0.043	161	55.7%	48.2%	-0.012	-0.003	0.042	123
	$\theta_1 = 0.0$			-0.027	-0.002	0.090	63			-0.037	-0.003	0.100	41
	$\theta_2 = 0.0$			0.018	0.002	0.062	122			0.019	0.003	0.058	93
	$\theta_3 = 0.0$			0.006	0.001	0.044	234			0.007	0.001	0.043	191
	$\alpha = -0.5$			-0.509	-0.503	0.150	58			-0.524	-0.508	0.160	43
	$\phi_1 = -0.4$			-0.369	-0.385	0.093	61			-0.358	-0.382	0.104	41
	$\phi_2 = 0.0$			-0.012	-0.001	0.054	98			-0.010	-0.001	0.049	82
20	$\phi_3 = 0.0$	73.8%	66.8%	-0.011	-0.003	0.043	157	55.9%	48.4%	-0.012	-0.003	0.042	121
	$\theta_1 = 0.0$			-0.027	-0.002	0.090	61			-0.037	-0.003	0.100	40
	$\theta_2 = 0.0$			0.018	0.002	0.062	118			0.019	0.003	0.058	89
	$\theta_3 = 0.0$			0.006	0.001	0.044	217			0.007	0.001	0.043	179

Thinning	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -0.5$			-0.532	-0.510	0.167	40			-0.538	-0.513	0.167	38
	$\phi_1 = -0.4$			-0.352	-0.381	0.110	38			-0.350	-0.380	0.112	36
	$\phi_2 = 0.0$			-0.010	-0.001	0.047	78			-0.009	-0.001	0.045	76
5	$\phi_3 = 0.0$	47.7%	40.9%	-0.012	-0.003	0.040	118	44.4%	37.5%	-0.011	-0.002	0.039	108
	$\theta_1 = 0.0$			-0.042	-0.004	0.106	35			-0.045	-0.005	0.108	32
	$\theta_2 = 0.0$			0.020	0.004	0.057	84			0.020	0.003	0.056	77
	$\theta_3 = 0.0$			0.006	0.001	0.041	195			0.006	0.001	0.040	171
	$\alpha = -0.5$			-0.532	-0.510	0.167	38			-0.538	-0.513	0.167	36
	$\phi_1 = -0.4$			-0.352	-0.381	0.110	36			-0.350	-0.380	0.112	34
	$\phi_2 = 0.0$			-0.010	-0.001	0.047	78			-0.009	-0.001	0.045	76
10	$\phi_3 = 0.0$	47.7%	40.9%	-0.012	-0.003	0.040	117	44.2%	37.5%	-0.011	-0.002	0.039	107
	$\theta_1 = 0.0$			-0.042	-0.004	0.106	34			-0.045	-0.005	0.108	32
	$\theta_2 = 0.0$			0.020	0.004	0.057	79			0.020	0.003	0.056	75
	$\theta_3 = 0.0$			0.006	0.001	0.041	188			0.006	0.001	0.040	166
	$\alpha = -0.5$			-0.532	-0.510	0.167	36			-0.538	-0.514	0.167	34
	$\phi_1 = -0.4$			-0.352	-0.381	0.110	34			-0.350	-0.380	0.112	32
	$\phi_2 = 0.0$			-0.010	-0.001	0.047	78			-0.009	-0.001	0.045	76
20	$\phi_3 = 0.0$	47.6%	41.3%	-0.012	-0.003	0.040	116	44.3%	37.7%	-0.011	-0.002	0.039	107
	$\theta_1 = 0.0$			-0.042	-0.004	0.106	33			-0.045	-0.005	0.108	31
	$\theta_2 = 0.0$			0.020	0.004	0.057	77			0.020	0.003	0.056	72
	$\theta_3 = 0.0$			0.006	0.001	0.041	176			0.006	0.001	0.040	158

Table 6: RJMCMC Simulation Results for GAR(2) Models considering thinning $\{5, 10, 20\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Thinning	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1$			-0.942	-0.966	0.136	74			-0.949	-0.969	0.128	68
	$\phi_1 = 0.0$			-0.036	-0.006	0.082	70			-0.033	-0.006	0.075	65
	$\phi_2 = -0.4$			-0.394	-0.396	0.063	210			-0.391	-0.395	0.069	146
5	$\phi_3 = 0.0$	98.7%	97.2%	-0.021	-0.006	0.049	172	95.9%	89.2%	-0.019	-0.006	0.045	140
	$\theta_1 = 0.0$			0.034	0.005	0.083	76			0.031	0.004	0.076	75
	$\theta_2 = 0.0$			-0.002	-0.000	0.062	194			-0.004	-0.000	0.067	127
	$\theta_3 = 0.0$			0.011	0.004	0.048	264			0.010	0.004	0.045	202
	$\alpha = -1$			-0.942	-0.966	0.136	70			-0.949	-0.969	0.128	65
	$\phi_1 = 0.0$			-0.036	-0.006	0.082	69			-0.033	-0.006	0.075	64
	$\phi_2 = -0.4$			-0.394	-0.396	0.063	202			-0.391	-0.395	0.069	140
10	$\phi_3 = 0.0$	98.7%	97.2%	-0.021	-0.006	0.049	162	95.8%	89.1%	-0.019	-0.006	0.045	136
	$\theta_1 = 0.0$			0.034	0.005	0.083	75			0.031	0.004	0.076	73
	$\theta_2 = 0.0$			-0.002	-0.000	0.062	189			-0.004	-0.000	0.067	124
	$\theta_3 = 0.0$			0.011	0.004	0.048	252			0.010	0.004	0.045	196
	$\alpha = -1$			-0.942	-0.966	0.136	68			-0.949	-0.969	0.129	63
	$\phi_1 = 0.0$			-0.036	-0.006	0.083	68			-0.033	-0.006	0.075	65
	$\phi_2 = -0.4$			-0.394	-0.396	0.063	195			-0.391	-0.395	0.069	136
20	$\phi_3 = 0.0$	98.7%	97.4%	-0.021	-0.006	0.049	154	95.6%	89.5%	-0.019	-0.006	0.045	132
	$\theta_1 = 0.0$			0.034	0.005	0.083	73			0.031	0.004	0.076	71
	$\theta_2 = 0.0$			-0.002	-0.000	0.062	184			-0.004	-0.000	0.067	122
	$\theta_3 = 0.0$			0.011	0.004	0.048	237			0.010	0.004	0.045	189
Thinning	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1$			-0.956	-0.971	0.126	64			-0.957	-0.971	0.128	61
	$\phi_1 = 0.0$			-0.031	-0.005	0.069	68			-0.031	-0.006	0.069	62
	$\phi_2 = -0.4$			-0.388	-0.395	0.072	127			-0.387	-0.394	0.075	116
5	$\phi_3 = 0.0$	92.1%	81.5%	-0.018	-0.006	0.043	130	87.2%	76.8%	-0.018	-0.006	0.042	119
	$\theta_1 = 0.0$			0.029	0.004	0.070	80			0.029	0.005	0.070	73
	$\theta_2 = 0.0$			-0.007	-0.001	0.069	109			-0.008	-0.001	0.072	98
	$\theta_3 = 0.0$			0.009	0.004	0.043	177			0.009	0.004	0.043	163
	$\alpha = -1$			-0.956	-0.971	0.126	61			-0.957	-0.971	0.128	58
	$\phi_1 = 0.0$			-0.031	-0.005	0.069	68			-0.031	-0.006	0.069	62
	$\phi_2 = -0.4$			-0.388	-0.395	0.072	121			-0.387	-0.394	0.076	110
10	$\phi_3 = 0.0$	92.1%	81.6%	-0.018	-0.006	0.043	128	87.2%	76.8%	-0.018	-0.006	0.042	117
	$\theta_1 = 0.0$			0.029	0.004	0.070	78			0.029	0.005	0.070	72
	$\theta_2 = 0.0$			-0.007	-0.001	0.069	106			-0.008	-0.001	0.072	97
	$\theta_3 = 0.0$			0.009	0.004	0.043	173			0.009	0.004	0.043	159
	$\alpha = -1$			-0.956	-0.971	0.127	60			-0.957	-0.971	0.129	57
	$\phi_1 = 0.0$			-0.031	-0.005	0.070	68			-0.031	-0.006	0.069	63
	$\phi_2 = -0.4$			-0.388	-0.395	0.072	118			-0.387	-0.394	0.076	106
20	$\phi_3 = 0.0$	92.0%	81.8%	-0.018	-0.006	0.043	126	87.0%	77.2%	-0.018	-0.006	0.042	114
	$\theta_1 = 0.0$			0.029	0.004	0.070	76			0.029	0.005	0.070	70
	$\theta_2 = 0.0$			-0.007	-0.001	0.069	104			-0.008	-0.001	0.072	95
	$\theta_3 = 0.0$			0.009	0.004	0.043	166			0.009	0.004	0.043	156

3.2.2 GMA(q) models

Considering the GAR(1) and GAR(2) models presented in Section 3.1, Tables 7 and 8 display the simulation results obtained by applying a thinning of size $\{5, 10, 20\}$ prior to inference. Analogously to the results for GAR models, applying a thinning did not present a significant effect on point estimate, and little to no improvement in the effective sample size.

3.2.3 Summary

In summary, considering the scenarios studied in the paper, we have evidence that the use of a burn-in before proceeding with inference is very effective in improving point estimation and model selection, whereas using a thinning approach does not significantly improve effective sample size and its use is not recommended. Furthermore, the use of a burn-in mitigates the dependence in the scale hyperparameter otherwise observed in the results, allowing for a more reliable use of the method.

Table 7: Simulation Results for the GMA(1) model considering thinning $\{5, 10, 20\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Thinning	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -0.5$			-0.277	-0.301	0.181	11			-0.356	-0.389	0.153	11
	$\phi_1 = 0.0$			-0.075	-0.061	0.066	14			-0.044	-0.028	0.050	33
	$\phi_2 = 0.0$			-0.006	-0.002	0.033	29			-0.005	-0.002	0.025	40
5	$\phi_3 = 0.0$	89.3%	88.4%	-0.001	0.001	0.029	36	94.0%	93.7%	-0.003	-0.002	0.022	48
	$\theta_1 = -0.5$			-0.402	-0.408	0.080	202			-0.430	-0.437	0.071	610
	$\theta_2 = 0.0$			-0.011	-0.005	0.049	412			-0.005	-0.002	0.041	829
	$\theta_3 = 0.0$			-0.002	-0.001	0.040	1123			0.000	-0.000	0.036	1323
	$\alpha = -0.5$			-0.277	-0.301	0.181	11			-0.356	-0.389	0.153	11
	$\phi_1 = 0.0$			-0.075	-0.061	0.066	15			-0.044	-0.028	0.050	35
	$\phi_2 = 0.0$			-0.006	-0.002	0.033	30			-0.005	-0.002	0.025	43
10	$\phi_3 = 0.0$	89.2%	88.5%	-0.001	0.001	0.029	37	94.1%	93.7%	-0.003	-0.002	0.022	50
	$\theta_1 = -0.5$			-0.402	-0.408	0.080	148			-0.430	-0.437	0.071	444
	$\theta_2 = 0.0$			-0.011	-0.005	0.049	303			-0.005	-0.002	0.041	606
	$\theta_3 = 0.0$			-0.002	-0.001	0.040	756			0.000	-0.000	0.036	962
	$\alpha = -0.5$			-0.277	-0.301	0.181	11			-0.356	-0.389	0.153	12
	$\phi_1 = 0.0$			-0.075	-0.061	0.066	15			-0.044	-0.028	0.050	36
	$\phi_2 = 0.0$			-0.006	-0.002	0.033	31			-0.005	-0.002	0.025	44
20	$\phi_3 = 0.0$	89.3%	88.3%	-0.001	0.001	0.029	37	94.0%	93.7%	-0.003	-0.002	0.022	52
	$\theta_1 = -0.5$			-0.401	-0.408	0.081	103			-0.430	-0.437	0.072	294
	$\theta_2 = 0.0$			-0.011	-0.005	0.049	230			-0.005	-0.002	0.041	426
	$\theta_3 = 0.0$			-0.001	-0.001	0.040	515			0.000	-0.000	0.036	640
Thinning	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -0.5$			-0.352	-0.384	0.159	15			-0.358	-0.389	0.154	13
	$\phi_1 = 0.0$			-0.049	-0.034	0.054	34			-0.046	-0.031	0.052	46
	$\phi_2 = 0.0$			-0.004	-0.001	0.024	53			-0.005	-0.001	0.023	55
5	$\phi_3 = 0.0$	94.9%	95.0%	-0.001	-0.001	0.021	58	95.7%	95.3%	-0.001	-0.000	0.019	62
	$\theta_1 = -0.5$			-0.426	-0.432	0.075	535			-0.428	-0.435	0.074	543
	$\theta_2 = 0.0$			-0.008	-0.003	0.041	735			-0.006	-0.002	0.041	618
	$\theta_3 = 0.0$			-0.001	-0.000	0.035	1267			-0.001	-0.001	0.034	1120
	$\alpha = -0.5$			-0.352	-0.384	0.159	15			-0.358	-0.389	0.154	13
	$\phi_1 = 0.0$			-0.049	-0.034	0.054	35			-0.046	-0.031	0.052	48
	$\phi_2 = 0.0$			-0.004	-0.001	0.024	55			-0.005	-0.001	0.023	55
10	$\phi_3 = 0.0$	95.0%	95.0%	-0.001	-0.001	0.021	60	95.7%	95.2%	-0.001	-0.000	0.019	66
	$\theta_1 = -0.5$			-0.426	-0.432	0.075	384			-0.428	-0.435	0.074	409
	$\theta_2 = 0.0$			-0.007	-0.003	0.042	560			-0.006	-0.002	0.041	503
	$\theta_3 = 0.0$			-0.001	-0.000	0.035	967			-0.001	-0.001	0.034	890
	$\alpha = -0.5$			-0.352	-0.384	0.159	15			-0.358	-0.389	0.154	14
	$\phi_1 = 0.0$			-0.049	-0.034	0.054	36			-0.046	-0.031	0.052	48
	$\phi_2 = 0.0$			-0.004	-0.001	0.024	57			-0.005	-0.001	0.023	57
20	$\phi_3 = 0.0$	95.1%	95.0%	-0.001	-0.001	0.021	63	95.8%	95.2%	-0.001	-0.000	0.019	70
	$\theta_1 = -0.5$			-0.426	-0.432	0.075	264			-0.428	-0.435	0.075	284
	$\theta_2 = 0.0$			-0.007	-0.003	0.042	417			-0.006	-0.002	0.041	383
	$\theta_3 = 0.0$			-0.001	-0.001	0.035	676			-0.001	-0.001	0.034	650

Table 8: RJMCMC Simulation Results for GMA(2) Models considering thinning $\{5, 10, 20\}$ and $\sigma \in \{0.5, 5, 10, 15\}$.

Thinning	Pars	$\sigma = 0.5$						$\sigma = 5$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1$			-0.813	-0.843	0.191	31			-0.929	-0.951	0.121	110
	$\phi_1 = 0.0$			-0.050	-0.027	0.072	32			-0.029	-0.007	0.067	94
	$\phi_2 = 0.0$			-0.009	-0.003	0.047	48			-0.005	-0.001	0.048	206
5	$\phi_3 = 0.0$	98.8%	99.1%	-0.020	-0.010	0.044	59	99.4%	99.4%	-0.014	-0.006	0.038	266
	$\theta_1 = 0.0$			0.040	0.017	0.072	93			0.025	0.003	0.065	108
	$\theta_2 = 0.6$			0.586	0.586	0.054	168			0.587	0.588	0.052	455
	$\theta_3 = 0.0$			0.026	0.012	0.049	132			0.019	0.006	0.044	185
	$\alpha = -1$			-0.813	-0.843	0.191	30			-0.929	-0.951	0.122	105
	$\phi_1 = 0.0$			-0.050	-0.027	0.072	32			-0.029	-0.007	0.067	94
	$\phi_2 = 0.0$			-0.009	-0.003	0.047	47			-0.005	-0.001	0.048	204
10	$\phi_3 = 0.0$	98.7%	99.1%	-0.020	-0.010	0.044	59	99.4%	99.4%	-0.014	-0.006	0.038	262
	$\theta_1 = 0.0$			0.040	0.017	0.072	86			0.025	0.003	0.065	107
	$\theta_2 = 0.6$			0.586	0.586	0.054	160			0.587	0.588	0.052	430
	$\theta_3 = 0.0$			0.026	0.012	0.049	122			0.019	0.006	0.044	176
	$\alpha = -1$			-0.813	-0.843	0.191	30			-0.929	-0.951	0.122	102
	$\phi_1 = 0.0$			-0.050	-0.027	0.072	32			-0.029	-0.007	0.067	93
	$\phi_2 = 0.0$			-0.009	-0.003	0.047	47			-0.005	-0.001	0.048	202
20	$\phi_3 = 0.0$	98.7%	99.1%	-0.020	-0.010	0.044	58	99.5%	99.4%	-0.014	-0.006	0.038	252
	$\theta_1 = 0.0$			0.040	0.017	0.072	75			0.025	0.003	0.065	103
	$\theta_2 = 0.6$			0.586	0.586	0.055	151			0.587	0.587	0.053	396
	$\theta_3 = 0.0$			0.026	0.012	0.049	111			0.019	0.006	0.044	167
Thinning	Pars	$\sigma = 10$						$\sigma = 15$					
		HPD	Quant	Mean	Med	SD	ESS	HPD	Quant	Mean	Med	SD	ESS
	$\alpha = -1$			-0.836	-0.863	0.169	32			-0.841	-0.868	0.165	32
	$\phi_1 = 0.0$			-0.042	-0.022	0.061	35			-0.040	-0.021	0.057	35
	$\phi_2 = 0.0$			-0.008	-0.003	0.041	45			-0.008	-0.004	0.040	45
5	$\phi_3 = 0.0$	99.2%	99.2%	-0.019	-0.010	0.039	53	99.0%	98.8%	-0.019	-0.009	0.038	50
	$\theta_1 = 0.0$			0.033	0.014	0.060	135			0.030	0.013	0.057	145
	$\theta_2 = 0.6$			0.585	0.586	0.053	238			0.586	0.586	0.054	255
	$\theta_3 = 0.0$			0.023	0.010	0.044	179			0.022	0.009	0.042	167
	$\alpha = -1$			-0.836	-0.863	0.169	32			-0.841	-0.868	0.166	32
	$\phi_1 = 0.0$			-0.042	-0.022	0.061	35			-0.040	-0.021	0.057	35
	$\phi_2 = 0.0$			-0.008	-0.003	0.041	46			-0.008	-0.004	0.040	45
10	$\phi_3 = 0.0$	99.2%	99.2%	-0.019	-0.010	0.039	54	99.0%	98.8%	-0.018	-0.009	0.038	50
	$\theta_1 = 0.0$			0.033	0.014	0.060	122			0.030	0.013	0.057	134
	$\theta_2 = 0.6$			0.585	0.586	0.053	220			0.586	0.586	0.054	234
	$\theta_3 = 0.0$			0.023	0.010	0.044	161			0.022	0.009	0.042	152
	$\alpha = -1$			-0.836	-0.863	0.170	31			-0.841	-0.867	0.166	31
	$\phi_1 = 0.0$			-0.042	-0.022	0.061	35			-0.040	-0.021	0.057	36
	$\phi_2 = 0.0$			-0.008	-0.003	0.041	46			-0.008	-0.004	0.040	46
20	$\phi_3 = 0.0$	99.2%	99.2%	-0.019	-0.010	0.039	54	99.0%	98.8%	-0.018	-0.009	0.038	50
	$\theta_1 = 0.0$			0.033	0.014	0.061	110			0.030	0.013	0.057	116
	$\theta_2 = 0.6$			0.585	0.586	0.054	203			0.586	0.586	0.055	211
	$\theta_3 = 0.0$			0.023	0.010	0.044	140			0.022	0.009	0.043	138

4 Application

In this section we present an application of the proposed methodology to analyzed the automobile production in Brazil between January 1993 and December 2013, which yields a sample size of $n = 252$ observations. The same data was considered in de Andrade et al. (2015). As in the mentioned work, the data was divided by 1,000 to reduce its magnitude and rounded to the nearest integer whenever necessary. The data is freely available from the ANFAVEA's - the Brazilian National Association of Motor Vehicle Manufacturers - website: `http://www.anfavea.com.br`. In de Andrade et al. (2015) the authors fit a negative binomial GARMA$(1,1)$ model to the data under a Bayesian framework. We are especially interested in model selection, conducted using information criteria as guideline in the aforementioned paper. Instead, we shall conduct model selection using the proposed RJMCMC approach.

The time series plot is presented in Figure 1 (left) and reveal the presence of a visible increasing trend. To account for that, de Andrade et al. (2015) considered a logarithmic trend as covariate in the model. However, considering the data directly, a linear trend yields a better fit. To see why is this the case, let $y_1, \cdots, y_n$ denote the observed time series. We fit the following linear models to the data:

$$\text{M1}: \ y_t = a_0 + a_1 \log(t) + e_t \quad \text{and} \quad \text{M2}: \ y_t = b_0 + b_1 t + e_t,$$

where e_t denotes a generic error term. The ordinary least squares estimates of the models are $\hat{a}_0 = -53.74$, $\hat{a}_1 = 44.89$, $\hat{b}_0 = 59.41$ and $\hat{b}_1 = 0.72$. The time series plot along with the fitted values for M1 and M2 are shown on Figure 1 (right). For M1, $R^2 = 0.54$, with residual standard error of 40.06, while for M2 we have $R^2 = 0.79$ and residual standard error of 26.9. These results favor the linear trend as a better fit for the long term growth observed in the time series. However, since GARMA models are defined in a GLM fashion, it may be the case that the linear trend does not outperform the logarithmic one when the GARMA structure is considered. To decide which one is more appropriate to model the data, we shall embed the trend term into the RJMCMC strategy, including the choice of trend selection along with model selection. The most complex GARMA(p_m, q_m) we shall consider consists of random component given by (6) along with deterministic component given by

$$g(\mu_t) = \beta_0 + \beta_1 t + \beta_2 \log(t) + \sum_{j=1}^{p_m} \phi_j \big[\log(Y^*_{t-j}) - \beta_1(t-j) - \beta_2 \log(t-j)\big] + \sum_{j=1}^{q_m} \theta_j r_{t-j}, \quad (8)$$

with $r_t := \log(Y^*_t) - \log(\mu_t)$ and $Y^*_t = \max\{0.3, Y_t\}$, that is, we set $c = 0.3$. Parameter β_0 is always included in the sampler, while all other parameters are targets for model transition.

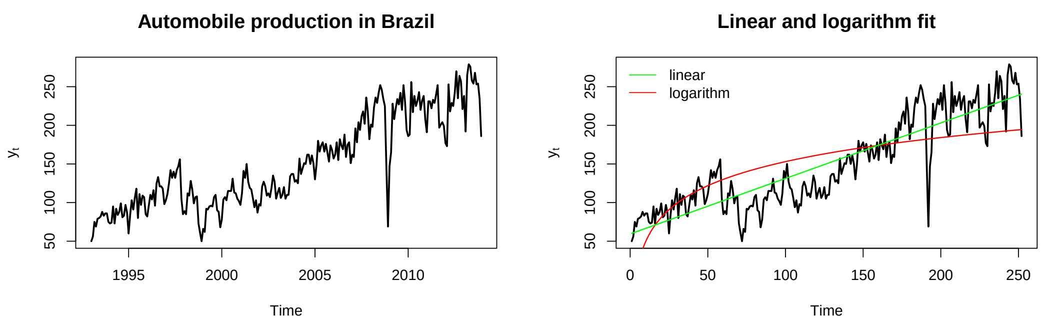

Figure 1: Automobile production in Brazil from January 1993 to December 2013. Shown are the time series plot alone in the left and along the fitted linear and logarithm trends on the right.

We found that convergence is very slow in this scenario, so that the RJMCMC is configured to produce a single chain containing 300,000 iterations, from which the first 295,000 are discarded as burn-in. The scale hyperparameter is set to 5, $m = 150$ just as in de Andrade et al. (2015) and the inclusion probability for each parameter is set to 0.5. All parameters are initialized in `Nimble` as 0. The prior distributions are given by: $\beta_0 \sim N(0, 0.3^2)$, $\phi_i \sim N(0, 0.2^2)$, $\theta_i \sim N(0, 0.2^2)$, $\beta_j \sim N(0, 16)$, for $i \in \{1, 2, 3\}$ and $j \in \{1, 2\}$.

The first exercise comprises setting $p_m = q_m = 3$ and running the RJMCMC. The results are presented in Table 9 and the time series plot of the generated chain are show in Figure 2. In the last column of Table 9 we present Geweke's convergence diagnostic (GCD), which is a convergence diagnostic for Markov chains based on a test for equality of the means of the first 10% and last 50% part of a Markov chain (Geweke, 1991). The values displayed are the z-scores calculated under the assumption that the two parts of the chain are asymptotically independent. We observe that all values are smaller than 1.96 in absolute value, indicating that the chain of each parameter converged to its target distribution, considering a 95% confidence level. From point estimates, the first thing we notice is that the RJMCMC selected the

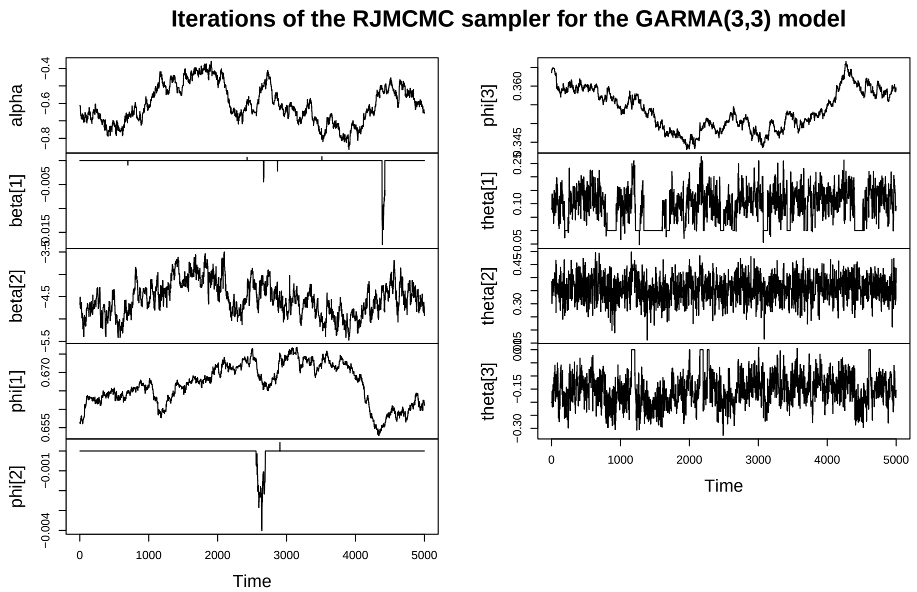

Figure 2: Iterations of the RJMCMC sampler.

logarithm long term growth, keeping β_1 out of the model in almost all iterations. This also happens with parameters ϕ_2, excluded in almost all iterations. Besides β_1 and ϕ_2, θ_1 is also non-significant considering a 95% confidence level HPD credibility interval, although it has been selected many times to be included in the model. All other parameters can be considered significant according to the HPD credible interval. Median and mean estimates are also very close together indicating symmetry of the target distribution. The roots of the characteristic polynomial for the AR component are all greater than 1.236, hence lying outside of the unitary circle.

In Figure 3, we present the reconstructed conditional mean μ_t based on the (mean) estimated values along with the original time series. This seemingly delayed pattern is commonly seen in GARMA models containing autoregressive components. As expected, μ_t accompanies y_t very closely, indicating that the model is a good fit. In de Andrade et al. (2015) based on information criteria, the authors selected a BN-GARMA$(1, 1)$ with a logarithm trend as the best model among those considered in the paper. By using the proposed RJMCMC approach,

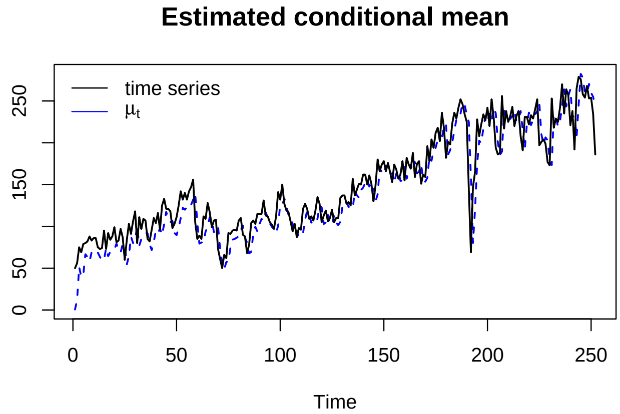

Figure 3: Reconstructed conditional mean.

Table 9: Results from fitting a BN-GARMA$(3, 3)$ model defined in (8).

Par	Mean	Median	SD	HPD CI (95%)	GCD
β_0	-0.606	-0.620	0.109	$[-0.786, -0.402]$	-1.231
β_1	0.000	0.000	0.001	–	1.229
β_2	-4.542	-4.571	0.377	$[-5.203, -3.750]$	-1.763
ϕ_1	0.667	0.667	0.005	$[0.657, 0.676]$	-0.982
ϕ_2	0.000	0.000	0.000	–	0.696
ϕ_3	0.354	0.353	0.005	$[0.344, 0.362]$	1.820
θ_1	0.093	0.099	0.063	$[0.000, 0.196]$	-0.077
θ_2	0.358	0.360	0.047	$[0.267, 0.445]$	0.638
θ_3	-0.154	-0.156	0.060	$[-0.285, -0.049]$	-0.983

we selected a more complex model, which is technically a BN-GARMA$(3, 3)$, but with coefficients ϕ_2 and θ_1 equal zero. The method also selected the logarithm long term growth as the "correct" one for the data. Unfortunately, a deeper comparison between the results of the present work and those presented in de Andrade et al. (2015) cannot be done as some key information to allow for a fair comparison are missing in the mentioned paper. For instance, there is no indication regarding the value of the constant c applied, nor about the number of iterations and the burn-in used in the paper.

5 Conclusion

In this paper, we tackle the problem of order selection in GARMA models for count time series from a Bayesian perspective, using the approach known as *Reversible Jump Markov Chain Monte Carlo* (RJMCMC). The study successfully achieved its main objective of investigating the selection of GARMA count models in the Bayesian context, through the RJMCMC approach. The sensitivity analysis regarding the choice of hyperparameters for the priors was also addressed, providing valuable insights into the method's robustness and flexibility. The RJMCMC simulations revealed that the implementation of a burn-in is consistently beneficial, resulting in notable improvements in all cases and metrics. This effect is particularly evident in the significant reduction of the impact of σ, making the results more reliable, with a notable improvement in the correct identification of models. Applying a burn-in showed significant improvements for GAR(1), while for GAR(2) the benefits were less pronounced, and for the GMA(q) model, the influence when applying a burn-in on the point estimate is significantly less notable compared to the GAR(p) model.

In contrast, the application of thinning between lags did not produce substantial improvements in point estimation or effective sample size, indicating that application of this procedure in the context of GARMA models is not advisable. In section 4, we address the empirical application in real-world datasets, which demonstrated the practical relevance of the proposed method, highlighting its ability to handle real-world situations.

In conclusion, the simulations and applications confirm the effectiveness of the RJMCMC method in the selection of GARMA count models for time series, highlighting the utility of the adaptation period and the limited contribution of lag spacing in the studied scenarios. As future work, we intend to explore model selection in GARMA models with continuous distributions under the context of RJMCMC. Furthermore, we contemplate the development of a package or script dedicated to sharing the computational implementation of the method, thereby making the approach more accessible to both researchers and professionals interested in the field.

References

Bayer, F. M., Bayer, D. M., and Pumi, G. (2017). Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*, 555:385–396.

Bayer, F. M., Pumi, G., Pereira, T. L., and Souza, T. C. (2023). Inflated beta autoregressive moving average models. *Computational And Applied Mathematics*, 42:183.

Benaduce, H. S. and Pumi, G. (2023). SYMARFIMA: a dynamical model for conditionally symmetric time series with long range dependence mean structure. *Journal of Statistical Planning and Inference*, 225:71–88.

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

Brännäs, K. and Johansson, P. (1994). Time series count data regression. *Communications in Statistics-Theory and Methods*, 23(10):2907–2925.

Casarin, R., Valle, L. D., and Leisen, F. (2010). Bayesian Model Selection for Beta Autoregressive Processes. *Bayesian Analysis*, 7(2):385–410.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 93–115.

Davis, R. A., Dunsmuir, W. T., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3):491–505.

de Andrade, B. S., Andrade, M. G., and Ehlers, R. S. (2015). Bayesian GARMA models for count data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1(4):192–205.

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., and Paganin, S. (2023). *NIMBLE User Manual*. R package manual version 1.0.1.

Eğrioğlu, E. and Günay, S. (2010). Bayesian model selection in ARFIMA models. *Expert Systems with Applications*, 37(12):8359–8364.

Freeland, R. K. and McCabe, B. P. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of time series analysis*, 25(5):701–722.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Staff Report 148, Federal Reserve Bank of Minneapolis.

Grande, A. F., Pumi, G., and Cybis, G. B. (2023). Bayesian analysis of beta autoregressive moving average models.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(4):711–732.

Liesenfeld, R., Nolte, I., and Pohlmeier, W. (2006). Modelling financial transaction price movements: a dynamic integer count data model. *Empirical Economics*, 30:795–825.

Melo, M. and Alencar, A. (2020). Conway–Maxwell–Poisson Autoregressive Moving Average Model for Equidispersed, Underdispersed, and Overdispersed Count Data. *Journal of Time Series Analysis*, 41(6):830–857.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Pumi, G., Valk, M., Bisognin, C., Bayer, F. M., and Prass, T. S. (2019). Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference*, 200:196 – 212.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rocha, A. V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models. *TEST*, 18(3):529–545.

Sales, L. O., Alencar, A. P., and Ho, L. L. (2022). The BerG generalized autoregressive moving average model for count time series. *Computers & Industrial Engineering*, 168:108104.

Snow, G. (2020). *TeachingDemos: Demonstrations for Teaching and Learning*. R package version 2.12.

Troughton, P. T. (1999a). *Simulation methods for linear and nonlinear time series models with application to distorted audio signals*. PhD thesis, University of Cambridge.

Troughton, P. T. (1999b). *Simulation methods for linear and nonlinear time series models with application to distorted audio signals*. PhD thesis, University of Cambridge.

Weiß, C. H. (2007). Controlling correlated processes of poisson counts. *Quality and reliability engineering international*, 23(6):741–754.

Zeger, S. L. and Qaqish, B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics*, 44(4):1019–1031.