UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
DOUTORADO EM ESTATÍSTICA

VINÍCIUS TEODORO SCHER

ESSAYS ON DOUBLE BOUNDED TIME SERIES ANALYSIS

Recife
2022

VINÍCIUS TEODORO SCHER

ESSAYS ON DOUBLE BOUNDED TIME SERIES ANALYSIS

Doctoral dissertation submitted to the Graduate Program in Statistics, Department of Statistics, Universidade Federal de Pernambuco as a requirement to obtain a Doctoral degree in Statistics. Concentration area: Mathematical Statistics

Advisor: Francisco Cribari-Neto
Co-advisor: Fábio Mariano Bayer

Recife

2022

Catalogação na fonte
Bibliotecária Nataly Soares Leite Moro, CRB4-1722

S326e Scher, Vinícius Teodoro
Essays on double bounded time series analysis / Vinícius Teodoro Scher. – 2022.
96 f.: il., fig., tab.

Orientador: Francisco Cribari Neto.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2022.
Inclui referências e apêndices.

1. Estatística matemática. 2. βARMA. 3. Bootstrap. 4. Simulação de Monte Carlo. 5. Testes portmanteau I. Cribari Neto, Francisco (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- CCEN 2022 - 47

VINICIUS TEODORO SCHER

Essays on Double Bounded Time Series Analysis

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 18 de fevereiro de 2022.

BANCA EXAMINADORA

Francisco Cribari Neto
UFPE

Patricia Leone Espinheira Ospina
UFPE

Raydonal Ospina Martinez
UFPE

Airlane Pereira Alencar
USP

Pedro Alberto Morettin
USP

To my father, with love.

AGRADECIMENTOS

Inicialmente, quero dedicar toda minha jornada de pós-graduação a minha mãe, Maria, e a meu pai, Olavo (in memoriam). Obrigado por sempre terem acreditado em meus sonhos, provendo todo o suporte afetivo e financeiro nos momentos em que mais deles precisei. A vocês, meu eterno amor e minha eterna gratidão por tudo que sou. Obrigado por me terem dado o conforto de um lar. Por mais difícil que fosse prover alimento em nossa mesa, nunca nada me foi negado quando o assunto envolvia meu futuro profissional.

Meu agradecimento especial a meu orientador, professor Francisco Cribari-Neto. O caminho da pós-graduação por muitas vezes é árduo, fazendo-nos repensar se é isso mesmo que desejamos. Sem os seus conselhos e sua mão para me levantar nas vezes em que tropecei, eu não teria conseguido chegar até aqui. Obrigado por todo carinho, paciência e dedicação em me conduzir nessa travessia. Eu gostaria que todos os meus colegas tivessem a sorte que eu tive em ter um verdadeiro mentor e guia nessa longa jornada. Você foi a luz nos momentos de escuridão e a inspiração profissional que me motivou a perseverar. Que seu coração continue gigante por toda eternidade! Também quero agradecer a meu coorientador, professor Fábio Bayer, por me ter aconselhado a seguir na carreira acadêmica após o término da graduação. Se eu não houvesse seguido seu conselho, eu não estaria aqui e não teria alcançado tantas conquistas.

Meu muito obrigado aos professores do Departamento de Estatística da Universidade Federal de Pernambuco (UFPE), em especial aos professores Alex Dias, Gauss Cordeiro, Klaus Vasconcellos, Patrícia Espinheira e Raydonal Ospina. Carregarei comigo seus ensinamentos e conselhos; vocês fazem parte de cada conquista minha até o momento. Agradeço ao Prof. Guilherme Pumi pela preciosa colaboração e parceria científica.

Um agradecimento especial a Kalline Costa, que durante grande parte do doutorado esteve ao meu lado, sendo uma grande companheira, amiga e namorada. Acalmou meu coração nos momentos difíceis e me presentou diariamente com um amor puro, verdadeiro e genuíno. Obrigado por caminhar comigo.

Agradeço a meus colegas e irmãos de caminhada Ana Cristina Guedes, César Diogo, Jucelino Matos, Yuri Araújo. Sua amizade e seu companheirismo transformaram meus dias em algo muito melhor, dando-me forças para seguir adiante. Levarei vocês em meu coração para sempre.

Agradeço ao casal de amigos Sarah Rito e Jairo Maciel, que me abraçaram quando mais

precisei, aquecendo meu coração e me despertando para uma nova vida. Vocês são incríveis! Muito obrigado por todo o amor.

Nos momentos em que precisei desabafar, a amiga Viviane Sales sempre esteve por perto. Obrigado, Vivi! Que ainda possamos debater por muito e muitos anos essa grande viagem chamada vida.

Aos parceiros de Castelo, Adson Belém, Lucas Tavares e Kalebe, obrigado pelos ensinamentos e pelo companheirismo.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE) e ao Centro Nacional de Supercomputação da Universidade Federal do Rio Grande do Sul (CESUP/UFRGS).

ABSTRACT

Two important steps in time series analysis are model selection and diagnostic analysis. We address the issue of performing diagnostic analysis through portmanteau testing inferences using time series data that assume values in the standard unit interval. Our focus lies in the class of beta autoregressive moving average (βARMA) models. In particular, we wish to test the goodness-of-fit of such models. We consider several testing criteria that have been proposed for Gaussian time series models and two new tests that were recently introduced in the literature. We derive the asymptotic null distribution of the two new test statistics in two different scenarios, namely: when the tests are applied to an observed time series and when they are applied to residuals from a fitted βARMA model. It is worth noticing that our results imply the asymptotic validity of standard portmanteau tests in the class of βARMA models that are, under the null hypothesis, asymptotically equivalent to the two new tests. We use Monte Carlo simulation to assess the relative merits of the different portmanteau tests when used with fitted βARMA. The simulation results we present show that the new tests are typically more powerful than a well known test whose test statistic is also based on residual partial autocorrelations. Overall, the two new tests perform quite well. We also model the dynamics of the proportion of stored hydroelectric energy in South of Brazil. The results show that the βARMA model outperforms three alternative models and an exponential smoothing algorithm. We also consider the issue of performing model selection with double bounded time series. We evaluate the effectiveness of βARMA model selection strategies based on different information criteria. The numerical evidence for autoregressive, moving average, and mixed autoregressive and moving average models shows that, overall, a bootstrap-based model selection criterion is the best performer. An empirical application which we present and discuss shows that the most accurate out-of-sample forecasts are obtained using bootstrap-based model selection. The βARMA model is tailored for use with fractional time series, i.e., time series that assume values in $(0, 1)$. We introduce a generalization of the model in which both the conditional mean and the conditional precision evolve over time. The standard βARMA model, in which precision is constant, is a particular case of our model. The more general formulation of the model includes a parsimonious submodel for the precision parameter. We present the model log-likelihood function, the score function, and Fisher's information matrix. We use the proposed model to forecast

future levels of stored hydrolectric energy in the South of Brazil. Our results show that more accurate forecasts are typically obtained by allowing the precision parameter to evolve over time.

Palavras-chave: βARMA; bootstrap; model selection; monte carlo simulation; portmanteau test; time series.

RESUMO

Duas etapas importantes na modelagem de séries temporais são seleção de modelos e análise de diagnóstico. No que diz respeito à análise de diagnóstico, nós abordamos a realização de inferências via testes portmanteau utilizando séries temporais que assumem valores no intervalo da unitário padrão. Nosso foco reside na classe de modelos beta autorregressivos e de médias móveis (βARMA). Em particular, desejamos testar a adequacidade de tais modelos. Nós consideramos diversos testes que foram propostos para modelos de séries temporais gaussianas e dois novos testes recentemente introduzidos na literatura. Derivamos a distribuição nula assintótica das duas novas estatísticas de teste em dois cenários diferentes, a saber: quando os testes são aplicados a uma série temporal observada e quando são aplicados a resíduos de um modelo βARMA. Vale a pena notar que nossos resultados implicam a validade assintótica dos testes portmanteau padrão na classe de modelos βARMA que são, sob hipótese nula, assintoticamente equivalente aos dois novos testes. Usamos simulação de Monte Carlo para avaliar os méritos relativos dos diferentes testes portmanteau quando usados conjuntamente com o modelo βARMA. Os resultados de simulação que apresentamos mostram que os novos testes são tipicamente mais poderosos que um teste bem conhecido, cuja estatística de teste também é baseada em autocorrelações parciais dos resíduos. No geral, os dois novos testes funcionam muito bem. Adicionalmente, modelamos a dinâmica da proporção de energia hidrelétrica armazenada no sul do Brasil. Os resultados mostram que o modelo βARMA supera três modelos alternativos e um algoritmo de suavização exponencial. Num segundo estudo, avaliamos a eficácia de estratégias de seleção de modelos com base em diferentes critérios de informação no modelo βARMA. A evidência numérica para modelos autorregressivos, de médias móveis e mistos (autorregressivos e de médias móveis) mostra que, em geral, um critério de seleção de modelos baseado em bootstrap apresenta o melhor desempenho. Nossa aplicação empírica mostra que as previsões mais precisas são obtidas usando seleção de modelo baseada em bootstrap. O modelo βARMA é adequado para uso com séries temporais fracionárias, ou seja, séries temporais que assumem valores em $(0, 1)$. Nós propomos uma generalização do modelo em que tanto a média condicional quanto a precisão condicional evoluem ao longo do tempo. O modelo βARMA padrão, em que a precisão é constante, é um caso particular do nosso modelo. A formulação mais geral do modelo inclui um submodelo parcimonioso para o parâmetro de precisão. Apresentamos

a função de log-verossimilhança do modelo, a função escore e a matriz de informação de Fisher. Utilizamos o modelo proposto para prever níveis futuros de energia hidroelétrica armazenada no Sul do Brasil. Nossos resultados mostram que previsões mais precisas são obtidas ao se permitir que o parâmetro de precisão evolua ao longo do tempo.

Palavras-chave: βARMA; bootstrap; seleção de modelos; séries temporais; simulação de monte carlo; testes portmanteau.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 GOODNESS-OF-FIT TESTS FOR βARMA HYDROLOGICAL TIME SERIES MODELING

1.1 INTRODUCTION

The beta regression model was introduced by Ferrari and Cribari-Neto (2004) for modeling dependent variables that assume values in the standard unit interval; see also Cribari-Neto and Zeileis (2010). Based on that model, Rocha and Cribari-Neto (2009) introduced the beta autoregressive moving average (βARMA) model, which is a dynamic model tailored for time series that assume values in $(0, 1)$, such as rates and proportions. Doubly bounded random variables are typically asymmetrically distributed and inferences based on the Gaussian assumption may be quite inaccurate. In the βARMA model, the variable of interest (y) is assumed to follow the beta law, its mean being impacted by a set of covariates and also subject to autoregressive and moving average dynamics. Novel features of the model are that it requires no data transformation and that βARMA fitted values and out-of-sample forecasts will never fall outside the standard unit interval.

Diagnostic analysis is of paramount importance in time series modeling. It is performed after the model has been identified and fitted. Different model validation strategies can be used. Perhaps the most commonly used validation strategy involves portmanteau testing inference. Such tests are based on statistics that use residual autocorrelations. They seek to detect any existing serial correlation in the residuals obtained from the fitted model.

Since the seminal article by Box and Pierce (1970), considerable attention has been devoted to tests that use residual autocorrelations to assess goodness-of-fit. Ljung and Box (1978) showed that a simple modification to the test statistic proposed by Box and Pierce (1970) considerably reduces the distribution location bias and improves the quality of the asymptotic approximation used in the test. Their statistic is widely used by practitioners; see, e.g., Chiogna and Gaetan (2005). Monti (1994) proposed to base portmanteau testing inference on a test statistic that uses residual partial autocorrelations rather than residual autocorrelations. Dufour and Roy (1986) introduced a non-parametric portmanteau test based on rank autocorrelations. Their test is particularly useful when the underlying distribution of the time series is unknown and it tends to deliver accurate infe-

rences under non-normality. Additional tests based on transformed sample autocorrelations were proposed by Kwan and Sim (1996a), Kwan and Sim (1996b). Peña and Rodriguez (2002) introduced a test based on the determinant of the residual autocorrelation matrix. They approximated the test statistic null distribution using the gamma distribution. Such an approximation can be poor in some situations. To circumvent such a shortcoming, Lin and McLeod (2006) proposed using bootstrap resampling to estimate the test statistic null distribution.

All aforementioned portmanteau tests were developed for standard ARMA models, i.e., for models used with variables that assume values in the real line. How do such tests perform when used with βARMA models? To the best of our knowledge, this question remains unanswered. The contribution of our work to the literature is twofold. First, we investigate the accuracy of portmanteau testing inference in the class of βARMA models. We also consider two new portmanteau tests proposed by Scher (2017) which are based on residual partial autocorrelations. Our numerical evidence shows that all tests are nearly free of size distortions when coupled with bootstrap resampling. By size distortion we mean the difference between exact and nominal type I error frequencies. Second, we derive the asymptotic null distribution of the two new test statistics in two different scenarios, namely: when the tests are applied to an observed time series and when they are applied to the residuals from a fitted βARMA model. A novel aspect of our proofs is that they also hold for the test statistics that were proposed in the Gaussian time series literature and that can be shown to be asymptotically equivalent to the two new test statistics under the null hypothesis. Hence, it follows that some portmanteau tests that were proposed for Gaussian ARMA models can also be used with βARMA models.

The simulation evidence we present shows that the two new tests are particularly powerful. In some scenarios of our numerical experiments, such tests were more powerful (i.e., more capable of detecting model misspecification) than all competing tests.

Our motivation in the present chapter lies in a hydrological empirical problem: we wish to model the time series dynamics of the proportion of stored hydroelectric energy in the South of Brazil. The data range from January 2001 to October 2016. We produce out-of-sample forecasts using a fitted βARMA model, another dynamic model tailored for time series that assume values in the standard unit interval, a Gaussian ARMA model, a Gaussian AR model, and an exponential smoothing algorithm. It is shown that the

βARMA yields the most accurate short term forecasts. Indeed, the six period ahead mean absolute forecasting error obtained with βARMA model is over 16% smaller than that of the second best performing method. Prior to using the fitted βARMA model for forecasting we validate it on the basis of portmanteau testing inference. We shall return to this application in Section 1.5.

The chapter is organized as follows. Section 1.2 presents the βARMA model and its main properties. In Section 1.3 we review some portmanteau tests for model adequacy that have been used in the literature. Two new tests are reviewed in Section 1.4. The asymptotic null distribution of the two new test statistics is derived under two different settings, namely: when the test statistics are computed using an observed time series and when they are computed using βARMA residuals. Monte Carlo simulation evidence is presented in Section 1.5. An empirical hydrological application is presented and discussed in Section 1.6. Finally, Section 1.7 contains some concluding remarks.

1.2 THE MODEL

The βARMA Rocha and Cribari-Neto (2009) model is a dynamic model based on the class of beta regression models Ferrari and Cribari-Neto (2004). It is useful for dealing with time series data that assume values in the standard unit interval, $(0,1)$, such as rates and proportions. The model includes autoregressive and moving average dynamics and also a set of regressors. It accommodates distributional asymmetries and non-constant dispersion. Unlike the standard ARMA model, fitted values and out-of-sample forecasts produced using βARMA are guaranteed to lie inside the standard unit interval.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ be an n-vector of time series random variables, where each y_t, for $t = 1, \ldots, n$, given the previous information set $\mathcal{F}_{t-1}$, follows the beta law, as parameterized in Ferrari and Cribari-Neto (2004), with conditional mean μ_t and precision parameter ϕ. The conditional density of y_t given $\mathcal{F}_{t-1}$ is

$$f(y_t \mid \mathcal{F}_{t-1}) = \frac{\Gamma(\phi)}{\Gamma(\mu_t\phi)\Gamma((1-\mu_t)\phi)} y_t^{\mu_t\phi-1}(1-y_t)^{(1-\mu_t)\phi-1}, \quad 0 < y_t < 1, \tag{1.1}$$

where $0 < \mu_t < 1$, $\phi > 0$ and $\Gamma(\cdot)$ is the gamma function. The conditional mean and the conditional variance of the y_t are, respectively, $\mathbb{E}(y_t \mid \mathcal{F}_{t-1}) = \mu_t$ and $\mathrm{var}(y_t \mid \mathcal{F}_{t-1}) = \mu_t(1-\mu_t)/(1+\phi)$. Note that μ_t is the mean of y_t and ϕ is a precision parameter, in the sense that for a fixed μ_t the variance of y_t decreases as ϕ increases.

By assuming that the variable of interest follows the above beta law, Rocha and Cribari-Neto (2009) proposed the following βARMA(p,q) model:

$$g(\mu_t) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i \left\{ g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} \right\} + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{1.2}$$

where $\boldsymbol{x}_t \in \mathbb{R}^c$ is a set of non-random covariates at time t, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_c)^\top \in \mathbb{R}^c$ is a vector of coefficients related to the covariates and $g : (0,1) \to \mathbb{R}$ is a twice differentiable strictly monotonic link function. Here, $\alpha \in \mathbb{R}$ is a scalar parameter and p, $q \in \mathbb{N}$ are, respectively, the autoregressive and moving average orders. Additionally, r_t is the error term and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ are the autoregressive and moving average vectors of parameters, respectively. The covariates are only required to be non-random and to satisfy (1.2) for the model to be well defined. Typically, the covariates are included in the model when some deterministic behavior (such as a cyclic or seasonal component) is present in the data dynamics; see Section 1.4.2 for details.

The formulation of Model (1.2) is similar to that in Benjamin, Rigby and Stasinopoulos (2003), who introduced the class of generalized autoregressive moving average (GARMA) models by extending previous results by Li (1994) and Zeger and Qaqish (1988) on non-Gaussian time series modeling. In both classes of models, the error term (r_t) is defined in residual fashion, that is, the errors do not drive the stochastic process like innovations in standard ARMA models. The Gaussian ARMA process is driven by the realization of a white noise error from a normal distribution. There is no white noise error disturbance in βARMA and GARMA models. The time series realizations come from a conditional distribution and the moving average error, r_t, is defined as the difference between an observed quantity (y_t or $g(y_t)$) and the corresponding model-based quantity (μ_t or $g(\mu_t)$, respectively). Notice that only past values of r_t are included Model (1.2). The two standard formulations for r_t are: (i) error on the original scale: $y_t - \mu_t$, and (ii) error on the predictor scale: $g(y_t) - \eta_t$, where $\eta_t = g(\mu_t)$. In what follows, we shall consider the latter. The error term r_t is $\mathcal{F}_{t-1}$-measurable. It is noteworthy that the error in the original scale ($r_t = y_t - \mu_t$) follows a martingale difference which implies that r_t has unconditional mean zero and is unconditionally uncorrelated. Notice that $\mathbb{E}(y_t|\mathcal{F}_{t-1}) = \mu_t$ which implies that $\mathbb{E}(r_t|\mathcal{F}_{t-1}) = 0$ and, hence, $\mathbb{E}(r_t) = 0$. Additionally, for $h \geq 1$, $\mathrm{cov}(r_t, r_{t+h}) = \mathbb{E}(r_t r_{t+h}) = \mathbb{E}(\mathbb{E}(r_t r_{t+h}|\mathcal{F}_{t+h-1})) = \mathbb{E}(r_r \mathbb{E}(r_{t+h}|\mathcal{F}_{t+h-1})) = 0$. Since r_t assumes values in $(-1,1)$, its unconditional variance is finite. Rocha and Cribari-Neto (2009) show that the error on the predictor scale has mean approximately equal to zero,

variance approximately equal to $(\partial\mu_t/\partial\eta_t)^2[\mu_t(1-\mu_t)/(1+\phi)]$ and that such errors are approximately orthogonal.

The estimation of the parameters that index the βARMA model is typically performed by conditional maximum likelihood Andersen (1970). The conditional log-likelihood function (given the first a observations) is $\ell = \sum_{t=a+1}^{n} \log f(y_t \mid \mathcal{F}_{t-1})$, where $a = \max\{p, q\}$ and $f(y_t \mid \mathcal{F}_{t-1})$ is presented in Equation (1.1).

When the model contains moving average components, it is necessary to take into account the recursive structure of log-likelihood derivatives Benjamin, Rigby and Stasinopoulos (1998). Using the predictor scale error, such derivatives are given by Rocha and Cribari-Neto (2017):

$$\begin{aligned}
\frac{\partial\eta_t}{\partial\alpha} &= 1 - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{t-j}}{\partial\alpha}, \quad \frac{\partial\eta_t}{\partial\boldsymbol{\beta}} = \boldsymbol{x}_t^\top - \sum_{i=1}^{p} \varphi_i \boldsymbol{x}_{t-i}^\top - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{t-j}}{\partial\boldsymbol{\beta}}, \\
\frac{\partial\eta_t}{\partial\varphi_i} &= g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top\boldsymbol{\beta} - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{t-j}}{\partial\varphi_i}, \quad i = 1, \ldots, p, \\
\frac{\partial\eta_t}{\partial\theta_l} &= g(y_{t-l}) - \eta_{t-l} - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{t-j}}{\partial\theta_l}, \quad l = 1, \ldots, q.
\end{aligned}$$

Starting values for η_t can be obtained by setting $\eta_t = g(y_t)$ and the derivatives of η with respect to the model parameters equal to zero for $t = 1, \ldots, q$ Benjamin, Rigby and Stasinopoulos (1998).

Bayesian model selection for the βARMA model was developed by Casarin, Valle and Leisen (2012) and bias-corrected maximum likelihood of the parameters that index the model was considered by Palm and Bayer (2018). An extension of the model that incorporates seasonal dynamics, the βSARMA model, was recently proposed by Bayer, Cintra and Cribari-Neto (2018) and an extension of the model for compositional data, the DARMA model ('D' stands for Dirichlet), was developed by Zheng and Chen (2017). A dynamic model for doubly bounded random variables based on an alternative law — the Kumarawasmy law — was introduced by Bayer, Bayer and Pumi (2017). In what follows we shall focus on the standard, baseline βARMA model.

1.3 STANDARD PORTMANTEAU TESTS

Portmanteau tests are commonly used in time series analysis to assess goodness-of-fit. The test statistics are based on residual autocorrelations and the fitted model is

taken as a good representation of the data when such autocorrelations are jointly negligible. In what follows we shall briefly present some well known portmanteau tests adapted for the βARMA model. Let w_t be a stationary time series of interest and let $\rho_k = \text{cor}(w_t, w_{t+k})$ be the kth order autocorrelation. For a user selected integer $m > 0$, we are interested in testing

$$\begin{aligned} &\mathcal{H}_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0 \\ &\mathcal{H}_1 : \text{at least one } \rho_i \neq 0. \end{aligned} \tag{1.3}$$

That is, the interest lies in testing that the first m autocorrelations are jointly equal to zero.

Oftentimes, w_t is the tth residual from a fitted time series model. In that case, the null hypothesis under test is that first m autocorrelations of the residuals from a fitted time series (ARMA or βARMA) are jointly equal to zero and its rejection is evidence of model misspecification. Let $\hat{r}_1, \ldots, \hat{r}_n$ denote the fitted model's residuals. The kth residual (sample) autocorrelation is $\hat{\rho}_k = \sum_{t=k+1}^{n} \hat{r}_t \hat{r}_{t-k} / \sum_{t=1}^{n} \hat{r}_t^2$, $k = 1, \ldots, m$. When the model is correctly specified, the residuals are expected to be nearly uncorrelated.

In Gaussian ARMA models, under the null hypothesis, the time series is a sequence of independent random variables whereas in βARMA model such a series is comprised of non-correlated variables.

Different residuals can be computed in the context of βARMA models. For instance, one can define residuals based on one the following discrepancies: $g(y_t) - g(\mu_t)$ and $y_t - \mu_t$. Recall that if, say, w_t follows an ARMA process, then aw_t follows an ARMA process of the same order $\forall a \neq 0$. It is important to notice, however, that transformations of ARMA processes are generally not ARMA processes, and in the specific cases where they are, the order of the process usually changes Linka (1988). In the present context, we notice that, when $y_t \sim \beta\text{ARMA}(p, q)$ without covariates, conditionally on $\mathcal{F}_{t-1}$, the sequence $\{(g(y_t), r_t)\}_t$ satisfies the ARMA(p, q) difference equations. Working with residuals based on $y_t - \mu_t$ may be potentially problematic since $\{(y_t, r_t)\}_t$ may not satisfy the difference equations of an ARMA model. However, the asymptotic results we shall state later (Theorem 1.4.2.1 and 1.4.2.2) may still hold as the following simple, yet useful case exemplifies. Suppose that $y_t \sim \beta\text{ARMA}(p, q)$, consider the logit link function, i.e., $g(x) = \log(x) - \log(1 - x)$, and consider the residual

$$\hat{r}_t = a_t(n)(y_t - \hat{\mu}_t) \tag{1.4}$$

for a sequence of random variables $\{a_t(n)\}_n$ such that, conditionally on $\mathcal{F}_{t-1}$, $a_t(n) \xrightarrow{p} a_t \in \mathbb{R}$ as n tends to infinity, for all t. By expanding $\log(1-x)$ into its power series around 0, we can write

$$g(y_t) - g(\mu_t) = y_t - \mu_t + \log\left(\frac{y_t}{\mu_t}\right) + \sum_{k=2}^{\infty} \frac{y_t^k - \mu_t^k}{k!}.$$

Notice that $0 < \mathrm{var}(y_t) \leq 1/[4(\phi+1)] < 1/4$, so that, if ϕ is large, the variance of y_t is very small and y_t will be close to μ_t with high probability. As a result, $\log(y_t/\mu_t) \approx 0$ and the summation, which involves powers of numbers in $(0,1)$, can also be expected to be negligible. We thus expect that $y_t - \mu_t \approx g(y_t) - g(\mu_t)$. If, in addition (as often happens in applications), the sequence a_t is nearly constant, we expect that the behavior of $a_t(y_t - \mu_t)$ should not be far from a constant times $g(y_t) - g(\hat{\mu}_t)$. As a consequence, portmanteau tests based on $g(y_t) - g(\hat{\mu}_t)$ and on (1.4) should behave similarly.

We shall now present some portmanteau test statistics that were proposed in the context of Gaussian time series and that can be used with time series that follow the beta law. Box and Pierce (1970) introduced a test statistic which, under the null hypothesis, is approximately χ^2 distributed in large samples. The associated test, however, was shown to have poor small sample performance. A variant of Box-Pierce test statistic was considered by Ljung and Box (1978):

$$Q_{LB} = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k}.$$

Monti (1994) proposed replacing the sample autocorrelations in the above test statistic by sample partial autocorrelations. The limiting null distribution of both test statistics is χ^2_{m-p-q}.

A non-parametric portmanteau test statistic based on rank autocorrelations was introduced by Dufour and Roy (1986). Let R_t be the rank of $\hat{r}_t$. The kth residual rank autocorrelation is $\tilde{\rho}_k = \sum_{t=1}^{n-k}(R_t - \bar{R})(R_{t+k} - \bar{R}) / \sum_{t=1}^{n}(R_t - \bar{R})^2$, $1 \leq k \leq n-1$, where $\bar{R} = n^{-1}\sum_{t=1}^{n} R_t = (n+1)/2$ and $\sum_{t=1}^{n}(R_t - \bar{R})^2 = n(n^2-1)/12$ if all ranks are distinct. Since $R_1, \ldots, R_n$ are interchangeable, when $\hat{r}_1, \ldots, \hat{r}_n$ are interchangeable and continuous, the mean of $\tilde{\rho}_k$ is $\mu_k = -(n-k)/[n(n-1)]$ for $1 \leq k \leq n-1$. The authors also provide an expression for $\tilde{\sigma}_k^2$, the variance of $\tilde{\rho}_k$. The portmanteau test statistic is $Q_{DR} = (\tilde{\boldsymbol{\rho}} - \boldsymbol{\mu})^\top D_2^{-1} (\tilde{\boldsymbol{\rho}} - \boldsymbol{\mu})$, where $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, \ldots, \tilde{\rho}_m)^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^\top$ and $D_2 = \mathrm{diag}\{\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_m^2\}$. Under $\mathcal{H}_0$, Q_{DR} is asymptotically distributed as χ^2_{m-p-q}.

Kwan and Sim (1996a) considered the situation in which a portmanteau test is applied to an observed time series, and not to residuals from a fitted model. They transformed the autocorrelations in order to reduce the dispersion bias of Q_{LB}. Fisher (1921) proposed transforming $\hat{\rho}_k$ as $z_{1k} = 0.5\log((1+\hat{\rho}_k)/(1-\hat{\rho}_k))$, and Hotelling (1953) introduced the following two transformations: $z_{2k} = z_{1k} - (3z_{1k} + \hat{\rho}_k)/[4(n-k)]$ and $z_{3k} = z_{2k} - 23z_{1k} + 33\hat{\rho}_k - 5\hat{\rho}_k^3/96(n-k)^2$, where z_{ik}, for $i = 1,2,3,4$, is normally distributed with $\mathbb{E}(z_{ik}) \approx 0$, $\mathrm{var}(z_{1k}) \approx (n-k-3)^{-1}$, $\mathrm{var}(z_{2k}) \approx (n-k-1)^{-1}$ and $\mathrm{var}(z_{3k}) \approx (n-k-1)^{-1}$. Using these approximations, Kwan and Sim (1996a) proposed three modified portmanteau test statistics: $Q_{KWi} = \sum_{k=1}^{m}(n-k-\tau_i)z_{ik}^2$, $i = 1,2,3$, where $\tau_1 = 3$ and $\tau_2 = \tau_3 = 1$. Under $\mathcal{H}_0$, they are asymptotically distributed as χ^2_m. The number of degrees of freedom is m because the test is, as noted above, applied to an observed time series.

Kwan and Sim (1996b) introduced a fourth test statistic based on a variance-stabilizing transformation proposed by Jenkins (1954): $z_{4k} = \sin^{-1}(\hat{\rho}_k)$. Here, $\mathbb{E}(z_{4k}) \approx 0$ and $\mathrm{var}(z_{4k}) \approx (n-k)^{-2}(n-k-1)$. The test statistic is $Q_{KW4} = \sum_{k=1}^{m}[(n-k)^2/(n-k-1)]z_{4k}^2$. Under $\mathcal{H}_0$, it is asymptotically distributed as χ^2_m.

When the sample size is large relative to m, the means of Q_{KW1}, Q_{KW2} and Q_{KW3} are approximately equal to $m - m(m+4)/n$ and that the mean of Q_{KW4} is approximately equal to $m - m(m+1)/n$. Kwan and Sim (1996a) noticed that these results suggest that, for fixed n, the means of the four test statistics are smaller than m. (Such results were obtained for Gaussian processes. In our simulations, we computed the means of the four test statistics and noticed that the approximations also hold for βARMA processes.) The authors then proposed to modify the tests critical values using

$$\mathbb{E}(Q_{KWi}) = \sum_{k=1}^{m}(n-k-\tau_i)\left\{\mathbb{E}(\hat{\rho}_k^2) + \frac{2}{3}\mathbb{E}(\hat{\rho}_k^4)\right\}, \quad i = 1,2,3,$$

$$\mathbb{E}(Q_{KW4}) = \sum_{k=1}^{m}\frac{(n-k)^2}{(n-k-1)}\left\{\mathbb{E}(\hat{\rho}_k^2) + \frac{1}{3}\mathbb{E}(\hat{\rho}_k^4)\right\}.$$

Expressions for $\mathbb{E}(\hat{\rho}_k^2)$ and $\mathbb{E}(\hat{\rho}_k^4)$ can be found in Davies, Triggs and Newbold (1977) and in Ljung and Box (1978). The tests can then be performed as follows when applied to an observed time series: Reject the null hypothesis at the $\gamma \times 100\%$ significance level ($0 < \gamma < 1$) if $Q_{KWi} \geq \chi^2_{1-\gamma,\mathbb{E}(Q_{KWi})}$, $i = 1,\ldots,4$. When the test statistics are computed from ARMA residuals, the null hypothesis is rejected if $Q_{KWi} \geq \chi^2_{1-\gamma,\mathbb{E}(Q_{KWi})-p-q}$, $i = 1,2,3,4$.

Peña and Rodriguez (2002) proposed a different portmanteau test statistic which is based on the determinant of the residual autocorrelation matrix: $\hat{D}_m = n\left(1 - \left|\hat{R}_m\right|^{1/m}\right)$,

where $\hat{R}_m$ is the $(m+1)\times(m+1)$ sample autocorrelation matrix. The determinant of $\hat{R}_m$, $|\hat{R}_m|$, is the estimated generalized variance of the standardized residuals. The authors proposed approximating the asymptotic null distribution of $\hat{D}_m$ by the gamma distribution. The approximation, however, can be quite poor. A modified test statistic was also proposed by the authors: $D_m = n\left(1-|\ddot{R}_m|^{1/m}\right)$, where $\ddot{R}_m$ is obtained by replacing $\hat{\rho}_k^2$ with $\ddot{\rho}_k^2 = (n+2)(n-k)^{-1}\hat{\rho}_k^2$. McLeod and Jimenez (1984) noted a shortcoming of D_m: $\ddot{R}_m$ is not always positive definite.

Lin and McLeod (2006) recommended performing the Peña-Rodriguez test using bootstrap resampling. The numerical evidence in their paper shows that the test performs considerably better when one does so.

1.4 TWO NEW PORTMANTEAU TESTS

In this section we present two new test portmanteau statistics proposed by Scher (2017) that are based on residual partial autocorrelations. We then prove that their limiting null distribution is χ^2_{m-p-q} in the class of βARMA models. When the statistics are computed using observed time series their limiting distribution under the null hypothesis of no serial correlation is χ^2_m. A novel aspect of our proofs is that they also hold for other test statistics that were proposed in the Gaussian time series literature which are asymptotically equivalent to our test statistics under the null hypothesis.

1.4.1 Two new test statistics

Two new portmanteau tests statistics were proposed by Scher (2017): Q_1 and Q_4. They are based on residual partial autocorrelations. The motivation for using partial autocorrelations in portmanteau test statistics stems from the fact that when the order of moving average dynamics is underestimated, the sum of squared partial autocorrelations is likely to be larger that of squared autocorrelations; see, e.g., Monti (1994, pp. 778–779).

As before, $\hat{\pi}_k$ denote the kth residual partial autocorrelation. Scher (2017) explored the use of the aforementioned variance-stabilizing transformations but now applied to partial autocorrelations, and then used the transformed partial autocorrelations to construct portmanteau test statistics. Consider the transformation introduced by Fisher

(1921):

$$\hat{z}_{1k} = \frac{1}{2}\log\left(\frac{1+\hat{\pi}_k}{1-\hat{\pi}_k}\right), \quad k = 1, \ldots, m.$$

The corresponding modified test statistic can be written as

$$Q_1 = \sum_{k=1}^{m}(n-k-3)\hat{z}_{1k}^2. \tag{1.5}$$

The second test statistic proposed by Scher (2017) makes use of the transformation introduced by Jenkins (1954), namely:

$$\hat{z}_{4k} = \sin^{-1}(\hat{\pi}_k).$$

Using it, Scher (2017) arrived at the following portmanteau test statistic:

$$Q_4 = \sum_{k=1}^{m}\frac{(n-k)^2}{n-k-1}\hat{z}_{4k}^2. \tag{1.6}$$

We shall prove that under appropriate conditions and under the null hypothesis, when the test is applied to an observed time series, Q_1 and Q_4 are asymptotically distributed as χ^2_m; when the test statistics are computed using βARMA residuals, they are, under the null hypothesis, asymptotically distributed as χ^2_{m-p-q}. Such results are proved in Subsection 1.4.2. We suggest that the tests critical values be corrected using the approach outlined in the previous section.

The finite sample performances of the two tests presented above can be improved with the aid of bootstrap resampling. A similar approach can be applied to other portmanteau tests. In what follows we shall use the bootstrap method to improve the finite sample performances of the following tests: Q_{LB}, Q_M, Q_{DR}, Q_{KW1}, Q_{KW2}, Q_{KW3}, Q_{KW4}, Q_1 and Q_4. Bootstrap inference is performed as described in Lin and McLeod (2006).

1.4.2 Asymptotic null distribution of the new portmanteau test statistics

We shall now derive the asymptotic distribution of the statistics Q_1 and Q_4 given in (1.5) and (1.6), respectively, under the null hypothesis (1.3) of no correlation up to a predetermined lag $m > 0$. We separate the discussion into two cases: the test is applied to an observed time series and the test is applied to the residuals from a fitted βARMA model. As we have mentioned heuristically before, the asymptotic null distributions are

different under these two cases. More specifically, when the time series under evaluation comes from a fitted model, it is obtained from a model with parameters replaced by estimates which impacts the number of degrees of freedom of the test statistic asymptotic null distribution. For further details, see Box and Pierce (1970), Ljung and Box (1978), Ljung (1986) and the references therein.

1.4.2.1 Observed time series

We start by analyzing the case where the time series under scrutiny is an observed time series, i.e., it is not a set of residuals from a time series model fit. Under the null hypothesis the series is comprised of uncorrelated and identically distributed random variables with finite second moment. As before, for $k > 0$, let $\hat{\rho}_k$ and $\hat{\pi}_k$ denote the kth order sample autocorrelation and the kth sample partial autocorrelation, respectively. Note that both $\hat{\rho}_k$ and $\hat{\pi}_k$ depend on n.

Theorem 1.4.2.1. *Under the null hypothesis, for $i = 1, 4$, the statistics Q_i and Q_{KWi} are asymptotically equivalent and*

$$Q_i \xrightarrow{d} \chi^2_m,$$

as n tends to infinity.

The proof of the above theorem and the proofs of the results that follow are presented in the Appendix 4.

1.4.2.2 βARMA residuals

In order to obtain the asymptotic null distribution of the two test statistics proposed by Scher (2017) we need to rely on the asymptotic theory for the conditional maximum likelihood estimator (CMLE) in the framework of βARMA models. The asymptotic theory is mentioned in Rocha and Cribari-Neto (2009) and more details are given in Rocha and Cribari-Neto (2017). The idea behind the asymptotic theory of the CMLE in the context of βARMA models is to write η_t as a linear combination of the regressors $\boldsymbol{x}_t$ and past values of the process itself (in autoregressive fashion), allowing for possible ancillary terms, in the spirit of Benjamin, Rigby and Stasinopoulos (2003) (Equation (3)) and Fokianos and Kedem (2004) (Equation (5) in particular).

In what follows we summarize the conditions required for the consistency and asymptotic normality of the conditional maximum likelihood estimator (CMLE) in βARMA models. We start with conditions regarding the model's systematic component.

A1. The roots of the autoregressive polynomial $\Phi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$ lie outside the unit circle.

A2. In the parametric space, there exists an open neighborhood U around the true parameters $\boldsymbol{\theta}$ where the roots of the moving average polynomial $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ lie outside the unit circle.

A3. The polynomials $\Theta(\cdot)$ and $\Phi(\cdot)$ have no common roots.

Assumptions A1 to A3 are standard in the context of usual ARMA models and they provide sufficient smoothness for the CMLE's asymptotic theory to hold. Now let $\boldsymbol{\kappa} = (\phi, \alpha, \beta_1, \ldots, \beta_c, \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q)^\top$ and $\boldsymbol{d}_t = (\boldsymbol{x}_t^\top, \ldots, \boldsymbol{x}_1^\top, \ldots, y_t, \ldots, y_1)^\top$, where $\boldsymbol{x}_t^\top$ is an c-dimensional set of non-random covariates. In addition to Conditions A1–A3, we make the following assumptions:

P1. The inverse link function g^{-1} is of class $\mathcal{C}^2$ and satisfies $\left|\partial g^{-1}(x)/\partial x\right| \neq 0$, for all $x \in \mathbb{R}$.

P2. The parametric space Ω is an open set in $\mathbb{R}^{c+q+p+2}$ and the true parameter $\boldsymbol{\kappa}_0$ lies in Ω.

P3. For each t, $\boldsymbol{d}_t$ almost surely belongs to a compact set Υ and there exists $n_0 \in \mathbb{N}$ such that, for all $n > n_0$, $\sum_{t=1}^n \boldsymbol{d}_t \boldsymbol{d}_t^\top$ is positive definite with probability 1. Additionally, $g^{-1}(\eta_t)$ is almost surely well-defined for all $\boldsymbol{d}_t \in \Upsilon$ and $\boldsymbol{\kappa} \in \Omega$.

P4. There exists a probability measure λ such that $\int \boldsymbol{z}\boldsymbol{z}^\top \lambda(d\boldsymbol{z})$ is positive definite and such that the weak convergence

$$\frac{1}{n}\sum_{t=1}^{n} I(\boldsymbol{d}_{t-1} \in A) \underset{n\to\infty}{\longrightarrow} \lambda(A)$$

holds for all λ-continuity sets A under (1.2) with $\boldsymbol{\kappa} = \boldsymbol{\kappa}_0$.

Under A1–A3 and P1–P4, the CMLE is consistent and asymptotically normal with variance-covariance matrix equal to the inverse of the information matrix. For further details, as well as a discussion of the conditions above, we refer readers to Fahrmeir and Kaufmann (1985), Shao (1992), Fokianos and Kedem (2004), Pumi *et al.* (2019) and references therein.

To set the notation, let $\hat{r}_1, \ldots, \hat{r}_n$ be the residuals obtained from a correctly specified βARMA(p, q) model, computed as $\hat{r}_t = g(y_t) - g(\hat{\mu}_t)$, where $\hat{\mu}_t$ is obtained by evaluating (1.2) at the maximum-likelihood estimates.

Theorem 1.4.2.2. *Let Q_i, $i = 1, 4$, be computed using residuals obtained from a correctly specified $\beta ARMA(p, q)$ model that satisfies Assumptions A1–A3 and P1–P4. Under the null hypothesis,*

$$Q_i \xrightarrow{d} \chi^2_{m-p-q},$$

as n tends to infinity.

Using the proof of the above result it is possible to establish the limiting null distribution of Q_{LB} and Q_M when such test statistics are computed using βARMA residuals, as indicated in the next result.

Corolário 1.4.2.1. *Let Q_{LB} and Q_M be computed using residuals from a correctly specified $\beta ARMA(p, q)$ model that satisfies Assumptions A1–A3 and P1–P4. Under the null hypothesis,*

$$Q_{LB} \xrightarrow{d} \chi^2_{m-p-q} \quad \textit{and} \quad Q_M \xrightarrow{d} \chi^2_{m-p-q},$$

as n tends to infinity.

We close this section by noting that one can force b coefficients to equal zero when fitting a βARMA model ($b < p + q$), obtain the residuals and perform portmanteau testing inferences. In that case, the asymptotic null distribution of Q_1 and Q_4 is $\chi^2_{m-p-q+b}$. The proof is essentially the same as the one presented above, in view of McLeod (1978).

1.5 NUMERICAL EVIDENCE

Several simulation experiments were carried out to investigate the finite sample performances of the different portmanteau tests in the class of βARMA models. All simulations were performed using the R statistical computing environment Team (2021). We use the standardized ordinary residual defined by Ferrari and Cribari-Neto (2004), i.e., $\hat{r}_t = (y_t - \hat{\mu}_t)/\sqrt{\widehat{\text{var}}(y_t)}$, where $\widehat{\text{var}}(y_t) = \hat{\mu}_t(1 - \hat{\mu}_t)/(1 + \hat{\phi})$. Here, $\hat{\mu}_t$ is obtained by evaluating μ_t at the maximum likelihood estimates and $\hat{\phi}$ is the maximum likelihood estimate of ϕ. In the simulations that follow, we use $\phi = 120$ so that $\text{var}(y_t) < 1/484 \approx 0.00207$. Under the conditions stated in Subsection 1.4.2, if we take $a_t(n) = \widehat{\text{var}}(y_t)^{-\frac{1}{2}} \xrightarrow{p} \text{var}(y_t)^{-\frac{1}{2}} = a_t$, for all t, as n tends to infinity; recall (1.4). It follows that Theorem 1.4.2.2 holds.

All tests are performed at 10%, 5% and 1% significance levels. For brevity, however, we shall only present results for $\gamma = 5\%$. The sample sizes are $n \in \{50, 250, 500\}$, the values of m we used are $m \in \{5, 10, 15, 20, 25\}$, the number of Monte Carlo replications is $5,000$. Separate simulations were performed for each value of m. Log-likelihood maximization was performed using the BFGS quasi-Newton method with analytic first derivatives. Starting values for the parameters were selected as follows: (i) all moving average parameters were set equal to zero, and (ii) the values for the autoregressive parameters were selected by regressing $g(y_t)$ on $g(y_{t-1}), \ldots, g(y_{t-p})$ using ordinary least squares. Beta random number generation was performed based on the Mersenne Twister uniform random generator Matsumoto and Nishimura (1998). All simulations were carried out using the logit link function.

Our Monte Carlo simulations entailed considerable computational cost. They were run at the National Supercomputing Center at Federal University of Rio Grande do Sul (CESUP/UFRGS). The hardware used was a cluster of computers with 64 blades of processing, 15.97 Tflops and 174TB RAM running the SUSE Enterprise Server Linux operating system. Our code made use of parallel computing and our simulations ran on 3 nodes with 24 clusters. To achieve reproducibility, we used the *doRNG* R package in conjunction with `foreach` loops. By using parallel computing, we were able to reduce execution time by approximately 78%.

At the outset, we ran size simulations using asymptotic critical values. We considered the following data generating mechanisms: βAR(1) (with $\varphi \in \{0.2, 0.5, 0.8\}$), βMA(1) (with $\theta \in \{0.2, 0.5, 0.8\}$) and βARMA(1,1) (with $\varphi = 0.2$ and $\theta \in \{0.2, 0.5, 0.8\}$). The tests displayed considerable size distortions in some cases when the sample size was small ($n = 50$) but performed well with larger sample sizes ($n = 250$ or $n = 500$). We shall not present these results for brevity. Instead, we shall focus on size simulations of tests that employ bootstrap resampling. We note that small sample size distortions also take place when the tests are used with Gaussian ARMA models. Lin and McLeod (2006), e.g., recommend using bootstrap resampling when performing the portmanteau test proposed by Peña and Rodriguez (2002) with Gaussian models when $n < 1,000$.

We shall now investigate the effectiveness of bootstrap resampling when coupled with portmanteau tests in the class of βARMA models. We shall consider the following bootstrap tests: Q_{LB}, Q_M, Q_{DR}, Q_{KW1}, Q_{KW2}, Q_{KW3}, Q_{KW4}, Q_1 and Q_4. Since Q_{KW1},

Q_{KW2} and Q_{KW3} behave similarly, we shall only report results for Q_{KW1}. All results are based on $1,000$ bootstrap samples and $n = 50$. We shall not report results for $n \in \{250, 500\}$ because in large samples the bootstrap tests behave similarly to the corresponding standard tests.

Table 1 contains the null rejection rates of the bootstrap portmanteau tests, including the bootstrap variant of the Peña-Rodriguez's test Lin and McLeod (2006), for the βAR(1) model with $\varphi \in \{0.2, 0.5, 0.8\}$. Table 2 presents the bootstrap tests null rejection rates obtained using the βMA(1) model with $\theta \in \{0.2, 0.5, 0.8\}$ for data generation. In both cases, all tests are now nearly size distortion free.

A second set of Monte Carlo simulations was carried out to evaluate the tests nonnull behavior, i.e., to evaluate the tests powers. Data generation is now carried out under the alternative hypothesis and the interest lies in examining the tests ability to detect that the model specification is in error. The true data generating process is βARMA(1,1), and the fitted model is βAR(1). The sample sizes are $n \in \{50, 250\}$ and the values of m range from 3 to 25. Since Q_{KW1}, Q_{KW2}, Q_{KW3} and Q_{KW4} behave similarly, we shall only report results on Q_{KW4}. The Q_1 and Q_4 tests proposed by Scher (2017) also behave similarly, and for that reason we shall only consider Q_4. Since some of the tests are liberal, we shall base all tests on exact (estimated from the size simulations) critical values rather than on asymptotic critical values. By doing so, we force all tests to have correct size.

The plots in the top row of Figure 1 display the empirical powers of Q_{LB}, Q_M, Q_{DR}, Q_{KW4} and Q_4 for $\varphi = 0.5$ and $\theta = 0.5$. Notice that when $n = 50$ (left panel) Q_4, the test introduced by Scher (2017) outperforms the remaining tests for all values of m. When the sample size is large ($n = 250$, right panel), the Q_4 test considerably outperforms all other tests for all values of m. The Q_{DR} test is the worst performer in both cases ($n = 50$ and $n = 250$). The plots in the bottom row of Figure 1 present results obtained using $\varphi = 0.2$ and $\theta = 0.8$ in the βARMA data generating process. The conclusions are similar to those from the previous case with Q_4 outperforming the competition regardless of the samples size. Visual inspection of the graphics in the bottom row of Figure 1 reveals that the choice of m considerably impacts the powers of the tests. In particular, the tests ability to detect model misspecification weakens as m grows. The same happens under Gaussian models Kwan, Sim and Wu (2005).

How does the value of θ impact the tests powers? In order to answer that

Table 1 – Null rejection rates of the bootstrap portmanteau tests at the 5% nominal level obtained using the βAR(1) model with $\varphi \in \{0.2, 0.5, 0.8\}$ and $n = 50$.

Test	$\varphi = 0.2$					$\varphi = 0.5$					$\varphi = 0.8$				
	m					m					m				
	5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
$\hat{D}_m$	4.8	4.9	5.3	5.4	5.0	4.1	5.0	5.1	5.0	4.6	4.7	4.8	4.6	4.6	4.8
Q_{LB}	5.0	5.0	4.8	4.8	4.9	4.7	5.0	5.0	5.5	5.2	5.1	4.8	4.9	4.7	4.8
Q_M	5.0	4.9	5.2	5.2	5.1	4.6	5.2	5.0	5.0	5.5	5.1	5.2	4.7	5.1	4.8
Q_{DR}	5.3	4.8	5.0	4.6	4.8	5.1	5.1	5.1	4.8	5.0	5.0	4.9	4.7	5.0	5.0
Q_{KW1}	5.0	5.0	4.7	4.5	4.9	4.7	5.0	5.1	5.3	5.3	5.0	4.9	4.7	5.0	5.0
Q_{KW4}	5.0	5.0	4.7	4.6	4.8	4.7	5.0	5.1	5.3	5.3	5.0	4.9	4.7	5.0	5.0
Q_1	5.1	4.9	5.2	4.9	5.2	4.6	4.9	5.2	5.0	5.3	4.9	5.0	4.8	5.1	5.2
Q_4	5.1	4.9	5.2	5.0	5.3	4.6	4.9	5.2	5.0	5.2	4.9	5.0	4.9	5.2	5.2

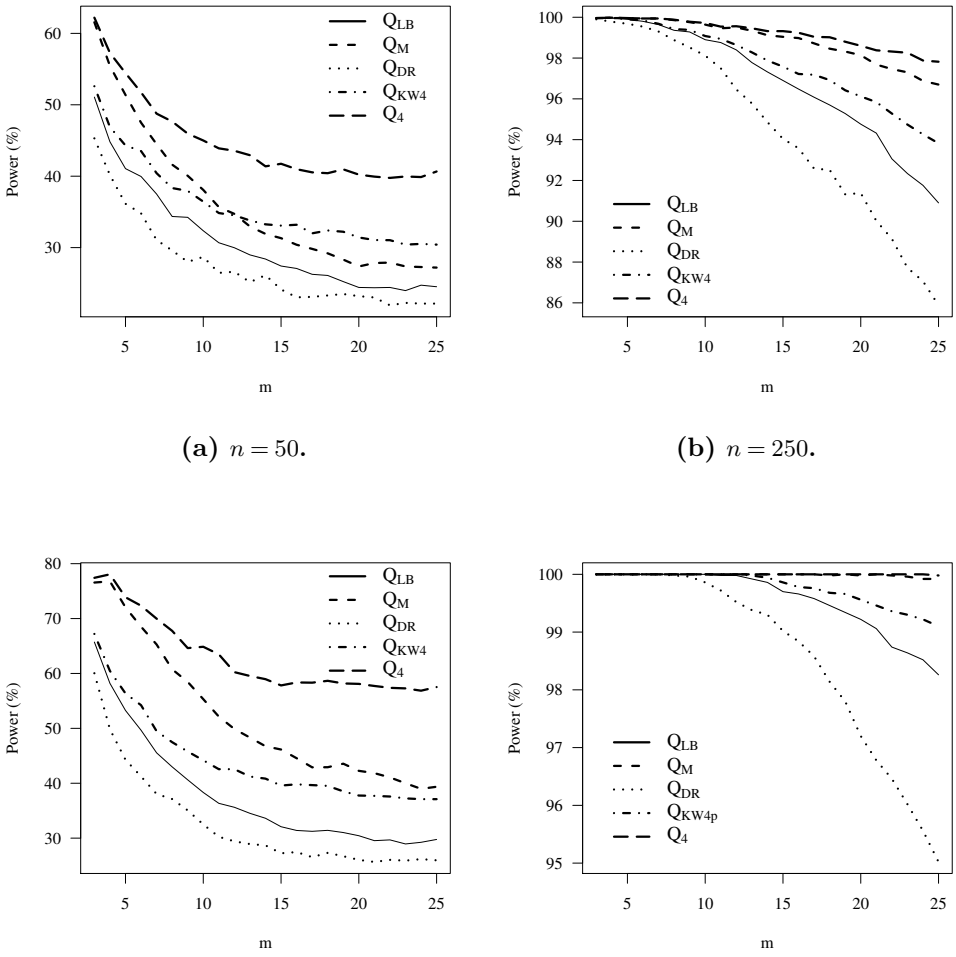
Source: The author (2020).

Table 2 – Null rejection rates of the bootstrap portmanteau tests at the 5% nominal level obtained using the βMA(1) model with $\theta \in \{0.2, 0.5, 0.8\}$ and $n = 50$.

Test	$\theta = 0.2$					$\theta = 0.5$					$\theta = 0.8$				
	m					m					m				
	5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
$\hat{D}_m$	4.8	5.2	5.4	5.3	5.2	4.5	4.9	4.7	5.0	4.9	4.8	5.2	5.1	5.0	5.0
Q_{LB}	5.1	5.0	5.0	5.0	4.8	5.4	4.8	5.1	5.0	5.2	5.3	5.1	5.3	5.6	5.6
Q_M	5.1	4.8	5.0	5.3	5.2	5.4	4.9	5.3	5.3	5.2	5.3	4.9	4.9	5.4	5.3
Q_{DR}	5.1	5.1	4.8	5.0	4.6	5.0	4.8	4.8	4.6	4.6	4.9	5.3	5.0	5.0	5.0
Q_{KW1}	5.3	5.0	5.2	5.2	5.0	5.4	4.9	4.9	5.1	5.0	5.4	5.1	5.2	5.4	5.6
Q_{KW4}	5.3	5.0	5.2	5.2	5.0	5.4	4.9	5.0	5.1	5.1	5.4	5.1	5.1	5.4	5.6
Q_1	5.1	4.8	5.1	5.3	5.0	5.4	4.8	4.9	5.2	5.2	5.2	4.7	4.9	5.2	5.2
Q_4	5.1	4.8	5.1	5.3	5.0	5.4	4.8	4.9	5.3	5.2	5.2	4.7	4.9	5.2	5.0

Source: The author (2020).

Figure 1 – Powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4}, Q_4 tests at the 5% nominal level when the fitted model is βAR(1) and the true data generating process is βARMA(1,1) with $\varphi = 0.5$, $\theta = 0.5$ (top row) and $\varphi = 0.2$, $\theta = 0.8$ (bottom row).



(a) $n = 50$. **(b)** $n = 250$.

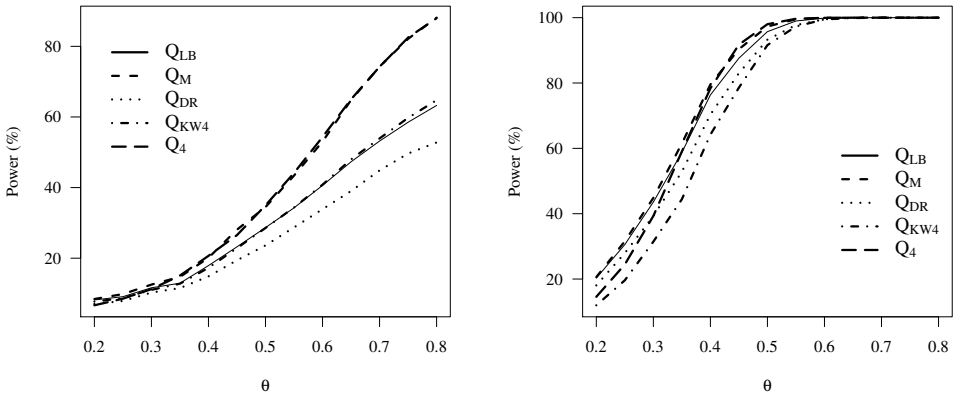
(c) $n = 50$. **(d)** $n = 250$.

Source: The author (2020).

question we ran simulations using different values of the moving average parameter when generating the data. The value of m is fixed at 5, the value of the autoregressive parameter (φ) is fixed at 0.2, and two sample sizes are used: $n \in \{50, 250\}$. The tests estimated powers are displayed in Figure 2 (left panel for $n = 50$ and right panel for $n = 250$). The Q_4 and Q_M tests are the clear winners when the sample size is small, especially when the value of θ is large; Q_{DR} is the test with smallest power. When the sample size is large, Q_4 and Q_M remain the most powerful tests, but not by much, and Q_{KW4} becomes the worst performing test.

We now move to the situation where the true data generating process is βARMA(1,1) but the fitted model is βMA(1). In the previous case the fitted model

Figure 2 – Powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4}, Q_4 tests at the 5% nominal level when the fitted model is βAR(1) and the true model is βARMA(1,1) with $\varphi = 0.2$ and different values of θ.



(a) $n = 50$. **(b)** $n = 250$.

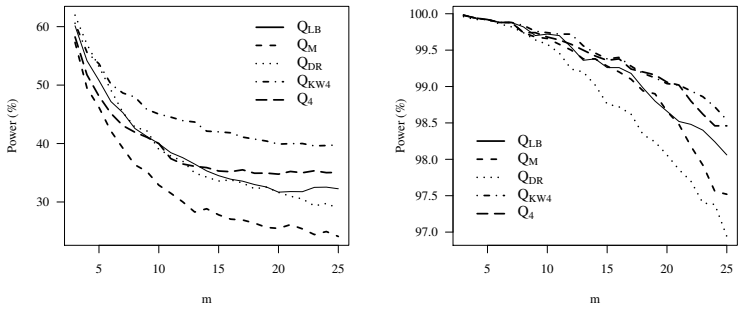
Source: The author (2020).

was incorrectly specified because it failed to take into account relevant moving average dynamics. In contrast, model misspecification now stems from failing to account for important autoregressive dynamics. For brevity, we shall only consider two scenarios. The tests powers are displayed in plots in the top row of Figure 3 which show the estimated powers of Q_{LB}, Q_M, Q_{DR}, Q_{KW4} and Q_4 for $\varphi = 0.5$ and $\theta = 0.5$. The tests ability to detect model misspecification increases with the sample size, as expected, and decreases with m. When the sample size is small (left panel), Q_{KW4} is the most powerful test, followed by Q_4 when m is large. When the sample size is large, Q_{KW4} and Q_4 are the winners, but not by much when m is small.

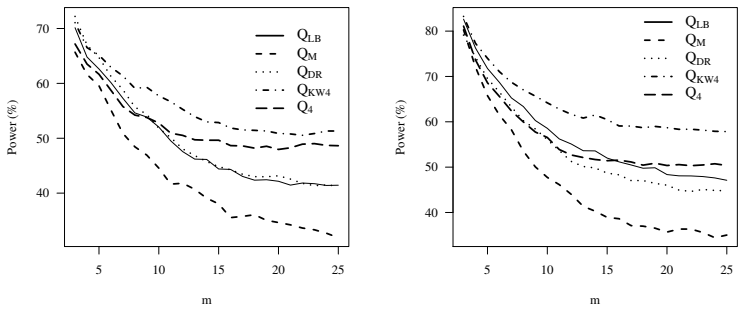
We shall now investigate whether the previous set of results are sensitive to the parameter values in the autoregressive and moving average dynamics. The plots in the middle row of Figure 3 present the tests empirical powers for $n = 50$: (i) $\varphi = 0.5$ and $\theta = 0.8$ (left panel) and (ii) $\varphi = 0.8$ and $\theta = 0.5$ (right panel). As in the previous case, Q_{KW4} is the most powerful test, Q_4 being the runner-up when $m > 10$ in the left panel and when $m > 16$ in the right panel. Overall, these results are similar to those presented in the previous figure.

Figure 3 (bottom row, left panel) presents the empirical powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4} and Q_4 tests for $\varphi = 0.8$ and $\theta = 0.2$. The results obtained using $\varphi = 0.8$ and $\theta = 0.8$ are displayed in Figure 3 (bottom row, right panel). In both cases, $n = 50$ and Q_{KW4} is the winner. In the right panel Q_4 is the runner-up and in the left panel Q_4 is

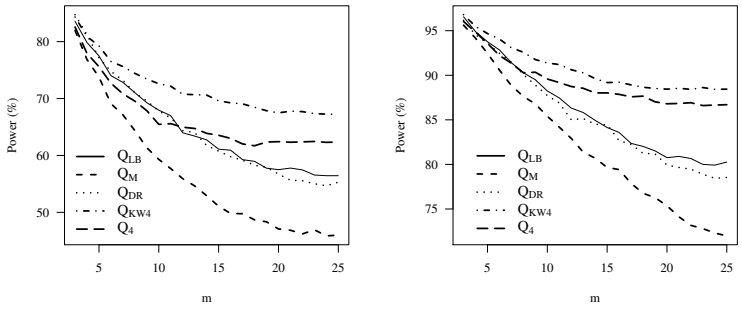
Figure 3 – Powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4}, Q_4 tests at the 5% nominal level when the fitted model is βMA(1) and the true data generating process is βARMA(1,1) with $\varphi = 0.5$, $\theta = 0.5$ (top row), $\varphi = 0.5$, $\theta = 0.8$ (left panel, middle row), $\varphi = 0.8$, $\theta = 0.5$ (right panel, middle row), $\varphi = 0.8$, $\theta = 0.2$ (left panel, bottom row) and $\varphi = 0.8$, $\theta = 0.8$ (right panel, bottom row).



(a) $n = 50$. **(b)** $n = 250$.



(c) $n = 50$. **(d)** $n = 50$.



(e) $n = 50$. **(f)** $n = 50$.

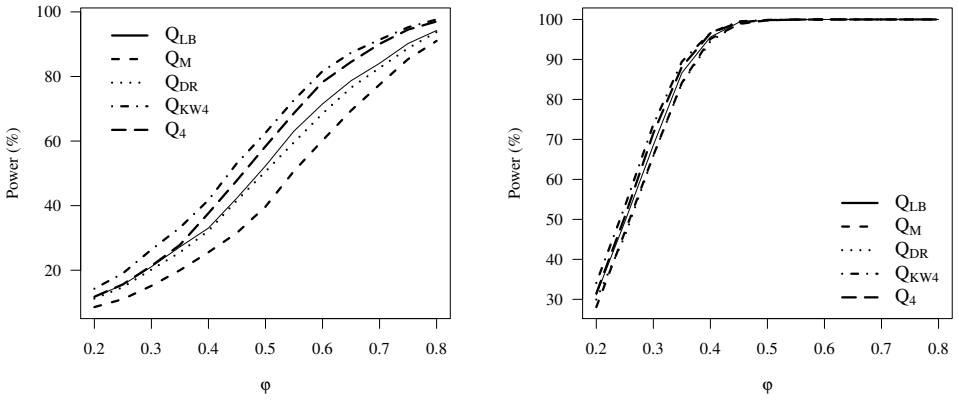
Source: The author (2020).

the second best performing test when m is large. Interestingly, in both cases Q_M is the worst performer. Recall that the Q_M and Q_4 test statistics use residual partial partial autocorrelations. Even though we do not present results for $n = 250$, we note that when $\varphi = 0.8$ all tests displayed power of 100%.

Previously, we evaluated the tests nonnull behavior for different values of

the moving average parameter. We shall now evaluate how the tests powers change with the value of the autoregressive parameter. Here, the data generating process is βARMA$(1,1)$, the βARMA$(0,1)$ model is estimated, $\theta = 0.8$ and $m = 25$. The tests powers for $n \in \{50, 250\}$ and different values of φ are presented in Figure 4. When $n = 50$ (left panel), Q_{KW4} and Q_4 are the most powerful tests for all values of φ, especially when φ is large, and Q_M is the worst performer. When $n = 250$ (right panel), Q_{KW4} and Q_4 are the winners for all values of φ, but not by much.

Figure 4 – Powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4}, Q_4 tests at the 5% nominal level when the fitted model is βMA(1) and the true model is βARMA$(1,1)$ with $\theta = 0.8$ and different values of φ.



(a) $n = 50$. **(b)** $n = 250$.

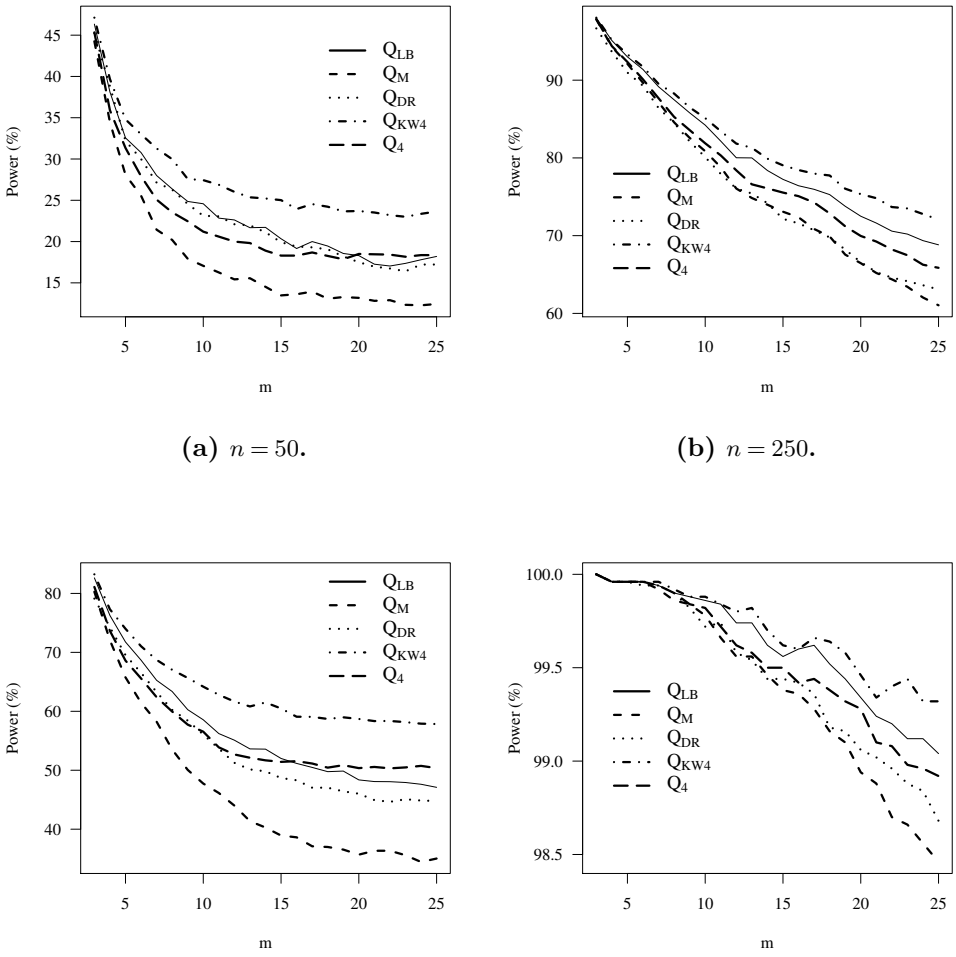
Source: The author (2020).

We shall now consider a different model specification error, namely: the true data generating process is βARMA$(2,1)$ but the fitted model is βARMA$(1,1)$, i.e., some existing autoregressive dynamics is not accounted for. The plots in the top row of Figure 5 display the empirical powers of Q_{LB}, Q_M, Q_{DR}, Q_{KW4} and Q_4 for $\boldsymbol{\varphi} = (0.2, 0.5)^\top$ and $\theta = 0.2$. All the tests become considerably more powerful when n goes from 50 to 250. Again, the tests powers decrease with m. In both cases, the Q_{KW4} is the clear winner.

The plots in the bottom row of Figure 5 display the empirical powers of Q_{LB}, Q_M, Q_{DR}, Q_{KW4} and Q_4 for $\boldsymbol{\varphi} = (0.2, 0.8)^\top$ and $\theta = 0.2$. The Q_{KW4} test again outperforms the competition. It is noteworthy that Q_M (which, like Q_4, is based on residual partial autocorrelations) is the worst performer. We also note that Q_4 is the second best performer when the sample size is small and m is large.

Finally, we shall consider a data generation process of higher order. The data

Figure 5 – Powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4}, Q_4 tests when the fitted model is βARMA$(1,1)$ and the true model is βARMA$(2,1)$ with $\boldsymbol{\varphi} = (0.2, 0.5)^\top$, $\theta = 0.2$ (top row) and $\boldsymbol{\varphi} = (0.2, 0.8)^\top$, $\theta = 0.2$ (bottom row).



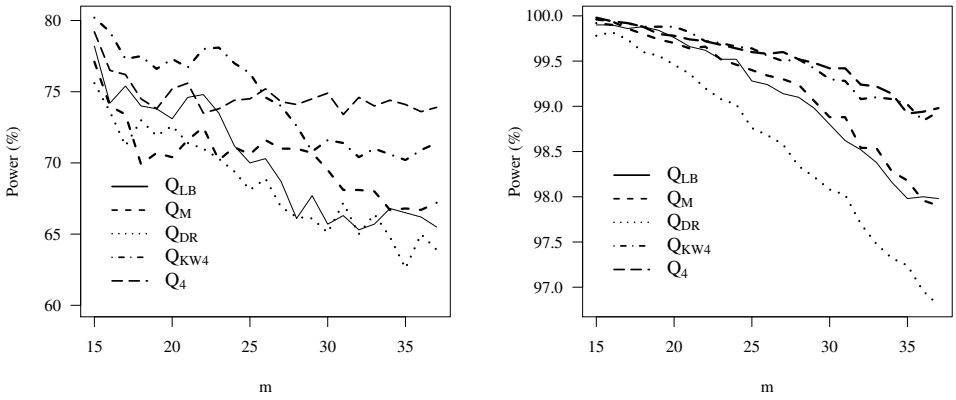
(a) $n = 50$. **(b)** $n = 250$.

(c) $n = 50$. **(d)** $n = 250$.

Source: The author (2020).

are generated using the βARMA$(4,4)$ dynamics but the βARMA$(2,2)$ model is estimated. Here, $\boldsymbol{\varphi} = (0.15, 0.2, 0.25, 0.3)^\top$ and $\boldsymbol{\theta} = (0.07, 0.13, 0.21, 0.33)^\top$. The tests powers for $n = 50$ were lower than in the previous simulations, the best (worst) performing test being Q_4 (Q_{DR}) whose nonnull rejection rates are approximately equal to 25% (below 20%). In Figure 6 we present the tests estimated powers for $n \in \{100, 250\}$. When $n = 100$, Q_{KW4} is the most powerful test for values of m up to 25 and Q_4 is the best performer for larger values of m. When $n = 250$, Q_{KW4} and Q_4 are clearly the most powerful tests for all values m.

Figure 6 – Powers of the Q_{LB}, Q_M, Q_{DR}, Q_{KW4}, Q_4 tests at the 5% nominal level when the fitted model is βMA(1) and the true model is βARMA(1,1) with $\theta = 0.8$ and different values of φ.



(a) $n = 50$. **(b)** $n = 250$.

Source: The author (2020).

1.6 EMPIRICAL HYDROLOGICAL APPLICATION

We shall now turn to the empirical application briefly described in Section 1.1. The variable of interest is the proportion of stored hydroelectric energy ONS (2020) in South Brazil. The data are monthly averages from January 2001 to October 2016, thus covering 190 months ($n = 190$). The following six observations (November 2016 through April 2017) were used for evaluating forecasting accuracy. The computer code used in this case study is available at <https://github.com/vscher/barma>.

Table 3 contains some descriptive statistics. Notice the negative skewness and also the negative excess kurtosis, and recall that the beta density easily accommodates such features.

Table 3 – Descriptive statistics of the monthly average rates of stored energy in the South of Brazil.

min	max	median	mean	variance	asymmetry	excess kurtosis
0.2977	0.9862	0.7323	0.7069	0.0403	−0.3270	−1.1644

Source: The author (2020).

Model selection was performed using the Akaike Information Criterion (AIC) Akaike (1974). We considered all models with autoregressive and moving average dynamics up to the fourth order and logit link. The selected model was the βARMA(1,1) model, whose AIC equals to −307.9635. The maximum likelihood estimates of α, φ_1, θ_1 and ϕ are (standard errors in parentheses 0.3452 (0.0787), 0.5235 (0.0412), 0.3588 (0.0502) and

11.7593 (1.1910), respectively. All parameters are significantly different from zero at the 1% significance level using the z-test.

We shall now consider portmanteau testing inference, i.e., we shall use the different portmanteau tests to assess whether the fitted model adequately represents the time series data. We computed the tests p-values for $m \in \{3, \ldots, 30\}$. The only p-values smaller than 0.05 are those of the Q_{DR} test for small values of m (for values of m up to 5). Except for that, the p-values of all tests for all values of m are greater than 0.05. We take take as evidence that the βARMA$(1,1)$ model is correctly specified.

Next, we shall perform portmanteau testing inference using a different model. Since the βARMA$(1,1)$ model seems to adequately represent the data dynamics, the tests are expected to detect model misspecification when applied to a different model. At the outset, we fitted the βAR(1) model. Table 4 displays the tests p-values for $m \in \{5, 10, 15, 20, 25, 30\}$. Q_4 is the only test with all p-values below 0.05, that is, it is the only test that yields rejection of the null hypothesis of correct model specification regardless of the value of m. When the βMA(1) is fitted, all tests yield rejection of $\mathcal{H}_0$ at the 5% significance level for all values of m.

Table 4 – Q_M, Q_{DR}, Q_{KW4} and Q_4 p-values computed from the fitted βAR(1) model.

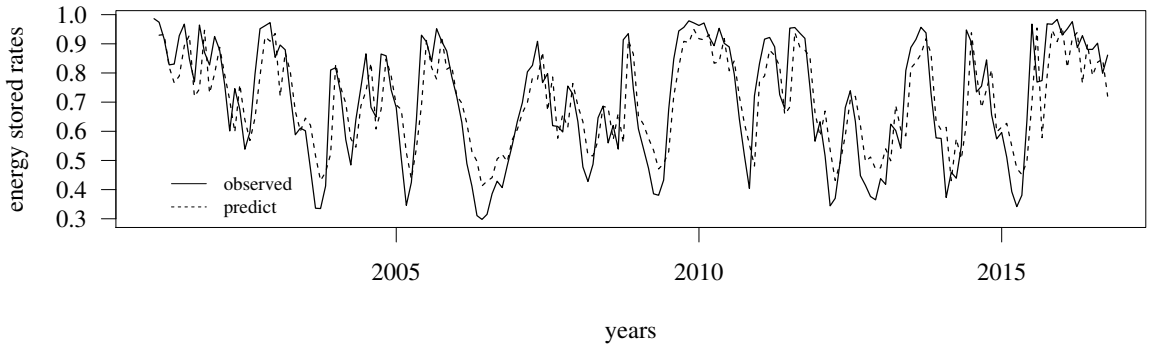
	$m=5$	$m=10$	$m=15$	$m=20$	$m=25$	$m=30$
Q_{LB}	0.006	0.006	0.026	0.058	0.128	0.182
Q_M	<0.001	0.002	0.008	0.042	0.054	0.065
Q_{DR}	<0.001	0.002	0.011	0.022	0.056	0.078
Q_{KW1}	0.005	0.005	0.021	0.043	0.087	0.120
Q_{KW4}	0.005	0.006	0.021	0.044	0.090	0.125
Q_4	<0.001	0.002	0.006	0.029	0.038	0.045

Source: The author (2020).

The final step in our empirical analysis involves forecasting. Indeed, stored energy forecasting is quite important for all institutions responsible for energy distribution. We produced forecasts using three different time series models, namely: the βARMA$(1,1)$, Gaussian ARMA$(1,1)$, and Gaussian AR(2) models. The latter two models were selected based on the AIC using the `auto.arima` function of the *forecast* package Hyndman and Khandakar (2008) of the R statistical computing environment Team (2021). We also used the KARMA$(1,1)$ model Bayer, Bayer and Pumi (2017); again, model selection was performed using the AIC. Finally, we considered the Holt exponential smoothing algorithm, as implemented in the `holt` function of the R *forecast* package. The observed time series

and the predicted values from the fitted βARMA$(1,1)$ model are presented in Figure 7. It is noteworthy that that the βARMA$(1,1)$ model is able to satisfactorily capture the data dynamics.

Figure 7 – Energy stored rates (solid lines) and predict values (dashed lines) computed from the fitted βARMA$(1,1)$ model.



Source: The author (2020).

We now move from in-sample to out-of-sample forecasting. We consider a horizon of 6 months, i.e., we wish to forecast the time series next six values. Forecasting accuracy is measured using the mean absolute error (MAE), i.e., the mean value of the absolute differences between observed and predicted values. The results are presented in Table 5 for $h \in \{1,\ldots,6\}$, h denoting the forecasting horizon. We note that the βARMA$(1,1)$ model yields forecasts that are more accurate than those obtained from the competing models and smoothing algorithm for all values of h. For instance, when forecasting the next three observations ($h = 3$), the βARMA$(1,1)$ MAE equals 0.1364 which is considerably smaller than the MAEs of the three competing models (0.1820, 0.1680 and 0.1783) of the exponential smoothing algorithm (0.2302).

Table 5 – Mean absolute forecasting errors from the βARMA$(1,1)$, ARMA$(1,1)$, AR(2) and KARMA$(1,1)$ models and from the Holt exponential smoothing algorithm.

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
βARMA$(1,1)$	0.1244	0.1444	0.1364	0.1484	0.1694	0.1839
ARMA$(1,1)$	0.1518	0.1828	0.1820	0.1982	0.2211	0.2364
AR(2)	0.1345	0.1690	0.1680	0.1830	0.2050	0.2198
KARMA$(1,1)$	0.1474	0.1788	0.1783	0.1950	0.2189	0.2352
Holt	0.1554	0.2110	0.2303	0.2629	0.2996	0.3264

Source: The author (2020).

1.7 CONCLUDING REMARKS

The βARMA model is particularly useful for modeling and forecasting time series data that assume values in the standard unit interval. It is thus useful for modeling several hydrological time series. The model naturally accommodates distributional asymmetry and nonconstant variance. It will always yield fitted values and out-of-sample forecasts that are positive and smaller than one. Additionally, no data transformation is needed prior to the analysis. The fitted model must be validated before it is used for forecasting. This is where our interest lies. Can the standard portmanteau tests be used with βARMA models? If so, how do they behave in finite samples? What is the impact of the choice of the truncation lag (m) on the tests null and nonnull behaviors? We reviewed several portmanteau tests that are available in the literature and two tests that were recently developed. The two new tests statistics proposed by Scher (2017) use residual partial autocorrelations instead of residual autocorrelations. We derived the asymptotic null distribution of the two new test statistics; more specifically, we proved that, under the null hypothesis, they are asymptotically distributed as χ^2_{m-p-q}. Our proof implies that some other well known test statistics (Q_{BP}, Q_{LM} and Q_M) are also asymptotically χ^2_{m-p-q} distributed under the null hypothesis when computed from βARMA residuals.

We presented Monte Carlo simulation results on the finite sample behaviors of the different portmanteau tests in the class of βARMA models. The tests size distortions were small when bootstrap resampling was used, especially when m is not very small.

The most interesting evidence from our numerical evaluations, however, relates to the tests powers. First, the choice of m impacts such powers: they typically decrease with m. Such an impact is well documented in the Gaussian literature; see, e.g., Kwan, Sim and Wu (2005). Second, a portmanteau test that proved to be robust under Gaussian data did not perform well when used with βARMA models. Third, overall, the tests introduced by Scher (2017) figure among the most powerful ones (the two new tests displayed similar nonnull behaviors, hence we only presented results for one of them). In particular, they were the best performers under pure autoregressive dynamics. It is noteworthy that whenever the two new tests were not the most powerful ones they were the next best performers as long as the value of m is not small. Fourth, overall, the tests proposed by Scher (2017) proved to be more powerful than an existing portmanteau test that also uses residual partial autocorrelations. Of course, the evidence we report is

restricted to the models considered in our numerical evaluation.

We also presented and discussed an empirical hydrological application. Our focus was on modeling and forecasting the proportion of stored hydroelectric energy in the Southern Region of Brazil. Such an empirical application showed the usefulness of portmanteau testing inference for model validation and also the usefulness of the class of βARMA time series models. It is noteworthy that the βARMA model used in the application yielded out-of-sample forecasts that were more accurate than those obtained using three alternative time series models and also an exponential smoothing algorithm.

In future work, we shall consider the use of the two test statistics proposed by Scher (2017) in other classes of dynamic models.

2 BETA AUTOREGRESSIVE MOVING AVERAGE MODEL SELECTION WITH APPLICATION TO MODELING AND FORECASTING STORED HYDROELECTRIC ENERGY

2.1 INTRODUCTION

The beta regression model was introduced by Ferrari and Cribari-Neto (2004) and has been extensively used with responses that assume values in the standard unit interval, $(0,1)$, such as rates, proportions and concentration indices. It is assumed that the responses are independently distributed, and hence the model is not useful for time series modeling. An extension of the model for serially dependent random variables was introduced by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017). It incorporates autoregressive and moving average dynamics, allows for the inclusion of fixed covariates and became known as the beta autoregressive moving average (βARMA) model. A novel feature of the model is that it accounts for the double bounded nature of the data and will never yield fitted values or out-of-sample forecasts that lie outside the standard unit interval. Additionally, the model also accounts for the inherent non-constant variance pattern of random variables in the standard unit interval, namely: the variance is smaller when the variable mean is close to zero or one, and is larger otherwise. The model can be used to produce out-of-sample forecasts of time series that assume values in $(0,1)$.

Even though the dynamic structure of the βARMA model is similar to that of the Gaussian ARMA model, there are some important differences between the two classes of models. For instance, unlike the latter, the errors in the former are not innovations that drive the stochastic process. Instead, the errors are, as we shall see, defined in a residual fashion as in the class of generalized autoregressive moving average (GARMA) models proposed by Benjamin, Rigby and Stasinopoulos (2003). Also, as noted earlier, the βARMA model is inherently heteroskedastic.

βARMA data modeling follows the standard Box-Jenkins approach, which consists of (i) model identification, (ii) parameter estimation, and (iii) diagnostic analysis; for details on the so-called Box-Jenkins approach, see Box *et al.* (2015). Parameter estimation is carried out by conditional maximum likelihood based on the underlying assumption that the variable of interest at each point in time follows the beta law; see

Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017). Diagnostic analysis based on portmanteau testing inferences for fitted βARMA models was developed by Scher *et al.* (2020). It remains to establish whether traditional model selection schemes work well when applied to βARMA data modeling and which model selection strategy is to be preferred, especially when the sample size is not large. The information criteria that are commonly employed for selecting models to be used for producing out-of-sample forecasts were not developed for dynamic models tailored to double bounded time series and then should be investigated in that context. As noted above, the βARMA dynamics is not driven by sequential realizations of white noise innovations as in Gaussian ARMA models. Given the different dynamic natures of the two processes, it is not clear that model selection strategies that perform best in traditional ARMA modeling will do so in βARMA modeling. It is thus important to assess the relative merits of different model selection strategies in the latter.

Practitioners may be tempted to resort to the most well known model selection practices when selecting a model for forecasting double bounded time series. This is the case, for example, of Melchior *et al.* (2021). The authors used the βARMA model to forecast mortality rates due to occupational accidents in three Brazilian states after performing model selection based on the largely well known Akaike information criterion. A relevant question is: Can more accurate forecasts be obtained in most applications that deal with double bounded data by performing model selection based on alternative criteria? Our empirical results indicate that more accurate model selection may translate into more accurate double bounded out-of-sample forecasts.

We performed simulations to evaluate the finite sample performance of different information criteria. The numerical evidence we report shows that βARMA model selection becomes considerably more accurate when it is based on bootstrap resampling. In several cases, the frequency of correct model identification based on a bootstrap information criterion greatly exceeds that of the second best performing criterion. We also evaluate the effectiveness of model selection based on measures of forecasting accuracy. The results favor a particular strategy based on directional forecasts, but they also indicate that more reliable model selection is achieved by using information criteria.

We present and discuss an empirical analysis in which the interest lies in forecasting in forecasting the future levels of stored hydroelectric energy in the South

of Brazil. Climate change has been adding uncertainty to hydropower generation and changing rainfall patterns and prolonged droughts have been making it increasingly difficult to assess future river flows. As a result, the use of stored hydroelectric energy has become increasingly important for hydropower generation. Interestingly, the most accurate forecasts were produced by βARMA models selected using bootstrap resampling. It is worth noticing that the same models were selected on the basis of the best performing strategies that employ measures of forecasting accuracy. The forecasts obtained from such models outperformed those computed from fitted models selected by alternative information criteria, in some cases by wide margins (e.g., over 1/3). There is, thus, agreement between our empirical and numerical results.

The chapter is organized as follows. Section 2.2 briefly presents the βARMA model. In Section 2.3 we review some model selection criteria that can be used to determine the orders of the autoregressive and moving average βARMA dynamics. In Section 2.4 we report the results of extensive Monte Carlo simulations that were performed to evaluate the accuracy of different βARMA model selection strategies in small to moderate sample sizes. An empirical analysis is presented and discussed in Section 2.5. Finally, some concluding remarks are offered in Section 2.6.

2.2 A DYNAMIC BETA MODEL

The βARMA model introduced by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017) is a dynamic model based on the beta regression model proposed by Ferrari and Cribari-Neto (2004). It is tailored for modeling random variables that assume values in $(0,1)$ and evolve over time. It can be used, for example, to predict the future behavior of rates, proportions and concentration indices.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ be an n-vector of time series random variables such that, given the previous information set $\mathcal{F}_{t-1}$ (the smallest σ-algebra such that the variables $y_1, \ldots, y_{t-1}$ are measurable), y_t follows the beta law indexed by its conditional mean μ_t and a precision parameter, ϕ, for $t = 1, \ldots, n$. The conditional density of y_t given $\mathcal{F}_{t-1}$ is

$$f(y_t|\mathcal{F}_{t-1}) = \frac{\Gamma(\phi)}{\Gamma(\mu_t\phi)\Gamma((1-\mu_t)\phi)} y_t^{\mu_t\phi-1}(1-y_t)^{(1-\mu_t)\phi-1}, \quad 0 < y_t < 1, \qquad (2.1)$$

$0 < \mu_t < 1$, $\phi > 0$, where $\Gamma(\cdot)$ is the gamma function. Here, $\mathbb{E}(y_t|\mathcal{F}_{t-1}) = \mu_t$ and $\mathrm{var}(y_t|\mathcal{F}_{t-1}) = \mu_t(1-\mu_t)/(1+\phi)$ are the conditional mean and conditional variance of y_t,

respectively. For given μ_t, the latter decreases as ϕ increases. Notice that the conditional variance of y_t is not constant; instead, it varies with μ_t. In particular, the conditional variance approaches zero as the conditional mean approaches zero or one.

The βARMA(p,q) model introduced by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017) assumes that y_t follows the above law with conditional mean such that

$$g(\mu_t) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i \left[g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} \right] + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{2.2}$$

where $\boldsymbol{x}_t^\top \in \mathbb{R}^c$ is a set of non-random covariates at time t, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_c)^\top \in \mathbb{R}^c$ is a vector of parameters, and $g : (0,1) \mapsto \mathbb{R}$ is a strictly monotonic and twice differentiable link function. Also, $\alpha \in \mathbb{R}$ is a scalar parameter and p, $q \in \mathbb{N}$ are the autoregressive and moving average orders associated with the $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ vectors of coefficients, respectively. Finally, r_t is an error term.

It is noteworthy the βARMA model structure in (2.2) is similar to that of the class of GARMA models; see Benjamin, Rigby and Stasinopoulos (2003). In both classes of models and unlike what happens in the Gaussian ARMA model, the error r_t is not an innovation that drives the stochastic process. Instead, r_t is defined in a residual fashion as $r_t = g(y_t) - g(\mu_t)$. Observe that, for fixed t, (2.2) only includes values of y_τ, $\boldsymbol{x}_\tau$ and r_τ for $\tau < t$, hence both μ_t and r_t are $\mathcal{F}_{t-1}$-measurable.

βARMA parameter estimation can be performed by conditional maximum likelihood. The conditional log-likelihood function, given the first $a = \max\{p,q\}$ observations, is $\ell \equiv \ell(\boldsymbol{\nu}_k, \phi|\boldsymbol{y}) = \sum_{t=a+1}^{n} \log f(y_t|\mathcal{F}_{t-1})$, where $\boldsymbol{\nu}_k = (\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$ is the vector of mean parameters and $f(y_t|\mathcal{F}_{t-1})$ is the beta density given in (2.1). For simplicity, we shall write $\ell(\boldsymbol{\nu}_k, \phi|\boldsymbol{y})$ as $\ell(\boldsymbol{\nu}_k|\boldsymbol{y})$ since ϕ is a fixed precision scalar. The model's score function and information matrix can be found in Rocha and Cribari-Neto (2017). Bias-corrected conditional maximum likelihood estimation was considered in Palm and Bayer (2018).

An extended version of the βARMA model that accommodates seasonal dynamics was proposed by Bayer, Cintra and Cribari-Neto (2018) and a version of the model that includes fractional integration was introduced by Pumi *et al.* (2019). Bayesian dynamic beta modeling was developed by Casarin, Valle and Leisen (2012) and Silva, Migon and Correia (2011); the former considers Bayesian model selection for beta autoregressive processes. In what follows, we shall work with the standard, baseline βARMA model in

the realm of frequentist statistical inference.

2.3 MODEL SELECTION STRATEGIES

Model selection aims at selecting a statistical model from a set of candidate models on the basis of a given data set. The selected model is typically used for out-of-sample forecasting provided that it yields a good data fit. The most commonly used model selection strategy is that based on criteria that penalize increases in the model's dimension. Typically, a set of models are fitted to the data, a given criterion is computed for each fitted model, and the model that displays the minimal criterion value is selected. Let $\boldsymbol{\nu}_k$ be the model's k-dimensional parameter vector, its conditional maximum likelihood estimator being denoted by $\hat{\boldsymbol{\nu}}_k$, and the maximized log-likelihood function being written as $\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k)$. In what follows, we shall present model selection strategies based on (i) information criteria that penalize model augmentation and (ii) forecasting accuracy measures.

The most commonly used criterion was introduced by Akaike (1974). It was obtained by minimizing the Kullback-Leibler distance between two densities and became known as the Akaike Information Criterion (AIC). The author showed that the model that minimizes minus two times the expected log-likelihood is the closest model to the true model according to the Kullback–Leibler information. He then used $-2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k)$ as an estimator of such a quantity, showed that its asymptotic bias is approximately equal to $-2k$ and arrived at the following information criterion:

$$\text{AIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + 2k.$$

We note that $2k$, the bias correcting term, can be viewed as a penalization term since it penalizes the model dimension augmentation when searching for the minimal criterion value. Based on data sets analyzed by Box and Jenkins, Ozaki (1978) showed that the use of Akaike's approach overcomes many difficulties of the identification procedure adopted in the authors' book. It was shown by Shibata (1976), however, that the AIC has a fixed overfitting probability asymptotically. As a consequence, the AIC tends to overestimate the model dimension. Several alternative criteria were then proposed aiming at achieving more accurate model selection.

A criterion that incorporates a small sample correction and is asymptotically

equivalent to the AIC was proposed by Sugiura (1978) and became known as $\mathrm{AIC_c}$; see also Hurvich and Tsai (1989). The corrected AIC is given by

$$\mathrm{AIC_c} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + 2k\left(\frac{n}{n-k-1}\right). \tag{2.3}$$

In essence, the new criterion includes an extra penalization term. According to Burnham and Anderson (2004), it should be preferred over the AIC unless $n/k > 40$ for the model with the largest value of k.

The Schwarz Information Criterion (SIC) was introduced by Schwarz (1978):

$$\mathrm{SIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + k\log(n). \tag{2.4}$$

A novel feature of this criterion is that it is consistent, i.e., the probability of selecting the true model tends to one as $n \to \infty$.

Model selection based on the SIC can be quite inaccurate in small samples. A modified version of the criterion was proposed by McQuarrie (1999). It incorporates a finite sample correction and can be expressed as

$$\mathrm{SIC_c} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + \frac{nk\log(n)}{n-k-1}.$$

The new criterion is asymptotically equivalent to the SIC and is expected to deliver superior model selection when the sample size is not large.

Hannan and Quinn (1979) focused on autoregressive model selection and proposed the Hannan-Quin Information Criterion (HQIC):

$$\mathrm{HQIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + 2k\log(\log(n)).$$

Like the AIC and the SIC, its small sample behavior may be poor. A modified version of the criterion was introduced by McQuarrie and Tsai (1998):

$$\mathrm{HQIC_c} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + \frac{2nk\log(\log(n))}{n-k-1}. \tag{2.5}$$

The Weighted-Average Information Criterion (WIC) was proposed by Wu and Sepulveda (1998). It is based on the criteria given in (2.3) and (2.4). Let $A_c = 2kn/(n-k-1)$ and $B = k\log n$. The WIC can be expressed as

$$\mathrm{WIC} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + \frac{A_c^2 + B^2}{A_c + B}.$$

The above expression can also be obtained by combining (2.3) and (2.5). The authors showed that the WIC behaves similarly to the $\mathrm{AIC_c}$ when the sample size is small and to the SIC in large samples. Like the SIC, the WIC is consistent.

Several authors considered the use of bootstrap resampling to estimate the penalty term used in the AIC criterion. We shall present two bootstrap-based criteria. They are of the form

$$\mathrm{EIC}i = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + \hat{B}_i,$$

$i = 1, 2$, where $\hat{B}_i$ is a bootstrap-based estimate of a penalty term that involves an expectation with respect to the distribution of the bootstrap sample. That is, the expected Kullback-Leibler discrepancy between the true and fitted models is estimated by means of data resampling. These criteria are said to be 'empirical' because their penalty terms are estimated from the data using the bootstrap method. They are usually referred to as 'Empirical Information Criteria'.

As noted by Shibata (1997, p. 379), a novel feature of bootstrap estimation of B_i is that it is free from any expansion whereas the AIC and related criteria are based on an expansion with respect to the model parameters. Hence, the bootstrap approach has wider applicability than the conventional bias correction. Additionally, we note that bootstrap bias correction is known to work well in other settings; see, e.g., Cribari-Neto, Frery and Silva (2002) and Ospina, Cribari-Neto and Vasconcellos (2006).

Let $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^\top$ denote the bootstrap sample and let N denote the number of bootstrap replications so that we have N bootstrap samples, each denoted as $\boldsymbol{y}^*(j)$, $j \in \{1, \ldots, N\}$. The bootstrap estimates of $\boldsymbol{\nu}_k$ are $\hat{\boldsymbol{\nu}}_k^*(j)$, $j \in \{1, \ldots, N\}$, where $\hat{\boldsymbol{\nu}}_k^*(j)$ is obtained by maximizing $\log f(\boldsymbol{y}^*(j)|\boldsymbol{\nu}_k)$. In what follows, $\mathbb{E}_*$ is used to denote expected value with respect to the distribution of $\boldsymbol{y}^*$.

Cavanaugh and Shumway (1997), in the context of Gaussian state space model selection, considered

$$B_1 = 2\left\{2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) - 2\,\mathbb{E}_*\left[\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k^*)\right]\right\},$$

which is estimated using bootstrap resampling. The corresponding model selection criterion is denoted as EIC1.

Shibata (1997) introduced the EIC2 bootstrap-based criteria. It uses

$$B_2 = 2\,\mathbb{E}_*\left\{2\log f(\boldsymbol{y}^*|\hat{\boldsymbol{\nu}}_k) - 2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k^*)\right\}.$$

As noted above, the bootstrap criteria use estimates of the penalty term obtained through data resampling. For instance, it can be shown that the EIC1 can be written as

$$\text{EIC1} = -2\log f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k) + 2\left\{\frac{1}{N}\sum_{j=1}^{N}\left[-2\log\frac{f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k^*(j))}{f(\boldsymbol{y}|\hat{\boldsymbol{\nu}}_k)}\right]\right\}.$$

EIC model selection for Gaussian autoregressive models was considered by Billah, Hyndman and Koehler (2005). The authors also considered EIC model selection for exponential smoothing. It is worth noticing that bootstrap-based model selection can be carried out in several different ways. For instance, Fenga (2017) uses bootstrap-resampling for Gaussian ARMA model selection by computing a model selection criterion (e.g., AIC) for each fitted model in each bootstrap replication, identifying the model that minimizes the criterion for that bootstrap time series, and then finally selecting the model on the basis of its relative frequency over the all bootstrap samples.

A second model selection approach involves the use of forecasting accuracy measures. The underlying idea is to remove a set of observations from the end of the series, forecast them using different models and select the model with the best forecasting performance. More specifically, the final s_f data points are removed, different models are fitted using the remaining $n - s_f$ observations, forecasts of the removed observations are produced, and a measure of forecasting accuracy is computed for each candidate model. The selected model is the one that displays the best forecasting performance. We shall present numerical evidence on such a strategy. Some forecasting accuracy measures that can be used for model selection are (i) mean absolute prediction errors (MAPE), (ii) root mean square error (RMSE), (iii) mean directional forecast (MDF), and (iv) rolling horizon weighted error (RHWE). The first two measures are well known and are routinely used for forecasting performance evaluation.

MDF-based model selection can be performed using a rolling window of n_r observations for parameter estimation and prediction. A sequence of $n - n_r - h$ out-of-sample h-step-ahead forecasts are produced and the corresponding forecasting errors are computed for each window terminating at observation $T \in \{n_r, \ldots, n-h\}$. The commonly used MDF measures are: (i) mean directional accuracy (MDA) and (ii) mean directional forecast value (MDV):

$$\text{MDA}_h = \frac{1}{n - n_r - h}\sum_{t=n_r}^{n-h} \mathbb{1}(Z_t = 1),$$

$$\text{MDV}_h = \frac{1}{n-n_r-h}\sum_{t=n_r}^{n-h}(-1)^{1-Z_t}|(y_{t+h}-y_t)/y_t|,$$

where $\mathbb{1}(\cdot)$ is the indicator function, $Z_t = \mathbb{1}(W_t = \widehat{W}_t)$ is the directional forecast, $W_t = \mathbb{1}(y_{t+h} - y_t > 0)$ is the realized direction and $\widehat{W}_t = \mathbb{1}(\hat{y}_t(h) - y_t > 0)$ is the predicted direction, $\hat{y}_t(h)$ denoting the forecast of y_{t+h} produced at time t. The MDF measure is computed for each candidate model and the selected model is that with largest MDA or MDV. For further details, see Blaskowitz and Herwartz (2009), Blaskowitz and Herwartz (2011) and Blaskowitz and Herwartz (2014).

The h-step ahead RHWE measure of forecasting accuracy was proposed in Poler and Mula (2011) for performing model selection: $\text{RHWE}_h = \sum_s \sum_t |e_t^s|^{\pi(\delta_1)}\zeta_h(\delta_1)\lambda(\delta_2)$, where $\delta_1 = t-s$ is the forecast forward, $\delta_2 = n-s$ is the forecast age, $e_t^s = y_t - \hat{y}_s(t-s)$ is the error of the forecast of y_t produced at time s, $\pi(\delta_1) \geq 1$ is the error power according to the forecast forward, $\zeta_h(\delta_1)$ is the multiplicative error factor according to the forecast forward ($\sum_{\delta_1}\zeta_h(\delta_1) = 1$) and $\lambda(\delta_2)$ is the multiplicative error factor according to the forecast age ($\sum_{\delta_2}\lambda(\delta_2) = 1$), $s \in \{n-s_f, \ldots, n-1\}$ and $t \in \{s+1, \ldots, \min\{s+h, n\}\}$.

Model selection strategies for non-dynamic beta regression models were investigated by Bayer and Cribari-Neto (2015), Bayer and Cribari-Neto (2017). In the next section we shall investigate model identification for dynamic beta models.

2.4 SIMULATION STUDY

The finite sample performances of the model selection strategies outlined in the previous section were already evaluated under different regression and time settings. It is not clear, however, how such criteria perform when used for βARMA model selection. In order to fill that gap, we shall report the results from extensive Monte Carlo simulations that were carried out to assess the accuracy of information criteria based model selection. We shall first focus on model selection based on information criteria and later consider model selection based on out-of-sample forecasting accuracy.

We consider pure autoregressive models, pure moving average models, and also models with both dynamics. The sample sizes are $n \in \{50, 150, 250\}$ and $\phi = 120$. We also report results obtained under smaller precision: $\phi = 12$. All simulations were carried out using the R statistical computing environment (versions 4.0.0 and 4.0.4); see Team (2021). The reported results were obtained using 5,000 Monte Carlo replications and

$N = 250$ bootstrap samples. This number of bootstrap replications is adequate since data resampling was used to estimate expected values, and not tail quantities as in confidence intervals and hypothesis tests; see Efron and Tibshirani (1986, Section 9). Bootstrap resampling was performed parametrically, i.e., we generated N bootstrap time series of size n from the fitted βARMA model after replacing the unknown parameters with their conditional maximum likelihood estimates.

Our Monte Carlo simulations are computationally challenging since they entail a very large number of log-likelihood numerical optimizations. The simulations were run at the National Center of Supercomputing of Universidade Federal do Rio Grande do Sul using a cluster of computers with 64 blades of processing, 15.97 TFLOPS, and 174-TB RAM that runs the SUSE Linux Enterprise Server operating system. We used parallel computing, and our simulations ran on three nodes with 24 clusters. By using parallel computing, we were able to reduce execution time by approximately 89%.

Log-likelihood maximization was carried out using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method with analytic first derivatives; see Nocedal and Wright (2006). Starting values for the parameters were selected as follows: (a) the starting values of all moving average parameters were zero, (b) the starting values of the autoregressive parameters were selected by regressing $g(y_t)$ on $g(y_{t-1}), \ldots, g(y_{t-p})$ using ordinary least squares, and (c) the starting value of ϕ was selected as in the beta regression model; see Ferrari and Cribari-Neto (2004). Beta random number generation was performed based on the Mersenne Twister uniform random generator. The logit link function was used in all data generating processes.

In what follows, we shall report the percentages of correct model selections achieved by using the following information criteria: AIC, AIC_c, SIC, SIC_c, HQIC, HQIC_c, WIC and EIC1. We only report results for EIC1 because the results for EIC2 were very similar. For simplicity, we shall refer to EIC1 as EIC.

At the outset, we focus on autoregressive processes. In particular, we consider $\beta\text{AR}(p)$ models, $p \in \{1, \ldots, 6\}$. The parameter values are: $\beta\text{AR}(1)$, $\varphi = 0.3$; $\beta\text{AR}(2)$, $\boldsymbol{\varphi} = (0.2, 0.4)^\top$; $\beta\text{AR}(3)$, $\boldsymbol{\varphi} = (0.2, -0.3, 0.4)^\top$; $\beta\text{AR}(4)$, $\boldsymbol{\varphi} = (0.2, -0.5, 0.4, -0.4)^\top$; $\beta\text{AR}(5)$, $\boldsymbol{\varphi} = (0.35, -0.4, 0.5, -0.45, 0.6)^\top$; and $\beta\text{AR}(6)$, $\boldsymbol{\varphi} = (0.45, -0.52, 0.65, -0.35, 0.4, -0.5)^\top$. The percentages of correct model selection are reported in Table 6. For each data generating process and sample size, the best result is in boldface. When the true model was $\beta\text{AR}(1)$,

βAR(2) or βAR(3), we fitted autoregressive models up to order six. When data were generated from the βAR(4), βAR(5) and βAR(6) models, we fitted autoregressive models up to order 7, 8 and 9, respectively.

The results in Table 6 show that the performances of all criteria improve as the sample size increases. When the sample size is small ($n = 50$), the criteria that incorporate finite sample corrections (AIC_c, SIC_c and HQIC_c) typically do not outperform the corresponding unmodified criteria. All information criteria perform quite well, even when the sample size is very small, when the true data generating process is βAR(1); all correct model selection percentages lie between 94.54% and 98.81%. The best overall performer is the EIC. In some cases, it outperforms the competition by wide margins. For instance, when $n = 50$ and the true model is βAR(3), its rate of correct model selection is 87.12% whereas that of the runner-up is 41.48%. Overall, under autoregressive dynamics, all model selection criteria perform well when $n \geq 150$, their success rates exceeding 93%. When the sample size is small ($n = 50$), model selection only works very well when $p = 1$, i.e., when the true model is βAR(1). The global winner under autoregressive dynamics is the EIC, the bootstrap-based criterion, and the worst performers are SIC and SIC_c.

Table 6 – Percentages of correct model selection, autoregressive models.

		βAR(1)	βAR(2)	βAR(3)	βAR(4)	βAR(5)	βAR(6)
$n = 50$	AIC	94.54	48.88	41.48	55.28	67.10	37.28
	AICc	95.32	47.10	38.32	50.98	63.69	27.17
	SIC	97.92	37.88	26.34	36.06	51.75	13.21
	SICc	**98.81**	34.02	20.30	30.22	40.31	15.91
	HQIC	96.06	44.62	36.12	46.14	61.98	25.95
	HQICc	97.02	42.14	31.08	43.38	56.34	16.61
	WIC	97.14	41.42	30.76	43.96	55.54	18.20
	EIC	98.31	**68.01**	**87.12**	**63.16**	**76.14**	**65.72**
$n = 150$	AIC	94.98	93.46	93.68	94.76	95.49	94.35
	AICc	95.24	93.80	94.14	94.78	96.21	95.22
	SIC	99.10	94.94	93.74	96.42	98.90	94.22
	SICc	**99.34**	94.64	93.32	96.02	**99.12**	93.36
	HQIC	97.44	95.14	95.00	95.79	97.32	96.30
	HQICc	97.78	95.42	95.16	96.47	97.64	96.41
	WIC	98.42	95.24	94.58	96.81	97.80	95.65
	EIC	98.62	**96.35**	**97.11**	**97.73**	98.53	**97.81**
$n = 250$	AIC	95.28	94.44	94.80	94.25	95.62	93.77
	AICc	95.56	94.76	95.02	94.32	96.05	94.45
	SIC	99.28	98.92	99.12	99.21	99.41	98.72
	SICc	**99.34**	**99.08**	**99.22**	**99.44**	**99.55**	**99.12**
	HQIC	97.72	97.52	97.68	97.22	97.84	97.10
	HQICc	97.76	97.72	97.86	97.84	98.24	97.24
	WIC	98.78	98.54	98.66	98.71	98.18	98.11
	EIC	97.78	97.61	98.7	98.91	99.45	98.23

Source: The author (2021).

We shall now move to moving average processes. We consider models with $q \in \{1, \ldots, 6\}$. The true parameter values are: $\theta = 0.5$ for $q = 1$, $\boldsymbol{\theta} = (0.2, 0.4)^\top$ for $q = 2$, $\boldsymbol{\theta} = (0.3, 0.2, 0.6)^\top$ for $q = 3$, $\boldsymbol{\theta} = (0.2, 0.3, -0.4, 0.6)^\top$ for $q = 4$, $\boldsymbol{\theta} = (0.15, 0.2, 0.3, 0.45, 0.5)^\top$ for $q = 5$, and $\boldsymbol{\theta} = (0.13, 0.19, 0.25, 0.3, 0.35, 0.5)^\top$ for $q = 6$. When the true model was βMA(1), βMA(2) or βMA(3), we fitted moving average models up to order six. When data were generated from the βMA(4), βMA(5) and βMA(6) models, we fitted moving average models up to order 7, 8 and 9, respectively. The percentages of correct model identification are presented in Table 7. The different model selection strategies become more accurate when the sample size is increased. Their performances deteriorate as q increases when the sample size is small ($n = 50$). For instance, the AIC and SIC rates of correct model identification drop from 61.62% and 76.40% to 18.52% and 26.15% as the order of the moving average dynamics increases from one to six, respectively.

Table 7 – Percentages of correct model selection, moving average models.

		βMA(1)	βMA(2)	βMA(3)	βMA(4)	βMA(5)	βMA(6)
$n = 50$	AIC	61.62	42.19	39.43	39.11	36.84	18.52
	AICc	65.45	43.32	39.72	39.89	31.32	19.41
	SIC	76.40	47.62	34.92	32.47	32.91	26.15
	SICc	83.65	46.45	39.68	37.29	35.49	23.37
	HQIC	66.96	50.18	37.35	38.82	31.74	28.61
	HQICc	72.57	49.29	36.13	37.66	32.18	26.88
	WIC	71.90	49.12	36.65	36.54	31.93	27.21
	EIC	**97.20**	**65.64**	**62.56**	**61.70**	**60.17**	**61.25**
$n = 150$	AIC	92.39	91.47	87.65	81.61	87.04	87.62
	AICc	92.83	92.23	87.18	82.42	87.14	82.10
	SIC	97.41	93.37	95.57	91.66	89.51	81.75
	SICc	**97.93**	94.11	**96.52**	92.91	88.29	86.54
	HQIC	95.54	93.34	91.23	86.54	88.13	82.21
	HQICc	95.61	93.70	92.45	87.81	89.47	81.10
	WIC	96.20	93.72	94.22	88.19	80.80	87.80
	EIC	97.50	**96.16**	95.84	**92.97**	**91.33**	**88.80**
$n = 250$	AIC	92.32	93.28	90.78	83.78	88.31	89.27
	AICc	92.63	93.58	90.92	84.72	89.23	88.11
	SIC	97.35	97.49	97.52	95.31	89.99	90.49
	SICc	97.35	**98.62**	97.34	95.85	89.05	91.28
	HQIC	96.10	96.31	94.23	89.79	89.81	86.11
	HQICc	96.73	96.43	94.97	90.25	89.74	86.23
	WIC	97.28	97.29	96.21	93.13	88.33	89.81
	EIC	**97.50**	97.88	**97.90**	**95.95**	**90.12**	**93.54**

Source: The author (2021).

Overall, the best performer is the EIC, the criterion that uses bootstrap resampling. In some cases, it outperforms the competing criteria by wide margins. For example, when $n = 50$ and $q = 1$, its rate of success is nearly 97% whereas that of the runner-up (AIC) is slightly above 61%. The difference becomes even more dramatic when

$q = 6$: 61.25% (EIC) vs 18.52% (runner-up, AIC).

We now consider models that include both autoregressive and moving average dynamics. We use three data generating processes and two scenarios (parameter values) for each process. The scenarios are indicated by the subscripts a and b next to the model orders. The parameter values of the three models under the two scenarios are: (i) βARMA$(1,1)_{\mathrm{a}}$, $\varphi = 0.3$, $\theta = 0.4$; βARMA$(1,1)_{\mathrm{b}}$, $\varphi = 0.3$, $\theta = 0.5$; (ii) βARMA$(1,2)_{\mathrm{a}}$, $\varphi = 0.35$, $\boldsymbol{\theta} = (0.2, 0.5)^\top$; βARMA$(1,2)_{\mathrm{b}}$, $\varphi = 0.35$, $\boldsymbol{\theta} = (0.2, 0.6)^\top$; (iii) βARMA$(2,1)_{\mathrm{a}}$, $\boldsymbol{\varphi} = (0.1, 0.5)^\top$, $\theta = 0.4$; βARMA$(2,1)_{\mathrm{b}}$, $\boldsymbol{\varphi} = (0.1, 0.6)^\top$, $\theta = 0.4$. When searching for the best model, we considered all combinations of p and q with each ranging from 0 to 3, except for $(0,0)$. It should be noted that this is a more challenging situation since the search for the best fitting model includes pure AR, pure MA and also models with AR and MA components. It is thus expected that larger sample sizes are needed in order to achieve reliable model selection. The rates of correct model identification (expressed as percentages) for the different criteria are presented in Table 8.

Table 8 – Percentages of correct model selection, autoregressive moving average models; the subscript next to the model order (a or b) identifies the scenario.

		$(1,1)_a$	$(1,1)_b$	$(1,2)_a$	$(1,2)_b$	$(2,1)_a$	$(2,1)_b$
$n = 50$	AIC	28.31	21.67	23.39	34.88	26.34	29.18
	AICc	28.82	23.59	25.51	35.59	26.99	29.02
	SIC	36.19	30.42	26.18	35.82	22.15	31.51
	SICc	37.45	32.73	26.45	35.96	22.60	33.29
	HQIC	35.24	24.16	23.15	34.18	23.47	32.05
	HQICc	36.52	26.14	23.98	34.77	23.89	33.63
	WIC	32.29	26.20	23.78	34.19	23.90	34.07
	EIC	**47.12**	**51.39**	**43.12**	**51.92**	**42.95**	**59.85**
$n = 150$	AIC	61.46	60.16	64.92	70.81	52.45	64.06
	AICc	61.84	61.04	66.89	72.25	54.50	64.50
	SIC	70.42	68.52	71.12	79.67	53.22	65.78
	SICc	70.91	69.18	72.06	80.92	54.18	65.89
	HQIC	65.69	63.40	63.54	75.42	49.61	65.12
	HQICc	66.09	64.21	64.89	75.88	51.19	65.98
	WIC	67.58	64.50	70.54	79.05	53.58	64.52
	EIC	**78.17**	**78.54**	**76.41**	**82.17**	**80.09**	**80.16**
$n = 250$	AIC	71.92	67.66	78.23	82.20	64.18	81.10
	AICc	73.09	67.89	78.86	83.01	64.89	81.99
	SIC	82.64	83.63	84.12	89.45	65.66	84.42
	SICc	83.47	85.07	85.19	89.81	66.05	85.06
	HQIC	78.52	80.21	83.25	85.42	72.45	83.12
	HQICc	79.69	80.35	83.88	86.07	73.09	84.03
	WIC	82.55	81.98	84.03	86.58	71.20	85.91
	EIC	**91.24**	**90.18**	**92.37**	**91.55**	**89.45**	**91.68**

Source: The author (2021).

As with pure autoregressive or pure moving average models, all model selection

strategies become more accurate as the sample size increases. Again, the overall winner was the EIC. Indeed, it was the best performer in all simulations, i.e., for all combinations of model order and scenario. In some situations, the EIC outperformed the competition by very wide margins. For instance, when $n = 50$ and $(2,1)_b$ (i.e., βARMA$(2,1)$ process and Scenario b), its rate of success was 59.85% whereas that of the second best performer (WIC) was only 34.07%. Even when $n = 150$ ($n = 250$) the EIC was far superior in some cases: success rate of 80.09% vs 54.50% (89.45% vs 73.09%) of the runner-up, which was the AIC_c (HQIC_c) under $(2,1)_a$.

We computed the average percentage of correct model selection for each criterion and each sample size and also the global figure; the latter was computed using all three sample sizes. The results are presented in Table 9. It is clear from these figures that the EIC is the best performed in all cases. When $n = 50$ it outperforms all alternative model selection strategies by wide margins; its average rate of correct model identification exceeds 64% whereas that of the runner-up (AIC) is approximately 42%. The relative advantage of the EIC over the competing criteria decreases as the sample size increases. For instance, when $n = 250$, the rates of success are above 95% for the EIC and over 92% for the second best performer (SIC_c). The EIC overall frequency of correct model selection (nearly 83%) considerably exceeds those of all other criteria, the runner-up being the SIC_c (nearly 73%).

Table 9 – Average percentages of correct model selection.

	$n = 50$	$n = 150$	$n = 250$	Global
AIC	41.45	81.57	86.17	69.73
AICc	40.62	81.90	86.61	69.71
SIC	38.66	85.29	91.82	71.92
SICc	38.55	85.84	92.21	72.20
HQIC	40.93	83.15	90.02	71.36
HQICc	40.01	83.73	90.45	71.39
WIC	39.71	84.40	91.40	71.83
EIC	**64.62**	**90.23**	**95.07**	**83.29**

Source: The author (2021).

In Table 6 (7) we presented the percentages of correct model specification for AR (MA) models obtained by only searching over AR (MA) models. The number of candidate models considered in the best fitting model search (A) ranged from 6 to 9 depending on the order of the true model. We shall now consider the more challenging case in which the true data generating process is AR or MA and the model search is performed over AR, MA and ARMA models. We consider the following true models:

(i) βAR(2) with $\boldsymbol{\varphi} = (0.2, 0.4)^\top$ and (ii) βMA(2) with $\boldsymbol{\theta} = (0.2, 0.4)^\top$. In each case, $A = 21$ candidate models are fitted, namely: AR models with p ranging from 1 to 6, MA models with q ranging from 1 to 6, and ARMA models with p and q ranging from 1 to 3. The sample size is $n = 150$. The percentages of correct model specification are presented in Table 10. We also report the differences between the results obtained under pure AR or MA model search and wider model search (Δ). The following conclusions can be drawn from the figures in Table 10. First, the success rates of all model selection criteria are now smaller ($\Delta < 0$). This was expected since more candidate models are considered in the search for the best fitting model. Second, the EIC is the best performer in both cases, i.e., under AR and MA data generating mechanisms. Third, under both dynamics, the EIC is the criterion with the smallest success rate reductions. For instance, under AR (MA) dynamics, its percentage of correct model determination dropped 17.38% (15.81%) whereas the corresponding figures for the alternative criteria range from 34.14% to 49.29% (34.46% to 47.48%). The AIC and the AICc are the criteria most impacted by the increase in the number of candidate models. Fifth, the performance ranks of the different criteria are the same as a before.

Table 10 – Percentages of correct model selection under wider model search.

	βAR(2)		βMA(2)	
	$A = 21$	Δ	$A = 21$	Δ
AIC	44.17	−49.29	43.99	−47.48
AICc	45.42	−48.38	45.99	−46.24
SIC	54.45	−40.49	58.91	−34.46
SICc	53.22	−41.42	59.63	−34.48
HQIC	57.59	−37.55	54.05	−39.29
HQICc	61.23	−34.14	55.04	−38.66
WIC	60.35	−34.92	57.52	−36.20
EIC	**78.97**	−17.38	**80.35**	−15.81

Source: The author (2021).

The simulation results presented above were obtained using $\phi = 120$ which is the same precision value used in Scher *et al.* (2020). We shall now investigate the impact of a smaller precision on the rates of correct model specification. To that end, we consider $n = 150$ and three data generating processes: βAR(2), βMA(1) and βARMA$(1,1)_b$. The percentages of correct model identification are presented in Table 11. For ease of comparison, the table also contains the differences in the figures obtained under the two scenarios (Δ). The results reported in Table 11 lead to several interesting conclusions. First, all model selection strategies become less accurate under small precision ($\Delta < 0$ in

all cases). Second, βARMA model selection accuracies are more impacted than those of βAR and βMA processes. Third, the EIC is the best performer under all data generating processes with $\phi = 12$. Fourth, the EIC went from runner-up to best performer under moving average dynamics when the precision parameter values was reduced. Fifth, EIC displays the smallest losses in accuracy.

Table 11 – Percentages of correct model selection in a small precision scenario ($\phi = 12$).

	βAR(2)		βMA(1)		βARMA(1,1)$_{\text{b}}$	
	$\phi = 12$	Δ	$\phi = 12$	Δ	$\phi = 12$	Δ
AIC	81.92	−11.54	78.04	−14.35	20.42	−39.74
AICc	83.74	−10.06	79.24	−13.59	21.15	−39.89
SIC	93.62	−1.32	86.05	−11.36	24.88	−43.64
SICc	91.14	−3.50	86.78	−11.15	24.91	−44.27
HQIC	91.28	−3.86	79.62	−15.92	22.36	−41.04
HQICc	92.44	−2.98	85.04	−10.57	23.17	−41.04
WIC	93.06	−2.18	86.42	−9.78	23.51	−40.99
EIC	**95.32**	−1.03	**89.86**	−7.64	**53.52**	−25.02

Source: The author (2021).

The numerical evidence presented above shows that all model selection criteria yield more accurate model identification as the sample size increases. More importantly, it reveals that there is much to be gained by resorting to bootstrap resampling when searching for the best model in the wider class of βARMA(p, q) models or when attention is restricted to βAR(p) or βMA(q) processes. When the EIC, the bootstrap-based criterion, was not the best performer it was a very close runner-up. In contrast, in several situations, the EIC was not only the best performer but did so by wide margins.

We shall now move to model selection based on out-of-sample forecasting accuracy. We use $n = 150$ and $h \in \{3, 6\}$. Model selection is based on MAPE, RMSE, MDF and RHWE. Two out-of-sample MDF measures are also considered: MDA and MDV. We use a rolling window of 100 observations for parameter estimation and prediction. A sequence of $50 - h$ out-of-sample h-step-ahead forecasts are produced and the corresponding forecasting errors are computed for each window terminating at observation $T \in \{100, \ldots, 150 - h\}$. When computing the RHWE measure, we set $\pi(\delta_1) = 1$ and $s_f = 10$, the corresponding weights being $\zeta_3(\delta_1) = \{0.5, 0.33, 0.17\}$, $\zeta_6(\delta_1) = \{0.29, 0.24, 0.20, 0.14, 0.09, 0.04\}$ and $\lambda(\delta_2) = \{0.02, 0.04, 0.05, 0.07, 0.09, 0.11, 0.13, 0.15, 0.16, 0.18\}$.

In what follows, we consider two scenarios when selecting a model based on an out-of-sample forecasting accuracy measure, namely: (i) the model that displays the highest overall accuracy is selected ('scenario 1'), and (ii) the model that displays the

highest accuracy among the models that pass a diagnostic test is selected ('scenario 2'). The motivation for carrying out a diagnostic test prior to forecasting is to only consider models for which there is no evidence of model misspecification. We use the Q_4 test portmanteau diagnostic test proposed by Scher *et al.* (2020) which performs well when used with fitted βARMA models. The lag truncation parameter is $m = 13$ and all testing inferences are carried out at the 5% significance level.

Table 12 contains the percentages of correct model selection obtained using the aforementioned out-of-sample forecasting accuracy criteria. For each criterion, the top and bottom rows correspond, respectively, to scenarios 1 and 2. The data generating processes are βAR(2), βMA(2) and βARMA$(1,1)_b$. The true parameter values are as before, and $\phi = 120$. The results obtained with MDA and MDV were very similar, and for brevity we shall only present those relative to MDV. The results in Table 12 lead to interesting conclusions. First, model selection based on MAPE and RMSE are the least accurate strategies, with rates of correct model selection that range from 17.26% to 23.30%. Second, RHWE model selection is slightly more successful than that based on the previous criteria. Third, MDV model selection is considerably more accurate than all alternative forecasting-based strategies, the corresponding rates of correct model identification fluctuating between 48.01% and 58% (approximately).

Table 12 – Percentages of correct model selection based on out-of-sample forecasting model selection criteria; top and bottom rows are for scenarios 1 and 2, respectively.

Criterion	βAR(2)			βMA(2)			βARMA$(1,1)_b$		
	$h=1$	$h=3$	$h=6$	$h=1$	$h=3$	$h=6$	$h=1$	$h=3$	$h=6$
MAPE	17.26	17.54	17.46	19.54	22.06	21.88	20.22	20.80	21.50
	18.20	19.09	17.56	20.19	21.95	22.90	20.18	22.74	21.95
RMSE	17.26	18.02	16.88	19.54	22.50	23.30	20.22	21.74	22.91
	18.20	18.88	17.62	20.19	24.23	24.68	20.18	24.21	24.24
RHWE	25.32	25.20	24.12	28.80	28.16	27.44	29.35	30.04	30.84
	27.74	26.33	26.25	29.00	30.00	30.37	31.75	32.16	31.61
MDV	48.01	56.58	55.62	56.18	56.38	56.02	57.14	57.05	58.03
	48.10	57.16	56.40	56.65	56.61	56.64	57.89	57.68	59.01

Source: The author (2021).

In some cases, MDV model selection is over twice more accurate than that based on the runner up criterion. For instance, under βMA(2) dynamics and $h = 3$, the MDV rate of correct model selection is 56.38% whereas that of the second best performer (RHWE) is 28.16%. Fourth, the results obtained with $h \in \{1,3,6\}$ are similar, except for MDV in the βAR(2) model where the rate of correct model selection is clearly smaller for

$h = 1$. Fifth, model selection based on information criteria is considerably more accurate than that performed on the basis of measures of forecasting accuracy, especially under pure AR and pure MA dynamics. Interestingly, MDV model selection is slightly less accurate than that based on the AIC (the worst performing model selection criterion) under βARMA$(1,1)$ dynamics (58.06% with $h = 6$ vs 60.16%). In contrast, under βAR(2) and βMA(2) dynamics, MDV model selection ($h = 6$) is considerably less accurate than that based on the AIC: 55.62% vs 93.46% and 56.02% vs 91.47%, respectively. Sixth, the RHWE and MDV model selection always benefit from a prior screening based on the Q_4 diagnostic test, especially the former. The largest increase in the rate of successful model selection that follows from the diagnostic test screening is 2.93% (RWME, βMA(2), $h = 6$).

2.5 FORECASTING STORED HYDROELECTRIC ENERGY

We shall, in what follows, present and discuss an empirical analysis. Hydroelectricity is a renewable energy source, and it is widely used in Brazil. There are two types of reservoirs: accumulation and water line. The former are usually located at the headwaters of rivers, in places of high waterfalls, since their large sizes allow for the accumulation of substantial amounts of water which function as stocks to be used in drought periods. They also allow hydroelectric power plants to rapidly respond to fluctuations in the demand for electricity. It is also worth noticing that hydroelectric power (hydro) is environmentally friendly since the hydroelectric life cycle produces very small amounts of greenhouse gases and hydro plants do not release pollutants into the air. Climate change, nonetheless, has been adding uncertainty to hydropower generation. Changing rainfall patterns and prolonged droughts have been making it increasingly difficult to assess future river flows. As a result, the use of accumulation reservoirs has become increasingly important for hydropower generation. Stored energy is the energy value of the accumulated water, that is, how much energy (in Megawatt monthly) can be generated from the stored volume of water expressed as a proportion of the total capacity. Stored energy forecasting is very important for companies in charge of energy distribution. Our interest lies in modeling the proportion of stored hydroelectric energy ONS (2020) in South Brazil and producing out-of-sample forecasts. We shall also evaluate the impact of model selection on the accuracy of the out-of-sample forecasts without regressors.

The data are monthly averages from July 2000 to April 2018, thus spanning

214 months. The final six observations were removed from the data to be used for forecasts evaluation. Hence, the effective sample size is $n = 208$ with the time series going up to October 2018. A shorter range of this time series was recently modeled by Scher *et al.* (2020). They used the AIC and focused on diagnostic (portmanteau) testing. In what follows, we shall use a longer time series and consider subsets of the data, namely: the first $n = 75$, $n = 150$ and $n = 208$ (complete data) observations. Some descriptive statistics are presented in Table 13: minimal value, maximal value, median, mean, variance, coefficient of asymmetry and coefficient of excess kurtosis. There is negative asymmetry and negative excess kurtosis. The mean level of stored energy is 0.7016 and the maximal level is close to one (0.9862). The time series data plot (top panel), correlogram (bottom left panel) and the partial correlogram (bottom right panel) can be found in Figure 8. The sample autocorrelations do not decay slowly towards zero, and hence there is no indication of long memory behavior. Also, the sample partial autocorrelations show no evidence of seasonal fluctuations. We note that the data contain several observations that are close to the upper standard unit interval limit.

Figure 9 contains the histogram (left panel) and boxplot (right panel) of the data. The former shows the prevalence of values close to one. Both graphs show that there is asymmetry in the data.

Table 13 – Descriptive statistics, stored hydroelectric energy in South Brazil, $n = 214$.

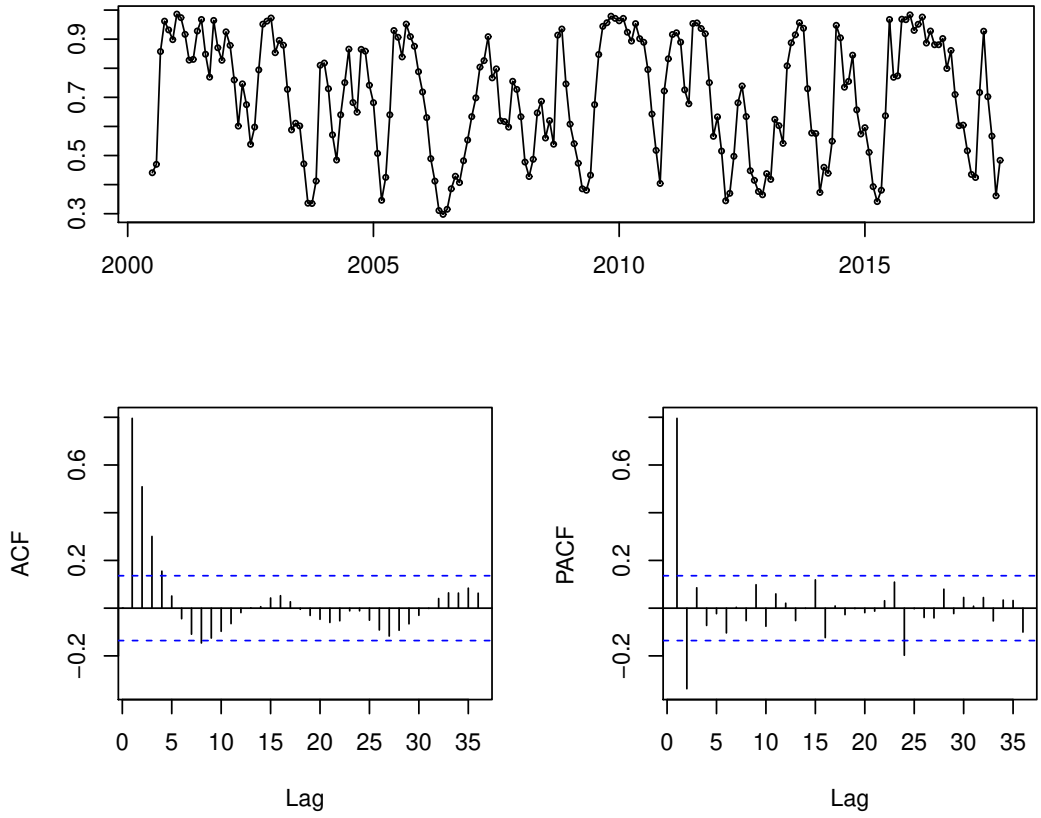
min	max	median	mean	variance	asymmetry	kurtosis
0.2977	0.9862	0.7265	0.7016	0.0404	−0.2714	−1.2180

Source: The author (2021).

As noted earlier, a novel feature of βARMA models is that they will never yield out-of-sample forecasts that lie outside $(0, 1)$. Such improper forecasts may be obtained, however, when using Gaussian ARMA models or an exponential smoothing algorithm. To illustrate that, we computed the first six out-of-sample forecasts from Gaussian ARMA models identified using the AIC and from the Holt algorithm for all subsamples of our time series with at least 24 observations (i.e., $n \geq 24$). In 12 of such subsamples, there was at least one forecast that exceeded one.

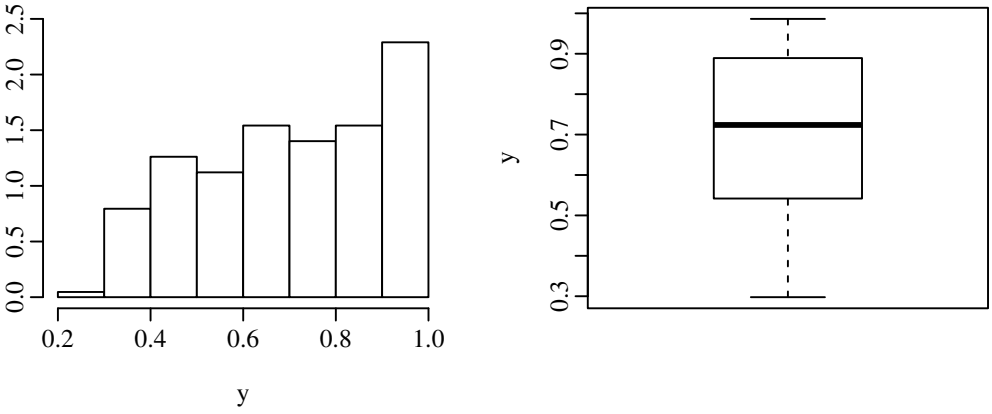
We shall now search for the best fitting βARMA model using different model selection criteria. Our main interest lies in selecting a model to be used for out-of-sample forecasting. In practice, forecasts are only produced based on models that display good

Figure 8 – Stored hydroelectric energy in South Brazil: time series data (top panel), correlogram (bottom left panel) and partial correlogram (bottom right panel).



Source: The author (2021).

Figure 9 – Histogram (left panel) and boxplot (right panel) of the data.



Source: The author (2021).

data fit, in particular, based on models that pass diagnostic testing. Hence, the selected models are submitted to portmanteau diagnostic testing based on the Q_4 test statistic proposed by Scher *et al.* (2020). The lag truncation parameter value used in the test statistic was $m = \lceil \sqrt{n} \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function. That is, $m \in \{9, 13, 14\}$ for

$n \in \{75, 150, 208\}$. We test the null hypothesis that the first m residual autocorrelations equal zero. Hence, under the null hypothesis the data dynamics are fully captured by the fitted model. When the null hypothesis is rejected, in contrast, the residuals are serially correlated and thus there is evidence of model misspecification. Models for which the correct model specification is rejected at the 5% significance level by the portmanteau test are discarded. When that happens, the next best fitting model according to the model selection criterion is selected. Even though we do not present results for the models that were discarded by the diagnostic analysis, we note that the out-of-sample forecasts produced by such models were typically less accurate than those obtained from the models we shall use in the empirical analysis that follows.

We consider three different data ranges. By Samples I and II we mean that the time series consists of the first 75 and 150 observations, respectively. Sample III refers to the complete sample (i.e., all 208 data points). When the time series only includes the first 75 observations ($n = 75$, Sample I), the EIC selects the βAR(3) model and the βARMA(1,1) model is chosen by all other criteria. Sample II includes the first 150 data points ($n = 150$). The model selected by the EIC is βAR(3). The βARMA(2,1) model is chosen by all the other criteria. The final scenario we consider is Sample III in which all 208 observations are used ($n = 208$). The EIC identifies the βARMA(2,3) model. A different model is selected by the remaining criteria, namely: βARMA(1,1).

The point estimates of α and ϕ for the models selected by the EIC (other criteria) in Samples I, II and II are, respectively, (i) 0.2380 and 12.0681 (0.4311 and 10.4292), (ii) 0.2366 and 14.0793 (0.3727 and 13.2730), and (iii) 0.1526 and 11.7536 (0.3739 and 11.1327). The point estimates (standard errors in parentheses) of the AR and MA parameters are presented in Table 14. For each sample, the top model was selected by the EIC and the bottom model was identified using the alternative criteria.

Our interest lies in forecasting y_{n+h}, for $h \geq 1$. Forecasting accuracy is assessed using MAPEs. For $h \in \{1, \ldots, 6\}$, $\text{MAPE}(h) = h^{-1} \sum_{j=1}^{h} |y_{n+j} - \hat{y}_n(j)|$, where $\hat{y}_n(j)$ denotes the forecast of y_{n+j} made at time n. The MAPEs obtained using the βARMA models identified by the different model selection criteria are presented in Table 15. The smallest MAPE for each value of h in each sample in Table 15 is displayed in boldface. The figures in this table lead to interesting conclusions. At the outset, consider the smallest sample size ($n = 75$, Sample I). The βARMA identified by the EIC outperformed the

Table 14 – Point estimates (standard errors in parentheses).

Sample	Model	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
I	βAR(3)	0.8825	−0.3783	0.1767	–	–	–
		(0.0444)	(0.0721)	(0.0592)	–	–	–
	βARMA(1,1)	0.4040	–	–	0.4317	–	–
		(0.0828)	–	–	(0.0957)	–	–
II	βAR(3)	0.9490	−0.3385	0.0871	–	–	–
		(0.0326)	(0.0511)	(0.0421)	–	–	–
	βARMA(2,1)	0.2501	0.2250	–	0.6485	–	–
		(0.1227)	(0.1035)	–	(0.1160)	–	–
III	βARMA(2,3)	0.2703	−0.4743	–	0.4778	−0.1512	0.1564
		(0.0812)	(0.0686)	–	(0.0920)	(0.0592)	(0.0609)
	βARMA(1,1)	0.4944	–	–	0.2562	–	–
		(0.0484)	–	–	(0.0612)	–	–

Source: The author (2021).

corresponding model chosen by all other criteria in all forecasting horizons, and by wide margins. For instance, when $h \in \{2,3\}$, the MAPEs of the forecasts made using the former were approximately 34% and 38% smaller than those obtained using the latter. When Sample II was used ($n = 150$), the dynamic beta model selected by the EIC (βAR(3)) outperformed that identified by all remaining criteria (βARMA(2,1)) for $h \in \{1,2,3,4,6\}$, i.e., it only fared worse for $h = 5$ and by a narrow margin (less than 4%). In some cases, the forecasts obtained from the βAR(3) model were considerably more precise than those from the competing beta model; e.g., MAPE nearly 20% smaller for $h = 1$. Next, we consider the situation in which the models were fitted using all data points ($n = 208$, Sample III). Here, the βARMA(2,3) model identified by the EIC yielded the most accurate forecasts for $h \in \{1,2,3,5,6\}$.

Table 15 – Mean absolute prediction errors.

Sample	Model	MAPE					
		$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
I	βAR(3)	**0.0990**	**0.0939**	**0.0801**	**0.0621**	**0.0611**	**0.0761**
	βARMA(1,1)	0.1296	0.1339	0.1222	0.0995	0.0853	0.0930
II	βAR(3)	**0.0254**	**0.0836**	**0.0591**	**0.0542**	0.0666	**0.0788**
	βARMA(2,1)	0.0316	0.0841	0.0647	0.0547	**0.0642**	0.0792
III	βARMA(2,3)	**0.0064**	**0.0228**	**0.0738**	**0.0734**	**0.0604**	**0.0585**
	βARMA(1,1)	0.0088	0.0369	0.0775	0.0748	0.0621	0.0586

Source: The author (2021).

In this section, we evaluated the forecasting accuracy of βARMA models identified using different model selection criteria. Overall, the EIC is the winner. In most cases, this criterion was able to select the best performing βARMA model. Its use yielded considerable gains in forecasting accuracy in some situations. For instance, when the smallest sample size was used (Sample I, $n = 75$), the MAPEs of the βAR(3) model

selected by the EIC for $h \in \{1, \ldots, 6\}$ were approximately 24%, 30%, 34%, 38%, 28% and 18% smaller, respectively, than those obtained with the βARMA(1,1) model that was selected by all other criteria.

It is worth noticing that in each sample (Samples I, II and III), different βARMA models were identified by (i) EIC and (ii) all other information criteria. That is, the EIC identified a model different from that selected by all other criteria. Recall that all identified models were sequentially submitted to portmanteau diagnostic testing. The selected model according to each criterion is the first model in the ordered (best to worst) list of models that passes diagnostic testing. It was only after diagnostic testing that one βARMA model was selected by the EIC and a different model was selected by the remaining information criteria.

We have also carried out model selection on the basis of the out-of-sample model selection criteria described in Section 2.3. The rolling windows of observations in MDF (n_r) are 50, 100 and 150 for Samples I, II and III, respectively, and the forecasting horizons are $h \in \{1, 3, 6\}$. As before, the Q_4 portmanteau diagnostic test was performed prior to forecasting. RHWE is computed using the same parameters as in Section 2.4. For all three sample sizes, the RHWE and MDF measures selected the same model identified by the EIC. This result was obtained using both MDA and MDV. The MAPE and RMSE measures selected models that are different from the ones identified on the basis of the EIC and all other information criteria. The forecasts from such models were uniformly less accurate than those obtained from the model selected by the EIC (and RHWE and MDF).

As a final exercise, we investigated the sensitivity of the different model selection strategies to the presence of outliers in the data. To that end, we introduced outliers into the complete time series (Sample III, $n = 208$). At the outset, we introduced a single outlier into the data as follows: (i) we multiplied $y_{52} = 0.8649$ by $a \in \{0.75, 0.50, 0.25\}$, (ii) we multiplied $y_{104} = 0.5407$ by $a \in \{1.75, 1.50, 1.25, 0.75, 0.50, 0.25\}$, and (iii) we multiplied $y_{156} = 0.8649$ by $a \in \{0.75, 0.50, 0.25\}$. We chose to modify the values of cases 52, 104 and 156, transforming them into atypical data points, because they are located at 25%, 50% and 75% of the time series length. Subsequently, we introduced three outliers into the data by replacing y_{52} and y_{156} with $0.25 \times y_{52}$ and $0.25 \times y_{156}$, respectively, and y_{104} with (i) $1.75 \times y_{104}$ and (ii) $0.25 \times y_{104}$. Model selection was not impacted by such outliers: in all cases, the EIC selected the βARMA(2,3) model and all other criteria selected the

βARMA$(1,1)$ model, as with the unperturbed data. In contrast, the presence of outliers in the data noticeably impacted the short term forecasts ($h \in \{1,2\}$) and overall had little impact on the forecasts when $h \in \{3,\ldots,6\}$. In particular, all one-step ahead forecasts became less accurate when the time series included one or three outliers. In future research we shall further investigate the impact of model misspecification and data anomalies on βARMA model selection and forecasting.

2.6 CONCLUDING REMARKS

The βARMA model is a dynamic model introduced by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017). It is tailored for use with time series that assume values in the standard unit interval such as rates, proportions and concentration indices. Parameter estimation is performed by conditional maximum likelihood. Diagnostic checking based on portmanteau tests in βARMA models was developed by Scher *et al.* (2020). Thus, it remained to investigate model identification prior to out-of-sample forecasting in that class of models. That was our chief goal in this chapter.

We considered βARMA model selection based on different information criteria. Since such criteria were not developed for dynamic models tailored to double bounded time series, it is important to investigate their usefulness in that context. We performed extensive and computer intensive simulations to estimate the rates of correct model identification of several criteria for different sample sizes and also by separately considering (i) autoregressive, (ii) moving average, and (iii) autoregressive and moving average dynamics. The numerical evidence we reported showed that all criteria yield more accurate model identification as the sample size increases. More importantly, it showed that model selection can be made substantially more accurate in samples of small to moderate sizes by using bootstrap resampling. In some cases, the frequency of correct model identification was more than double of that achieved by using criteria that are not resampling-based. We also considered model selection based on measures of forecasting accuracy. Our results showed that a measure based on directional forecasts leads to model selection that is more accurate than that obtained using alternative measures. Also, model selection guided by information criteria is more reliable than that guided by forecasting accuracy measures.

We also presented and discussed an empirical application. Our goal was to model and forecast the future behavior of the share of stored hydroelectric energy in South

Brazil. We used different samples that corresponded to different sample sizes (75, 150 and 208 observations). Interestingly, in all cases the bootstrap model selection criterion identified a model that was different from that selected on the basis of all other criteria. Model selection based on directional forecasting accuracy agrees with that performed using bootstrap resampling. In nearly all situations, the forecasts obtained with the models selected with the aid of the bootstrap-based information criterion were more accurate than those yielded by the models identified by the competing information criteria, in some cases by a wide margin (e.g., over 1/3).

Finally, a word of caution is in order. The simulation evidence we present provides a nice insight on the different model selection strategies' ability to recover the true model, and then on their reliability. In our empirical analysis, in contrast, we focus on out-of-sample forecasting. Here, selecting the true model might be less important than assessing the quality of the forecasts. Interestingly, in both settings (simulations and real data analysis) the best results were obtained using the same model selection strategy.

3 GENERALIZED βARMA MODEL FOR DOUBLE BOUNDED TIME SERIES FORECASTING

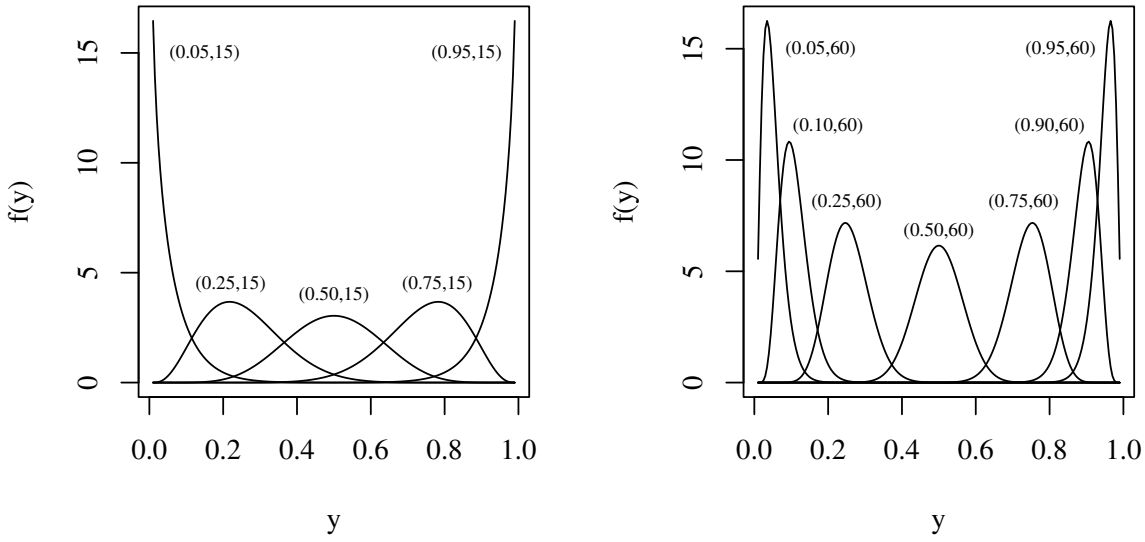
3.1 INTRODUCTION

The βARMA model was introduced by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017) as a dynamic extension of the class of beta regression models proposed by Ferrari and Cribari-Neto (2004); see Cribari-Neto and Zeileis (2010) and Douma and Weedon (2019). Both models are tailored for use with double bounded random variables, i.e., random variables that assume values in (a,b), where a and b are known and finite. Focus is typically placed on random variables that assume values in the standard unit interval, $(0,1)$, since when $a \neq 0$ and/or $b \neq 1$, one can subtract a from the variable of interest and then divide it by $b-a$ to obtain a random variable with support in the standard unit interval. The βARMA model is useful for modeling the behavior of double bounded random variables that evolve over time and for predicting their future levels. It incorporates autoregressive and moving average dynamics, allows for the inclusion of fixed covariates, and is heteroskedastic. A novel feature of the model is that it will never yield improper forecasts, i.e., forecasts that lie outside $(0,1)$. The model was successfully used by Melchior *et al.* (2021) to forecast mortality rates due to occupational accidents in three Brazilian states. Their results showed that the βARMA forecasts were typically more accurate than those obtained using alternative approaches.

In the βARMA model, it is assumed that the variable of interest follows the beta law parameterized in the terms of its mean, μ, and a precision parameter, ϕ; see beta density in Ferrari and Cribari-Neto (2004), Equation (4). The beta distribution is quite flexible, since its density can assume many shapes depending on the parameter values. The uniform distribution is a special case obtained with $\mu = 0.5$ and $\phi = 2$. The two panels in Figure 10 contain plots of beta densities for different values of (μ,ϕ). The density is symmetric when $\mu = 0$ and asymmetric otherwise. It can be J-shaped and also inverted J-shaped. It can also be skewed to the right or to the left. Thus, the beta law can easily accommodate distributional asymmetries. Additionally, since the beta variance is $\mu(1-\mu)/(1+\phi)$ and the mean parameter evolves over time, the βARMA model is naturally heteroskedastic, as noted above. It follows that the density shape and the distribution variance change over time, thus making the model quite flexible. It also takes into account the double bounded nature of the time series and, as a result, it will never yield forecasts

that lie outside the standard unit interval, as noted earlier.

Figure 10 – Beta densities for different values of (μ, ϕ).



Source: The author (2022).

βARMA diagnostic analysis was developed by Scher *et al.* (2020). The authors proposed a portmanteau test statistic, Q_4, that works well when used to test whether a fitted βARMA model is correctly specified. The test statistic uses partial residual autocorrelations and rejection of the null hypothesis is taken as evidence of model misspecification. The authors proved that, under the null hypothesis of correct model specification, the limiting distribution of their test statistic is χ^2_{m-p-q}, where m is the number of lags (number of partial autocorrelations), and p and q are the autoregressive and moving average orders, respectively.

Model selection was investigated by Cribari-Neto, Scher and Bayer (2022). They considered different βARMA model selection strategies and found that the best results are typically obtained by using an information criterion based on data (bootstrap) resampling. They also presented an empirical analysis in which the main goal was to model the proportion of stored hydroelectric energy in South Brazil. Furthermore, the authors showed that more accurate model selection typically translates into more accurate forecasts. Bayesian model selection for beta autoregressive processes was developed by Casarin, Valle and Leisen (2012).

Our goal in this chapter is to introduce a generalized, more flexible formulation of the βARMA model. In the standard formulation of the model, the beta law mean (conditional on previous information) evolves over time and the precision parameter is assumed to be globally fixed. By contrast, in the generalized formulation of the model we introduce both the conditional mean and the conditional precision are allowed to evolve dynamically. The generalized model contains two submodels, one for the mean and another for the precision. We use a parsimonious formulation for the dynamic structure that drives the precision over time.

The generalized βARMA model is used to produce out-of-sample forecasts of the time series modeled by Cribari-Neto, Scher and Bayer (2022) and also of an updated series that includes more recent data. Out-of-sample forecasts were also produced from models fitted to subsets of the data that range from 45 to 245 data points in steps of five observations, totaling 41 sample sizes. We compare such forecasts to those yielded by standard, fixed precision βARMA models. The results show that the generalized model's forecasts were typically more accurate than those from the standard βARMA model, especially the short-term ones (up to three steps ahead). In some cases, the gains in forecasting accuracy were sizable. The generalized model contains two submodels that allow for the two parameters that index the beta law (conditional on previous information) to evolve over time, whereas in the standard model only one beta law parameters (the mean) displays time series dynamics. The proposed model thus has an additional layer of flexibility, since it allows the beta density shape to evolve more freely over time. Our results show that such an additional layer of flexibility translates into forecasts that are typically more accurate than those obtained by imposing fixed precision.

The remainder of the chapter is structured as follows. In Section 3.2, we introduce the generalized βARMA model and in Section 3.3 we develop maximum likelihood inference for the model's parameters. In particular, we present the log-likelihood function, the score function, and Fisher's information matrix. Section 3.4 contains an empirical analysis in which we model and forecast the future levels of stored hydroelectric energy using the generalized and standard βARMA models. We consider different sample sizes and perform model selection using three strategies. The results show that out-of-sample forecasts obtained by allowing for varying precision are typically more accurate than those obtained under the assumption of fixed precision. Concluding remarks are offered in

Section 3.5.

3.2 THE GENERALIZED βARMA MODEL

The βARMA model proposed by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017) is a dynamic time series model for use with fractional data, i.e., doubly bounded data in the interval $(0,1)$, such as rates and proportions. In the following, we will introduce a generalization of the model with the goal of improving the accuracy of out-of-sample forecasts, especially short-term ones. The new model comprises two sub-models, one for the conditional mean of the process and one for the conditional precision.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ be an n-vector of time series random variables such that each y_t, $t = 1, \ldots, n$, conditionally on the set of previous information $\mathcal{F}_{t-1}$, follows the beta law with mean μ_t and precision ϕ_t. Here, $\mathcal{F}_{t-1}$ is the smallest σ-algebra such that the variables $y_1, \ldots, y_{t-1}$ are measurable. The conditional density of y_t, given $\mathcal{F}_{t-1}$, is

$$f(y_t|\mathcal{F}_{t-1}) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\Gamma((1-\mu_t)\phi_t)} y_t^{\mu_t\phi_t-1}(1-y_t)^{(1-\mu_t)\phi_t-1}, \quad 0 < y_t < 1,$$

where $\Gamma(\cdot)$ is the gamma function, $0 < \mu_t < 1$ and $\phi_t > 0$. The conditional mean and the conditional variance of y_t are, respectively, $\mathbb{E}(y_t|\mathcal{F}_{t-1}) = \mu_t$ and $\mathrm{var}(y_t|\mathcal{F}_{t-1}) = \mu_t(1-\mu_t)/(1+\phi_t)$.

Let $g_1 : (0,1) \mapsto \mathbb{R}$ be a strictly increasing and twice differentiable link function, such as the logit, probit, cauchit, log-log, and complementary log-log functions. In the βARMA model,

$$g_1(\mu_t) = \alpha_1 + \boldsymbol{x}_t^\top\boldsymbol{\beta} + \sum_{i=1}^{p}\varphi_i\left[g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top\boldsymbol{\beta}\right] + \sum_{j=1}^{q}\theta_j r_{t-j}, \tag{3.1}$$

where $\alpha_1 \in \mathbb{R}$ and $p, q \in \mathbb{N}$ are the autoregressive and moving average orders. Here, $\boldsymbol{x}_t \in \mathbb{R}^c$ is a set of non-random covariates at time t and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_c)^\top \in \mathbb{R}^c$ is a vector of parameters. Also, r_t is an error term which can be specified in the original scale, $y_t - \mu_t$, or in the predictor scale, $g_1(y_t) - g_1(\mu_t)$; in what follows, we will consider the latter.

In the standard formulation of the model, the precision parameter is assumed to be constant for all observations, i.e., $\phi_t = \phi\ \forall t$. We generalize the model by allowing the precision parameter to evolve over time. We use a parsimonious parametric structure aiming at improving short-term forecasting accuracy. Let $g_2 : \mathbb{R}_+ \mapsto \mathbb{R}$ be a strictly increasing and twice-differentiable link function, such as the log and square root functions.

Following Cribari-Neto and Zeileis (2010), we also consider the identity link function. The dynamic submodel for the precision parameter is specified as

$$g_2(\phi_t) = \alpha_2 + \delta z_{t-1}, \tag{3.2}$$

where $\alpha_2 \in \mathbb{R}$, $\delta \in \mathbb{R}$ and $z_t = y_t(1-y_t)$. The tth precision is $\phi_t = \exp(\alpha_2 + \delta z_{t-1})$, $\phi_t = (\alpha_2 + \delta z_{t-1})^2$ and $\phi_t = \alpha_2 + \delta z_{t-1}$ for the log, square root and identity link functions, respectively. The standard βARMA model is obtained as a particular case of our model by letting g_2 be the identity link and setting $\delta = 0$.

The rationale for the proposed extension of the βARMA model is as follows. For a given precision value, the variance of y_t increases with $\mu_t(1-\mu_t)$, being maximal at $\mu_t = 0.5$ and approaching zero as μ_t approaches zero or one. The precision submodel of the generalized βARMA model includes $y_{t-1}(1-y_{t-1})$ as an explanatory variable. When its value increases, there is some evidence of a variability increase in the previous period, and the model responds by decreasing the value of the current precision. It is expected that $\alpha_2 > 0$ and $\delta < 0$. The intercept (α_2) determines the maximal precision level, which is given by $g_2^{-1}(\alpha_2)$, and δz_{t-1} determines how the precisions fluctuate below it. Notice that $\delta < 0$ implies $\phi_t < g_2^{-1}(\alpha_2)$ $\forall t$. Consider, e.g., $y_{t-1} = 0.5$ (0.05 or 0.95). Then, $z_{t-1} = y_{t-1}(1-y_{t-1}) = 0.1875$ (0.0475). As long as $\delta < 0$, the current precision decreases whenever the previous value of the process moves towards the middle of standard unit interval, and increases otherwise. The value of δ determines the magnitude of the changes in the precision levels between consecutive time periods. As an example, suppose $\alpha_2 = 20$, $\delta = -39$ and the link function is identity. (These values are close to the estimates obtained in the next section for one of the time series we model.) When $y_{t-1} = 0.6$ we get $\phi_t = 10.64$, whereas when $y_{t-1} = 0.95$ we obtain $\phi_t = 18.1475$. Finally, we note that in the standard βARMA model changes to the beta density shape over time are only driven by changes in μ_t, whereas in the more general formulation of the model they follow from changes in μ_t and ϕ_t. There is thus greater flexibility in the generalized βARMA framework since the beta density shape may evolve more freely over time.

3.3 PARAMETER ESTIMATION

Parameter estimation for the generalized βARMA model given in (3.1) and (3.2) is performed by conditional maximum likelihood. The model can be expressed more

concisely as $g_1(\mu_t) = \eta_{1t}$ and $g_2(\phi_t) = \eta_{2t}$, where η_{1t} and η_{2t} are the mean and precision linear predictors. Let $\boldsymbol{\nu} = (\alpha_1, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top, \alpha_2, \delta)^\top$ be the k-dimensional parameter vector, where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_c)^\top$, with $k = p+q+c+3 < n$. The total conditional log-likelihood function for the parameter vector $\boldsymbol{\nu}$, given the first $a = \max\{p, q\}$ observations, is

$$\ell \equiv \ell(\mu_t, \phi_t) = \sum_{t=a+1}^{n} \ell_t(\mu_t, \phi_t),$$

where $\log f(y_t|\mathcal{F}_{t-1}) = \ell_t(\mu_t, \phi_t)$ and

$$\begin{aligned}\ell_t(\mu_t, \phi_t) = &\log\Gamma(\phi_t) - \log\Gamma(\mu_t, \phi_t) - \log\left((1-\mu_t)\phi_t\right) + (\mu_t\phi_t - 1)\log(y_t) \\ &+ \{(1-\mu_t)\phi_t - 1\}\log(1-y_t).\end{aligned}$$

The conditional maximum likelihood estimators of the model parameters cannot be expressed in closed form. Point estimates can be obtained by numerically maximizing ℓ using, say, a Newton or quasi-Newton optimization algorithm. In what follows we will use the BFGS quasi-Newton algorithm with analytical first derivatives; for details, see Nocedal and Wright (2006). When the model has moving average components, it is necessary account for the recursive structure of the derivatives of ℓ; see Rocha and Cribari-Neto (2017).

In the following, we will present closed-form expressions for the conditional score function and for the conditional (expected) information matrix. The latter is useful, for instance, for obtaining standard errors for the maximum likelihood point estimates, interval estimation and for hypothesis testing inferences.

3.3.1 Conditional score vector

Let $\boldsymbol{U} = (U_{\alpha_1}, \boldsymbol{U}_{\boldsymbol{\varphi}}^\top, \boldsymbol{U}_{\boldsymbol{\theta}}^\top, \boldsymbol{U}_{\boldsymbol{\beta}}^\top, U_{\alpha_2}, U_\delta)^\top$ be the conditional score vector. In order to obtain a closed-form expression for it, we need to obtain the derivatives of ℓ with respect to each parameter. The derivative of ℓ with respect to α_1 is

$$\frac{\partial \ell}{\partial \alpha_1} = \sum_{t=a+1}^{n} \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \alpha_1}.$$

It is important to note that $\partial\mu_t/\partial\eta_{1t} = 1/g_1'(\mu_t)$ and

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} = \phi_t \left\{ \log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)] \right\},$$

where $\psi(\cdot)$ is the digamma function. Let $y_t^* = \log(y_t/(1-y_t))$ and $\mu_t^* = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$. Thus,

$$\frac{\partial \ell}{\partial \alpha_1} = \sum_{t=a+1}^{n} \phi_t(y_t^* - \mu_t^*)\frac{1}{g_1'(\mu_t)}\left(1 + \sum_{j=1}^{q}\theta_j\frac{\partial r_{t-j}}{\partial \alpha_1}\right).$$

Let $\boldsymbol{s}$ be the $(n-a)$-dimensional vector with ith element given by

$$\frac{\partial \eta_{1(i+a)}}{\partial \alpha_1} = 1 + \sum_{j=1}^{q}\theta_j\frac{\partial r_{i+a-j}}{\partial \alpha_1},$$

$\boldsymbol{y}^* = (y_{a+1}^*, \ldots, y_n^*)^\top$ and $\boldsymbol{\mu}^* = (\mu_{a+1}^*, \ldots, \mu_n^*)^\top$. It thus follows that

$$U_{\alpha_1} = \boldsymbol{s}^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*),$$

where $\Phi = \mathrm{diag}\{\phi_{a+1}, \ldots, \phi_n\}$ and $T_1 = \mathrm{diag}\{1/g_1'(\mu_{a+1}), \ldots, 1/g_1'(\mu_n)\}$.

Additionally, for $l \in \{1, \ldots, c\}$,

$$\frac{\partial \ell}{\partial \beta_l} = \sum_{t=a+1}^{n} \phi_t(y_t^* - \mu_t^*)\frac{1}{g_1'(\mu_t)}\left(x_{tl} - \sum_{i=1}^{p}\varphi_i x_{(t-i)l} + \sum_{j=1}^{q}\theta_j\frac{\partial r_{t-j}}{\partial \beta_l}\right).$$

Let M be the $(n-a)\times c$ matrix whose ith row is

$$\frac{\partial \eta_{1(i+a)}}{\partial \boldsymbol{\beta}} = \boldsymbol{x}_{i+a} - \sum_{i=1}^{p}\varphi_i \boldsymbol{x}_{i+a} + \sum_{j=1}^{q}\theta_j\frac{\partial r_{i+a-j}}{\partial \boldsymbol{\beta}}.$$

We obtain

$$\boldsymbol{U}_{\boldsymbol{\beta}} = M^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

For $i \in \{1, \ldots, p\}$, we have

$$\frac{\partial \ell}{\partial \varphi_i} = \sum_{t=a+1}^{n} \phi_t(y_t^* - \mu_t^*)\frac{1}{g_1'(\mu_t)}\left(g_1(y_{t-i}) + \sum_{j=1}^{q}\theta_j\frac{\partial r_{t-j}}{\partial \varphi_i}\right).$$

Let P be the $(n-a)\times p$ matrix whose (i,j)th element is

$$\frac{\partial \eta_{1(i+a)}}{\partial \varphi_j} = g_1(y_{i+a-j}) + \sum_{l=1}^{q}\theta_l\frac{\partial r_{i+a-l}}{\partial \varphi_j}.$$

Thus,

$$\boldsymbol{U}_{\boldsymbol{\varphi}} = P^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

The derivative of ℓ with respect to θ_j, for $j \in \{1, \ldots, q\}$, is given by

$$\frac{\partial \ell}{\partial \theta_j} = \sum_{t=a+1}^{n} \phi_t(y_t^* - \mu_t^*)\frac{1}{g_1'(\mu_t)}\left(r_{t-j} + \sum_{j=1}^{q}\theta_j\frac{\partial r_{t-j}}{\partial \theta_j}\right).$$

Let R be the $(n-a)\times q$ matrix whose (i,j)th element is

$$\frac{\partial \eta_{1(i+a)}}{\partial \theta_j} = r_{i+a-j} + \sum_{l=1}^{q} \theta_l \frac{\partial r_{i+a-l}}{\partial \theta_j}.$$

Therefore,

$$\boldsymbol{U_\theta} = R^\top \Phi T_1 (\boldsymbol{y}^* - \boldsymbol{\mu}^*).$$

The conditional score function for α_2 is

$$\frac{\partial \ell}{\partial \alpha_2} = \sum_{t=a+1}^{n} \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \alpha_2}.$$

Here, $\partial \phi_t / \partial \eta_{2t} = 1/g_2'(\phi_t)$ and

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} = \mu_t(y_t^* - \mu_t^*) + \log(1-y_t) - \psi((1-\mu_t)\phi_t) + \psi(\phi_t).$$

Note that $g_2'(\phi_t) = 1$ when the identity precision link function is used. Let

$$\begin{aligned} H = \operatorname{diag}\Big\{ & \mu_{a+1}(y_{a+1}^* - \mu_{a+1}^*) + \log(1-y_{a+1}) - \psi((1-\mu_{a+1})\phi_{a+1}) \\ & + \psi(\phi_{a+1}), \ldots, \mu_n(y_n^* - \mu_n^*) + \log(1-y_n) - \psi((1-\mu_n)\phi_n) \\ & + \psi(\phi_n) \Big\} \end{aligned}$$

and $T_2 = \operatorname{diag}\{1/g_2'(\phi_{a+1}), \ldots, 1/g_2'(\phi_n)\}$. Thus,

$$U_{\alpha_2} = \mathbf{1}_n^\top H T_2 \mathbf{1}_n,$$

where $\mathbf{1}_n$ is an $(n-a)\times 1$ vector of ones. When g_2 is the identity link, T_2 is the $(n-a)$-dimensional identity matrix, and hence $U_{\alpha_2} = \mathbf{1}_n^\top H \mathbf{1}_n$.

Finally,

$$\frac{\partial \ell}{\partial \delta} = \sum_{t=a+1}^{n} \left[\mu_t(y_t^* - \mu_t^*) + \log(1-y_t) - \psi((1-\mu_t)\phi_t) + \psi(\phi_t)\right] \frac{1}{g_2'(\phi_t)} z_{t-1}.$$

Let $\boldsymbol{\omega}$ is the $(n-a)$-dimensional vector given by $\boldsymbol{\omega} = (z_a, \ldots, z_{n-1})^\top$. Thus,

$$U_\delta = \boldsymbol{\omega}^\top H T_2 \mathbf{1}_n.$$

When the precision link function is identity, $U_\delta = \boldsymbol{\omega}^\top H \mathbf{1}_n$.

The elements of the score vector $\boldsymbol{U}$ can then be expressed in matrix form as

$$U_{\alpha_1} = \boldsymbol{s}^\top \Phi T_1 (\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad \boldsymbol{U_\beta} = M^\top \Phi T_1 (\boldsymbol{y}^* - \boldsymbol{\mu}^*),$$

$$\boldsymbol{U}_{\boldsymbol{\varphi}} = P^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad \boldsymbol{U}_{\boldsymbol{\theta}} = R^\top \Phi T_1(\boldsymbol{y}^* - \boldsymbol{\mu}^*),$$
$$U_{\alpha_2} = \mathbf{1}_n^\top H T_2 \mathbf{1}_n \quad \text{and} \quad U_\delta = \boldsymbol{\omega}^\top H T_2 \mathbf{1}_n.$$

The conditional maximum likelihood estimator of $\boldsymbol{\nu}$ is obtained as the solution to the system of equations for $\boldsymbol{U} = \mathbf{0}_k$, where $\mathbf{0}_k$ is the $k \times 1$ vector of zeros. As noted earlier, it cannot be expressed in closed-form and estimates can be obtained by numerically maximizing the conditional log-likelihood function. Starting values for the parameters can be selected as follows: (i) all moving average parameters are set equal to zero, (ii) the values for the autoregressive parameters and α_1 are selected by regressing $g_1(y_t)$ on a constant and $g_1(y_{t-1}), \ldots, g_1(y_{t-p})$ using ordinary least squares, (iii) δ is set equal to zero, and (iv) α_2 is set equal to $g_2^{-1}(\phi^0)$, where ϕ^0 is computed as described on page 805 of Ferrari and Cribari-Neto (2004).

3.3.2 Conditional information matrix

In order to obtain the conditional Fisher information matrix, we need to compute the expected values of the second order log-likelihood derivatives.

Let $\boldsymbol{\lambda} = (\alpha_1, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$ and $\boldsymbol{\gamma} = (\alpha_2, \delta)^\top$. We have

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda_i \partial \lambda_j} &= \sum_{t=a+1}^{n} \frac{\partial}{\partial \mu_t}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \lambda_j}\right)\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \lambda_i} \\
&= \sum_{t=a+1}^{n}\left[\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \mu_t^2}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \lambda_j} + \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{\partial}{\partial \mu_t}\left(\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \lambda_j}\right)\right] \\
&\quad \times \frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \lambda_i},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} &= \sum_{t=a+1}^{n} \frac{\partial}{\partial \phi_t}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \phi_t}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \gamma_j}\right)\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \gamma_i} \\
&= \sum_{t=a+1}^{n}\left[\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \phi_t^2}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \gamma_j} + \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \phi_t}\frac{\partial}{\partial \phi_t}\left(\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \gamma_j}\right)\right] \\
&\quad \times \frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \gamma_i},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda_i \partial \gamma_j} &= \sum_{t=a+1}^{n} \frac{\partial}{\partial \phi_t}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \gamma_j}\right)\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \lambda_i} \\
&= \sum_{t=a+1}^{n}\left[\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \phi_t \partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \gamma_j} + \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{\partial}{\partial \phi_t}\left(\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \gamma_j}\right)\right] \\
&\quad \times \frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \lambda_i}.
\end{aligned}$$

Since $\mathbb{E}(\partial\ell_t(\mu_t,\phi_t)/\partial\mu_t|\mathcal{F}_{t-1}) = \mathbb{E}(\partial\ell_t(\mu_t,\phi_t)/\partial\phi_t|\mathcal{F}_{t-1}) = 0$, we have

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\lambda_i\partial\lambda_j}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=a+1}^{n}\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\middle|\mathcal{F}_{t-1}\right)\left(\frac{\partial\mu_t}{\partial\eta_{1t}}\right)^2\frac{\partial\eta_{1t}}{\partial\lambda_j}\frac{\partial\eta_{1t}}{\partial\lambda_i},$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\gamma_i\partial\gamma_j}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=a+1}^{n}\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\middle|\mathcal{F}_{t-1}\right)\left(\frac{\partial\phi_t}{\partial\eta_{2t}}\right)^2\frac{\partial\eta_{2t}}{\partial\gamma_j}\frac{\partial\eta_{2t}}{\partial\gamma_i},$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\lambda_i\partial\gamma_j}\middle|\mathcal{F}_{t-1}\right) = \sum_{t=a+1}^{n}\mathbb{E}\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t\partial\mu_t}\middle|\mathcal{F}_{t-1}\right)\left(\frac{\partial\mu_t}{\partial\eta_{1t}}\right)\left(\frac{\partial\phi_t}{\partial\eta_{2t}}\right)\frac{\partial\eta_{1t}}{\partial\lambda_j}\frac{\partial\eta_{2t}}{\partial\gamma_i}.$$

Using

$$\begin{aligned}
\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2} &= -\phi_t^2[\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)],\\
\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2} &= -\mu_t^2(\psi'(\mu_t\phi_t)) - (1-\mu_t)^2[\psi'((1-\mu_t)\phi_t)] + \psi'(\phi_t),\\
\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t\partial\mu_t} &= \psi((1-\mu_t)\phi_t) - \psi(\mu_t\phi_t) + (1-\mu_t)\phi_t\psi'((1-\mu_t)\phi_t)\\
&\quad - \mu_t\phi_t\psi'(\mu_t\phi_t) + \log\left(\frac{y_t}{1-y_t}\right),
\end{aligned}$$

we obtain

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2\ell}{\partial\lambda_i\partial\lambda_j}\middle|\mathcal{F}_{t-1}\right) &= -\sum_{t=a+1}^{n}\frac{A_t}{g_1'(\mu_t)^2}\frac{\partial\eta_{1t}}{\partial\lambda_j}\frac{\partial\eta_{1t}}{\partial\lambda_i},\\
\mathbb{E}\left(\frac{\partial^2\ell}{\partial\gamma_i\partial\gamma_j}\middle|\mathcal{F}_{t-1}\right) &= -\sum_{t=a+1}^{n}\frac{B_t}{g_2'(\phi_t)^2}\frac{\partial\eta_{2t}}{\partial\gamma_j}\frac{\partial\eta_{2t}}{\partial\gamma_i},\\
\mathbb{E}\left(\frac{\partial^2\ell}{\partial\lambda_i\partial\gamma_j}\middle|\mathcal{F}_{t-1}\right) &= -\sum_{t=a+1}^{n}\frac{C_t}{g_1'(\mu_t)g_2'(\phi_t)}\frac{\partial\eta_{1t}}{\partial\lambda_j}\frac{\partial\eta_{2t}}{\partial\gamma_i},
\end{aligned}$$

where $A_t = \phi_t^2[\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)]$, $B_t = \mu_t^2(\psi'(\mu_t\phi_t)) + (1-\mu_t)^2[\psi'((1-\mu_t)\phi_t)] - \psi'(\phi_t)$ and $C_t = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t) - (1-\mu_t)\phi_t\psi'((1-\mu_t)\phi_t) + \mu_t\phi_t\psi'(\mu_t\phi_t) - \log(y_t/(1-y_t))$.

Let $W_1 = \mathrm{diag}\{w_{1(a+1)},\ldots,w_{1(n)}\}$, $W_2 = \mathrm{diag}\{w_{2(a+1)},\ldots,w_{2(n)}\}$ and $W_3 = \mathrm{diag}\{w_{3(a+1)},\ldots,w_{3(n)}\}$, with

$$w_{1(t)} = \frac{A_t}{g_1'(\mu_t)^2}, \quad w_{2(t)} = \frac{B_t}{g_2'(\phi_t)^2} \quad \text{and} \quad w_{3(t)} = \frac{C_t}{g_1'(\mu_t)g_2'(\phi_t)}.$$

Thus,

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2\ell}{\partial\alpha_1^2}\middle|\mathcal{F}_{t-1}\right) &= -\boldsymbol{s}^\top W_1\boldsymbol{s}, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\alpha_1}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1\boldsymbol{s},\\
\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\varphi}\partial\alpha_1}\middle|\mathcal{F}_{t-1}\right) &= -P^\top W_1\boldsymbol{s}, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\alpha_1}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_1\boldsymbol{s},
\end{aligned}$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\alpha_1\partial\alpha_2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{s}^\top W_3\mathbf{1}_n, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\alpha_1\partial\delta}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{s}^\top W_3\boldsymbol{\omega},$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 M, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\varphi}^\top}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 P,$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^\top}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_1 R, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\alpha_2}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_3\mathbf{1}_n,$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\delta}\middle|\mathcal{F}_{t-1}\right) = -M^\top W_3\boldsymbol{\omega}, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}^\top}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_1 P,$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\theta}^\top}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_1 R, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\varphi}\partial\alpha_2}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_3\mathbf{1}_n,$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\varphi}\partial\delta}\middle|\mathcal{F}_{t-1}\right) = -P^\top W_3\boldsymbol{\omega}, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_1 R,$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\alpha_2}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_3\mathbf{1}_n, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\boldsymbol{\theta}\partial\delta}\middle|\mathcal{F}_{t-1}\right) = -R^\top W_3\boldsymbol{\omega},$$
$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\delta\partial\alpha_2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{\omega}^\top W_2\mathbf{1}_n, \quad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\delta^2}\middle|\mathcal{F}_{t-1}\right) = -\boldsymbol{\omega}^\top W_2\boldsymbol{\omega}.$$

The joint conditional Fisher information matrix can be expressed as

$$K \equiv K(\boldsymbol{\nu}) = \begin{bmatrix} K_{\alpha_1\alpha_1} & K_{\alpha_1\boldsymbol{\beta}} & K_{\alpha_1\boldsymbol{\varphi}} & K_{\alpha_1\boldsymbol{\theta}} & K_{\alpha_1\alpha_2} & K_{\alpha_1\delta} \\ K_{\boldsymbol{\beta}\alpha_1} & K_{\boldsymbol{\beta}\boldsymbol{\beta}} & K_{\boldsymbol{\beta}\boldsymbol{\varphi}} & K_{\boldsymbol{\beta}\boldsymbol{\theta}} & K_{\boldsymbol{\beta}\alpha_2} & K_{\boldsymbol{\beta}\delta} \\ K_{\boldsymbol{\varphi}\alpha_1} & K_{\boldsymbol{\varphi}\boldsymbol{\beta}} & K_{\boldsymbol{\varphi}\boldsymbol{\varphi}} & K_{\boldsymbol{\varphi}\boldsymbol{\theta}} & K_{\boldsymbol{\varphi}\alpha_2} & K_{\boldsymbol{\varphi}\delta} \\ K_{\boldsymbol{\theta}\alpha_1} & K_{\boldsymbol{\theta}\boldsymbol{\beta}} & K_{\boldsymbol{\theta}\boldsymbol{\varphi}} & K_{\boldsymbol{\theta}\boldsymbol{\theta}} & K_{\boldsymbol{\theta}\alpha_2} & K_{\boldsymbol{\theta}\delta} \\ K_{\alpha_2\alpha_1} & K_{\alpha_2\boldsymbol{\beta}} & K_{\alpha_2\boldsymbol{\varphi}} & K_{\alpha_2\boldsymbol{\theta}} & K_{\alpha_2\alpha_2} & K_{\alpha_2\delta} \\ K_{\delta\alpha_1} & K_{\delta\boldsymbol{\beta}} & K_{\delta\boldsymbol{\varphi}} & K_{\delta\boldsymbol{\theta}} & K_{\delta\alpha_2} & K_{\delta\delta} \end{bmatrix},$$

where $K_{\alpha_1\alpha_1} = \boldsymbol{s}^\top W_1\boldsymbol{s}$, $K_{\alpha_1\boldsymbol{\beta}} = K_{\boldsymbol{\beta}\alpha_1}^\top = M^\top W_1\boldsymbol{s}$, $K_{\alpha_1\boldsymbol{\varphi}} = K_{\boldsymbol{\varphi}\alpha_1}^\top = P^\top W_1\boldsymbol{s}$, $K_{\alpha_1\boldsymbol{\theta}} = K_{\boldsymbol{\theta}\alpha_1}^\top = R^\top W_1\boldsymbol{s}$, $K_{\alpha_1\alpha_2} = K_{\alpha_2\alpha_1}^\top = \boldsymbol{s}^\top W_3\mathbf{1}_n$, $K_{\alpha_1\delta} = K_{\delta\alpha_1}^\top = \boldsymbol{s}^\top W_3\boldsymbol{\omega}$, $K_{\boldsymbol{\beta}\boldsymbol{\beta}} = M^\top W_1 M$, $K_{\boldsymbol{\beta}\boldsymbol{\varphi}} = K_{\boldsymbol{\varphi}\boldsymbol{\beta}}^\top = M^\top W_1 P$, $K_{\boldsymbol{\beta}\boldsymbol{\theta}} = K_{\boldsymbol{\theta}\boldsymbol{\beta}}^\top = M^\top W_1 R$, $K_{\boldsymbol{\beta}\alpha_2} = K_{\alpha_2\boldsymbol{\beta}}^\top = M^\top W_3\mathbf{1}_n$, $K_{\boldsymbol{\beta}\delta} = K_{\delta\boldsymbol{\beta}}^\top = M^\top W_3\boldsymbol{\omega}$, $K_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = P^\top W_1 P$, $K_{\boldsymbol{\varphi}\boldsymbol{\theta}} = K_{\boldsymbol{\theta}\boldsymbol{\varphi}}^\top = R^\top W_1 P$, $K_{\boldsymbol{\varphi}\alpha_2} = K_{\alpha_2\boldsymbol{\varphi}}^\top = P^\top W_3\mathbf{1}_n$, $K_{\boldsymbol{\varphi}\delta} = K_{\delta\boldsymbol{\varphi}}^\top = P^\top W_3\boldsymbol{\omega}$, $K_{\boldsymbol{\theta}\boldsymbol{\theta}} = R^\top W_1 R$, $K_{\boldsymbol{\theta}\alpha_2} = K_{\alpha_2\boldsymbol{\theta}}^\top = R^\top W_3\mathbf{1}_n$, $K_{\boldsymbol{\theta}\delta} = K_{\delta\boldsymbol{\theta}}^\top = R^\top W_3\boldsymbol{\omega}$, $K_{\alpha_2\alpha_2} = \mathbf{1}_n^\top W_2\mathbf{1}_n$, $K_{\alpha_2\delta} = K_{\delta\alpha_2}^\top = \boldsymbol{\omega}^\top W_2\mathbf{1}_n$ and $K_{\delta\delta} = \boldsymbol{\omega}^\top W_2\boldsymbol{\omega}$.

The parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are not orthogonal (i.e., Fisher's information matrix is not block diagonal). When n is large, $\hat{\boldsymbol{\nu}}$, the conditional maximum likelihood estimator of $\boldsymbol{\nu}$, is approximately distributed as $\mathcal{N}_k(\boldsymbol{\nu}, K^{-1}(\boldsymbol{\nu}))$.

3.4 OUT-OF-SAMPLE FORECASTING EVALUATION

The generalized βARMA model adds an additional layer of flexibility to the standard formulation of the model since it allows the precision parameter to evolve over time. To what extent such additional flexibility translates into more accurate out-of-sample short-term forecasts? We will answer this question using the data analyzed by Cribari-Neto, Scher and Bayer (2022) and also an an updated version of the time series used by these authors. Their focus was on (fixed precision) βARMA model selection. Their results indicate that model selection based on an Empirical Information Criterion (EIC) which makes use of parametric bootstrap resampling typically outperforms those based on alternative strategies.

Our interest is in modeling and forecasting the proportion of stored hydroelectric energy in South Brazil. Stored energy is the energy value of the accumulated water, i.e., how much energy (in Megawatt monthly) can be generated from the stored volume of water expressed as a proportion of the total hydroelectric power plant capacity. Our study is structured in three parts: (i) we use the monthly averages of stored energy from July 2000 to April 2018, totaling 214 observations, with the final six observations reserved for forecast evaluation; the complete data contain $n = 208$ observations corresponding to the July 2000 to October 2017 period; (ii) we use data from July 2000 to May 2021, totaling 251 observations, with the final six data points reserved for forecast evaluation, hence $n = 245$; (iii) we use 41 sample sizes in a sequential forecasting analysis; the sample sizes are $n \in \{45, 50, \ldots, 245\}$. The data in the first part of our forecasting exercise are the same as used by Cribari-Neto, Scher and Bayer (2022). In the second and third parts of the experiment, we work with an updated version of the time series. In all cases, our goal is to perform a comparative analysis between standard (fixed precision) and generalized (variable precision) βARMA forecasts. All estimations, descriptive analyses, and graphical analyses were carried out using the R statistical computing environment; see Team (2021). EIC-based model selection was performed using 1,000 bootstrap replications.

In all empirical analyses that follow, model selection was performed using three information criteria, namely AIC (Akaike), SIC (Schwarz) and EIC. For details on the former two criteria, see Burnham and Anderson (2004) and Choi (1992); for details on the EIC, see Cavanaugh and Shumway (1997) and Cribari-Neto, Scher and Bayer (2022). We search for the best model by considering all combinations of p and q such that $p, q = 0, \ldots, 4$,

except, of course, for the $(0,0)$ model. The Q_4 portmanteau test proposed by Scher *et al.* (2020) was performed using the residuals from the fitted selected model. Following Cribari-Neto, Scher and Bayer (2022), models for which the correct model specification is rejected by the Q_4 test at the 5% significance level are discarded. When that happens, the next best fitting model according to the model selection criterion is selected. In all fitted models, g_1 is the logit link and in all fitted generalized βARMA models, g_2 is the identity link. We also considered the log and square root precision link functions; these results will not be shown for brevity. We note that slightly more accurate forecasts were obtained using the identity precision link.

Table 16 presents descriptive statistics for the two time series ($n = 214$ and $n = 251$). We report the maximal and minimal values and also the means, medians, variances, coefficients of skewness, and coefficients of excess kurtosis. The longer time series displays smaller minimal value, mean and median. It also displays less skewness and excess kurtosis.

Table 16 – Descriptive statistics, stored hydroelectric energy in South Brazil.

n	min	max	median	mean	variance	skewness	kurtosis
214	0.2977	0.9862	0.7265	0.7016	0.0404	−0.2714	−1.2180
251	0.1495	0.9862	0.6477	0.6613	0.0461	−0.2169	−1.0072

Source: The author (2022).

In the first part of our study, we follow Cribari-Neto, Scher and Bayer (2022) and split the series into three subsets of distinct sample sizes, namely $n = 75$ (Sample I), $n = 150$ (Sample II), and $n = 208$ (Sample III). The Q_4 test statistic was computed using 9, 13, 14 lags for 75, 150 and 208 observations, respectively. We performed model selection for the generalized βARMA model. When $n = 75$ (Sample I), the EIC selected the βAR(3) model, whereas the AIC and SIC selected the βARMA$(1,1)$ model. When $n = 150$ (Sample II), the EIC selected the βAR(3) model and the βARMA$(1,1)$ model was selected by the other information criteria; the latter differs from the βARMA$(2,1)$ model selected by the same criteria under fixed precision in Cribari-Neto, Scher and Bayer (2022). Finally, when $n = 208$, the βARMA$(2,3)$ model was selected by the EIC whereas the remaining criteria selected the βARMA$(1,1)$ model. Table 17 presents the values of the AIC, SIC and EIC for the βARMA models with fixed and variable precision (denoted by ϕ and ϕ_t, respectively) selected by these criteria. The smallest values for each criterion

are denoted in boldface. It is noteworthy that three model selection criteria favor the generalized models over the standard models in all three samples.

Table 17 – Model selection criteria values for the selected standard and generalized models.

Sample	Criterion	ϕ	ϕ_t
I	AIC	-113.8075	$\mathbf{-118.4186}$
	SIC	-104.5375	$\mathbf{-106.8311}$
	EIC	-123.9971	$\mathbf{-129.8971}$
II	AIC	-247.5301	$\mathbf{-255.6199}$
	SIC	-235.4876	$\mathbf{-240.5667}$
	EIC	-260.1577	$\mathbf{-267.2730}$
III	AIC	-322.4067	$\mathbf{-332.0913}$
	SIC	-309.0565	$\mathbf{-315.4036}$
	EIC	-340.0288	$\mathbf{-349.9947}$

Source: The author (2022).

Table 18 contains the conditional maximum likelihood estimates (standard errors in parentheses) of the parameters that index the selected generalized βARMA models. For each sample, the models above and below the horizontal line are those selected by the EIC and AIC/BIC, respectively. As expected, for all models $\hat{\alpha}_2 > 0$ and $\hat{\delta} < 0$.

Next, we tested the null hypothesis of constant precision versus the alternative hypothesis of variable precision, i.e., we test $\mathcal{H}_0 : \delta = 0$ versus $\mathcal{H}_1 : \delta \neq 0$. The likelihood ratio test p-values for the selected generalized βARMA models in the three sample sizes are presented in Table 19. In all cases, fixed precision is rejected at the 5% significance level and in some cases rejection takes place at 1%. There is thus clear evidence in favor of variable dispersion.

We will now move to forecasting evaluation. Using each fixed precision and variable precision selected model for each sample size, we produced forecasts of y_{n+h}, $h \in \{1, \ldots, 6\}$. That is, we produced forecasts of the next six observations. Table 20 contains the mean absolute prediction errors (MAPEs) of the forecasts which are computed as $\text{MAPE}(h) = h^{-1} \sum_{j=1}^{h} |y_{n+j} - \hat{y}_n(j)|$, with $\hat{y}_n(j)$ denoting the forecast of y_{n+j} made at time n. The best result for each sample size and forecasting horizon is displayed in boldface.

The figures in Table 20 convey important information. First, in all sample sizes, more accurate forecasts were obtained using the generalized βARMA model. Second, in

Table 18 – Point estimates (standard errors in parentheses), generalized models.

Sample	Model	$\hat{\alpha}_1$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\alpha}_2$	$\hat{\delta}$
I	βAR(3)	0.1854	0.9324	−0.3675	0.1958	–	–	–	22.4072	−51.5107
		(0.1136)	(0.1041)	(0.1270)	(0.0926)	–	–	–	(6.8816)	(33.9520)
	βARMA(1,1)	0.4029	0.5639	–	–	0.3957	–	–	22.2637	−54.7338
		(0.1502)	(0.0990)	–	–	(0.1096)	–	–	(6.5113)	(31.1118)
II	βAR(3)	0.2066	0.9598	−0.3343	0.1093	–	–	–	23.7350	−52.1579
		(0.0719)	(0.0758)	(0.0974)	(0.0687)	–	–	–	(5.3322)	(25.2859)
	βARMA(1,1)	0.3240	0.6094	–	–	0.3546	–	–	24.5026	−58.0225
		(0.0954)	(0.0676)	–	–	(0.0826)	–	–	(5.3217)	(24.7131)
III	βARMA(2,3)	0.1512	0.4042	−0.5861	–	−0.5391	−0.1726	0.2357	21.4448	−47.7560
		(0.0619)	(0.1543)	(0.1160)	–	(0.1648)	(0.0862)	(0.0898)	(4.1466)	(19.5878)
	βARMA(1,1)	0.3505	0.5795	–	–	0.3075	–	–	20.1116	−43.8440
		(0.0833)	(0.0606)	–	–	(0.0737)	–	–	(3.7753)	(17.6981)

Source: The author (2022).

Table 19 – p-values of the likelihood ratio test of constant precision ($\mathcal{H}_0 : \delta = 0$).

Sample	Model	p-value
I	βAR(3)	0.0463
	βARMA(1,1)	0.0135
II	βAR(3)	0.0226
	βARMA(1,1)	0.0024
III	βARMA(2,3)	0.0018
	βARMA(1,1)	< 0.0001

Source: The author (2022).

Table 20 – Mean absolute prediction errors, first empirical analysis.

Sample	Model	Precision	MAPE					
			$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
I	βAR(3)	ϕ_t	**0.0773**	**0.0653**	**0.0492**	**0.0465**	**0.0538**	**0.0737**
		ϕ	0.0990	0.0939	0.0801	0.0621	0.0611	0.0761
	βARMA(1,1)	ϕ_t	0.1202	0.1261	0.1197	0.1027	0.0823	0.0848
		ϕ	0.1296	0.1339	0.1222	0.0995	0.0853	0.0930
II	βAR(3)	ϕ_t	**0.0140**	**0.0669**	**0.0568**	**0.0461**	**0.0559**	**0.0727**
		ϕ	0.0254	0.0836	0.0591	0.0542	0.0666	0.0788
	βARMA(1,1)	ϕ_t	0.0248	0.0823	0.0583	0.0549	0.0680	0.0791
	βARMA(2,1)	ϕ	0.0316	0.0841	0.0647	0.0547	0.0642	0.0792
III	βARMA(2,3)	ϕ_t	**0.0027**	**0.0133**	**0.0709**	0.0713	0.0575	0.0594
		ϕ	0.0064	0.0228	0.0738	0.0734	0.0604	0.0585
	βARMA(1,1)	ϕ_t	0.0051	0.0370	0.0750	**0.0701**	**0.0567**	**0.0568**
		ϕ	0.0088	0.0369	0.0775	0.0748	0.0621	0.0586

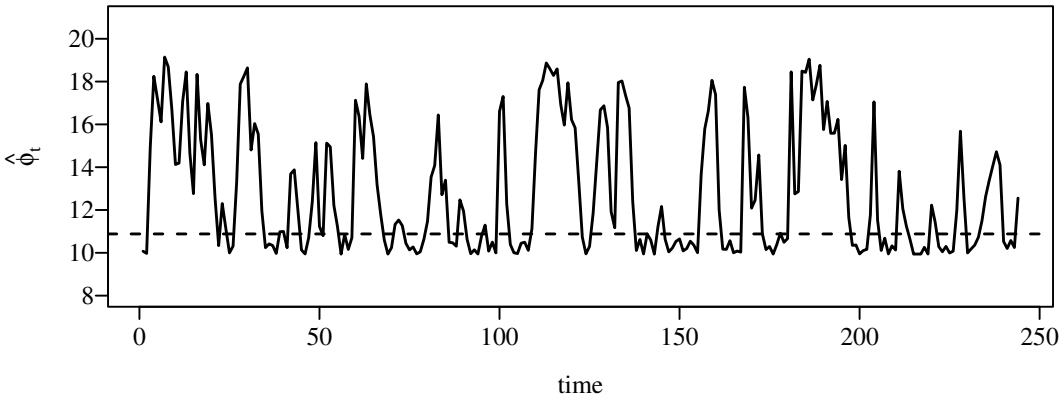
Source: The author (2022).

Samples I and II, the models selected by the EIC yielded the best results. In Sample III, the best results for $h \in \{1,2,3\}$ (short-term forecasting) were obtained using the EIC for generalized βARMA model selection, whereas the most accurate forecasts for $h \in \{4,5,6\}$ were yielded by the variable precision model selected by the AIC and BIC. Third, in some cases, the gains in forecasting accuracy achieved by allowing for variable precision are large. Consider, e.g., Sample I, $h = 2$ and βAR(3) (model selected by the EIC). The MAPE of the forecasts from the generalized model is over 30% smaller than that of the standard model's forecasts; in Sample III, $h = 1$ and βARMA(2,3), the gain in MAPE is of nearly 58%.

We will now move to the second part of our empirical investigation, in which we use more recent data. Here, the data range from July 2000 to May 2021, totaling 251 observations, and the final six observations are reserved for forecasting evaluation. Thus, the effective sample size is $n = 245$. The Q_4 portmanteau test statistic used to assess model misspecification employs 15 lags. The βARMA(1,1) model was selected by all three

criteria under both variable and fixed precision. The AIC, BIC and EIC values for the variable (fixed) precision model are, respectively, -381.0816, -374.0319 and -392.3007 (-372.7727, -367.1329 and -380.5489). The three criteria favor the generalized βARMA model. The p-value of the likelihood ratio test of $\mathcal{H}_0 : \delta = 0$ is 0.0015. We thus reject the null hypothesis of fixed precision at the 1% significance level. For brevity, we only report precision estimates (standard errors in parentheses). For the standard model, $\hat{\phi} = 10.8825$ (0.9674). For the generalized model, $\hat{\alpha}_2 = 19.6795$ (3.6415) and $\hat{\delta} = -38.9409$ (17.0931); again, as expected, these estimates are positive and negative, respectively. Figure 11 contains an index plot of $\hat{\phi}_t$ with a dashed horizontal line at the estimated precision parameter estimate from the fitted standard model (10.8825). The minimal and maximal estimated precisions from the generalized model are 9.9385 and 19.1435, respectively, the average precision being 12.8286. As expected, the largest estimated precisions are associated with observations that are close to an endpoint of the standard unit interval. For instance, there are 51 estimated precisions in excess of 16, and they all coincide with observations that are somewhat close one; the minimum value of these 51 data points is 0.8958.

Figure 11 – Estimated precisions from the fitted generalized βARMA model.

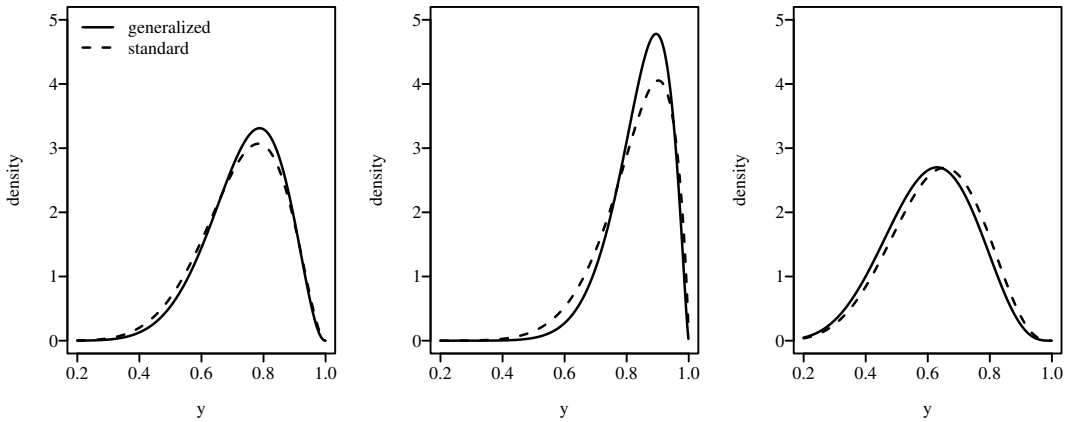


Source: The author (2022).

As noted earlier, changes in the beta density shape across observations are only driven by μ_t in the standard βARMA model; by contrast, in the more general formulation of the model they are driven by μ_t and ϕ_t. The general model thus has an additional layer of flexibility since it allows the beta density shape to evolve more freely over time. In order to exemplify that, we present in Figure 12 the beta density functions evaluated at the mean and precision estimates obtained from the two models ($\hat{\mu}_t$ and $\hat{\phi}_t$ for the

generalized model and $\hat{\mu}_t$ and $\hat{\phi}$ for the standard model) for observations 48 (left panel), 49 (middle panel) and 50 (right panel). Note that the shape of the beta density changes more intensely over the three time periods when we consider the mean and precision estimates obtained from the generalized model.

Figure 12 – Estimated beta densities for observations 48, 49, 50.



Source: The author (2022).

Again, forecasts of y_{n+h} for $h \in \{1, \ldots, n\}$ were produced using the generalized and standard βARMA models. The MAPEs are presented in Table 21. The forecasts obtained under variable precision were more accurate for all forecasting horizons. When $h = 1$, the gain in accuracy from using the generalized model exceeded 62%.

Table 21 – Mean absolute prediction errors, second empirical analysis.

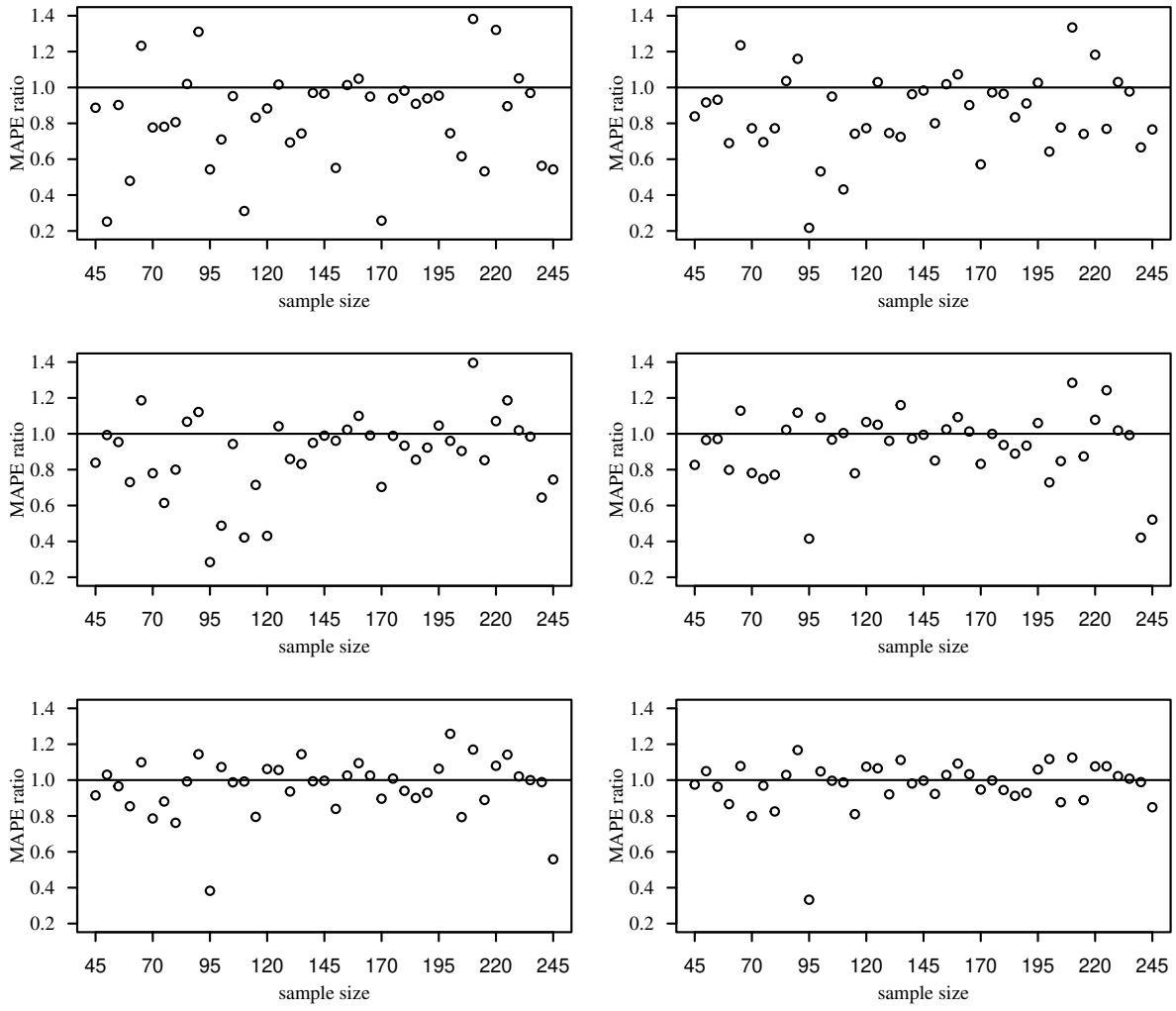
Sample	Model	Precision	MAPE					
			$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
$n=245$	βARMA(1,1)	ϕ_t	**0.0062**	**0.0561**	**0.0809**	**0.0720**	**0.0658**	**0.0641**
		ϕ	0.0167	0.0590	0.0834	0.0756	0.0666	0.0644

Source: The author (2022).

In the third and final part of our empirical analysis, we consider 41 sample sizes that range from $n = 45$ to $n = 245$ in steps of five observations (i.e., $n \in \{45, 50, \ldots, 245\}$). For each sample size, a generalized and a standard βARMA model were selected using the EIC (the bootstrap-based model selection criterion); as before, the Q_4 portmanteau test of correct model specification was performed on the residuals from both models. For each sample size, forecasts of the next six observations ($h \in \{1, \ldots, 6\}$) were produced using each model, MAPEs were computed for each set of forecasts, and the ratios between the MAPEs of the forecasts from the generalized and standard models were computed. That is, we computed, for each n, $\text{MAPE}_r(h) = \text{MAPE}_g(h)/\text{MAPE}_s(h)$, where the subscripts '$g$' and

's' stand for 'generalized' and 'standard', respectively. Values of $\text{MAPE}_r(h)$ smaller than one (greater than one) favor the generalized (standard) model. Overall, the results clearly favor the generalized model, especially when the interest lies in short-term forecasting. For instance, for $h=1$ ($h=2$) [$h=3$], the aforementioned ratio was smaller than one in 78.05% (75.61%) [73.17%] of the 41 sample sizes. In Figure 13 we present, in six panels, plots of $\text{MAPE}_r(h)$ against the sample size, each panel corresponding to a forecasting horizon. Points that lie below (above) the horizontal line drawn at 1.0 are indicative of better (worse) forecasting accuracy of the generalized βARMA model relative to the standard model. The former clearly outperforms the latter, especially for $h \in \{1,2,3\}$.

Figure 13 – $\text{MAPE}_r(h)$ **vs** $n \in \{45,50,\ldots,245\}$**:** $h=1$**: top left;** $h=2$**: top right;** $h=3$**: middle left;** $h=4$**: middle right;** $h=5$**: bottom left;** $h=6$**: bottom right.**



Source: The author (2022).

We also computed the mean values of $\text{MAPE}_r(h)$ (i.e., average over the 41 sample sizes) for each h. The figure for $h=1$ ($h=2$) [$h=3$] was 0.8348 (0.8560) [0.8859]. It

is thus clear that the generalized model yielded forecasts that were, on average, considerably more accurate than those from the standard model. It also outperformed the standard model for larger forecasting horizons ($h \in \{4,5,6\}$), but by smaller margins; e.g., the mean MAPE ratio for $h = 4$ ($h = 5$) [$h = 6$] was 0.9324 (0.9626) [0.9741]. It is natural for the gains in forecasting accuracy achieved by the generalized model to be more pronounced for small forecasting horizons, since the two sets of forecasts (generalized and standard) converge to $\bar{y}$, the average value of y_t's, as h increases.

3.5 CONCLUDING REMARKS

The βARMA model proposed by Rocha and Cribari-Neto (2009), Rocha and Cribari-Neto (2017) is useful for modeling random variables that assume values in $(0,1)$ and evolve over time. The model extends the class of beta regressions so that it can be used with non-independent, serially correlated random variables. Like its regression counterpart, it accommodates distributional asymmetries, accounts for heteroskedasticity, and does not yield improper predictions. Both models are based on the assumption that the variable of interest is beta-distributed. In the dynamic model, such an assumption is made for the variable of interest at each time period conditional on the set of previous information. As is well known, the beta law is very flexible, since its density can assume many shapes depending on the parameter values. Also, both models are based on the beta parametrization proposed by Ferrari and Cribari-Neto (2004) according to which the beta density is indexed by the distribution mean and a precision parameter.

The standard formulation of the βARMA model allows the mean parameter to evolve over time, but imposes that the precision is globally fixed, i.e., it is assumed that the precision parameter has same value for all observations. In this chapter, we introduced a more general formulation of the model which allows the two parameters that index the beta law (conditional mean and conditional precision) to vary over time. We developed maximum likelihood inference for the proposed model. In particular, we presented closed-form expressions for the model's conditional log-likelihood function, conditional score function and conditional Fisher's information matrix.

The proposed model was used to model and forecast future levels of stored hydroelectric energy in the South of Brazil. Different sample sizes were considered and, for each sample size, generalized and standard βARMA models were selected using three

different information criteria. One of the model selection criteria we used is based on bootstrap resampling. In most configurations, more accurate out-of-samples forecasts were obtained by using the more general formulation of the model. In some cases, the gains in forecasting accuracy were large, especially for short-term forecasting. It is noteworthy that the general formulation of the model is more flexible in the sense that changes in the beta density shape over time are driven by two parameters (mean and precision), and not only by a single parameter (mean) like in the standard model formulation. We encourage practitioners who wish to model and forecast double bounded time series to use the generalized βARMA model proposed in this chapter.

REFERENCES

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, IEEE, v. 19, n. 6, p. 716–723, 1974.

ANDERSEN, E. B. Asymptotic properties of conditional maximum-likelihood estimators. **Journal of the Royal Statistical Society B**, JSTOR, v. 32, n. 2, p. 283–301, 1970.

BAYER, F. M.; BAYER, D. M.; PUMI, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. **Journal of Hydrology**, v. 555, p. 385–396, 2017.

BAYER, F. M.; CINTRA, R. J.; CRIBARI-NETO, F. Beta seasonal autoregressive moving average models. **Journal of Statistical Computation and Simulation**, v. 88, n. 15, p. 2961–2981, 2018.

BAYER, F. M.; CRIBARI-NETO, F. Bootstrap-based model selection criteria for beta regressions. **TEST**, v. 24, n. 4, p. 776–795, 2015.

BAYER, F. M.; CRIBARI-NETO, F. Model selection criteria in beta regression with varying dispersion. **Communications in Statistics - Simulation and Computation**, v. 46, n. 4, p. 729–746, 2017.

BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, M. Generalized autoregressive moving average models. **Journal of the American Statistical Association**, v. 98, n. 461, p. 214–223, 2003.

BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, M. D. Fitting non-Gaussian time series models. In: SPRINGER. **COMPSTAT**. New York, 1998. p. 191–196.

BILLAH, B.; HYNDMAN, R. J.; KOEHLER, A. B. Empirical information criteria for time series forecasting. **Journal of Statistical Computation and Simulation**, v. 75, n. 10, p. 831–840, 2005.

BLASKOWITZ, O.; HERWARTZ, H. Adaptive forecasting of the euribor swap term structure. **Journal of Forecasting**, v. 28, n. 7, p. 575–594, 2009.

BLASKOWITZ, O.; HERWARTZ, H. On economic evaluation of directional forecasts. **International of Journal of Forecasting**, v. 27, n. 4, p. 1058–1065, 2011.

BLASKOWITZ, O.; HERWARTZ, H. Testing the value of directional forecasts in the presence of serial correlation. **International of Journal of Forecasting**, v. 30, n. 1, p. 30–42, 2014.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time Series Analysis: Forecasting and Control**. 5th. ed. Hoboken: Wiley, 2015.

BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 65, n. 332, p. 1509–1526, 1970.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: Understanding AIC and BIC in model selection. **Sociological Methods Research**, v. 33, n. 2, p. 261–304, 2004.

CASARIN, R.; VALLE, L. D.; LEISEN, F. Bayesian model selection for beta autoregressive processes. **Bayesian Analysis**, v. 7, n. 2, p. 385–410, 2012.

CAVANAUGH, J. E.; SHUMWAY, R. H. A bootstrap variant of AIC for state-space model selection. **Statistica Sinica**, JSTOR, v. 7, n. 2, p. 473–496, 1997.

CHIOGNA, M.; GAETAN, C. Mining epidemiological time series: An approach based on dynamic regression. **Statistical Modelling**, v. 5, p. 309–325, 2005.

CHOI, B. S. **ARMA Model Identification**. New York: Springer-Verlag, 1992. ISBN 3540977953.

CRIBARI-NETO, F.; FRERY, A. C.; SILVA, M. F. Improved estimation of clutter properties in speckled imagery. **Computational Statistics & Data Analysis**, v. 40, n. 4, p. 801–824, 2002.

CRIBARI-NETO, F.; SCHER, V. T.; BAYER, F. M. Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. **International Journal of Forecasting**, 2022. - forthcoming.

CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. **Journal of Statistical Software**, v. 34, p. 1–24, 2010.

DAVIES, N.; TRIGGS, C. M.; NEWBOLD, P. Significance levels of the Box-Pierce portmanteau statistic in finite samples. **Biometrika**, Biometrika Trust, v. 64, n. 3, p. 517–522, 1977.

DOUMA, J. C.; WEEDON, J. T. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. **Methods in Ecology and Evolution**, v. 10, n. 9, p. 1412–1430, 2019.

DUFOUR, J.-M.; ROY, R. Generalized portmanteau statistics and tests of randomness. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 15, n. 10, p. 2953–2972, 1986.

EFRON, B.; TIBSHIRANI, R. J. Bootstrap methods for standard errors, condifence intrevals and other measures of statistical accuracy. **Statistical Science**, v. 1, n. 1, p. 54–77, 1986.

FAHRMEIR, L.; KAUFMANN, H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. **The Annals of Statistics**, v. 13, n. 1, p. 342–368, 1985.

FENGA, L. Bootstrap order determination for ARMA models: A comparison between different model selection criteria. **Journal of Probability and Statistics**, v. 2017, p. Article ID 1235979, 2017.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799–815, 2004.

FISHER, R. A. On the "probable error" of a coefficient of correlation deduced from a small sample. **Metron**, v. 1, p. 3–32, 1921.

FOKIANOS, K.; KEDEM, B. Partial likelihood inference for time series following generalized linear models. **Journal of Time Series Analysis**, v. 25, n. 2, p. 173–197, 2004.

HANNAN, E. J.; QUINN, B. G. The determination of the order of an autoregression. **Journal of the Royal Statistical Society B**, Wiley Online Library, v. 41, n. 2, p. 190–195, 1979.

HOTELLING, H. New light on the correlation coefficient and its transforms. **Journal of the Royal Statistical Society B**, JSTOR, v. 15, n. 2, p. 193–232, 1953.

HURVICH, C. M.; TSAI, C. L. Regression and time series model selection in small samples. **Biometrika**, v. 76, n. 2, p. 297–307, 1989.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. **Journal of Statistical Software**, v. 26, n. 3, p. 1–22, 2008.

JENKINS, G. M. An angular transformation for the serial correlation coefficient. **Biometrika**, JSTOR, v. 41, n. 1/2, p. 261–265, 1954.

KENDALL, M.; STUART, A. **The Advanced Theory of Statistics**. 4th. ed. London: Griffin, 1977. v. 1.

KWAN, A. C. C.; SIM, A.-B. On the finite-sample distribution of modified portmanteau tests for randomness of a Gaussian time series. **Biometrika**, Biometrika Trust, v. 83, n. 4, p. 938–943, 1996.

KWAN, A. C. C.; SIM, A.-B. Portmanteau tests of randomness and Jenkins' variance-stabilizing transformation. **Economics Letters**, Elsevier, v. 50, n. 1, p. 41–49, 1996.

KWAN, A. C. C.; SIM, A.-B.; WU, Y. A comparative study of the finite-sample performance of some portmanteau tests for randomness of a time series. **Computational Statistics & Data Analysis**, Elsevier, v. 48, n. 2, p. 391–413, 2005.

LI, W. K. Time series models based on generalized linear models: Some further results. **Biometrics**, v. 50, n. 2, p. 506–511, 1994.

LIN, J.-W.; MCLEOD, A. I. Improved Peña-Rodriguez portmanteau test. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 3, p. 1731–1738, 2006.

LINKA, A. On transformations of multivariate ARMA processes. **Kybernetika**, v. 24, n. 2, p. 122–129, 1988.

LJUNG, G. M. Diagnostic testing of univariate time series models. **Journal of the American Statistical Association**, v. 73, n. 3, p. 725–730, 1986.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, Biometrika Trust, v. 65, n. 2, p. 297–303, 1978.

MATSUMOTO, M.; NISHIMURA, T. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. **ACM Transactions on Modeling and Computer Simulation (TOMACS)**, ACM, v. 8, n. 1, p. 3–30, 1998.

MCLEOD, A. I. On the distribution of residual autocorrelations in Box-Jenkins models. **Journal of the Royal Statistical Society B**, JSTOR, v. 40, n. 3, p. 296–302, 1978.

MCLEOD, A. I.; JIMENEZ, C. Nonnegative definiteness of the sample autocovariance function. **The American Statistician**, Taylor & Francis Group, v. 38, n. 4, p. 297–298, 1984.

MCQUARRIE, A. D. A small-sample correction for the Schwarz SIC model selection criterion. **Statistics & Probability Letters**, Elsevier, v. 44, n. 1, p. 79–86, 1999.

MCQUARRIE, A. D. R.; TSAI, C.-L. **Regression and Time Series Model Selection**. London: World Scientific, 1998.

MELCHIOR, C.; ZANINI, R. R.; GUERRA, R. R.; ROCKENBACH, D. A. Forecasting Brazilian mortality rates due to occupational accidents using autoregressive moving average approaches. **International Journal of Forecasting**, v. 37, n. 2, p. 825–837, 2021.

MONTI, A. C. A proposal for a residual autocorrelation test in linear models. **Biometrika**, Biometrika Trust, v. 81, n. 4, p. 776–780, 1994.

NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization**. 2nd. ed. New York: Springer, 2006.

ONS. **Operador Nacional do Sistema Elétrico – Energia Armazenada**. 2020. <http://www.ons.org.br/historico/energia_armazenada.aspx>.

OSPINA, R.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved point and interval estimation for a beta regression model. **Computational Statistics & Data Analysis**, v. 51, n. 2, p. 960–981, 2006. Erratum: 55, 2011, 2445.

OZAKI, T. On the order determination of ARIMA models. **Applied Statistics**, v. 26, n. 3, p. 290–301, 1978.

PALM, B.; BAYER, F. M. Bootstrap-based inferential improvements in beta autoregressive moving average model. **Communications in Statistics - Simulation and Computation**, v. 47, n. 4, p. 977–996, 2018.

PEñA, D.; RODRIGUEZ, J. A powerful portmanteau test of lack of fit for time series. **Journal of the American Statistical Association**, American Statistical Association, v. 97, n. 458, p. 601–610, 2002.

PIERCE, D. A. Residual correlations and diagnostic checking in dynamic-disturbance time series models. **Journal of the American Statistical Association**, v. 67, n. 339, p. 636–640, 1972.

POLER, R.; MULA, J. Forecasting model selection through out-of-sample rolling horizon weighted error. **Expert Systems with Applications**, v. 38, n. 12, p. 14778–14785, 2011.

PUMI, G.; VALK, M.; BISOGNIN, C.; BAYER, F. M.; PRASS, T. S. Beta autoregressive fractionally integrated models. **Journal of Statistical Planning and Inference**, v. 200, p. 196–212, 2019.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **TEST**, Springer, v. 18, n. 3, p. 529–545, 2009.

ROCHA, A. V.; CRIBARI-NETO, F. Erratum to: Beta autoregressive moving average models. **TEST**, Springer, v. 26, n. 2, p. 451–459, 2017.

SCHER, V. T. **Portmanteau testing inference in beta autoregressive moving average models**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2017.

SCHER, V. T.; CRIBARI-NETO, F.; PUMI, G.; BAYER, F. M. Goodness-of-fit tests for βARMA hydrological time series modeling. **Environmetrics**, Wiley Online Library, v. 31, n. 3, p. e2607, 2020.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SHAO, J. Asymptotic theory in generalized linear models with nuisance scale parameters. **Probability Theory and Related Fields**, v. 91, n. 1, p. 25–41, 1992.

SHIBATA, R. Selection of the order of an autoregressive model by Akaike's information criterion. **Biometrika**, v. 63, n. 1, p. 117–126, 1976.

SHIBATA, R. Bootstrap estimate of Kullback-Leibler information for model selection. **Statistica Sinica**, JSTOR, v. 7, n. 2, p. 375–394, 1997.

SILVA, C. Q. da; MIGON, H. S.; CORREIA, L. T. Dynamic Bayesian beta models. **Computational Statistics & Data analysis**, v. 55, n. 6, p. 2074–2089, 2011.

SUGIURA, N. Further analysts of the data by Akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 7, n. 1, p. 13–26, 1978.

TEAM, R. C. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.

WU, T.-J.; SEPULVEDA, A. The weighted average information criterion for order selection in time series and regression models. **Statistics & Probability Letters**, Elsevier, v. 39, n. 1, p. 1–10, 1998.

ZEGER, S. L.; QAQISH, B. Markov regression models for time series: A quasi-likelihood aapproach. **Biometrics**, v. 44, n. 4, p. 1019–1032, 1988.

ZHENG, T.; CHEN, R. Dirichlet ARMA models for compositional time series. **Journal of Multivariate Analysis**, v. 158, p. 31–46, 2017.

APPENDIX A - PROOF OF THEOREMS 1.4.2.1 AND 1.4.2.2 AND OF COROLLARY 1.4.2.2

Demonstração. (Theorem 1.4.2.1). Under the null hypothesis, the same argument applied in the proof of Lemma 1 in Monti (1994) holds so that

$$\hat{\pi}_k = \hat{\rho}_k + O_p(n^{-1}), \tag{3.3}$$

which implies that $\hat{\pi}_k = \hat{\rho}_k + o_p(n^{-\delta})$ for $1/2 < \delta < 1$. We start by showing that $\hat{z}_{1k} = z_{1k} + o_p(n^{-1/2})$. Under $\mathcal{H}_0$, there exists $n_0 > 0$ such that, for $n > n_0$, $\hat{\rho}_k$ is contained in a compact subinterval of $[-1,1]$, say $U = [-M, M]$, for $0 < M < 1$, with probability tending to one. For $n > n_0$ and for any $a \in \mathbb{R}$ with $a \neq 0$,

$$\exp\left\{\frac{\sqrt{n}}{2}\Big[\log(1 \pm a\hat{\pi}_k) - \log(1 \pm a\hat{\rho}_k)\Big]\right\} = \left(\frac{1 \pm a\hat{\pi}_k}{1 \pm a\hat{\rho}_k}\right)^{\frac{\sqrt{n}}{2}} = \left(1 + \frac{o_p(n^{-\delta})}{1 \pm a\hat{\rho}_k}\right)^{\frac{\sqrt{n}}{2}} \xrightarrow{p} 1,$$

as n tends to infinity, since $\delta > 1/2$. Upon applying the logarithm function to the above expression, the result follows from the Continuous Mapping Theorem, with $a = 1$. Using Slutsky's Theorem we conclude that Kendall and Stuart (1977, p. 419)

$$\sqrt{(n-k-3)}\hat{z}_{1k} = \sqrt{(n-k-3)}\Big(z_{1k} + o_p(n^{-1/2})\Big) \xrightarrow{d} \mathcal{N}(0,1).$$

Hence, Q_1 and Q_{KW1} are asymptotically equivalent and $Q_1 \xrightarrow{d} \chi^2_m$.

We shall show that $\hat{z}_{4k} = z_{4k} + o_p(n^{-\delta})$. By using the integral representation $\sin^{-1}(x) = \int_0^x (1-z^2)^{-\frac{1}{2}} dz$, for any $a \neq 0$ we obtain

$$\begin{aligned}\sin^{-1}(a\hat{\pi}_k) &= \sin^{-1}\left(a\hat{\rho}_k + o_p(n^{-\delta})\right) = \int_0^{a\hat{\rho}_k} \frac{1}{\sqrt{1-z^2}} dz + \int_{a\hat{\rho}_k}^{a\hat{\rho}_k + o_p(n^{-\delta})} \frac{1}{\sqrt{1-z^2}} dz \\ &= \sin^{-1}(a\hat{\rho}_k) + R_n.\end{aligned}$$

For n sufficiently large, $|\hat{\rho}_k + o_p(n^{-\delta})| < M + \varepsilon < 1$, for some $0 < \varepsilon < 1 - M$, so that

$$0 \leq R_n = \int_{a\hat{\rho}_k}^{a\hat{\rho}_k + o_p(n^{-\delta})} \frac{1}{\sqrt{1-z^2}} dz \leq \frac{1}{\sqrt{1-(M+\varepsilon)^2}} \Big| o_p(n^{-\delta}) \Big| = o_p(n^{-\delta}),$$

and the result follows with $a = 1$. Using Slutsky's Theorem, it follows that $\hat{z}_{4k}$ and z_{4k} are asymptotic equivalent and so are Q_4 and Q_{KW4}. Hence, by Kwan and Sim (1996b), $Q_4 \xrightarrow{d} \chi^2_m$ follows. □

Demonstração. (Theorem 1.4.2.2). Let us assume for the moment that no covariates are present in the model. We start by showing that $Q_1 \xrightarrow{d} \chi^2_{m-p-q}$. Let $\boldsymbol{\rho}_{(k)} = (\rho_1, \ldots, \rho_k)^\top$ and

$\hat{\boldsymbol{\rho}}_{(k)} = (\hat{\rho}_1, \ldots, \hat{\rho}_k)^\top$ be the true and sample autocorrelation vector up to lag k, respectively. Since $g(\mu_t) = g(y_t) - r_t$, then

$$g(y_t) = \alpha + \sum_{i=1}^{p} \varphi_i g(y_{t-i}) + \sum_{j=0}^{q} \theta_j r_{t-j} \quad \Longleftrightarrow \quad \Phi(L) g(y_t) = \alpha + \Theta(L) r_t, \tag{3.4}$$

where L is the lag (backshift) operator and $\theta_0 = 1$. By A1 and A3, (3.4) is invertible at the true value of $\boldsymbol{\theta}$ as well as in U. Hence, for sufficiently large n, (3.4) is invertible at the CMLE estimate $\hat{\boldsymbol{\theta}}$. In view of this, following McLeod (1978), under Assumptions A1 through A3 there exists an idempotent matrix Q of rank $p+q$ such that $\hat{\boldsymbol{\rho}}_{(k)} = (I_k - Q)\boldsymbol{\rho}_{(k)} + O_p(n^{-1})$, where I_k is the k-dimensional identity matrix. Under the null hypothesis, (3.3) is still valid, so that by letting $\hat{\boldsymbol{\pi}}_{(k)} = (\hat{\pi}_1, \ldots, \hat{\pi}_k)^\top$ and following the same steps as in the proof of Theorem 1.4.2.1, we establish that

$$\hat{\boldsymbol{\pi}}_{(k)} = (I_k - Q)\hat{\boldsymbol{\rho}}_{(k)} + o_p(n^{-\delta}).$$

Let $\hat{\boldsymbol{z}}_{1(m)} = (\hat{z}_{11}, \ldots, \hat{z}_{1m})^\top$ and $\boldsymbol{z}_{1(m)} = (z_{11}, \ldots, z_{1m})^\top$. Using the argumentation employed in the proof of Theorem 1.4.2.1 (with a being the appropriate entry of the matrix $I_k - Q$), we conclude that

$$\sqrt{(n-k-3)}\hat{\boldsymbol{z}}_{1(m)} = \sqrt{(n-k-3)}\Big(\boldsymbol{z}_{1(m)} + o_p(n^{-1/2})\Big) \xrightarrow{d} \mathcal{N}_m\Big(\mathbf{0}, I_m - Q\Big),$$

where $\mathcal{N}_m(\mathbf{0}, \Sigma)$ denotes the m-variate normal distribution with mean vector $\mathbf{0} = (0, \ldots, 0)^\top \in \mathbb{R}^m$ and covariance matrix Σ. Finally, since $I_m - Q$ is an idempotent matrix of rank $p+q$, it follows that

$$Q_1 = (n-k-3)\hat{\boldsymbol{z}}_{1(m)}^\top \hat{\boldsymbol{z}}_{1(m)} \xrightarrow{d} \chi^2_{m-p-q}.$$

Next, we shall prove that $Q_4 \xrightarrow{d} \chi^2_{m-p-q}$. Again, under $\mathcal{H}_0$, (3.3) holds and under Assumptions A1 through A3, the representation in (3.4) is valid. Let $\hat{\boldsymbol{z}}_{4(m)} = (\hat{z}_{41}, \ldots, \hat{z}_{4m})^\top$ and $\boldsymbol{z}_{4(m)} = (z_{41}, \ldots, z_{4m})^\top$. By making use of the argumentation employed in the proof of Theorem 1.4.2.1 with the appropriate coordinate a, we conclude that

$$\left(\frac{n-k}{n-k-1}\right)\hat{\boldsymbol{z}}_{4(m)} \xrightarrow{d} \mathcal{N}_m\Big(\mathbf{0}, I_m - Q\Big),$$

and the result follows. If covariates are present in the model, since $r_t = g(y_t) - g(\mu_t)$, (1.2) implies that $S_t = g(y_t) - \boldsymbol{x}_t^\top \boldsymbol{\beta}$ satisfies the ARMA(p,q) difference equations:

$$S_t = \alpha + \sum_{i=1}^{p} \varphi_i S_{t-i} + \sum_{j=0}^{q} \theta_j r_{t-j}, \tag{3.5}$$

with $\theta_0 = 1$. The result follows from the arguments outlined in Section 3 of Pierce (1972), in light of Box and Pierce (1970) and McLeod (1978). $\square$

Demonstração. (Corollary 1.4.2.2). The asymptotic null distribution of Q_{LB} follows from using the reasoning presented in Section 3 in Ljung and Box (1978). Finally, the asymptotic null distribution of Q_M follows from the arguments used in Lemma 1 of Monti (1994). □