Beta Autoregressive Moving Average Model with the Aranda-Ordaz Link Function

Carlos E. F. Manchini [1] , Diego Ramos Canterle [2,3] , Guilherme Pumi [3,4] and Fábio M. Bayer [1,4,*]

[1] Departamento de Estatística and LACESM, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil; carlos.manchini@ufpe.br
[2] Instituto de Matemática e Estatística, Universidade de São Paulo, Sao Paulo 05508-220, Brazil; canterle@ime.usp.br
[3] Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre 90035-003, Brazil; guilherme.pumi@ufrgs.br
[4] Programa de Pós-Graduação em Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre 90035-003, Brazil
* Correspondence: bayer@ufsm.br

Abstract: In this work, we introduce an extension of the so-called beta autoregressive moving average (βARMA) models. βARMA models consider a linear dynamic structure for the conditional mean of a beta distributed variable. The conditional mean is connected to the linear predictor via a suitable link function. We propose modeling the relationship between the conditional mean and the linear predictor by means of the asymmetric Aranda-Ordaz parametric link function. The link function contains a parameter estimated along with the other parameters via partial maximum likelihood. We derive the partial score vector and Fisher's information matrix and consider hypothesis testing, diagnostic analysis, and forecasting for the proposed model. The finite sample performance of the partial maximum likelihood estimation is studied through a Monte Carlo simulation study. An application to the proportion of stocked hydroelectric energy in the south of Brazil is presented.

1. Introduction

Gaussianity is by far the most commonly used hypothesis in statistics. It is easy to find in the literature applications of Gaussian time series in contexts where it is neither natural nor adequate to suppose normality of the underlying data distribution. Simple examples are strictly positive data such as prices, counting phenomena or double bounded data, such as rates and proportions whose support is $(0, 1)$. In these situations, normality is obviously not an adequate hypothesis to assume. The consequences of using a Gaussian time series model were it is not reasonable may be grave, especially when the focus lies on forecasting. In these cases, it is a common problem to obtain predicted values that lie outside the natural bounds of the data.

To overcome these issues, non-Gaussian time series models have been extensively explored in the literature. For instance, autoregressive models for integer valued time series were introduced in [1]. A broad class of dynamic models for non-Gaussian time series based on generalized linear models (GLM) [2] was considered in [3], which called them generalized autoregressive moving average (GARMA) models. Traditional GLM methods can serve as an inspiration for time series models, but there are important distinctions and technicalities to keep in mind [4]. Other studies in this direction can be found in [5–8].

Modeling rates and proportions observed over time is a common problem in many areas of application. By nature, such time series are limited to the interval $(0, 1)$ and, hence,

Gaussianity is an assumption that should be avoided [9]. In this direction, the class of beta autoregressive moving average (βARMA) models, proposed by [10], introduces a GLM-like dynamic model for time series restricted to $(0, 1)$. The βARMA model assume that, conditionally to its past, the variable of interest follows a beta distribution while the conditional mean is modeled through a dynamical time dependent linear structure accommodating an ARMA-like term and a linear combination of exogenous covariates. The conditional mean is connected to the linear predictor through a suitable fixed link function.

The beta distribution is well known for its flexibility, being able to model asymmetric behaviors such as bathtub and J-shaped and inverted J-shaped densities, among other. For this reason, the literature has seen a growing interest in beta-based models in the last decade and several improvements and generalizations of the βARMA model have been proposed. For instance, some generalizations propose the use of different specifications for the systematic component. Ref. [11] proposed the class of βSARMA models which introduces a seasonal ARMA structure in the systematic components. Ref. [12] proposed the class of βARFIMA models by considering a long range dependent specification for the systematic component. Some goodness-of-fit tests for βARMA are proposed in [13], while prediction intervals and model selection criteria are recently explored in [14,15], respectively. In [16], the inflated βARMA model was introduced for modeling time series data that assume values in the intervals $(0, 1]$, $[0, 1)$ or $[0, 1]$.

A common feature in the aforementioned models, i.e., the connection between the conditional mean with the linear predictor is made by a suitable link function, ensures that the modeled conditional mean values do not fall outside its natural bounds. Typical choices for responses taking values in $(0, 1)$ are the logit and cloglog fixed links. However, misspecification of the link function may cause distortions in parameter estimation [2]. A simple solution to this problem is to apply a parametric link function. This adds flexibility to the model and improves the finite sample performance of maximum likelihood estimation in the context of GLM [17], compared to the canonical ones. In the literature, we can find some models that apply parametric link functions in the context of regression models. For instance, for binary response problems, [18] proposed the use a modified Box-Cox transformation as link function, while [19] introduced a modified two parameter link that allows free manipulation of both tails of the link. More flexible approaches, which treat the entire link function to be estimated from the data, were considered in [20–22]. In the context of beta regression with variable dispersion, [23] proposed the use of parametric link function for the specification of both, mean and dispersion. Recently, ref. [24] proposed the use of the Aranda-Ordaz parametric link function [25] in the context of Kumaraswamy regression. Other applications of parametric link functions in regression models can be found in [26–29].

Despite the relative growth of the literature on parametric link function in regression models, to the best of our knowledge, there are no time series models considering parametric link functions. In this direction, this works generalizes the βARMA model by introducing a parametric link function in the conditional mean specification. Given that a time series following a βARMA lies on $(0, 1)$, a suitable and widely known link (as discussed in [23]) is the Aranda-Ordaz asymmetric link function, introduced in [25]. The Aranda-Ordaz link function depends on a parameter λ which must be known, or, ideally, estimated from the data. To do that, we propose a partial maximum likelihood approach to estimate λ along with all other parameters in the model.

Different from time series analysis, in regression analysis, the interpretability of the fitted model is of great interest. Some commonly applied link functions, such as the logit and logarithm, allow for a simple interpretation of model parameters, which is no longer the case when we consider a parametric specification for the link function. However, in time series models, prediction is usually considered more important than model interpretability, which favors the use of a parametric link function, since it usually allows for a better model fitting and, often, a superior forecasting performance. In this sense, the proposed βARMA with Aranda-Ordaz link function provides two advantages over the standard βARMA

model: (i) better model fitting due to more flexibility when considering the relationship between the mean and the linear predictor and (ii) robustness against inferential distortions usually attributed to link misspecification.

The paper is organized as follows. In the next section, we introduce the βARMA model with the Aranda-Ordaz link function. In Section 3, we propose a partial maximum likelihood approach to estimate the model parameters and discuss some of its properties. In Section 4, we present some diagnostic and goodness-of-fit tools and discuss forecasting in the context of the proposed model. A Monte Carlo simulation study is presented in Section 5, while a real data application is presented in Section 6. Our conclusions are presented in Section 7 and some technical details are discussed in the Appendices A and B.

2. Proposed Model

Let $\{y_t\}_{t \in \mathbb{Z}}$ denote a time series of interest and let $\{x_t\}_{t \geq 1}$ denote a set of k-dimensional exogenous covariates, possibly time dependent and random. Let $\mathscr{F}_t$ denote the σ-field representing the history of the model known to the researcher up to time t, that is, the sigma-field generated by $(y_t, x'_t, y_{t-1}, x'_{t-1}, y_{t-2}, x'_{t-2}, \dots)$. We assume that the conditional distribution of y_t given $\mathscr{F}_{t-1}$ is $\text{Beta}(\mu_t, \phi)$ parameterized as in [9], with density:

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

where $\Gamma(\cdot)$ is the Gamma function, $0 < \mu_t < 1$ and $\phi > 0$. It is easy to show that:

$$\mathbb{E}(y_t|\mathscr{F}_{t-1}) = \mu_t \quad \text{and} \quad \text{Var}(y_t|\mathscr{F}_{t-1}) = \frac{\mu_t(1-\mu_t)}{1+\phi}.$$

Observe that the parameter ϕ acts as a precision parameter, since the higher the ϕ, the smaller the variance of y_t, for a fixed μ_t. Note that the variance of y_t changes for each t, so the model is naturally heteroscedastic. Although, in principle, it could be possible to consider a variable dispersion parameter, as in [23]; however, the majority of works in the literature assume a constant precision/dispersion parameter ([3,9–11,15,30], to name just a few) so, for simplicity, we shall do the same.

Let $g(\cdot, \lambda) : (0,1) \to \mathbb{R}$ be a twice continuously differentiable one to one link function, possibly depending on a parameter λ. Consider the following specification for the model's systematic component:

$$\eta_t := g(\mu_t, \lambda) = \alpha + x_t^\top \beta + \sum_{i=1}^p \varphi_i \big[g(y_{t-i}, \lambda) - x_{t-i}^\top \beta \big] + \sum_{j=1}^q \theta_j r_{t-j}, \qquad (2)$$

where $\alpha \in \mathbb{R}$ is an intercept, $\beta = (\beta_1, \dots, \beta_k)^\top \in \mathbb{R}^k$ is a k-dimensional vector of parameter associated with the covariates, and $\varphi := (\varphi_1, \dots, \varphi_p)^\top$ and $\theta := (\theta_1, \dots, \theta_q)^\top$ are p and q-dimensional vectors of parameters associated with the autoregressive and moving average components, respectively. The error term is defined as $r_t = g(y_t, \lambda) - g(\mu_t, \lambda)$. Observe that if we substitute $g(y, \lambda)$ for a fixed (non-parametric) link function $g(y)$, such as logit or probit links, then we obtain the βARMA model of [10].

In (2), when g indeed depends on λ we have a parametric link function. The choice of the parametric link is an important one and due to its flexibility, we shall apply the asymmetric Aranda-Ordaz family of link functions in (2), which has the form:

$$\eta_t = g(\mu_t, \lambda) = \log\left(\frac{(1-\mu_t)^{-\lambda} - 1}{\lambda} \right), \qquad (3)$$

for $\lambda > 0$. Observe that $\mu_t = g^{-1}(\eta_t, \lambda) = 1 - \big[1 + \lambda \exp(\eta_t)\big]^{-\frac{1}{\lambda}}$ and that g is infinitely differentiable away from zero on λ and μ. Figure 1 shows the Aranda-Ordaz link function for several values of λ, highlighting its asymmetric behavior, which allows an asymmetric

relationship between μ_t and η_t. We note that it is an increasing and monotonic function, where positive (or negative) changes in the linear predictor imply a positive (or negative) effect on the response mean. The traditional logit is a particular case obtained when $\lambda = 1$ and the cloglog is obtained as a limiting case for $\lambda \to 0$.
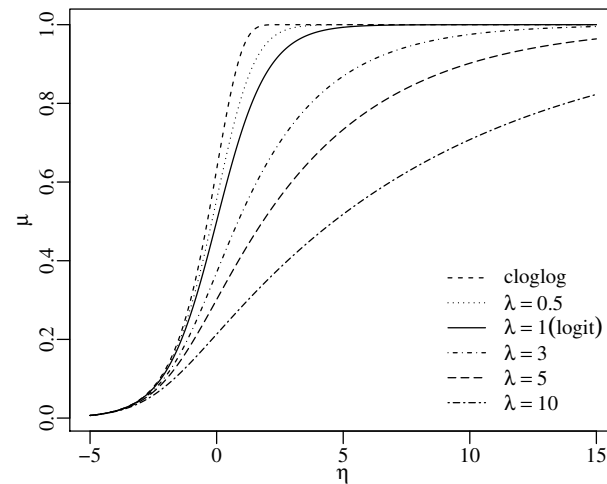

Figure 1. The Aranda-Ordaz link function for different values of λ.

The proposed βARMA model with Aranda-Ordaz link function, hereafter referred to as βARMA$_\lambda$, is defined by (1), (2), and (3). The Aranda-Ordaz parameter influences the conditional mean μ_t, which affects the response Y_t through $Y_t | \mathscr{F}_{t-1} \sim \mathrm{Beta}(\mu_t, \phi)$. Figure 1 illustrates how the parameter λ affects μ_t as a function of η_t. For $0 < \lambda < 1$, the effect on μ_t is mild compared to cloclog and logit, primarily impacting η_t values between 0 and 4 by slightly accelerating or decelerating the increase of μ_t relative to η_t. For $\lambda > 1$, the impact of λ is more pronounced, with larger λ values dampening the speed at which changes in η_t affect μ_t. Thus, for $\lambda > 1$, μ_t becomes less sensitive to variations in η_t, smoothing Y_t's conditional mean compared to the logit and cloglog functions.

3. Partial Likelihood Inference

Let $\{(x_t, y_t)\}_{t=1}^n$ be a sample from the proposed βARMA$_\lambda$ model and let $\boldsymbol{\omega} := (\phi, \lambda, \alpha, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$ be the $(p + q + k + 3)$-dimensional vector of unknown parameters in the model, and $\Omega = (0, \infty)^2 \times \mathbb{R}^{1+k+p+q}$ is the parametric space, such that $\boldsymbol{\omega} \in \Omega$. In this section, we derive a partial maximum likelihood estimation (PMLE) approach to estimate $\boldsymbol{\omega}$. Given $\mathscr{F}_{t-1}$, the partial log-likelihood function is given by:

$$\ell(\boldsymbol{\omega}) := \sum_{t=m+1}^n \log f(y_t | \mathscr{F}_{t-1}) = \sum_{t=m+1}^n \ell_t(\mu_t, \phi), \tag{4}$$

where $m = \max\{p, q\}$ and

$$\ell_t(\mu_t, \phi) := \log\big(\Gamma(\phi)\big) - \log\big(\Gamma(\mu_t \phi)\big) - \log\big(\Gamma\big[(1 - \mu_t)\phi\big]\big)$$
$$+ (\mu_t \phi - 1)\log(y_t) + \big[(1 - \mu_t)\phi - 1\big]\log(1 - y_t).$$

The estimator $\widehat{\boldsymbol{\omega}}$ is defined as:

$$\widehat{\boldsymbol{\omega}} := \arg\sup_{\boldsymbol{\omega} \in \Omega} \big\{\ell(\boldsymbol{\omega})\big\}. \tag{5}$$

The solution of (5) is obtained by solving $U(\boldsymbol{\omega}) = \mathbf{0}$, where $U(\boldsymbol{\omega})$ is the partial score vector and $\mathbf{0} := (0, \ldots, 0)^\top \in \mathbb{R}^{p+q+k+3}$. Closed-form expressions of the partial score vector are presented in Appendix A.

The non-linear system $U(\boldsymbol{\omega}) = \mathbf{0}$ has no closed-form solution. To obtain approximate solutions, it is necessary to numerically maximize the partial log-likelihood function given in (4). In this work, we consider the so-called BFGS method [31] with analytical first derivatives to do so. The procedure requires initialization. We initialize $\lambda = 1$ (logit particular case) and $\boldsymbol{\theta} = \mathbf{0}$ in all cases. Parameters α, $\boldsymbol{\beta}$, and $\boldsymbol{\varphi}$ are initialized as the ordinary least squares estimates of the regression problem:

$$g(y_t, 1) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i g(y_{t-i}, 1) + \varepsilon_t,$$

where ε_t denotes an error term. Let $(\alpha^0, \boldsymbol{\beta}^0, \boldsymbol{\varphi}^0)^\top$ denote the initial value of $(\alpha, \boldsymbol{\beta}, \boldsymbol{\varphi})^\top$, the starting value of ϕ is given by:

$$\phi^0 := \frac{1}{n} \sum_{t=1}^{n} \frac{\mu_t^0(1 - \mu_t^0)}{\mathfrak{s}^0} - 1,$$

where $\mu_t^0 := g^{-1}\big(\alpha^0 + \boldsymbol{x}_t^\top \boldsymbol{\beta}^0 + \sum_{i=1}^{p} \varphi_i^0 [g(y_{t-i}, 1) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta}^0], 1\big)$ and $\mathfrak{s}^0 := \dfrac{1}{n-1} \sum_{t=1}^{n} (y_t - \mu_t^0)^2$.

In addition to point inference, it is also interesting to build confidence intervals and carry out hypothesis tests. In this sense, we need the asymptotic distribution of the partial maximum likelihood estimator. The seminal work of [32] established the asymptotic theory of the maximum likelihood in the context of traditional GLM under mild conditions. These results were generalized for the case of GLM with parametric link functions in [33]. For GARMA-like models, a rigorous asymptotic theory was established in [34,35]. Ref. [12] presented the asymptotic theory for the PMLE in the context of βARFIMA models. Recently, Ref. [24] established the asymptotic theory of the MLE for GLM-like models with the Aranda-Ordaz parametric link based on the Kumaraswamy distribution, which is not a member of the exponential family.

In the context of βARMA$_\lambda$ model, specification (1) and the fact that the information matrix is not block diagonal imply that the parameter estimation must be jointly performed via the log-likelihood function. To derive the asymptotic theory for the PMLE in the context of the present work, a similar argument as in [33,34] can be applied, with a few modifications. Under suitable conditions, closely related to the ones presented in [34], it can be shown that the partial maximum likelihood estimator $\widehat{\boldsymbol{\omega}}$ is consistent and:

$$\sqrt{n}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, K^{-1}(\boldsymbol{\omega})),$$

as n tends to infinity, where $\mathcal{N}(\mathbf{0}, \Sigma)$ denotes the $(p + q + k + 3)$-variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix Σ. The matrix $K(\boldsymbol{\omega})$ is a positive definite and invertible matrix, the analogous to the information matrix per observation in the context of i.i.d. samples. Under suitable conditions, $K_n(\boldsymbol{\omega})/n \longrightarrow K(\boldsymbol{\omega})$ in probability, where $K_n(\boldsymbol{\omega})$ is the cumulative conditional information matrix presented in the Appendix B.

Let ω_r denote the r-th component of $\boldsymbol{\omega}$, and with the previous results, it is possible to obtain approximate confidence interval to ω_r and standard Z statistics to test hypothesis like $\mathcal{H}_0 : \omega_r = \omega_r^0$ vs. $\mathcal{H}_1 : \omega_r \neq \omega_r^0$ [36]. Versions of other commonly applied test statistics, such as the Wald [37], likelihood ratio [38], Rao's score [39], and the gradient [40] statistic, can be similarly defined. Their asymptotic distributions will be χ^2 with the corresponding degrees of freedom imposed by the restrictions under $\mathcal{H}_0$. More general forms of these hypothesis tests can be performed similarly to traditional regression models.

4. Diagnostics and Forecasting

In this section, we consider some diagnostic tools useful in determining whether a fitted model succeeded in capturing the data dynamics. Also, we detail a method to produce forecasts based on a fitted model. Besides the joint significance of the parameters

obtained using the tests presented in the previous section, we can also discuss residual analysis. A priori, there is no imposed distributional structure for the error term r_t in (2), but it is quite common to look at some goodness-of-fit statistics.

The so-called deviance statistics is commonly applied as a goodness-of-fit measure of a given model [4]. It can be shown that, if the model is correctly specified, then the deviance statistics is asymptotically $\chi^2_{n-(p+q+k+3)}$ distributed [3,4]. Model selection can be performed by using adapted versions of the AIC [41] and BIC [42]. A lower AIC and BIC values are associated with more suitable models. Residual analysis is a fundamental diagnostic tool in statistical modeling. There are several ways to define the residuals for the proposed model, such as standardized, quantile [43], and deviance residuals. As considered in [13], for the βARMA, and in [9], for the beta regression model, we suggest using the standardized ordinary residual.

Another useful tool in diagnostic analysis are Portmanteau tests. If the model is correctly specified, it is expected that the residuals behave like a white noise [4]. Portmanteau tests for βARMA models were rigorously studied by [13]. In this context, similar arguments to those in [13] can be applied to show that the Ljung–Box statistics to test the null hypothesis that the first s residual autocorrelations are zero will be asymptotically χ^2_{s-p-q} distributed, under mild assumptions.

Since the proposed model is an extension of the βARMA, h_0-step ahead forecasts can be obtained similarly. Let $\{(x_t, y_t)\}_{t=1}^n$ be a sample from the proposed βARMA$_\lambda$ model. Let $\widehat{\omega}$ be the PMLE based on the sample and $\widehat{\mu}_t$ be μ_t evaluated at $\widehat{\omega}$ and $\widehat{r}_t := \left[g(y_t, \widehat{\lambda}) - g(\widehat{\mu}_t, \widehat{\lambda}) \right] I(1 \leq t \leq n)$. For $h = 1, 2, \ldots, h_0$, h-step ahead forecasts are given by:

$$\widehat{\mu}_{n+h} := g^{-1}\left(\widehat{\alpha} + x_{n+h}^\top \widehat{\beta} + \sum_{i=1}^p \widehat{\varphi}_i \left(\left[g(y_{n+h-i}, \widehat{\lambda}) \right] - x_{n+h-i}^\top \widehat{\beta} \right) + \sum_{j=1}^q \widehat{\theta}_j \widehat{r}_{n+h-j}, \widehat{\lambda} \right),$$

where $\quad \left[g(y_t, \lambda) \right] := \begin{cases} g(\widehat{\mu}_t, \widehat{\lambda}), & \text{if } t > n, \\ g(y_t, \widehat{\lambda}), & \text{if } t \leq n. \end{cases}$

5. Numerical Experiments

In this section, we present a Monte Carlo simulation study to assess the finite sample properties of the PMLE for the proposed model parameters. We consider 10,000 simulated replications of a process $\{y_t\}_{t=1}^n$, for $n \in \{100, 300, 500, 1000\}$ following the βARMA$_\lambda(p, q)$. The following two scenarios were considered in the simulation:

1. βARMA$_\lambda(2, 2)$ with two covariates, where the parameters are set as $\alpha = 0.5$, $\beta = (0.3, -1)^\top$, $\varphi = (-0.5, 0.3)^\top$, $\theta = (0.4, -0.1)^\top$, $\phi \in \{20, 120\}$, and $\lambda = 1.5$;
2. βARMA$_\lambda(2, 1)$ with one covariate, where the parameters are $\alpha = -0.5$, $\beta_1 = -1$, $\varphi = (-0.4, 0.2)^\top$, $\theta_1 = 0.3$, $\phi \in \{20, 120\}$, and $\lambda = 0.5$.

In order to induce a deterministic seasonality into the model, we consider $x_t = (\sin(2\pi t/12), \cos(2\pi t/12))^\top$ in Scenario 1 and $x_t = \cos(2\pi t/12)$ in Scenario 2. All simulations were performed using R version 3.5.2 [44]. Tables 1 and 2 present the simulation results for Scenario 1, while Tables 3 and 4 for Scenario 2 by varying the precision parameter $\phi \in \{20, 120\}$. Presented are the mean estimate, relative bias (RB), standard error (SE), and mean square error (MSE).

As expected, in general, as n increases, the mean of the estimated values converges to the true value of the parameter and the RB, SE, and MSE decrease. Note that for $\phi = 20$, the relative bias and standard errors are higher than the ones when $\phi = 120$. This behavior is expected since decreasing precision implies an increase in the variability of y_t. For small sample sizes such as $n = 100$, there is considerable bias in all scenarios. Just as in the case of the beta regression [45] and βARFIMA [12], $\widehat{\phi}$ is slightly biased, especially in small samples.

Table 1. Simulation results for the proposed PMLE under Scenario 1 with $\phi = 120$.

	α 0.5	β_1 0.3	β_2 -1	φ_1 -0.5	φ_2 0.3	θ_1 0.4	θ_2 -0.1	λ 1.5	ϕ 120
				$n = 100$					
Mean	0.693	0.253	-1.050	-0.488	0.281	0.308	-0.036	1.820	132.753
RB (%)	38.631	-15.801	4.965	-2.328	-6.299	-22.944	-64.227	21.361	10.627
SE	2.660	0.319	0.424	0.603	0.689	0.616	0.529	3.887	20.858
MSE	7.115	0.104	0.182	0.364	0.474	0.387	0.284	15.212	597.688
				$n = 300$					
Mean	0.549	0.282	-1.022	-0.549	0.290	0.416	-0.064	1.545	124.046
RB (%)	9.772	-6.107	2.197	9.707	-3.466	3.876	-35.550	3.021	3.371
SE	0.394	0.038	0.164	0.407	0.391	0.427	0.328	0.500	10.420
MSE	0.158	0.002	0.027	0.168	0.153	0.183	0.109	0.253	124.945
				$n = 500$					
Mean	0.532	0.282	-1.016	-0.524	0.324	0.397	-0.093	1.532	122.285
RB (%)	6.429	-5.978	1.648	4.734	7.847	-0.840	-6.583	2.104	1.904
SE	0.321	0.027	0.124	0.327	0.315	0.342	0.265	0.379	7.810
MSE	0.104	0.001	0.016	0.108	0.100	0.117	0.070	0.144	66.225
				$n = 1000$					
Mean	0.539	0.283	-1.016	-0.523	0.330	0.400	-0.095	1.531	121.149
RB (%)	7.747	-5.601	1.619	4.640	9.954	0.099	-4.635	2.044	0.957
SE	0.226	0.019	0.086	0.225	0.217	0.233	0.181	0.260	5.395
MSE	0.053	0.001	0.008	0.051	0.048	0.054	0.033	0.069	30.430

Table 2. Simulation results for the proposed PMLE under Scenario 1 with $\phi = 20$.

	α 0.5	β_1 0.3	β_2 -1	φ_1 -0.5	φ_2 0.3	θ_1 0.4	θ_2 -0.1	λ 1.5	ϕ 20
				$n = 100$					
Mean	2.197	0.161	-1.295	-0.657	0.450	0.363	-0.004	3.855	22.678
RB (%)	339.318	-46.389	29.453	31.354	49.996	-9.144	-96.077	156.990	13.389
SE	7.309	0.960	1.079	0.952	1.372	0.634	0.538	10.157	3.650
MSE	56.303	0.942	1.251	0.932	1.904	0.404	0.298	108.704	20.491
				$n = 300$					
Mean	0.910	-0.595	0.302	0.448	-0.044	20.698	1.964	0.253	-1.098
RB (%)	81.931	19.038	0.792	12.098	-56.266	3.491	30.966	-15.671	9.753
SE	3.505	0.447	0.604	0.429	0.325	1.702	4.698	0.450	0.529
MSE	12.454	0.208	0.365	0.186	0.109	3.384	22.290	0.205	0.289
				$n = 500$					
Mean	0.624	0.282	-1.040	-0.566	0.292	0.439	-0.063	1.576	20.400
RB (%)	24.788	-6.031	4.038	13.232	-2.588	9.711	-37.279	5.073	2.002
SE	0.812	0.093	0.282	0.331	0.320	0.337	0.257	0.955	1.297
MSE	0.675	0.009	0.081	0.114	0.102	0.115	0.067	0.918	1.841
				$n = 1000$					
Mean	0.557	0.282	-1.015	-0.546	0.308	0.423	-0.078	1.511	20.191
RB (%)	11.455	-5.869	1.485	9.203	2.685	5.827	-22.212	0.730	0.953
SE	0.397	0.038	0.169	0.226	0.216	0.231	0.176	0.503	0.881
MSE	0.161	0.002	0.029	0.053	0.047	0.054	0.031	0.253	0.813

Table 3. Simulation results for the proposed PMLE under Scenario 2 with $\phi = 120$.

	α -0.5	β_1 -1.0	φ_1 -0.4	φ_2 0.2	θ_1 0.3	λ 0.5	ϕ 120
$n = 100$							
Mean	-0.433	-1.048	-0.314	0.160	0.207	0.629	130.483
RB (%)	-13.354	4.850	-21.569	-20.199	-30.895	25.739	8.735
SE	0.248	0.119	0.389	0.119	0.413	0.394	19.982
MSE	0.066	0.016	0.159	0.016	0.179	0.171	509.160
$n = 300$							
Mean	-0.518	-1.015	-0.408	0.149	0.307	0.534	122.941
RB (%)	3.636	1.511	1.946	-25.272	2.486	6.765	2.451
SE	0.159	0.073	0.206	0.067	0.212	0.246	10.199
MSE	0.025	0.006	0.042	0.007	0.045	0.062	112.665
$n = 500$							
Mean	-0.531	-0.427	0.149	0.329	121.462	0.525	-1.011
RB (%)	6.222	6.823	-25.370	9.795	1.219	5.020	1.068
SE	0.127	0.144	0.050	0.149	7.765	0.197	0.059
MSE	0.017	0.022	0.005	0.023	62.427	0.040	0.004
$n = 1000$							
Mean	-0.541	-1.008	-0.442	0.147	0.345	0.518	120.628
RB (%)	8.180	0.753	10.413	-26.418	14.834	3.574	0.523
SE	0.088	0.041	0.094	0.035	0.098	0.139	5.375
MSE	0.009	0.002	0.011	0.004	0.012	0.020	29.288

Table 4. Simulation results for the proposed PMLE under Scenario 2 with $\phi = 20$.

	α -0.5	β_1 -1.0	φ_1 -0.4	φ_2 0.2	θ_1 0.3	λ 0.5	ϕ 20
$n = 100$							
Mean	-0.091	-1.240	-0.345	0.197	0.211	1.332	21.745
RB (%)	-81.898	24.005	-13.710	-1.590	-29.671	166.329	8.727
SE	2.411	1.075	0.444	0.180	0.390	5.924	3.253
MSE	5.979	1.213	0.200	0.032	0.160	35.784	13.629
$n = 300$							
Mean	-0.406	-1.080	-0.381	0.179	0.282	0.733	20.565
RB (%)	-18.707	7.985	-4.638	-10.545	-5.972	46.634	2.824
SE	0.248	0.150	0.204	0.074	0.209	0.464	1.710
MSE	0.070	0.029	0.042	0.006	0.044	0.270	3.244
$n = 500$							
Mean	-0.448	-1.051	-0.387	0.175	0.292	0.653	20.289
RB (%)	-10.381	5.118	-3.291	-12.493	-2.803	30.606	1.445
SE	0.196	0.118	0.147	0.055	0.152	0.371	1.264
MSE	0.041	0.017	0.022	0.004	0.023	0.161	1.681
$n = 1000$							
Mean	-0.483	-1.030	-0.398	0.171	0.306	0.592	20.136
RB (%)	-3.365	3.032	-0.486	-14.723	1.915	18.495	0.678
SE	0.143	0.086	0.099	0.039	0.102	0.272	0.878
MSE	0.021	0.008	0.010	0.002	0.010	0.083	0.789

Overall, all parameters are reasonably estimated even in small samples. In all scenarios, the MSE uniformly decreases as *n* increases, which represents numerical evidence of the PMLE's consistency. Among all parameters, the estimation of α is the one presenting the highest relative bias. The simulation results provide numerical evidence supporting the theory discussed in Section 3.

6. Application

In this section, we showcase the usefulness of the proposed βARMA$_\lambda$ model in a real data application. The data comprehend the monthly proportion of stocked hydroelectric energy in the south of Brazil from January 2000 to May 2022, yielding a total of 269 observations. The last six months of data were reserved for out-of-sample forecasting purposes, and hence, for modeling purposes the sample size is $n = 263$. This is an updated time series that was also modeled in [13]. The data are freely available at the Operador Nacional do Sistema Elétrico repository (http://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao/dados-gerais/, accessed on 29 October 2024). The hydrological study is essential for the adequate distribution of energy and to supply the consumption demand of the population. Forecasts are widely used by institutions to prevent energy shortages. All the results presented in this section can be accessed by a friendly web application available at http://ufsm.shinyapps.io/appBARMA/, accessed on 29 October 2024. In this online application, users can upload any time series data to fit using the proposed model.

Figure 2 presents the time series plot of the data, as well as its sample autocorrelation (ACF) and partial autocorrelation (PACF) functions. From the time series plot, a clear yearly seasonal pattern can be observed. In order to capture this seasonality, we shall employ $x_t = \sin(2\pi t / 12)$ as covariate in the model.

Additionally, we consider a further step in the optimization procedure that explores different initial values of λ (Aranda-Ordaz parameter); this step selects the model that has the smallest AIC value. The algorithm is described by:

1. Let $\boldsymbol{\lambda}^0 = (\lambda_1^0, \ldots, \lambda_L^0)^\top \in (0, \infty)^L$ be a vector of initial values associated with the Aranda-Ordaz link function parameter.
2. Fit L models, one for each λ_l^0, for $l \in \{1, \ldots, L\}$.
3. Calculate the AIC for each fitted model and choose the one with the lowest value.

For this application, we considered $\boldsymbol{\lambda}^0 = (0.5, 1, 3, 5)^\top$.

Model selection is carried out in a similar fashion to the iterative Box and Jenkins methodology [46]. To select the model, first we compare the AIC of the following competing models: βAR$_\lambda$(1), βAR$_\lambda$(2), βAR$_\lambda$(3), βMA$_\lambda$(1), βMA$_\lambda$(2), βMA$_\lambda$(3), βARMA$_\lambda$(1, 1), βARMA$_\lambda$(2, 1), βARMA$_\lambda$(3, 1), βARMA$_\lambda$(1, 2), βARMA$_\lambda$(1, 3), βARMA$_\lambda$(2, 2), βARMA$_\lambda$(3, 2), βARMA$_\lambda$(2, 3), and βARMA$_\lambda$(3, 3). The model selected by the AIC was a βARMA$_\lambda$(2, 1) given by:

$$\eta_t = g(\mu_t, \lambda) = \alpha + \beta_1 \sin(2\pi t / 12) + \varphi_2 \big[g(y_{t-2}, \lambda) - \beta_1 \sin(2\pi(t-2)/12) \big] + \theta_1 r_{t-1}.$$

The parameter φ_1 was not significant. Table 5 presents a summary of the fitted model with estimated parameters, standard errors, *Z*-statistics, and *p*-values, as well as AIC, BIC, deviance, and Ljung–Box test for the residuals. We also test the null hypothesis $\mathcal{H}_0 : \lambda = 1$ (logit), rejecting the null hypothesis at the 1% significance level, indicating that the logit function is not adequate for the data.

To carry on the residual analysis, we consider the standardized ordinary residual. The diagnostic plots presented in Figure 3 suggest that the residuals do not exhibit any pattern and are uncorrelated, as confirmed by the Ljung–Box test [13]. Figure 4 shows the normal plot with simulated envelope of the residuals considering the proposed model and the standard βARMA model with different fixed link functions. Among all competitors, the proposed model was the only one capable of fitting the data adequately, with the residuals

lying uniformly within the 95% confidence region. The plots and tests further support the hypothesis that the model is correctly specified.
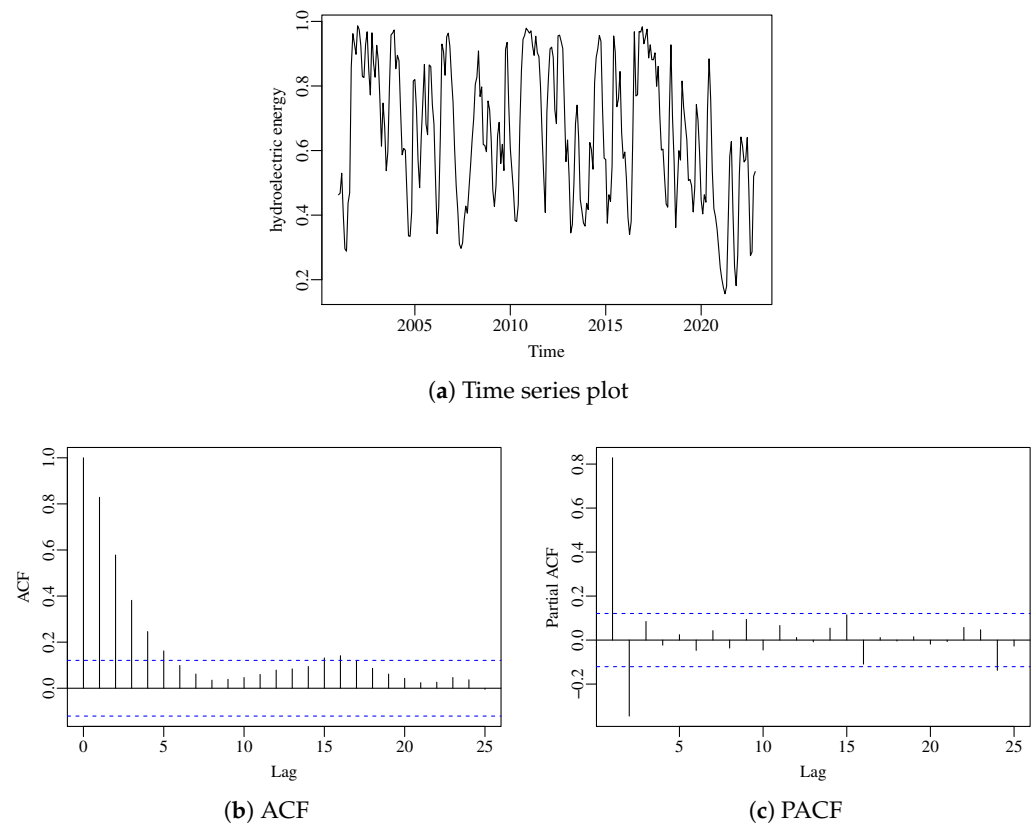

(**a**) Time series plot

(**b**) ACF

(**c**) PACF

Figure 2. Time series and related plots for the hydroelectric energy.

Table 5. Summary statistics for the fitted $\beta\mathrm{ARMA}_\lambda(2,1)$.

	Estimate	Std. Error	Z Statistic	*p*-Value
α	0.616	0.249	2.478	0.013
φ_2	0.445	0.109	4.090	<0.001
θ_1	0.840	0.062	13.634	<0.001
β_1	−0.424	0.163	2.600	0.009
λ	1.920	0.264	7.284 *	<0.001
ϕ	12.482	1.073	-	-
Log-likelihood = 209.941; $D = 228.188$				
AIC $= -414.301$; BIC $= -396.440$				
Ljung–Box test: *p*-value = 0.456				

* Z-statistic calculated under $\mathcal{H}_0 : \lambda = 1$.

Figure 5a presents the time series plot of the data and the predicted values (in-sample forecasts), while Figure 5b presents the six-step out-of-sample forecasts. In order to make a comparison, we have added to the plot out-of-sample forecasts for the fitted $\beta\mathrm{ARMA}(2,1)$ model coupled with four different link functions, namely, the fitted Aranda-Ordaz ($\beta\mathrm{ARMA}_\lambda$), the logit, the probit, and the cloglog. It is noteworthy that the $\beta\mathrm{ARMA}_\lambda$ model yielded more accurate out-of-sample forecasts than the competitors.

(**a**) Observed vs. fitted values

(**b**) Quantile residuals

(**c**) Residual ACF

(**d**) Residual PACF
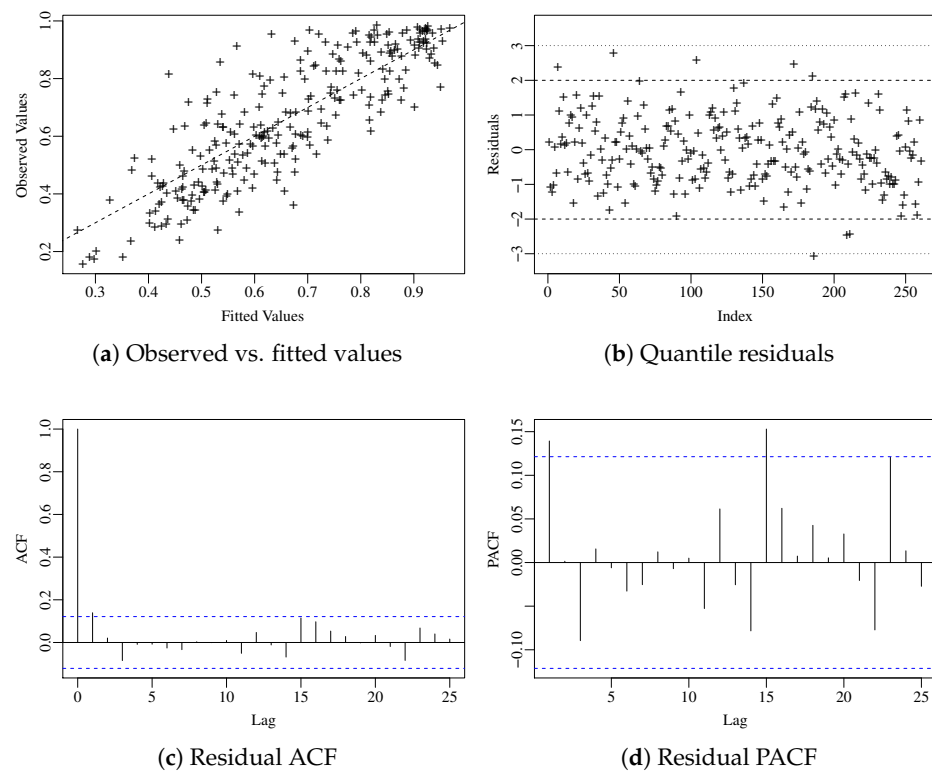
Figure 3. Diagnostic plots for the $\beta ARMA_\lambda(2,1)$ residuals.

(**a**) $\beta ARMA_\lambda(2,1)$

(**b**) $\beta ARMA(2,1)$ with logit link function

(**c**) $\beta ARMA(2,1)$ with probit link function

(**d**) $\beta ARMA(2,1)$ with cloglog link function
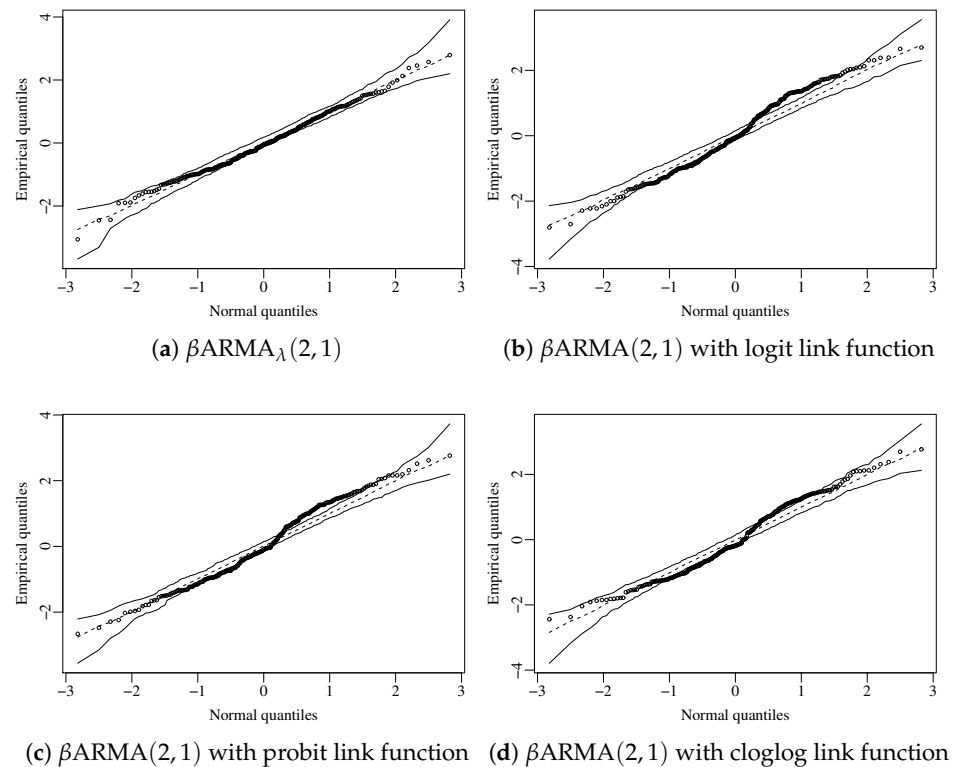
Figure 4. Normal plot with simulated envelope for different fitted models.

Table 6 presents in-sample and out-of-sample forecasting root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) measures between observed y_t and fitted $\widehat{\mu}_t$ values, for all t, for the four fitted models. In addi-

tion, we present the forecast measures by modeling the time series with smaller periods, namely $n \in \{253, 243, 233\}$. We observe that the proposed βARMA_λ model outperforms the competitor models in all metrics in the in-sample case. Considering out-of-sample forecasting, the proposed model outperforms in most cases, with lower performance only in MAPE for the logit and probit link functions, and in RMSE for the logit link function when considering $n = 233$. The usual logit link function performs better than our proposal only in two metrics of one prediction scenario. Our proposal performs better in most cases, and when it does not, it is still very competitive. The Aranda-Ordaz link function with $\lambda = 1.920$ has a heavier right tail compared to the other links, which may explain its superior forecasting performance.

(**a**) In-sample forecasts (fitted values)   (**b**) Out-of-sample forecasts
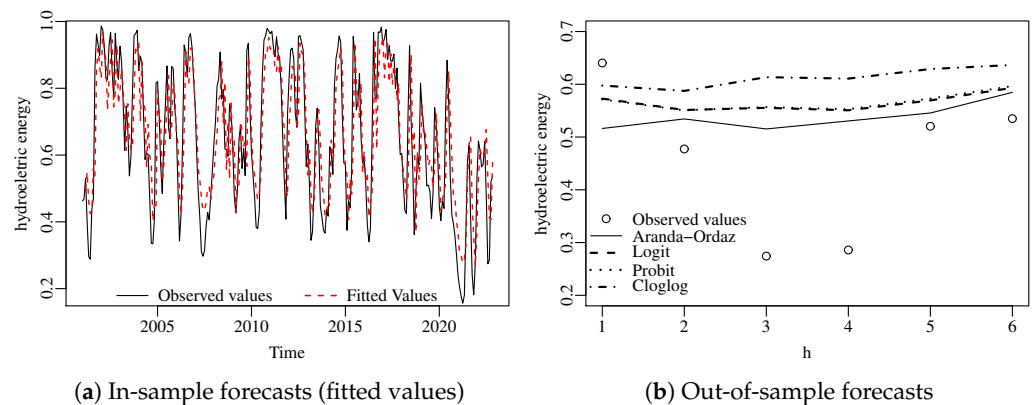
Figure 5. Observed in-sample and out-of-sample forecasts for the hydroelectric energy data.

Table 6. In-sample and out-of-sample ($h = 6$) forecasting measures for the $\beta\text{ARMA}_\lambda(2, 1)$ model compared to the βARMA model with other fixed links (best figures are in bold).

Link	In-Sample			Out-of-Sample		
	RMSE	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**
			$n = 263$			
Aranda-Ordaz	**0.116**	**0.092**	**17.026%**	**0.153**	**0.124**	**36.474%**
logit	0.155	0.130	24.395%	0.166	0.132	40.233%
probit	0.145	0.121	22.466%	0.167	0.134	40.676%
cloglog	0.145	0.122	22.899%	0.207	0.171	51.138%
			$n = 253$			
Aranda-Ordaz	**0.116**	**0.093**	**16.871%**	**0.103**	**0.073**	**13.384%**
logit	0.155	0.130	24.082%	0.113	0.088	15.619%
probit	0.144	0.121	22.246%	0.107	0.081	14.539%
cloglog	0.146	0.125	23.755%	0.109	0.083	15.225%
			$n = 243$			
Aranda-Ordaz	**0.115**	**0.092**	**15.976%**	**0.183**	**0.157**	**59.158%**
logit	0.152	0.129	22.349%	0.198	0.174	69.918%
probit	0.142	0.121	20.768%	0.186	0.160	64.178%
cloglog	0.142	0.121	21.130%	0.185	0.158	63.833%
			$n = 233$			
Aranda-Ordaz	**0.114**	**0.092**	**15.372%**	0.265	**0.233**	53.918%
logit	0.150	0.128	21.526%	**0.263**	0.239	**52.931%**
probit	0.140	0.119	19.820%	0.267	0.240	53.906%
cloglog	0.140	0.120	20.008%	0.268	0.241	54.016%

Moreover, besides presenting the best forecast performance, the flexibility of the Aranda-Ordaz link allows for the proposed βARMA_λ model to circumvent problems steaming from link function misspecification, being more robust and facilitating the construction of an adequate model for practitioners.

7. Conclusions

In this work, we considered an alternative way to model bounded time series by extending the relationship between the random component and linear predictors in the context of βARMA models. To do that, we introduced the Aranda-Ordaz parametric link function in place of the traditional fixed links. The parameter on the Aranda-Ordaz link is estimated along the other βARMA parameters by partial maximum likelihood. We discussed large sample inferences and presented a Monte Carlo simulation study showcasing the estimator's performance in finite sample sizes. Based on the proposed methodology, we discussed residual analysis, hypothesis testing, and forecasting. Finally, a real data application to the proportion of stocked hydroelectric energy in the south of Brazil was presented, showcasing the usefulness of the proposed model, which has outperformed the competing ones.

Author Contributions: Conceptualization, D.R.C. and F.M.B.; methodology, C.E.F.M., D.R.C., G.P. and F.M.B.; software, C.E.F.M., D.R.C. and F.M.B.; formal analysis, C.E.F.M., D.R.C., G.P. and F.M.B.; investigation, C.E.F.M., D.R.C., G.P. and F.M.B.; resources, C.E.F.M., D.R.C., G.P. and F.M.B.; writing—original draft preparation, C.E.F.M., D.R.C., G.P. and F.M.B.; writing—review and editing, C.E.F.M., D.R.C., G.P. and F.M.B.; supervision, D.R.C. and F.M.B.; project administration, F.M.B.; funding acquisition, F.M.B. All authors have read and agreed to the published version of the manuscript.

Appendix A. Score Vector

The conditional score vector will be derived in this appendix. In what follows, all equalities should be understood to hold almost surely. Let $\boldsymbol{\delta} = (\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$, we have:

$$U(\boldsymbol{\delta}) = \frac{\partial \ell(\boldsymbol{\omega})}{\partial \delta_i} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_i} = \phi \sum_{t=m+1}^{n} (y_t^* - \mu_t^*) \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_i}.$$

It is easy to see that $\frac{\partial \mu_t}{\partial \eta_t} = \exp(\eta_t)[1 + \lambda \exp(\eta_t)]^{-\frac{(1+\lambda)}{\lambda}}$, $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$ and $\mu_t^* = \psi(\mu_t \phi) - \psi[(1-\mu_t)\phi]$, where $\psi(\cdot) = \frac{d}{dx} \log(\Gamma(\cdot))$ is the digamma function. Proceeding similarly as in [10,30], the derivatives $\frac{\partial \eta_t}{\partial \delta_i}$ are given by:

$$\frac{\partial \eta_t}{\partial \alpha} = 1 + \sum_{j=1}^{q} \theta_j \frac{\partial r_{t-j}}{\partial \alpha} = 1 - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \alpha},$$

$$\frac{\partial \eta_t}{\partial \beta_l} = x_{tl} - \sum_{i=1}^{p} \varphi_i x_{(t-i)l} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_l}, \quad l = 1, \ldots, k,$$

$$\frac{\partial \eta_t}{\partial \varphi_i} = g(y_{t-i}, \lambda) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \varphi_i}, \quad i = 1, \ldots, p,$$

$$\frac{\partial \eta_t}{\partial \theta_j} = g(y_{t-j}, \lambda) - \eta_{t-j} - \sum_{l=1}^{q} \theta_l \frac{\partial \eta_{t-l}}{\partial \theta_j}, \quad j = 1, \ldots, q,$$

where x_{tr} is the r-th coordinate of $\boldsymbol{x}_t$. Finally:

$$\frac{\partial \ell(\boldsymbol{\omega})}{\partial \lambda} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \lambda},$$

where

$$\rho_t := \frac{\partial \mu_t}{\partial \lambda} = \frac{[1 + \lambda \exp(\eta_t)]^{-\frac{1}{\lambda}}}{\lambda} \left\{ \frac{1}{\exp(-\eta_t) + \lambda} - \frac{\log[1 + \lambda \exp(\eta_t)]}{\lambda} \right\}.$$

The score vector is given by $U(\boldsymbol{\omega}) = \left(U_\alpha(\boldsymbol{\omega}), U_{\boldsymbol{\beta}}(\boldsymbol{\omega})^\top, U_\phi(\boldsymbol{\omega}), U_{\boldsymbol{\varphi}}(\boldsymbol{\omega})^\top, U_{\boldsymbol{\theta}}(\boldsymbol{\omega})^\top, U_\lambda(\boldsymbol{\omega}) \right)^\top$, which can be compactly rewritten in matrix form as:

$$U_\alpha(\boldsymbol{\omega}) = \phi \boldsymbol{a}^\top \boldsymbol{T}(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad U_{\boldsymbol{\beta}}(\boldsymbol{\omega}) = \phi \boldsymbol{M}^\top \boldsymbol{T}(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad U_{\boldsymbol{\varphi}}(\boldsymbol{\omega}) = \phi \boldsymbol{P}^\top \boldsymbol{T}(\boldsymbol{y}^* - \boldsymbol{\mu}^*),$$
$$U_{\boldsymbol{\theta}}(\boldsymbol{\omega}) = \phi \boldsymbol{R}^\top \boldsymbol{T}(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \quad U_\lambda(\boldsymbol{\omega}) = \phi \boldsymbol{\rho}^\top (\boldsymbol{y}^* - \boldsymbol{\mu}^*),$$
$$U_\phi(\boldsymbol{\omega}) = \sum_{t=m+1}^{n} \left[\mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi[(1 - \mu_t)\phi] + \psi(\phi) \right],$$

where $\boldsymbol{y}^* := (y_{m+1}^*, \ldots, y_n^*)^\top$, $\boldsymbol{\mu}^* := (\mu_{m+1}^*, \ldots, \mu_n^*)^\top$, $\boldsymbol{a} := \left(\frac{\partial \eta_{m+1}}{\partial \alpha}, \ldots, \frac{\partial \eta_n}{\partial \alpha} \right)^\top$, $\boldsymbol{T} :=$ diag$\left\{ \frac{\partial \mu_{m+1}}{\eta_{m+1}}, \ldots, \frac{\partial \mu_n}{\partial \eta_n} \right\}$, $\boldsymbol{M}$, $\boldsymbol{P}$, and $\boldsymbol{R}$ are $(n-m) \times k$, $(n-m) \times p$ and $(n-m) \times q$ matrices, respectively, with (i,j)-th entry given by $M_{i,j} := \frac{\partial \eta_{i+m}}{\partial \beta_j}$, $P_{i,j} := \frac{\partial \eta_{i+m}}{\partial \varphi_j}$, $R_{i,j} := \frac{\partial \eta_{i+m}}{\partial \theta_j}$ and $\boldsymbol{\rho} := (\rho_{m+1}, \ldots, \rho_n)^\top$.

Appendix B. Information Matrix

In this appendix, we derive the cumulative conditional information matrix and all equalities should be understood to hold almost surely. The cumulative conditional information matrix is given by:

$$K_n(\boldsymbol{\omega}) = - \sum_{t=m+1}^{n} \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top} \middle| \mathcal{F}_{t-1} \right).$$

Let δ_i and δ_j be proxies for $\alpha, \boldsymbol{\beta}, \boldsymbol{\varphi}$ or $\boldsymbol{\theta}$. We have:

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \delta_i \partial \delta_j} = \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_j} \right) \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_i}$$
$$= \left[\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_j} + \frac{\partial \ell_t(\mu, \phi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_j} \right) \right] \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \delta_i}.$$

It is easy to show that $\mathbb{E}(\partial \ell_t(\mu_t, \phi)/\partial \mu_t | \mathcal{F}_{t-1}) = 0$ and $\mathbb{E}(y_t^* | \mathcal{F}_{t-1}) = \mu_t^*$. Hence:

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \delta_i \partial \delta_j} \middle| \mathcal{F}_{t-1} \right) = \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) \left(\frac{\partial \mu_t}{\partial \eta_t} \right)^2 \frac{\partial \eta_t}{\partial \delta_j} \frac{\partial \eta_t}{\partial \delta_i},$$

where $\partial \mu_t / \partial \eta_t$ and $\partial \eta_t / \partial \delta_j$ were derived in Section 3. Observe that

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} = -\phi^2 \left\{ \psi'(\mu_t \phi) + \psi'[(1 - \mu_t)\phi] \right\}.$$

Upon defining $v_t := \phi^2 \left\{ \psi'(\mu_t \phi) + \psi'[(1 - \mu_t)\phi] \right\}$, we have:

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \delta_i \partial \delta_j} \middle| \mathcal{F}_{t-1} \right) = -v_t \left(\frac{\partial \mu_t}{\partial \eta_t} \right)^2 \frac{\partial \eta_t}{\partial \delta_j} \frac{\partial \eta_t}{\partial \delta_i}.$$

The derivative of $U_\phi(\omega)$ with respect to δ_i is given by:

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \phi \partial \delta_i} = \frac{\partial}{\partial \phi}\left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t}\right)\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i} = [(y_t^* - \mu_t^*) - c_t]\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i},$$

where $c_t := \phi(\partial \mu_t^*/\partial \phi) = \phi\{\psi'(\mu_t\phi)\mu_t - \psi'[(1-\mu_t)\phi](1-\mu_t)\}$. Observe further that:

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \phi \partial \delta_i}\bigg| \mathscr{F}_{t-1}\right) = -c_t \frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i} \quad \text{and} \quad \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \phi^2}\bigg| \mathscr{F}_{t-1}\right) = -s_t,$$

where $s_t := \psi'(\mu_t\phi)\mu_t^2 + \psi'[(1-\mu_t)\phi](1-\mu_t)^2 - \psi'(\phi)$. The derivative of $U_\lambda(\omega)$ with respect to λ is given by:

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \lambda^2} = \frac{\partial}{\partial \lambda}\left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \lambda}\right) = \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2}\left(\frac{\partial \mu_t}{\partial \lambda}\right)^2 + \frac{\partial \ell_t(\mu, \phi)}{\partial \mu_t}\frac{\partial^2 \mu_t}{\partial \lambda^2}.$$

We then arrive at:

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \lambda^2}\bigg| \mathscr{F}_{t-1}\right) = \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2}\left(\frac{\partial \mu_t}{\partial \lambda}\right)^2 = -v_t \rho_t^2.$$

The derivatives with respect to λ and ϕ are given by:

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \lambda \partial \phi}\bigg| \mathscr{F}_{t-1}\right) = \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t \partial \phi}\bigg| \mathscr{F}_{t-1}\right)\frac{\partial \mu_t}{\partial \lambda} = -c_t \rho_t.$$

The derivatives with respect to δ_i and λ, are given by:

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \delta_i \partial \lambda} = \frac{\partial}{\partial \lambda}\left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i}\right)$$

$$= \frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2}\frac{\partial \mu_t}{\partial \lambda}\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i} + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t}\frac{\partial}{\partial \lambda}\left(\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i}\right),$$

and, hence:

$$\mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \delta_i \partial \lambda}\bigg| \mathscr{F}_{t-1}\right) = -v_t \rho_t \frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \delta_i}.$$

Let $V := \text{diag}\{v_{m+1}, \ldots, v_n\}$, $C := \text{diag}\{c_{m+1}, \ldots, c_n\}$, $S := \text{diag}\{s_{m+1}, \ldots, s_n\}$, the conditional cumulative information matrix can be written as:

$$K_n(\omega) = \begin{pmatrix} K_{\alpha,\alpha} & K_{\alpha,\beta} & K_{\alpha,\phi} & K_{\alpha,\varphi} & K_{\alpha,\theta} & K_{\alpha,\lambda} \\ K_{\beta,\alpha} & K_{\beta,\beta} & K_{\beta,\phi} & K_{\beta,\varphi} & K_{\beta,\theta} & K_{\beta,\lambda} \\ K_{\phi,\alpha} & K_{\phi,\beta} & K_{\phi,\phi} & K_{\phi,\varphi} & K_{\phi,\theta} & K_{\phi,\lambda} \\ K_{\varphi,\alpha} & K_{\varphi,\beta} & K_{\varphi,\phi} & K_{\varphi,\varphi} & K_{\varphi,\theta} & K_{\varphi,\lambda} \\ K_{\theta,\alpha} & K_{\theta,\beta} & K_{\theta,\phi} & K_{\theta,\varphi} & K_{\theta,\theta} & K_{\theta,\lambda} \\ K_{\lambda,\alpha} & K_{\lambda,\beta} & K_{\lambda,\phi} & K_{\lambda,\varphi} & K_{\lambda,\theta} & K_{\lambda,\lambda} \end{pmatrix},$$

where $K_{\alpha,\alpha} = a^\top V T^2 a$, $K_{\alpha,\beta} = K_{\beta,\alpha}^\top = aVT^2 M$, $K_{\alpha,\phi} = K_{\phi,\alpha}^\top = a^\top CT\mathbf{1}$, $K_{\alpha,\varphi} = K_{\varphi,\alpha}^\top = a^\top VT^2 P$, $K_{\alpha,\theta} = K_{\theta,\alpha}^\top = a^\top VT^2 R$, $K_{\alpha,\lambda} = K_{\lambda,\alpha}^\top = a^\top VT\rho$, $K_{\beta,\beta} = M^\top VT^2 M$, $K_{\beta,\phi} = K_{\phi,\beta}^\top = M^\top CT\mathbf{1}$, $K_{\beta,\varphi} = K_{\varphi,\beta}^\top = M^\top VT^2 P$, $K_{\beta,\theta} = K_{\theta,\beta}^\top = M^\top VT^2 R$, $K_{\beta,\lambda} = K_{\lambda,\beta}^\top = M^\top VT\rho$, $K_{\phi,\phi} = \text{tr}(S)$, $K_{\phi,\varphi} = K_{\varphi,\phi}^\top = P^\top CT\mathbf{1}$, $K_{\phi,\theta} = K_{\theta,\phi}^\top = R^\top CT\mathbf{1}$, $K_{\phi,\lambda} = K_{\lambda,\phi}^\top = C\rho$, $K_{\varphi,\varphi} = P^\top VT^2 P$, $K_{\varphi,\theta} = K_{\theta,\varphi}^\top = P^\top VT^2 R$, $K_{\varphi,\lambda} = K_{\lambda,\varphi}^\top = P^\top VT\rho$, $K_{\theta,\theta} = R^\top VT^2 R$, $K_{\theta,\lambda} = K_{\lambda,\theta}^\top = R^\top VT\rho$, and $K_{\lambda,\lambda} = \rho^\top V\rho$, with $\mathbf{1}$ denoting a $(n-m)$ vector of ones. From

the cumulative conditional information matrix we conclude that the model parameters are not orthogonal, contrarily to some linear and some GARMA models [3].

References

1. McKenzie, E. Some simple models for discrete variate time series. *J. Am. Water Resour. Assoc.* **1985**, *21*, 645–650. [CrossRef]
2. McCullagh, P.; Nelder, J. *Generalized Linear Models*, 2nd ed.; Chapman and Hall: Boca Raton, FL, USA, 1989.
3. Benjamin, M.A.; Rigby, R.A.; Stasinopoulos, D.M. Generalized autoregressive moving average models. *J. Am. Stat. Assoc.* **2003**, *98*, 214–223. [CrossRef]
4. Kedem, B.; Fokianos, K. *Regression Models for Time Series Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
5. Janacek, G.; Swift, A. A class of models for non-normal time series. *J. Time Ser. Anal.* **1990**, *11*, 19–31. [CrossRef]
6. Tiku, M.L.; Wong, W.K.; Vaughan, D.C.; Bian, G. Time series models in non-normal situations: Symmetric innovations. *J. Time Ser. Anal.* **2000**, *21*, 571–596. [CrossRef]
7. Jung, R.C.; Kukuk, M.; Liesenfeld, R. Time series of count data: Modeling, estimation and diagnostics. *Comput. Stat. Data Anal.* **2006**, *51*, 2350–2364. [CrossRef]
8. Ribeiro, T.F.; Peña-Ramírez, F.A.; Guerra, R.R.; Alencar, A.P.; Cordeiro, G.M. Forecasting the proportion of stored energy using the unit Burr XII quantile autoregressive moving average model. *Comput. Appl. Math.* **2024**, *43*, 27. [CrossRef]
9. Ferrari, S.L.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
10. Rocha, A.V.; Cribari-Neto, F. Beta autoregressive moving average models. *TEST* **2009**, *18*, 529; Erratum in *TEST* **2017**, *26*, 451–459. [CrossRef]
11. Bayer, F.M.; Cintra, R.J.; Cribari-Neto, F. Beta seasonal autoregressive moving average models. *J. Stat. Comput. Simul.* **2018**, *88*, 2961–2981. [CrossRef]
12. Pumi, G.; Valk, M.; Bisognin, C.; Bayer, F.M.; Prass, T.S. Beta autoregressive fractionally integrated moving average models. *J. Stat. Plan. Inference* **2019**, *200*, 196–212. [CrossRef]
13. Scher, V.T.; Cribari-Neto, F.; Pumi, G.; Bayer, F.M. Goodness-of-fit tests for βARMA hydrological time series modeling. *Environmetrics* **2020**, *31*, e2607. [CrossRef]
14. Palm, B.G.; Bayer, F.M.; Cintra, R.J. Prediction intervals in the beta autoregressive moving average model. *Commun. Stat.-Simul. Comput.* **2023**, *52*, 3635–3656. [CrossRef]
15. Cribari-Neto, F.; Scher, V.T.; Bayer, F.M. Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. *Int. J. Forecast.* **2023**, *39*, 98–109. [CrossRef]
16. Bayer, F.M.; Pumi, G.; Pereira, T.L.; Souza, T.C. Inflated beta autoregressive moving average models. *Comput. Appl. Math.* **2023**, *42*, 183. [CrossRef]
17. Czado, C. On selecting parametric link transformation families in generalized linear models. *J. Stat. Plan. Inference* **1997**, *61*, 125–140. [CrossRef]
18. Guerrero, V.M.; Johnson, R.A. Use of the Box-Cox transformation with binary response models. *Biometrika* **1982**, *69*, 309–314. [CrossRef]
19. Czado, C. Parametric link modification of both tails in binary regression. *Stat. Pap.* **1994**, *35*, 189–201. [CrossRef]
20. Mallick, B.K.; Gelfand, A.E. Generalized Linear Models with Unknown Link Functions. *Biometrika* **1994**, *81*, 237–245. [CrossRef]
21. Newton, M.A.; Czado, C.; Chappell, R. Bayesian Inference for Semiparametric Binary Regression. *J. Am. Stat. Assoc.* **1996**, *91*, 142–153. [CrossRef]
22. Muggeo, V.M.; Ferrara, G. Fitting generalized linear models with unspecified link function: A P-spline approach. *Comput. Stat. Data Anal.* **2008**, *52*, 2529–2537. [CrossRef]
23. Canterle, D.R.; Bayer, F.M. Variable dispersion beta regressions with parametric link functions. *Stat. Pap.* **2019**, *60*, 1541–1567. [CrossRef]
24. Pumi, G.; Rauber, C.; Bayer, F.M. Kumaraswamy regression model with Aranda-Ordaz link function. *TEST* **2020**, *29*, 1051–1071. [CrossRef]
25. Aranda-Ordaz, F.J. On two families of transformations to additivity for binary response data. *Biometrika* **1981**, *68*, 357–363. [CrossRef]
26. Koenker, R.; Yoon, J. Parametric links for binary choice models: A Fisherian–Bayesian colloquy. *J. Econom.* **2009**, *152*, 120–130. [CrossRef]
27. Ramalho, E.A.; Ramalho, J.J.; Murteira, J.M. Alternative estimating and testing empirical strategies for fractional regression models. *J. Econ. Surv.* **2011**, *25*, 19–68. [CrossRef]
28. Flach, N. Generalized Linear Models with Parametric Link Families in R. Ph.D. Thesis, Department of Mathematics, Technische Universität München, München, Germany, 2014.
29. Dehbi, H.M.; Cortina-Borja, M.; Geraci, M. Aranda-Ordaz quantile regression for student performance assessment. *J. Appl. Stat.* **2016**, *43*, 58–71. [CrossRef]
30. Bayer, F.M.; Bayer, D.M.; Pumi, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. *J. Hydrol.* **2017**, *555*, 385–396. [CrossRef]
31. Nocedal, J.; Wright, S. *Numerical Optimization*; Springer: New York, NY, USA, 1999.

32. Fahrmeir, L.; Kaufmann, H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Stat.* **1985**, *13*, 342–368. [CrossRef]
33. Czado, C.; Munk, A. Noncanonical links in generalized linear models—When is the effort justified? *J. Stat. Plan. Inference* **2000**, *87*, 317–345. [CrossRef]
34. Fokianos, K.; Kedem, B. Partial likelihood inference for time series following generalized linear models. *J. Time Ser. Anal.* **2004**, *25*, 173–197. [CrossRef]
35. Fokianos, K.; Kedem, B. Prediction and Classification of non-stationary categorical time series. *J. Multivar. Anal.* **1998**, *67*, 277–296. [CrossRef]
36. Pawitan, Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*; Oxford University Press: Oxford, UK, 2001.
37. Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* **1943**, *54*, 426–482. [CrossRef]
38. Neyman, J.; Pearson, E.S. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika* **1928**, *20*, 175–240.
39. Rao, C.R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Camb. Philos. Soc.* **1948**, *44*, 50–57.
40. Terrell, G.R. The gradient statistic. *Comput. Sci. Stat.* **2002**, *34*, 206–215.
41. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
42. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
43. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244. [CrossRef]
44. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
45. Ospina, R.; Cribari-Neto, F.; Vasconcellos, K.L.P. Improved point and intervalar estimation for a beta regression model. *Comput. Stat. Data Anal.* **2006**, *51*, 960–981. [CrossRef]
46. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.