Structural Generalized Autoregressive Score models via robust optimization

Matheus Alves Pereira dos Santos
Pontifícia Universidade Catolica do Rio de Janeiro
Terminal da PUC - Gavea, Rio de Janeiro
matheusapds@id.uff.br

Davi Michel Valladão
Pontifícia Universidade Catolica do Rio de Janeiro
Terminal da PUC - Gavea, Rio de Janeiro
davimv@puc-rio.br

RESUMO

Os Modelos *Generalized Autoregressive Score* (GAS) oferecem uma estrutura versátil para modelar séries temporais, permitindo a variação dos parâmetros da distribuição ao longo do tempo. Embora amplamente aplicados em diversas áreas, esses modelos apresentam desafios na estimação devido ao grande número de parâmetros envolvidos. No entanto, ao tratá-los como problemas de otimização, é possível simplificar esse processo e aprimorar a precisão, ao incorporar técnicas como regularização e robustificação. Para gerar modelos robustos às mudanças de regime, considera-se a técnica de robustificação na amostra, na qual o modelo é estimado com base na pior subamostra possível de tamanho K. Utilizando dados mensais de carga elétrica de diferentes sistemas brasileiros, os resultados demonstram que a utilização do modelo GAS como um problema de otimização robusta resulta em melhorias nas três métricas de erro (MAPE, MASE e MAE) em comparação com os resultados obtidos pelo modelo Auto ARIMA.

PALAVRAS CHAVE. Séries Temporais, Modelos GAS, Otimização Robusta.

EST&MP – Estatística e Modelos Probabilísticos, BDA – Big Data e Analytics, PM – Programação Matemática

ABSTRACT

The Generalized Autoregressive Score (GAS) Models provide a versatile framework for modeling time series, accommodating dynamic variations in distribution parameters over time. Despite their widespread application across diverse domains, these models present challenges in estimation due to their numerous parameters. However, by treating them as optimization problems, we have the potential to simplify the estimation process and enhance accuracy, once this approach allows for the inclusion of techniques like regularization and robustification. In the context of generating models that are robust to regime changes, a robustification technique is considered in the sample, where the model is estimated based on the worst possible sub-sample of size K. Utilizing monthly electric load data from different Brazilian systems, our findings demonstrate that employing the GAS model as a robust optimization problem leads to improvements in all three error metrics (MAPE, MASE and MAE) compared to the results obtained from the Auto ARIMA model.

KEYWORDS. Time Series, GAS Models, Robust Optimization.

EST&MP – Estatística e Modelos Probabilísticos, BDA – Big Data e Analytics, PM – Programação Matemática

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

1. Introduction

Time series play a pivotal role in data-driven decision-making processes. Among their contributions is possible to cite the ability to construct future scenarios that can be integrated into optimization models. This underscores the significance of accurate predictions when addressing such problems. As a result, various methodologies for time series forecasting have been developed over time, including models like the Seasonal Autoregressive Integrated Moving Average (SARIMA) model [Box e Jenkins, 1970] and the State Space model [Harvey, 1990]. A potential limitation of both these methodologies lies in their assumption of data following a Normal distribution.

To address this limitation, Creal et al. [2013] and Harvey [2013] independently introduced the score-driven or GAS framework. This flexible framework facilitates the development of time series models with time-varying parameters, accommodating discrete or continuous distributions beyond Gaussian. In addition to the distribution flexibility, the dynamics of time-varying parameters can be described akin to ARMA models or even through unobservable components, such as State Space. This versatility positions GAS models as a highly adaptable framework, successfully applied in diverse domains including risk analysis [Patton et al., 2019], renewable energy [Hoeltgebaum et al., 2018], and finance [Ayala e Blazsek, 2018].

Despite the widespread use and demonstrated effectiveness of these models, some are rooted in methods from the 1980s and 1990s, constrained by the optimization technologies available at the time. However, advancements in computational capacity and optimization techniques have alleviated concerns around using exact methods in specific cases. Consequently, formulating time series models as optimization problems taps into the wealth of theory and modern techniques in this field, enhancing model accuracy [Bertsimas e Dunn, 2019a].

Another recurring challenge in time series analysis pertains to abrupt regime changes in the data, triggered by factors like economic crises, natural disasters, and pandemics. Traditional time series models are not inherently equipped to handle such characteristics, posing difficulties in generating accurate predictions.

Once again, the optimization-centric approach aids in tackling this challenge. Robust optimization techniques have proven to bolster prediction accuracy during regime change periods [Bertsimas e Paskov, 2020]. Although the referenced paper demonstrates the approach's ability to enhance model robustness without compromising performance in typical periods, the showcased results pertain to an Autoregressive Model of Order p, one of the simplest methods in time series forecasting.

Against this backdrop, this paper aims to examine whether the robustness offered by the methodology proposed by [Bertsimas e Paskov, 2020] remains intact when applied to more intricate models. To this end, we focus on a GAS model. In contrast to most GAS literature, which employs ARMA-like dynamics for time-varying parameters, as implemented in [Bodin et al., 2020], we consider dynamics driven by unobservable components, as in [Sarlo et al., 2023].

2. Theoretical Background

The aim of this section is to provide a more general development of the robust time series methodology proposed in [Bertsimas e Paskov, 2020] and make a overview about the structural GAS framework.

2.1. Robust Framework

In general, a time series model can be formulated as optimization problems, such as

$$\underset{\boldsymbol{\theta} \in \Theta}{\text{Min}} \quad \sum_{t=1}^{T} g(y_t, \boldsymbol{\theta}) \tag{1}$$

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

where $\boldsymbol{\theta}$ is the vector of model's parameters that may vary over time, $g(y_t, \boldsymbol{\theta})$ is a general loss function that depends on the observed time series (y_t) and $\boldsymbol{\theta}$, lastly Θ is the set that represents the domain of model's parameters. It is important to note that depending on the model, Θ may characterize constraints or describe the time-varying nature of the parameters.

Extending the concept proposed by Bertsimas e Paskov [2020], which exposes the model to worst training subsets according to some criteria and estimates it to perform well across others subsets scenarios, we can extend Equation (1) as

$$\underset{\boldsymbol{\theta}\in\Theta}{\text{Min}} \quad \underset{\mathbf{z}\in Z}{\text{Max}} \quad \sum_{t=1}^{T} z_t g(y_t, \boldsymbol{\theta}), \tag{2}$$

where $\mathbf{z} = (z_1, \ldots, z_T)$ is a vector of indicator variables $z_t \in \{0, 1\}, \forall t = 1, \ldots, T$. The idea behind Equation (2) is to train the model using the observed values that maximize the error. In other words, for a given model, with optimized parameters, the inner maximization problem selects the sub-sample that leads to the highest error for this model. This selection is made by the binary variable z_t that is contained in the Z set, which describes the criteria that govern this selection. Note that when $z_t = 1$ implies that y_t will be used in the training set. Solving this problem generates a model trained on the most challenging training set under some criterion, resulting in an improved robustness and stability during regime changes.

In this paper, the Z set will be defined in a way to provide the hardest sub-sample of length K. Note that as the inner maximization problem is linear in $\mathbf{z}$, solving the problem in Equation (2) is equivalent to optimizing over the convex hull of Z [Bertsimas e Dunn, 2019b] that is defined as

$$conv(Z) = \left\{ z : \sum_{t=1}^{T} z_t = k, 0 \le z_i \le 1, t \in [1, T] \right\}. \tag{3}$$

This modification simplifies the optimization process, because it avoids the use of integer variables. Using the definition in Equation (3) it is possible to rewrite the optimization problem as in Equation (4).

Regardless of how Z is defined, Equation (2) cannot be directly solved due to the nested min-max structure. To address this issue, a reformulation using robust optimization and duality principles is presented,transforming the problem into a computationally tractable minimization problem. For detailed information, refer to [Bertsimas e Paskov, 2020].

To make this reformulation, the duality theory is employed to convert the maximization into a minimization problem. This involves introducing dual variables:

$$\begin{aligned} \underset{z_t}{\text{Max}} \quad & \sum_{t=1}^{T} z_t g(y_t, \boldsymbol{\theta}) & \\ \text{s.t.} \quad & \sum_{t=1}^{T} z_t = K, & \times(\delta) \\ & 0 \le z_t \le 1, \quad \forall t = 1 \cdots T & \times(u_t) \end{aligned} \tag{4}$$

Using the dual variables presented in Equation (4), we obtain:

$$\sum_{t=1}^{T} z_t g(y_t, \Theta) \le \delta \sum_{t=1}^{T} z_t + \sum_{t=1}^{T} z_t u_t \le K\delta + \sum_{t=1}^{T} u_t \tag{5}$$

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

Therefore, this inequality holds if and only if $(\delta + u_t) \geq g(y_t, \boldsymbol{\theta})$ for all $t = 1, \cdots, T$. Here, it is important to remember that the vector of parameters $\boldsymbol{\theta}$, that here, is considered to be given, can vary over time or not. Consequently, the robust counterpart of the problem stated in Equation (2) is given by:

$$\begin{aligned} \underset{\boldsymbol{\theta}\in\Theta, \delta\geq 0, u_t \geq 0}{\text{Min}} \quad & K\delta + \sum_{t=1}^{T} u_t \\ \text{s.t.} \quad & \delta + u_t \geq g(y_t, \boldsymbol{\theta}) \qquad \forall t = 1, \cdots, T \end{aligned} \tag{6}$$

In the robust model, presented in Equation (6) it was introduced additional variables and constraints to account for robustness considerations. The non-negative parameter δ represents the worst-case deviation from the loss function considered, and u_t are non-negative slack variables. The objective function includes a penalty term $K\delta$ and a sum of slack variables $\sum_{t=1}^{T} u_t$ to ensure that the worst-case deviations and slacks are minimized.

The constraints enforce that the sum of δ and u_t is greater than or equal to loss function considered $g(y_t, \boldsymbol{\theta})$ at each time point. This ensures that the robust model provides a good fit to the observed data while accounting for potential deviations. By solving the optimization problem in Equation (6), we can estimate the parameters $\boldsymbol{\theta}$ of the robust time series model while considering robustness aspects.

By performing all the manipulations showed previously, the robust time series model presented in Equation (2) has been reformulated into a computationally tractable form, which can be readily solved using commercial solvers. It is important to note that the development outlined in this section considers a general formulation for time series models, making it applicable to various types of methodologies.

The subsequent step involves defining the time series model to be utilized within this robust formulation. As mentioned previously, a Structural Generalized Autoregressive Score (GAS) model will be considered to investigate whether this stable framework remains effective for more complex time series models.

2.2. Generalized Autoregressive Score Model Specification

The Generalized Autoregressive Score (GAS) model belongs to the class of observation-driven models that incorporate time-varying parameters. The first step to define a GAS model, is to choose an appropriate predictive distribution for the considered data, that can be, discrete or continuous. This is one of the advantages of this framework, once it is possible to define Non-Gaussian models.

To define the model mathematically, let's suppose y_t is a univariate time series of interest, $f_{t|t-1}$ is the vector of the parameters of the predictive distribution that is assumed to be time-varying, and θ is a vector of fixed parameters. Supposing that $\mathbf{Y}_t = \{y_1, \cdots, y_t\}$ and $\mathbf{F}_t = \{f_0, f_{1|0}, \cdots, f_{t|t-1}\}$, the available information in time t is given by $\{\mathbf{Y}_{t-1}, \mathbf{F}_{t-1}\}$. So, in the GAS framework, it is possible to say that y_t is satisfactorily modeled considering the conditional or predictive density\probability function given by

$$y_t \sim p(y_t | f_{t|t-1}, \mathbf{Y}_{t-1}, \mathbf{F}_{t-1}; \theta). \tag{7}$$

Once chosen an adequate predictive distribution, the next step is to define the time-varying parameters' dynamic. Here appears the second advantage of the GAS framework, since it is possible to describe this dynamic using a ARMA-like equation or even stochastic unobserved components. As mentioned earlier, in this article we will focus on the latter. Now, to illustrate how to define this

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

kind of model, let suppose a very simple example, when the predictive distribution has only one time-varying parameter, denoted by f_t, which vary over time following a random walk dynamic.

$$\begin{aligned} h(f_{t|t-1}) &= m_{t|t-1} \\ m_{t|t-1} &= m_{t-1|t-2} + \kappa s_{t-1} \end{aligned} \tag{8}$$

Here, $h(\cdot)$ represents an appropriate link function that ensures the estimates of the parameter f_t adhere to their respective domains. In Equation (8), it is possible to notice that $h(f_{t|t-1})$ is equal to a unobserved component denoted by m_t, that describes a random walk dynamic. Still in relation to Equation (8) it is evident that the update mechanism of m_t relies on s_{t-1}, which is defined as:

$$s_{t|t-1} = I_{t|t-1}^{-d} \cdot \nabla_t, \tag{9}$$

where $I_{t|t-1}^{-d}$ represents the scaled Fisher Information Matrix and ∇_t is the score of the predictive distribution, defined as:

$$\nabla_t = \frac{\partial \ln \left(p(y_t | f_{t|t-1}, \mathbf{F}_{t-1}, \mathbf{Y}_{t-1}, \theta) \right)}{\partial f_t}. \tag{10}$$

For more information about how the mechanism described in Equation (9) is capable to correctly update the models estimates, see Caivano et al. [2016].

It is important to note that the scale parameter d typically takes values in the set $\left\{0, \frac{1}{2}, 1\right\}$, and if $d = \frac{1}{2}$, $I_{t|t-1}^{-\frac{1}{2}}$ results from the Cholesky decomposition of $I_{t|t-1}^{-1}$ [Bodin et al., 2020]. Finally, the parameter estimation in this model is performed via maximum likelihood. Naturally, this framework allows for the definition of numerous models with different dynamics.

3. Proposed Robust Structural Generalized Autoregressive Score Model

In this paper, with the objective of evaluating the additional accuracy that the robustification method presented in Sub-section 2.1 will provide to the GAS model, we will consider monthly data of electric load in the Brazilian North and South power systems from 2015 to 2021. Figure 1 shows that this data exhibits a seasonal pattern throughout the year, and does not seem to be a strong increasing or decreasing trend. However, starting from 2020, there is a noticeable change in the usual pattern, likely due to the COVID-19 pandemic.

To appropriately model this data, the GAS model needs to be defined accordingly. Considering the characteristics of the data shown in Figure 1, the proposed specification for the structural GAS model, considering a generic predictive distribution is as follows:

$$\begin{aligned} f_{t+1|t} &= c + m_{t+1|t} + \gamma_{t+1|t} \\ m_{t+1|t} &= \phi m_{t|t-1} + \kappa_m s_t, \quad -1 \leq \phi \leq 1 \\ \gamma_{t+1|t} &= \sum_{j=1}^{M_1} \left[\lambda_{1j} \cos\left(\frac{2\pi j t}{12}\right) + \lambda_{2j} \sin\left(\frac{2\pi j t}{12}\right) \right] \end{aligned} \tag{11}$$

Here, m_t and γ_t represent the trend and seasonal components, respectively. The trend component follows an AR(1) process, once the data seems to have a stationary trend, while the deterministic seasonal component is modeled using trigonometric terms. The time-varying parameter f_t is defined as the sum of a constant and the two aforementioned components.

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

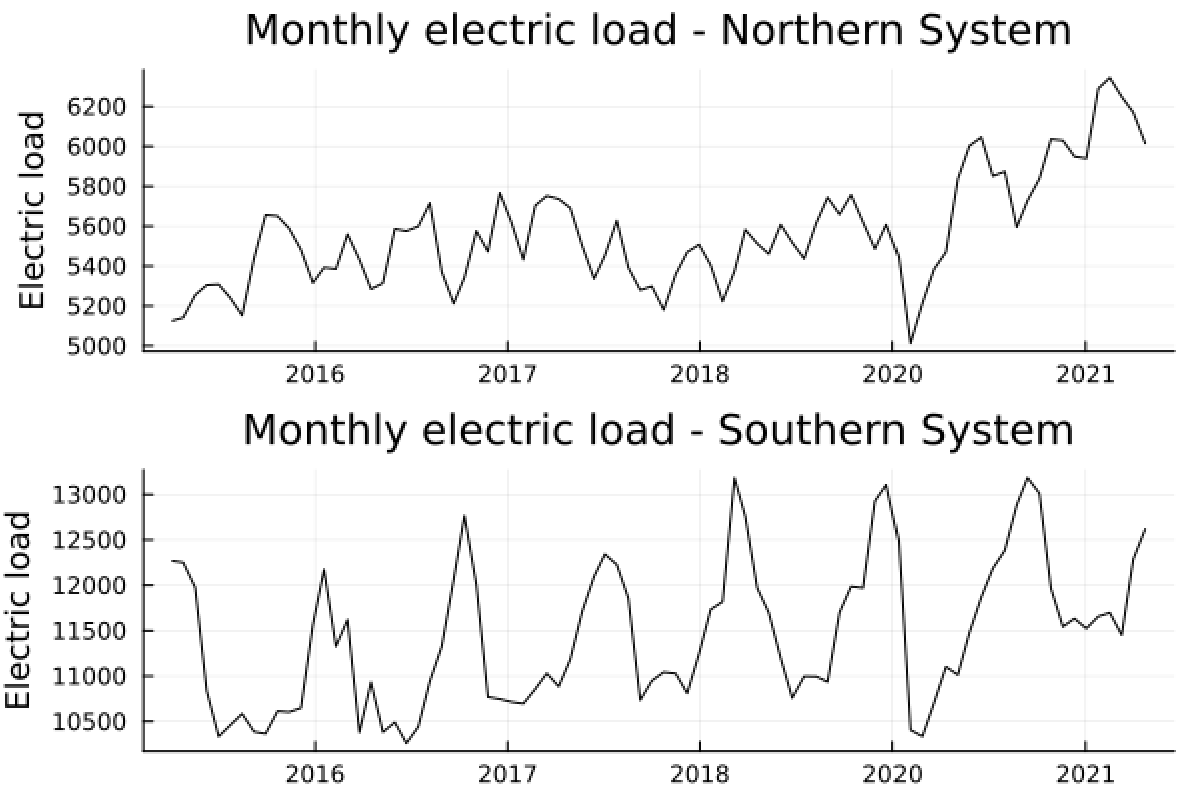

Figure 1: Monthly electric load of the Brazilian North and South power systems

Now, formulating the model in Equation (11) as a Non-Linear optimization problem (NLP), we have:

$$
\begin{aligned}
\underset{\Theta}{\text{Min}} \quad & -\sum_{t=1}^{T} \ln\left(p(y_t|f_{t|t-1}, \mathbf{F}_{t-1}, \mathbf{Y}_{t-1}, \theta)\right) & \\
s.t \quad & f_{t+1|t} = c + m_{t+1|t} + \gamma_{t+1|t} & \forall t = 1, \cdots, T \\
& m_{t+1|t} = \phi m_{t|t-1} + \kappa_m s_t & \forall t = 1, \cdots, T \\
& \gamma_{t+1|t} = \sum_{j=1}^{M_1} \left[\lambda_{1j} \cos\left(\frac{2\pi j t}{12}\right) + \lambda_{2j} \sin\left(\frac{2\pi j t}{12}\right)\right] & \forall t = 1, \cdots, T \\
& -1 \leq \phi \leq 1 & \\
& \kappa_m \geq 0 &
\end{aligned}
\tag{12}
$$

Here, $\boldsymbol{\Theta} = \{\boldsymbol{f}, \mathbf{m}, \boldsymbol{\gamma}, c, \phi, \kappa_m, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \theta\}$. It is essential to highlight that with this formulation, the model is classified as a NLP, and because of that, it has to be solved by a interior points based solver and there is no guarantee of global optimization. So, to avoid problems with local optimal points a good initialization is necessary when dealing with this kind of problems.

So, once we have formulated the proposed structural GAS model as a optimization problem, it is possible to use the robustification technique showed in Equation (6) and formulate a robust

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

structural GAS model as

$$
\begin{aligned}
&\underset{\boldsymbol{\Theta},\delta\geq 0,u_t\geq 0}{\text{Min}} K\delta + \sum_{t=1}^{T} u_t \\
&s.t \quad \delta + u_t \geq -\sum_{t=1}^{T} \ln\left(p(y_t|f_{t|t-1},\mathbf{F}_{t-1},\mathbf{Y}_{t-1},\theta)\right) && \forall t = 1,\cdots,T \\
&\qquad f_{t+1|t} = c + m_{t+1|t} + \gamma_{t+1|t} && \forall t = 1,\cdots,T \\
&\qquad m_{t+1|t} = \phi m_{t|t-1} + \kappa_m s_t && \forall t = 1,\cdots,T \\
&\qquad \gamma_{t+1|t} = \sum_{j=1}^{M_1}\left[\lambda_{1j}\cos\left(\frac{2\pi jt}{12}\right) + \lambda_{2j}\sin\left(\frac{2\pi jt}{12}\right)\right] && \forall t = 1,\cdots,T \\
&\qquad -1 \leq \phi \leq 1 \\
&\qquad \kappa_m \geq 0
\end{aligned}
\tag{13}
$$

It is interesting to notice that, the use of this robustification methodology does not change the classification of the optimization problem. In other words, the formulation shown in Equation (13) has an improvement in robustness with the same level of complexity as its non-robust formulation shown in Equation (12).

Once we have a distribution generic formulation for the proposed structural GAS model, and its robust version, it is necessary to define a predictive distribution to conclude the models definition. To being able to complete our illustrative example, we will consider a Normal distribution where the mean parameter μ_t varies over time. In other words, we assume that the predictive distribution is given by:

$$p(y_t|\mathbf{Y}_{t-1},\mathbf{F}_{t-1},\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(y_t-\mu_t)^2}{\sigma^2}}, \quad \sigma^2 > 0 \tag{14}$$

To complete the definition of the proposed GAS models, we need to compute the appropriate standard score s_t for the Normal distribution using Equations (9) and (10). First, let us consider the logarithm of $p(y_t|Y_{t-1},F_{t-1},\sigma^2)$:

$$l(\mu_t,\sigma^2|y_t) = \ln p(y_t|\mathbf{Y}_{t-1},\mathbf{F}_{t-1},\sigma^2) = -\ln(\sigma) - \frac{1}{2}\ln(2\pi) - \frac{(y_t-\mu_t)}{2\sigma^2} \tag{15}$$

Using Equation (15), we can derive the expressions for ∇_t and $I_{t|t-1}$. The derivation of these results will be omitted here, but they can be found in [Bodin et al., 2020]. The expressions for ∇_t and $I_{t|t-1}$ are as follows:

$$\nabla_t = \begin{bmatrix} \frac{y_t-\mu_t}{\sigma^2} \\ \frac{-1}{2\sigma^2}\left(1-\frac{(y_t-\mu)^2}{\sigma^2}\right)\end{bmatrix} \tag{16}$$

$$I_{t|t-1} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4}\end{bmatrix} \tag{17}$$

It is important to notice that, once only the μ parameter is considered to be time-varying, only the element of the s vector that correspond to the mean parameter will be considered (s_μ). Formulating the model in Equation (13) to consider the Gaussian distribution we have the robust

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

Normal model in Equation (18). Note that, in this formulation we have a constraint that says that the variance parameter (σ^2) is strictly greater than zero. On the other hand, this type of restriction cannot be implemented in commercial solvers. To overcame this limitation, the implementation of this constraint is done by considering that σ^2 is greater or equal to a small number denoted as τ. Another relevant characteristic of this formulation is that, once the domain of a parameter is guaranteed by constraints on the model, it is not necessary to consider link functions, which can potentially simplifies the specification of the scaled score function (s_t).

$$
\begin{aligned}
&\underset{\boldsymbol{\Theta},\delta\geq 0,u_t\geq 0}{\text{Min}} K\delta + \sum_{t=1}^{T} u_t \\
s.t \quad & \delta + u_t \geq -\sum_{t=1}^{T} l(\mu_t, \sigma^2|y_t) && \forall t = 1,\cdots,T \\
& \mu_{t+1|t} = c + m_{t+1|t} + \gamma_{t+1|t} && \forall t = 1,\cdots,T \\
& m_{t+1|t} = \phi m_{t|t-1} + \kappa_m s_{\mu t} && \forall t = 1,\cdots,T \\
& \gamma_{t+1|t} = \sum_{j=1}^{M_1}\left[\lambda_{1j}\cos\left(\frac{2\pi jt}{12}\right) + \lambda_{2j}\sin\left(\frac{2\pi jt}{12}\right)\right] && \forall t = 1,\cdots,T \\
& -1 \leq \phi \leq 1 \\
& \sigma^2 > 0 \\
& \kappa_m \geq 0
\end{aligned} \tag{18}
$$

4. Results Analysis

The implementation of the proposed models were done using JuMP.jl [Lubin et al., 2023], a mathematical programming framework available in Julia programming language. As mentioned before, the proposed models are classified as NLPs, so it is necessary to resort to a solver based on the interior point method. This was done using the Ipopt.jl, that is a package that implements the Ipopt algorithm [Wächter e Biegler, 2006].

To evaluate the gain in prediction accuracy of the Robust GAS model (18), against its non robust version, we will use the monthly electric load data for the North and Southern regions of Brazil. Taking advantage of the regime change induced by the Covid-19 pandemic, we can compare the forecast accuracy of the robust model with the Non-Robust model during typical and atypical periods. Here, the typical period refers to the pre-pandemic period from 2015 to 2019. The atypical period encompasses the years from 2020 onwards, as evident from the distinctive behavior observed in Figure 1.

The evaluation of the model's accuracy will be based on twelve-step ahead predictions using twelve rolling windows. For the first window of the typical periods, the model will be estimated using data from 2015 to 2017, and predictions will be made for 2018. For the second rolling window, the estimation period will include data of January 2018, and predictions will be made for the subsequent twelve steps ahead. This process will be repeated for the remaining windows, providing predictions for 2019. A similar procedure will be followed for the atypical periods, where the first estimation period will span from 2017 to 2019, and predictions will be made for 2020 and 2021. This process will be repeated for both systems. In addition to the two versions of the GAS model, the Auto ARIMA model [Hyndman e Khandakar, 2008], implemented using the StatsForecast.py

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

package [Federico Garza, 2022], will also be computed as a benchmark. The accuracy of the models will be compared using the MAPE, MASE, and MAE error metrics for quantitative analysis.

It is worth mentioning the model specifications used for all periods. The number of harmonics used for seasonality ($M1$) was set to 6, the K parameter of the robust model was defined as 70% of the length of the training period, and the scale parameter d was set to 1. The obtained results are summarized in Table 1, Table 2, and Table 3.

Table 1: Comparison of mean MAPE of each model for typical and atypical scenarios for both systems.

Scenario	Non-Robust Model	Robust Model	Auto ARIMA
North Typical	0.0382	0.0369	0.0387
North Atypical	0.0574	0.0549	0.0574
South Typical	0.0197	0.0253	0.0292
South Atypical	0.0323	0.0304	0.0390

Regarding the Table 1, which compares the mean value of the mean absolute percentage error (MAPE) between the models, the robust model consistently outperforms the non-robust model and the Auto ARIMA model in terms of accuracy. This is true for all analyzed scenarios, with the exception of the typical period of the Southern system, in which the non-robust model presented a slightly better result than the robust one. It is important to highlight that both methodologies based on the GAS model presented lower mean MAPE values than Auto ARIMA, indicating its superior predictive capacity and lower error percentages.

Table 2: Comparison of mean MASE of each model for typical and atypical scenarios for both systems.

Scenario	Non-Robust Model	Robust Model	Auto ARIMA
North Typical	1.6732	1.6310	1.7295
North Atypical	2.3831	2.2683	2.3869
South Typical	0.5407	0.6931	0.8135
South Atypical	1.0018	0.9476	1.2633

Turning to Table 2, which presents the mean Mean Absolute Scaled Error (MASE), the results confirm the robustness of the Robust Model. It consistently demonstrates lower mean MASE values compared to the Non-Robust Model and the Auto ARIMA model for all scenarios, again with exception of the typical period for the Southern system. This indicates that the Robust Model provides more accurate predictions and better captures the underlying patterns and dynamics of the time series data for atypical periods, without compromising its quality during periods.

Still on the Table 2, it is important to point out that a MASE value greater than 1 indicates that the model's performance is worse than the naive model, that is used as benchmark. In this case, it suggests that the models, including the Non-Robust Model, Robust Model, and Auto ARIMA model, have difficulty in accurately capturing the underlying patterns and dynamics of the time series data. The MASE values above 1 indicate that the models' predictions have a larger mean absolute error relative to the scale of the data, compared to the benchmark model.

While having a MASE greater than 1 may seem undesirable, it is important to note that the benchmark model used for comparison can vary, and different benchmarks may yield different interpretations. Additionally, the performance of a model should be evaluated in relation to other models and their respective MASE values. In the context of the presented analysis, the focus is on

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

comparing the performance of the different models, rather than achieving a MASE value below 1.

Therefore, although the models in the analysis may have MASE values above 1, the emphasis lies in assessing their relative performance. The comparison of the models reveals that the Robust Model consistently outperforms the Non-Robust Model and the Auto ARIMA model, demonstrating its superiority in terms of accuracy measures. This suggests that, among the evaluated models, the Robust Model provides the most accurate predictions despite the MASE values above 1.

It is worth noting that the interpretation of the MASE values should also consider the specific characteristics of the data and the forecasting task at hand. In some cases, achieving a MASE value below 1 may be more feasible or realistic, depending on the nature of the time series and the objectives of the forecasting exercise.

Table 3: Comparison of mean MAE of each model for typical and atypical scenarios for both systems.

Scenario	Non-Robust Model	Robust Model	Auto ARIMA
North Typical	205.6790	199.6790	210.4886
North Atypical	335.8411	321.1833	337.5163
South Typical	232.7111	297.9745	348.6002
South Atypical	372.0748	349.1153	455.2260

Moving on to Table 3, which showcases the mean Mean Absolute Error (MAE), once again, the Robust Model outperforms the Non-Robust Model and the Auto ARIMA model for the majority of the analyzed scenarios. The lower absolute errors associated with the Robust Model highlight its ability to minimize the discrepancy between predicted values and actual values, making it a favorable choice for accurate time series forecasting.

In summary, the evaluation of the Robust Model's performance using the Brazilian monthly electric load data for the North and South regions demonstrates its superior accuracy compared to the Non-Robust Model and the Auto ARIMA model, during atypical periods, without relevant losses during typical periods. The proposed GAS methodologies consistently outperforms the Auto ARIMA in terms of MAPE, MASE, and MAE measures for both typical and atypical periods in both systems. These results highlight the robustness and effectiveness of the proposed robust formulation in capturing the complexities and uncertainties present in the electric load data, particularly during atypical periods such as the Covid-19 pandemic.

5. Conclusion

In conclusion, this paper has successfully achieved its objectives by proposing a generic formulation for robustifying time series models based on the work of Bertsimas e Paskov [2020]. The case study conducted using the Brazilian monthly electric load data for the North and South systems consistently demonstrated the superiority of the Robust Model in terms of accuracy measures, including MAPE, MASE, and MAE, for atypical scenarios and a comparable result to its non-robust version during typical periods. The Robust Model exhibited robustness, reliability, and enhanced predictive capabilities, positioning it as a promising approach for time series forecasting applications, even in the presence of regime changes.

The findings of this study align with the results presented in Bertsimas e Paskov [2020], indicating that the proposed robustification methodology remains effective for more complex time series models, such as the proposed GAS model. This highlights the versatility and applicability of the methodology across various time series forecasting contexts.

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

As a future endeavor, the expansion of the GAS definition to encompass other distributions and incorporate different dynamics will be pursued. This pursuit aims to increase the generality of this methodology, allowing its application in diverse fields to enhance the accuracy of predictions and the decision-making process.

By extending the GAS model to encompass a wider range of distributions and dynamics, we aim to enhance its utility and relevance in practical forecasting scenarios. This ongoing development will contribute to advancing the field of time series forecasting and enable more accurate predictions in real-world applications across various domains.

In summary, this paper has made significant contributions to the field of time series forecasting by turning a robust formulation previously proposed more generic and demonstrating its effectiveness through a comprehensive case study. The consistent superiority of the Robust Model in terms of accuracy measures supports its potential for practical implementation and highlights its value in improving decision-making processes. As further advancements are made, the robustification methodology can be applied in various domains to enhance prediction accuracy and drive more informed decision-making.

References

Ayala, A. e Blazsek, S. (2018). Score-driven copula models for portfolios of two risky assets. *The European Journal of Finance*, 24(18):1861–1884. URL `https://doi.org/10.1080/1351847X.2018.1464488`.

Bertsimas, D. e Dunn, J. (2019a). *Machine Learning Under a Modern Optimization Lens*. Dynamic Ideas LLC. ISBN 9781733788502. URL `https://books.google.com.br/books?id=g3ZWygEACAAJ`.

Bertsimas, D. e Dunn, J. (2019b). *Machine Learning Under a Modern Optimization Lens*, chapter 17, p. 333–348. Dynamic Ideas LLC.

Bertsimas, D. e Paskov, I. (2020). Time series that are robust to regime changes. *Journal of Machine Learning Research*, 21:1–31.

Bodin, G., Saavedra, R., Fernandes, C., e Street, A. (2020). Scoredrivenmodels.jl: a julia package for generalized autoregressive score models. URL `https://github.com/LAMPSPUC/ScoreDrivenModels.jl`.

Box, G. e Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Caivano, M., Harvey, A., e Luati, A. (2016). Robust time series models with trend and seasonal components. *SERIEs*, 7(1):99–120. ISSN 1869-4195. URL `https://doi.org/10.1007/s13209-015-0134-1`.

Creal, D., Koopman, S. J., e Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Federico Garza, C. C. K. G. O., Max Mergenthaler Canseco (2022). StatsForecast: Lightning fast forecasting with statistical and econometric models. PyCon Salt Lake City, Utah, US 2022. URL `https://github.com/Nixtla/statsforecast`.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

Hoeltgebaum, H., Fernandes, C., e Street, A. (2018). Generating joint scenarios for renewable generation: The case for non-gaussian models with time-varying parameters. *IEEE Transactions on Power Systems*, 33(6):7011–7019.

Hyndman, R. J. e Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(3):1–22. URL `https://www.jstatsoft.org/index.php/jss/article/view/v027i03`.

Lubin, M., Dowson, O., Garcia, J. D., Huchette, J., Legat, B., e Vielma, J. P. (2023). Jump 1.0: Recent improvements to a modeling language for mathematical optimization. *Mathematical Programming Computation*. In press.

Patton, A. J., Ziegel, J. F., e Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413. ISSN 0304-4076. URL `https://www.sciencedirect.com/science/article/pii/S030440761930048X`.

Sarlo, R., Fernandes, C., e Borenstein, D. (2023). Lumpy and intermittent retail demand forecasts with score-driven models. *European Journal of Operational Research*, 307(3):1146–1160. ISSN 0377-2217. URL `https://www.sciencedirect.com/science/article/pii/S0377221722007743`.

Wächter, A. e Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57.

DOI: 10.59254/sbpo-2023-174883
https://proceedings.science/p/174883?lang=pt-br