UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

METODOLOGIA DE COVARIÂNCIA GINI APLICADA A ESTIMAÇÃO DE PARÂMETROS EM MODELOS AUTO REGRESSIVOS COM CAUDAIS LONGOS

Lucas Chilelli da Silva

Itajubá

Julho 2019

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Lucas Chilelli da Silva

METODOLOGIA DE COVARIÂNCIA GINI APLICADA A ESTIMAÇÃO DE PARÂMETROS EM MODELOS AUTO REGRESSIVOS COM CAUDAIS LONGOS

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção como parte dos requisitos para a obtenção do título de Mestre em Ciências em Engenharia de Produção.

Orientador: Prof. Dr. Rafael Coradi Leme

Itajubá

Julho 2019

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Lucas Chilelli da Silva

METODOLOGIA DE COVARIÂNCIA GINI APLICADA A ESTIMAÇÃO DE PARÂMETROS EM MODELOS AUTO REGRESSIVOS COM CAUDAIS LONGOS

Dissertação aprovada por banca examinadora em 19 de julho de 2019, conferindo ao autor o título de Mestre em Ciências em Engenharia de Produção.

Banca Examinadora:
Prof. Dr. Fernando Luiz Cyrino Oliveira (PUC-Rio)
Prof. Dr. Antonio Fernando Branco Costa (UNIFEI)
Prof. Dr. Rafael Coradi Leme (Orientador)

Itajubá
Julho 2019

Dedido este trabalho especialmente para minha mãe Inês Chilelli Alves e meu pai, Edson Inácio da Silva, os quais me forneceram a base moral para sempre buscar um mundo melhor, os quais também sempre me incentivaram ao desenvolvimento intelectual. Dedico este trabalho também a todos os pesquisadores, os quais mesmo em meio a dificuldades e frustações, nunca abandonam a infindável busca pela verdade

AGRADECIMENTOS

Agradeço primeiramente ao professor Dr. Rafael Coradi Leme por me guiar nesta jornada de descobrimentos, sempre estimulando os questionamentos e a busca por soluções que fomentassem a contínua busca pela verdade, pilastra central do método científico. Agradeço também ao professor Dr. Edson de Oliveira Pamplona por permitir meus primeiros contatos com o meio acadêmico do curso de Pós-Graduação em Engenharia de Produção da Universidade Federal de Itajubá, contato o qual me permitiu desenvolver minhas ideias e conceitos, tornando possível chegar ao ponto derradeiro desta jornada.

Sou grato à CAPES e ao MEC que por meio da sua bolsa de fomento permitiram que eu me dedicasse integralmente às atividades de pesquisa, e pudesse empurrar os horizontes deste projeto além do que seria possível em outras circunstâncias. Agradeço também aos meus amigos e companheiros da Pós-Graduação, Fabrício Almeida, Rodrigo Leite, Laila Alves, entre outros, os quais, por meio de sua experiência e comprometimento, permitiram que eu mudasse linhas de pensamento e abordagens, estas mudanças foram fundamentais para o desenvolvimento deste trabalho. Agradeço aos amigos os quais a vida me aproximou durante o tempo, por seu incentivo e confiança no meu trabalho, sempre me empurrando em busca de algo novo. Agradeço à Ana Carolina F Nogueira por toda ajuda e incentivo durante o tempo de projeto, sempre confiando na minha capacidade, mesmo nos momentos em que eu mesmo duvidava delas, seu companheirismo foi crucial para o prosseguir desta pesquisa.

Agradeço aos membros convidados da banca, por disponibilizarem seu tempo e conhecimento para apresentação deste trabalho.

Por último, agradeço novamente a meus pais, por permitirem por meio de seu trabalho e esforço, que eu chegasse até aqui.

... um cientista deve ser como uma criança. Se ele vê algo, deve dizer o que está vendo, independentemente daquilo ser o que ele imaginava ver ou não. Ver primeiro, testar depois. Mas sempre ver primeiro. Senão você só vai ver o que você espera ver."

Douglas Adams - Até mais, e obrigado pelos peixes!

RESUMO

A metodologia Gini foi apresentada há mais de um século pelo estatístico italiano Corrado Gini, durante quase um século, ela foi aplicada para medir a desigualdade na distribuição de renda. Há cerca de 3 décadas, trabalhos dissertando sobre sua utilização em outras áreas e sobre sua capacidade na estimação de parâmetros e na modelagem de séries que apresentam distribuições as quais fogem da normalidade começaram a ser publicados. Com base nestes trabalhos novos métodos começaram a ser apresentados e vêm ganhando maturidade ao longo do tempo. A intenção deste trabalho foi desenvolver um modelo capaz de aplicar os métodos presentes na literatura pertinente para estimar os parâmetros de modelos para séries Auto Regressivas cujas distribuições adjacentes fogem da normalidade apresentam características de caudas longas, com grande presença de valores extremos. Para avaliar o desempenho destes modelos embasados nos métodos de Gini, foram também estimados modelos Auto Regressivos pela métodologia clássica, fundamentados pelos Mínimos Quadrados e critérios de seleção de melhor modelo pelo metodo de Critério de Seleção de Akaike. Os resultados mostraram que o emprego da metodologia Gini na modelagem de tais séries apresenta grandes vantagens e superioridade na previsão da série, permitindo estimar modelos por meio das correlações de Gini e que fornecem melhores previsões tanto em modelos de ordem igual a um ou dois, quanto em modelos com ordem superior à dois.

Palavras-chaves: Séries Temporais, Previsão, Modelagem.

ABSTRACT

The Gini Methodology was present more than one century ago by the Italian Statistician Corrado Gini, for almost 80 years it was used to measure income distribution and inequality. Around 3 decades ago, some researchers published works about the use of the Gini in other areas of study and its capacity on parameters estimation and modelling series with distributions departing from normality. Based on these works, new methods started to arise e have become more mature over time. The aim of this work is to develop a model capable of applying the methods found on the literature to estimate the parameters of Auto Regressive models with heavy-tailed underlying distributions and extreme values. To evaluate the performance of these models the results were compared to models created using the classical Auto Regressive estimation, through the Ordinary Least Squares Method and the Akaike model selection criteria. The results showed that the Gini on time series modelling presents great advantages and superiority on the forecast, through its correlations and other metrics is possible to conclude the Gini was a good estimator on models with low order, such as one or two, but also on model with order higher than two.

Key-words: Time Series, Forecast, Modelling

LISTA DE ILUSTRAÇÕES

Figura 4.1 – Funções de Densidade de Probabilidade Log-Normal com $\theta = 0$ para valores selcionados de ω^2. Fonte: (MONTGOMERY, 2008, pág. 146) 33
Figura 4.2 – Exemplo de distribuição Pareto. 34
Figura 4.3 – Exemplo de distribuição Weibull. 35
Figura 4.4 – Exemplo de distribuição Bimodal. 35
Figura 4.5 – Exemplo de distribuição Burr. 36

LISTA DE TABELAS

Tabela 3.1 – Valores, ranks e distribuição cumulativa de uma variável X hipotética. 19
Tabela 3.2 – Valores, ranks e distribuição cumulativa de uma variável Y hipotética. 19

Tabela 4.1 – Valores dos parâmetros aleatórios e utilizados para gerar os modelos Auto Regresivos. 37

Tabela 5.1 – MAPE dos modelos de séries Bimodais 42
Tabela 5.2 – MSD dos modelos de séries Bimodais 43
Tabela 5.3 – MAE dos modelos de séries Bimodais 43
Tabela 5.4 – MSE dos modelos de séries Bimodais 44
Tabela 5.5 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Bimodais. 44
Tabela 5.6 – Coeficientes de Gini dos resíduos nos modelos com séries Bimodais. 45
Tabela 5.7 – MAPE dos modelos de séries Weibull 46
Tabela 5.8 – MSD dos modelos de séries Weibull 46
Tabela 5.9 – MAE dos modelos de séries Weibull 46
Tabela 5.10 – MSE dos modelos de séries Weibull 47
Tabela 5.11 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Weibull. 47
Tabela 5.12 – Coeficientes de Gini dos resíduos nos modelos com séries Weibull. . 47
Tabela 5.13 – MAPE dos modelos de séries Pareto 48
Tabela 5.14 – MSD dos modelos de séries Pareto 49
Tabela 5.15 – MAE dos modelos de séries Pareto 49
Tabela 5.16 – MSE dos modelos de séries Pareto 50
Tabela 5.17 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico. 50
Tabela 5.18 – Coeficientes de Gini dos resíduos nos modelos com séries Pareto . . 51
Tabela 5.19 – MAPE dos modelos de séries Log-Normal 52
Tabela 5.20 – MSD dos modelos de séries Log-Normal 52
Tabela 5.21 – MAE dos modelos de séries Log-Normal 52
Tabela 5.22 – MSE dos modelos de séries Log-Normal 53
Tabela 5.23 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Log-Normal. 53
Tabela 5.24 – Coeficientes de Gini dos resíduos nos modelos com séries Log-Normal. 53
Tabela 5.25 – MAPE dos modelos de séries Burr 55
Tabela 5.26 – MSD dos modelos de séries Burr 55
Tabela 5.27 – MAE dos modelos de séries Burr 56

Tabela 5.28 – MSE dos modelos de séries Burr 56
Tabela 5.29 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Burr. 57
Tabela 5.30 – Coeficientes de Gini dos resíduos nos modelos com séries Burr . . . 57

LISTA DE ABREVIATURAS E SIGLAS

ACF	Função de Auto Correlação *Auto Correlation Function*
PACF	Função de Auto Correlação Parcial *Partial Auto Correlation Function*
GMD	Diferença Média de Gini *Gini Mean Difference*
iid	independente e identicamente distribuído
OLS	Método dos Mínimos Quadrados
AR	Processo Auto Regressivo
MAPE	Erro Percentual Absoluto Médio (*Mean Absolute Percente Error*)
MSD	Desvio Médio Quadrado (*Mean Squared Deviation*)
MAE	Erro Médio Absoluto (*Mean Absolute error*)
MSE	Erro Médio Quadrado (*Mean Squared Error*)
MPE	Erro Percentual Médio (*Mean Percentage Error*)
ANOVA	Análise de Variância (*Analysis Of Variance*)
Gini Min/Min	Método Gini de Minimização do Resíduo
Gini SP/SP	Método Gini Semi-Paramétrico
Clássico/C	Método Clássico de estimação dos parâmetros AR
co-Ginis	Covariâncias de Gini

LISTA DE SÍMBOLOS

$\sum$	Letra grega sigma maiúsculo representando somatório
Π	Letra grega pi maiúsculo representando produtório
μ	Letra grega mi representando a média
σ	Letra grega sigma minúsculo representando desvio padrão
t_x	Representa o tempo t em algum momento x positivo.
y_t	Observação no tempo t.
ln	Logaritmo natural ou neperiano
γ	Letra grega gamma representando correlação Gini
e_t	Resíduo
$\hat{y}_t$	Valor estimado pelo modelo para y_t
ϵ	Ruído branco
β	Parâmetros daa regressão
F_x	Função cumulativa de densidade
cov	Covariância
ϕ	Parâmetros de um proceso linear
F_{z_t}	Operador *Forward*
B_{z_t}	Operador *Backward*.
a_t	Observações da série auto regressiva no tempo t.
var	variância
$F(X)$	Função cumulativa dos *ranks* de x
$\hat{G}$	Variável relacionado à metodologia Gini
R_{e_i}	Função cumulativa dos *ranks* dos resíduos do modelo
$E\|x\|$	Valor esperado de x
∞	infinito

Δ	Letra grega Delta representando variação
$\int$	Símbolo que representa a operação de integração.
γ_k	Função de Auto Covariância

SUMÁRIO

Agradecimentos v

Resumo vii

Abstract viii

Lista de Ilustrações viii

Lista de Tabelas ix

Lista de Abreviaturas e Siglas xii

Lista de Símbolos xiv

1 Introdução 1

1.1 Introdução 1

1.2 Considerações Iniciais 1

1.3 Motivação do Trabalho 3

1.3.1 Similaridades Entre o GMD e a Variância 3

1.4 Objetivo 4

1.5 Contribuição do Trabalho 4

1.6 Estrutura do Trabalho 5

2 Séries Temporais 7

2.1 Considerações Iniciais 7

2.2 Séries Temporais 7

2.3 Previsões 7

2.4 Elementos Estatísticos e Numéricos das Séries Temporais 8

2.4.1 Estacionaridade de Séries Temporais 8

2.4.2 Ruído Branco 9

2.4.3 Funções de Autocovariância e Autocorrelação 9

2.5 Análise de Regressão e Previsões 9

2.5.1 Mínimos Quadrados em Modelos de Regressão 10

2.5.2 Regressões Múltiplas 10

2.6 Processos Auto Regressivos 10

2.6.1 Processos Auto Regressivos de Primeira e Segunda Ordem 11

2.6.2 Função de Auto Correlação Parcial (PACF) 13

3 Metodologia Gini 14
3.1 Considerações Iniciais 14
3.2 Gini e o GMD 14
3.3 Representações da Diferença Média de Gini 15
3.3.1 Formulação baseada em integrais da Distribuição Cumulativa . . . 15
3.3.2 Formulação baseada na covariância 16
3.4 Aplicações do GMD 17
3.4.1 Correlação Gini 19
3.5 Regressão Gini Simples 21
3.5.1 A Representação Semi-Paramétrica 21
3.5.2 Regressão Gini por meio de Minimização 22
3.6 Regressões Gini Múltiplas 23
3.6.1 Abordagem Semi-Paramétrica 23

4 Metodologia Experimental e Procedimentos Descritivos 25
4.1 Considerações Iniciais 25
4.2 Metodologia Gini Aplicada à Séries Temporais 25
4.2.1 Auto Covariância Gini 25
4.2.2 Função de Auto Correlação de Gini (ACF de Gini) 26
4.2.3 Função de Auto Correlação Parcial de Gini (PACF de Gini) 27
4.2.4 Sistema de Covariâncias 28
4.2.4.1 O sistema Gini-Yule-Walker 30
4.3 Séries Sintéticas 30
4.3.1 Distribuições Com Caudas Longas 31
4.3.2 Séries Log-Normais 32
4.3.3 Séries Pareto 33
4.3.4 Séries Weibull 34
4.3.5 Séries Bimodais 35
4.3.6 Séries Burr 36
4.4 Transformação dos dados em Séries Temporais 36
4.5 Seleção de Ordem do Modelo 37
4.6 Avaliação do Modelo 38
4.6.1 Coeficiente de Gini 39
4.7 Modelo Computacional 39
4.7.1 Fluxo do Modelo 40

5 Análise dos Resultados 41
5.1 Resultados 41
5.2 Modelos gerados a partir de Séries com distribuições Bimodais 42
5.2.1 Desempenho dos modelos 42

5.2.2 Considerações dos resultados para Séries Bimodais 45
5.3 Modelos gerados a partir de Séries com distribuições Weibull 45
5.3.1 Resultados dos Modelos 45
5.3.2 Considerações dos Resultados para Séries Weibull 48
5.4 Modelos gerados a partir de Séries com distribuições Pareto 48
5.4.1 Considerações dos resultados para Séries Pareto 51
5.4.2 Modelos gerados a partir de Séries com distribuições Log-Normal . 51
5.4.3 Resultados dos Modelos 51
5.4.4 Considerações dos resultados para Séries Log-Normal 54
5.5 Modelos gerados a partir de Séries com distribuições distribuições Burr . . 54
5.5.1 Resultados dos modelos 54
5.5.2 Considerações dos resultados para Séries Burr 57

6 Conclusão 59

Conclusão 59

Referências 61

Apêndices 66

APÊNDICE A Códigos Originais do Modelo Computacional 67
A.1 Modelo Computacional 67
A.1.1 FuncCreateMatrixRank - Transformações nos Dados 68
A.1.2 FuncsiGACFOne e FuncsiGACFTwo - ACF de Gini 68
A.1.3 FuncCalcTheGiniCoefs - PACF de Gini 69
A.1.4 FuncCoefGini - Correlação Gini 70
A.1.5 Coeficiente e Correlações Gini 70
A.1.6 FuncFindBestARIMA - Seleção do Modelo 71
A.1.7 FuncGiniPureAR - Gini Minimização 73
A.1.8 FuncGiniARSemiParametric - Gini Semi-Paramétrico 74
A.1.9 FuncErrorMetrics - Cálculo dos Erros 75
A.1.10 FuncNormalityResiduals - Teste de Normalidade 77
A.1.11 Gráficos de ACF e PACF 77
A.1.12 Gráficos de ACF e PACF 80

1 INTRODUÇÃO

1.1 INTRODUÇÃO

Este capítulo tem como objetivo apresentar a proposta principal deste trabalho, assim como o porque de sua escolha. Com base nos paralelos com a metodologia clássica de previsão, apresentamos oportunidades de estudo com base nos métodos recentemente apresentados na literatura para utilização de técnicas estatísticas desenvolvidas no século passado e apresentadas por Corrado Gini para desenvolver estudos de estimação de parâmetros em séries temporais. Além disso, busca-se familiarizar o leitor com o tema em questão, descrever a proposta do trabalho, justificar sua importância e apresentar suas limitações.

1.2 CONSIDERAÇÕES INICIAIS

Em qualquer campo de estudo o qual se preocupa em coletar e analisar dados, existem duas medidas pincipais as quais são sempre computadas e interpretadas: medidas de centralidade ou localização, e medidas de dispersão ou variabilidade (CARCEA, 2018). Uma das mais utilizadas medidas de dispersão é a variância, ou desvio padrão. Junto com a variância existem outras medidas, tais como, desvio absoluto em relação à mediana, desvio médio absoluto, etc. De todas as medidas, variância e desvio padrão são as mais utilizadas hoje, devido a sua facilidade de implementação.

A literatura atual relacionada ao estudo das séries temporais, medem dispersão e associação, majoritariamente, por meio da variância e covariância. Entretanto, estas análises geralmente se apoiam em premissas as quais podem ter sua validade questionada. Por exemplo, a auto covariância e a auto correlação, baseadas na covariância e correlação de uma variável contra sua versão variada no tempo assume simetria em suas variáveis como consequência da própria definição de covariância (SHELEF; SCHECHTMAN, 2016).

Qualquer tipo de modelagem (inclusive a de séries temporais), é altamente dependente das premissas do modelo. Para séries estacionárias a modelagem da função de auto correlação e da função de auto correlação parcial fornecem a descrição necessária da estrutura de dependência presente nos dados, quando os momentos de segunda ordem são finitos. De acordo com Carcea (2018) a premissa de segundos momentos não é válida para séries temporais com caudas longos, fazendo com que as funções que definem essas características não sejam mais bem definidas como entidades da população.

Exemplos clássicos para os quais essas premissas falham são, por exemplo, séries com distribuições caudas longas. Existem muitos trabalhos sendo desenvolvidos buscando

solucionar problemas com dados que exibem características tais como longos intervalos de dependência, não-linearidade e caudas longos (DAVIS; RESNICK, 1996). Comportamentos de caudas longas podem surgir de diferentes fontes, tais como em finanças, economia, telecomunicações entre outras áreas. Feigin e Resnick (1999) ressaltam o fato de que a correta modelagem de tais conjuntos de dados é importante, pois as simulações efetivas de modelos complexos em larga escala demandam a modelagem precisa de dados com caracaterísticas caudas longas quando estes são a entrada do modelo.

Recentemente, algumas alternativas vêm surgindo para contornar o problema com premissas de ordem superior a 1 (um) na modelagem de tais séries. Uma alternativa de interesse foi a **Diferença Média de Gini** (do inglês *Gini Mean Difference - GMD*), tal como uma medida de dispersão, a qual foi introduzida por Serfling (2010). Outros autores, tais como Shelef e Schechtman (2011) e Shelef (2016) também trabalharam sobre esta abordagem, mostrando como a metodologia pode beneficar os métodos de previsão. Estes conceitos também podem ser extendidos para a covariância e correlação.

A metodologia Gini, como será referenciada neste trabalho, é uma metodologia baseada na classificação (*ranks*), a qual leva em conta não só as observações em si, mas também a posição ocupada por elas na distribuição cumulativa de probabilidade. o GMD compartilha muitas propriedades com a variância e difere em outras (YITZHAKI, 2003). Ambos os índices utilizam todas as observações, de maneira que nada é perdido, porém o GMD é menos sensível a observações extremas. Devido a este fato Serfling (2010) afirma que o GMD pode ser mais apropriado para modelagem de séries com caudas longos.

O GMD foi primeiramente introduzido por Corrado Gini em 1912 em seu livro *Variabilità e Mutabilità* como uma medida alternativa de variabilidade. O GMD e os parâmetros que são derivados dele (tais como a correlação Gini) tem sido utilizados nas áreas de distribuição de renda por mais de um século (SHLOMO; EDNA, 2013). A utilização dos métodos e Gini são justificados quando o investigador não está preparado para impor, sem qualquer questionamento, o conveniente mundo da normalidade. Quando a distribuição adjacente não é multivariada normal os métodos apresentados por Gini são capazes de revelar quando algumas relações entre variáveis aleatórias são simétricas ou não, assim como quando a população é estratificada e a qual extensão.

No trabalho de Yitzhaki (2003) o autor argumenta que o GMD é muito mais informativo para distribuições não-normais, quando comparado à variância. Enquanto a variância é superior para distribuições que satisfazem a normalidade, o GMD pode ser mais informativo em relação às distribuições subjacentes da população.

1.3 MOTIVAÇÃO DO TRABALHO

A vasta aplicabilidade das técnicas de previsão em séries temporais para os mais diversos problemas do mundo real, como previsão de demanda energética, variações de oferta, custos de ativos, previsão de chuvas, planejamento estratégico e muitas outras áreas, torna essencial o estudo do desempenho das previsões nos mais diversos cenários de distribuições estatísticas.

Um destes cenários engloba um conjunto específico de distribuições estatísticas, as quais possuem uma densidade elevada de valores extremos, chamadas de caudas longas. Um exemplo deste tipo de distribuição seria no campo de ciências atuárias, onde este tipo de dado aparece em modelos de perda, tal como é apresentado no trabalho de Klugman *et al.* (2012), e em finanças médicas Rosenberg *et al.* (2007).

Encontrar um solução viável para a problemática de previsão em séries com caudas longas é a motivação principal deste trabalho, o qual investiga a contribuição do GMD para estimar os parâmetros em modelos Auto Regressivo para tais séries. A metodologia Gini vem como uma opção devido à sua fácil aplicabilidade e construção por meio de técnicas baseadas na covariância e já consolidadas na literatura, como as funções de Auto Correlação, Covariância e Correlação Parcial. Essa construção é possível pois os métodos Gini compartilham características com as técnicas de Minímos Quadrados e covariância, comumente utilizadas no estudo de séries temporais.

Para estudar o desempenho dos métodos Gini, foram utilizadas séries sintéticas, principalmente pela capacidade de dinamizar o perfil de distribuições passíveis de estudo sem requerer a obtenção de dados reais, necessitando apenas garantir que os parâmetros estatísticos intrínsecos de tais séries sejam respeitados. A utilização de séries sintéticas e a quantidade de trabalhos que utilizam estas séries é muito grande. Nas áreas de previsão energética e disponibilidade de recursos estas séries são muito utilizadas, exemplos são os trabalhos de Gemignani *et al.* (2018), Detzel e Mine (2011), Carapellucci e Giordano (2013), Pereira e Souza (2014), os quais se utilizam de tais séries para estudas metolodogias de demanda e recorrência de oferta de recursos para a geração renovável.

Além destes motivos e como já citado antes, o Gini e os métodos clássicos compartilham varias similaridades, como apresentado na seção 1.3.1 a seguir.

1.3.1 Similaridades Entre o GMD e a Variância

A primeira similaridade entre o GMD e a variância é o fato de que ambos podem ser escritos como covariâncias. A variância de X é $cov(X, X)$, enquanto o GMD de X é $cov(X, F(X))$. Esta similaridade serve como base para a habilidade de traduzir o universo da variância para o universo Gini.

A segunda similaridade é o fato da decomposição da variância de uma combinação linear de variáveis aleatórias ser um caso especial da decomposição do GMD da mesma combinação. A decomposição do GMD inclui alguns parâmetros extras os quais provêm informações adicionais a respeito da distribuição subjacente. Quando estes parâmetros adicionais são iguais a zero, a decomposição do GMD e da variância possuem estruturas idênticas. Esta propriedade faz o GMD adequado para teste de suposições implícitas que levam a conveniência de se utilizar a variância. Esta propriedade é também base para a afirmação colocada por Lambert e Decoster (2005) a qual alega que o "Gini revela mais".

A terceira similaridade é o fato que ambos, a variância e o GMD são baseados nas distâncias medias entre todos os pares de observações como apresentado em Daniels (1944), alternativamente, calculando a média das distâncias entre escolhas aleatórias de duas variáveis i.i.d. Entretanto, a diferença entre elas está na função de distância utilizada. Os efeitos das funções de distância nas propriedades dos indices nas regressões Gini (DANIELS, 1948), apresentadas mais a frente neste capítulo.

1.4 OBJETIVO

A pergunta principal deste trabalho é:

- Seria a metodologia Gini uma boa estimadora de parâmetros em modelos Auto Regressivos com distribuições não-normais?

Os objetivos secundário são:

- Verificar se a aplicação da metodologia Gini é viável;
- Construir um modelo de domínio público para estimação de parâmetros pela metodologia Gini;
- Verificar possíveis pontos fortes/fracos do método.

1.5 CONTRIBUIÇÃO DO TRABALHO

Neste trabalho buscamos investigar as respostas da metodologia Gini aplicada para estimação de parâmetros em séries temporais com distribuições não-normais e caudas longas, tais como Pareto, Burr, Log-Normal, entre outras. Assim como características Auto Regressivas, por meio da metodologia apresentadas nos próximos capítulos e o estabelecimento de paralelos entre o que aqui chamaremos de "metodologia clássica' de estimação de parâmetros, será possível utilizar os componentes de Auto Correlação, Auto Correlação Parcial, entre outras propriedades para obter estimativas dos coeficientes do modelo

Auto Regressivo. Uma outra característica aqui é que o Gini, por sua definição, resulta em duas estimações de correlação, uma olhando para frente (*Forward*) e outra olhando para trás (*Backward*) na série, tais diferenças entre os estimadores, quando existirem, podem fornecer informações adicionais.

Desta forma, busca-se aqui estudar qual a resposta da aplicação da metodologia Gini na estimação de tais parâmetros, seus pontos fortes e fracos, assim como quais seriam os melhores cenários para aplicação da mesma ao invés dos métodos clássicos. Busca-se também entender se alguma característica intrínseca destas distribuições é melhor estimada pelos métodos Gini e quais seriam suas aplicabilidades nos diversos contextos de utilização de séries temporais e modelos de previsão.

1.6 ESTRUTURA DO TRABALHO

O presente trabalho é estruturado em 6 capítulos, sendo eles:

1. Introdução: Com o objetivo de apresentar o tema ao leitos, assim como estabelecer algumas fundamentações e premissas básicas que justificam a escolha do tema e como ele será desenvolvido.

2. Séries Temporais: Este capítulo busca fundamentar as técnicas aplicadas que fundamentam este trabalho e que também permitem estabelecer os paralelos para que utilizemos a metodologia Gini em comparação aos métodos clássicos para entender sua performance.

3. Metodologia Gini: Nesta seção é apresentada a metodologia desenvolvida por Corrado Gini e como, ao longos dos anos, novas descobertas permitiram que ela fosse aplicada à previsão e estimação de parâmetros em séries temporais, além disso, buca-se fundamentar sua aplicação, algumas limitações, assim como pontos fortes do método.

4. Metodologia: Neste capítulo é apresentado como ligamos os métodos tradicionais de variância e covariância aos métodos Gini e como eles podem ser aplicados nas séries temporais, além disso, apresentam-se as distribuições estatísticas utilizadas na modelagem do trabalho e como é feita a estimativa de performance, ajuste e qualidade dos modelos obtidos.

5. Resultados: Após a apresentação de toda metodologia e teoria necessária para a construção dos modelos utilizados neste trabalho, são apresentados e discutidos os resultados obtidos da modelagem.

6. Conclusão: Por fim, analiza-se a performance do método utilizada e discute-se quais são as oportunidades abertas por este trabalho e algumas ideias de como proceder em trabalhos futuros para estimação de parâmetros e aplicação da metodologia.

2 SÉRIES TEMPORAIS

2.1 CONSIDERAÇÕES INICIAIS

Neste capítulo, apresenta-se a fundamentação teórica sobre o processo de estudo e análise de séries temporais, além dos métodos relacionados à estimação de parâmetros e quais os procedimentos são necessários para que uma correta análise seja direcionada ao conjunto de dados em estudo. Este capítulo é importante para a compreensão de como a metodologia Gini poderá ser empregada para melhorar a capacidade preditiva dos métodos clássico já estabelecidos.

2.2 SÉRIES TEMPORAIS

Uma série temporal é uma coleção de dados gravados durante um período de tempo, tal como segundos, minutos, horas. Existem diversos exemplos de séries temporais: séries diárias de ventos, variações de preço de ações na bolsa de valores, uma reação química ao longo do tempo. Exemplos de séries temporais existem nos mais diversos campos do conhecimento, tais como economia, engenharia, negócios, ciências naturais e ciências sociais. Neste trabalho, a atenção é dada análises de métodos voltados a promoção da capacidade de previsão de modelos matemáticos aplicados para tais séries.

Uma parte importante na análise de séries temporais é a seleção de um modelo adequado para o modelo probabilístico (ou uma classe de modelo) dos dados. Para proceder com a natureza imprevisível das observações futuras é comum supor que cada observação yt é uma realização de certo valor agregado para a variável aleatória Y_t no tempo.

2.3 PREVISÕES

Uma previsão é uma predição de algum evento futuro. A previsão de séries temporais concentra-se em identificar eventos futuros baseados em eventos conhecidos, usualmente gravados em períodos uniformes de tempo (ZHENG; KUSIAK, 2009). Os problemas de previsão são comumente classificados como de curto, médio e longo prazo. De acordo com Montgomery (2008) previsões de curso prazo envolvem eventos com apenas alguns períodos (dias, semanas ou meses) no futuro. Eventos de médio prazo se estendem a um ou dois anos no futuro, enquanto eventos de longo prazo podem se estender por muitos anos. Previsões de curto e médio prazo são requeridas em atividades variando de gestão operacional a planejamento de finanças e seleção de novas pesquisas e projetos de desenvolvimento. Previsões de longo prazo impactam atividades como planejamento estratégico. As previsões de curto e médio prazo são baseadas, tipicamente, em identificar,

modelar e extrapolar padrões encontrados em dados históricos. Devido ao fato destes dados históricos manterem inércia e não variarem drasticamente ao longo dos anos, métodos estatísticos são muitos úteis em sua previsão.

Um conceito de grande importância para análise de séries temporais aplicadas através da metodologia Gini, o conceito de erro de previsão. A diferença entre a observação yt e o valor obtido pelo ajuste de um modelo de série temporal ao dado, ou valor do ajuste y^t, definido acima, é chamado de resíduo e é denotado por

$$\hat{e}_t = y_t - \hat{y}_t \tag{2.1}$$

2.4 ELEMENTOS ESTATÍSTICOS E NUMÉRICOS DAS SÉRIES TEMPORAIS

Nesta seção serão abordados os conceitos estatísticos fundamentais para que as análises desenvolvidas neste trabalho possam ser abordadas com objetividade e clareza, desta maneira, apresentar-se-á de forma sucinta a teoria necessária, referenciando-se algumas demonstrações.

2.4.1 Estacionaridade de Séries Temporais

Um tipo de série temporal muito importante é conhecido como estacionário. De maneira simples, uma série temporal $Y_t, t = 0, \pm1, ...,$ é dita estacionária caso tenha propriedades estatísticas similares quando "deslocamos a série no tempo" $X_{t+h}, t = 0, \pm1, ...,$ para cada inteiro h (BROCKWELL, 2002). Restringindo a atenção para o fato dessas propriedades dependerem apenas dos momentos de primeira e segunda ordem (aqui está uma das vantagens do Gini, já que o método não precisa dos momentos de segunda ordem). Uma série temporal é dita estritamente estacionária quando suas propriedades não são drasticamente alteradas com o deslocamento no tempo. Outra propriedade destas séries é que a distribuição conjunta de probabilidade entre $(X_1, ..., X_n)$ e $(X_{1+h}, ..., X_{n+h})$ para todos os inteiros h e $n > 0$ é a mesma.

A série temporal também pode ser chamada de fracamente estacionária, este caso ocorre quando

- $u_x(t)$ é independente de t
- $y_x(t + h, t)$ é independente de t para cada h.

Uma abordagem mais profunda deste fenômeno pode ser encontrado no livro de Montgomery (2008).

2.4.2 Ruído Branco

O exemplo mais fundamental de um processo estacionário é uma sequência de variáveis aleatórias $i.i.d$ denotadas, por exemplo, como $a_1, ..., a_t$ o qual assumimos possuir média zero e variância σ_a^2. Este processo é estritamente estacionário e referido como ruído branco. Devido à independência, podemos assumir que a_t é não correlacionado e sua função de autocovariância é simplesmente

$$\gamma_k = E\left[a_t a_{t+k}\right] = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{2.2}$$

2.4.3 Funções de Autocovariância e Autocorrelação

A premissa de estacionaridade também implica que a distribuição conjunta de probabilidade $p(z_{t_1}, z_{t_2})$ é a mesma para todos os tempos t_1, t_2, os quais são constantes nos intervalos. Desta forma, a covariância entre os valores z_t e z_{t+k} separados por um intervalo k, conhecido como *lag*, deve ser o mesmo para todos t sob a premissa de **estacionaridade**. Esta covariância será chamada de autocovariância no *lag* k e é definida por

$$\gamma_k = \text{cov}\left[z_t, z_{t+k}\right] = E\left[\left(z_t - \mu\right)\left(z_{t+k} - \mu\right)\right] \tag{2.3}$$

Similarmente, a autocorrelação no *lag* k é

$$\begin{aligned} \rho_k &= \frac{E\left[\left(z_t - \mu\right)\left(z_{t+k} - \mu\right)\right]}{\sqrt{E\left[\left(z_t - \mu\right)^2\right] E\left[\left(z_{t+k} - \mu\right)^2\right]}} \\ &= \frac{E\left[\left(z_t - \mu\right)\left(z_{t+k} - \mu\right)\right]}{\sigma_z^2} \end{aligned} \tag{2.4}$$

Lembrando que, para um processo estacionário, a variância $\sigma_z^2 = \gamma_0$ é a mesma no tempo $t+k$ e em t. Assim, a autocorrelação no *lag* k, isto é, a correlação entre z_t e z_{t+k}, é

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{2.5}$$

conceitos mais detalhados do desenvolvimento dessas funções podem ser encontrados em Box *et al.* (2008).

2.5 ANÁLISE DE REGRESSÃO E PREVISÕES

A análise de regressão é uma técnica estatística para modelar e investigar as relações entre uma variável de saída ou resposta e uma ou mais variáveis preditoras ou regressoras. O resultado final do estudo de regressão é, geralmente, um modelo utilizado

para prever valores futuros de uma variável de resposta, dados valores específicos da(s) variável(is) preditora(s).

2.5.1 Mínimos Quadrados em Modelos de Regressão

Um dos grandes paralelos entre a metodologia de Gini e os métodos clássicos baseados em momentos de primeira e segunda ordem, é a estrutura de estimação dos parâmetros do modelo. Neste contexto é introduzido o Método dos Mínimos Quadrados, do inglês Ordinary Least Squares (OLS). Considerando (Y, X) uma variável bivariada aleatória com função de densidade $f(y, x)$, assumimos $f(x), F_x$, μ_x e γ_x^2 denotando a densidade, a distribuição cumulativa, o valor esperado e a variância de X, respectivamente. Assumindo também que os momentos de primeira e segunda ordem existem. Temos $g(x) = EY|X = x$ como a curva de regressão, onde $g'(x)$ é o coeficiente angular definido como

$$\mathrm{g'(x)} = \frac{\partial \mathrm{E\{Y|X = x\}}}{\partial \mathrm{x}} \tag{2.6}$$

A prova deste conceito pode ser encontrada em Shlomo e Edna (2013, pág. 136).

2.5.2 Regressões Múltiplas

O objetivo das regressões múltiplas é estudar as relações entre diversas variáveis explanatórias e a variável dependente. A extensão do modelo de uma variável explanatória em várias traz consigo uma série de complicações. Por exemplo, na regressão múltiplas é preciso considerar os efeitos das relações entre as variáveis explanatórias nos estimadores.

Um modelo de regressão múltipla pode ser apresentado, de forma simples, como

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \tag{2.7}$$

onde Y é a variável dependente e x_1 e x_2 são dois regressores que representam variáveis que influenciam na resposta Y, ε é o termo de erro. O termo linear é utilizado pois 2.7 é uma função linear dos parâmetros β_0, β_1 e β_2.

Na regressão simples de X_k, com T_k, $k, j = 1, ..., K$, os coeficientes de regressão são

$$\beta_{\varepsilon \mathrm{j}} = \frac{\mathrm{cov}\,(\varepsilon, \mathrm{T_j})}{\mathrm{cov}\,(\mathrm{X_j}, \mathrm{T_j})}, \quad \beta_{\mathrm{kj}} = \frac{\mathrm{cov}\,(\mathrm{X_k}, \mathrm{T_j})}{\mathrm{cov}\,(\mathrm{X_j}, \mathrm{T_j})} \tag{2.8}$$

A dedução deste resultado pode ser encontrado explicitamente no livro de Shlomo e Edna (2013, pág. 179).

2.6 PROCESSOS AUTO REGRESSIVOS

Considere a representação de um processo, possível apenas se um número finito de pesos π forem diferentes de zero, por exemplo, $\pi_1 = \phi_1, \pi_2 = \phi_2, \ldots, \pi_p = \phi_p$, and $\pi_k = 0$

para $k > p$, então o processo resultante é chamado de modelo autoregressivo de ordem p (WEI, 2006, pag. 53). Esse processo auto regressivo é representado pela sigla $AR(p)$ e dado por

$$\begin{gathered} \tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \\ \text{ou} \\ (1 - \phi_1 B - \cdots - \phi_p B^p)\, \tilde{z}_t = \phi(B) \tilde{z}_t = a_t \end{gathered} \tag{2.9}$$

onde p é chamado de "ordem" do processo. Um processo auto regressivo de primeira ordem ($p = 1$) pode ser representado por

$$\begin{gathered} (1 - \phi_1 B)\, \tilde{z}_t = a_t \\ \text{também descrito por} \\ \tilde{z}_t = (1 - \phi_1 B)^{-1} a_t = \textstyle\sum_{j=0}^{\infty} \phi_1^j a_{t-j} \end{gathered} \tag{2.10}$$

Essa representação provém que uma série infinita a direita converge, segundo um senso apropriado. Consequentemente,

$$\psi(B) = (1 - \phi_1 B)^{-1} = \sum_{j=0}^{\infty} \phi_1^j B^j \tag{2.11}$$

Um conceito essencial para a análise da regressão Gini é o conceito de invertibilidade. A invertibilidade deve ser tratada com atenção nos processos auto regressivos pois ela é uma condição necessária quando precisamos associar eventos presentes com acontecimentos passados de uma maneira sensível. Está condição está diretamente ligada ao estudo da metodologia Gini e seus dois coeficientes de correlação que podem proporcionar uma análise para frente (*Forward*) e para trás (*Backward*) na série. As definições e propriedades necessárias são apresentadas em Box *et al.* (2008, pág. 56)

2.6.1 Processos Auto Regressivos de Primeira e Segunda Ordem

Um processo auto regressivo de primeira ordem é representado por

$$\begin{aligned} \tilde{z}_t &= \phi_1 \tilde{z}_{t-1} + a_t \\ &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \end{aligned} \tag{2.12}$$

onde $|\psi| < 1$, como já apresentado anteriormente para que o processo seja estacionário.

Um processo auto regressivo de segunda ordem $AR(2)$ é a extensão de 2.12 pela inclusão do termo z_{t-2}

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \tag{2.13}$$

A função de auto correlação pode ser dada por

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k > 0 \tag{2.14}$$

com valores iniciais $\rho_0 = 1$ e $\rho_1 = \phi_1/(1-\phi_2)$.

Uma maneira prática de representar tais processos uma vez que possuímos a função de auto correlação, é através das equações de Yule-Walker. Consideremos a equação a seguir como uma função de correlação que satisfaz a forma de equação de diferenças sem os choques aleatórios (ruído branco).

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \quad k > 0 \tag{2.15}$$

Quando substituímos $k = 1, 2, ..., p$ na equação acima, obtemos um conjunto de equações lineares para ϕ_1, ϕ_2, ϕ_p em termos deρ_1, ρ_2, ρ_p, isto é

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_{p2} + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p
\end{aligned} \tag{2.16}$$

Para obter os estimadores dos parâmetros, substituímos as auto correlações teóricas ρ_k pela auto correlação estimada ρ_k. Escrevendo

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad \mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \tag{2.17}$$

A solução de 2.16 para os parâmetros ϕ em termos de auto correlações podem ser escritas como

$$\phi = \mathbf{P}_p^{-1}\rho_p \tag{2.18}$$

Substituindo $p = 2$ em 2.16, as equações de Yule-Walker são

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 \\
\rho_2 &= \phi_1\rho_1 + \phi_2
\end{aligned} \tag{2.19}$$

as quais, quando resolvidas para $\phi_1 e \phi_2$, resultam em

$$\begin{aligned}
\phi_1 &= \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\
\phi_2 &= \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}
\end{aligned} \tag{2.20}$$

As equações 2.19 podem ser resolvidas para expressar ρ_1 e ρ_2 em termos de ϕ_1 e ϕ_2, resultando em

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1-\phi_2} \\ \rho_2 &= \phi_2 + \frac{\phi_1^2}{1-\phi_2} \end{aligned} \tag{2.21}$$

2.6.2 Função de Auto Correlação Parcial (PACF)

A princípio, é comum não sabermos a ordem de um processo auto regressivo para ajustar uma série temporal observada. A função de auto correlação parcial (do inglês Partial Autocorrelation Function – PACF) é um dispositivo que explora o fato de, onde houver função de auto correlação infinita em extensão para um processo AR(p), a mesma pode, por sua natureza ser descrita em termos de p funções não nulas de auto correlações.

A função de auto correlação parcial pode ser derivada da forma como é apresentado em Brockwell (1991),

$$\rho_j = \phi_{k1}\rho_{j-1} + \cdots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \quad j = 1, 2, \ldots, k \tag{2.22}$$

Através do conjunto de equações de Yule-Walker, 2.22 pode ser reescrita de forma compacta como em 2.18

Uma outra maneira de entender a função de auto correlação parcial é como a representação da auto correlação entre ρ_t e ρ_{t+k} após a dependência linear mutua causada nas variáveis intermediarias $\rho_{t+1}, \rho_{t+2}, ..., \rho_{t+k-1}$ ter sido removida. Essa representação pode ser dada por

$$\text{Corr}\left(\rho_t, \rho_{t+k} | \rho_{t+1}, \ldots, \rho_{t+k-1}\right) \tag{2.23}$$

Para um processo auto regressivo de ordem p, uma função de auto correlação parcial ϕ_{kk} será não nula para k menor ou igual a p e zero para k maior que p. A literatura comumente chama essa relação de *cutoff* para *lags* maiores que p. A função de auto correlação parcial ϕ_{kk} dada por 2.18 é definida para qualquer processo estacionário em função das auto correlações ρ_k. Mais detalhes sobre suas propriedades e limitações podem ser encontradas em Box *et al.* (2008, pág. 67).

3 METODOLOGIA GINI

3.1 CONSIDERAÇÕES INICIAIS

Neste capítulo, apresenta-se a fundamentação teórica sobre a metodologia Gini e alguns exemplos de aplicações. Também são apresentados conceitos da metodologia Gini que fazem paralelo aos métodos de *Ordinary Least Square* (OLS) e como a aplicação dos métodos de covariância Gini podem ser úteis quando as distribuições estudadas apresentam distribuições caudas longas, uma vez que a metodologia Gini é menos sensível a *outliers*

O intuito principal deste capítulo está no entendimento de como funciona o método e fundamentar a teoria apresentada no próximo capítulo em que é feita a construção de modelos Auto Regressivos com base na covariância Gini.

3.2 GINI E O GMD

A chamada Diferença Média de Gini, para fins contextuais aqui chamada de GMD (*Gini Mean Difference*), foi introduzida por Corrado Gini em 1912 como uma alternativa para medidas de variabilidade. O GMD e os parâmetros que são derivados dele (tais como o Coeficiente de Gini, referido por alguns como Razão de Concentração) tem sido utilizado na área de distribuição de renda por quase um século e existem evidências de que a introdução do GMD é anterior à esta data (HARTER, 1978). Em outras áreas ocorrem aparições esporádicas deste método e suas aplicações parecem ser redescobertas sob diferentes nomes. De acordo com Shlomo e Edna (2013), o GMD pode ser representado de, pelo menos, 14 formas diferentes. Cada representação pode ser atribuída a uma interpretação única e levar, naturalmente, a diferentes ferramentas analíticas, tais como métricas L1, teorias de ordem estatística, teoria do valor extremo, curvas de concentração, entre outros. Algumas representações se mantêm apenas para valores não-negativos, enquanto outros precisam de ajustes para lidar com distribuições discretas.

Devido a esta diversidade em interpretações e seu detalhamento, a abordagem deste trabalho se limitará na análise das técnicas relacionadas à metodologia Gini em relação a imitar as técnicas baseadas na variância, por meio da reposição da variância pelo GMD e suas variantes. O objetivo desta abordagem é mostrar que grande parte das medidas de variabilidade construídas com a variância podem ser replicadas por meio da utilização do GMD. Com base nesta mentalidade, a atenção do desenvolvimento apresentado no trabalho será construída apoiando-se na representação embasada na covariância.

A utilização do GMD como uma medida de variabilidade é justificada em ocasiões

onde o investigador não está pronto para afirmar, sem questionamentos, a conveniência da normalidade. Quando a distribuição é univariada e normal, a média amostral e a variância são estatísticas suficientes para descrever suas características e o GMD é redundante. O mesmo ocorre quando é preciso lidar com distribuições multivariadas, nas quais os casos em que a normalidade é completamente distribuída por médias individuais, suas variâncias e coeficientes de correlação de Pearson. Nestes casos, o GMD e seus equivalentes nada podem adicionar ao entendimento ou análise dos dados. Entretanto, quando a distribuição foge da normalidade ou assume-se que a distribuição estudada não pode ser assumida como normal, então, como apresentado por Lambert e Decoster (2005), o GMD pode enriquecer a análise. Esse enriquecimento pode ser descrito através, por exemplo, da verificação da relação simetria ou não entre variáveis aleatórias, quando uma população é estratificada e a qual extensão, quando a premissa de linearidade na análise de regressão é suportada pelos dados.

3.3 REPRESENTAÇÕES DA DIFERENÇA MÉDIA DE GINI

A representação do GMD como medida de variabilidade já é conhecida há mais de um século. Algumas delas são válidas apenas para distribuições contínuas, enquanto outras apenas podem ser aplicadas para variáveis não-negativas.

Algumas das fórmulas para o GMD, dependendo dos elementos envolvidos são:

(A) Formulação baseada em integrais de funções de distribuição cumulativas;

(B) Formulações baseadas em covariância.

O foco deste trabalho se dará sobre estas duas formulações, as quais serão apresentadas afim de ilustrar as capacidades da metodologia de Gini.

3.3.1 Formulação baseada em integrais da Distribuição Cumulativa

Nesta seção é apresentada uma interpretação do GMD formulado através de integrais da distribuição cumulativa. A equação básica necessária para desenvolver tal representação tem fundação na expressão do valor esperado de uma distribuição. Assumimos X sendo uma variável contínua distribuída no intervalo $[a, \infty)$. Então o valor esperado de X é dado por

$$\mu = \mathrm{a} + \int_{\mathrm{a}}^{\infty} [1 - \mathrm{F(x)}]\mathrm{dx} \tag{3.1}$$

Prova: A definição padrão do valor esperado é $\mu = \int_{\mathrm{a}}^{\infty} \mathrm{xf(x)dx}$. Utilizando integração por partes com $u = x$ e $v = -[1 - F(x)]$ obtemos 3.1. Utilizando 3.1 e o fato de

que a distribuição cumulativa de um mínimo de duas variáveis idêntica e independentemente distribuídas ($i.i.d$) podem ser expressadas como $1-[1-F(x)]^2$, podemos escrever $\Delta = 2\mu - 2\mathrm{E}\left[\mathrm{Min}\left\{\mathrm{X}_1, \mathrm{X}_2\right\}\right]$ (para mais detalhes vide (SHLOMO; EDNA, 2013)) como

$$\Delta = 2\int[1-\mathrm{F(t)}]\mathrm{dt} - 2\int[1-\mathrm{F(t)}]^2\mathrm{dt} \tag{3.2}$$

e combinando as duas integrais

$$\Delta = 2\int \mathrm{F(t)}[1-\mathrm{F(t)}]\mathrm{dt} \tag{3.3}$$

Mais detalhes podem ser encontrados em Dorfman (1979). Considerando $F_n(x)$ a distribuição cumulativa empírica de X, baseada em uma amostra de n observações. Conforme apresentado por (SERFLING, 1980), para um dado x, $F_n(x)$ é a média amostral de N $i.i.d$ variáveis de Bernoulli com $p = F(x)$. A variância de $F_n(x)$ é igual à

$$\sigma^2_{\mathrm{F_n(x)}} = \mathrm{F(x)}[1-\mathrm{F(x)}]/\mathrm{n} \tag{3.4}$$

e o GMD pode ser interpretado como 2 n $\int \sigma^2_{\mathrm{F_n(x)}}\mathrm{dx}$

Uma variação similar desta fórmula é

$$\Delta = 2\mathrm{nE}\left\{\int\left[\mathrm{F_n(x)} - \mathrm{F(x)}\right]^2 \mathrm{dx}\right\} \tag{3.5}$$

a qual é o critério original de Cramer-Von Mises-Smirnov para teste de bondade de ajuste de uma distribuição. De alguma forma 3.5 pode ser visto como uma aproximação dual para os momentos centrais de uma distribuição. Momentos centrais são lineares em probabilidade e funções potência de desvios da variação em relação ao valor esperado. No GMD, a função potência é aplicada ao desvio da distribuição cumulativa do seu valor esperado enquanto a linearidade é aplicada à própria variável. Essa interpretação também sugere uma possível explicação para a robustez de tal aproximação

Finalmente, reescrevemos 3.3 como

$$\Delta = 2\int_a^\infty \left[\int_a^x \mathrm{f(t)}\mathrm{dt}\int_x^\infty \mathrm{f(t)}\mathrm{dt}\right]\mathrm{dx} \tag{3.6}$$

tal forma é como Wold (1935) a apresentou.

Uma apresentação adicional também dada por Wold (1935) a qual é valida para variáveis não negativas é

$$\Delta = 2\int_0^\infty \left[\int_0^t \mathrm{F(u)}\mathrm{du}\right]\mathrm{dF(t)} \tag{3.7}$$

3.3.2 Formulação baseada na covariância

O conceito de variância pode ser vastamente encontrado na literatura, dentre suas propriedades, vamos utilizar a qual se refere a ela como um caso especial da covariância, uma vez que ela pode ser escrita como $var(X) = cov(X, X)$. Uma vez que essas

propriedades são definidas, é possível se utilizar de outras ferramentas matemáticas para definer a correlação Gini e a decomposição do GMD de uma combinação linear de variáveis aleatórias, o que nos levará a um ponto cuminante deste trabalho, as regressões Gini e estimação dos parâmetros de series temporais por meio da metodologia Gini.

Começando com a expressão 3.3 e aplicando integração por partes a ela, com $v = F(t)[1 - F(t)]$ e $u = t$, temos que, após remover os zeros e rearranjar os termos,

$$\Delta = 2\int \mathrm{F(t)[1-F(t)]dt} = 4\int \mathrm{t(F(t)-0.5]f(t)dt} \tag{3.8}$$

Lembrando que o valor esperado de F, o qual e uniformemente distribuído em $[0, 1]$ é 0,5, podemos reescrever 3.8 como

$$\Delta = 4\mathrm{E\{X(F(X)-E[F(X)])\}} = 4\,\mathrm{cov[X, F(X)]} \tag{3.9}$$

A equação 3.9 que será amplamente utilizada neste trabalho, a qual é a representação da covariância do GMD, ela nos permite calcular o GMD usando qualquer pacote de regressão simples. Lembrando que $F(X)$ é uniformemente distríbuido em $[0, 1]$. Portanto, $cov[F(X), F(X)] = 1/2$ e o GMD pode ser apresentado como

$$\Delta = (1/3)\,\mathrm{cov[X, F(X)]/\,cov[F(X), F(X)]} \tag{3.10}$$

A equação 3.9, pode ser utilizada para mostrar que R-regressions são, na verdade, baseados na minimização do GMD dos resíduos da regressão característica a qual será abordada mais a frente neste trabalho.

3.4 APLICAÇÕES DO GMD

Dada duas variáveis aleatórias, tem-se um grande interesse em medir a correlação ou associação ou concordância entre estas variáveis (GILI; BETTUZZI, 1987). Este propósito pode ser generalizado pela ideia de um "grau de concordância" entre a ordem e a classificação de duas variáveis.

Medidas de associação são tratadas na literatura como paramétricas e não-paramétricas. Na abordagem paramétrica o método amplamente utilizado é o coeficiente de medida de correlação produto-momento de Pearson, p, o qual é baseado na covariância normalizada entre duas variáveis. No ambiente não-paramétrico a medida mais comum é o coeficiente de correlação de Spearman, o qual baseia-se na amostra normalizada da covariância entre classificações (por exemplo, distribuições cumulativas) das variáveis. O objetivo desta seção é definir a correlação Gini, a qual, de certa forma, está localizada entre as duas correlações citadas (SHLOMO; EDNA, 2013). Ela é baseada na covariância normalizada

entre uma variável e a classificação da outra variável. Vale enfatizar o fato de que existem duas correlações de Gini entre cada par de variáveis, dependendo da qual é tomada com variável e qual é classificada.

De maneira a ilustrar as propriedades de medidas de associação, é importante listar as propriedades desejadas para uma boa medida de associação:

1. Não-paramétrico: uma medida desejável de associação deve medir um conceito útil de associação, independente da distribuição subjacente.
2. Limites conhecidos: Para saber afirmar se a associação é forte ou não, são necessários pontos de referência. Pontos úteis são o limite superior, o limite inferior e um ponto médio, no qual não há associação ou independência entre as variáveis.
3. Uma medida desejável deve ser capaz de detectar relações monotônicas e não-monotônicas, assim como pontos de inflexão nas relações, caso existam (a ideia principal aqui é buscar pontos de inflexão, comuns em séries temporais).
4. Explicação intuitiva: Uma medida se torna desejável quando possui uma explicação intuitiva de seu resultado, o que permite uma fácil elucidação por meio do pesquisador.

Quando nos utilizamos de medidas de dispersão ou de associação entre variáveis, encontramos uma grande barreira para a confiabilidade de diversos métodos, o problema da simetria. A maioria das medidas de associação são simétricas. Desta maneira, essa propriedade não é razoável quando estamos trabalhando com distribuições não simétricas ou quando a relação para a qual estamos olhando for não-simétrica tal qual as análises de regressão onde os papéis da variável dependente e explanatória não forem simétricos.

Um dos maiores problemas, enfatizados neste trabalho é exatamente o fato da assimetria quando analisamos a relação entre variáveis. Uma vez que quando impusermos simetria para o coeficiente de correlação, a habilidade de utilizá-lo para o teste de simetria é perdida (SHLOMO; EDNA, 2013). Outro grande problema é relacionado à decomposição da medida de variabilidade de uma combinação linear de variáveis aleatórias. A medida de variabilidade de uma combinação linear pode ser decomposta em contribuições individuais e contribuições de diversas variáveis simultaneamente, as quais são medidas por métricas de associação. Neste caso, uma boa medida de associação permitirá a decomposição. A decomposição também permite evitar contagens duplicadas por meio da classificação de contribuições à variabilidade em contribuições individuais e aquelas que não podem ser associadas com uma variável em particular.

3.4.1 Correlação Gini

Como já explicitado antes, existem diversas maneiras de representar o GMD. Por isso, também existem diversas maneiras de representar a correlação Gini. Visando uma apresentação enxuta, a representação aqui descrita será apenas a baseada na covariância. Esta abordagem se faz conveniente para estabelecermos paralelos entre a correlação Gini e os coeficientes de Pearson e Spearman.

Em geral, a correlação Gini é baseada em uma mistura de variantes e distribuições cumulativas. Daniels (1944), Stuart (1954), Kendall (1962), Kendall (1955) e Barnett *et al.* (1976) dão exemplos de medidas baseadas em tais misturas.

Inicialmente, começamos com a definição da correlação Gini como equivalente da covariância. Lembrando que existem duas covariâncias Gini entre cada par de variáveis aleatórias. As quais serão chamadas de co-Ginis. Estas são definidas por

$$\mathrm{Gcov(X,Y)} = \mathrm{cov(X,G(Y))}; \quad \mathrm{Gcov(Y,X)} = \mathrm{cov(Y,F(X))} \tag{3.11}$$

As correlações, as quais são os co-Ginis normalizados são escritas como

$$\Gamma_{\mathrm{X,Y}} = \frac{\mathrm{cov(X,G(Y))}}{\mathrm{cov(X,F(X))}}; \quad \Gamma_{\mathrm{Y,X}} = \frac{\mathrm{cov(Y,F(X))}}{\mathrm{cov(Y,G(Y))}} \tag{3.12}$$

As tabelas 3.1 e 3.2 mostram exemplos hipotéticos de valores das variáveis X e Y e suas distribuições cumulativas ao redor da mediana (sendo a mediana o maior peso das observações).

Tabela 3.1 – Valores, ranks e distribuição cumulativa de uma variável X hipotética.

X	10	11	13	14	16	19	25	26	28	29	30
Ranks	1	2	3	4	5	6	5	4	3	2	1
F(X)	0.0278	0.0556	0.0833	0.1111	0.1389	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278

Tabela 3.2 – Valores, ranks e distribuição cumulativa de uma variável Y hipotética.

Y	14.34	15.57	18.04	19.28	21.74	25.45	32.85	34.08	36.55	37.79	39.02
Ranks	1	2	3	4	5	6	5	4	3	2	1
G(Y)	0.0278	0.0556	0.0833	0.1111	0.1389	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278

$$\mathrm{Gcov(X,Y)} = \mathrm{cov(X,G(Y))}; \quad \mathrm{Gcov(Y,X)} = \mathrm{cov(Y,F(X))} \tag{3.13}$$

Em geral essas relações não são simétricas em X e Y. Além do mais, $Gcov(X,Y)$ e $Gcov(Y,X)$ podem, até mesmo, ter sinais diferentes. Por outro lado,existem instâncias importantes do conceitos de assimetria, como por exemplo, analises de regressão e o conceito de elasticidade, onde a propriedade de assimetria é vista como uma vantagem.

Outra maneira de analisar a correlação Gini é através da representação encontrada nos trabalhos de Hoeffding (1948), Schweizer e Wolff (1981) e Schechtman e Yitzhaki (1999) que fazem relação às correlações de Pearson e Spearman. Sendo $K(X,Y)$ denotando a distribuição conjunta de X e Y, então os coeficientes de correlação Person p, Spearman r_s e Gini τ podem ser unificados através da expressão

$$\begin{aligned} \rho_{\mathrm{X,Y}} &= \frac{\iint(\mathrm{K(x,y)-F(x)G(y))dxdy}}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}} \\ \mathrm{r_{s,X,Y}} &= 12\iint(\mathrm{K(x,y)-F(x)G(y))dF(x)dG(y)} \end{aligned} \tag{3.14}$$

e

$$\Gamma_{\mathrm{X,Y}} = \frac{\iint(\mathrm{K(x,y)-F(x)G(y))dxdG(y)}}{\mathrm{cov(X,F(X))}} \tag{3.15}$$

Estas representações mostram que a correlação Gini é, na verdade, uma mistura das propriedades das correlações de Pearson e Spearman: sendo similar a Pearson em X e a Spearman em Y.

Segundo as definições encontradas em Schechtman e Yitzhaki (1999)], Schechtman *et al.* (2011), Yitzhaki (2003), e Serfling e Xiao (2007). É possível apresentar as seguintes propriedades da correlação Gini.

1. $-1 <= \Gamma_{x,y} <= 1$
2. Se Y é uma função de incremento monotônico de X, então ambos $\Gamma_{x,y}$ e$\Gamma_{y,x}$ igualam-se a $+1(-1)$;
3. Se X e Y são estatisticamente independentes então $\Gamma_{x,y} = \Gamma_{y,x}, x = 0$;
4. $\Gamma_{x,y} = -\Gamma_{x,-y} = \Gamma_{-x,y} = \Gamma_{-x,-y}$;
5. $\Gamma_{x,y}$ é invariante sob todas as transformações estritamente monotônicas de Y;
6. $\Gamma_{x,y}$ é invariante sob mudanças de escala e localização de X;
7. $\Gamma_{x,y}$ é simétrico em (X,Y) se $(aX+b, cY+d)$ é intercambiável para alguma constante a, b, c, d com a e $c > 0$;
8. Se (X,Y) seguem uma distribuição bivariada normal com parâmetros $(\mu_x, \mu_y, \Gamma_x^2, \Gamma_y^2, \rho)$ então $\Gamma_{x,y} = \Gamma_{y,x} = \rho$, onde ρ é o coeficiente de correlação de Pearson.

$$\hat{G}\,\mathrm{cor}(Y,X) = \frac{\mathrm{cov}(Y,R(X))}{\mathrm{cov}(Y,R(Y))} = \frac{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)\left(R\left(X_i\right)-\overline{R}(X)\right)/(n-1)}{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)\left(R\left(Y_i\right)-\overline{R}(Y)\right)/(n-1)} \tag{3.16}$$

3.5 REGRESSÃO GINI SIMPLES

A covariância entre a variável dependente e a variável explanatória é o conceito fundamental em uma regressão. Baseado na metodologia Gini, é possível interpretar dois métodos diferentes a partir do GMD. O primeiro método é baseado na possibilidade de expressão a covariância Gini entre as variáveis como uma soma ponderada das inclinações da curva de regressão, essa abordagem é chamada de semi-paramétrica. O segundo método é baseado na minimização do GMD dos resíduos.

3.5.1 A Representação Semi-Paramétrica

Nesta abordagem a estimação dos parâmetros baseados em Gini e os estimadores possuem grandes paralelos com o método dos mínimos quadrados (OLS). Ambos os modelos não requerem a especificação de uma forma de modelo, este método é de interesse quando necessitamos estimar as inclinações médias sem necessitar de um modelo de estimação (SHLOMO; EDNA, 2013, pag 134). Este método é chamado de semi-paramétrico pois ele não requer a premissa de linearidade ou qualquer premissa da distribuição dos dados, este é mais um paralelo entre o método e o OLS. A única diferença entre o OLS e os coeficiente da regressão Gini semi-paramétrica está nos pesos referentes a cada inclinação. Partindo de uma distribuição bivariada (X, Y) a qual segue uma distribuição contínua com momentos de primeira e segunda ordem contínuos, sem quaisquer premissas adicionais, nem mesmo a de que há uma relação linear entre as duas variáveis. Afim de construir um modelo linear de Y, baseado em X. O preditor linear será

$$\hat{\mathrm{Y}} = \alpha + \beta \mathrm{X} \tag{3.17}$$

Onde alpha e β são constantes arbitrárias impostas. O resíduo deste processo será então

$$\varepsilon = \mathrm{Y} - \hat{\mathrm{Y}} \equiv \mathrm{Y} - \alpha - \beta \mathrm{X} \tag{3.18}$$

Todas as propriedades dos resíduos são derivadas das propriedades de (Y, X). Por meio das propriedades da covariância, obtemos

$$\mathrm{cov}(\mathrm{Y}, \mathrm{X}) \equiv \mathrm{cov}(\alpha + \beta X + \varepsilon, \mathrm{X}) \equiv \beta \,\mathrm{cov}(\mathrm{X}, \mathrm{X}) + \mathrm{cov}(\varepsilon, \mathrm{X}) \tag{3.19}$$

Impondo a ortogonalidade $cov(\varepsilon, X) = 0$, 3.19 será alterada para uma equação facilmente solucionável do tipo

$$\beta = \frac{\mathrm{cov}(\mathrm{Y}, \mathrm{X})}{\mathrm{cov}(\mathrm{X}, \mathrm{X})} \tag{3.20}$$

Vemos aqui que 3.20 é similar em estrutura ao coeficiente obtido por meio do OLS. No OLS a restrição de ortogonalidade de $cov(\varepsilon, X) = 0$ é derivada da minimização da variância do termo de erro. Fazendo um paralelo com os métodos aplicados por Gini, substituímos os termos relacionados a covariância pelo relacionados à covariância Gini, resultando no equivalente da regressão Gini semi-paramétrica:

$$\beta_{\mathrm{N}} = \frac{\mathrm{cov(Y, F(X))}}{\mathrm{cov(X, F(X))}} \tag{3.21}$$

Por meio das propriedades da covariância, obtemos outro importante termo

$$\mathrm{cov}\left(\varepsilon_{\mathrm{N}}, \mathrm{F(X)}\right) = 0 \tag{3.22}$$

onde ε_{N} é o resíduo da regressão Gini semi-paramétrica.

Com β da expressão 3.17 estimado é possível estimar o parâmetro α de duas formas:

1. Fazendo a linha de regressão passar pela média das variáveis;
2. Estimar alpha por meio da minimização da soma dos desvios absolutos entre os resíduos e uma constante, ocasionando a passagem da linha de regressão pela mediana.

Esta abordagem permite distinguir a separação na escolha dos critérios utilizados para determinar a inclinação e o critério para estimar o termo constante.

3.5.2 Regressão Gini por meio de Minimização

Esta abordagem se faz por meio da minimização do GMD dos resíduos. Neste caso é necessário a especificação de uma função alvo e um modelo, desta forma, assume-se que o modelo é linear. Os parâmetros e estimadores relacionados à abordagem de minimização do GMD serão denotadas por M. Partindo da mesma premissa de um modelo linear como em 3.17 e resíduos 3.18, consideremos que o termo α é estimado por meio das médias das variáveis. Schechtman *et al.* (2011) apresentam, através da representação do GMD, e impondo uma restrição da média residual ser igual a zero nos permite apresentar o GMD do erro amostral como uma função de β, o qual, com base na expressão da covariância Gini, pode ser escrito como

$$G_e(b) = \frac{1}{n}\,\mathrm{cov}(e, R(e)) \tag{3.23}$$

onde R representa o vetor de classificação dos termos do erro. Minimizar 3.23 é o equivalente de minimizar

$$\sum e_i R\left(e_i\right) \tag{3.24}$$

Com base nas propriedades da variância e o fato de que $\varepsilon = \mathrm{Y} - \alpha - \beta\mathrm{X}$

$$G_e(b) = \left(\frac{4}{n}\right)\{\mathrm{Cov}(y, R(e)) - b\,\mathrm{Cov}(x, R(e))\} \tag{3.25}$$

Supondo que para um dado β, calculemos o GMD do termo de erro

$$G_e(b) = \sum_{i,j} |e_i - e_j| \tag{3.26}$$

A única propriedade aqui requerida é que a minimização do termo de erro resulta em uma condição de ortogonalidade, tal qual a da equação normal do OLS, dada por

$$R_m, R(ei) \text{ e } x'r_M = 0 \tag{3.27}$$

A equação 3.27 mostra que a covariância amostral entre a classificação dos resíduos e o valor da variável explanatória são levados a zero como um resultado da minimização do GMD dos resíduos. Olkin e Yitzhaki (1992) ressaltam que o caso 3.27 se mantem para casos de regressões simples. Os conceitos relativos às regressões múltiplas serão apresentados mais à frente.

3.6 REGRESSÕES GINI MÚLTIPLAS

Nesta seção serão apresentados os conceitos necessários para desenvolver a Regressão Múltipla Gini, a qual foi utilizado na modelagem das séries temporais estudadas neste trabalho. Desta forma são apresentados os conceitos da Regressão Simples de Gini com seus paralelos no OLS para transferir o método semi-paramétrico e de minimização para o cenário múltiplo.

3.6.1 Abordagem Semi-Paramétrica

Considerando $(Y, X_1, ..., X_k)$ sendo uma $(K + 1)$ variável aleatória, com valores esperados $(\mu_1, \mu_2, ..., \mu_{ik})$, respectivamente, e matriz de variância-covariância finita E, o modelo genérico da curva de regressão será então

$$g\left(x_1, \ldots, x_K\right) = E\left\{Y | X_1 = x_1, \ldots, X_K = x_K\right\} \tag{3.28}$$

Buscamos estimar uma aproximação linear da curva de regressão, a qual é composta por inclinações condicionais, ou seja, cada inclinação de Y em X_i é condicional aos outros X no modelo. Conforme os conceitos apresentados na sessão de regressões lineares múltiplas com base no OLS, estabelecemos os parâmetros (inclinações condicionais), os

quais são interpretados como a solução do sistema de equações lineares envolvendo as inclinações univariada e os dados.

Schechtman *et al.* (2011) apresentam o desenvolvimento deste problema como um vetor de coeficientes da regressão, β_n, dado por

$$\beta_N = [E(V'X)]^{-1} E(V'Y) \tag{3.29}$$

onde $\beta_n = \beta_{n1}, ..., \beta_{nk}$ é um $(Kx1)$ vetor coluna dos coeficientes da regressão, V é uma matrix (nxK) das distribuições cumulativas de $X_1, ..., X_k$ (em desvios dos valores esperados), Y é um vetor $(nx1)$da variável dependente e X é uma matriz (nxK) dos desvios da variável explanatória de seus valores esperados. Os elementos de $E(V'Y)$ e $E(V'X)$ são, respectivamente,

$$\text{COV}(Y, F(X_k)) \text{ e } \text{COV}(X_j, F(X_k)) \tag{3.30}$$

Contudo, esta metodologia possui uma restrição. Para ter uma solução, a ordem de $V'X$ deve ser igual à de K, o número de variáveis explanatórias. Isso representa um problema no GMD e nas regressões Gini, pois a utilização de várias transformações monotônicas nas variáveis explanatórias no modelo de regressão não mudam as classificações das observações; consequentemente, as colunas representando as distribuições cumulativas das variáveis explanatórias podem ser idênticas, resultando em multicolinearidade.

Os estimadores naturais dos coeficientes da regressão são baseados na substituição das distribuições cumulativas por distribuições empíricas

$$b_N = [v'x]^{-1} (v'y) \tag{3.31}$$

Onde v é a matrix dos elementos $[n^{-1}(r(x_{ik})) - 1/2]$, e r_{xik} é a ordem de xik entre $x_{1k}, ..., x_{nk}$. Schechtman e Yitzhaki (2008) apresentam a prova de que b_n é um estimador consistente de β_n.

Uma vez estimados os coeficientes da regressão Gini, o termo constante pode ser estimado pela minimização da função dos resíduos. A função exata utilizada irá determinar se a regressão passa pela média, mediana ou qualquer outro quantil.

4 METODOLOGIA EXPERIMENTAL E PROCEDIMENTOS DESCRITIVOS

4.1 CONSIDERAÇÕES INICIAIS

Nos capítulos anteriores, foram discutidos os conceitos fundamentais e necessários para o estudo da estimação de parâmetros para modelo auto regressivos por meio da metodologia de covariância Gini.

4.2 METODOLOGIA GINI APLICADA À SÉRIES TEMPORAIS

Conforme apresentado no capítulo 2, para distribuições univariadas F, uma importante alternativa para o desvio padrão como uma medida de dispersão foi introduzida por Gini (1912) conhecida por GMD:

$$\alpha(F) = E\,|X_1 - X_2| \tag{4.1}$$

Sendo a covariância Gini representada por

$$\alpha(X) = 2\,\mathrm{Cov}(X, 2F(X) - 1) = 4\,\mathrm{Cov}(X, F(X)) \tag{4.2}$$

E a correlação Gini

$$\begin{aligned} G\,\mathrm{cor}(Y,X) = \Gamma(Y,X) &= \frac{G\,\mathrm{cov}(Y,X)}{G\,\mathrm{cov}(Y,Y)} = \frac{\mathrm{COV}\,(Y, F_X(X))}{\mathrm{COV}\,(Y, F_Y(Y))} \\ G\,\mathrm{cor}(X,Y) = \Gamma(X,Y) &= \frac{G\,\mathrm{cov}(X,Y)}{G\,\mathrm{cov}(X,X)} = \frac{\mathrm{COV}\,(X, F_Y(Y))}{\mathrm{COV}\,(X, F_X(X))} \end{aligned} \tag{4.3}$$

4.2.1 Auto Covariância Gini

Considerando um modelo auto regressivo discreto $t(t = 0, \pm1, \pm2, ...)$. A auto covariância entre Y_t e Y_{t-s} é definida como λ_t, $t - s = COV(Y_t, Y_{t-s})$ para qualquer *lag* $s(s = 0, \pm1, \pm2, ...)$. Seguindo o modelo de Serfling (2010) e Shelef e Schechtman (2011) são definidos duas auto covariância Gini para o *lag* s como

$$\gamma^{G_1}_{(t,t-s)} = \mathrm{COV}\,(Y_t, F\,(Y_{t-s})) \text{ and } \gamma^{G_2}_{(t_{t-s})} = \mathrm{COV}\,(Y_{t-s}, F\,(Y_t)) \tag{4.4}$$

Essas auto covariância podem ser vistas como as auto covariâncias Gini olhando para frente e para trás na série. Assumindo essas séries estritamente estacionárias, as

distribuições conjuntas de $(Y_{t1}, ..., Y_{tk})$ e $(Y_{t1+s}, ..., Y_{tk+s})$ são iguais para todos os inteiros positivos e para todos os $(t_1, ..., t_k)$, $s \in Z$ (BROCKWELL, 1991). Consequentemente as condições se mantêm para todo t, s

$$\because \text{COV}\left(Y_t, F\left(Y_{t-s}\right)\right) = \text{COV}\left(Y_{t-j}, F\left(Y_{t-j-s}\right)\right) = \gamma_{(s)}^{G_1} \tag{4.5}$$

e

$$\text{COV}\left(Y_{t-s}, F\left(Y_t\right)\right) = \text{COV}\left(Y_{t-j-s}, F\left(Y_{t-j}\right)\right) = \gamma_{(s)}^{G_2} \tag{4.6}$$

onde todos $\gamma_{(s)}^{G_1}$ e $\gamma_{(s)}^{G_2}$ são independentes do tempo. Portanto, para um processo auto regressivo de primeira ordem, do tipo $Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$

$$\gamma_{(s)}^{G_1} = \text{COV}\left(\phi_0 + \phi_1 Y_{t-1} + \varepsilon_t, F\left(Y_{t-s}\right)\right) = \text{COV}\left(\phi_1 Y_{t-1}, F\left(Y_{t-s}\right)\right) = \phi_1^s \gamma_{(s=0)}^{G_1} \tag{4.7}$$

porém,

$$\gamma_{(s)}^{G_2} = \text{COV}\left(Y_{t-s}, F\left(\phi_0 + \phi_1 Y_{t-1} + \varepsilon_t\right)\right) \tag{4.8}$$

não precisa necessariamente ser igual a $\gamma_{(s)}^{G_1}$.

4.2.2 Função de Auto Correlação de Gini (ACF de Gini)

A função de auto correlação (ACF) apresentada pelo método clássico entre Y_t e Y_{t-s} (sob estrita estacionaridade)

$$ACF\left(Y_t, Y_{t-s}\right) = \rho_s = \gamma_s / \gamma_0 \tag{4.9}$$

Essa função é comumente apresentada versus o lag, como já mencionado antes, sendo uma forma de identificar o modelo para um conjunto de dados. Quando estamos trabalhando com amostras a versão mais utilizada da ACF é

$$\hat{\rho}_s = \frac{\sum_{t=1}^{T-s}\left(Y_{t+s} - \overline{Y}\right)\left(Y_t - \overline{Y}\right)}{\sum_{t=1}^{T}\left(Y_t - \overline{Y}\right)^2} \tag{4.10}$$

Uma versão diferente da ACF sugerida por Davis e Resnick (1985) para o caso de séries com distribuições caudais longas. Outros autores, como Feigin e Resnick (1999) ressaltam que cuidado deve ser tomado quando se busca ajustar modelos para séries com caudas longas.

Utilizando como base a equação 4.9 as duas ACF de Gini, tendo ordem s, são:

$$\begin{aligned} \text{Gini} - ACF\left(Y_t, Y_{t-s}\right) = \rho_{(s)}^{G_1} = \frac{\gamma_{(s)}^{(s)}}{\gamma_{(s=0)}^{G_1}} \text{ e} \\ Gini - ACF\left(Y_{t-s}, Y_t\right) = \rho_{(s)}^{G_2} = \frac{\gamma_{(s)}^{G_2}}{\gamma_{(s=0)}^{G_2}} \end{aligned} \tag{4.11}$$

Conforme apresentado na equação 4.7, $\rho_{(s)}^{G_1} = \phi_1^s = \rho_{(s)}$ indicando que a primeira auto correlação de Gini é igual a ACF. Desta maneira, esta é uma opção para calcular a ACF quando momentos de segunda ordem não existirem. Uma diferença nos as duas ACF de Gini, quando existir, implica que uma medida assimétrica, tal como a apresentada por Gini, por ser mais apropriada e irá oferecer mais informação sobre a distribuição adjacente (SHELEF; SCHECHTMAN, 2016). Nos casos onde o erro da função é uma variável normal e $i.i.d$ então podemos assumir Y_t e Y_{t-s} como uma combinação linear de variáveis normais. Por outro lado, caso as ACF de Gini sejam diferentes, isto indica que Y_t e Y_{t-s} não são intercambiáveis, resultando em uma análise diferente quando olhamos a série para frente ou para trás. A grande importância de identificar esses casos é por que ele descarta a possibilidade de modelar os dados como um processo Gaussiano.

A maioria desses estimadores não são consistentes para os métodos comumente utilizados para estimação da auto correlação. Assim, segundo Shelef e Schechtman (2011) a alternativa é estimar as funções de auto correlação Gini, as quais são mais consistentes com a correlação amostral.

$$\hat{\rho}_{(s)}^{G_1} = \frac{\sum_{t=1}^{T-s}\left(Y_{t+s} - \overline{Y}\right)\left(R\left(Y_t\right) - \overline{R}\left(Y_{1:(T-s))}^{T}\right)}{\sum_{t=1}^{T}\left(Y_t - \overline{Y}\right)\left(R\left(Y_t\right) - \overline{R}\left(Y_{1:T}\right)\right)} \tag{4.12}$$

$$\hat{\rho}_{(s)}^{G_2} = \frac{\sum_{t=1}^{T-s}\left(Y_t - Y\right)\left(R\left(Y_{t+s}\right) - R\left(Y_{(s+1):T)}\right)\right.}{\sum_{t=1}^{T}\left(Y_t - \overline{Y}\right)\left(R\left(Y_t\right) - \overline{R}\left(Y_{1:T}\right)\right)} \tag{4.13}$$

onde $R\left(Y_t\right)$ é o classificação de Y_t/N e $\overline{R}\left(Y_{i:j}\right) = \sum_{t=i}^{j} R\left(Y_t\right)/(j-i+1)$. Quando observamos a equação 4.10, $-Y$ representa o estimador da média do processo. Além disso, com o aumento de s (ordem do processo) o numerador terá poucos componentes e convergirá para zero, como esperado em um processo sob estrita estacionaridade.

4.2.3 Função de Auto Correlação Parcial de Gini (PACF de Gini)

A representação da PACF de Gini vem do método da variância, o qual provê o seguinte sistema de equações

$$\rho_{(j)} = \phi_{s1}\rho_{(j-1)} + \ldots + \phi_{ss}\rho_{(j-s)}, \text{ para todo } j = 1, 2, \ldots, s \tag{4.14}$$

na equação 4.14, a PACF é definida como o último coeficiente ϕ_{ss}, o qual representa a auto correlação entre Y_t e Y_{t-s} após o ajuste para o efeito das variáveis intermediárias, como já discutido anteriormente no capítulo 2.6.2, tal que $\phi_{ss} = cor(Y_t, Y_{t-s}|Y_{t-1}, ..., Y_{t-s+1})$.

De maneira similar (como quase tudo apresentado em relação ao método Gini) podemos representar a PACF de Gini como sendo a auto correlação de Gini entre Y_t e Y_{t-s} após o ajuste para o efeito das variáveis intermediárias, aqui representada por

$\phi_{ss}^{G_1} = Gcorr(Y_t, Y_{t-s}|Y_{t-1}, ..., Y_{t-s+1})$. A PACF de Gini também é definida como o último coeficiente da auto regressão parcial de Gini com ordem s (e média centrada em zero) definida por

$$Y_r = \phi_{s1}^{G_1} Y_{t-1} + \phi_{s2}^{G_1} Y_{t-2} + \ldots + \phi_{s(s-1)}^{G_1} Y_{t-s+1} + \phi_{ss}^{G_1} Y_{t-s} + \varepsilon_t \tag{4.15}$$

A covariância de Gini entre Y_t e Y_{t-j} é

$$G\,\mathrm{cov}\,(Y_t, Y_{t-j}) = \mathrm{COV}\,(Y_t, F\,(Y_{t-j})) = \gamma_{(j)}^{G_1} = \phi_{s1}^{G} \gamma_{(j-1)}^{G_1} + \ldots + \phi_{ss}^{G} \gamma_{(j-s)}^{G_1} \tag{4.16}$$

O que resulta em

$$\rho_{(j)}^{G_1} = \phi_{s1}^{G_1} \rho_{(j-1)}^{G_1} + \ldots + \phi_{ss}^{G_1} \rho_{(j-s)}^{G} \tag{4.17}$$

Essa solução abre caminho para o entendimento de que a substituição de cada ACF ρ na equação 4.14 por seu primeiro representante na GACF $\phi_s^{G_1}$, caso utilizemos a segunda "parcela" do GACF $\phi_s^{G_2}$ ao invés do primeiro, temos um versão adicional do PACF de Gini, o qual será de grande relevância para os estudos desenvolvidos neste trabalho, permitindo o estudo das séries temporais para frente e para trás.

Com base no método das equações de Yule-Walker é possível fazer a aproximação das estimativas sucessivas do PACF em processos auto regressivos utilizando ρ 's como estimadores da auto correlação teórica (WEI, 2006). Segundo Box *et al.* (2008, pg 67) ressalta que estes estimadores são sensíveis a errors de arredondamento e não devem ser utilizados caso os parâmetros estiverem próximos aos limites da não estacionaridade. Desta maneira Shelef e Schechtman (2011) sugerem a estimação das duas PACF de Gini utilizando o seguinte Sistema de equações

$$\hat{\rho}_{(j)}^{G_1} = \hat{\phi}_{s1}^{G_1} \hat{\rho}_{(j-1)}^{G_1} + \hat{\phi}_{s2}^{G_1} \hat{\rho}_{(j-2)}^{G_1} + \cdots + \hat{\phi}_{s(s-1)}^{G_1} \hat{\rho}_{(j-s+1)}^{G_1} + \hat{\phi}_{ss}^{G_1} \hat{\rho}_{(j-s)}^{G_1} \tag{4.18}$$

e

$$\hat{\rho}_{(j)}^{G_2} = \hat{\phi}_{s1}^{G_2} \hat{\rho}_{(j-1)}^{G_2} + \hat{\phi}_{s2}^{G_2} \hat{\rho}_{(j-2)}^{G_2} + \cdots + \hat{\phi}_{s(s-1)}^{G_2} \hat{\rho}_{(j-s+1)}^{G_2} + \hat{\phi}_{ss}^{G_2} \hat{\rho}_{(j-s)}^{G_2} \tag{4.19}$$

As quais devem ser resolvidas para os dois últimos coeficientes $\hat{\phi}_{ss}^{G_1}$ e $\hat{\phi}_{SS}^{G_2}$ para todo $s = 1, 2, \ldots$

4.2.4 Sistema de Covariâncias

Como visto na seção 3.6, a metodologia Gini presente na literatura apresenta duas abordagens para a estimação dos parâmetros de uma regressão, o método **Semi-**

paramétrico e o de **Minimização**. Esta seção apresenta o método utilizado para obter o sistema de equações que resultam nos parâmetros do processo AR, via um sistema linear de covariância Gini e o sistema Yule-Walker (2.6.1)

Em geral, dado um sistema linear de p equações com coeficientes dados por uma certa população de parâmetros $\phi_1, \ldots, \phi_p$, a matriz inversa resulta em fórmulas explícitas destes parâmetros. Considere

$$\eta = \sum_{j=1}^{p} \phi_j \alpha_j + \varepsilon \tag{4.20}$$

com ϵ independente para $\alpha_1, \ldots, \alpha_p$. Segundo Serfling (2010) buscamos então, o sistema linear

$$a_i = \sum_{j=1}^{p} b_{ij} \phi_j, i = 1, \ldots, p \tag{4.21}$$

ou em forma matricial

$$a = B\phi \tag{4.22}$$

com $a = (a_1, \ldots, a_p)^{\mathrm{T}}$, $\phi = (\phi_1, \ldots, \phi_p)^{\mathrm{T}}$ e $B = (b_{ij})_{p \times p}$. Para qualquer função $Q(\alpha_1, \ldots, \alpha_p)$, teremos

$$\operatorname{Cov}(\eta, Q(\alpha_1, \ldots, \alpha_p)) = \sum_{j=1}^{p} \phi_j \operatorname{Cov}(\alpha_j, Q(\alpha_1, \ldots, \alpha_p)) \tag{4.23}$$

provemos que essas covariâncias são finitas.

Qualquer escolha de p funções $Q_i(\alpha_1, \ldots, \alpha_p), 1 \leq i \leq p$ em 4.23 resulta em um sistema do tipo 4.22

$$\begin{aligned} a_i &= \operatorname{Cov}(\eta, Q_i(\alpha_1, \ldots, \alpha_p), 1 \leq i \leq p \\ b_{ij} &= \operatorname{Cov}(\xi_j, Q_i(\alpha_1, \ldots, \alpha_p)), 1 \leq i, j \leq p \end{aligned} \tag{4.24}$$

Para alguma função g, escolhemos estas p funções para serem da forma $Q_i(\alpha_1, \ldots, \alpha_p) = g(\alpha_i), 1 \leq i \leq p$ e obtemos

$$\begin{aligned} a_i &= \operatorname{Cov}(\eta, g(\alpha_i)), 1 \leq i \leq p \\ b_{ij} &= \operatorname{Cov}(\alpha_j, g(\alpha_i)), 1 \leq i, j \leq p \end{aligned} \tag{4.25}$$

Aplicando recursivamente o mecanismo $(\eta, \alpha_1, \ldots, \alpha_p) = (X_t, X_{t-1}, \ldots, X_{t-p})$ tal qual no modelo $AR(p)$, obtemos tanto a abordagem de Mínimos Quadrados sob premissas de segunda-ordem (variância), quanto a abordagem Gini sob premissas de apenas primeira-ordem (média/mediana).

4.2.4.1 O sistema Gini-Yule-Walker

Com premissas de primeira-ordem e utilizando $g(\alpha_i) = 2(2F_X(\alpha_i) - 1)$ juntamente com $Q_i(X_{t-1}, \ldots, X_{t-p}) = 2\left(2F_{X_{t-i}}(X_{t-i}) - 1\right)$, obtemos

$$a_i = \beta(X_t, X_{t-i}) = \gamma^{(G)}(i), 1 \leq i \leq p, \text{ e } b_{ij} = \beta(X_{t-j}, X_{t-i}) = \gamma^{(\mathrm{G})}(i-j) 1 \leq i, j \leq p \tag{4.26}$$

Com essas considerações, podemos afirmar que 4.22 resulta no sistema *Gini-Yule-Walker* para $\phi_1, \ldots, \phi_p$. Para $p = 1$ essa solução é simplesmente

$$\phi_1 = \gamma^{(\mathrm{G})}(1)/\gamma^{(\mathrm{G})}(0) \tag{4.27}$$

O sistema *Gini-Yule-Walker* possue a mesma estrutura computacional do sistema de mínimos quadrados, porem requer apenas premissas de primeira-ordem (CARCEA; SERFLING, 2015).

4.3 SÉRIES SINTÉTICAS

Como já enfatizado anteriormente, a metodologia Gini requer apenas premissas de primeira ordem para estimação dos parâmetros da regressão, ou modelo AR(p), desta maneira, este trabalho buscou avaliar o comportamento e os resultados gerados pela aplicação da metodologia na estimação de parâmetros para séries com distribuições não-normais e nas quais momentos de segunda ou terceira ordem não são parâmetros que refletem o comportamento amostral.

A escolha do tamanho amostral foi arbitrário, sendo escolhido um tamanho igual a **duzentas observações** (o qual poderia representar pouco mais que seis meses de observações), com populações seguindo distribuições dos seguintes tipos:

- Pareto;
- Log-Normal;
- Weibull;
- Bimodal;
- Burr.

O processo de criação de cada uma das séries levou em conta a distribuição dos parâmetros intrínsecos a cada distribuição, os quais definem seu formato, buscando obter

séries com caudas longas e formas variadas para que a avaliação da metodologia Gini seja diversificada para um conjunto de casos, algumas caracaterísticas das séries geradas são:

- Os valores dentro de cada distribuição são sempre positivos e maiores que zero;
- Foram geradas 22 séries apresentando caudas longas;
- Para validar a não-normalidade de tais séries, foram aplicados testes de normalidade de Shapiro e Anderson-Darling, os quais mostraram que nenhuma das séries se aproximam da normalidade.

Algumas características essenciais das séries estudadas são apresentadas a seguir. A metodologia de transformação das séries em processos AR(p) é apresentada na seção 4.4.

4.3.1 Distribuições Com Caudas Longas

Suponha que tenhamos uma sequência de variáveis $i.i.d$ $X_1, X_2, ...$ de uma distribuição desconhecida F Denotamos o máximo das primeiras n observações como $M_n = \max(X_1, \ldots, X_n)$. Além disso, suponha que podemos encontrar sequências de números reais $a_n > 0$ e b_n tal que $(M_n - b_n)/a_n$, a sequência de máximos normalizados convergem em distribuição. Isto é

$$P\{(M_n, -b_n)/a_n \leq x\} = F^n(\text{anx} + b_n) \to H(x), \text{ as } n \to \infty \tag{4.28}$$

caso esta condição se mantenha, dizemos que F está no domínio máximo de atração de H e escrevemos $F \in \text{MDA}(H)$.

Fisher e Tippett (1928) mostram que

$$F \in \text{MDA}(H) \Rightarrow H \text{ é do tipo } H_\xi \text{ para algum } \xi \tag{4.29}$$

consequentemente, se sabemos que o máximo normalizado ajustado converge em distribuição, então a distribuição limite deve ser uma distribuição de valor extremo para algum valor de parâmetros ξ, μ e σ.

A classe de distribuições F para as quais a condição 4.28 se mantém é grande. Uma variedade de condições equivalentes podem ser derivadas (veja, por exemplo, Falk e Reiss (1994)). Como apresentado em McNeil (1997) se $\xi > 0, F \in \text{MDA}(H_\xi)$ apenas e apenas se $1 - F(x) = x^{-1/\xi}L(x)$, para algum função pouco variável $L(x)$. Este resultado diz essencialmente que caso a cauda de d.f. $F(x)$ decaia como uma função potência, então essa distribuição é considerada no domínio de caudais longos. A classe de distribuições neste domínio (decaimento da cauda como uma função potência) é grande, e inclui Pareto, Burr, LogGamma, Cauchy e distribuções-t, e também variados modelos de mistura. Estas

distribuições serão amplamente utilizadas ara testar a capacidade da metodologia Gini em estimar modelos auto regressivos, uma vez que, como apresentado no capítulo 3 o Gini é menos sensível a *outliers* e por isso pode permitir melhores estimativas quando os dados apresentam distribuições com estas características.

O principal objetivo de uma análise de valores extremos é estimar a probabilidade de eventos os quais são mais extremos do que qualquer outro que já tenha sido observado (FERREIRA, 2012). Como exemplo, podemos citar eventos naturais com decorrência longa, por exemplo, enchentes de 50 anos, entre outros eventos.

4.3.2 Séries Log-Normais

Variáveis em um sistema que seguem uma relação exponencial $x = exp(w)$. Se o expoente é uma variável alatória W, então $X = exp(W)$ é uma variável aleatória com uma distribuição de interesse. O caso especial ocorre quando W tem uma distribuição normal. Neste caso, a distribuição de X é chamada de Distribução Log-Normal. O nome segue a transformação $ln(X) = W$. Isto é, o logarítmo natural de X é normalmente distribuído.

Considerando W como tendo uma distribuição normal com média θ e variância ω^2, então $X = \exp(W)$ é uma variável aleatória Log-Normal, com função de densidade de probabilidade

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \theta)^2}{2\omega^2}\right] \quad 0 < x < \infty| \tag{4.30}$$

a média e variância de X são

$$E(X) = e^{\theta+\omega^2/2} \quad \text{e} \quad V(X) = e^{2\theta+\omega^2}\left(e^{\omega^2} - 1\right) \tag{4.31}$$

Devore (2010) destaca o fato de que os parâmetros de média e variância não são referentes a X, e sim a $ln(X)$. É comum nos referirmos a θ e variância ω^2 como os parâmetros de localização e de escala, respectivamente.

Para distribuições Log-Normal, a maneira mais precisa de estimar os parâmetros são transformações logarítmicas. As médias e desvio padrões empíricos dos dados são calculados e então inversamente transformados, métodos mais robustos, porém menos eficientes podem ser obstidos pela mediana e quartis dos dados (LIMPERT *et al.*, 2018). A figura 4.1 mostra exemplos de distribuições Log-Normal.

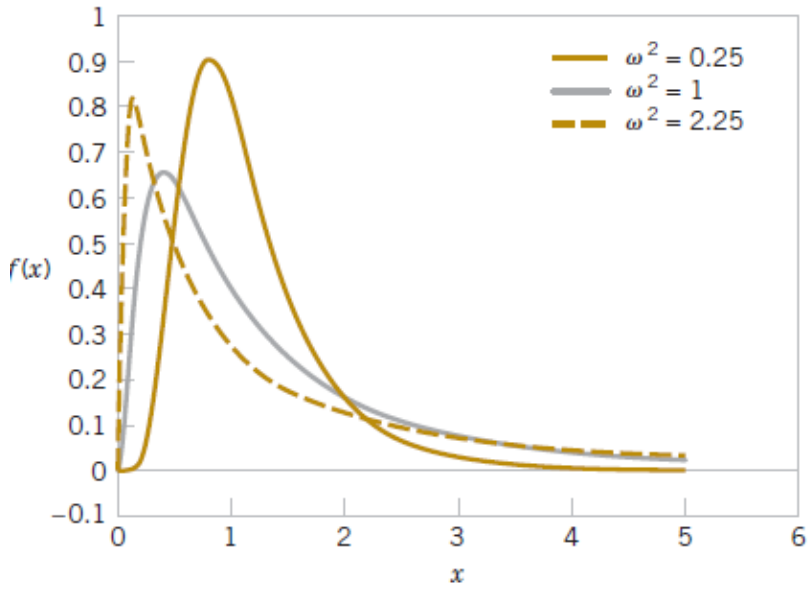

Figura 4.1 – Funções de Densidade de Probabilidade Log-Normal com $\theta = 0$ para valores selcionados de ω^2. Fonte: (MONTGOMERY, 2008, pág. 146)

Este trabalho buscou testar a metodologia para diversas séries com características Log-Normal com tamanho amostral n = 200. Essas séries foram geradas variando os parâmetros de localização e escala, estas distribuições serão exploradas na seção de resultados.

4.3.3 Séries Pareto

Uma grande variedade de variáveis socioeconômicas possuem distribuição as quais posuem caudas longas e são razoavelmente ajustadas em uma distribuição de Pareto (YEH *et al.*, 1988). Modelo extremos são ferramentas preciosas que nos permitem extrapolações deste tipo. Como observado por Moore (1897), muitas variáveis econômicas possuem distribuições com caudais longos e não são bem modeladas pela curva normal. Ao invés disto, ele propos uma modelo, posteriormente nomeado em sua homenagem, a distribuição de Pareto, a qual será discutida nesta seção.

O estudo de Ghitany *et al.* (2018) afirma que o uso de séries com longas caudas a direita é de vital importância no mercado de seguros, onde as distribuições de Pareto e Log-Normal são amplamente utilizadas em modelos de perda, confiabilidade de seguros, seguros contra incêncios e catastrofes.

A caracaterística marcante desta distribuição é que sua função de sobrevivência $P(X > x)$ decresce com uma potência negativa de x como $x \to \infty$, tal como em

$$P(X > x) \sim cx^{-\alpha}, \text{ as } x \to \infty \tag{4.32}$$

Generalizações das distribuições de Pareto foram amplamente propostas para mo-

delar variáveis econômicas (como pode ser encontrado em (ARNOLD, 2008)).

Ferreira (2012) apresenta a distribuição Pareto clássica (também chamada de Pareto(I)) como tendo a função de sobrevivência da forma

$$\overline{F}_X(x) = (x/\sigma)^{-\alpha}, x > \sigma \tag{4.33}$$

onde $\sigma > 0$ é o parâmetro de escala e $\alpha > 0$ é o parâmetro de forma (ou desigualdade). Uma maneira de escrever a variável X caso a mesma tenha a distribuição 4.33 é $X \sim P(I)(\sigma, \alpha)$.

Um exemplo gráfico de uma distribuição Pareto é apresentada na figura 4.2

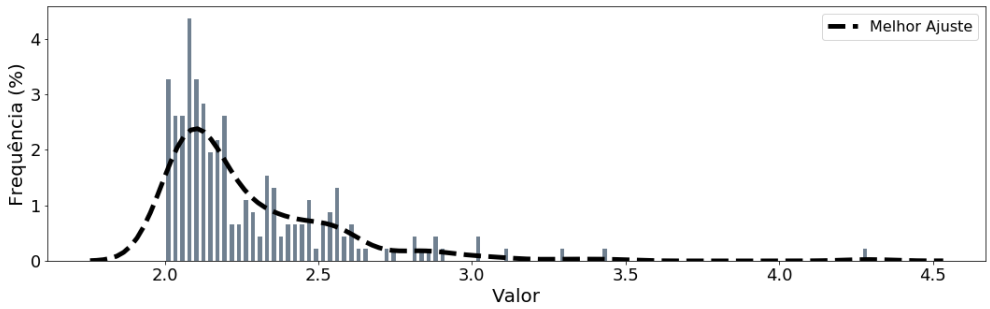

Figura 4.2 – Exemplo de distribuição Pareto.

4.3.4 Séries Weibull

Distribuições de Weibull são comuns na engenharia, medicina, ciências sociais, finanças, etc. Ela é comumente utilizada para modelar o tempo até uma falha em diversos sistemas físicos. Os parâmetros na distribuição provêm grande flexibilidade na modelagens de sistemas em que as falhas aumentam com o tempo.

A variável aleatória X com função de densidade de probabilidade

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right], \quad \text{para } x > 0 \tag{4.34}$$

é uma variável aleatória Weibull com parâmetro de escala $\delta > 0$ e fator de forma $\beta > 0$ (MONTGOMERY, 2008).

Com $\beta < 1$, a distribuição de Weibull é particularmente adequada para séries temporais com caudas longas (os quais serão aplicados neste trabalho), onde os valores distantes dos máximos de probabilidades ainda são comuns.

A distribuição de Weibull é assimétrica, desta forma, probabilidades de eventos ocorridos antes da moda não são os mesmos após ela. Para sua utilização, é essencial conhecer os fatores de escala e forma.

Kizilersü *et al.* (2018) destaca três regimes que descrevem os formatos.

1. Para $\beta < 1$, as probabilidades tendem a infinito conforme o tempo (t) aproxima zero.

2. Quando $\beta = 1$, a Weibull é nada mais do que uma distribuição exponencial, a qual é finita no ponto inicial.

3. Quando $\beta > 1$ encontramos a distribuição com uma "lombada", como uma curva em forma de sino, parecida com uma distribuição normal, exceto pelo fato de ser assimétrica.

Um exemplo gráfico de uma distribuição Weibull é apresentada na figura 4.3

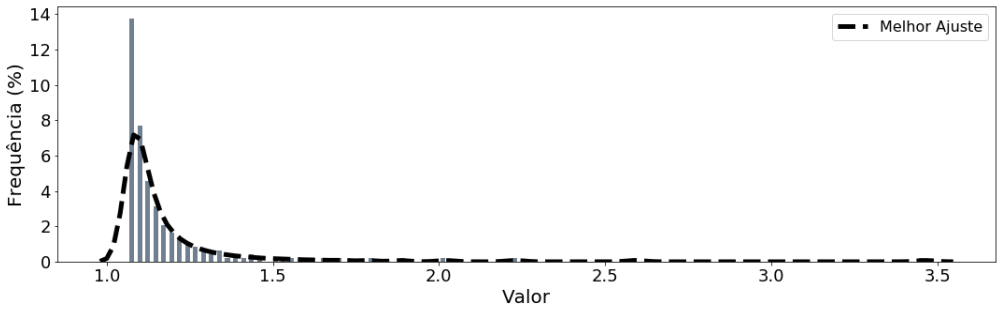

Figura 4.3 – Exemplo de distribuição Weibull.

4.3.5 Séries Bimodais

Não é incomum encontrar na natureza, processos que possuam distribuições bimodais. Por isso, estas distribuições também serão testadas neste trabalho. Muitas distribuições Bimodais podem ter sua origem da mistura de duas distribuições normais.

O desenvolvimento de distribuições bimodais deste trabalho, teve sua fundamentação no trabalho de Eisenberger (1964) o qual apresenta diversas metodologias de teste de unimodalidade e testes para séries bimodais.

A figura 4.4 apresenta um exemplo de série bimodal utilizada neste trabalho.

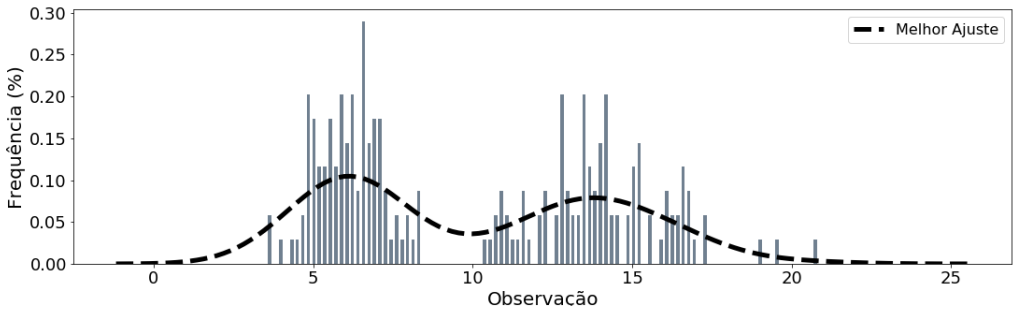

Figura 4.4 – Exemplo de distribuição Bimodal.

4.3.6 Séries Burr

Burr (1942) desenvolveu uma função de densidade com um domínio positivo, a qual é capaz de representar uma série de distribuições de probabilidade, baseados nos 4 primeiros momentos dos dados. Esta distribuição também pode ser facilmente transferida para qualquer distribuição normal e não-normal.

As utilizações desta distribuição são amplas, Chung-Ho e Chao-Yu (2017) aplicam esta distribuição para investigar os efeitos da não-normalidade em estudos de controle estatístico de qualidade, como por exemplo, cartas de controle e planos amostrais. Os autores buscaram determinar parâmetros ótimos de processo nos problemas de configuração de nível.

Outro exemplo de aplicação das distribuições de Burr é apresentada por Taylor (2017) que utiliza esta distribuição para representar tempos de viagem observados e explica como ela pode ser utilizada para desenvolver uma expressão exata para a taxa de confiabilidade na estimação de tempos de viagem.

A função de distribuição cumulativa da função de densidade de Burr é

$$F(y) = 1 - \frac{1}{(1+y^c)^k}, y \geq 0 \tag{4.35}$$

onde $c > 1$ e $k > 1$. Tomando a primeira derivativa na equação 4.35 obtemos a função de densidade de probabilidade de Burr, dada por

$$f(y) = \frac{cky^{c-1}}{(1+y^c)^{k+1}}, y \geq 0 \tag{4.36}$$

Diferentes combinações de c e k cobrem uma ampla gama de funções de densidade.

Um exemplo gráfico de uma distribuição Burr é apresentada na figura 4.5

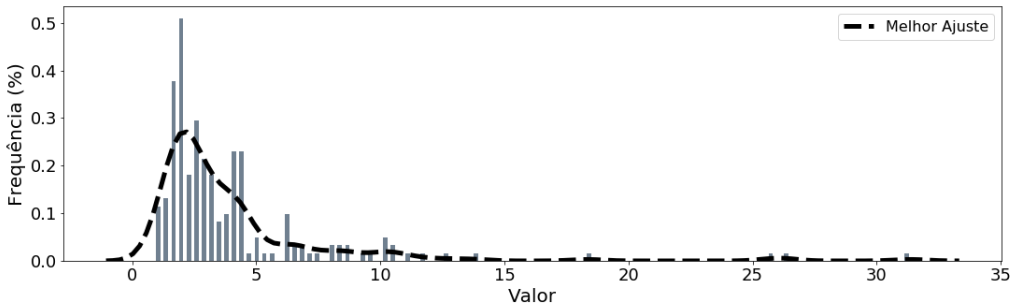

Figura 4.5 – Exemplo de distribuição Burr.

4.4 TRANSFORMAÇÃO DOS DADOS EM SÉRIES TEMPORAIS

Apenas gerar dados não garante que os mesmos tenham características auto regressivas, uma vez que não existe o fator de "movimentação no tempo" aplicado a eles.

Consequentemente, faz-se necessário aplicar um filtro linear ao mesmos, utilizando parâmetros estabelecidos e valores prévios para prever os próximos. Com base na metodologia apresentada até aqui, é possível aplicar este filtro ao conjunto de dados para criar séries temporais que possuem características auto regressivas e também distribuições com caudas longas. Com essa mescla de métodos, foram obtidas as séries finais, as quais analisaremos no capítulo 5.

Neste trabalho, foi estabelecido que a ordem máxima dos modelos seria $p = 4$. Portanto, foram gerados aleatoriamente 4 conjuntos de parâmetros de ordem $t, t = 1, 2, 3, 4$, seus valores são apresentados na tabela 4.1 a seguir.

Tabela 4.1 – Valores dos parâmetros aleatórios e utilizados para gerar os modelos Auto Regresivos.

Ordem 1	Ordem 2	Ordem 3	Ordem 4
0.57	0.76	0.79	0.44
	0.52	0.41	0.57
		0.68	0.62
			0.67

Os parâmetros foram escolhidos aleatóriamente dentro do domínio positivo $\phi > 0$. Para converter os conjuntos de dados em processos auto regressivos foram aplicadas as seguintes operações:

- Aplicação do filtro linear com base nos parâmetros ϕ da tabela 4.1. O filtro linear nada mais é do que gerar a série através da inserção de cada valor da distribuição na série e relacionando-o com as observações através dos pesos associados ϕ.
- Utilização de um processo de Monte Carlo para aplicar o erro padrão (ruído branco) em todas as séries. Lembrando que o erro tem média zero e variância σ_a^2, no caso deste trabalho, $\sigma_a^2 = 1$. Desta forma a cada série foi gerada um conjunto de 200 ruídos, os quais foram somados à cada respectiva observação.
- As 22 séries sintéticas geradas passaram pelos processos dos itens anteriores, sendo aplicadas à todas ordens $p = 1, 2, 3, 4$. Como resultado, obtiveram-se 88 séries temporais sintéticas.

4.5 SELEÇÃO DE ORDEM DO MODELO

A correta estimativa da ordem de um modelo é vital para seu sucesso de previsão, desta maneira, é necessário avaliar diversas combinações de possíveis ordens, afim de encontrar a que melhor representa o modelo de dados. Uma das etapas deste trabalho foi testar todos os modelos possíveis dentro de um certo intervalo de ordens. Para isto o Critério de Informação de Akaike (AIC) foi utilizado com critério de seleção, e com base em seu resultado, foi escolhida a ordem para determinado conjunto de dados.

Mesmo que os dados sintéticos já possuam suas ordens definidas, esta é uma maneira de testar se em cenários reais, qual seria a capacidade do método AIC em estimar a correta ordem para modelos com distribuições não-normais.

Esta abordagem de seleção de modelo foi proposta por (Akaike, 1974). Em sua implementação um intervalo de potenciais modelos é estimado pormétodos de máxima verossimilhança, e para cada modelo, o critério AIC, dado por

$$\mathrm{AIC}_{p,q} = \frac{-2\ln(\text{ maximized likelihood }) + 2r}{n} \tag{4.37}$$

é avaliado. Onde $\hat{\sigma}_a^2$ representa a máxima verossimilhança estimada de σ_a^2, e $r = p+q+1$ denota o número de parâmetros estimados do model, incluindo o termo constante (BROCKWELL, 2002). Na expressão acima, o primeiro termo descreve a menos $2/n$ vezes o log de verossimilhança maximizada, enquanto o segundo termo é um "fator de penalização" pelo inclusão de parâmetros adicionais no modelo. Na abordagem do critério de informação, modelos que resultam em um valor mínimo para o critério são preferidos. O AIC é comparado entre vários modelos afim de escolher o melhor.

Uma desvantagem desta abordagem é que vários modelos precisam ser estimados, o que pode ser custoso do ponto de vista computacional. Neste trabalho, esta foi a abordagem selecionada e devido ao tamanho dos dados e otimização do modelo, não houveram prejuízos em tempos de processamento justificassem a busca por outra alternativa.

Esta metodologia é comumente utlizada na literatura e pode ser encontra nos mais diversos trabalhos e áreas. Ebrahim *et al.* (2018), por exemplo, utiliza o AIC para selecionar modelos e testar a performance de aços na industria por meio de series temporais estacionárias de acidentes e associações com fatores de risco no tempo. Enquanto Ioannidis (2011) propõe variações do AIC para estimar modelos autoregressivos de séries espectrais de densidade em séries temporais.

4.6 AVALIAÇÃO DO MODELO

Para medir e comparar a eficiência da aplicação da metodologia Gini nos modelos de previsão, foram escolhidas quatro métricas de erro. Essas métricas possuem particularidades que permitem uma análise distinta de cada resultado em relação ao que elas nos informam. Desta forma podemos entender melhor as características dos resultados do modelo.

É importante definir cuidadosamente o meios de medir performance. Existem muitas medidas estatísticas que descrevem quão bem um modelo é ajustado para os dados, em geral essas técnicas utilizam o resíduo e não refletem a capacidade de previsão do modelo para valores futuros.

Medidas de acurácia de previsão devem sempre ser utilizadas com parte dos esforços de validação do modelo. Quando mais de uma técnica de previsão é utilizada para determinada aplicação, essas medidas de acurácia da previsão podem também ser utilizadas para discriminar entre os modelos. Este é exatamente o foco da utilização destas técnicas (MONTGOMERY, 2008).

Para a avaliação de desempenho dos modelos, foram utilizadas quatro métricas bastante conhecidas.

- Desvio Padrão Médio (MAD);
- Erro Médio Quadrado (MSE);
- Erro Percentual Absoluto Médio (MAPE);
- Desvio Médio Quadrado (MSD).

Os erros calculados para cada modelo possuem tolerância de $\pm 1\%$.

4.6.1 Coeficiente de Gini

O coeficiente é utilizado como uma medida de desigualdade, sua interpretação como apresentado em Shlomo e Edna (2013) é de que o Coeficiente Gini é o GMD dividido pelo dobro da média. Neste caso, a média deve ser positiva.

Com base nesta definição e de que já sabemos qual a expressão que define o GMD, torna-se fácil entender a expressão que define o coeficiente de Gini, utilizado no trabalho, o qual é apresentado na equação 4.38.

$$G = \frac{\sum_{i=1}^{n}(2i - n - 1)x_i}{n\sum_{i=1}^{n} x_i} \tag{4.38}$$

Neste trabalho, o coeficiente de Gini será utilizado para estudar os resíduos de cada modelo gerado e entender o quão desiguais estes são em relação a um modelo ideal. Além do Coeficiente Gini, serão também utilizadas as Correlações Gini, apresentadas no capítulo 3.

4.7 MODELO COMPUTACIONAL

Para estimar diversas séries sintéticas com rapidez e versatilidade, foi necessário desenvolver um modelo computacional que abordasse todos os tópicos e necessidades dest trabalho. Desta forma, foi criado um modelo de gráfico e matemático, com base na linguagem Python 3. O modelo envolve todas as etapas necessárias, desde a geração das séries sintéticas até a estimação dos modelos Auto Regressivos, estimação dos erros,

gráficos e tabelas apresentados neste trabalho. Este modelo está disponível online sob o repositório do *GitLab* de domínio dos autores, sob o endereço https://gitlab.com/chilelli/gini-autoregressive-estimator.

Esta seção busca explicar ao leitor a lógica do modelo computacional desenvolvido, na forma de um algoritmo descritivo de cada passo e sua finalidade a cada etapa.

4.7.1 Fluxo do Modelo

O modelo compuatcional envolve três módulos distintos, assim criados visando facilitar a análise de *bugs* ou outros problemas decorrentes de características específicas dos testes feitos com o mesmo. Estes três módulos são:

- Módulo de Séries Sintéticas: Código que cria o as séries sintéticas.
- Módulo de Funções de Cálculo: Conjunto de funções que performam os cálculos Gini, Clássico e também a criação dos gráficos.
- Macro de Envelopagem e Execução do Processo: Código que organiza e invoca as funções em cada etapa lógica do processo, afim de obter o resultado final.

Visando facilitar o entendimento do leitor para com todas as váriaveis, funções e demais entradas e fluxos do processo, foi adotada uma lista de prefixos que seguem cada nome de varíavel, função, macro, módulo, *set* ou outro componente do código. Esta descrição está contida no Apêndice A.

Todas as etapas do modelo já foram descritas na seções anteriores, portanto, não serão enfatizadas aqui.

5 ANÁLISE DOS RESULTADOS

5.1 RESULTADOS

Este capítulo discute os resultados obtidos através da aplicação da metodologia Gini para a estimação dos parâmetros nos modelos AR(p). Conforme apresentado no capítulo 3 a metodologia Gini permite a estimação de dois coeficientes de correlação, um olhando a séries para frente $t_0, t_1, ..., t_n$ e outro olha a série para trás $t_n, t_{n-1}, ..., t_0$, aqui eles serão chamado de Operador *Forward* e Operador *Backward*, respectivamente. O método Semi-Paramétrico permite a obtenção de dois estimadores em consequência de suas propriedades, já o método de minimização devido às suas próprias características, resulta apenas em um resultado. Assim sendo, a estruturação deste capítulo será feita de maneira a apresentar separadamente cada tipo de distribuição com cauda longa (similar ao feito na seção 4.3.1), sob cada métrica de erro e coeficiente Gini.

Por questões estéticas, algumas tabelas e figuras possuem abreviações, de maneira a elucidar essas terminologias apresentadas neste capítulo, segue abaixo uma lista com seus respectivos significados.

- Gini Min/Min: Método Gini de Minimização do Resíduo;
- Gini SP/SP: Método Gini Semi-Paramétrico;
- Clássico/C: Método Clássico de estimação dos parâmetros AR.

As tabelas apresentadas a seguir possuem as seguintes características:

- Elas possuem os resultados do erro de previsão para cada modelo estimado, sob cada operador. A coluna “Modelos Gini Superiores” mostra o número de modelos Gini os quais obtiveram erros de previsão menores que o método clássico com tolerância de $\pm 1\%$. Nos casos onde a relação entre o método clássico e os modelos Gini ficam dentro da tolerância, é considerado que não há vantagem na utilização do Gini. No caso deste trabalho, os possíveis valores são 0 (nenhum modelo Gini foi superior ao método clássico), 1 (um modelo Gini foi superior), 2 (dois modelos Gini foram superiores) e 3 (Todos os modelos Gini foram superiores);
- As tabelas do tipo 5.5 apresentam a sumarização dos campos “Modelos Gini Superiores” das tabelas da seção, sob cada métrica de erro.

5.2 MODELOS GERADOS A PARTIR DE SÉRIES COM DISTRIBUIÇÕES BIMODAIS

Aqui serão apresentados os resultados obtidos para os modelos Bimodais. Para facilitar o entendimento do leitor, as médias e desvios padrões das distribuições normais formadoras da bimodal estão contidas nos conjuntos **m** e **d**, respectivamente.

Os campos em destaque nas tabelas mostram os modelos em que o método clássico foi superior ou igual ao Gini, esses casos serão abordados assim como os opostos, em que o desempenho Gini foi superior.

5.2.1 Desempenho dos modelos

Os erros calculados do MAPE são apresentados na tabela 5.1.

Tabela 5.1 – MAPE dos modelos de séries Bimodais

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Bimodal m = [4,12], d = [2,4], p = 1	1.18	1.37	1.27	1.37	0
Bimodal m = [4,12], d = [2,4], p = 2	1.54	1.35	1.35	1.30	3
Bimodal m = [4,12], d = [2,4], p = 3	3.22	2.96	3.05	2.84	3
Bimodal m = [4,12], d = [2,4], p = 4	2.39	2.05	2.36	1.96	3
Bimodal m = [5,13], d = [1,2], p = 1	1.09	1.21	0.99	0.99	2
Bimodal m = [5,13], d = [1,2], p = 2	2.12	2.00	2.03	1.71	3
Bimodal m = [5,13], d = [1,2], p = 3	1.71	1.61	1.66	1.56	3
Bimodal m = [5,13], d = [1,2], p = 4	1.60	1.46	1.53	1.50	3
Bimodal m = [6,14], d = [1,2], p = 1	1.00	1.05	1.00	1.00	2
Bimodal m = [6,14], d = [1,2], p = 2	1.78	1.72	1.64	1.60	3
Bimodal m = [6,14], d = [1,2], p = 3	1.97	1.82	1.91	1.76	3
Bimodal m = [6,14], d = [1,2], p = 4	1.60	1.47	1.59	1.53	3
Bimodal m = [7,15], d = [1,4], p = 1	1.06	1.11	1.08	1.10	0
Bimodal m = [7,15], d = [1,4], p = 2	1.64	1.46	1.53	1.45	3
Bimodal m = [7,15], d = [1,4], p = 3	3.50	3.55	3.29	3.12	2
Bimodal m = [7,15], d = [1,4], p = 4	1.79	1.79	1.74	1.85	1

O MAPE dos modelos apresentou bons resultado na grande maioria dos casos, os modelos gerados a partir da metodologia Gini foram igualadas apenas quando a ordem do modelo era baixa, igual a 1. Nos outros casos vemos que o Gini teve umo desempenho superior ao método clássico.

O MSD apresentou características similares ao MAPE, as quais serão discutidas conjuntamente ao final desta seção. Seus resultados estão contidos na tabela 5.2

Tabela 5.2 – MSD dos modelos de séries Bimodais

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Bimodal m = [4,12], d = [2,4], p = 1	235	237	236	237	0
Bimodal m = [4,12], d = [2,4], p = 2	388	376	377	380	3
Bimodal m = [4,12], d = [2,4], p = 3	455	436	431	450	3
Bimodal m = [4,12], d = [2,4], p = 4	454	443	435	449	3
Bimodal m = [5,13], d = [1,2], p = 1	249	249	248	248	3
Bimodal m = [5,13], d = [1,2], p = 2	352	352	349	361	0
Bimodal m = [5,13], d = [1,2], p = 3	527	504	503	516	3
Bimodal m = [5,13], d = [1,2], p = 4	432	420	407	411	3
Bimodal m = [6,14], d = [1,2], p = 1	262	263	263	262	0
Bimodal m = [6,14], d = [1,2], p = 2	417	400	400	401	3
Bimodal m = [6,14], d = [1,2], p = 3	406	398	385	399	3
Bimodal m = [6,14], d = [1,2], p = 4	480	468	436	481	2
Bimodal m = [7,15], d = [1,4], p = 1	239	237	237	237	0
Bimodal m = [7,15], d = [1,4], p = 2	391	390	388	399	2
Bimodal m = [7,15], d = [1,4], p = 3	481	469	465	470	3
Bimodal m = [7,15], d = [1,4], p = 4	401	396	373	418	2

Os resultados do MAE para séries Bimodais é apresentado na tabela 5.3.

Tabela 5.3 – MAE dos modelos de séries Bimodais

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Bimodal m = [4,12], d = [2,4], p = 1	1.08	1.09	1.09	1.09	0
Bimodal m = [4,12], d = [2,4], p = 2	1.39	1.38	1.38	1.39	0
Bimodal m = [4,12], d = [2,4], p = 3	1.51	1.49	1.48	1.51	2
Bimodal m = [4,12], d = [2,4], p = 4	1.51	1.50	1.49	1.51	2
Bimodal m = [5,13], d = [1,2], p = 1	1.12	1.12	1.12	1.12	0
Bimodal m = [5,13], d = [1,2], p = 2	1.33	1.34	1.33	1.36	0
Bimodal m = [5,13], d = [1,2], p = 3	1.62	1.60	1.60	1.62	3
Bimodal m = [5,13], d = [1,2], p = 4	1.47	1.46	1.44	1.45	3
Bimodal m = [6,14], d = [1,2], p = 1	1.14	1.15	1.15	1.15	0
Bimodal m = [6,14], d = [1,2], p = 2	1.44	1.42	1.42	1.42	3
Bimodal m = [6,14], d = [1,2], p = 3	1.43	1.42	1.40	1.42	3
Bimodal m = [6,14], d = [1,2], p = 4	1.55	1.55	1.49	1.57	1
Bimodal m = [7,15], d = [1,4], p = 1	1.09	1.09	1.09	1.09	1
Bimodal m = [7,15], d = [1,4], p = 2	1.40	1.41	1.41	1.43	0
Bimodal m = [7,15], d = [1,4], p = 3	1.55	1.54	1.54	1.54	3
Bimodal m = [7,15], d = [1,4], p = 4	1.42	1.42	1.38	1.46	1

O MAE, assim como o MSD e o MAPE, mostrou que nas ocasiões em os métodos Gini foram igualados ou parcialmente superados, foram em modelos de baixa ordem.

O MSE, assim como será visto na análise das outras séries modeladas, foi a métrica que mais indicou falhas nos modelos Gini. Esta característica será discutida ao final do capítulo. Os resultados do MSE para as séries em questão é apresentado na tabela 5.4.

Tabela 5.4 – MSE dos modelos de séries Bimodais

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Bimodal m = [4,12], d = [2,4], p = 1	2.31	2.33	2.32	2.33	0
Bimodal m = [4,12], d = [2,4], p = 2	3.71	3.74	3.74	3.78	0
Bimodal m = [4,12], d = [2,4], p = 3	4.00	3.97	3.93	3.99	1
Bimodal m = [4,12], d = [2,4], p = 4	3.94	3.87	3.81	3.92	3
Bimodal m = [5,13], d = [1,2], p = 1	2.14	2.15	2.15	2.15	0
Bimodal m = [5,13], d = [1,2], p = 2	3.19	3.27	3.23	3.31	0
Bimodal m = [5,13], d = [1,2], p = 3	4.41	4.27	4.27	4.33	3
Bimodal m = [5,13], d = [1,2], p = 4	3.66	3.61	3.55	3.56	3
Bimodal m = [6,14], d = [1,2], p = 1	2.28	2.30	2.29	2.29	0
Bimodal m = [6,14], d = [1,2], p = 2	3.68	3.60	3.61	3.62	3
Bimodal m = [6,14], d = [1,2], p = 3	3.85	3.77	3.76	3.79	3
Bimodal m = [6,14], d = [1,2], p = 4	4.19	4.07	3.96	4.16	3
Bimodal m = [7,15], d = [1,4], p = 1	2.23	2.23	2.23	2.23	0
Bimodal m = [7,15], d = [1,4], p = 2	3.67	3.80	3.74	3.85	0
Bimodal m = [7,15], d = [1,4], p = 3	4.21	4.12	4.11	4.18	3
Bimodal m = [7,15], d = [1,4], p = 4	3.57	3.55	3.39	3.69	2

O desempenho dos modelos Gini em relação ao método clássico para as séries com distribuições Bimodais é apresentada na tabela 5.5.

Tabela 5.5 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Bimodais.

	Modelos Gini Superiores			
Série	**MAPE**	**MSD**	**MAE**	**MSE**
Bimodal m = [4,12], d = [2,4],p = 1	0	0	0	0
Bimodal m = [4,12], d = [2,4], p = 2	3	3	0	0
Bimodal m = [4,12], d = [2,4], p = 3	3	3	2	1
Bimodal m = [4,12], d = [2,4], p = 4	3	3	2	3
Bimodal m = [5,13], d = [1,2], p = 1	2	3	0	0
Bimodal m = [5,13], d = [1,2], p = 2	3	0	0	0
Bimodal m = [5,13], d = [1,2], p = 3	3	3	3	3
Bimodal m = [5,13], d = [1,2], p = 4	3	3	3	3
Bimodal m = [6,14], d = [1,2], p = 1	2	0	0	0
Bimodal m = [6,14], d = [1,2], p = 2	3	3	3	3
Bimodal m = [6,14], d = [1,2], p = 3	3	3	3	3
Bimodal m = [6,14], d = [1,2], p = 4	3	2	1	3
Bimodal m = [7,15], d = [1,4], p = 1	0	0	1	0
Bimodal m = [7,15], d = [1,4], p = 2	3	2	0	0
Bimodal m = [7,15], d = [1,4], p = 3	2	3	3	3
Bimodal m = [7,15], d = [1,4], p = 4	1	2	1	2

O coeficiente Gini calculado dos resíduos do modelo são apresentados na tabela 5.6. É interessante observar que os valores dos coeficientes são muito proximos para os modelos em que o desempenho do método clássico foi superior ou igual ao da metodologia Gini, nos casos em que o Gini foi superior os valores dos coeficientes diferem consideravelmente em alguns casos.

Tabela 5.6 – Coeficientes de Gini dos resíduos nos modelos com séries Bimodais.

		Forward		Backward	
Série	Clássico	Gini SP	Gini Min	Gini SP	Gini Min
Bimodal m = [4,12], d = [2,4], p = 1	0.26	0.26	0.26	0.26	0.26
Bimodal m = [4,12], d = [2,4], p = 2	0.20	0.20	0.20	0.21	0.20
Bimodal m = [4,12], d = [2,4], p = 3	0.18	0.17	0.16	0.21	0.16
Bimodal m = [4,12], d = [2,4], p = 4	0.19	0.25	0.24	0.25	0.24
Bimodal m = [5,13], d = [1,2], p = 1	0.21	0.21	0.20	0.21	0.20
Bimodal m = [5,13], d = [1,2], p = 2	0.17	0.16	0.17	0.21	0.17
Bimodal m = [5,13], d = [1,2], p = 3	0.19	0.19	0.18	0.22	0.18
Bimodal m = [5,13], d = [1,2], p = 4	0.19	0.21	0.19	0.20	0.19
Bimodal m = [6,14], d = [1,2], p = 1	0.27	0.27	0.27	0.27	0.27
Bimodal m = [6,14], d = [1,2], p = 2	0.21	0.20	0.20	0.20	0.20
Bimodal m = [6,14], d = [1,2], p = 3	0.18	0.24	0.20	0.27	0.20
Bimodal m = [6,14], d = [1,2], p = 4	0.19	0.22	0.21	0.22	0.21
Bimodal m = [7,15], d = [1,4], p = 1	0.26	0.26	0.26	0.26	0.26
Bimodal m = [7,15], d = [1,4], p = 2	0.20	0.21	0.21	0.22	0.21
Bimodal m = [7,15], d = [1,4], p = 3	0.20	0.24	0.22	0.26	0.22
Bimodal m = [7,15], d = [1,4], p = 4	0.16	0.19	0.15	0.18	0.15

5.2.2 Considerações dos resultados para Séries Bimodais

Com base nas tabelas desta seção, é possível observar que o Gini foi superior ao método clássico na previsão das séries Bimodais com caudas longas. Uma métrica que não conseguiu indicar claramente esso desempenho, como pode ser observado na tabela 5.4 foi o MSE, em parte esta "falha" se deve ao fato de que o MSE assume simetria ou normalidade nos resíduos, fato que não aconteceu nesta modelagem, uma vez que as séries em sua totalidade, diferem da normalidade.

5.3 MODELOS GERADOS A PARTIR DE SÉRIES COM DISTRIBUIÇÕES WEIBULL

Os modelos Weibull apresentaram resultados extremamente favoráveis ao Gini, onde sob a ótica de três métricas quase não houveram modelos Gini inferiores ao clássico. A discussão dos resultados gerais destas séries é feita na seção 5.3.2.

5.3.1 Resultados dos Modelos

Como será constatado nas próximas seções e destacado na discussão, o MSE não é a melhor métrica para avaliação dos resultados do modelo, uma vez que ela não consegue medir com exatidão as características resíduais, uma vez que estes são não-normais. Na tabela 5.11 é possível constatar essa informação e ver que as outras três métricas são bastante favoráveis ao Gini.

O MAPE da tabela 5.7 para os modelos Weibull apresentou apenas uma resultado onde o Gini foi inferior, e como já era esperado, para um modelo de baixa ordem.

Tabela 5.7 – MAPE dos modelos de séries Weibull

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Weibull $\delta = 10$,$\beta = 0.5$, p = 1	0.98	0.98	0.98	0.98	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 2	2.93	4.44	2.11	2.20	2
Weibull $\delta = 10$, $\beta = 0.5$, p = 3	3.30	3.41	3.04	3.03	2
Weibull $\delta = 10$, $\beta = 0.5$, p = 4	4.09	3.54	3.79	3.51	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 1	1.28	1.21	1.25	1.27	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 2	1.91	1.70	1.58	1.41	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 3	1.76	1.56	1.64	1.54	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 4	2.12	1.95	2.04	2.05	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 1	1.46	1.44	1.05	1.05	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 2	1.60	1.49	1.49	1.38	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 3	1.61	1.62	1.60	1.52	2
Weibull $\delta = 8$, $\beta = 0.3$, p = 4	2.17	2.01	2.05	2.25	2
Weibull $\delta = 9$, $\beta = 0.4$, p = 1	6.37	4.17	6.26	6.46	2
Weibull $\delta = 9$, $\beta = 0.4$, p = 2	2.76	2.60	2.08	2.10	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 3	3.36	3.26	2.96	3.20	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 4	3.92	3.70	3.82	4.04	2

Assim como o MAPE o MSD na tabela 5.8 também foi favorável.

Tabela 5.8 – MSD dos modelos de séries Weibull

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Weibull $\delta = 10$,$\beta = 0.5$, p = 1	230	228	227	227	3
Weibull $\delta = 10$, $\beta = 0.5$, p = 2	431	425	422	422	3
Weibull $\delta = 10$, $\beta = 0.5$, p = 3	537	516	516	521	3
Weibull $\delta = 10$, $\beta = 0.5$, p = 4	438	432	408	445	2
Weibull $\delta = 11$, $\beta = 0.6$, p = 1	233	233	233	232	0
Weibull $\delta = 11$, $\beta = 0.6$, p = 2	437	428	427	427	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 3	488	464	458	463	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 4	404	379	372	382	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 1	212	210	207	207	1
Weibull $\delta = 8$, $\beta = 0.3$, p = 2	361	352	351	353	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 3	557	551	545	561	2
Weibull $\delta = 8$, $\beta = 0.3$, p = 4	453	442	426	464	2
Weibull $\delta = 9$, $\beta = 0.4$, p = 1	274	271	272	272	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 2	412	399	393	394	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 3	518	492	494	494	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 4	427	453	401	480	1

O MAE em 5.9 mostrou uma quantidade maior de problemas com o Gini, porém o padrão de desempenho nos modelos de ordem $p = 1, 2$ se repete, assim como acontece com o MSE na tabela 5.10.

Tabela 5.9 – MAE dos modelos de séries Weibull

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Weibull $\delta = 10$, $\beta = 0.5$, p = 1	1.07	1.07	1.07	1.07	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 2	1.47	1.47	1.46	1.46	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 3	1.64	1.62	1.62	1.63	3
Weibull $\delta = 10$, $\beta = 0.5$, p = 4	1.48	1.48	1.44	1.51	1
Weibull $\delta = 11$, $\beta = 0.6$, p = 1	1.08	1.09	1.08	1.08	0
Weibull $\delta = 11$, $\beta = 0.6$, p = 2	1.48	1.47	1.47	1.47	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 3	1.56	1.54	1.53	1.53	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 4	1.42	1.39	1.38	1.40	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 1	1.03	1.03	1.02	1.02	0
Weibull $\delta = 8$, $\beta = 0.3$, p = 2	1.34	1.33	1.33	1.34	0
Weibull $\delta = 8$, $\beta = 0.3$, p = 3	1.67	1.67	1.66	1.69	1
Weibull $\delta = 8$, $\beta = 0.3$, p = 4	1.50	1.50	1.47	1.54	1
Weibull $\delta = 9$, $\beta = 0.4$, p = 1	1.17	1.17	1.17	1.17	0
Weibull $\delta = 9$, $\beta = 0.4$, p = 2	1.44	1.42	1.41	1.41	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 3	1.61	1.58	1.58	1.58	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 4	1.46	1.52	1.43	1.56	1

Tabela 5.10 – MSE dos modelos de séries Weibull

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Weibull $\delta = 10$, $\beta = 0.5$, p = 1	2.09	2.09	2.09	2.09	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 2	3.91	3.94	3.93	3.94	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 3	4.36	4.27	4.26	4.30	3
Weibull $\delta = 10$, $\beta = 0.5$, p = 4	3.64	3.62	3.49	3.70	2
Weibull $\delta = 11$, $\beta = 0.6$, p = 1	1.98	2.00	1.99	2.00	0
Weibull $\delta = 11$, $\beta = 0.6$, p = 2	3.67	3.69	3.71	3.77	0
Weibull $\delta = 11$, $\beta = 0.6$, p = 3	4.09	4.02	4.00	4.02	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 4	3.50	3.37	3.27	3.40	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 1	1.92	1.93	1.94	1.94	0
Weibull $\delta = 8$, $\beta = 0.3$, p = 2	3.34	3.31	3.30	3.34	2
Weibull $\delta = 8$, $\beta = 0.3$, p = 3	4.75	4.78	4.75	4.79	0
Weibull $\delta = 8$, $\beta = 0.3$, p = 4	4.22	4.21	4.10	4.35	2
Weibull $\delta = 9$, $\beta = 0.4$, p = 1	2.29	2.29	2.29	2.29	0
Weibull $\delta = 9$, $\beta = 0.4$, p = 2	3.52	3.50	3.56	3.55	1
Weibull $\delta = 9$, $\beta = 0.4$, p = 3	4.51	4.42	4.39	4.40	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 4	3.94	3.99	3.80	4.20	1

Tabela 5.11 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Weibull.

	Modelos Gini Superiores			
Série	**MAPE**	**MSD**	**MAE**	**MSE**
Weibull $\delta = 10$, $\beta = 0.5$, p = 1	0	3	0	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 2	2	3	0	0
Weibull $\delta = 10$, $\beta = 0.5$, p = 3	2	3	3	3
Weibull $\delta = 10$, $\beta = 0.5$, p = 4	3	2	1	2
Weibull $\delta = 11$, $\beta = 0.6$, p = 1	3	0	0	0
Weibull $\delta = 11$, $\beta = 0.6$, p = 2	3	3	3	0
Weibull $\delta = 11$, $\beta = 0.6$, p = 3	3	3	3	3
Weibull $\delta = 11$, $\beta = 0.6$, p = 4	3	3	3	3
Weibull $\delta = 8$, $\beta = 0.3$, p = 1	3	1	0	0
Weibull $\delta = 8$, $\beta = 0.3$, p = 2	3	3	0	2
Weibull $\delta = 8$, $\beta = 0.3$, p = 3	2	2	1	0
Weibull $\delta = 8$, $\beta = 0.3$, p = 4	2	2	1	2
Weibull $\delta = 9$, $\beta = 0.4$, p = 1	2	3	0	0
Weibull $\delta = 9$, $\beta = 0.4$, p = 2	3	3	3	1
Weibull $\delta = 9$, $\beta = 0.4$, p = 3	3	3	3	3
Weibull $\delta = 9$, $\beta = 0.4$, p = 4	2	1	1	1

O coeficiente Gini para cada modelo é apresentado na figura 5.12

Tabela 5.12 – Coeficientes de Gini dos resíduos nos modelos com séries Weibull.

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Gini Min**
Weibull $\delta = 10$, $\beta = 0.5$, p = 1	0.27	0.27	0.27	0.27	0.27
Weibull $\delta = 10$, $\beta = 0.5$, p = 2	0.27	0.23	0.29	0.29	0.29
Weibull $\delta = 10$, $\beta = 0.5$, p = 3	0.24	0.23	0.24	0.24	0.24
Weibull $\delta = 10$, $\beta = 0.5$, p = 4	0.17	0.22	0.18	0.21	0.18
Weibull $\delta = 11$, $\beta = 0.6$, p = 1	0.20	0.20	0.20	0.21	0.20
Weibull $\delta = 11$, $\beta = 0.6$, p = 2	0.25	0.25	0.26	0.27	0.26
Weibull $\delta = 11$, $\beta = 0.6$, p = 3	0.28	0.33	0.32	0.33	0.32
Weibull $\delta = 11$, $\beta = 0.6$, p = 4	0.16	0.24	0.22	0.24	0.22
Weibull $\delta = 8$, $\beta = 0.3$, p = 1	0.24	0.24	0.23	0.23	0.23
Weibull $\delta = 8$, $\beta = 0.3$, p = 2	0.24	0.25	0.28	0.28	0.28
Weibull $\delta = 8$, $\beta = 0.3$, p = 3	0.22	0.22	0.22	0.25	0.22
Weibull $\delta = 8$, $\beta = 0.3$, p = 4	0.21	0.24	0.23	0.23	0.23
Weibull $\delta = 9$, $\beta = 0.4$, p = 1	0.22	0.23	0.22	0.22	0.22
Weibull $\delta = 9$, $\beta = 0.4$, p = 2	0.27	0.27	0.28	0.28	0.28
Weibull $\delta = 9$, $\beta = 0.4$, p = 3	0.26	0.29	0.29	0.30	0.29
Weibull $\delta = 9$, $\beta = 0.4$, p = 4	0.19	0.22	0.16	0.25	0.16

5.3.2 Considerações dos Resultados para Séries Weibull

Como dito no inicio da seção, os resultados para os modelos Weibull se mostraram totalmente favoráveis à aplicação dos métodos Gini para previsão de séries geradas a partir de dados com distribuições Weibull. Um ponto importante a se destacar aqui é que as séries de baixa ordem continuam não sendo bem modeladas sob a metodologia Gini. Casos onde $p = 1$ ou $p = 2$ apresentam resultados similares, assim como medidas de desigualdade dos resíduos e da correlação entre os mesmos e as observações das séries. Em casos onde $p > 2$ existe superioridade do Gini.

Outro ponto a ser destacado aqui, é que as séries com distribuições Weibull são melhor modeladas pelo método de minimização dos resíduos, uma vez que em grande parte dos resultados, este método é melhor que o Gini Semi-Paramétrico. Grande parte desta comprovação talvez se deva a premissa de linearidade, necessária ao Gini Semi-Paramétrico, os quais podem induzir o modelo a uma estimação errônea dos resultados. Para á serie AR(2) $\delta = 9$, $\beta = 0.4$ até mesmo o MSE teve uma leve tendência em favor do Gini.

5.4 MODELOS GERADOS A PARTIR DE SÉRIES COM DISTRIBUIÇÕES PARETO

Assim como as séries Weibull, os modelos para séries geradas a partir de dados com distribuições Pareto foram extremamente favoráveis em relação a metodologia Gini, onde quase todos os modelos feitos com base na metodologia Gini são superiores aos do método clássico.

O MAPE dos modelos para essas séries é apresentado na tabela 5.13. Aqui é possível ver que o Gini conseguiu modelar com eficiência até mesmo as séries de ordem $p = 1, 2$.

Tabela 5.13 – MAPE dos modelos de séries Pareto

		Forward		**Backward**	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Pareto $\sigma = 11$, $\alpha = 2$, p = 1	1.04	1.02	1.01	1.02	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 2	2.82	2.40	2.24	2.27	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 3	3.23	2.81	2.97	2.76	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 4	1.98	1.72	1.99	1.78	1
Pareto $\sigma = 2$, $\alpha = 2$, p = 1	1.06	1.04	1.04	1.05	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 2	5.88	3.92	4.60	3.93	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 3	3.49	3.47	3.43	3.22	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 4	3.22	2.83	3.32	3.11	1
Pareto $\sigma = 5$, $\alpha = 2$, p = 1	1.35	1.38	1.27	1.31	2
Pareto $\sigma = 5$, $\alpha = 2$, p = 2	1.41	1.28	1.29	1.22	3
Pareto $\sigma = 5$, $\alpha = 2$, p = 3	7.86	4.98	8.58	3.90	1
Pareto $\sigma = 5$, $\alpha = 2$, p = 4	4.78	3.53	5.08	3.30	1
Pareto $\sigma = 8$, $\alpha = 2$, p = 1	1.41	1.47	1.34	1.36	2
Pareto $\sigma = 8$, $\alpha = 2$, p = 2	2.07	1.93	1.78	1.67	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 3	2.06	1.96	2.04	1.84	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 4	1.83	1.70	1.74	1.75	3

O MSD, apresentado na tabela 5.14, assim como o MAPE, mostra superioridade nas previsões dos modelos Gini.

Tabela 5.14 – MSD dos modelos de séries Pareto

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Pareto $\sigma = 11$, $\alpha = 2$, p = 1	277	273	273	273	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 2	449	438	430	430	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 3	473	453	446	461	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 4	452	444	423	441	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 1	239	237	237	237	0
Pareto $\sigma = 2$, $\alpha = 2$, p = 2	396	393	393	394	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 3	565	572	550	575	1
Pareto $\sigma = 2$, $\alpha = 2$, p = 4	422	415	395	436	2
Pareto $\sigma = 5$, $\alpha = 2$, p = 1	249	249	249	249	2
Pareto $\sigma = 5$, $\alpha = 2$, p = 2	430	423	425	433	2
Pareto $\sigma = 5$, $\alpha = 2$, p = 3	527	520	507	527	3
Pareto $\sigma = 5$, $\alpha = 2$, p = 4	415	407	389	414	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 1	224	225	221	223	2
Pareto $\sigma = 8$, $\alpha = 2$, p = 2	431	417	414	413	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 3	496	485	485	493	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 4	413	388	376	415	2

O MAE das séries Pareto, apresentado na tabela 5.14, repete o padrão encontrado nas séries Weibull na tabela 5.9.

Tabela 5.15 – MAE dos modelos de séries Pareto

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Pareto $\sigma = 11$, $\alpha = 2$, p = 1	1.18	1.17	1.17	1.17	0
Pareto $\sigma = 11$, $\alpha = 2$, p = 2	1.50	1.49	1.47	1.47	1
Pareto $\sigma = 11$, $\alpha = 2$, p = 3	1.54	1.52	1.50	1.53	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 4	1.50	1.50	1.47	1.50	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 1	1.09	1.09	1.09	1.09	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 2	1.41	1.41	1.41	1.41	0
Pareto $\sigma = 2$, $\alpha = 2$, p = 3	1.68	1.70	1.67	1.71	1
Pareto $\sigma = 2$, $\alpha = 2$, p = 4	1.45	1.46	1.42	1.49	1
Pareto $\sigma = 5$, $\alpha = 2$, p = 1	1.12	1.12	1.12	1.12	0
Pareto $\sigma = 5$, $\alpha = 2$, p = 2	1.47	1.46	1.47	1.48	0
Pareto $\sigma = 5$, $\alpha = 2$, p = 3	1.62	1.62	1.60	1.64	1
Pareto $\sigma = 5$, $\alpha = 2$, p = 4	1.44	1.44	1.41	1.45	1
Pareto $\sigma = 8$, $\alpha = 2$, p = 1	1.06	1.07	1.06	1.07	0
Pareto $\sigma = 8$, $\alpha = 2$, p = 2	1.47	1.45	1.45	1.44	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 3	1.58	1.57	1.57	1.58	0
Pareto $\sigma = 8$, $\alpha = 2$, p = 4	1.44	1.41	1.38	1.46	2

O MSE em 5.16, como já esperado, é desfavorável à análise, segundo esta métrica o desempenho do método Gini foi ruim, quando comparado as outras medidas.

Tabela 5.16 – MSE dos modelos de séries Pareto

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Pareto $\sigma = 11$, $\alpha = 2$, p = 1	2.33	2.32	2.32	2.32	0
Pareto $\sigma = 11$, $\alpha = 2$, p = 2	3.94	3.92	3.94	3.94	0
Pareto $\sigma = 11$, $\alpha = 2$, p = 3	4.01	3.95	3.92	3.98	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 4	4.23	4.19	4.03	4.15	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 1	2.04	2.04	2.04	2.04	0
Pareto $\sigma = 2$, $\alpha = 2$, p = 2	3.62	3.66	3.66	3.67	0
Pareto $\sigma = 2$, $\alpha = 2$, p = 3	4.88	4.85	4.83	4.87	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 4	3.69	3.68	3.53	3.81	2
Pareto $\sigma = 5$, $\alpha = 2$, p = 1	2.03	2.04	2.04	2.04	0
Pareto $\sigma = 5$, $\alpha = 2$, p = 2	4.08	4.12	4.12	4.21	0
Pareto $\sigma = 5$, $\alpha = 2$, p = 3	4.66	4.62	4.58	4.66	3
Pareto $\sigma = 5$, $\alpha = 2$, p = 4	3.50	3.48	3.38	3.55	2
Pareto $\sigma = 8$, $\alpha = 2$, p = 1	1.86	1.89	1.89	1.90	0
Pareto $\sigma = 8$, $\alpha = 2$, p = 2	3.65	3.63	3.66	3.68	1
Pareto $\sigma = 8$, $\alpha = 2$, p = 3	4.20	4.16	4.15	4.19	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 4	3.79	3.67	3.60	3.83	2

O desempenho de cada grupo de modelo é apresentado na figura 5.17

Tabela 5.17 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico.

	Modelos Gini Superiores			
Série	**MAPE**	**MSD**	**MAE**	**MSE**
Pareto $\sigma = 11$, $\alpha = 2$, p = 1	3	3	0	0
Pareto $\sigma = 11$, $\alpha = 2$, p = 2	3	3	1	0
Pareto $\sigma = 11$, $\alpha = 2$, p = 3	3	3	3	3
Pareto $\sigma = 11$, $\alpha = 2$, p = 4	1	3	3	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 1	3	0	3	0
Pareto $\sigma = 2$, $\alpha = 2$, p = 2	3	3	0	0
Pareto $\sigma = 2$, $\alpha = 2$, p = 3	3	1	1	3
Pareto $\sigma = 2$, $\alpha = 2$, p = 4	1	2	1	2
Pareto $\sigma = 5$, $\alpha = 2$, p = 1	2	2	0	0
Pareto $\sigma = 5$, $\alpha = 2$, p = 2	3	2	0	0
Pareto $\sigma = 5$, $\alpha = 2$, p = 3	1	3	1	3
Pareto $\sigma = 5$, $\alpha = 2$, p = 4	1	3	1	2
Pareto $\sigma = 8$, $\alpha = 2$, p = 1	2	2	0	0
Pareto $\sigma = 8$, $\alpha = 2$, p = 2	3	3	3	1
Pareto $\sigma = 8$, $\alpha = 2$, p = 3	3	3	0	3
Pareto $\sigma = 8$, $\alpha = 2$, p = 4	3	2	2	2

Os coefiientes Gini para os modelos de séries Pareto está presente na figura 5.18. Aqui é possível notar que o padrão das ocorrências anteriores se repete para os casos em que o Gini não é superior ao método clássico.

Tabela 5.18 – Coeficientes de Gini dos resíduos nos modelos com séries Pareto

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Gini Min**
Pareto $\sigma = 11$, $\alpha = 2$, p = 1	0.24	0.24	0.24	0.24	0.24
Pareto $\sigma = 11$, $\alpha = 2$, p = 2	0.24	0.26	0.29	0.29	0.29
Pareto $\sigma = 11$, $\alpha = 2$, p = 3	0.21	0.23	0.21	0.25	0.21
Pareto $\sigma = 11$, $\alpha = 2$, p = 4	0.17	0.26	0.27	0.26	0.27
Pareto $\sigma = 2$, $\alpha = 2$, p = 1	0.27	0.27	0.27	0.27	0.27
Pareto $\sigma = 2$, $\alpha = 2$, p = 2	0.26	0.26	0.26	0.26	0.26
Pareto $\sigma = 2$, $\alpha = 2$, p = 3	0.17	0.19	0.15	0.20	0.15
Pareto $\sigma = 2$, $\alpha = 2$, p = 4	0.17	0.23	0.21	0.24	0.21
Pareto $\sigma = 5$, $\alpha = 2$, p = 1	0.24	0.24	0.24	0.24	0.24
Pareto $\sigma = 5$, $\alpha = 2$, p = 2	0.24	0.26	0.26	0.28	0.26
Pareto $\sigma = 5$, $\alpha = 2$, p = 3	0.22	0.22	0.22	0.23	0.22
Pareto $\sigma = 5$, $\alpha = 2$, p = 4	0.20	0.22	0.21	0.22	0.21
Pareto $\sigma = 8$, $\alpha = 2$, p = 1	0.26	0.26	0.26	0.26	0.26
Pareto $\sigma = 8$, $\alpha = 2$, p = 2	0.22	0.23	0.25	0.25	0.25
Pareto $\sigma = 8$, $\alpha = 2$, p = 3	0.24	0.26	0.24	0.25	0.24
Pareto $\sigma = 8$, $\alpha = 2$, p = 4	0.14	0.14	0.14	0.14	0.14

5.4.1 Considerações dos resultados para Séries Pareto

Os resultados dos modelos com séries Pareto foram bastante próximos aos encontrados para as séries Weibull, onde grande parte dos modelos Gini foi superior. Os ajustes de cada modelo se mostraram muito próximos para as séries com superioridade Gini, enquanto alguns resultados, de ordem mais baixa ($p = 1$ ou $p = 2$) apresentaram resultados de certa forma, equiparados. Este ponto, como já destacado, está presente também nos outros resultados.

5.4.2 Modelos gerados a partir de Séries com distribuições Log-Normal

Os resultados para as séries Log-Normal foram os que mais se afastaram dos outros, mesmo com uma quantidade maior de modelos clássicos melhores que o Gini, o padrão de desempenho em baixa ordem se repete. Mais uma vez é possível ver que o desempenho Gini em ordens $p >= 3$.

5.4.3 Resultados dos Modelos

O MAPE na tabela 5.19 mostra os resultados para as séries Log-Normal.

Tabela 5.19 – MAPE dos modelos de séries Log-Normal

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Log Normal $\theta = 3$, $\omega = 0.9$, p = 1	1.06	1.02	1.00	1.00	3
Log Normal $\theta = 3$, $\omega = 0.9$, p = 2	1.36	1.50	1.29	1.33	2
Log Normal $\theta = 3$, $\omega = 0.9$, p = 3	1.55	1.69	1.36	1.40	2
Log Normal $\theta = 3$, $\omega = 0.9$, p = 4	2.11	2.78	2.55	2.53	0
Log Normal $\theta = 3$, $\omega = 1$, p = 1	1.06	1.01	1.01	1.01	3
Log Normal $\theta = 3$, $\omega = 1$, p = 2	1.00	1.32	1.01	1.27	0
Log Normal $\theta = 3$, $\omega = 1$, p = 3	1.56	1.43	1.05	1.23	3
Log Normal $\theta = 3$, $\omega = 1$, p = 4	2.05	1.63	1.60	1.58	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 1	1.01	1.26	1.11	1.10	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 2	1.95	4.66	1.65	3.47	1
Log Normal $\theta = 4$, $\omega = 0.9$, p = 3	2.78	1.91	1.50	2.64	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 4	5.63	4.87	4.07	5.43	3
Log Normal $\theta = 4$, $\omega = 1$, p = 1	1.01	1.03	1.02	1.02	0
Log Normal $\theta = 4$, $\omega = 1$, p = 2	1.90	2.02	1.77	1.77	2
Log Normal $\theta = 4$, $\omega = 1$, p = 3	1.37	1.54	1.61	1.72	0
Log Normal $\theta = 4$, $\omega = 1$, p = 4	2.55	1.51	1.33	2.01	3

Tanto o MAPE, quanto o MSD e MAE nas tabelas 5.20 e 5.21 respectivamente, mostram um perfil parecido para as análises, com os modelos de baixa ordem tentendo a serem melhor modelados pelo método clássico.

Tabela 5.20 – MSD dos modelos de séries Log-Normal

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Log Normal $\theta = 3$, $\omega = 0.9$, p = 1	232	231	231	231	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 2	508	513	510	509	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 3	669	514	483	482	3
Log Normal $\theta = 3$, $\omega = 0.9$, p = 4	709	679	671	671	3
Log Normal $\theta = 3$, $\omega = 1$, p = 1	220	218	218	218	0
Log Normal $\theta = 3$, $\omega = 1$, p = 2	593	581	575	577	3
Log Normal $\theta = 3$, $\omega = 1$, p = 3	1282	746	738	711	3
Log Normal $\theta = 3$, $\omega = 1$, p = 4	930	725	686	774	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 1	267	268	267	267	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 2	578	570	577	568	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 3	1127	700	672	682	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 4	806	618	532	657	3
Log Normal $\theta = 4$, $\omega = 1$, p = 1	232	231	231	231	0
Log Normal $\theta = 4$, $\omega = 1$, p = 2	463	421	423	424	3
Log Normal $\theta = 4$, $\omega = 1$, p = 3	704	598	532	566	3
Log Normal $\theta = 4$, $\omega = 1$, p = 4	667	441	407	478	3

Tabela 5.21 – MAE dos modelos de séries Log-Normal

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Log Normal $\theta = 3$, $\omega = 0.9$, p = 1	1.08	1.08	1.08	1.08	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 2	1.59	1.61	1.60	1.60	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 3	1.83	1.62	1.57	1.56	3
Log Normal $\theta = 3$, $\omega = 0.9$, p = 4	1.88	1.85	1.84	1.84	3
Log Normal $\theta = 3$, $\omega = 1$, p = 1	1.05	1.05	1.05	1.05	0
Log Normal $\theta = 3$, $\omega = 1$, p = 2	1.72	1.71	1.70	1.70	3
Log Normal $\theta = 3$, $\omega = 1$, p = 3	2.53	1.95	1.94	1.90	3
Log Normal $\theta = 3$, $\omega = 1$, p = 4	2.16	1.91	1.86	1.98	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 1	1.16	1.16	1.16	1.16	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 2	1.70	1.69	1.70	1.69	1
Log Normal $\theta = 4$, $\omega = 0.9$, p = 3	2.37	1.89	1.85	1.86	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 4	2.01	1.77	1.64	1.83	3
Log Normal $\theta = 4$, $\omega = 1$, p = 1	1.08	1.08	1.08	1.08	0
Log Normal $\theta = 4$, $\omega = 1$, p = 2	1.52	1.46	1.46	1.46	3
Log Normal $\theta = 4$, $\omega = 1$, p = 3	1.88	1.74	1.64	1.70	3
Log Normal $\theta = 4$, $\omega = 1$, p = 4	1.83	1.50	1.44	1.56	3

O MSE novamente não conseguiu refletir o resultados das medidas anteriores, como pode ser constatado na tabela 5.22.

Tabela 5.22 – MSE dos modelos de séries Log-Normal

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Log Normal $\theta = 3$, $\omega = 0.9$, p = 1	2.16	2.17	2.17	2.17	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 2	5.65	5.71	5.69	5.68	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 3	8.12	5.43	5.01	5.09	3
Log Normal $\theta = 3$, $\omega = 0.9$, p = 4	11.67	12.23	11.88	11.8	0
Log Normal $\theta = 3$, $\omega = 1$, p = 1	1.98	1.99	1.99	1.99	0
Log Normal $\theta = 3$, $\omega = 1$, p = 2	14.06	14.77	14.06	14.6	0
Log Normal $\theta = 3$, $\omega = 1$, p = 3	29.20	10.74	6.09	8.24	3
Log Normal $\theta = 3$, $\omega = 1$, p = 4	20.04	17.89	18.84	18.0	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 1	2.34	2.36	2.35	2.35	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 2	11.80	12.38	11.85	12.0	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 3	21.07	9.22	6.59	7.98	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 4	13.41	8.96	9.17	9.48	3
Log Normal $\theta = 4$, $\omega = 1$, p = 1	1.96	1.96	1.96	1.96	0
Log Normal $\theta = 4$, $\omega = 1$, p = 2	4.81	3.98	3.99	3.99	3
Log Normal $\theta = 4$, $\omega = 1$, p = 3	9.53	6.56	5.44	5.83	3
Log Normal $\theta = 4$, $\omega = 1$, p = 4	9.23	5.03	4.73	6.07	3

Os resultados de desempenho dos modelos para séries com distribuições Log-Normal são apresentados na tabela 5.23

Tabela 5.23 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Log-Normal.

	Modelos Gini Superiores			
Série	**MAPE**	**MSD**	**MAE**	**MSE**
Log Normal $\theta = 3$, $\omega = 0.9$, p = 1	3	0	0	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 2	2	0	0	0
Log Normal $\theta = 3$, $\omega = 0.9$, p = 3	2	3	3	3
Log Normal $\theta = 3$, $\omega = 0.9$, p = 4	0	3	3	0
Log Normal $\theta = 3$, $\omega = 1$, p = 1	3	0	0	0
Log Normal $\theta = 3$, $\omega = 1$, p = 2	0	3	3	0
Log Normal $\theta = 3$, $\omega = 1$, p = 3	3	3	3	3
Log Normal $\theta = 3$, $\omega = 1$, p = 4	3	3	3	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 1	0	0	0	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 2	1	3	1	0
Log Normal $\theta = 4$, $\omega = 0.9$, p = 3	3	3	3	3
Log Normal $\theta = 4$, $\omega = 0.9$, p = 4	3	3	3	3
Log Normal $\theta = 4$, $\omega = 1$, p = 1	0	0	0	0
Log Normal $\theta = 4$, $\omega = 1$, p = 2	2	3	3	3
Log Normal $\theta = 4$, $\omega = 1$, p = 3	0	3	3	3
Log Normal $\theta = 4$, $\omega = 1$, p = 4	3	3	3	3

Os coeficientes de Gini dos resíduos são apresentados na tabela 5.24

Tabela 5.24 – Coeficientes de Gini dos resíduos nos modelos com séries Log-Normal.

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Gini Min**
Log Normal $\theta = 3$, $\omega = 0.9$, p = 1	0.24	0.24	0.24	0.24	0.24
Log Normal $\theta = 3$, $\omega = 0.9$, p = 2	0.17	0.14	0.19	0.18	0.19
Log Normal $\theta = 3$, $\omega = 0.9$, p = 3	0.11	0.07	0.08	0.08	0.08
Log Normal $\theta = 3$, $\omega = 0.9$, p = 4	0.08	0.05	0.06	0.06	0.06
Log Normal $\theta = 3$, $\omega = 1$, p = 1	0.24	0.23	0.23	0.23	0.23
Log Normal $\theta = 3$, $\omega = 1$, p = 2	0.25	0.10	0.25	0.11	0.25
Log Normal $\theta = 3$, $\omega = 1$, p = 3	0.08	0.04	0.11	0.05	0.11
Log Normal $\theta = 3$, $\omega = 1$, p = 4	0.06	0.05	0.04	0.06	0.04
Log Normal $\theta = 4$, $\omega = 0.9$, p = 1	0.23	0.23	0.23	0.23	0.23
Log Normal $\theta = 4$, $\omega = 0.9$, p = 2	0.29	0.10	0.29	0.15	0.29
Log Normal $\theta = 4$, $\omega = 0.9$, p = 3	0.09	0.05	0.08	0.06	0.08
Log Normal $\theta = 4$, $\omega = 0.9$, p = 4	0.07	0.06	0.05	0.06	0.05
Log Normal $\theta = 4$, $\omega = 1$, p = 1	0.21	0.21	0.21	0.21	0.21
Log Normal $\theta = 4$, $\omega = 1$, p = 2	0.19	0.17	0.20	0.20	0.20
Log Normal $\theta = 4$, $\omega = 1$, p = 3	0.10	0.07	0.09	0.08	0.09
Log Normal $\theta = 4$, $\omega = 1$, p = 4	0.10	0.08	0.07	0.07	0.07

5.4.4 Considerações dos resultados para Séries Log-Normal

Os resultados observados para as séries com distribuição Log-Normal mostram novamente a tendência de superioridade da metodologia Gini em modelos com ordem $p > 2$, para os modelos de ordens inferiores, é possível observar que boa parte dos resultados apresentam o método clássico como um estimador suficiente. O perfil dos coeficientes Gini continua seguindo o padrão, em séries onde o método clássico apresentou superioridade, vemos que os coeficientes Gini não variam.

Os modelos onde o método clássico teve desempenho melhor que o Gini não apresentam diferenças tão significativas em seus resíduos, já quando a situação é oposta, existem grandes diferenças no comportamento residual, isso pode ser um dos motivos dos problemas com as medidas do MSE.

5.5 MODELOS GERADOS A PARTIR DE SÉRIES COM DISTRIBUIÇÕES DISTRIBUIÇÕES BURR

Os resultados com as séries Burr foram favoráveis ao Gini, assim como as Bimodais, que mostraram superioridade, porém com alguns casos de exceção, mas nenhum deles em ordens $p < 3$.

5.5.1 Resultados dos modelos

O MAPE é apresentado na tabela 5.25. No caso destes resultados é possível ver que o desempenho do Operador *Forward* foi muito inferior ao do Operador *Backward*.

Tabela 5.25 – MAPE dos modelos de séries Burr

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Burr c = 2, k = 6, p = 1	1.31	1.48	1.42	1.51	0
Burr c = 2, k = 6, p = 2	3.38	3.16	2.32	2.07	3
Burr c = 2, k = 6, p = 3	4.49	3.86	3.85	3.43	3
Burr c = 2, k = 6, p = 4	1.48	1.51	1.43	1.61	1
Burr c = 2, k = 7, p = 1	1.12	1.24	1.10	1.11	2
Burr c = 2, k = 7, p = 2	1.68	1.64	1.46	1.35	3
Burr c = 2, k = 7, p = 3	3.68	3.20	3.90	3.17	1
Burr c = 2, k = 7, p = 4	1.66	1.53	1.62	1.58	3
Burr c = 3, k = 6, p = 1	1.08	1.12	1.05	1.08	1
Burr c = 3, k = 6, p = 2	7.24	7.25	5.49	5.78	1
Burr c = 3, k = 6, p = 3	2.33	2.13	2.09	2.05	3
Burr c = 3, k = 6, p = 4	2.82	2.44	2.70	2.40	3
Burr c = 3, k = 7, p = 1	1.00	1.01	1.01	1.00	0
Burr c = 3, k = 7, p = 2	2.38	2.09	2.25	2.01	3
Burr c = 3, k = 7, p = 3	2.30	2.13	2.14	1.99	3
Burr c = 3, k = 7, p = 4	1.58	1.49	1.53	1.38	3
Burr c = 4, k = 6, p = 1	1.03	1.00	1.05	1.08	1
Burr c = 4, k = 6, p = 2	2.91	2.79	2.03	2.01	3
Burr c = 4, k = 6, p = 3	2.48	2.35	2.43	2.19	3
Burr c = 4, k = 6, p = 4	1.84	1.61	1.79	1.71	3
Burr c = 4, k = 7, p = 1	1.02	1.00	1.01	1.01	1
Burr c = 4, k = 7, p = 2	1.84	2.02	1.76	1.84	1
Burr c = 4, k = 7, p = 3	2.56	2.47	2.39	2.35	3
Burr c = 4, k = 7, p = 4	1.43	1.28	1.38	1.33	3

É possível ver um diferença entre as medidas do MSD em 5.26 e do MAE em 5.27, essa diferença requer uma análise mais profunda de problemáticas com a dispersão e comportamento residual.

Tabela 5.26 – MSD dos modelos de séries Burr

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Burr c = 2, k = 6, p = 1	241	241	241	241	0
Burr c = 2, k = 6, p = 2	438	429	425	425	3
Burr c = 2, k = 6, p = 3	563	534	534	539	3
Burr c = 2, k = 6, p = 4	386	389	350	406	1
Burr c = 2, k = 7, p = 1	274	277	275	275	0
Burr c = 2, k = 7, p = 2	390	381	379	381	3
Burr c = 2, k = 7, p = 3	465	454	453	455	3
Burr c = 2, k = 7, p = 4	393	396	354	428	1
Burr c = 3, k = 6, p = 1	248	248	247	247	2
Burr c = 3, k = 6, p = 2	450	447	443	444	3
Burr c = 3, k = 6, p = 3	546	517	516	515	3
Burr c = 3, k = 6, p = 4	460	439	422	447	3
Burr c = 3, k = 7, p = 1	226	224	224	224	0
Burr c = 3, k = 7, p = 2	402	408	400	421	1
Burr c = 3, k = 7, p = 3	524	500	497	499	3
Burr c = 3, k = 7, p = 4	465	447	437	466	2
Burr c = 4, k = 6, p = 1	264	265	265	265	0
Burr c = 4, k = 6, p = 2	381	378	362	363	3
Burr c = 4, k = 6, p = 3	443	446	425	442	2
Burr c = 4, k = 6, p = 4	443	413	410	419	3
Burr c = 4, k = 7, p = 1	287	285	285	285	3
Burr c = 4, k = 7, p = 2	386	376	379	378	3
Burr c = 4, k = 7, p = 3	536	514	508	516	3
Burr c = 4, k = 7, p = 4	375	381	339	395	1

Tabela 5.27 – MAE dos modelos de séries Burr

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Burr c = 2, k = 6, p = 1	1.10	1.10	1.10	1.10	0
Burr c = 2, k = 6, p = 2	1.48	1.47	1.46	1.47	3
Burr c = 2, k = 6, p = 3	1.68	1.65	1.65	1.65	3
Burr c = 2, k = 6, p = 4	1.39	1.41	1.34	1.44	1
Burr c = 2, k = 7, p = 1	1.17	1.18	1.18	1.18	0
Burr c = 2, k = 7, p = 2	1.40	1.39	1.38	1.39	1
Burr c = 2, k = 7, p = 3	1.52	1.52	1.52	1.52	3
Burr c = 2, k = 7, p = 4	1.40	1.42	1.35	1.48	1
Burr c = 3, k = 6, p = 1	1.11	1.12	1.11	1.11	0
Burr c = 3, k = 6, p = 2	1.50	1.50	1.49	1.49	3
Burr c = 3, k = 6, p = 3	1.65	1.62	1.62	1.62	3
Burr c = 3, k = 6, p = 4	1.52	1.50	1.47	1.51	3
Burr c = 3, k = 7, p = 1	1.06	1.06	1.06	1.06	0
Burr c = 3, k = 7, p = 2	1.42	1.44	1.43	1.47	0
Burr c = 3, k = 7, p = 3	1.62	1.59	1.59	1.59	3
Burr c = 3, k = 7, p = 4	1.53	1.51	1.49	1.54	2
Burr c = 4, k = 6, p = 1	1.15	1.15	1.15	1.15	0
Burr c = 4, k = 6, p = 2	1.38	1.39	1.36	1.36	2
Burr c = 4, k = 6, p = 3	1.49	1.51	1.47	1.50	1
Burr c = 4, k = 6, p = 4	1.49	1.45	1.44	1.46	3
Burr c = 4, k = 7, p = 1	1.20	1.20	1.20	1.20	0
Burr c = 4, k = 7, p = 2	1.39	1.38	1.38	1.38	3
Burr c = 4, k = 7, p = 3	1.64	1.62	1.61	1.62	3
Burr c = 4, k = 7, p = 4	1.37	1.39	1.32	1.42	1

O MSE na tabela 5.28 como já esperado, mostra o Goini inferior em grande parte dos casos para os modelos de baixa ordem.

Tabela 5.28 – MSE dos modelos de séries Burr

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Modelos Gini Superiores**
Burr c = 2, k = 6, p = 1	2.28	2.29	2.29	2.29	0
Burr c = 2, k = 6, p = 2	4.08	4.13	4.18	4.23	0
Burr c = 2, k = 6, p = 3	4.62	4.52	4.53	4.59	3
Burr c = 2, k = 6, p = 4	3.49	3.47	3.18	3.60	2
Burr c = 2, k = 7, p = 1	2.34	2.36	2.35	2.35	0
Burr c = 2, k = 7, p = 2	3.73	3.76	3.77	3.80	0
Burr c = 2, k = 7, p = 3	4.10	4.10	4.08	4.09	3
Burr c = 2, k = 7, p = 4	3.49	3.44	3.36	3.59	2
Burr c = 3, k = 6, p = 1	2.22	2.23	2.22	2.22	0
Burr c = 3, k = 6, p = 2	4.14	4.15	4.19	4.17	0
Burr c = 3, k = 6, p = 3	4.50	4.38	4.37	4.40	3
Burr c = 3, k = 6, p = 4	4.09	3.90	3.89	3.97	3
Burr c = 3, k = 7, p = 1	1.99	1.99	1.99	1.99	0
Burr c = 3, k = 7, p = 2	3.65	3.73	3.69	3.80	0
Burr c = 3, k = 7, p = 3	4.18	4.11	4.06	4.14	3
Burr c = 3, k = 7, p = 4	4.23	4.09	4.08	4.17	3
Burr c = 4, k = 6, p = 1	2.29	2.30	2.30	2.31	0
Burr c = 4, k = 6, p = 2	3.48	3.58	3.61	3.63	0
Burr c = 4, k = 6, p = 3	3.61	3.62	3.57	3.63	1
Burr c = 4, k = 6, p = 4	3.78	3.67	3.60	3.73	3
Burr c = 4, k = 7, p = 1	2.47	2.47	2.47	2.47	0
Burr c = 4, k = 7, p = 2	3.54	3.55	3.55	3.55	0
Burr c = 4, k = 7, p = 3	4.21	4.14	4.11	4.20	3
Burr c = 4, k = 7, p = 4	3.31	3.33	3.16	3.38	1

O desempenho para os modelos com séries Burr estão descritos na tabela 5.29.

Tabela 5.29 – Comparação do desempenho dos modelos Gini em relação aos modelos do método clássico para as séries Burr.

	Modelos Gini Superiores			
Série	**MAPE**	**MSD**	**MAE**	**MSE**
Burr c = 2, k = 6, p = 1	0	0	0	0
Burr c = 2, k = 6, p = 2	3	3	3	0
Burr c = 2, k = 6, p = 3	3	3	3	3
Burr c = 2, k = 6, p = 4	1	1	1	2
Burr c = 2, k = 7, p = 1	2	0	0	0
Burr c = 2, k = 7, p = 2	3	3	1	0
Burr c = 2, k = 7, p = 3	1	3	3	3
Burr c = 2, k = 7, p = 4	3	1	1	2
Burr c = 3, k = 6, p = 1	1	2	0	0
Burr c = 3, k = 6, p = 2	1	3	3	0
Burr c = 3, k = 6, p = 3	3	3	3	3
Burr c = 3, k = 6, p = 4	3	3	3	3
Burr c = 3, k = 7, p = 1	0	0	0	0
Burr c = 3, k = 7, p = 2	3	1	0	0
Burr c = 3, k = 7, p = 3	3	3	3	3
Burr c = 3, k = 7, p = 4	3	2	2	3
Burr c = 4, k = 6, p = 1	1	0	0	0
Burr c = 4, k = 6, p = 2	3	3	2	2
Burr c = 4, k = 6, p = 3	3	2	1	1
Burr c = 4, k = 6, p = 4	3	3	3	3
Burr c = 4, k = 7, p = 1	1	3	0	0
Burr c = 4, k = 7, p = 2	1	3	3	0
Burr c = 4, k = 7, p = 3	3	3	3	3
Burr c = 4, k = 7, p = 4	3	1	1	1

Os resultados para os coeficientes Gini dos resíduos nos modelos de séries com distribuições Burr estão contidos na tabela 5.30.

Tabela 5.30 – Coeficientes de Gini dos resíduos nos modelos com séries Burr

		Forward		Backward	
Série	**Clássico**	**Gini SP**	**Gini Min**	**Gini SP**	**Gini Min**
Burr c = 2, k = 6, p = 1	0.24	0.23	0.24	0.23	0.24
Burr c = 2, k = 6, p = 2	0.22	0.20	0.27	0.26	0.27
Burr c = 2, k = 6, p = 3	0.20	0.18	0.19	0.24	0.19
Burr c = 2, k = 6, p = 4	0.19	0.26	0.27	0.26	0.27
Burr c = 2, k = 7, p = 1	0.24	0.24	0.24	0.24	0.24
Burr c = 2, k = 7, p = 2	0.17	0.15	0.18	0.20	0.18
Burr c = 2, k = 7, p = 3	0.21	0.21	0.21	0.22	0.21
Burr c = 2, k = 7, p = 4	0.18	0.19	0.16	0.21	0.16
Burr c = 3, k = 6, p = 1	0.23	0.23	0.23	0.23	0.23
Burr c = 3, k = 6, p = 2	0.24	0.24	0.23	0.23	0.23
Burr c = 3, k = 6, p = 3	0.23	0.22	0.22	0.22	0.22
Burr c = 3, k = 6, p = 4	0.19	0.20	0.19	0.20	0.19
Burr c = 3, k = 7, p = 1	0.21	0.21	0.21	0.21	0.21
Burr c = 3, k = 7, p = 2	0.25	0.26	0.26	0.28	0.26
Burr c = 3, k = 7, p = 3	0.22	0.28	0.29	0.28	0.29
Burr c = 3, k = 7, p = 4	0.17	0.25	0.24	0.29	0.24
Burr c = 4, k = 6, p = 1	0.29	0.29	0.29	0.29	0.29
Burr c = 4, k = 6, p = 2	0.25	0.24	0.24	0.24	0.24
Burr c = 4, k = 6, p = 3	0.18	0.21	0.16	0.22	0.16
Burr c = 4, k = 6, p = 4	0.20	0.25	0.26	0.25	0.26
Burr c = 4, k = 7, p = 1	0.30	0.30	0.30	0.30	0.30
Burr c = 4, k = 7, p = 2	0.20	0.19	0.21	0.21	0.21
Burr c = 4, k = 7, p = 3	0.18	0.19	0.17	0.21	0.17
Burr c = 4, k = 7, p = 4	0.16	0.25	0.26	0.25	0.26

5.5.2 Considerações dos resultados para Séries Burr

Os resultados observados para as séries com distribuição Burr mostram que este tipo de série foi a que mais apresentou desempenhos variáveis através dos métodos. Mas

aqui a constatação de que o Gini performa melhor em séries com ordem $p > 2$ ficou clara novamente. As análises dos coefientes de correlações Gini mostraram que quando destas ocorrências, estes são claramente iguais ou divergem pouco em valor.

6 CONCLUSÃO

A metodologia Gini se mostrou bastante adequada na previsão de séries temporais que apresentam distribuições com caudas longas. Na grande maioria dos resultados houve superioridade da previsão Gini. Um fato a ser levantado é que o Gini teve melhor desempenho que o método clássico em modelos de ordem superior ($p >= 3$), porém, mesmo nos modelos de ordem inferior ($p = 1, 2$), em grande parte dos casos estudados, o desempenho da metodologia foi igual ou pouco inferior à do método clássico, o que mostra sua robustez. Outro ponto a ser levantado aqui é a capacidade dos modelos Gini em obterem melhores resultados quando não se necessita da suposição de modelos a serem estabelecidos, permitindo ao investigador uma maior liberdade quando realizando estudos de investigação em séries temporais, as quais ele desconhece características intrínsecas.

Outro fato importante observado foi que a utilização do MSE como métrica para este estudo, mostrou que esta não é uma boa métrica a ser utilizada e caso fosse a única, levaria a conclusões totalmente precipitadas em relação aos modelos. Isso pode ser causado pelas características intrínsecas de cálculo do MSE, as quais são dependentes de variância, enquanto as outras métricas são dependentes apenas dos desvios dos próprios resíduos em relação aos valores observados, fator o qual é ponto forte da metodologia Gini. Uma opção em aberto e que pode ser uma boa oportunidade para estudos futuros, seria a aplicação de outras técnicas de mensuração de erros para os modelos aqui criados, tais como o RMSE e o SMAPE, técnicas essas que podem proporcionar outros tipos de análises e conclusões para iluminar ainda mais o entendimento e utilização dos métodos de Gini.

Com base nos métodos Gini também é possível identificar locais na ACF e PACF onde existem grandes tendências de ocorrência de valores extremos ao longo da distribuição da série. Muitas ACF e PACF foram capazes de identificar *lags* nos quais ocorriam a influência de valores extremos. Nos casos em que existem grandes diferenças entre os ACF e PACF nos operadores *Backward* e *Forward* podemos constatar que Y_t e Y_{t-s} não são "trocaveis", resultando em diferentes resultados sob os dois operadores. Isso pode ser constatado pela diferença nos resultados dos modelos onde esse fato ocorre.

Um fato interessante decorrente deste estudo foi a capacidade do AIC em prever corretamente a ordem dos modelos sintéticos, mesmo estes fugindo da normalidade, e que segundo os próprios entendimentos do método mostram que sua robustez pode ser questionada em cenários de não-normalidade. Para entender esta afirmação são necessários estudos mais aprofundados, os quais não foram o foco deste trabalho.

São necessárias maiores investigações a respeito de como melhorar o desempenho do método Gini para modelos de baixa ordem ($p = 1, 2$) e também como utilizar as outras

intepretações de aplicação da metodologia em conjunto com as técnicas estudadas neste trabalho para, talvez, desenvolver modelos ainda mais robustos e como capacidade de rpevisão melhorada para as séries aqui estudadas.

Uma grande oportunidade para desenvolvimentos futuros é a aplicação e replicação deste estudo com séries reais, como em energias renováveis, finanças e planejamento. Atentando-se ao fato de que as séries devem respeitar os requisitos desta metodologia. Este trabalho mostrou que os métodos Gini podem oferecer diversas vantagens e maior robustez ao processo de previsão, o que pode ser benéfico nas mais diversas áreas do conhecimento. Transportar este estudo para aplicações reais pode ser a última porta para a generalização das aplicações dos métodos de covariância Gini em outros campos.

REFERÊNCIAS

Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, December 1974. ISSN 0018-9286. Citado na página 38.

ARNOLD, B. C. Pareto and generalized pareto distributions. In: _____. *Modeling Income Distributions and Lorenz Curves*. New York, NY: Springer New York, 2008. p. 119–145. ISBN 978-0-387-72796-7. Citado na página 34.

BARNETT, V.; GREEN, P. J.; ROBINSON, A. Concomitants and correlation estimates. *Biometrika*, v. 63, p. 323–328, 1976. Citado na página 19.

BOX, G. E.; M., G.; JENKINS, G. C. R. *Time series analysis : forecasting and control*. 4. ed. [S.l.]: John Wiley, 2008. 756 p. ISBN 978-0-470-27284-8. Citado 4 vezes nas páginas 9, 11, 13 e 28.

BROCKWELL, D. R. A. *Time series: theory and methods*. 2. ed. [S.l.]: Springer, 1991. Citado 2 vezes nas páginas 13 e 26.

BROCKWELL, P. J. *Introduction to time series and forecasting*. 2. ed. [S.l.]: Springer, 2002. 449 p. ISBN 0-387-95351-5. Citado 2 vezes nas páginas 8 e 38.

BURR, I. W. Cumulative frequency functions. *Ann. Math. Statist.*, The Institute of Mathematical Statistics, v. 13, n. 2, p. 215–232, 06 1942. Disponível em: <https://doi.org/10.1214/aoms/1177731607>. Citado na página 36.

CARAPELLUCCI, R.; GIORDANO, L. A new approach for synthetically generating wind speeds: A comparison with the markov chains method. *Energy*, v. 49, p. 298 – 305, 2013. ISSN 0360-5442. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0360544212007943>. Citado na página 3.

CARCEA, M.; SERFLING, R. A gini autocovariance function for time series modelling. *Journal of Time Series Analysis*, v. 36, n. 6, p. 817–838, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12130>. Citado na página 30.

CARCEA, M. D. Gini autocovariance function used for time series with heavy-tail distributions. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 10, n. 3, p. e1428, 2018. Citado na página 1.

CHUNG-HO, C.; CHAO-YU, C. Optimum process mean, standard deviation and specification limits settings under the burr distribution. *Engineering Computations*, v. 34, n. 1, p. 66–76, 2017. Citado na página 36.

DANIELS, H. E. The relation between measures of correlation in the universe of sample permutation. *Biometrika*, v. 33, p. 129–135, 1944. Citado 2 vezes nas páginas 4 e 19.

DANIELS, H. E. A property of rank correlation. *Biometrika*, v. 35, p. 416–417, 1948. Citado na página 4.

DAVIS, R. A.; RESNICK, S. More limit theory for the sample correlation function of moving averages. *Stoch Process Their Appl*, v. 20, p. 257–279, 1985. Citado na página 26.

DAVIS, R. A.; RESNICK, S. I. Limit theory for bilinear processes with heavy-tailed noise. *The Annals of Applied Probability*, The Institute of Mathematical Statistics, v. 6, n. 4, p. 1191–1210, 1996. Citado na página 2.

DETZEL, D.; MINE, M. Generation of daily synthetic precipitation series: Analyses and application in la plata river basin. *The Open Hydrology Journal*, v. 5, 05 2011. Citado na página 3.

DEVORE, J. L. *Probability and Statistics for Engineering and the Sciences*. [S.l.]: Brooks/Cole, 2010. 776 p. ISBN 978-0-538-73352-6. Citado na página 32.

DORFMAN, R. A formula for the gini coefficient. *Review of Economics and Statistics*, v. 61, p. 146–149, 1979. Citado na página 16.

EBRAHIM, N.; GHOLAMHOSSEIN, H.; MEHDI, J.; HOSSEIN, F.; MORTEZA, M. Safety performance evaluation in a steel industry: A short-term time series approach. *Safety Science*, v. 110, p. 285 – 290, 2018. ISSN 0925-7535. Citado na página 38.

EISENBERGER, I. Genesis of bimodal distributions. *Technometrics*, Taylor and Francis, v. 6, n. 4, p. 357–363, 1964. Disponível em: <https://www.tandfonline.com/doi/abs/10-.1080/00401706.1964.10490199>. Citado na página 35.

FALK, M.; REISS, R.-D. Functional laws of small numbers. In: _____. *Extreme Value Theory and Applications: Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1 Gaithersburg Maryland 1993*. Boston, MA: Springer US, 1994. p. 337–354. ISBN 978-1-4613-3638-9. Citado na página 31.

FEIGIN, P. D.; RESNICK, S. I. Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes*, v. 1, p. 391–422, 1999. Citado 2 vezes nas páginas 2 e 26.

FERREIRA, M. On the extremal behavior of a pareto process: an alternative for armax modeling. *Kybernetika*, v. 1, 01 2012. Citado 2 vezes nas páginas 32 e 34.

FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, v. 24, n. 2, p. 180–190, 1928. Citado na página 31.

GEMIGNANI, M.; ROSTEGUI, G.; KAGAN, N. e. a. Solar radiation synthetic series for power purchase agreements. *Environ Sci Pollut Res*, 2018. Citado na página 3.

GHITANY, M. E.; GóMEZ-DéNIZ, E.; NADARAJAH, S. A new generalization of the pareto distribution and its application to insurance data. *Journal of Risk and Financial Management*, v. 11, n. 1, 2018. ISSN 1911-8074. Disponível em: <http://www.mdpi.com/1911-8074/11/1/10>. Citado na página 33.

GILI, A.; BETTUZZI, G. *Italian contributions to the methodology of statistics*. [S.l.]: Societa Italiana di Statistica, 1987. 231–243 p. Citado na página 17.

HARTER, H. L. A chronological annotated bibliography of order statistics. Washington, DC: U.S. Government Printing Office., v. 1, 1978. Citado na página 14.

HOEFFDING, W. A class of statistics with asymptotically normal distribution. *Annals of Mathematical statistics*, v. 19, p. 293–325, 1948. Citado na página 20.

IOANNIDIS, E. E. Akaike's information criterion correction for the least-squares autoregressive spectral estimator. *Journal of Time Series Analysis*, v. 32, n. 6, p. 618–630, 2011. Citado na página 38.

KENDALL, M. G. *Rank correlation methods*. 2. ed. [S.l.]: Hafner Publishing Company, 1955. 549 p. Citado na página 19.

KENDALL, M. G. *Rank correlation methods*. 3. ed. [S.l.]: Griffin and Co, 1962. 549 p. Citado na página 19.

KIZILERSü, A.; KREER, M.; THOMAS, A. W. The weibull distribution. *Significance*, v. 15, n. 2, p. 10–11, 2018. Citado na página 34.

KLUGMAN, S. A.; PANJER, H. H.; E., G. *Loss Models: From Data to Decisions*. 4. ed. [S.l.]: Wiley, 2012. ISBN 978-1-118-31532-3. Citado na página 3.

LAMBERT, P. J.; DECOSTER, A. The gini coefficient reveals more. *International Journal of Statistics*, Metron, v. 1, p. 373–400, 2005. Citado 2 vezes nas páginas 4 e 15.

LIMPERT, E.; STAHEL, W.; ABBT, M. Log-normal distributions across the sciences: Keys and clues. *Significance*, v. 51, n. 5, p. 341–352, 2018. Citado na página 32.

MCNEIL, A. J. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, Cambridge University Press, v. 27, n. 1, p. 117–137, 1997. Citado na página 31.

MONTGOMERY, D. C. *Introduction to time series analysis and forecasting*. [S.l.]: John Wiley, Sons, 2008. 469 p. ISBN 978-0-4 71-65397-4. Citado 6 vezes nas páginas viii, 7, 8, 33, 34 e 39.

MOORE, H. Cours d'Économie politique. by vilfredo pareto, professeur à l'université de lausanne. vol. i. pp. 430. i896. vol. ii. pp. 426. i897. lausanne: F. rouge. *The ANNALS of the American Academy of Political and Social Science*, v. 9, n. 3, p. 128–131, 1897. Citado na página 33.

OLKIN, I.; YITZHAKI, S. Gini regression analysis. *International Statistical Review*, v. 60, p. 185–196, 1992. Citado na página 23.

PEREIRA, G.; SOUZA, R. C. Long memory models to generate synthetic hydrological series. *Mathematical Problems in Engineering*, v. 2014, p. 1–8, 07 2014. Citado na página 3.

ROSENBERG, M. A.; FREES, E. W.; SUN, J.; H., P.; JR., J.; ROBINSON, J. Predictive modeling with longitudinal data. *North American Actuarial Journal*, Routledge, v. 11, n. 3, p. 54–69, 2007. Disponível em: <https://doi.org/10.1080/10920277.2007.10597466>. Citado na página 3.

SCHECHTMAN, E.; YITZHAKI, S. On the proper bounds of the gini correlation. *Economics Letters*, v. 63, p. 133–138, 1999. Citado na página 20.

SCHECHTMAN, E.; YITZHAKI, S. Calculating the extended gini coefficient from grouped data: a covariance presentation approach. *Bulletin of Statistics and Economics*, v. 2, p. 64–69, 2008. Citado na página 24.

SCHECHTMAN, E.; YITZHAKI, S.; TAINA, P. Gini's multiple regressions: two approaches and their interaction. *Metron*, p. 67–99, 2011. Citado 3 vezes nas páginas 20, 22 e 24.

SCHWEIZER, B.; WOLFF, E. F. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, v. 9, p. 879–885, 1981. Citado na página 20.

SERFLING, R. Fitting autoregressive models via yule-walker equations allowing heavy tail innovations. 08 2010. Citado 3 vezes nas páginas 2, 25 e 29.

SERFLING, R.; XIAO, P. A contribution to multivariate l-moments: L-comoment matrices. *Journal of Multivariate Analysis*, v. 98, p. 1765–1781, 10 2007. Citado na página 20.

SERFLING, R. J. *Approximation theorems of mathematical statistics*. [S.l.]: John Wiley and Sons, 1980. Citado na página 16.

SHELEF, A. A gini-based unit root test. *Computational Statistics and Data Analysis*, v. 100, p. 763 – 772, 2016. ISSN 0167-9473. Citado na página 2.

SHELEF, A.; SCHECHTMAN, E. A gini-based methodology for identifying and analyzing time series with non-normal innovations. 2011. Citado 4 vezes nas páginas 2, 25, 27 e 28.

SHELEF, A.; SCHECHTMAN, E. A gini-based time series analysis and test for reversibility. *Statistical Papers*, 11 2016. Citado 2 vezes nas páginas 1 e 27.

SHLOMO, Y.; EDNA, S. *The Gini Methodology*. [S.l.]: Springer, 2013. 549 p. ISBN 978-1-4614-4719-1. Citado 8 vezes nas páginas 2, 10, 14, 16, 17, 18, 21 e 39.

STUART, A. The correlation between variate-values and ranks in samples from a continuous distributions. *British Journal of Statistical Psychology*, v. 7, p. 37–44, 1954. Citado na página 19.

TAYLOR, M. A. Fosgerau's travel time reliability ratio and the burr distribution. *Transportation Research Part B: Methodological*, v. 97, p. 50 – 63, 2017. ISSN 0191-2615. Citado na página 36.

WEI, W. W. S. *Time Series analysis: univariate and multivariate methods*. 2. ed. [S.l.]: Person Education, 2006. Citado 2 vezes nas páginas 11 e 28.

WOLD, H. A study of the mean difference, concentration curves and concentration ratios. *Metron*, v. 12, n. 2, p. 39–58, 1935. Citado na página 16.

YEH, H.-C.; ARNOLD, B. C.; ROBERTSON, C. A. Pareto processes. *Journal of Applied Probability*, Applied Probability Trust, v. 25, n. 2, p. 291–301, 1988. ISSN 00219002. Disponível em: <http://www.jstor.org/stable/3214437>. Citado na página 33.

YITZHAKI, S. Gini's mean difference: A superior measure of variability for non-normal distributions. *Metron - International Journal of Statistics*, LXI, p. 285–316, 02 2003. Citado 2 vezes nas páginas 2 e 20.

ZHENG, A.; KUSIAK, A. Prediction of wind farm power ramp rates: A data-mining approach. *Journal of Solar Energy Engineering*, Elsevier, v. 131, n. 10, p. 1–8, 2009. Citado na página 7.

Apêndices

APÊNDICE A – CÓDIGOS ORIGINAIS DO MODELO COMPUTACIONAL

A.1 MODELO COMPUTACIONAL

Este apêndice tem como objetivo apresentar todos os códigos e funções utilizadas no desenvolvimento deste trabalho. Como apresentado no capítulo 4, a linguagem utilizada foi Python 3.

Este códigos também podem ser encontrados no repositório do *GitLab*, sob o endereço https://gitlab.com/chilelli/gini-autoregressive-estimator.

Para facilitar a leitura e entendimento, foi criada uma lista de regras de nomenclatura para as diversas variáveis do modelo. Esta lista é apresentada a seguir.

- *STRING* (Texto) = sAlgumaString
- INTEIROS = iAlgumInteiro
- REAL / DECIMAL = nAlbumNúmero
- *BOOLEAN* (Lógico) = bAlgumBoolean
- CONJUNTO DE *STRING* = ssAlgumConjuntoDeString
- CONJUNTO DE INTEIROS / REAIS = siAlgumConjuntoDeInteiros
- *DATASET* = dAlgumDataset
- FUNCÃO = FuncAlgumafunção
- MÓDULO = modAlgumModulo

Visando manter um padrão alguns termos foram mantidos com base na terminologia internacional, para esclarecer ao leitor, seguem definições sucintas de cada um deles:

- *STRING*: Texto
- *BOOLEAN*: Operador lógico, Verdadeiro ou Falso
- *DATASET*: Matriz de dados com propriedades especiais para operações complexas e tratamentos específicos.

Um ponto importante a se levantar aqui é que o modelo foi originalmente construído para lidar com qualquer série temporal, sejam elas auto regressivas, médias móveis ou os modelos ARIMA, desta forma, muitas partes dos códigos apresentam parâmetros ou argumentos que parecem redundantes quando se considera a modelagem de série puramente auto regressivas, porém, sua concepção original foi para lidar com a mais diversa gama de séries, por isso, eles apresentam esta estrutura.

A.1.1 FuncCreateMatrixRank - Transformações nos Dados

Esta função é responsável por criar a matriz contendo os *ranks*, as observações centralizadas na média (pode ser na mediana, basta trocar a função mean() por median()), junto a esta função está embutido um teste Dick-Fuller, para teste a necessidade de estacionarizar a série, no caso, fazendo uma diferença entre os valores. No caso deste trabalho, como as séries sintéticas já eram estacionárias, esta parte da função ficou "inativa".

```
def FuncCreateMatrixRank(dMatrix_ranks, dData, sCol):

    from statsmodels.tsa.stattools import adfuller
    from pandas import Series
    from numpy import diff

    for f in range(len(sCol)):
        dMatrix_ranks[sCol] = dData[sCol] - dData[sCol].mean()
        Fuller = adfuller(dMatrix_ranks[sCol])

        if Fuller[4]['10%'] < Fuller[0]:
            dif = Series(diff(dMatrix_ranks[sCol], n=1))
            dMatrix_ranks[sCol][1:] = dif

    dMatrix_ranks.drop(dMatrix_ranks.index[0], inplace=True)
    return(dMatrix_ranks)
```

A.1.2 FuncsiGACFOne e FuncsiGACFTwo - ACF de Gini

As duas funções apresentadas abaixo são as responsáveis por calcular os Gini-ACF com bases na covariâncias calculadas e fornecidas pelo algoritmo da seção **??**.

```
def FuncsiGACFOne(siGACF_1, dMatrix_ranks, sCol, iNLags, siGini_Covs):

    from numpy import cov

    for b in range(1,iNLags+1):
```

```
        dCumLag = (dMatrix_ranks[sCol][:-b].rank(method='average',
                                    na_option='top'))/len(dMatrix_ranks[sCol][:-b])
        nNumerator = cov(dMatrix_ranks[sCol][b:],dCumLag)[0][1]
        nDenominator = cov(dMatrix_ranks[sCol],dMatrix_ranks['CDF_' +
            sCol])[0][1]
        siGini_Covs[sCol].append(4*nNumerator)
        siGACF_1[sCol].append(nNumerator/nDenominator)

    return(siGACF_1, siGini_Covs)


def FuncsiGACFTwo(siGACF_2, dMatrix_ranks, sCol, iNLags, siGini_Covs):

    from numpy import cov

    for b in range(1,iNLags+1):
        dCumLag = (dMatrix_ranks[sCol][:-b].rank(method='average',
                        na_option='top'))/len(dMatrix_ranks[sCol][:-b])
        nNumerator = cov(dMatrix_ranks[sCol][:-b],dMatrix_ranks['CDF_' +
            sCol][b:])[0][1]
        nDenominator = cov(dMatrix_ranks[sCol][:-b], dCumLag)[0][1]
        siGini_Covs[sCol].append(4*nNumerator)
        siGACF_2[sCol].append(nNumerator/nDenominator)
    return(siGACF_2, siGini_Covs)
```

A.1.3 FuncCalcTheGiniCoefs - PACF de Gini

Esta é a função responsável por calcular os coeficientes em cada *lag* do sistema Yule-Walker, a função segue a metodologia apresentada e obtém apenas o último termo ϕ_{kk} em cada iteração do modelo até o *lag* máximo, definido por iNLags.

```
def FuncCalcTheGiniCoefs(iNLags, siGACF, sCol, siPartial, siReverse):
    for t in range(2, iNLags+1):
            up = siGACF[sCol][t-1] - sum([c*d for c,d in
                zip(siPartial[t-1],siReverse[-(t-1):])])
            down = 1 - sum([c*d for c,d in
                zip(siPartial[t-1],siGACF[sCol][:(t-1)])])
            siPartial[t][t] = up/down

            for j in range(1,t):
                siPartial[t][j] = siPartial[t-1][j] - siPartial[t][t] *
                    siPartial[t-1][t-j]
```

```
    return(siPartial)
```

A.1.4 FuncCoefGini - Correlação Gini

Com os resultados da função que cálcula os coeficientes do Sistema Yule-Walker, é possível obter os coeficientes de correlação Gini, que servirão com parâmetros do modelo Auto Regressivo com ordem p.

```
def FuncCoefGini(sCol, siGACF_1, siGACF_2):
    siCoef_Gar_1, siCoef_Gar_2 = ([] for x in range(2))

    siCoef_Gar_1.append(siGACF_1[sCol][0]) # Lag 1 Partial Autocorrelation
    siCoef_Gar_1.append((siGACF_1[sCol][1] - siGACF_1[sCol][0]**2)/(1 -
        siGACF_1[sCol][0]**2)) # Lag 2 Partial Autocorrelation

    siCoef_Gar_2.append(siGACF_2[sCol][0]) # Lag 1 Partial Autocorrelation
    siCoef_Gar_2.append((siGACF_2[sCol][1] - siGACF_2[sCol][0]**2)/(1 -
        siGACF_2[sCol][0]**2)) # Lag 2 Partial Autocorrelation

    return(siCoef_Gar_1, siCoef_Gar_2)
```

A.1.5 Coeficiente e Correlações Gini

Com base na metodologia apresentada, foram construídas duas funções com o objetivo de calcular os coeficientes de Gini entre dois conjuntos de dados, elas são apresentadas a seguir.

```
def FuncGiniCorrOne(snX, snY):

    from numpy import cov
    from pandas import Series

    dCumLagY = Series((snY.rank(method='average', na_option='top')/len(snY)) -
        1/2)
    dCumLagX = Series((snX.rank(method='average', na_option='top')/len(snX)) -
        1/2)

    nNumerator = cov(snX, dCumLagY)[0][1]
    nDenominator = cov(snX, dCumLagX)[0][1]
    nGiniCorrelation = nNumerator/nDenominator
    return(nGiniCorrelation)
```

```
def FuncGiniCorrTwo(snX, snY):
    from numpy import cov
    from pandas import Series

    dCumLagY = Series((snY.rank(method='average',na_option='top')/len(snY)) -
        1/2)
    dCumLagX = Series((snX.rank(method='average',na_option='top')/len(snX)) -
        1/2)

    nNumerator = cov(snY, dCumLagX)[0][1]
    nDenominator = cov(snY, dCumLagY)[0][1]
    nGiniCorrelation = nNumerator/nDenominator
    return(nGiniCorrelation)
```

Para calcular o coeficiente de Gini foi utilizado o seguinte algoritmo, contribuição da usuária **Olivia Guest** presente no seguinte https://github.com/oliviaguest/gini no *GitHub*.

```
def fGiniCoefficient(array):
    import numpy as np

    array = array.values
    array = array.flatten()
    if np.amin(array) < 0:
        array -= np.amin(array)
    array += 0.0000001
    array = np.sort(array)
    index = np.arange(1,array.shape[0]+1)
    n = array.shape[0]
    return ((np.sum((2 * index - n - 1) * array)) / (n * np.sum(array)))
        #Coeficiente Gini
```

A.1.6 FuncFindBestARIMA - Seleção do Modelo

Este seção é de vital importância para o trabalho, o fluxograma dela foi apresentado na seção **??**. Este é o algoritmo que testa todas as possíveis combinações de parâmetros do modelo.

A princípio ele foi construído para testar modelos ARIMA, por isso existe um esquema de iteração com três termos, p, d e q, os quais são os parâmetros de um modelo ARIMA clássico. No caso do código abaixo, d e q são definidos com valor zero, caracteri-

zando assim, um processo puramente Auto Regressivo.

```
def FuncFindBestARIMA(iPmin, iPmax, sCol, dMatrix_ranks, dData_central):

    from statsmodels.tsa.arima_model import ARIMA
    from itertools import product
    from numpy import arange

    dct = {}

    siP = arange(iPmin, iPmax, 1)
    d = q = [0]

    siPDQ = list(product(siP,d,q))

    dct[sCol] = []
    for siParam in siPDQ:
        try:
            model = ARIMA(dMatrix_ranks[sCol].dropna(), order = siParam)
            results = model.fit()
            dct[sCol].append({'P': siParam, 'AIC': results.aic})
        except:
            continue

    from pandas import concat
    from pandas import DataFrame

    dict_of_best_arima = {k: DataFrame(v) for k,v in dct.items()}
    dBestArima = concat(dict_of_best_arima, axis=1)

    siArrange = []
    siArrange.append(dBestArima[sCol]['AIC'].idxmin())

    dBestParam = []
    for append_best in range(len(siArrange)):
        dBestParam.append(dBestArima[sCol]['P'][siArrange[append_best]])

    siForecast = []
    dFits = DataFrame()

    model = ARIMA(dData_central[sCol].dropna(), order = dBestParam[0])
    model_fit = model.fit(trend='c')
```

```
    dFits['Res_'+ sCol] = model_fit.resid
    dFits['Fitted_'+ sCol] = model_fit.fittedvalues
    siForecast.append(model_fit.forecast(steps=1))

    return(dBestParam, siForecast, dFits, model_fit.arparams)
```

A.1.7 FuncGiniPureAR - Gini Minimização

Esta função é a responsável por calcular o modelo AR com base na metodologia Gini de minimização. O algoritmo de minimização escolhido foi o BFGS, como já apresentado na seção ??.

```
def FuncGiniPureAR(dData_central, sCol, dBestParam, dMatrix_GAR_Coef1):

    from pandas import DataFrame, Series
    from numpy import arange, mean, zeros, empty, cov
    from scipy.optimize import minimize

    siForecasts = {}
    sDataAR = DataFrame()


    siForecasts[sCol] = []
    dDataGini = dData_central[sCol]

    if dBestParam[0][0] != 0:

        sDataAR[sCol] = dDataGini


        dDataGini.reset_index(drop=True,inplace=True)
        dMatrix_of_Lags = DataFrame(index = dDataGini.index.values, columns =
            arange(1,dBestParam[0][0]+1,1))
        for p in range(1, dBestParam[0][0]+1):
            dMatrix_of_Lags[p][(dBestParam[0][0]-p): -p] =
                dDataGini[(dBestParam[0][0]-p): -p] #The n# of columns is due
                to the order of this specific model.
        dMatrix_of_Lags = dMatrix_of_Lags.apply(lambda x:
            Series(x.dropna().values))
        sisiAR_Coeff = dMatrix_GAR_Coef1[sCol][:dBestParam[0][0]]

        error = Series()
```

```
        dDataGini = dDataGini[dBestParam[0][0]:]
        dDataGini = dDataGini.reset_index(drop=True)

        def f(params):
            rho = empty(dBestParam[0][0])
            rho = params
            error = dDataGini - Series.sum(dMatrix_of_Lags * rho, axis=1)
            rank_error = error.rank(method='average',
                na_option='top')/len(error)
            return cov(error, rank_error)[0][1]

        alpha = minimize(f, zeros(dBestParam[0][0]), method='BFGS', tol=1e-10,
            constraints={'type': 'eq', 'fun' : mean(error) - 0})

        for p in range(dBestParam[0][0]):
            siForecasts[sCol].append(0)
        siForecasts[sCol].extend(Series.sum(dMatrix_of_Lags*alpha.x[:],axis=1)
            )


    siResidualsGini = {}
    siResidualsGini['Residual' + sCol] =
        (dData_central[sCol][dBestParam[0][0]:] -
        siForecasts[sCol][dBestParam[0][0]:])

    return(siForecasts, sisiAR_Coeff, siResidualsGini)
```

A.1.8 FuncGiniARSemiParametric - Gini Semi-Paramétrico

Esta função é a responsável por calcular o modelo AR com base na metodologia Gini Semi-Paramétrico, apresentada na seção 3.6.1.

```
def FuncGiniARSemiParametric(dData_central, sCol, dBestParam,
    dMatrix_GAR_Coef1):

    from pandas import DataFrame, Series
    from numpy import arange

    dForecasts_SP = {}

    dForecasts_SP[sCol] = []
    dDataGini = dData_central[sCol]
```

```
    dDataGini.reset_index(drop=True,inplace=True)
    dMatrix_of_Lags = DataFrame(index = dDataGini.index.values, columns =
        arange(1,dBestParam[0][0]+1,1))
    for p in range(1, dBestParam[0][0]+1):
        dMatrix_of_Lags[p][(dBestParam[0][0]-p): -p] =
            dDataGini[(dBestParam[0][0]-p): -p]
    dMatrix_of_Lags = dMatrix_of_Lags.apply(lambda x:
        Series(x.dropna().values))
    siAR_Coeff = dMatrix_GAR_Coef1[sCol][:dBestParam[0][0]]

    for p in range(dBestParam[0][0]):
        dForecasts_SP[sCol].append(0)
    dForecasts_SP[sCol].extend(Series.sum(dMatrix_of_Lags*siAR_Coeff,axis=1))

    siResidualsGini = {}
    siResidualsGini['Residual'+sCol] = (dData_central[sCol][dBestParam[0][0]:]
        - dForecasts_SP[sCol][dBestParam[0][0]:])

    return(dForecasts_SP, siAR_Coeff, siResidualsGini)
```

A.1.9 FuncErrorMetrics - Cálculo dos Erros

Para avaliar a performance de cada modelo, foram construídas estruturas para armazenar estes resultados. A função a seguir é responsável por calcular os erros de cada modelo de um forma que permita fácil comparação e entendimento por parte do pesquisador.

```
def FuncErrorMetrics(sCol, dData_central, dFits, dForecasts, dBestParam,
    dForecasts_SP):

    import pandas as pd
    from sklearn.metrics import mean_absolute_error,mean_squared_error

    siMAE, siMSE, siMAPE, siMSD = ({} for dic in range(4))

    siMAE['ARIMA_'+ sCol], siMSE['ARIMA_' + sCol], siMAPE['ARIMA_' + sCol],
        siMSD['ARIMA_' + sCol] = ([] for y in range(4))
    siMAE['GINI_M_'+ sCol], siMSE['GINI_M_' + sCol], siMAPE['GINI_M_' + sCol],
        siMSD['GINI_M_' + sCol] = ([] for y in range(4))
    siMAE['GINI_SP_'+ sCol], siMSE['GINI_SP_'+ sCol], siMAPE['GINI_SP_'+
```

```
    sCol], siMSD['GINI_SP_'+ sCol] = ([] for y in range(4))

siMAE['ARIMA_'+sCol].append(mean_absolute_error(dData_central[sCol],
    dFits['Fitted_'+sCol]))
siMSE['ARIMA_'+sCol].append(mean_squared_error(dData_central[sCol],
    dFits['Fitted_'+sCol]))
#Mean Absolute Percentage Error (siMAPE)
siMAPE['ARIMA_'+sCol].append((1/len(dFits))*sum(abs(((dData_central[sCol]
    - dFits['Fitted_'+sCol])/(dData_central[sCol])))))
#Mean Squared Deviation
siMSD['ARIMA_'+sCol].append((1/len(dFits))*sum(abs(dData_central[sCol] -
    dFits['Fitted_'+sCol]))**2)

siMAE['GINI_M_'+sCol].append(mean_absolute_error(
dData_central[sCol][dBestParam[0][0]:],dForecasts[sCol][dBestParam[0][0]:]))

siMSE['GINI_M_'+sCol].append(mean_squared_error(
dData_central[sCol][dBestParam[0][0]:],
    dForecasts[sCol][dBestParam[0][0]:]))

#Mean Absolute Percentage Error (siMAPE)
siMAPE['GINI_M_'+sCol].append((1/len(dForecasts[sCol][dBestParam[0][0]:]))*
sum(abs(((dData_central[sCol][dBestParam[0][0]:] -
    dForecasts[sCol][dBestParam[0][0]:])/
(dData_central[sCol][dBestParam[0][0]:])))))

#Mean Squared Deviation
siMSD['GINI_M_'+sCol].append((1/len(dForecasts[sCol][dBestParam[0][0]:]))*
sum(abs(dData_central[sCol][dBestParam[0][0]:] -
    dForecasts[sCol][dBestParam[0][0]:]))**2)

siMAE['GINI_SP_'+sCol].append(mean_absolute_error(
dData_central[sCol][dBestParam[0][0]:],
    dForecasts_SP[sCol][dBestParam[0][0]:]))

siMSE['GINI_SP_'+sCol].append(mean_squared_error(
dData_central[sCol][dBestParam[0][0]:],
    dForecasts_SP[sCol][dBestParam[0][0]:]))

#Mean Absolute Percentage Error (siMAPE)
siMAPE['GINI_SP_'+sCol].append((1/len(
dForecasts_SP[sCol][dBestParam[0][0]:]))*sum(abs(((
```

```
    dData_central[sCol][dBestParam[0][0]:] -
        dForecasts_SP[sCol][dBestParam[0][0]:])/
    (dData_central[sCol][dBestParam[0][0]:])))))

    #Mean Squared Deviation
    siMSD['GINI_SP_'+sCol].append((1/len(
    dForecasts_SP[sCol][dBestParam[0][0]:]))*sum(abs(
    dData_central[sCol][dBestParam[0][0]:] -
        dForecasts_SP[sCol][dBestParam[0][0]:]))**2)

    dError_metrics = pd.DataFrame.from_dict([siMAE, siMSE, siMAPE, siMSD])
    dError_metrics.index = ['MAE', 'MSE', 'MAPE', 'MSD']

    return(dError_metrics)
```

A.1.10 FuncNormalityResiduals - Teste de Normalidade

O estudo dos resíduos foi feito por meio do teste Shapiro de normalidade, apresentado a seguir.

```
def FuncNormalityResiduals(dFits, sCol):
    from scipy.stats import shapiro

    siResiduals_Shapiro = shapiro(dFits)
    if (siResiduals_Shapiro[1]<0.05):
        iIsnormal = 0
    else:
        iIsnormal = 1

    return(siResiduals_Shapiro, iIsnormal)
```

A.1.11 Gráficos de ACF e PACF

Para entender o comportamento das séries foram gerados os gráficos de Auto Correlação e Auto Correlação Parcial, apresentados a seguir.

Gráficos de Auto Correlação:

```
def FuncPlotAutoCorrelations(dMatrix_ranks, iNLags, sCol, siGACF_1, siGACF_2,
    sOperator, sGoodOrBad):
```

```
from matplotlib.pyplot import legend, title, ylabel, xlabel, xticks,
    yticks, show, rcParams, subplot, stem, axhline, ylim, savefig, figure
from statsmodels.graphics.tsaplots import plot_acf
from numpy import arange, sqrt

param = ''.join(map(str, [int(s) for s in sCol.split("_")[:-2] if
    s.isdigit()]))
sName = sCol.split("_")[0] + param
significance = (+1.96)/sqrt(len(dMatrix_ranks))

rcParams["figure.figsize"] = [15,5]

plot_acf(dMatrix_ranks, lags = iNLags,
        alpha = 0.05)
title('Pearson ACF ', fontsize = 22)
ylabel('Autocorrelation', fontsize = 18)
xlabel('Lags', fontsize = 18)
xticks(fontsize = 16)
yticks(fontsize = 16)
xticks(arange(0,iNLags+1,1), arange(0,iNLags+1,1))
show()

figure(figsize=(18,6))

subplot(1, 2, 1)
stem(siGACF_1)
title('Forward ACF de Gini ', fontsize = 20)
axhline(significance, linestyle='--',color='k')
axhline(-significance, linestyle='--',color='k')
xticks(arange(0,iNLags+1,1), arange(1,iNLags+1,1))
ylabel('Auto Correlacao', fontsize = 18)
xlabel('Lags', fontsize = 18)
xticks(fontsize = 16)
yticks(fontsize = 16)
ylim(top=1)

subplot(1, 2, 2)
stem(siGACF_2)
title('Backward ACF de Gini', fontsize = 20)
axhline(significance, linestyle='--',color='k')
axhline(-significance, linestyle='--',color='k')
xticks(arange(0,iNLags+1,1), arange(1,iNLags+1,1))
```

```
    xlabel('Lags', fontsize = 18)
    xticks(fontsize = 16)
    yticks(fontsize = 16)
    ylim(top=1)

    show()
```

Gráficos de Auto Correlação Parcial:

```
def FuncPlotPartialAutocorr(dMatrix_ranks, iNLags, sCol, dMatrix_GAR_Coef1,
    dMatrix_GAR_Coef2, sOperator, sGoodOrBad):

    from matplotlib.pyplot import legend, title, ylabel, xlabel, xticks,
        yticks, show, rcParams, subplot, stem, axhline, ylim, savefig, figure
    from statsmodels.graphics.tsaplots import plot_pacf
    from numpy import arange, sqrt

    iParam = ''.join(map(str, [int(s) for s in sCol.split("_")[:-2] if
        s.isdigit()]))
    sName = sCol.split("_")[0] + iParam
    nnSignificance = (+1.96)/sqrt(len(dMatrix_ranks))

    figure(figsize=(10,5))
    plot_pacf(dMatrix_ranks.values, lags = iNLags, title = 'PACF de Pearson')
    title('PACF de Pearson', fontsize = 22)
    ylabel('Auto Correlacao Parcial', fontsize = 18)
    xlabel('Lags', fontsize = 18)
    xticks(fontsize = 16)
    yticks(fontsize = 16)
    xticks(arange(0,iNLags+1,1), arange(0,iNLags+1,1))
    show()

    figure(figsize=(18,6))

    subplot(1, 2, 1)
    stem(dMatrix_GAR_Coef1)
    title('Forward PACF de Gini', fontsize = 20)
    axhline(nnSignificance, linestyle='--',color='k')
    axhline(-nnSignificance, linestyle='--',color='k')
    xticks(arange(0,iNLags+1,1), arange(1,iNLags+1,1))
    ylabel('Auto Correlacao Parcial', fontsize = 18)
    xlabel('Lags', fontsize = 18)
```

```
    xticks(fontsize = 16)
    yticks(fontsize = 16)
    ylim(top=1)

    subplot(1, 2, 2)
    stem(dMatrix_GAR_Coef2)
    title('Backward PACF de Gini', fontsize = 20)
    axhline(nnSignificance, linestyle='--',color='k')
    axhline(-nnSignificance, linestyle='--',color='k')
    xticks(arange(0,iNLags+1,1), arange(1,iNLags+1,1))
    xlabel('Lags', fontsize = 18)
    xticks(fontsize = 16)
    yticks(fontsize = 16)
    ylim(top=1)
    show()
```

A.1.12 Gráficos de ACF e PACF

Por fim, foram gerados os gráficos de distribuição residual e Gráfico Q-Q, através do código abaixo.

```
def FuncPlotProbQQ(siResidual, siFitted, dData, sCol, sMethod, sOperator,
    sGoodOrBad):

    import matplotlib.pyplot as plt
    from seaborn import residplot
    import scipy.stats as stats

    iParam = ''.join(map(str, [int(s) for s in sCol.split("_")[:-2] if
        s.isdigit()]))
    sName2 = sCol.split('_')[0] + iParam

    if sMethod.split(' ')[1] == 'Semi-Parametrico':
        sName = 'SP'
    if sMethod.split(' ')[1] == 'Minimizacao':
        sName = 'Min'
    if sMethod.split(' ')[1] == 'Classico':
        sName = 'C'

    plt.figure(figsize=(20, 8))

    plt.subplot(1, 2, 1)
```

```
    residplot(dData, siFitted, color = 'k')
    plt.ylabel('y', fontsize = 20)
    plt.xlabel('x', fontsize = 20)
    plt.xticks(fontsize = 15)
    plt.yticks(fontsize = 15)

    plt.subplot(1, 2, 2)
    stats.probplot(siResidual, dist="norm", plot=plt)
    plt.ylabel('Quartis Amostrais', fontsize = 16)
    plt.xticks(fontsize = 15)
    plt.yticks(fontsize = 15)

    plt.tight_layout()
    plt.show()
```