

RICARDO TAVARES ANTUNES DE OLIVEIRA

**Modelagem e simulação computacional da
combinação de preditores de séries temporais
por meio de cópulas**

RECIFE-PE
FEVEREIRO - 2014



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

Modelagem e simulação computacional da combinação de preditores de séries temporais por meio de cópulas

Dissertação de mestrado apresentada ao
Curso de Pós-Graduação em Informática
Aplicada da Universidade Federal Rural de
Pernambuco, como requisito parcial para
obtenção do grau de Mestre em Informática
Aplicada.

Área de Concentração: Modelagem Computacional
Orientador: Tiago Alessandro Espínola Ferreira
Co-orientador: Paulo Renato Alves Firmino

RECIFE-PE
FEVEREIRO - 2014

INSERIR FICHA CATALOGRÁFICA

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

Modelagem e simulação computacional da combinação de preditores de séries
temporais por meio de cópulas

Ricardo Tavares Antunes de Oliveira

Dissertação julgada adequada para obtenção
do título de Mestre em Informática Aplicada,
defendida e aprovada por unanimidade em
26/02/2014 pela Banca Examinadora.

Orientador:

Prof. DSc. Tiago Alessandro Espínola Ferreira
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. DSc. Paulo Renato Alves Firmino
Universidade Federal do Cariri

Prof. DSc. Silvana Boca Negra
Universidade Federal Rural de Pernambuco

Prof. DSc. George Darmiton da Cunha Cavalcanti
Universidade Federal de Pernambuco

À
minha querida mãe, à minha
família, e a todas as pessoas
que acreditaram em mim.

Agradecimentos

Primeiramente e acima de tudo, agradeço a DEUS por estar sempre ao meu lado nos momentos bons e ruins, por me proteger, abençoar, por iluminar minhas trajetórias e principalmente pelas dificuldades. Agradeço o SENHOR por mais essa vitória em minha vida.

À minha querida mãe Maria Tavares, por estar ao meu lado me apoiando durante todo o percurso dos meus estudos, dando força e me encorajando, pois sem seu apoio nada disso seria possível ou mesmo faria sentido.

À toda a minha família, pela torcida em favor do meu sucesso, não só profissional mas também como vitória pessoal. Em especial pela minha amada vó Maria de Lourdes pelas suas orações e energia positiva, para que pudesse conquistar este humilde sonho.

Agradeço imensamente ao meu orientador Tiago Alessandro, ser extraordinário, de qualidades ímpares, por ser uma referência para mim tanto como pesquisador como ser humano, por confiar e acreditar em mim para desenvolver tal tema. Por todos os ensinamentos, pela paciência e amizade. Pois sei que sem sua ajuda não seria possível a conclusão deste trabalho.

Ao sempre disponível Prof. Paulo Firmino, ser fantástico e pesquisador exemplar, que me orientou desde o início deste trabalho. Não existem palavras que podem descrever meus sinceros agradecimentos por ter feito de mim a pessoa que me tornei. Tive o imenso privilégio de conhecer e trabalhar com essa pessoa formidável. Que fez toda a diferença para conclusão deste trabalho, meu eterno agradecimento.

Aos amigos, em especial aqueles que acompanharam de perto as dificuldades e realizações nesse tempo: Allisson Dantas, Anselmo Lacerda, Fillippo Régis e Thaíze Fernandes. Muito obrigado por existirem.

Finalmente, meus agradecimentos aos professores e colegas da Universidade Federal Rural de Pernambuco, pelo ambiente acolhedor e produtivo. À Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco por ter investido em minha formação e ao Programa em Informática Aplicada PPGIA, pelos recursos disponibilizados.

Resumo

Sob o prisma da Estatística-Computacional, esta dissertação trata do problema de combinação não linear de modelos de previsão de séries temporais. Neste contexto, suas principais características e propriedades básicas são apresentadas. Alguns dos principais métodos de previsão de séries temporais presentes na literatura são descritos. As propostas apresentadas nesta dissertação são mostradas por meio de três casos de estudo utilizando um formalismo matemático conhecido como cópulas pela sua capacidade de poder medir dependência e combinar modelos de predição sem perda de informação. A combinação dos modelos de previsão ocorre por meio do formalismo matemático de cópulas que apresenta resultados convincentes e motivadores através de três casos de estudo presentes nesta dissertação. No primeiro é apresentado um pequeno caso para combinar as previsões dos modelos de previsão via cópula de Gumbel-Hougaard. No segundo caso de estudo é proposto um estimador combinado construído através da função multivariada não-paramétrica de Cacoullos para um caso simples de combinação. No terceiro e último caso de estudo são apresentados os principais resultados desta dissertação, em que, é realizado um experimento que compara estimadores combinados construídos levando em consideração várias séries temporais e inúmeros modelos de previsão, tal que, são simuladas diversas situações. Nesse sentido, experimentos computacionais realizados demonstram que o estimador combinado construído via cópula obteve melhores resultados quando comparado com os modelos individuais e o método de combinação linear.

Abstract

This dissertation discusses the problem of combining models for time series forecasting. In this context, the main characteristics and basic properties of combined models are presented. Some of the main methods of time series forecasting present in the literature are described. Computational experiments compare diverse copulas-based combined models. First of all, an algorithm is presented to combine predictions of the predictive models via Gumbel-Hougaard copula. In the second case, it is proposed a combined estimator constructed via non parametric Cacoullos multivariate functions. In the third and final case of study, the main results of this dissertation are presented, in which an experiment that compares combined estimators constructed taking into account thousands of time series and numerous forecasting models were simulated. Thus, computational experiments show that the combined estimator constructed via copula obtained better results compared with the individual models and the linear combination method.

Lista de Figuras

4.1	Arquitetura da abordagem proposta. A abordagem clássica são os modelos de previsão comuns. A modelagem dos erros prepara os resíduos para servir de entrada no procedimento de agregação onde são combinados para gerar uma previsão combinada [39].	30
5.1	A série temporal u_t representada por $(-)$, os respectivos modelos individuais ARIMA $(X_{t,1}$ e $X_{t,2})$ são representados por $(*)$ e (Δ) , e o estimador combinado \hat{U}_t e SA por (\square) e (\bullet) . A linha tracejada indica o início do horizonte de previsão ($t > 200$).	44
5.2	A série temporal u_t representada por $(-)$, os respectivos modelos individuais ARIMA $(X_{t,1}$ e $X_{t,2})$ são representados por $(*)$ e (Δ) , e o estimador combinado \hat{U}_t e SA por (\square) e (\bullet) . Nesta Figura, é levado em consideração apenas o desempenho das previsões no período ($t > 200$).	45
5.3	A série temporal u_t representada por $(-)$, enquanto $(*)$ e (Δ) representam respectivamente os modelos ARIMA individuais $X_{t,1}$ e $X_{t,2}$, o estimador combinado \hat{U}_t e SA são ilustrado por (\square) e (\bullet) . Esta figura leva em consideração apenas o desempenho dos modelos no período onde ($t > n$).	49
5.4	Parâmetro de suavização (λ) pelo EQM do estimador combinado.	51
5.5	Ilustra o desempenho do estimador combinado \hat{U}_t^{**} representado por (\square) ajustado com o parâmetro λ que minimiza sua acurácia para estimar a série temporal u_t	53

5.6 Desempenho do estimador combinado \hat{U}_t^* representado por (□) ajustado com o parâmetro λ que maximiza o EQM de \hat{U}_t^* para estimar a série temporal u_t . 54

Lista de Tabelas

3.1	Dados fictícios das variáveis Y_1 e Y_2 para ilustração do funcionamento do <i>ranking</i> das cópulas Empíricas.	22
3.2	<i>Ranking</i> gerado a partir das variáveis Y_1 e Y_2	22
5.1	Erro quadrático médio e desvio padrão dos resíduos dos modelos ARIMA individuais $(x_{t,1}, x_{t,2})$, estimador combinado $(\hat{U}_t$ e SA.	43
5.2	Desempenho dos modelos individuais, SA e estimador combinado construído via cópula de Cacoullos para o conjunto de exemplos ($n = 500$) e o conjunto de previsão ($m = 30$).	50
5.3	Avaliação de desempenho baseada no erro quadrático médio do estimador combinado construído via cópula de Cacoullos, onde \hat{U}_t^* e \hat{U}_t^{**} são os estimadores combinados ajustados com o valor de λ que regula o desempenho do estimador.	52
5.4	Erro quadrático médio dos modelos individuais e dos estimadores combinados via cópula de Cacoullos. Onde \hat{U}_t^* e \hat{U}_t^{**} foram ajustados com parâmetro λ adequado e inadequado.	54
5.5	Combinação de situações possíveis de acontecer.	61
5.6	Sumário do teste comparativo entre os estimadores Correto e Clássico.	62
5.7	Sumário do teste comparativo entre o estimador combinado Correto e o método SA.	63

5.8	Sumário do teste comparativo entre o estimador combinado Clássico e o método SA.	64
5.9	Sumário do teste comparativo entre o estimador combinado Cacoullos e Correto.	65
5.10	Sumário do teste comparativo entre o estimador combinado Cacoullos e o método SA.	67
5.11	Sumário do teste comparativo entre o estimador combinado Cacoullos e Clássico.	68

Lista de símbolos, abreviaturas e siglas

Símbolos	Descrição
ARIMA(\cdot)	Modelo Auto-Regressivo Integrado de Média Móveis
AR(\cdot)	Modelo Auto-Regressivo
a_t	Ruído branco
α_i	Vetor de parâmetros das distribuições marginais dos resíduos do modelo i
$\alpha_{i,j}$	Representa o j -ésimo parâmetro da i -ésima distribuição marginal
β	Parâmetro de dependência da cópula Normal
B	Operador de translação para o passado
$C(\cdot)$	Função cópula
$c_{\mathbf{E}_t}(\cdot)$	Função de densidade de cópula dos erros
d	Ordem de diferenciação
Δ^d	Operador de diferenças de ordem d
$E(\cdot)$	Função média
$ef(\cdot)$	Eficiência relativa
\mathbf{E}_t	Representa o vetor de erros aleatórios dos modelos de previsão no instante t
$E_{t,i}$	Representa o erro da t -ésima observação do i -ésimo modelo de previsão
\mathbf{e}_t	Vetor de erros observados diante de u_t
$e_{t,i}$	Respectivo erro cometido pelo i -ésimo modelo no instante t
EQM	Erro quadrático médio de um dado modelo

Continua na próxima página.

Tabela 1 – Continuação a partir da página anterior.

Símbolos	Descrição
EQM_R	Erro quadrático médio relativo de um dado modelo
$EQM_R(\cdot)$	Função do erro quadrático médio relativo de um dado modelo
$\overline{EQM_R}$	Média do EQM_R
\emptyset	São os parâmetros do modelo MA
$F_{E_{t,i}}(\cdot)$	Distribuição acumulada marginal de $E_{t,i}$
$F(\cdot)$	Função de distribuição acumulada
$F_Y(\cdot)$	Função de distribuição acumulada da variável aleatória Y
$f(\cdot)$	Função de Cacoullos
\mathbf{fl}	Saída da função de densidade de cópula
γ	Vetor de parâmetros da cópula
k	Quantidade de modelos de previsão (ou variáveis aleatórias)
k_f	Dimensionalidade do espaço de medição
L	Saída da função de máxima verosimilhança
l	Vetor de tamanho n_l com valores igualmente espaçados
$\lambda_{exemplo}$	Parâmetro λ utilizado no conjunto de exemplos
λ_{teste}	Parâmetro λ utilizado no conjunto de teste
m	Quantidade de observações da série temporal que almeja-se prever
μ	Representa a média
$MA(\cdot)$	Modelo Média Móveis
n	Quantidade de observações da série temporal usada para ajuste do estimador combinado (ou variável aleatória)
n_f	Número de padrões de treinamento
n_l	Quantidade de posições do vetor l
n_{caso}	Quantidade de casos utilizados nas simulações
ω_i	Peso dado para o i -ésimo modelo individual
$P(\cdot)$	Função de probabilidade
$p_{Y_i}(\cdot)$	Função de densidade de probabilidade marginal de Y_i
$p_{\mathbf{Y}}(\cdot)$	Função de densidade de probabilidade conjunta de \mathbf{Y}
$p_{E_t}(\cdot)$	Função de densidade de probabilidade marginal dos erros de E_t

Continua na próxima página.

Tabela 1 – Continuação a partir da página anterior.

Símbolos	Descrição
$\varphi(\cdot)$	Função geradora da cópula
φ^{-1}	Função inversa
$\Phi_k(\cdot)$	Função de distribuição acumulada multivariada normal
$\psi(B)$	Operador auto-regressivo estacionário
ϕ	São os coeficientes auto-regressivos
p	Ordem do modelo AR
q	Ordem do modelo MA
$R_i(\cdot)$	Função que retorna a ordem de $y_{t,i}$ no vetor $(y_{1,i}, y_{2,i}, \dots, y_{5,i})$
σ	Desvio padrão
σ_i	Desvio padrão do i -ésimo modelo
σ_{ij}	Covariância entre os modelos X_i e X_j
s	Índice do maior elemento contido no vetor \mathbf{fl}
t	Índice da observação da série temporal (ou variável aleatória)
θ	Parâmetro de dependência da cópula Gumbel-Hougaard
\mathbf{T}	Vetor de padrões de treinamento
T_i	É o i -ésimo padrão de treinamento em \mathbf{T}
$\Theta(B)$	Operador de médias móveis estacionárias
u_t	Observação da série temporal no instante t
U_t	Série temporal
\hat{u}_t	Valor estimado da série temporal no instante t
\hat{U}_t^a e \hat{U}_t^b	Modelos de previsão a e b para U_t
\tilde{U}_t	Ponto do instante t da série temporal com média nula
$Var(\cdot)$	Função variância
\mathbf{V}	Vetor das distribuições marginais acumuladas
v_i	Valor da distribuição acumulada marginal para Y_i (ou X_i)
\mathbf{X}_t	Representa o vetor de estimadores para u_t
\mathbf{x}_t	Vetor de previsões dos k modelos para u_t
\mathbf{Y}	Vetor de variáveis aleatórias
Y_i	É a i -ésima variável aleatória de \mathbf{Y}

Continua na próxima página.

Tabela 1 – Continuação a partir da página anterior.

Símbolos	Descrição
y_i	Instância da i -ésima variável Y_i
$y_{t,i}$	É a t -ésima observação da variável Y_i
z	Quantidade de parâmetros da marginal
$1(\cdot)$	Função indicadora

Sumário

1	Introdução	1
1.1	Motivação	3
1.2	Objetivos	4
1.3	Contribuições	4
1.4	Organização do trabalho	5
2	Modelos para previsão de séries temporais	7
2.1	Combinação de Modelos de previsão de séries temporais	8
2.2	Combinações lineares	9
2.3	Metodologia Box & Jenkins	10
2.3.1	Modelos Auto-Regressivos de Ordem p	10
2.3.2	Modelo de Médias Móveis de Ordem q	11
2.3.3	Modelos mistos: auto-regressivos com médias móveis	12
2.3.4	Modelo auto-regressivo integrado e de médias móveis	12
2.3.5	Algumas aplicações modelos Box & Jenkins	13
2.4	Redes Neurais Artificiais	14

2.4.1	Redes neurais artificiais no problema de previsão de séries temporais	14
2.5	Medidas de desempenho	15
2.6	Testes de hipóteses	16
2.6.1	Teste de Kolmogorov-Smirnov	16
2.6.2	Teste t -Student	16
2.6.3	Teste de Wilcoxon	17
3	Cópuas	18
3.1	Fundamentos de cópuas	18
3.2	Conceitos Básicos	19
3.3	Famílias de Cópuas	20
3.3.1	Cópuas Elípticas e Arquimedianas	20
3.4	Cópuas Empíricas	21
3.4.1	Cópuas de Cacoullos	22
3.5	Cópuas Gumbel	24
3.6	Cópuas Normais	25
3.7	Estimação de parâmetros	26
3.7.1	Método de máxima verossimilhança	26
3.7.2	Função de inferência de marginais	27
3.7.3	Métodos semi e não-paramétricos	28
4	Modelo proposto	29
4.1	Arquitetura da abordagem proposta	29

4.2	Conceitos básicos e definições	31
4.3	Cópulas	32
4.4	Inferência dos parâmetros de cópulas	33
4.5	Combinação de previsões	35
4.6	Modelo Clássico	35
4.7	Algoritmo de simulação	36
4.8	Exemplo de combinação via modelo proposto	38
5	Experimentos e resultados	41
5.0.1	Combinação via cópula de Gumbel-Hougaard	41
5.0.2	Combinação via cópula de Cacoullos	45
5.1	Uma análise comparativa entre estimadores combinados	55
5.1.1	Caso de estudo	55
5.1.2	Procedimento de simulação	57
5.1.3	Aplicação do teste de hipóteses	58
5.1.4	Resultados	59
5.2	Considerações finais	69
6	Conclusões	70
6.1	Contribuições	71
6.2	Limitações	72
6.3	Trabalhos futuros	73
6.4	Trabalhos publicados	74

A Cálculos para obter a equação de estimação dos pesos da combinação linear 81

I Código-fonte das simulação via cópula de Gumbel-Hougaard 84

Capítulo 1

Introdução

As diversas experiências ocorridas no decorrer da vida de pessoas e organizações mostram um inegável pretexto psicológico para buscar suporte de especialistas o tempo todo, principalmente antes de fazer uma decisão importante. Isso ocorre de forma natural para a maioria das pessoas, em particular se a decisão envolve dinheiro ou saúde. É normal perguntar a opinião de usuários antes de comprar um produto ou referência antes de contratar algum serviço, assim como pedir a opinião de diferentes médicos antes de passar por uma cirurgia séria. Esses são apenas alguns dos muitos exemplos que mostram a importância da opinião de vários especialistas.

Modelos especialistas podem ser encarados como humanos ou computacionais. Especialistas humanos podem ser vistos como médicos, advogados entre diversos profissionais que são tidos como peritos em determinado assunto ou área. Enquanto especialistas computacionais são *software* programados para desenvolver uma determinada atividade que normalmente é realizada por profissionais.

Esta dissertação aborda a combinação de modelos especialistas, em particular, com o objetivo de melhorar a confiança na decisão a ser tomada em relação ao uso de um único modelo. Por meio de algum processo de combinação, almeja-se chegar a uma decisão mais acurada do que aquela apresentada por um único modelo especialista.

No contexto aqui apresentado, modelos especialistas são encarados como preditores, onde

os mesmos serão empregados para resolver o problema de previsão de séries temporais. Nos estudos de incerteza de modelos para análise de previsão de séries temporais a literatura mostra ainda que a acurácia pode ser substancialmente melhorada pela combinação de modelos de previsão individuais. Autores como Jeong e Kim [28], Wallis [52], Dell'Aquila e Ronchetti [15], Amendola e Storti [1] e Lux [34] suportam essa afirmação. A literatura ainda apresenta diversos autores como R. T. A. Oliveira [39], T. F. Oliveira [41], Clemen [9] e Firmino *et al.*[22] que mostram meios de combinação de modelos para obterem resultados melhores quando comparados com os modelos individuais.

Podem ser vistos em R. T. A. Oliveira [39] e T. F. Oliveira [40] a combinação de modelos de previsão de séries temporais, onde o processo de combinação é realizado por meio do formalismo de cópula. Os resultados obtidos nos trabalhos, mostraram que o modelo combinado obteve melhor acurácia em relação aos modelos individuais.

Do exposto, busca-se mais especificamente com esta dissertação propor avanços no estado da arte de estudos sobre incerteza de modelos voltados a séries temporais, com interesse particular no uso do formalismo de cópulas aplicada a combinação não linear de preditores. Para tanto, estudar-se-ão modelos de previsão de séries temporais e estimadores de máxima verossimilhança associados à agregação de modelos ajustados a uma mesma série.

A previsão de determinados fenômenos auxilia na tomada de decisão no intuito de conhecer e analisar fenômenos futuros. Dentre seus objetivos destacam-se:

- Minimizar estragos causados por fenômenos da natureza, tais como: enchentes, terremotos, tornados, tsunames, por exemplo;
- Minimizar a quantidade de acidentes causados em: rodovias, espaço aéreo, em mar aberto, entre outros;
- Maximizar o retorno de investimento, como por exemplo em aplicações em ações da bolsa de valores.

Nesse sentido, são inúmeras as aplicações possíveis por meio de previsões de séries temporais, para auxiliar as diversas áreas do conhecimento a optar pela melhor decisão diante de determinadas situações.

São muitas as técnicas que buscam resolver o problema de previsão de séries temporais, porém, várias técnicas possuem algum erro de previsão associado, muitas vezes tal erro é suposto ser normalmente distribuído, conhecido na literatura como ruído branco. Desta forma, o interesse desta área é construir modelos de previsão que minimizem o erro de previsão, em outras palavras, construir um modelo que melhor se ajuste aos dados da série. A construção do modelo ótimo não é trivial e muitos preditores podem conduzir a previsões pouco acuradas e eficientes [13].

1.1 Motivação

De fato, estudos sobre incerteza de modelos têm estado na vanguarda da modelagem de séries temporais. Neste contexto, pesquisadores têm sido desafiados a combinar diversos modelos ajustados a uma mesma série com intuito de promover previsões agregadas. Estatisticamente estes estimadores agregados mostram-se superiores aos modelos individuais, tanto em termos de acurácia quanto eficiência. Uma das alternativas de agregação mais adotadas são as combinações lineares onde a estimativa resultante é dada por uma média ponderada das predições provenientes dos modelos individuais envolvidos. Revisões da literatura da área, tais como Clemen [9], Jeong [28] e Wallis [52], e pesquisas correlatas, como Dell’Aquila e Ronchetti [15], Amendola e Storti [1] e Lux [34], dão suporte a este argumento.

Neste trabalho, será apresentada uma metodologia alternativa para realizar previsões de séries temporais, baseada na combinação não linear de preditores. O procedimento utilizado é conhecido na literatura como cópulas [36]. Uma cópula pode ser definida como uma função que associa distribuições univariadas com o propósito de formar uma distribuição multivariada [2].

Dentre as principais vantagens na adoção de cópula destaca-se sua flexibilidade em medir dependências entre distribuições marginais pré-estabelecidas, em outras palavras, cópulas é capaz de medir a dependência que um modelo de previsão exercê sobre o outro. Na agregação via cópula a influência de um modelo é usualmente função de sua eficiência e dependência com os demais modelos. A eficiência do modelo está associada com a variância dos modelos, enquanto que sua dependência com outro modelo reflete o grau de similaridade de ambos.

Através desta abordagem, espera-se obter um estimador combinado mais acurado e eficiente em relação aos preditores individuais utilizados na combinação e às combinações lineares frequentemente adotadas.

1.2 Objetivos

Esta dissertação se dedicará a estudar cópulas para construção de estimadores voltados à agregação de preditores de séries temporais, com a finalidade de produzir previsões. Para tanto, a presente dissertação busca mais especificamente:

- i Avaliar o potencial de cópula aplicada à combinação de preditores ajustados a uma mesma série temporal;
- ii Implementar via linguagem de programação, a metodologia fundamentada em (i);
- iii Analisar comparativamente estimadores combinados construídos pelo item (ii).

Assim, de posse da distribuição de probabilidades inerente de cada modelo, propõe-se adotar cópulas para a elaboração da distribuição multivariada dos modelos, da qual será construído o estimador combinado para combinar preditores.

Deseja-se investigar até que ponto a metodologia proposta de fato mostra robustez. Essa pesquisa deseja analisar se o modelo combinado mostra-se mais acurado e eficiente em relação aos preditores individuais e outras alternativas de agregação. Almeja-se avaliar comparativamente o desempenho entre estimadores combinados, para isso, serão realizadas várias combinações considerando inúmeras séries temporais e modelos individuais de previsão.

1.3 Contribuições

Nessa dissertação deseja-se apresentar resultados que possam responder questões, como:

- i A combinação de preditores através do formalismo de cópulas é capaz de apresentar previsões estatisticamente eficientes e acuradas?

- ii O preditor combinado mostra-se mais eficiente e acurado em relação aos modelos individuais usados no procedimento de combinação?
- iii Construir um preditor combinado por meio da função multivariada de Cacoullos [6] é capaz de produzir previsões eficientes e acuradas? E ainda, as previsões obtidas são melhores em comparação aos preditores individuais?
- iv O ajuste do parâmetro de suavização da função multivariada de Cacoullos [6] conduz a melhores previsões?
- v Existe vantagem em estudar os erros cometidos pelos modelos?

1.4 Organização do trabalho

Esta dissertação apresenta mais quatro Capítulos. No Capítulo 2 será feita uma breve revisão bibliográfica sobre os conceitos básicos que envolvem o problema de previsão de séries temporais. Durante a revisão é dado destaque os modelos Box & Jenkins e as Redes Neurais Artificiais, por suas importâncias no contexto do problema de previsão, um vez que são encontrados inúmeros trabalhos na literatura que relatam o uso destes aplicados ao problema em estudo.

No Capítulo 3, são apresentados os principais conceitos envolvidos no âmbito de cópulas. Nesse sentido, são apresentadas ao longo deste trabalho as famílias de cópulas paramétricas Elípticas e Arquimedianas, a partir destas famílias serão abordadas respectivamente as cópulas Gumbel-Hougaard e Normal, além disso, a cópula Empírica também é levada em consideração nesta dissertação com ênfase especial à cópula de Cacoullos.

Posteriormente, no Capítulo 4 é apresentado o modelo proposto para combinação de modelos de previsão de séries temporais por meio de cópulas. Neste capítulo, são introduzidos os conceitos e definições adotados para agregação dos modelos de previsão, bem como, métodos para inferência de parâmetros de cópula. E ainda, a arquitetura utilizada para implementação do algoritmo de simulação da abordagem proposta é introduzido.

Capítulo 5 são mostrados alguns casos de estudo que mostram a utilidade do formalismo

de cópulas aplicada ao problema de previsão de séries temporais. Para tanto, foram implementados algoritmos de simulação no *software* R-Project [51], para diferentes casos de estudo. No primeiro caso apresentado nesta dissertação foi usada uma cópula Gumbel-Hougaard para construir um estimador combinado que acopla diferentes preditores individuais para prever uma série temporal. Em seguida, no segundo caso de estudo são apresentados os resultados obtidos com a combinação de preditores por meio da cópula de Cacoullos proposta nesta dissertação. A cópula proposta foi baseada na função não paramétrica multivariada de Cacoullos [6]. E ainda são apresentados resultados alcançados quando o parâmetro de suavização da função de Cacoulo é ajustada. Finalmente, no terceiro e último caso de estudo são apresentados os principais resultados desta dissertação, em que, é realizado um experimento que compara estimadores combinados construídos levando em consideração inúmeras situações distintas de combinação.

O Capítulo 6 apresenta as conclusões e propostas de trabalhos futuros.

Capítulo 2

Modelos para previsão de séries temporais

Uma sequência de observações ordenada cronologicamente a partir de um determinado índice é chamada de série temporal. Em geral, tal índice é o tempo, e a série temporal pode ser definida tanto no espaço contínuo como no espaço discreto. Na prática, uma série temporal é montada a partir de observações medidas de um fenômeno, como: quantidade de chuva mensal de uma região; número de acidentes rodoviários semanais; quantidade diária de crimes em uma cidade; valor de uma ação a cada minuto em um pregão da bolsa de valores, entre outros [5].

Previsão de séries temporais tem como principal objetivo prever valores futuros ou passados de determinado evento, como por exemplo prever: o consumo de energia de uma residência no próximo mês; quantidade de vendas para o próximo ano; quantidade de produtos que serão necessários para repor um estoque, entre outros. A tomada de decisão baseada em previsões abrange diversas áreas, tais como: financeira, meio ambiente, produção, entre inúmeras outras [13].

Seja $(u_1, \dots, u_t, \dots, u_n)$ instâncias da série temporal $(U_1, \dots, U_t, \dots, U_n)$, onde o número de observações no tempo é dado por n , tal que, $n \geq 1$. O interesse no problema de previsão de séries temporais é prever o próximo valor de $(u_t, t > n)$ com base nas observações passadas, isto é, (u_1, u_2, \dots, u_n) [57, 5].

É importante antes de realizar qualquer análise verificar se a série temporal é estacionária. Uma série temporal é dita estacionária quando as observações amostradas no tempo flutuam sobre um nível constante (média do processo é constante) com um desvio padrão constante. Portanto, $E(U_t) = E(U_{t+1}) = E(U_{t+n})$ e $Var(U_t) = Var(U_{t+1}) = Var(U_{t+n})$, onde $E(\cdot)$ e $Var(\cdot)$ são, respectivamente as funções média e variância [5]. Na presente dissertação, serão geradas, via simulação, séries temporais estacionárias.

2.1 Combinação de Modelos de previsão de séries temporais

Como destacado por Chatfield [8], boa parte da literatura voltada a séries temporais implicitamente assume que existe um único modelo intrínseco à série. Contudo, o autor enfatiza que mesmo considerando a existência de um único modelo, este será raramente conhecido *a priori* e não haverá qualquer garantia que ele seja selecionado como o melhor a se ajustar à série observada. Autores como Neuman [37] comentam inclusive que a adoção de um único modelo pode conduzir a vieses estatísticos e a subestimação da real incerteza subjacente à série temporal. Com estes argumentos em mente a incerteza de modelos parece assumir um papel indispensável para a análise de séries temporais. Neste contexto, pesquisadores têm sido desafiados a combinar diversos modelos ajustados a uma mesma série com o intuito de promover previsões agregadas.

Muitas combinações de modelos são realizadas após a seleção dos melhores modelos para uma dada série temporal. De forma geral, a seleção de modelos pode conduzir a previsões mais acuradas. Autores como Wichard e Ogorzalek [54] abordam isso em seu trabalho sobre a seleção de modelos. Apesar da seleção de modelos não fazer parte do escopo dessa dissertação, a seleção de preditores para combinação mostra-se como uma alternativa valiosa para aprimorar a acurácia do modelo combinado.

2.2 Combinações lineares

Uma das alternativas de agregação mais adotadas são as combinações lineares onde a estimativa resultante é dada por uma média ponderada das previsões provenientes dos modelos individuais envolvidos. Desta maneira, sejam $X_{t,i}$ o i -ésimo modelo individual de previsão da série temporal u_t no instante t , e $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,i}, \dots, x_{t,k})$ as respectivas previsões. A combinação linear pode ser escrita da seguinte forma:

$$\hat{u}_t = \frac{1}{k} \sum_{i=1}^k \omega_i x_{t,i} \quad (2.1)$$

onde, k é a quantidade de modelos, ω_i é o peso dado para o i -ésimo modelo, tal que, $\sum_{i=1}^k \omega_i = 1$ [54].

Nas combinações lineares o peso dado a um modelo é função de sua eficiência e dependência em relação aos demais modelos. Uma medida de dependência frequentemente usada é a correlação entre os erros dos modelos [22]. A importância de um dado modelo está diretamente relacionada com a variância dos seus erros de previsão.

Uma das principais alternativas para a estimação dos pesos nas combinações lineares são os ditos estimadores de mínima variância [12], que sob a suposição de normalidade dos erros e da dependência dos modelos coincidem com os estimadores de máxima verossimilhança. Sob a suposição de normalidade dos erros não apenas a computação do peso de cada modelo torna-se atividade trivial, mas também a medida da dependência entre os modelos. Por exemplo, diante da estimativa de dois modelos (X_1 e X_2) cujos erros são supostos seguirem uma distribuição normal, tem-se como estimativa agregada de mínima variância para função (\hat{u}_t) na qual os pesos podem ser obtidos pela seguinte equação [22]:

$$\omega_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \quad \omega_2 = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

onde, σ_i^2 é a variância do modelo X_i e σ_{ij} é a covariância entre X_i e X_j . Os cálculos para obter ω_1 e ω_2 são expostos no Apêndice A.

Neste trabalho, será também estudada a combinação linear conhecida como “*Simple Average*” (SA), para compará-la com os estimadores combinados que serão apresentados no Capítulo 4. A estratégia SA consiste na combinação linear onde os pesos dados para os modelos individuais são iguais a $\frac{1}{k}$, em outras palavras, o SA é uma média aritmética simples.

2.3 Metodologia Box & Jenkins

A metodologia de Box & Jenkins [5] é uma das obras da literatura mais importantes e conhecidas na área de previsão de séries temporais. Os estudos de Box & Jenkins resultaram nos modelos Auto-Regressivo (AR), Médias Móveis (MA) e Auto-Regressivo Integrado e de Médias Móveis (ARIMA).

A metodologia de Box e Jenkins [5] propõe uma família de modelos de previsão conhecidos como ARIMA, em que, a partir deste seleciona-se o modelo que desempenha a melhor previsão de um dada série temporal. O modelo ARIMA(p, d, q) é composto dos modelos AR(p) e MA(q) combinados com uma integração de ordem d . Neste sentido um modelo ARIMA apresenta-se como um preditor de uma determinada série temporal(U_t) que relaciona-se com seus valores passados, isto é, $(U_{t-1}, \dots, U_{t-p})$ em termos do ruído, diferenciados d vezes [18].

2.3.1 Modelos Auto-Regressivos de Ordem p

São baseados nas p observações passadas da série temporal $U_{t-1}, U_{t-2}, \dots, U_{t-p}$, onde p determina o número de passos entre as observações passadas e a próxima mais um termo aleatório, dado por:

$$\tilde{U}_t = \phi_1 \tilde{U}_{t-1} + \phi_2 \tilde{U}_{t-2} + \dots + \phi_p \tilde{U}_{t-p} + a_t \quad (2.2)$$

onde, ϕ_i ($i = 1, \dots, p$) são os coeficientes auto-regressivos e a_t é o ruído branco. Foi subtraído o nível médio (μ) da série:

$$\tilde{U}_t = U_t - \mu \quad (2.3)$$

Desta forma, um operador auto-regressivo estacionário de ordem p pode ser escrito como:

$$\psi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.4)$$

onde a restrição de estacionariedade é que $|B| \leq 1$, tal que, B é o operador de translação para o passado ($B^4 \tilde{U}_t = \tilde{U}_{t-4}$), podendo ser reescrita a Equação 2.2 como:

$$\psi(B)\tilde{U}_t = a_t \quad (2.5)$$

Para um caso em particular quando o modelo auto-regressivo de ordem 1 ($AR(1)$) então:

$$\tilde{U}_t = \phi_1 \tilde{U}_{t-1} + a_t \quad (2.6)$$

2.3.2 Modelo de Médias Móveis de Ordem q

Um modelo $MA(q)$ é caracterizado pela combinação dos ruídos de a_t , ocorridos no período atual com os ocorridos no passado, sendo descrito:

$$\tilde{U}_t = a_t - \emptyset_1 a_{t-1} - \emptyset_2 a_{t-2} - \dots - \emptyset_q a_{t-q} \quad (2.7)$$

onde \emptyset_i ($i = 1, \dots, q$) são os parâmetros da estrutura e q é a ordem do modelo. Fazendo uso do fato que $\tilde{U}_t = U_t - \mu$. Desta forma, o operador de médias móveis, é dado por:

$$\Theta(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \dots - \emptyset_q B^q \quad (2.8)$$

então

$$\tilde{U}_t = \Theta(B)a_t \quad (2.9)$$

A Equação 2.10 apresenta uma versão trivial para o modelo $MA(1)$:

$$\tilde{U}_t = a_t - \theta_1 a_{t-1} \quad (2.10)$$

2.3.3 Modelos mistos: auto-regressivos com médias móveis

Para diversos casos encontrados na prática, pode ser necessário recorrer a uma grande quantidade de parâmetros quando se trata de modelos unicamente AR ou MA. Neste caso, a inclusão de termos auto-regressivos e de média móveis em um único modelo pode ser uma possível solução para obter um modelo com poucos parâmetros. Assim o modelo $ARMA(p, q)$ necessita de menos termos e pode ser escrito da seguinte forma:

$$\tilde{U}_t = \phi_1 \tilde{U}_{t-1} + \dots + \phi_p \tilde{U}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.11)$$

O modelo $ARMA(1, 1)$ é o mais simples, sendo dado por:

$$\tilde{U}_t = \phi_1 \tilde{U}_{t-1} + a_t - \theta_1 a_{t-1} \quad (2.12)$$

2.3.4 Modelo auto-regressivo integrado e de médias móveis

Uma determinada série temporal pode não ser estacionária quanto ao nível, isto é, quando esta oscila ao redor de um nível durante algum tempo e depois salta para outro nível temporal. A série pode não ser estacionária quanto à inclinação, quando a variação dos dados não

permanece a mesma para qualquer tempo da série temporal, isto é, a flutuação aumenta ou diminui ao longo do tempo [53, 18].

Os modelos vistos anteriormente são necessários quando a série é estacionária. Para os casos envolvendo séries temporais não estacionárias, ou seja, que alteram sua média ou mesmo seu desvio padrão ao passar do tempo, pode-se utilizar o modelo ARIMA(p, d, q) para torná-las estacionárias [53, 18]. O modelo ARIMA pode ser expresso da seguinte forma:

$$\psi(B)\Delta^d U_t = \Theta(B)a_t \quad (2.13)$$

onde o modelo ARIMA(p, d, q), sendo p e q as ordens de $\psi(B)$ e $\Theta(B)$ respectivamente, e d é a ordem das diferenças. Em que, $\psi(B)$ representa o operador auto-regressivo de ordem p , $\Theta(B)$ representa o operador médias móveis de ordem q , a_t é o ruído branco, enquanto $\Delta = 1 - B$ representa o operador diferença. Este operador diferença pode ser definido como:

$$U_t - U_{t-1} = U_t - BU_t = (1 - B)U_t = \Delta U_t \quad (2.14)$$

então,

$$\Delta^d = (1 - B)^d \quad (2.15)$$

Enfatizando que na maioria dos casos tomar uma ou duas diferenças é suficiente para tornar a série temporal estacionária [5].

2.3.5 Algumas aplicações modelos Box & Jenkins

Os modelos de Box & Jenkins são apresentados nas várias obras da literatura de diversas áreas de previsão de séries temporais. Felipe [16] relata a utilização de vários experimentos com diversos modelos diferentes de previsão de séries temporais, aplicados à previsão da série do preço diário da soja no Norte do Paraná. Nos estudos realizados pelo autor, o modelo ARIMA(5, 0, 0) mostrou-se mais acurado em relação aos demais modelos.

Silva [45] estuda variáveis climáticas, mais especificamente, aborda previsões direcionadas para temperatura média mensal da cidade de Uberlândia. Entre os diversos modelos utili-

zados pelo autor o modelo SARIMA(3, 1, 0)(0, 1, 1) mostrou-se mais adequado para prever a série temporal de temperatura média mensal.

2.4 Redes Neurais Artificiais

A modelagem de Redes Neurais Artificiais (RNA) vem sendo estudada desde que se reconheceu que o cérebro humano processa informações de maneira distinta do computador digital convencional. O cérebro humano é um “computador” altamente complexo, não-linear e paralelo, capaz de organizar seus constituintes estruturais em neurônios, de tal forma, que seja possível realizar o processamento de atividades como reconhecer padrões, percepção e controle motor, dentre outras. O cérebro é mais rápido que qualquer computador digital existente [24].

As RNAs foram inspiradas no cérebro humano. Elas são aplicadas em diferentes áreas da ciência, para solucionar problemas, tais como: reconhecimento de padrões, classificação, previsão de séries temporais, entre outras [17].

2.4.1 Redes neurais artificiais no problema de previsão de séries temporais

Revisões da literatura apresentam diversas técnicas computacionais para o problema de previsão de séries temporais, entre as mais conhecidas estão as redes neurais artificiais como apresentado nos trabalhos de Ferreira *et al.* [19], Zhang [56], Hill e William[25] e Tang *et al.* [50], outros métodos computacionais utilizados são mostrados em Jain *et al.* [27] e Kim *et al.* [31]. Os métodos computacionais apresentados podem ser vistos como especialistas que auxiliam na tomada de decisão, quando esta diz respeito a prever o valor mais provável de um ponto futuro da série temporal.

2.5 Medidas de desempenho

Na literatura existem diversas medidas de desempenho aplicadas à previsão de séries temporais, como pode ser visto em muitos trabalhos [22, 39, 41, 13]. Nessa dissertação, será utilizado o Erro Quadrático Médio (EQM) para medir o desempenho dos modelos, no entanto, poderia ser utilizada outra medida. A seguir é apresentada a equação do EQM, onde, n é o número de observações, u_t é a série temporal no instante t e \hat{u}_t é a série temporal estimada pelo modelo de previsão, este podendo ser combinado ou individual. O EQM pode ser escrito da seguinte maneira:

$$\text{EQM}(\hat{u}_t, u_t) = \frac{1}{n} \sum_{t=1}^n (u_t - \hat{u}_t)^2 \quad (2.16)$$

Existe ainda a diferença relativa entre erros quadráticos médios (EQM_R), dada pela diferença relativa entre o EQM de dois modelos quaisquer chamados neste trabalho por \hat{U}_t^a e \hat{U}_t^b . Logo, o EQM_R pode ser escrito da seguinte forma:

$$\text{EQM}_R(\text{EQM}^{\hat{U}_t^a}, \text{EQM}^{\hat{U}_t^b}) = \frac{\text{EQM}^{\hat{U}_t^a} - \text{EQM}^{\hat{U}_t^b}}{\text{EQM}^{\hat{U}_t^b}} \quad (2.17)$$

Nesse sentido, o $\overline{\text{EQM}_R}$ é a média do EQM_R e quando a hipótese ($\overline{\text{EQM}_R} = 0$) é aceita, tem-se que ambos os modelos (\hat{U}_t^a e \hat{U}_t^b) são estatisticamente iguais. Por outro lado, quando a hipótese ($\overline{\text{EQM}_R} > 0$) é aceita tem-se que o modelo \hat{U}_t^b é melhor, caso contrario, \hat{U}_t^a é melhor.

Neste trabalho o termo eficiência ou precisão são utilizados para indicar o quão bom pode ser um modelo de previsão. De modo geral, dados dois estimadores não viesados \hat{U}_t^a e \hat{U}_t^b , sendo $\text{Var}(\hat{U}_t^a) < \text{Var}(\hat{U}_t^b)$, então \hat{U}_t^a é dito mais eficiente ou preciso do que \hat{U}_t^b , e a eficiência relativa de \hat{U}_t^a , em relação a \hat{U}_t^b é dada por [4]:

$$ef(\hat{U}_t^a, \hat{U}_t^b) = \frac{\text{Var}(\hat{U}_t^b)}{\text{Var}(\hat{U}_t^a)} \quad (2.18)$$

Enquanto a acurácia é tida como sendo o grau de proximidade de uma estimativa de \hat{U}_t^a em relação ao valor real, em outras palavras, acurácia é a distância entre o valor estimado

por \hat{U}_t^a e o valor real, desta forma, é dito que quanto menor for a distância, mais acurado é o estimador. Tal medida de proximidade pode ser obtida pelo EQM [35].

2.6 Testes de hipóteses

2.6.1 Teste de Kolmogorov-Smirnov

O teste estatístico de Kolmogorov-Smirnov (KS) permite avaliar se uma amostra retirada da população segue determinada distribuição de probabilidade específica, tal como: normal, uniforme e exponencial, por exemplo. Embora seja possível aplicar o teste qui-quadrado para este fim, geralmente é melhor aplicar o teste KS que é uma alternativa mais poderosa do que o teste qui-quadrado, nesta situação [4]. Seja $F(x)$ a função de distribuição acumulada, para a qual se quer verificar o ajuste dos dados. As hipóteses são [10]:

H_0 : a amostra provém da distribuição $F(x)$;

H_1 : a amostra não provém da distribuição $F(x)$.

2.6.2 Teste t -Student

A t -Student é uma das distribuições mais importantes para a inferência sobre médias populacionais. O teste t -Student é apropriado para comparar dois conjuntos de dados quantitativos, em termos de seus valores médios, quando estes seguem uma distribuição normal. Mais especificamente estuda-se neste trabalho as hipóteses [4]:

H_0 : $\mu_1 = \mu_2$;

H_1 : $\mu_1 \neq \mu_2$.

onde, μ_1 é o valor esperado da diferença relativa entre erros quadráticos médios e $\mu_2 = 0$.

2.6.3 Teste de Wilcoxon

O teste estatístico T de Wilcoxon substitui o t -Student quando a amostra não provém de uma distribuição normal. A ideia que norteia este teste é a de que se um tratamento A qualquer produzir valores maiores do que outro tratamento qualquer nomeado como B , as diferenças entre A e B ($A - B$) de sinal positivo serão maiores em número e grau do que as diferenças de sinal negativo. Se ambos os tratamentos têm o mesmo efeito, as diferenças positivas e negativas devem se anular [7]. As hipóteses para duas médias quaisquer (μ_1 e μ_2) são:

$$H_0: \mu_1 = \mu_2;$$

$$H_1: \mu_1 \neq \mu_2.$$

Capítulo 3

Cópuas

3.1 Fundamentos de cópuas

Uma variável aleatória é tida como uma variável que tem um único valor associado ao espaço amostral determinado aleatoriamente a cada resultado do experimento realizado. Em geral, o valor de uma variável aleatória só é conhecido após a realização de um experimento, como o número de caras obtidos após o lançamento de duas moedas, por exemplo [4].

A distribuição de probabilidade de uma variável aleatória é a chance (probabilidade) de cada valor da variável ocorrer [4]. Logo, a distribuição de probabilidade de cada variável individualmente é chamada de distribuição marginal.

Uma cópula pode ser definida como uma função que associa distribuições de probabilidade marginais para formar uma distribuição de probabilidade marginal multivariada, com a finalidade de juntar todas as distribuições marginais sem perda de informação. Desta maneira, assume-se que a distribuição multivariada possui as informações de todas as marginais [2]. O conceito de cópuas surgiu quando Sklar [47] *apud* Nelsen [36] percebeu que a função de distribuição acumulada multivariada $F_{\mathbf{Y}}(y_1, \dots, y_i, \dots, y_k)$ para qualquer conjunto de variáveis aleatórias \mathbf{Y} pode ser escrita da seguinte forma:

$$F_{\mathbf{Y}}(y_1, \dots, y_i, \dots, y_k) = C(v_1, \dots, v_i, \dots, v_k) \quad (3.1)$$

onde v_i ($i = 1, 2, \dots, k$) dada pela Equação 3.3, é a distribuição da i -ésima variável em $\mathbf{Y} = (Y_1, \dots, Y_k)$. A Equação 3.1 acima trata-se do teorema de Sklar, que fundamenta cópulas e, de forma geral, afirma que:

Teorema 3.1. *Seja $\mathbf{Y} = (Y_1, \dots, Y_i, \dots, Y_k)$ um vetor de variáveis aleatórias com função de distribuição acumulada multivariada $F_{\mathbf{Y}}$ e acumuladas marginais $F_{Y_i}(y_i)$, $i = 1, 2, \dots, k$. Então a cópula C é uma função acumulada k -dimensional, sendo escrita da seguinte forma:*

$$F_{\mathbf{Y}}(y_1, y_2, \dots, y_k) = C(v_1, v_2, \dots, v_i, \dots, v_k) = P(Y_1 \leq y_1, \dots, Y_i \leq y_i, \dots, Y_k \leq y_k) \quad (3.2)$$

desta forma, se $v_1, v_2, \dots, v_i, \dots, v_k$ forem contínuos então C é único. ■

3.2 Conceitos Básicos

Cópulas são funções que medem a dependência entre variáveis [36]. Mais especificamente na modelagem estatística, considerando um vetor envolvendo k variáveis aleatórias, $\mathbf{Y} = (Y_1, \dots, Y_i, \dots, Y_k)$, tal que ($i = 1, 2, \dots, k$). Em geral, uma cópula $C(\cdot)$ é uma função de distribuição acumulada conjunta que opera sobre as distribuições acumuladas marginais das variáveis em \mathbf{Y} , onde para uma dada instância de \mathbf{Y} , $\mathbf{y} = (y_1, \dots, y_i, \dots, y_k)$, então $C(\cdot)$ é dada pela distribuição acumulada conjunta de \mathbf{Y} onde:

$$v_i = F_{Y_i}(y_i) \quad (3.3)$$

As distribuições marginais de entrada da cópula $C(\cdot)$ estão no intervalo $[0,1]$, isto é, $v_i \in [0, 1]$, ($i = 1, 2, \dots, k$). Desta forma, v_i é o valor da distribuição marginal dada pela variável Y_i em uma determinada instância y_i . Assim, resumidamente, cópulas proporcionam um modelo multivariado para $(Y_1, \dots, Y_i, \dots, Y_k)$ pelo mapeamento de $(y_1, \dots, y_i, \dots, y_k)$ para $(v_1, \dots, v_i, \dots, v_k)$ via Equação 3.3.

Sobre a função de densidade de probabilidade conjunta de \mathbf{Y} , $p_{\mathbf{Y}}(y_1, \dots, y_i, \dots, y_k)$,

pode-se demonstrar que:

$$p_Y(y_1, \dots, y_i, \dots, y_k) = \frac{\partial^k F_Y(y_1, \dots, y_k)}{\partial y_1 \cdots \partial y_k} = c(F_{Y_1}(y_1), \dots, F_{Y_k}(y_k)) \prod_{i=1}^k p_{Y_i}(y_i) \quad (3.4)$$

onde, $p_{Y_i}(y_i)$ é a função de densidade de probabilidade marginal de Y_i e $c(\cdot)$ é a função de densidade de cópula.

3.3 Famílias de Cópulas

Existem diversos tipos de famílias de cópulas presentes na literatura [32]. Sendo que neste trabalho serão apresentadas em particular apenas as cópulas das famílias Elíptica e Arquimediana, por fazerem parte do objeto de estudo realizado no decorrer desta dissertação. As cópulas Elípticas foram adotadas pelo fato de que entre elas existe a cópula Normal e esta será apresentada neste trabalho em casos que deseja-se estudar a distribuição normal das marginais utilizando uma função multivariada normal e a família Arquimediana foi selecionada por apresentar cópulas simples de serem manuseadas como a Gumbel-Hougaard, uma vez que, esta possui apenas um parâmetro de dependência, enquanto outras cópulas possuem vários parâmetros, como a cópula Normal que possui um matriz de dependência.

As cópulas paramétricas possuem um ou vários parâmetros de dependência que medem a correlação entre as marginais, estes valores são intervalados e tal intervalo é dado por uma restrição matemática previamente estabelecida pela cópula.

3.3.1 Cópulas Elípticas e Arquimedianas

As cópulas Elípticas são distribuições multivariadas que possuem diversas propriedades similares à função de distribuição normal multivariada. De fato, essa última é apenas uma instância da primeira. Estas se mostram atrativas por ser possível mensurar estruturas de dependência não normais conforme colocado por Leal [32].

A família das cópulas Arquimedianas são capazes de captar a estrutura de dependência

na cauda inferior ou superior de uma distribuição, nos casos em que existem assimetria no comportamento das variáveis em estudo.

Dentre as diversas cópulas da família Arquimedianas, destacam-se as de Gumbel, Clayton e Frank. Por exemplo, a cópula Gumbel-Hougaard por definição possui apenas um parâmetro de ajuste, tornando o manuseio dessa cópula simples.

3.4 Cópula Empírica

As cópulas empíricas são baseadas no conjunto de n amostras (instâncias) disponíveis, independentes do vetor \mathbf{Y} . Nelas, $C(\cdot)$ é aproximada pela distribuição não paramétrica $C_n(\cdot)$ [23]:

$$C_n(v_1, \dots, v_k) = \frac{1}{n} \sum_{t=1}^n 1 \left(\frac{R_1(y_{t,1})}{n+1} \leq v_1, \dots, \frac{R_i(y_{t,i})}{n+1} \leq v_i, \dots, \frac{R_k(y_{t,k})}{n+1} \leq v_k \right) \quad (3.5)$$

onde, $1(\cdot)$ denota a função indicadora¹, $y_{t,i}$ representa a t -ésima observação da variável Y_i e $R_i(y_{t,i})$ a posição (ordem ou *ranking*) do valor $y_{t,i}$ quando as observações da variável Y_i são postas em ordem crescente. Um dos principais argumentos em favor da cópula apresentada na Equação 3.5 reside na sua simplicidade[23].

De modo a simbolizar o funcionamento da cópula empírica baseada em ordem (*ranking*), foram elaboradas duas tabelas com dados fictícios. A Tabela 3.1 apresenta instâncias de duas variáveis Y_1 e Y_2 , os vetores $\mathbf{y}_t = (y_{t,1}, y_{t,2}), (t = 1, \dots, 5)$. A Tabela 3.2 mostra os *rankings* associados aos elementos de \mathbf{y}_t . Desta forma, $R_i(y_{t,i})$ representará a ordem de $y_{t,i}$ quando as instâncias da variável Y_i , forem ordenadas crescentemente. Pode-se observar que o menor valor apresentado pelas instâncias de Y_1 é $y_{1,1} = 0,0765$, logo $R_1(y_{1,1}) = 1$, ou seja, a menor ordem possível (dado que R_1 representa os *rankings* das instâncias da variável Y_1 e R_2 de Y_2). O maior valor em Y_1 é $y_{2,1} = 0,876$, logo $R_1(y_{2,1}) = 5$ (o número que representa a maior ordem para Y_1). Os demais casos seguem o mesmo raciocínio.

¹A função indicadora de um conjunto é responsável por indicar se o elemento pertence ao conjunto. Esta podendo ser representada na matemática da seguinte forma $1(\cdot)$.

Tabela 3.1: Dados fictícios das variáveis Y_1 e Y_2 para ilustração do funcionamento do *ranking* das cópulas Empíricas.

t	1	2	3	4	5
$y_{t,1}$	0,0765	0,876	0,161	0,210	0,428
$y_{t,2}$	2,729	0,200	0,548	0,246	0,551

Tabela 3.2: *Ranking* gerado a partir das variáveis Y_1 e Y_2 .

t	1	2	3	4	5
$R_1(y_{t,1})$	1	5	2	3	4
$R_2(y_{t,2})$	5	1	3	2	4

Esta cópula é capaz apenas de modelar a tendência da dependência entre duas ou mais variáveis (se positiva ou negativa), negligenciando aspectos mais específicos da forma funcional de tal dependência. Este comportamento pode ser explicado por uma possível limitação desta cópula. Para maiores detalhes recomenda-se a leitura de R. T. A. Oliveira [38] em que são apresentadas propriedades mais detalhadas sobre esta família de cópulas, através de várias simulações. Além disso, os autores estudam a influência da estrutura das distribuições acumuladas das marginais sobre a cópula. Mais especificamente, são estudadas as cópulas empíricas onde é aproximada tanto por estimativas de estatísticas de ordem baseadas nos dados (Equação 3.5), quanto por distribuições paramétricas. Lima [33] busca estudar a cópula Empírica aplicada a três tipos diferentes de dependências entre duas variáveis, sendo levado em consideração a dependência linear, parabólica e independência.

3.4.1 Cópula de Cacoullos

A presente cópula é uma abordagem inovadora proposta nesta dissertação que originou-se quando Cacoullos [6] estendeu os resultados obtidos pela função proposta por Parzen [43] para os casos multivariados. O Teorema 4.1, em Cacoullos [6], apresenta como foi

possível estender os resultados de Parzen para casos particulares onde o estimador de kernel multivariado é produto de kernels univariados [49]. Em particular, para os casos onde o kernel é Gaussiano o estimador multivariado pode ser expresso por:

$$f(\mathbf{T}) = \frac{1}{(2\pi)^{k_f/2} \lambda^{k_f}} \frac{1}{n_f} \sum_{i=1}^{n_f} \exp \left[-\frac{(\mathbf{T} - \mathbf{T}_i)^t (\mathbf{T} - \mathbf{T}_i)}{2\lambda^2} \right] \quad (3.6)$$

onde,

$i \equiv$ índice do padrão

$n_f \equiv$ número total de padrões de treinamento

$T_i \equiv$ i -ésimo padrão de treinamento em T

$\lambda \equiv$ parâmetro de suavização

$k_f \equiv$ dimensionalidade do espaço de medição

Nota-se que a função $f(\mathbf{T})$ possui o parâmetro λ conhecido como parâmetro de suavização, com $\lambda > 0$. Specht, Donald F. [48] *apud* Specht [49] envolve discussões sobre como estimar o melhor valor do parâmetro de suavização λ para um problema de classificação, onde o autor constata que na prática não é difícil encontrar um bom valor para λ , e ainda que a taxa de erro das classificações não é alterada drasticamente com pequenas alterações no parâmetro λ .

A função $f(\mathbf{T})$ é simplesmente a soma das distribuições normais multivariadas centradas em cada amostra (ponto) de treinamento [49]. A cópula proposta tem o funcionamento similar ao algoritmo *k-Nearest Neighbor* (k-NN) [20]. O k-NN a partir de um conjunto de exemplos (padrões de treinamento), consegue tomar decisões com base nos exemplos de treinamento para inferir sobre novos conjuntos de entrada (padrões de teste). Na literatura, são apresentados métodos de previsão de séries temporais utilizando o k-NN, como pode ser visto em Ferreo [20] que propõem uma abordagem para seleção de vizinho mais próximos chamado como *k-Nearest Neighbor - Time Series Prediction*, este sendo uma derivação do k-NN utilizado como método de previsão.

3.5 Cópula Gumbel

Há diversos tipos de cópulas paramétricas a serem consideradas do lado direito da Equação 3.2. A título de ilustração, a cópula $C(\cdot)$ é dita da família de cópula Arquimediana se admite a representação:

$$C(v_1, \dots, v_k) = \varphi^{-1}[\varphi(v_1) + \dots + \varphi(v_k)] \quad (3.7)$$

onde, $\varphi(v_i)$ é conhecida como a função geradora da cópula com inversa $\varphi^{-1}(\cdot)$. A família das cópulas Arquimedianas abrange uma grande variedade de estruturas de dependência (tais como: dependência linear, exponencial, parabólica, entre outras). Em particular, as cópulas Arquimedianas podem assumir dependência caudal assimétrica, sendo uma propriedade a favor de sua aplicação à modelagem de dados com estrutura de dependência assimétrica. Quando a função geradora é dada por:

$$\varphi(v_i) = (-\ln v_i)^\theta, \forall \theta \geq 1 \quad (3.8)$$

$\varphi^{-1}(t) = \exp\left(-t^{\frac{1}{\theta}}\right)$ e a Equação 3.7 resulta na chamada cópula de Gumbel-Hougaard [23].

Neste caso, a distribuição acumulada conjunta da cópula de Gumbel-Hougaard é dada pela seguinte equação:

$$F_Y(y_1, \dots, y_k) = C(v_1, \dots, v_k) = \exp\left\{-\left[\sum_{i=1}^k (-\ln v_i)^\theta\right]^{\frac{1}{\theta}}\right\} \quad (3.9)$$

onde, v_i é dado pela Equação 3.3.

Na cópula de Gumbel-Hougaard, o grau de dependência entre as variáveis em \mathbf{Y} é medida unicamente a partir do parâmetro θ . Para $\theta = 1$ obtêm-se a cópula de marginais independentes e caso $\theta \rightarrow \infty$ obtêm-se a cópula dedicada à dependência perfeita positiva. Desta

forma, a cópula de Gumbel-Hougaard permite modelar desde situações de independência até de dependência perfeita positiva. Ela trata de uma cópula de valores extremos e apresenta dependência na cauda superior.

3.6 Cópula Normal

A cópula Normal ou cópula Gaussiana faz parte da família das Elípticas, que é completamente simétrica e possui dependência zero para qualquer lado dos extremos da distribuição [3]. Sua dependência é medida pela matriz de dependência, em que, cada um dos elementos da matriz são usados para guardar o grau de dependência entre cada combinação de variáveis [44]. Segundo Renard e Lang [44] a cópula Normal é mais conveniente quando o número de dimensões, isto é, a quantidade de variáveis aleatórias é maior que dois ($k > 2$). O parâmetro de dependência β da cópula interpola entre -1 e 1, onde, ($\beta = 0$) indica independência, ($\beta = -1$) dependência perfeita negativa e ($\beta = 1$) dependência perfeita positiva.

A cópula Normal pode ser escrita da seguinte forma:

$$C(v_1, \dots, v_k) = \Phi_k(\varphi^{-1}(v_1), \dots, \varphi^{-1}(v_k) | \beta), \quad (3.10)$$

onde, $\varphi^{-1}(v_i), i = 1, \dots, k$, é a função inversa da distribuição acumulada unitária e Φ_k é a função de distribuição acumulada multivariada dada por:

$$\Phi_k(\varphi^{-1}(v_1), \dots, \varphi^{-1}(v_k) | \beta) = \int_{-\infty}^{\varphi^{-1}(v_1)} \dots \int_{-\infty}^{\varphi^{-1}(v_k)} \Phi(v_1) \dots \Phi(v_k) dv_1 \dots dv_k \quad (3.11)$$

Para os casos em que $\varphi^{-1}(v_i)$ é a inversa da distribuição normal univariada, a cópula Normal gera uma distribuição Normal multivariada. Entretanto, basta que ao menos uma das distribuições marginais não seja normal para que a distribuição conjunta resultante seja diferente [11].

A vantagem da cópula Normal consiste em sua simplicidade, uma vez que os dados

tenham sido transformados pela função $\varphi^{-1}(v_i)$. Renard e Lang [44] relatam que particularmente, uma cópula Normal não é capaz de modelar apropriadamente as estruturas de dependência mais complexas (em R. T. A. Oliveira [38] é apresentado experimentos com alguns tipos de dependência, em que, constatou-se que a dependência parabólica parece ser mais difícil de modelar em relação a linear), enquanto Panchenko [42] menciona que devido esta cópula ser radialmente simétrica e ter dependência zero na cauda, a mesma, pode ser inadequada para caracterizar a dependência entre séries financeiras.

3.7 Estimação de parâmetros

Na literatura, existem diversos tipos de funções de estimação e inferência sobre a cópula mais adequada aos dados observados [3, 29]. Em muitos trabalhos são citados tanto métodos não-paramétricos como as cópulas empíricas, quanto paramétricos de máxima verossimilhança (MV) e de maximização a partir das distribuições marginais, também conhecida como (IFM), acrônimo para “*Inference function for margins*” [23, 29]. Há ainda as cópulas semi-paramétricas, onde as marginais não-paramétricas são consideradas para o ajuste da cópula paramétrica [3]. Todas estas alternativas citadas serão brevemente introduzidas a seguir.

3.7.1 Método de máxima verossimilhança

O método de máxima verossimilhança (MMV) é usualmente adotado para estimar tanto os parâmetros das distribuições marginais quanto da cópula paramétrica $C(\cdot)$. O MMV estima simultaneamente os parâmetros das distribuições marginais de Y e de $C(\cdot)$. Os estimadores obtidos são aqueles que maximizam a função de log-verossimilhança resultante da Equação 3.4 diante dos dados evidenciados em \mathbf{y}_t [55]:

$$\begin{aligned}
L &= \sum_{t=1}^n \ln(p_{Y_1, \dots, Y_k}(y_{t,1}, \dots, y_{t,i}, \dots, y_{t,k} | \alpha_1, \dots, \alpha_i, \dots, \alpha_k, \gamma)) \\
&= \sum_{t=1}^n \ln(c(F_{Y_1}(y_{t,1} | \alpha_1), \dots, F_{Y_k}(y_{t,k} | \alpha_k) | \gamma)) + \sum_{t=1}^n \sum_{i=1}^k \ln(p_{Y_i}(y_{t,i} | \alpha_i))
\end{aligned} \tag{3.12}$$

onde α_i é o vetor que contém o conjunto de parâmetros de $F_{Y_i}(y_i)$ e γ é o vetor de parâmetros de dependência da cópula, tal como θ (Gumbel-Hougaard) ou β (Normal), por exemplo. Por sua vez, o vetor $\mathbf{y}_t = (y_{t,1}, \dots, y_{t,i}, \dots, y_{t,k})$ representa a t -ésima instância do vetor de variáveis \mathbf{Y} . Assim, via MMV, os parâmetros γ e α_i são estimados simultaneamente, de forma a maximizar a função L . Esta estimação conjunta torna a obtenção de estimativas para os parâmetros bastante complexa [32].

3.7.2 Função de inferência de marginais

O método conhecido como função de inferência de marginais ou do inglês “*Inference Function for Margins*” (IFM) permite separar o processo de estimação em duas etapas. Primeiramente, são estimados os parâmetros das marginais (α_i) para, em seguida, serem estimados os parâmetros da cópula (γ). Neste sentido, é comum que se adote o MMV em ambas as etapas de estimação, separando a soma da Equação 3.12 em duas funções a serem otimizadas em sequência [55]. Desta forma, o problema de inferência das marginais é tratado primeiramente de maneira usual, e pode-se recorrer a testes de ajuste tais como o de Kolmogorov-Smirnov (ver Simard [46]). Apenas em um segundo momento é que o modelo de cópulas é, de fato, tratado.

Esta alternativa de inferência mostra-se interessante por diversos motivos. Em termos práticos, pode-se estar interessado em estudar a dependência entre variáveis, cujo comportamento marginal tenha sido previamente modelado, tal como ocorre em problemas de incerteza de modelos [21], por exemplo. Por outro lado, embora as estimativas obtidas via IFM não sejam necessariamente equivalentes as do MMV, as diferenças são desprezíveis na grande parte dos casos [32]. Por fim, destaca-se ainda o fato de que ao decompor o problema de inferência, o IFM torna-se computacionalmente atrativas, já que se reduz a sua complexi-

dade em relação ao MMV. Para maiores detalhes sobre o IFM e sua comparação com MMV, sugere-se literatura mais específica [32, 29].

3.7.3 Métodos semi e não-paramétricos

Nos métodos não-paramétricos, as cópulas são construídas somente a partir da distribuição de frequência dos dados e modelos matemáticos paramétricos são negligenciados. Desta abordagem emergem as cópulas empíricas descritas na Equação 3.5. Alguns autores têm se dedicado a esclarecer a importância desta abordagem, principalmente devido à sua simplicidade de aplicação e propriedades estatísticas [23].

Como já destacado anteriormente, as cópulas empíricas não fazem uso de modelos paramétricos nem sobre as distribuições marginais das variáveis em \mathbf{Y} nem sobre a própria cópula. Nelas, a distribuição acumulada da variável Y_i é aproximada pela distribuição de frequências acumuladas de acordo com os dados, $R_i(y_{t,i})$, por sua vez baseada na ordem das observações em $y_{t,i}$. As cópulas empíricas foram introduzidas primeiramente por Deheuvels [14] *apud* Nelsen [36], chamadas na ocasião de “funções de dependência empírica”. Estas cópulas tendem a ser computacionalmente eficientes e são estimadores centrados para a verdadeira cópula à medida que o tamanho da amostra n cresce [30, 26].

Sobre as cópulas semi-paramétricas, nelas as marginais são usualmente aproximadas pela distribuição de frequências acumuladas $R_i(y_{t,i})$. Contudo, a cópula $C(\cdot)$, diferentemente do método não-paramétrico, se dá por um modelo matemático [3].

Embora pouco mencionada na literatura, é ainda possível que a modelagem envolva distribuições marginais paramétricas ajustadas aos dados e uma cópula não-paramétrica. Para tanto, bastaria substituir, na Equação 3.5, $R_i(y_{t,i})/(n + 1)$ pela função acumulada marginal de $Y_{t,i}$, ajustada via MMV, por exemplo. No trabalho de R. T. A. Oliveira [38], esta modalidade de cópula é denominada de “semi-empírica”.

Capítulo 4

Modelo proposto

O presente Capítulo irá introduzir o conceito de combinação de modelos de previsão de séries temporais por meio de cópulas. O intuito de combinar preditores via cópula, se mostra interessante, quando se pensa em construir um estimador combinado, baseado no método de máxima verossimilhança. Resumidamente, será apresentada a arquitetura da abordagem proposta, em seguida, são introduzidos os conceitos básicos e definições da abordagem. E finalmente, é mostrado o algoritmo de simulação utilizado para proceder a agregação das previsões.

4.1 Arquitetura da abordagem proposta

O presente estimador combinado pode ser especificado em três partes (ver Figura 4.1): A abordagem Clássica, Modelagem dos erros e Procedimento de Agregação de modelos de previsão através de cópulas. O primeiro passo, isto é, a Abordagem Clássica, são obtidas as previsões realizadas por cada modelo de previsão de séries temporais, tal que, são extraídas k previsões do banco de dados (BD) com seus respectivos *lags* relevantes, ou seja, primeiramente são extraídas as previsões utilizadas para ajuste do estimador combinado, para em seguida, retirar as previsões para teste do modelo. A Abordagem Clássica não será objeto de estudo deste trabalho, visto que, o objetivo apresentado aqui é combinar preditores pré-definidos. Para tanto, cada modelo pode ser visto como uma entrada para abordagem

proposta. Assim, os modelos individuais podem ser considerados como caixas pretas, onde os parâmetros e método utilizado para sua construção são negligenciados.

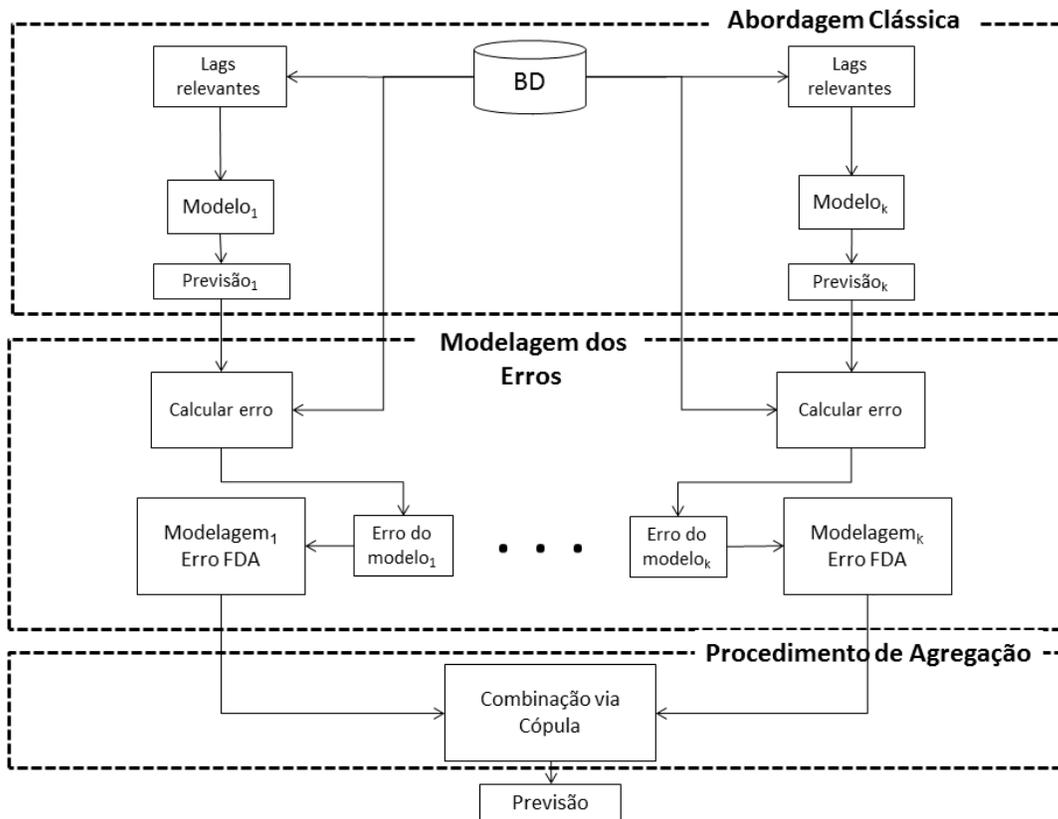


Figura 4.1: Arquitetura da abordagem proposta. A abordagem clássica são os modelos de previsão comuns. A modelagem dos erros prepara os resíduos para servir de entrada no procedimento de agregação onde são combinados para gerar uma previsão combinada [39].

No segundo passo é realizada a modelagem dos erros cometidos pelos modelos que serão utilizados como entrada para o passo de agregação. Logo, o segundo passo inicia calculando os erros cometidos por cada modelo através da Equação 4.1 ou 4.2, por exemplo. Em seguida, é calculada a função de distribuição acumulada marginal dos erros para cada modelo, como ilustrado na Figura 4.1.

Finalmente, o passo de agregação das previsões recebe como entrada as distribuições acumuladas marginais dos erros de cada modelo em estudo. Este procedimento provê uma combinação eventualmente não linear de previsões.

4.2 Conceitos básicos e definições

No contexto do presente trabalho, seja u_t a série temporal de interesse no tempo t e seja $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,i}, \dots, X_{t,k})$ o vetor de k preditores para u_t . Por sua vez, seja $\{u_t\}_{t=1}^n$ as observações da série temporal de tamanho n utilizadas para ajuste do estimador combinado e seja $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,i}, \dots, x_{t,k})$ as respectivas previsões, isto é estimativas de u_t através deste período de tempo. Finalmente, seja $\{u_t\}_{t=n+1}^{n+m}$ os valores dos pontos futuros da série temporal a serem preditos. Então, é possível inferir sobre as distribuições marginais de $X_{t,i}$ e a respectiva distribuição conjunta de \mathbf{X}_t a partir de $\{\mathbf{x}_t\}_{t=1}^n$ e $\{u_t\}_{t=1}^n$ via cópulas, a fim de construir um estimador combinado para u_t .

Desta forma, seja $\mathbf{E}_t = (E_{t,1}, \dots, E_{t,i}, \dots, E_{t,k})$ o vetor aleatório de erros de \mathbf{X}_t , que pode ser formulado pelo erro aditivo ou multiplicativo, por exemplo, sendo o erro aditivo dado pela seguinte equação:

$$E_{t,i} = X_{t,i} - u_t. \quad (4.1)$$

Enquanto que o erro multiplicativo é escrito da seguinte forma:

$$E_{t,i} = \frac{X_{t,i}}{u_t}. \quad (4.2)$$

Assim, $E_{t,i}$ é instanciado em função de $\{\mathbf{x}_t\}_{t=1}^n$ e $\{u_t\}_{t=1}^n$, isto através das equações:

$$e_{t,i} = x_{t,i} - u_t \quad (4.3)$$

ou

$$e_{t,i} = \frac{x_{t,i}}{u_t} \quad (4.4)$$

onde, $\{e_{t,i}\}_{t=1}^n$ é o conjunto de erros observados do i -ésimo preditor, em relação a série temporal u_t . Por sua vez, $\mathbf{e}_t = (e_{t,1}, \dots, e_{t,i}, \dots, e_{t,k})$ é o vetor de erros dos k preditores no instante t .

Neste contexto, a distribuição acumulada marginal dos erros de \mathbf{E}_t é dada por $F_{E_{t,i}}(e_{t,i})$ e a distribuição acumulada conjunta marginal dos erros é dada por $F_{E_{t,1}, \dots, E_{t,k}}(e_{t,1}, \dots, e_{t,k})$ e a distribuição acumulada marginal de $X_{t,i}$ para erros aditivos, por exemplo, é dado por:

$$F_{X_{t,i}}(x_{t,i}) = P(X_{t,i} < x_{t,i}) = P(E_{t,i} + u_t < x_{t,i}) = P(E_{t,i} < x_{t,i} - u_t) = F_{E_{t,i}}(e_{t,i}) \quad (4.5)$$

Assim a incerteza adjacente $X_{t,i}$ pode ser descrita em termos da incerteza intrínseca a $E_{t,i}$. Da mesma maneira, é fácil obter a função de densidade de probabilidade marginal de $X_{t,i}$ a partir de $E_{t,i}$. Para o erro aditivo, tem-se:

$$p_{X_{t,i}}(x_{t,i}) = p_{E_{t,i}}(e_{t,i}) \quad (4.6)$$

A Equação 4.6 será obtida da função de densidade de probabilidade de cópulas que será introduzida na Seção 4.3 deste Capítulo.

4.3 Cópulas

Como pode ser visto no Capítulo 3, uma cópula $\mathcal{C}(\cdot)$ é uma função de distribuição acumulada multivariada cujas distribuições marginais são uniformes no intervalo unitário entre $[0, 1]$. Assim, de acordo com o formalismo de cópulas [36] para qualquer vetor \mathbf{E}_t , existe uma distribuição acumulada conjunta,

$$F_{\mathbf{E}_t}(e_{t,1}, \dots, e_{t,k}) = \mathcal{C}(V_1 \leq v_1, \dots, V_k \leq v_k) \quad (4.7)$$

onde,

$$v_i = F_{E_{t,i}}(e_{t,i}) \quad (4.8)$$

Portanto, \mathbf{V} envolve distribuições marginais uniformes no intervalo $[0, 1]^k$, independentemente da distribuição marginal de $E_{t,i}$ [23]. Além disso, a partir da probabilidade e fun-

damentos de cópulas, tem-se que a função de densidade de probabilidade conjunta de \mathbf{E}_t é dada por:

$$p_{E_{t,1}, \dots, E_{t,k}}(e_{t,1}, \dots, e_{t,k}) = c_{E_{t,1}, \dots, E_{t,k}}(F_{E_{t,1}}(e_{t,1}), \dots, F_{E_{t,k}}(e_{t,k})) \prod_{i=1}^k p_{E_{t,i}}(e_{t,i}) \quad (4.9)$$

onde $p_{E_{t,i}}(e_{t,i})$ é a função de densidade de probabilidade marginal de $E_{t,i}$ e $c_{\mathbf{E}_t}(\cdot)$ representa a função de densidade de probabilidade de cópulas de \mathbf{E}_t .

No contexto deste trabalho, é suposto que $c_{\mathbf{E}_t}(\cdot)$ e $p_{E_{t,i}}(e_{t,i})$ são invariantes através do tempo, uma vez que é assumida que a série temporal é estacionária. Embora pareça restritivo a primeira vista, deve-se ressaltar que tal suposição está de acordo com a maioria das estratégias de modelagem de séries temporais. Na verdade, tem sido usualmente considerado que o componente $(E_{1,i}, \dots, E_{t,i}, \dots, E_{n,i})$ são independentes e identicamente distribuídos (*iid*). De fato, tal suposição tem sido uma condição necessária para a adequação de preditores. Por outro lado, deve-se destacar que a abordagem proposta é capaz de lidar com previsões tendenciosas (estimativas com algum erro associado), como poderá ser visto nos casos de estudo a seguir, uma vez que $X_{t,i}$ é mapeado para $E_{t,i}$.

No decorrer deste trabalho, C e c , introduzidos nesta Seção, assumirão diversos tipos de cópulas.

4.4 Inferência dos parâmetros de cópulas

Como dito anteriormente, o IFM é um estimador de máxima verossimilhança (MV) que se dividi em dois passos para estimar o valor dos parâmetros das marginais e da cópula, assim os parâmetros da cópula é o vetor γ e das marginais é o vetor de parâmetros $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,j}, \dots, \alpha_{i,z})$ onde z é quantidade de parâmetros.

A função de densidade de probabilidade da Equação 4.9, pode ser generalizada para

equação a seguir:

$$p_{\mathbf{E}_t}(\mathbf{e}_t | \alpha_1, \dots, \alpha_k, \gamma) = c_{\mathbf{E}_t}(F_{E_{t,1}}(e_{t,1} | \alpha_1), \dots, F_{E_{t,k}}(e_{t,k} | \alpha_k) | \gamma) \prod_{i=1}^k p_{E_{t,i}}(e_{t,i} | \alpha_i) \quad (4.10)$$

Desta maneira, a partir do conjunto de resíduos $(e_{1,i}, \dots, e_{n,i})$, o primeiro passo é inferir α_i via MV sobre $p_{E_{t,i}}(e_{t,i} | \alpha_i)$. Neste sentido, quando o erro seguir uma distribuição normal, o vetor de parâmetros $\hat{\alpha}_i$ é obtido por:

$$\begin{aligned} \mu_{e_{t,i}} &\leftarrow \frac{1}{n} \sum_{t=1}^n e_{t,i} \\ \sigma_{e_{t,i}} &\leftarrow \sqrt{\frac{1}{n-1} \sum_{t=1}^n (e_{t,i} - \mu_{e_{t,i}})^2} \\ \hat{\alpha}_i &\leftarrow \{\mu_{e_{t,i}}, \sigma_{e_{t,i}}\} \end{aligned} \quad (4.11)$$

onde, $\mu_{e_{t,i}}$ e $\sigma_{e_{t,i}}$ é a média e desvio padrão de $e_{t,i}$. Quando o erro seguir uma distribuição log-normal é obtido da seguinte forma:

$$\begin{aligned} \mu_{\log} &\leftarrow \frac{1}{n} \sum_{t=1}^n \log(e_{t,i}) \\ \sigma_{\log} &\leftarrow \sqrt{\frac{1}{n-1} \sum_{t=1}^n (\log(e_{t,i}) - \mu_{\log})^2} \\ \hat{\alpha}_i &\leftarrow \{\mu_{\log}, \sigma_{\log}\} \end{aligned} \quad (4.12)$$

onde, μ_{\log} e σ_{\log} é a média e desvio padrão do log de $e_{t,i}$.

Por sua vez, a partir das estimações de $(\alpha_{i,1}, \dots, \alpha_{i,j}, \dots, \alpha_{i,z})$, diz que $(\hat{\alpha}_{i,1}, \dots, \hat{\alpha}_{i,j}, \dots, \hat{\alpha}_{i,z})$ são as estimativas dos parâmetros das marginais. Posteriormente no segundo passo, é estimado o parâmetro γ , em que, $\hat{\gamma}$ é o valor que maximiza a função de cópulas, sendo obtido pela maximização de $c_{\mathbf{E}_t}(F_{E_{t,1}}(e_{t,1} | \hat{\alpha}_1), \dots, F_{E_{t,k}}(e_{t,k} | \hat{\alpha}_k) | \gamma)$. Maiores detalhes em relação ao método IFM pode ser visto em Joe and Xu [29].

4.5 Combinação de previsões

Uma vez que as estimativas de $(\hat{\alpha}_1, \dots, \hat{\alpha}_k, \hat{\gamma})$ são obtidas, pode-se reescrever a Equação 4.10 com uma explícita função de u_t tal como:

$$\begin{aligned} & p_{\mathbf{E}_t}(\mathbf{e}_t | \hat{\alpha}_1, \dots, \hat{\alpha}_k, \hat{\gamma}) \\ &= c_{\mathbf{E}_t} \left(F_{E_{t,1}}(e_{t,1} | \hat{\alpha}_1), \dots, F_{E_{t,k}}(e_{t,k} | \hat{\alpha}_k) | \hat{\gamma} \right) \prod_{i=1}^k p_{E_{t,i}}(e_{t,i} | \hat{\alpha}_i). \end{aligned} \quad (4.13)$$

Assim, a Equação 4.13 é utilizada enquanto o horizonte de previsão for $(t > n)$, onde n é o tamanho do horizonte de ajuste do modelo de cópula. Durante o período de previsão (isto é, o horizonte m) de $x_{t,i}$ é realizada a estimação de u_t . Então, o estimador combinado baseado no método de máxima verossimilhança de u_t , faz uso do conjunto de parâmetros $(\hat{\alpha}_1, \dots, \hat{\alpha}_k, \hat{\gamma})$ e do conjunto de previsões dos k modelos de predição $(x_{t,1}, \dots, x_{t,k})$ para prover o valor de \hat{u}_t , que maximiza a Equação 4.13. No entanto, devido ao fato de que a Equação pode conduzir para um problema intrínseco de otimização, neste caso, algoritmos de aproximação são requeridos.

4.6 Modelo Clássico

De modo geral, o chamado modelo Clássico neste trabalho refere-se ao preditor de séries temporais que negligencia um maior estudo sobre a distribuição dos erros cometidos pelos modelos de previsão utilizados na agregação. Neste sentido, o modelo Clássico assume que a distribuição marginal dos erros são normais e os erros foram gerados a partir de uma função multivariada normal. A literatura ainda apresenta diversos autores como Firmino [22] e T. F. Oliveira [41] que mostram meios de combinação através do modelo Clássico.

4.7 Algoritmo de simulação

As simulações foram conduzidas através do pacote de cópulas implementado no *software* R-Project [51] (versão 2.14.2), enfatizando que a semente aleatória utilizada nos experimentos simulados nas Seções 5.0.1 e 5.0.2 foi zero, enquanto na Seção 5.1 foi um. Salientando que todos os algoritmos que serão apresentados neste trabalho são determinísticos. Com base na arquitetura da abordagem proposta vista na Figura 4.1 é possível escrever o Algoritmo 1 que apresenta um pseudo-código das simulações que serão mencionadas neste Capítulo. Assim, para cada caso, se faz necessário instanciar esse algoritmo. Este algoritmo calcula os erros de cada modelo de previsão ($X_{t,1}$ e $X_{t,2}$, $k = 2$) por meio da função `CalculaErro()`, que por sua vez, calcula o erro utilizando a Equação 4.3 ou 4.4.

Estima os parâmetros das marginais através da Equação 4.11 ou 4.12. Posteriormente, o Algoritmo constrói a distribuição marginal dos erros (via função `pnorm()` ou `plnorm()` implementada no *software* R) passando como argumento $\hat{\alpha}_i$ e os erros dos modelos. Em seguida, os parâmetros da cópula são estimados através da Equação 4.10, em outras palavras, os parâmetros são estimados pelo método IFM. Logo após, c pode ser ajustado com qualquer cópula retornada pela função `AjustaCópula()`, com por exemplo Normal, Gumbel-Hougaard ou Cacoullos. Finalmente, a função `CombinarPrevisões()` realiza a combinação das previsões realizadas pelos modelos individuais $(x_{t,1}, \dots, x_{t,k})$.

A escolha entre aplicar uma equação e outra para calcular o erro ou construir a distribuição marginal irá depender da distribuição dos erros das previsões.

Entrada: $x_{t,1}, \dots, x_{t,k}, u_t, n, m$.

Resultado: Este algoritmo resulta em um vetor \hat{u}_t processado com as combinações de k previsões.

Início

Para $i \leftarrow 1$ **até** k **faça**

Para $t \leftarrow 1$ **até** n **faça**

$e_{t,i} \leftarrow \text{CalculaErro}(x_{t,i}, u_t)$;

fim para

$\hat{\alpha}_i \leftarrow$ é o vetor de parâmetros estimado via IFM;

$F_{e_{j,i}} \leftarrow \text{ConstróiDistribuiçãoMarginal}(e_{j,i}, \hat{\alpha}_i)$;

fim para

$\hat{\gamma} \leftarrow$ Ajusta o parâmetro da cópula utilizando $F_{e_{j,1}}, \dots, F_{e_{j,k}}$ via IFM ;

$c \leftarrow \text{AjustaCópula}(\hat{\alpha}_1, \dots, \hat{\alpha}_k, \hat{\gamma})$;

Para $t \leftarrow n + 1$ **até** $n + m$ **faça**

$\hat{u}_t \leftarrow \text{CombinarPrevisões}(x_{t,1}, \dots, x_{t,k}, F_{e_{j,1}}, \dots, F_{e_{j,k}}, \hat{\alpha}_1, \dots, \hat{\alpha}_k, c)$;

fim para

retorna \hat{u}_t ;

fim

Algoritmo 1: Este é um algoritmo genérico para construir um estimador combinado via cópula, para prever a série temporal u_t .

O Algoritmo 2 realiza a combinação de k modelos de previsão ajustados a uma mesma série temporal. Para tanto, foi definido um vetor \mathbf{l} de tamanho ($n_l = 1000$) igualmente espaçados entre o valor mínimo e máximo dos modelos de previsão e a série temporal, isto é, $\{x_{t,1}, x_{t,2}, u_t\}_{t=1}^n$. Para cada modelo é calculado o $e_{t,i}$ de acordo com o erro adotado. Por fim, a combinação ocorre através do procedimento de calcular a função de densidade de probabilidade de cópula passando como argumento os erros (e_1, \dots, e_k) , as distribuições marginais dos erros do conjunto de ajuste $(F_{e_{j,1}}, \dots, F_{e_{j,k}})$ e os parâmetros das marginais $(\hat{\alpha}_1, \dots, \hat{\alpha}_k)$, para posteriormente selecionar o elemento em l_t que maximiza a função \mathbf{fl} . Portanto, o elemento selecionado em l_s que maximiza a função de densidade de cópula é tomado como \hat{u}_t .

Função CombinarPrevisões(...)**Entrada:** $x_1, \dots, x_k, F_{e_{j,1}}, \dots, F_{e_{j,k}}, \hat{\alpha}_1, \dots, \hat{\alpha}_k, c.$ **Resultado:****Início** $n_l \leftarrow 1000;$ $l_t \leftarrow$ é um vetor de tamanho qualquer n_l com valores espaçados uniformemente entre o valor mínimo e máximo de $\{x_{t,1}, \dots, x_{t,k}, u_t\}_{t=1}^n$;**Para** $t \leftarrow 1$ **até** n_l **faça****Para** $i \leftarrow 1$ **até** k **faça** $e_i \leftarrow \text{CalculaErro}(x_i, l_t);$ **fim para** $\mathbf{fl} \leftarrow \text{c.densidade}(e_1, \dots, e_k, F_{e_{j,1}}, \dots, F_{e_{j,k}}, \hat{\alpha}_1, \dots, \hat{\alpha}_k);$ **fim para** $s \leftarrow$ é o índice do maior elemento contido no vetor \mathbf{fl} ;**retorna** l_s ;**fim****fim**

Algoritmo 2: Este algoritmo genérico realiza a combinação das previsões de k modelos individuais ajustados a uma mesma série temporal. Para tanto, a combinação ocorre utilizando os erros dos modelos através da função de densidade de cópula para obter o valor mais provável para série temporal u_t no instante $t + 1$.

4.8 Exemplo de combinação via modelo proposto

Nesta seção, será ilustrado um pequeno exemplo envolvendo apenas três pontos de uma série temporal, com intuito de realizar a combinação de dois modelos individuais de previsão. Nesse exemplo, será mostrado alguns detalhes sobre os procedimentos de combinação da abordagem proposta para agregar previsões.

Dadas as entradas para o Algoritmo 1, isto é, a série temporal ($u_t = -3, 26; -6, 97; -10$), os modelos individuais ($x_{t,1} = 16, 45; 9, 26; -15, 41; 0, 55$) e ($x_{t,2} = -18, 1; -35, 3; -40, 3; 31, 2$) ($t = 1, \dots, 4$), onde $x_{4,1}$ e $x_{4,2}$ (0,55 e 31,2) são as previsões

dos modelos individuais para a série temporal u_t no instante ($t = 4$) desconhecido. Nesse sentido, o tamanho do conjunto de treinamento e teste é $n = 3$ e $m = 1$.

Calcular erro dos modelos individuais

Primeiramente, o Algoritmo 1 calcula os erros cometidos por cada modelo individual. Para este exemplo, adotou-se o erro aditivo (sendo calculado pela Equação 4.3). Os erros obtidos de cada modelo foram $(e_{t,1} = 19, 7; 16, 23; -5, 42)$ e $(e_{t,2} = -14, 8; -28, 3; -30, 3)$, sendo $e_{t,1}$ o erro de $X_{t,1}$, bem como, $e_{t,2}$ é o erro de $X_{t,2}$.

Estimar os parâmetros das marginais ($\hat{\alpha}_i$)

Nesta etapa, é estimada a média e o desvio padrão dos erros de cada modelo individual de previsão ($e_{t,1}$ e $e_{t,2}$) através da Equação 4.11. Desta forma, os parâmetros (média e desvio padrão) do modelo $X_{t,1}$ são $(\hat{\alpha}_1 = 10, 2; 13, 6)$ e para $X_{t,2}$ são $(\hat{\alpha}_2 = -24, 5; 8, 45)$.

Construção das distribuições acumuladas marginais

Através dos erros e parâmetros das distribuições é possível construir a distribuição acumulada marginal dos erros de cada modelo. Para este exemplo, sendo $(F_{e_{j,1}} = 0, 77; 0, 67; 0, 07)$ e $(F_{e_{j,2}} = 0, 84; 0, 23; 0, 16)$ ($j = 1, \dots, n$).

Estimar parâmetro da cópula

O vetor de parâmetros da cópula é estimado via Equação 4.10. A quantidade de parâmetros do vetor γ varia de acordo com a cópula, conforme mencionado anteriormente. Neste exemplo, será construído um estimador combinado utilizado a cópula de Gumbel-Hougaard por possuir apenas um parâmetro. O valor estimado para γ foi 2,22. Posteriormente, é construído o estimador combinado via cópula de Gumbel-Hougaard utilizando o parâmetro γ estimado.

Combinar previsões

Passando os devidos argumentos para a função `CombinarPrevisões(·)` (ver Algoritmo 1), é possível estimar o próximo ponto da série temporal. Para tanto, é criado um vetor l_t com $n_l = 1000$ posições uniformemente espaçadas entre o valor máximo e mínimo do conjunto $\{x_{t,1}, x_{t,2}, u\}_{t=1}^n$, para este exemplo os limites mínimo e máximo são -40,32 e 16,50. Assim, o vetor l_t é preenchido com os seguintes valores $(-40, 32; -40, 26; \dots; 16, 44; 16, 50)$.

Posteriormente, é calculado o erro de previsão de cada modelo individual através da função `CalculaErro(x_i, l_t)`, porém, desta vez utilizando o vetor l_t ao invés da série temporal u_t , em outras palavras, calcula-se o erro desta maneira $(e_{t,1} = x_1 - l_t = 0,55 - (-40, 32); 0,55 - (-40, 26); \dots; 0,55 - 16, 50)$ ($t = 1, \dots, n_l$). Portanto, l_t contém os possíveis valores que a série temporal pode assumir. Desta forma, os erros obtidos são $(e_{t,1} = 40, 87; 40, 82; \dots; -15, 95)$ e $(e_{t,2} = 9, 11; 9, 05; \dots; -47, 66)$.

Em seguida, é obtida a densidade de cada erro $(e_{t,1}$ e $e_{t,2})$ através da Equação 4.13. No Algoritmo 2 a densidade é armazenada no vetor \mathbf{f} . Com isso, as densidades obtidas são $(\mathbf{f} = 4, 24; 4, 46; \dots; 1, 02; 1)$. Posteriormente, é obtido o índice (s) do elemento com maior valor contido em \mathbf{f} , tal índice, é utilizado para obter o elemento no vetor l_s com maior probabilidade de ser o próximo ponto da série temporal. Neste exemplo, o valor retornado para l_s é -8,55.

Assim, por meio dos passos apresentados anteriormente, obtém-se a estimativa do estimador combinado \hat{U}_t de -8,55 para o ponto futuro da série temporal, enquanto que os modelos individuais $X_{t,1}$ e $X_{t,2}$ previram 0,55 e 31,2. Na realidade o real valor da série temporal u_t no instante $(t = 4)$ foi de -13,67. Dessa maneira, em termos do EQM o estimador combinado obteve 25, enquanto $X_{t,1}$ e $X_{t,2}$ resultaram em 202,2 e 2.013,3. Logo, podemos verificar que o modelo proposto foi mais acurado quando comparado com os modelos individuais.

Capítulo 5

Experimentos e resultados

Este Capítulo apresenta alguns experimentos que fazem uso da cópula de Gumbel-Hougaard, Cacoullos e Normal para combinar modelos de previsão de séries temporais individuais. Será mostrado o potencial destas cópulas para combinar as previsões de modelos preditivos gerados via simulação. Primeiramente será introduzido dois pequenos casos de estudo, em que, cada caso leva em consideração apenas uma única série temporal. Na Seção 5.1 será apresentado um problema mais complexo que dedica-se a estudar vários modelos alternativos de cópulas aplicadas a inúmeras séries temporais.

5.0.1 Combinação via cópula de Gumbel-Hougaard

Conforme mencionado anteriormente a cópula Gumbel-Hougaard apresenta um modelo que capta dependência positiva, ou seja, capta dependência na cauda superior entre as variáveis. Esta fazendo parte da família Arquimediana de cópulas [36], formulada a partir da Equação 3.9, onde v_i é dada pela Equação 4.8. Assim, é possível genericamente acoplar a dependência positiva das distribuições marginais dos resíduos.

Caso de estudo

A fim de ilustrar a utilidade do estimador combinado, neste pequeno caso de estudo, serão gerados dois modelos ARIMA de previsão de séries temporais, chamados de $X_{t,1}$ e $X_{t,2}$, deste modo ($k = 2$). A série temporal u_t com 230 observações sendo ($n = 200$ e $m = 30$), foi gerada a partir de um modelo ARIMA(1,1,1), onde tanto o parâmetro da parte AR quanto MA obtiveram o valor 0,9.

Por sua vez, a série temporal u_t divide-se em duas partes: A primeira parte utilizada na fase de modelagem (treinamento) com n pontos ($t \leq n$), usados para construir o estimador combinado, isto é, estes pontos são usados para ajustar a cópula. A segunda parte utiliza os últimos m pontos ($n < t \leq n + m$) para realizar as previsões. Portanto, foram definidos os horizontes de ajuste e previsão que serão estudados nessa simulação.

Os erros dos modelos $X_{t,1}$ e $X_{t,2}$, isto é $E_{t,1}$ e $E_{t,2}$, obtidos por meio da Equação 4.1, foram gerados através da cópula Gumbel-Hougaard com parâmetro $\theta = 1,5$ (O valor de θ poderia ser qualquer outro, desde que, $1 \leq \theta \leq \infty$). Considerado que as marginais dos erros seguem uma distribuição normal, sendo que, $E_{t,1}$ possui $\mu = 10$ e $\sigma = 10$, enquanto o $E_{t,2}$ possui $\mu = -20$ e $\sigma = 5$. Então, deve-se enfatizar que a partir desta configuração marginal as previsões de $X_{t,1}$ são estatisticamente mais acuradas e menos eficientes que $X_{t,2}$, em outras palavras, $X_{t,1}$ é mais acurado por possuir um EQM inferior e menos eficiente por variar mais que $X_{t,2}$.

O método IFM é aplicado neste caso de estudo para ajustar os parâmetros (μ e σ) das distribuições marginais dos erros, além de ajustar o parâmetro de dependência da cópula Gumbel-Hougaard (θ).

Além do método de combinação de modelos via cópula chamado de \hat{U}_t , neste caso de estudo, também apresentará os resultados do modelo construído utilizando combinação linear (ver Seção 2.2) denominado de SA.

No Anexo I é exposto o código-fonte utilizado neste caso de estudo.

Resultados

A Tabela 5.1 apresenta a qualidade dos modelos em termos do erro quadrático médio e do desvio padrão dos resíduos de cada modelo, correspondendo as partes de ajuste dos modelos (treinamento) e a parte de previsão (teste). Pode-se ver que o estimador combinado expressivamente supera os modelos ARIMA individuais, quando o balanço entre a acurácia e eficiência é levada em conta pelo EQM. Por sua vez, sobre a eficiência (desvio padrão dos resíduos), o estimador combinado parece ser um pouco pior que $X_{t,2}$. Tal comportamento é principalmente por causa da incorporação de $X_{t,1}$ em \hat{U}_t , sendo que neste caso, $X_{t,1}$ é notavelmente menos eficiente que $X_{t,2}$ estatisticamente.

No que diz respeito ao EQM para parte de ajuste, o primeira modelo $X_{t,1}$ e o segundo modelo $X_{t,2}$ obtiveram 204,588 e 414,599, enquanto o estimador combinado e SA obtiveram 25,388 e 76,838. Para o conjunto de previsão, isto é, ($t > n$) o EQM obtido pelos modelos foi 180.247 e 435.324, já o estimador combinado e SA tiveram o desempenho de 20,850 e 69,217. Assim, o estimador combinado apresentou melhor desempenho neste caso de estudo específico.

Tabela 5.1: Erro quadrático médio e desvio padrão dos resíduos dos modelos ARIMA individuais ($x_{t,1}$ e $x_{t,2}$), estimador combinado (\hat{U}_t) e SA.

Estatística	Modelos	$t \leq n$	$n < t \leq n + m$
EQM	$X_{t,1}$	204,588	180,247
	$X_{t,2}$	414,599	435,324
	\hat{U}_t	25,388	20,850
	SA	76,838	69,217
Desvio Padrão dos Resíduos	$X_{t,1}$	10,815	10,139
	$X_{t,2}$	4,994	4,438
	\hat{U}_t	5,050	4,569
	SA	7,092	6,158

As Figuras 5.1 e 5.2 apresentam o desempenho dos modelos $X_{t,1}$, $X_{t,2}$, \hat{U}_t e SA. Particu-

larmente, percebe-se a partir da Figura 5.2 que em termos de acurácia o estimador combinado \hat{U}_t apresenta melhor ajuste à série temporal em estudo comparado com os demais modelos.

O presente caso de estudo teve o objetivo estudar o desempenho da cópula de Gumbel-Hougaard no problema de previsão de séries temporais e combinação de modelos. Este método traz a vantagem matemática de modelar a dependência entre preditores por meio de um único parâmetro. Enfatizando que para este caso específico o estimador combinado mostrou-se melhor em relação aos demais preditores.

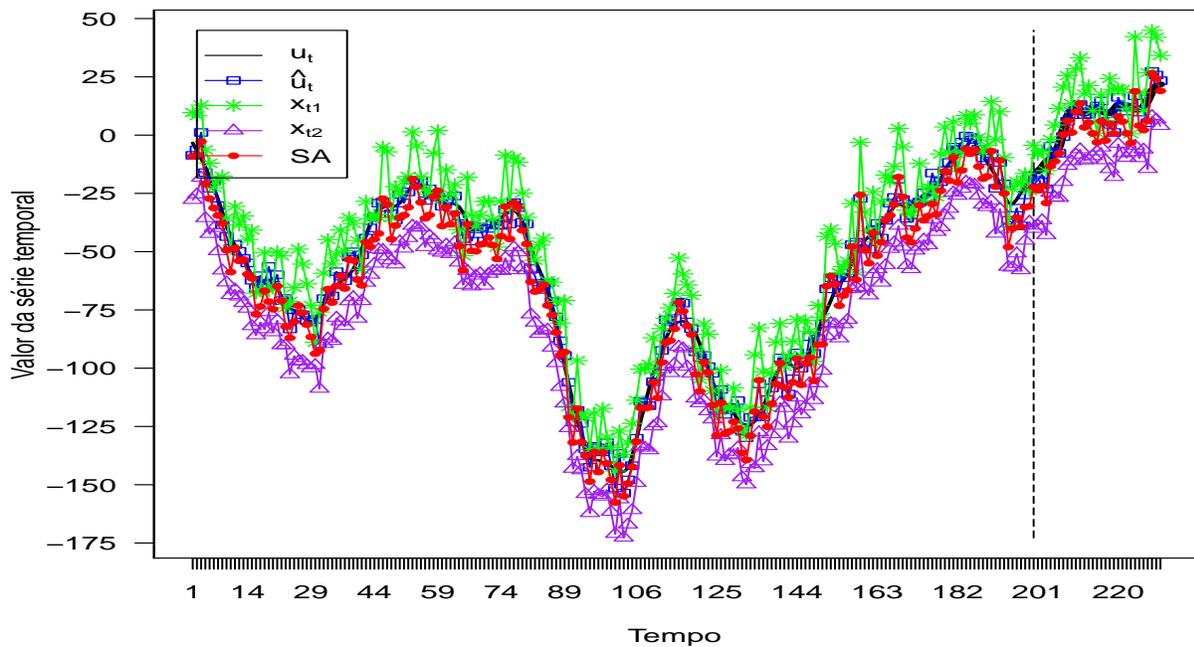


Figura 5.1: A série temporal u_t representada por (—), os respectivos modelos individuais ARIMA ($X_{t,1}$ e $X_{t,2}$) são representados por (*) e (Δ), e o estimador combinado \hat{U}_t e SA por (\square e \bullet). A linha tracejada indica o início do horizonte de previsão ($t > 200$).

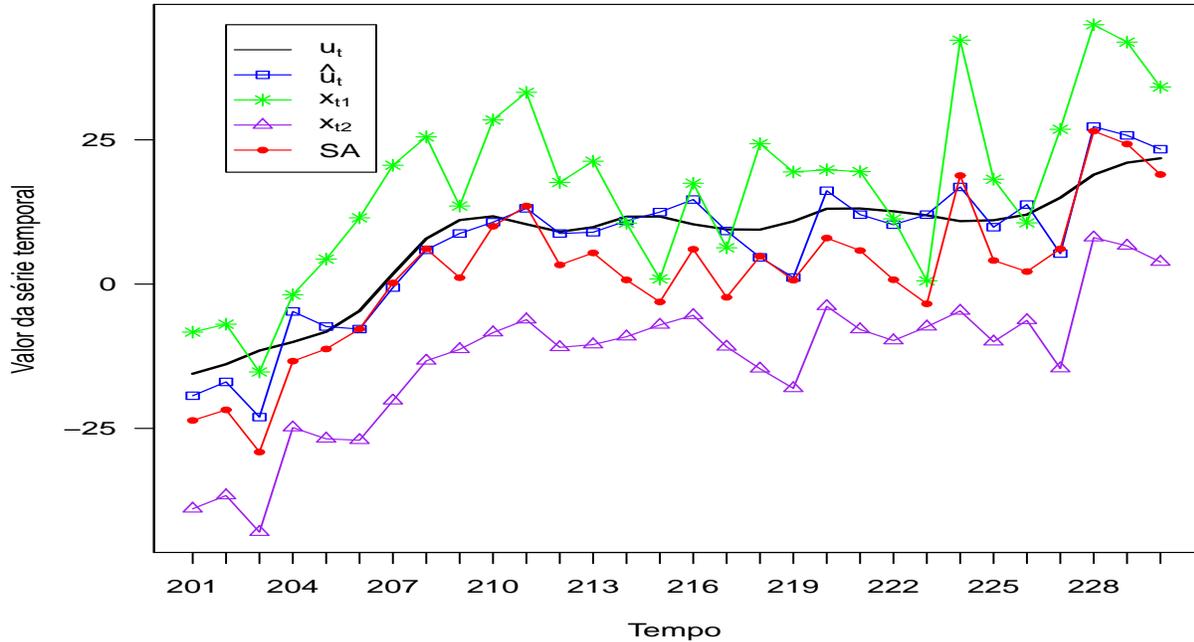


Figura 5.2: A série temporal u_t representada por (—), os respectivos modelos individuais ARIMA ($X_{t,1}$ e $X_{t,2}$) são representados por (*) e (Δ), e o estimador combinado \hat{U}_t e SA por (\square e \bullet). Nesta Figura, é levado em consideração apenas o desempenho das previsões no período ($t > 200$).

5.0.2 Combinação via cópula de Cacoullos

A cópula de Cacoullos proposta nesse caso de estudo, segue a mesma ideia apresentada nas Seções 4.1, 4.2, 4.3 e 4.5 onde são utilizados os erros dos modelos para estimar o valor de \hat{U}_t . O estimador combinado proposto pode ser expresso pela seguinte Equação [49]:

$$C(v_1, \dots, v_k) = \frac{1}{2\pi^{k/2}\lambda^k} \frac{1}{2} \sum_{t=1}^n \exp \left[- \sum_{i=1}^k \frac{(v_i - F_{e_{ti}})^2}{2\lambda^2} \right] \quad (5.1)$$

onde,

$k \equiv$ número de modelos de previsão de séries temporais

$\lambda \equiv$ parâmetro de suavização

$n \equiv$ número de observações da série temporal usada na parte de ajuste do estimador combinado

$v_i \equiv$ função acumulada marginal dos erros desconhecidos

$F_{e_{ti}} \equiv$ função acumulada marginal dos erros das observações de treinamento

Caso de estudo

A fim de ilustrar a utilidade da cópula de Cacoullos proposta nesta Seção, será apresentado em particular um caso de estudo para combinar as previsões dos modelos de predição de séries temporais nomeados como $X_{t,1}$ e $X_{t,2}$ ($t = 1, \dots, n$), onde ($k = 2$).

Serão realizados dois experimentos neste caso de estudo, onde cada experimento possui suas particularidades, para ilustrar o desempenho da cópula em diferentes situações. Para isso, uma série temporal foi utilizada para cada experimento. Em ambos os experimentos é adotado o erro aditivo (ver Equação 4.3), onde:

i No primeiro experimento, a série temporal foi gerada através do modelo ARIMA (1,0,1) com parâmetros (AR=0,824) e (MA=0,15), em que, a série resultante possui ($n + m = 530$) pontos. Por sua vez, a série temporal gerada foi dividida em duas partes, sendo a primeira parte utilizada para ajuste do estimador combinado com ($n = 500$) pontos (conjunto de exemplos do estimador). Por fim, a segunda parte fez uso de ($m = 30$) pontos destinados ao conjunto de teste para avaliar o desempenho do estimador combinado. Os modelos de previsão de séries temporais $X_{t,1}$ e $X_{t,2}$, utilizados neste experimento, foram obtidos através da soma da série temporal com os erros $E_{t,1}$ e $E_{t,2}$ simulados por meio da cópula Normal com parâmetro de dependência ($\beta = 0, 2$). Os erros gerados seguem uma distribuição normal com média ($\mu_{E_{t,1}}$ e $\mu_{E_{t,2}}$ iguais a -20), e desvio padrão ($\sigma_{E_{t,1}} = 3$) e ($\sigma_{E_{t,2}} = 5$). Nas simulações realizadas neste caso de estudo admiti-se que o parâmetro da cópula de Cacoullos assume o valor ($\lambda = 0, 5$).

ii O segundo experimento foi ajustado igual ao primeiro, com exceção do tamanho da série temporal e o parâmetro λ da cópula. Neste contexto, o tamanho da série é ($n + m = 130$) pontos, dividida em ($n = 100$) pontos para o conjunto de ajuste e ($m = 30$)

pontos para avaliar o estimador combinado. Este segundo experimento tem o objetivo de avaliar quanto o valor do parâmetro de suavização (λ) pode influenciar no desempenho do estimador combinado. Nesse sentido, o parâmetro de suavização assumirá diversos valores durante as simulações, de modo a obter o valor de λ que minimiza o EQM do estimador combinado.

Em ambos os experimentos o tamanho da série temporal, bem como, as partes de treinamento e teste foram definidas sem qualquer motivo específico. Neste sentido, poderia ter sido utilizado outras definições.

Neste experimento, além da combinação via cópula de Cacoullos também será realizada a combinação dos modelos individuais através do método de combinação linear SA. Desta forma, almeja-se avaliar a robustez do estimador combinado comparado com os modelos individuais e o método SA.

Uma instância do Algoritmo 1 executa as simulações com a cópula de Cacoullos. Neste caso de estudo não foi usada a função de densidade de probabilidade de cópulas previamente implementada no pacote “copula”, disponíveis no *software* R-Project, uma vez que, a cópula de Cacoullos é uma abordagem não paramétrica proposta neste trabalho. Para tanto, o Algoritmo 3 calcula a densidade de probabilidade da cópula de Cacoullos.

Sendo assim, a combinação das previsões via Algoritmo 2 ocorre por meio da função de densidade de probabilidade da cópula de Cacoullos, vista no Algoritmo 3, ao invés de utilizar uma função de densidade previamente implementada no R-Project para combinar as previsões.

O Algoritmo 3 constrói as distribuições marginais acumuladas dos erros desconhecidos, isto é, realiza a extrapolação dos pontos da série temporal de interesse, em que, l_j no Algoritmo 2 representa os possíveis valores que a série temporal pode assumir. Logo, l_s interpola entre o valor mínimo e máximo do conjunto $\{x_{t,1}, x_{t,2} \text{ e } u_t\}_{t=1}^n$. Assim, (e_1, \dots, e_k) são os erros da diferença entre $(x_1 - l_t, \dots, x_k - l_t)$ onde $(t = 1, \dots, n_l)$, e seja $(F_{e_{t,1}}, \dots, F_{e_{t,k}})$ o conjunto de ajuste (treinamento) e (v_i, \dots, v_k) é o conjunto de teste da cópula. Enfim, este algoritmo tem por finalidade calcular a densidade de probabilidade da cópula de Cacoullos.

Função dCacoullos(...)**Entrada:** $e_1, \dots, e_k, F_{e_{t,1}}, \dots, F_{e_{t,k}}, \hat{\alpha}_1, \dots, \hat{\alpha}_k$.**Resultado:** Resulta na densidade da cópula de Cacoullos.**Início** $v_i \leftarrow \text{ConstróiDistribuiçãoMarginal}(e_1, \hat{\alpha}_1) ;$ \vdots $v_k \leftarrow \text{ConstróiDistribuiçãoMarginal}(e_k, \hat{\alpha}_k) ;$ $\lambda \leftarrow 0.5 ;$ $c \leftarrow \frac{1}{2\pi^{k/2}\lambda^k} \frac{1}{2} \sum_{t=1}^n \exp \left[- \sum_{i=1}^k \frac{(v_i - F_{e_{t,i}})^2}{2\lambda^2} \right] ;$ **retorna** $c \prod_{i=1}^k p_{E_{t,i}}(e_{t,i} | \hat{\alpha}_i)$ **fim****fim**

Algoritmo 3: Este algoritmo calcula a densidade de probabilidade da cópula de Cacoullos.

Resultados do experimento (i)

Nesta Subsecção serão exibidos os resultados obtidos através da combinação das previsões de dois modelos de predição de séries temporais ($X_{t,1}$ e $X_{t,2}$).

A Tabela 5.2 mostra o desempenho de cada modelo para prever a série temporal. Analisando a tabela, fica claro a superioridade do estimador combinado em termos do EQM, onde o valor obtido por $\hat{U}_t \cong 11$ enquanto os modelos individuais $X_{t,1}$ e $X_{t,2}$ obtiveram aproximadamente 389 e 363, e o SA obteve 369 para ($n < t \leq n + m$).

A Figura 5.3 apresenta o desempenho dos modelos ($X_{t,1}$, $X_{t,2}$, \hat{U}_t e SA) para prever a série temporal. Nesta figura, fica notável a superioridade do estimador combinado representado por (\square), onde \hat{U}_t ajusta-se melhor a série temporal em relação ao demais modelos.

A Tabela 5.2 apresenta ainda o EQM e desvio padrão de cada modelo de previsão. Como pode ser visto, o modelo \hat{U}_t mostrou o menor desvio em relação aos demais modelos, assim como, a melhor acurácia.

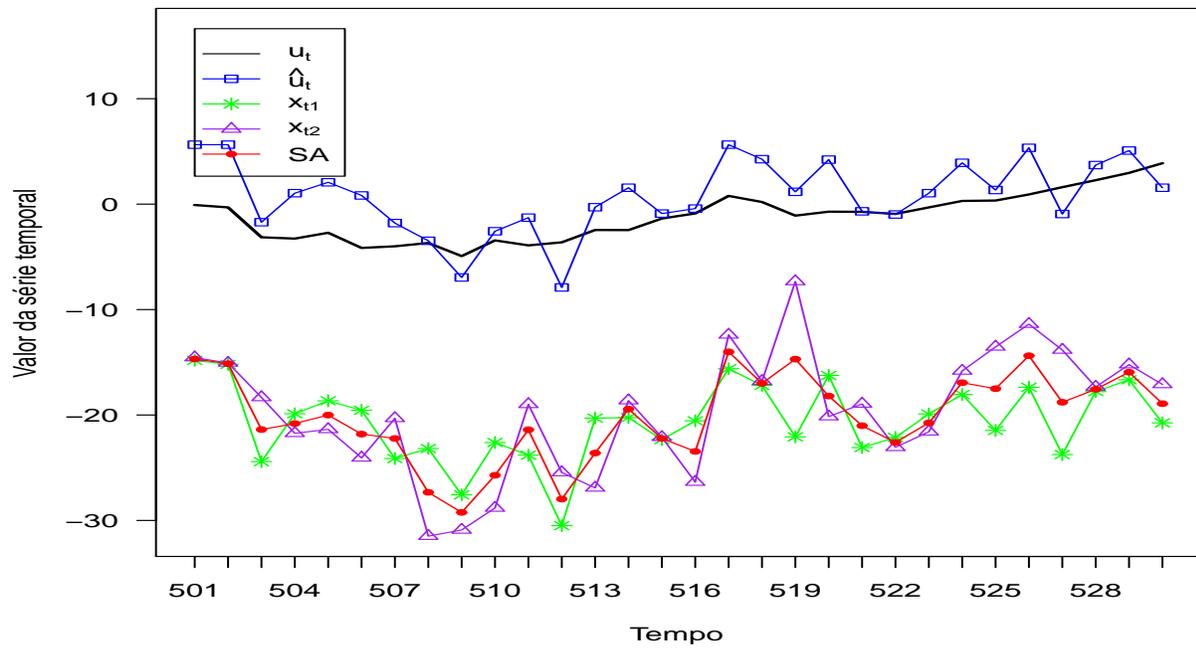


Figura 5.3: A série temporal u_t representada por $(-)$, enquanto $(*)$ e (Δ) representam respectivamente os modelos ARIMA individuais $X_{t,1}$ e $X_{t,2}$, o estimador combinado \hat{U}_t e SA são ilustrado por (\square) e (\bullet) . Esta figura leva em consideração apenas o desempenho dos modelos no período onde $(t > n)$.

Tabela 5.2: Desempenho dos modelos individuais, SA e estimador combinado construído via cópula de Cacoullos para o conjunto de exemplos ($n = 500$) e o conjunto de previsão ($m = 30$).

Estatística	Modelos	$n < t \leq n + m$
EQM	$X_{t,1}$	389,096
	$X_{t,2}$	363,240
	\hat{U}_t	10,532
	SA	369,405
Desvio Padrão dos Resíduos	$X_{t,1}$	3,049
	$X_{t,2}$	4,698
	\hat{U}_t	2,624
	SA	2,992

Resultados do experimento (ii)

A Figura 5.4 apresenta os resultados alcançados através dos testes realizados variando o valor de λ entre 0,01 até 1 e de 1 até 100, mais especificamente a Tabela 5.3 mostra o valor do parâmetro de suavização pelo EQM do estimador combinado. Analisando detalhadamente a Tabela, pode-se observar que as simulações mostram os resultados obtidos em termos do EQM através da combinação dos modelos $X_{t,1}$ e $X_{t,2}$. Quando o parâmetro de suavização assume valores no intervalo ($\lambda = 1, 2, \dots, 100$), a simulação *A* destina-se a analisar os resultados obtidos através do conjunto de treinamento, enquanto a simulação *B* refere-se ao conjunto de teste. Já as simulações *C* e *D* são referentes aos conjuntos de treinamento e de teste, variando o valor do parâmetro de suavização no intervalo ($\lambda = 0, 01; 0, 02; \dots; 1$).

No contexto deste trabalho, maximizar o desempenho do estimador combinado indica minimizar o erro quadrático médio cometido pelo estimador combinado, enquanto minimizar o desempenho segue o raciocínio inverso. Logo, pode-se notar na Figura 5.4 que quando o parâmetro de suavização varia entre (1 e 100) o valor que maximiza ($\lambda_{exemplo} = 1$ e $\lambda_{teste} = 1$) e minimiza ($\lambda_{exemplo} = 26$ e $\lambda_{teste} = 15$) o desempenho do estimador combinado. Por outro

lado, quando λ varia entre 0,01 até 1, o valor que maximiza ($\lambda_{exemplo} = 0,69$ e $\lambda_{teste} = 0,75$) e minimiza ($\lambda_{exemplo} = 0,01$ e $\lambda_{teste} = 0,01$) respectivamente.

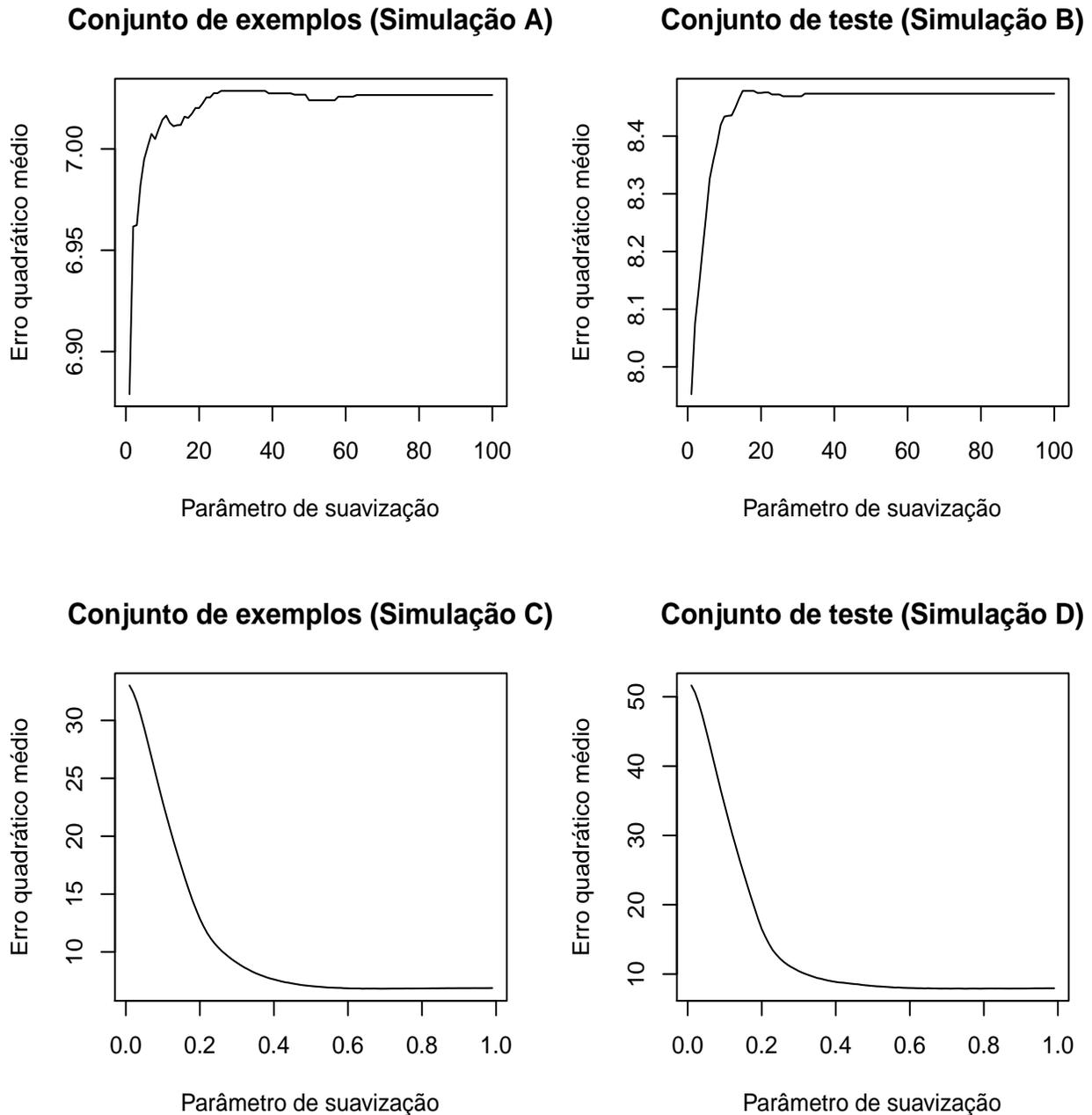


Figura 5.4: Parâmetro de suavização (λ) pelo EQM do estimador combinado.

A Tabela 5.3 mostra como a escolha incorreta para o parâmetro da cópula pode levar a resultados indesejáveis. Seja \hat{U}_t^* o estimador combinado ajustado com o valor de λ que

maximiza o desempenho de tal forma que conduz os resultados a obter um EQM menor. Enquanto \widehat{U}_t^{**} é o estimador combinado ajustado com o parâmetro λ que minimiza seu desempenho. Neste contexto, \widehat{U}_t^{**} conduz a resultados piores em relação a \widehat{U}_t^* , como pode ser visto na Tabela 5.3.

Com base nos resultados apresentados nas Tabelas 5.3 e 5.4 é razoável dizer que o parâmetro de suavização está diretamente relacionado com o desempenho do estimador combinado. Logo, a utilização de um parâmetro inadequado na combinação de modelos via cópula de Cacoullos conduz a resultados significativamente inferiores, comparado com os resultados obtidos utilizando um parâmetro adequado de suavização.

Tabela 5.3: Avaliação de desempenho baseada no erro quadrático médio do estimador combinado construído via cópula de Cacoullos, onde \widehat{U}_t^* e \widehat{U}_t^{**} são os estimadores combinados ajustados com o valor de λ que regula o desempenho do estimador.

Estatística	Modelos	$n < t \leq n + m$
EQM ($\lambda = 1, 2, \dots, 100$)	\widehat{U}_t^*	7,952
	\widehat{U}_t^{**}	8,478
EQM ($\lambda = 0.01, 0.02 \dots, 1$)	\widehat{U}_t^*	7,888
	\widehat{U}_t^{**}	51,637

Conforme mencionado por Donald F. [48] *apud* Specht [49] parece não ser difícil encontrar um valor adequado para λ quando este varia entre 1 e 100, bem como, os resultados alcançados com um ajuste melhor não são expressivos. Porém, quando o intervalo de variação passa a ser entre 0,01 e 1, os resultados são bem diferentes. Neste sentido, a discrepância dos resultados entre o estimador \widehat{U}_t^* e \widehat{U}_t^{**} são mais expressivas, sendo notável a necessidade de estimar um melhor valor de λ para \widehat{U}_t^{**} .

A Figura 5.5 ilustra o desempenho obtido pelo estimador combinado \widehat{U}_t^{**} ajustado com o parâmetro λ que minimiza o desempenho. Como pode ser visto, o estimador combinado representado por (\square), mostrou-se melhor comparado com os modelos individuais, embora tenha sido ajustado previamente com parâmetro de suavização que minimize seu desempenho. Enquanto, os modelos individuais $X_{t,1}$ e $X_{t,2}$ são parecidos pois possuem a mesma média e

desvios padrões similares. Apesar da figura apresentar apenas a parte de previsão da série, o estimador combinado mostrou-se superior também na parte de ajuste para $(t \leq n)$.

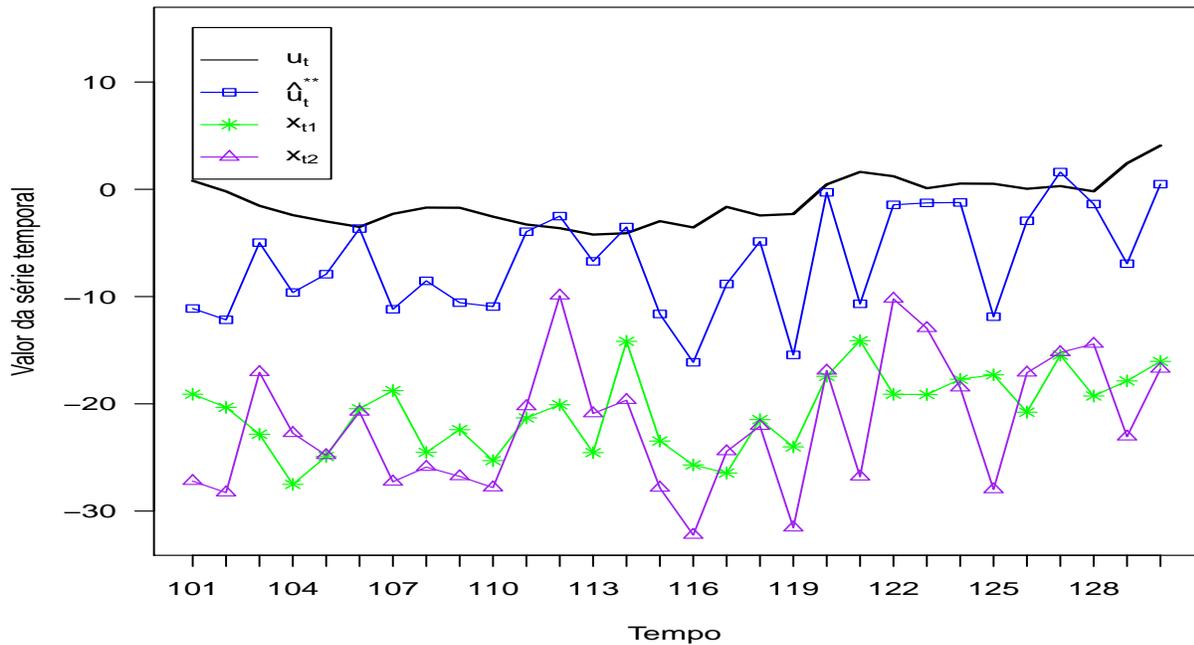


Figura 5.5: Ilustra o desempenho do estimador combinado \hat{U}_t^{**} representado por (\square) ajustado com o parâmetro λ que minimiza sua acurácia para estimar a série temporal u_t .

A Figura 5.6 apresenta a mesma série temporal e os mesmos modelos de previsão individuais $X_{t,1}$ e $X_{t,2}$, mostrados na Figura 5.5, porém o estimador combinado \hat{U}_t^* foi ajustado com um parâmetro λ adequado, tal parâmetro foi obtido através da análise resultante da Tabela 5.3. O estimador combinado \hat{U}_t^* desempenha melhor comportamento comparado com os modelos individuais e a Tabela 5.4 comprova essa superioridade do estimador quando ajustado de forma adequada. A Tabela 5.4 mostra ainda os resultados obtidos em termo do EQM para os modelos individuais ($X_{t,1}$ e $X_{t,2}$) e os estimadores combinados (\hat{U}_t^* e \hat{U}_t^{**}) para as partes de previsão $(n < t \leq n + m)$ dos modelos.

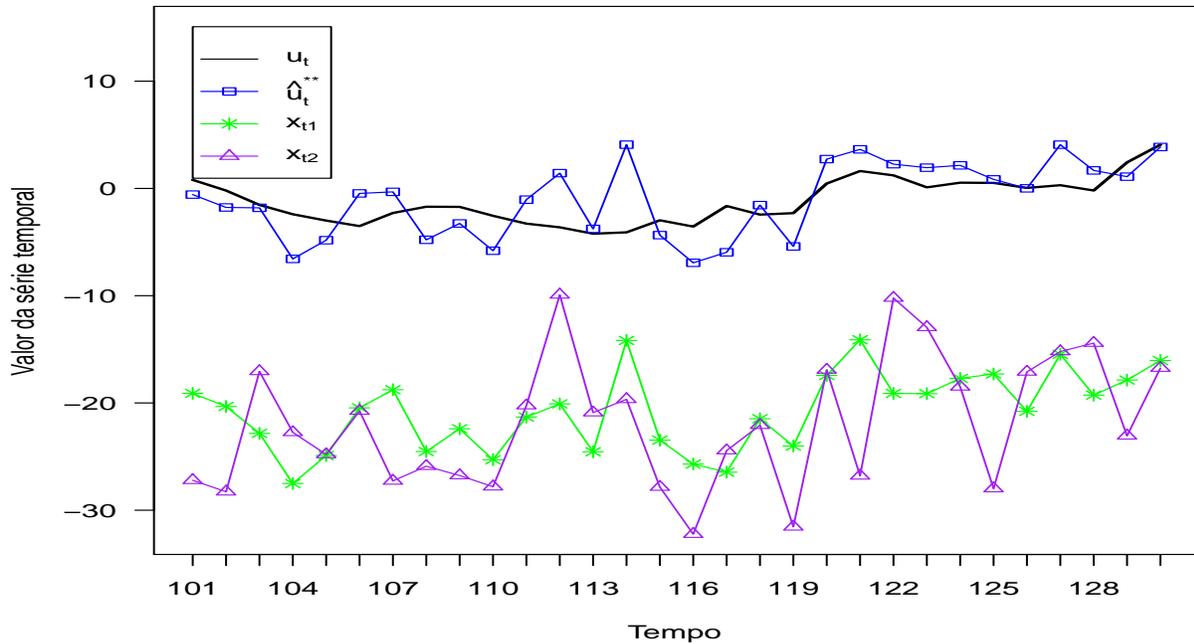


Figura 5.6: Desempenho do estimador combinado \hat{U}_t^* representado por (\square) ajustado com o parâmetro λ que maximiza o EQM de \hat{U}_t^* para estimar a série temporal u_t .

Tabela 5.4: Erro quadrático médio dos modelos individuais e dos estimadores combinados via cópula de Cacoullos. Onde \hat{U}_t^* e \hat{U}_t^{**} foram ajustados com parâmetro λ adequado e inadequado.

Estatística	Modelos	$n < t \leq n + m$
EQM	$X_{t,1}$	390,851
	$X_{t,2}$	464,405
	\hat{U}_t^*	7,888
	\hat{U}_t^{**}	51,637

Este caso de estudo envolveu relativamente uma pequena quantidade de preditores e apenas uma série temporal, mesmo assim, foi possível ilustrar o potencial deste método de combinação. Logo, é razoável dizer que para este caso de estudo em particular o estimador combinado por meio da cópula de Cacoullos obteve resultados mais atrativos, comparado

com os modelos individuais e o método de combinação linear apresentado. No segundo experimento, constatou-se que o parâmetro de suavização influencia diretamente na obtenção de bons resultados de estimação.

5.1 Uma análise comparativa entre estimadores combinados

A presente Seção tem o objetivo de realizar uma análise comparativa entre estimadores combinados, chamados neste trabalho como modelo Correto, Clássico, Cacoullou e *Simple Average* (SA). O interesse deste caso de estudo é avaliar se existe vantagem em utilizar um modelo ao invés de outro.

5.1.1 Caso de estudo

Este caso de estudo propõe através de simulações, analisar o desempenho dos estimadores combinados. Desta forma, almeja-se avaliar se existe vantagem em estudar as distribuições dos erros e de suas dependências. Neste sentido, o estimador combinado Correto é construído levando em consideração qual a distribuição dos erros cometidos pelos modelos, e qual a cópula que respectivamente gerou os erros dos modelos. Enquanto o estimador combinado Clássico é construído assumindo que a distribuição dos erros segue uma Normal, assim como, considera-se que os resíduos foram gerados por uma cópula Normal. Além disso, será avaliado o desempenho do estimador combinado via cópula de Cacoullou e SA com os outros estimadores. Neste caso de estudo foram simulados 1027 casos, em que, para cada caso foi gerada uma série temporal e dois modelos individuais de previsão, para então, combinar as previsões dos modelos.

Geração de série temporal

É gerada uma série temporal u_t , a partir de um modelo $ARIMA(p, d, q)$, onde tanto os argumentos (p e q), quanto os parâmetros dos modelos AR e MA são gerados aleatoriamente

nos intervalos 0, 1 ou 2 para os argumentos p e q , enquanto, d sempre assume o valor 0 respectivamente. Neste trabalho a série temporal pode ser gerada apenas com valores positivos, quando isto ocorre, é dito que a natureza da série temporal é *log*, caso contrario a natureza é dita *original* com a série assumindo valores negativos e positivos.

A série temporal gerada possui 1840 pontos. Sendo dividida em duas fases (ajuste e teste), tal que, a fase de ajuste (treinamento) a série dispõem de n_j pontos ($n_j = (50, 100, 200, 400, 800, 1600)$) ($j = 1, \dots, 6$). Para a fase de teste, a série temporal conta com m_j pontos ($m_j = (9, 18, 35, 71, 141, 282)$). Logo, para cada caso, a série temporal u_t é particionada em 6 partes, em que, cada parte é utilizada para combinar as previsões realizadas pelos modelos individuais.

Portanto, a amostra obtida da série temporal é dividida em duas partes (n_j e m_j): sendo que a primeira parte foi destinada a fase de ajuste do estimador combinado possuindo 85% da amostra, enquanto a segunda parte utilizada para fase de teste possui os 15% restantes da amostra. Mais detalhes serão apresentados na Subseção 5.1.2 sobre o procedimento de simulação.

Construção de modelos individuais

As distribuições dos erros ($E_{t,1}$ e $E_{t,2}$) dos modelos individuais são gerados aleatoriamente, e podem seguir uma distribuição log-normal ou normal. Neste contexto, quando os erros seguirem uma distribuição log-normal, estes são obtidos por meio da Equação 4.4, ou seja, utiliza-se o erro multiplicativo, para evitar a obtenção de um erro negativo. Caso contrário, quando a distribuição for normal, utiliza-se o erro aditivo através da Equação 4.3, enfatizando que poderia ter sido utilizado outro tipo de erro.

Desta forma, $E_{t,1}$ e $E_{t,2}$ são gerados aleatoriamente por meio da cópula Normal ou Gumbel-Hougaard (sendo esta selecionada aleatoriamente). Então o modelo é construído pela adição de $X_{t,i} = u_t + E_{t,i}$ ou produto de $X_{t,i} = u_t \times E_{t,i}$ para ($i = 1, \dots, k$). No caso de estudo em questão, os erros dos modelos são gerados pelas cópulas utilizando os respectivos parâmetros $\beta = 0,2$ para Normal e $\theta = 1,5$ para Gumbel-Hougaard. Considerou-se a média e desvio padrão dos erros sendo $\mu = 0,5$ e $\sigma = 0,5$ para o $E_{t,1}$, enquanto $E_{t,2}$ possui $\mu = 10$

e $\sigma = 0,25$. Enfatizando que, quando a natureza da série não for *original* aplica-se o log nas observações marginais. Então, a partir deste ajuste o modelo $X_{t,1}$ é estatisticamente mais acurado, porém menos eficiente que $X_{t,2}$.

Combinação de previsões

A série temporal é particionada em 6 partes (n_j e m_j) para ($j = 1, \dots, 6$), em que, para cada parte é realizada a combinação das previsões dos modelos individuais. Neste contexto, o estimador combinado Correto é modelado a partir da mesma cópula que gerou os erros, bem como, estuda a real distribuição dos erros dos modelos individuais (podendo ser log-normal ou normal). Enquanto o estimador combinado Clássico é construído assumindo que os erros seguem uma distribuição normal e que foram gerados a partir da cópula Normal. Os parâmetros das marginais, assim como, da cópula são obtidos por meio do método IFM (ver Seção 4.4). Desta forma, os estimadores realizam a combinação das predições obtidas pelos modelos individuais $X_{t,1}$ e $X_{t,2}$.

5.1.2 Procedimento de simulação

Este experimento foi conduzido através dos seguintes passos:

1. Neste primeiro momento, são geradas as séries temporais u_t , onde $n_{caso} = 1027$ é a quantidade de casos e conseqüentemente o número de séries que serão executados neste experimento. As séries temporais são geradas de forma aleatória, neste contexto, tanto os argumentos do modelo ARIMA (p , d e q) quanto os parâmetros (AR e MA) e a natureza da série temporal são obtidas de forma aleatória (ver metodologia Box & Jenkins [5]);
2. Em seguida, são constituídos os modelos ARIMA de previsão da série temporal u_t . Com base na natureza da série temporal são gerados os resíduos dos modelos. Quando a natureza da série temporal for *log* então a distribuição dos erros $E_{t,1}$ e $E_{t,2}$ podem seguir uma distribuição log-normal ou normal, mas quando a natureza da série for *original* a distribuição dos erros por padrão é normal. Logo, quando $E_{t,i}$ for log-

normal então $X_{t,i} = u_t \times E_{t,i}$, senão $X_{t,i} = u_t + E_{t,i}$. Enfatizando, que nos casos em que as distribuições marginais dos erros forem normais, então a cópula geradora dos resíduos é Gumbel-Hougaard;

3. A série temporal é originada com 1840 pontos, em que, será particionada em 6 partes e dividida entre o conjunto de ajuste (n_j) e de teste (m_j);
4. Em seguida, são calculados os erros dos modelos ARIMA gerados no passo 2, levando em consideração a respectiva distribuição dos erros;
5. Os parâmetros das distribuições marginais dos erros (μ e σ), assim como, os parâmetros de dependências da cópula, mais especificamente o parâmetro da cópula Gumbel-Hougaard (θ) e da cópula Normal (β) são estimados através do método IFM;
6. Posteriormente, o estimador combinado Correto é ajustado com base na cópula geradora dos resíduos e na distribuição dos resíduos. Em seguida, o estimador combinado Clássico é ajustado admitindo-se que a cópula geradora dos erros é normal, assim como, as distribuições dos erros. Enquanto o estimador combinado Cacoullos é configurado assumindo que as marginais são normais e a cópula de Cacoullos que gerou os erros. A estratégia SA utilizada neste trabalho não necessita de qualquer ajuste;
7. Finalmente, os estimadores combinados são utilizados para realizar a combinação das previsões obtidas através dos modelos individuais. Nas simulações foram tidas em consideração as previsões quando ($n_j < t \leq m_j$).

5.1.3 Aplicação do teste de hipóteses

A comparação realizada entre os estimadores foi baseada na análise estatística feita por meio do teste de hipótese Kolmogorov-Smirnov (KS) e Wilcoxon ou t-Student. Para tanto, foi utilizado o teste de normalidade através do KS para avaliar se o EQM_R segue uma distribuição normal, caso essa hipótese seja aceita é aplicado o teste de igualdade de médias t-Student, caso contrário, aplica-se o teste Wilcoxon para verificar se o EQM_R do estimador combinado \hat{U}_t^a é melhor em relação ao estimador \hat{U}_t^b , uma vez que, \hat{U}_t^a e \hat{U}_t^b são os estimadores combinados estudados nesta Seção. O nível de significância adotado neste trabalho foi de 5%.

Para calcular o p -valor do teste KS é utilizado o EQM_R dado pela Equação 2.17, sendo que para computar o EQM dos estimadores combinados, foi considerado apenas a parte de ajuste da série temporal ($t \leq n_j$). Enquanto no teste t-Student e Wilcoxon o EQM é calculado utilizando a parte de previsão da série temporal ($t > n_j$). A hipótese do teste KS:

H_0 : o EQM_R segue uma distribuição normal;

H_1 : o EQM_R não segue uma distribuição normal.

enquanto a hipótese para o teste t-Student ou Wilcoxon é dada por:

H_0 : o $EQM_R = 0$;

H_1 : o $EQM_R \neq 0$.

5.1.4 Resultados

Nesta Subsecção serão apresentados os resultados obtidos pelo experimento proposto. A Tabela 5.5 ilustra 13 possíveis situações que foram usadas neste trabalho, cada sigla representa uma situação, bem como, a sigla NNG indica que o $E_{t,1}$ e $E_{t,2}$ segue uma distribuição normal e os erros foram gerados por uma respectiva cópula Gumbel-Hougaard. Logo, o estimador combinado Correto é ajustado com a mesma situação utilizada no teste, por sua vez, o estimador combinado Clássico é sempre ajustado com a situação NNN ($E_{t,1}$ e $E_{t,2}$ são normais com cópula normal). A situação NNN é a única que não foi gerada nas simulações, uma vez que, através desta o estimador combinado Correto passaria a ser igual ao Clássico.

Nesta Subsecção serão apresentados os resultados obtidos pelo experimento proposto. A Tabela 5.5 ilustra 13 possíveis situações que foram usadas neste trabalho, cada sigla representa uma situação, bem como, a sigla NNG indica que o $E_{t,1}$ e $E_{t,2}$ segue uma distribuição normal e os erros foram gerados por uma respectiva cópula Gumbel-Hougaard. Logo, o estimador combinado Correto é ajustado com a mesma situação utilizada no teste, por sua vez, o estimador combinado Clássico é sempre ajustado com a situação NNN ($E_{t,1}$ e $E_{t,2}$ são normais com cópula normal). Por sua vez, a situação NNC indica que a distribuição dos erros é normal e foram gerados a partir da cópula de Cacoullos. As situações NNN e NNC são as únicas que não foram geradas nas simulações, uma vez que, através destas o estimador

combinado Clássico e Cacoullos passariam a ser iguais ao estimador Correto.

Neste contexto, a situação LLG indica que os erros seguem uma distribuição log-normal, sendo que tais erros foram gerados a partir da cópula Gumbel-Hougaard. Por sua vez, a sigla LNN representa a situação quando a distribuição do erro $E_{t,1}$ é log-normal, enquanto $E_{t,2}$ segue uma distribuição normal com a cópula geradora dos erros sendo normal.

A Tabela 5.6 apresenta os resultados obtidos por meio dos testes realizados. A coluna “Situações” expõe as siglas das situações listadas na Tabela 5.5 e a quantidade de casos que tais situações acontecem são exibidas na coluna “Amostra”, assim a situação NNG ocorre 641 vezes neste experimento, por exemplo. A coluna “Conjunto de treinamento” mostra as partições da série temporal, em outras palavras, é expressa seis séries temporais com tamanhos distintos para treinamento do modelo combinado. O $\overline{\text{EQM}}_R$ entre os dois estimadores combinados que estão sendo comparados é expresso na coluna “ $\overline{\text{EQM}}_R$ ”. O *p-valor* do teste KS é apresentado na coluna “Teste (KS)”. A coluna “Teste comparativo” apresenta o *p-valor* obtido através do teste t-Student ou Wilcoxon. Conforme mencionado anteriormente, a seleção do teste é feita com base no resultado do teste KS. Enquanto a coluna “Melhor modelo” apresenta o nome do estimador combinado mais acurado, com base no teste de hipótese. Neste sentido, de modo a exemplificar a interpretação da tabela, quando a situação for NNG à amostra possuirá 641 séries e dentre as séries temporais com conjunto de treinamento igual a 50 o $\overline{\text{EQM}}_R$ obtido é de 1.744298e-02, e o *p-valor* resultante do teste KS é 1.905183e-03, ou seja, passando no teste de normalidade, em seguida, é aplicado o teste t-Student que ocasiona no valor de 1.752771e-01 exibido na coluna “Teste comparativo”, logo não passando no teste t-Student, fazendo com que os estimadores sejam considerados equivalentes como pode ser visto na coluna “Melhor modelo”.

Tabela 5.5: Combinação de situações possíveis de acontecer.

Situação	Distribuição do $E_{t,1}$	Distribuição do $E_{t,2}$	Cópula
1. NNG	Normal	Normal	Gumbel-Hougaard
2. LLG	Log-normal	Log-normal	Gumbel-Hougaard
3. LLN	Log-normal	Log-normal	Normal
4. LNG	Log-normal	Normal	Gumbel-Hougaard
5. NLG	Normal	Log-normal	Gumbel-Hougaard
6. LNN	Log-normal	Normal	Normal
7. NLN	Normal	Log-normal	Normal
8. LNA	Log-normal	Normal	Gumbel-Hougaard ou Normal
9. NLA	Normal	Log-normal	Gumbel-Hougaard ou Normal
10. LLA	Log-normal	Log-normal	Gumbel-Hougaard ou Normal
11. AAN	Log-normal ou Normal	Log-normal ou Normal	Normal
12. AAG	Log-normal ou Normal	Log-normal ou Normal	Gumbel-Hougaard
13. AAA	Log-normal ou Normal	Log-normal ou Normal	Gumbel-Hougaard ou Normal
14. NNN ¹	Normal	Normal	Normal
15. NNC ¹	Normal	Normal	Cacoullos

Analisando a Tabela 5.6, observa-se que a situação NNG, para todos os conjuntos de previsão o teste rejeita a hipótese dos estimadores serem iguais (ver coluna “Teste comparativo”), respectivamente aceitando que são desiguais. Desta forma, através do $EQM_R(\hat{U}_t^a, \hat{U}_t^b)$, onde ($\hat{U}_t^a \equiv$ estimador Correto e $\hat{U}_t^b \equiv$ estimador Clássico) obtém-se que o estimador combinado Clássico é melhor quando o $(\overline{EQM}_R > 0)$ (ver coluna “ \overline{EQM}_R ”), caso contrário, o estimador combinado Correto mostra-se melhor. Assim, para a situação NNG o modelo Clássico mostrou-se melhor.

¹A situação NNN e NNC não é levada em conta neste trabalho.

Tabela 5.6: Sumário do teste comparativo entre os estimadores Correto e Clássico.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
NNG	641	50	1,744298e-02	1,905183e-03	1,752771e-01	Equivalentes
		100	7,055493e-03	2,809238e-02	2,380341e-01	Equivalentes
		200	8,487181e-03	7,329366e-01	1,37091e-03	Clássico
		400	4,531735e-03	3,426619e-01	1,266273e-02	Clássico
		800	5,499714e-03	4,729159e-01	2,862587e-05	Clássico
		1600	4,112164e-03	9,749134e-01	3,37382e-06	Clássico
LLG	37	50	3,522292e+04	4,022348e-01	1,383432e-02	Clássico
		100	3,942001e+04	3,910546e-02	1,455192e-11	Clássico
		200	3,919392e+04	2,659343e-02	1,455192e-11	Clássico
		400	5,161017e+04	7,585391e-02	7,172114e-07	Clássico
		800	9,82697e+04	5,062074e-01	7,548193e-06	Clássico
		1600	1,296791e+05	1,962315e-01	8,045482e-09	Clássico
LLN	83	50	1,612239e+05	2,39748e-03	5,463819e-15	Clássico
		100	9,234484e+05	4,932593e-10	3,73217e-15	Clássico
		200	6,413193e+05	1,7763e-05	2,549525e-15	Clássico
		400	9,548944e+05	2,23066e-11	3,73217e-15	Clássico
		800	3,700654e+05	4,233493e-05	7,999532e-15	Clássico
		1600	5,171036e+05	1,918828e-02	5,463819e-15	Clássico
LNG ou NLG	75	50	-5,342137e-01	8,181601e-08	9,561727e-13	Correto
		100	-6,052963e-01	6,475187e-07	4,34591e-12	Correto
		200	-7,240493e-01	4,955815e-08	1,032442e-12	Correto
		400	-8,191863e-01	2,287429e-04	5,610203e-14	Correto
		800	-9,521638e-01	1,895438e-05	5,38814e-14	Correto
		1600	-9,249994e-01	7,90651e-04	1,032442e-12	Correto
LNN ou NLN	191	50	-2,704842e-01	6,112777e-11	7,358434e-15	Correto
		100	-3,099307e-01	3,957945e-13	1,12091e-15	Correto
		200	-2,908496e-01	1,240119e-13	9,201678e-17	Correto
		400	-4,211137e-01	2,997602e-15	1,725317e-18	Correto
		800	-5,025811e-01	4,440892e-16	1,895008e-19	Correto
		1600	-5,038202e-01	0	2,770104e-19	Correto
LNA ou NLA	266	50	-3,44844e-01	0	2,826894e-26	Correto
		100	-3,932105e-01	0	2,34986e-26	Correto
		200	-4,129924e-01	0	5,811765e-28	Correto
		400	-5,333522e-01	0	8,90339e-31	Correto
		800	-6,293432e-01	0	8,879755e-32	Correto
		1600	-6,225737e-01	0	2,552398e-30	Correto
AAN	274	50	4,88377e+04	0	7,903233e-06	Clássico
		100	2,797305e+05	0	8,48403e-06	Clássico
		200	1,942681e+05	0	5,929988e-06	Clássico
		400	2,89256e+05	0	2,242147e-05	Clássico
		800	1,120997e+05	0	6,049349e-05	Clássico
		1600	1,566405e+05	0	3,246653e-05	Clássico
AAG	753	50	1,730703e+03	0	9,559718e-01	Equivalentes
		100	1,936918e+03	0	6,434353e-01	Equivalentes
		200	1,925798e+03	0	5,936831e-01	Equivalentes
		400	2,53588e+03	0	8,862952e-01	Equivalentes
		800	4,828567e+03	0	3,118549e-01	Equivalentes
		1600	6,371927e+03	0	9,270181e-02	Equivalentes
AAA	1027	50	1,429868e+04	0	2,598958e-01	Equivalentes
		100	7,605127e+04	0	5,907136e-01	Equivalentes
		200	5,324204e+04	0	5,442105e-01	Equivalentes
		400	7,903181e+04	0	9,490397e-01	Equivalentes
		800	3,344814e+04	0	7,091643e-01	Equivalentes
		1600	4,646306e+04	0	3,159928e-01	Equivalentes

De maneira geral, os resultados da Tabela 5.6 mostram a robustez do modelo Clássico para a maioria das situações apresentadas. Bem como, a importância de um modelo mais

sofisticado, baseado em cópula, para os casos em que as distribuições dos erros são diferentes (isto é, log-normal ou normal). Assim, existe vantagem em estudar a distribuição dos erros dos modelos para construção do estimador combinado com marginais distintas. E quando as marginais são iguais parece ser mais interessante utilizar o modelo Clássico ao invés de um modelo mais sofisticado. Para a situação AAA onde é levando em consideração toda a amostra o teste aponta que os modelos são equivalentes.

A Tabela 5.7 apresenta os resultados obtidos na comparação entre o estimador combinado Correto e o método SA. Os resultados alcançados neste experimento mostram a robustez do estimador combinado Correto, este apresentando melhor desempenho para todas as situações em relação ao SA.

Tabela 5.7: Sumário do teste comparativo entre o estimador combinado Correto e o método SA.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
NNG	641	50	-9,976922e-01	1,445001e-08	1,219068e-106	Correto
		100	-9,976813e-01	1,348504e-09	1,219068e-106	Correto
		200	-9,976806e-01	1,887379e-15	1,219068e-106	Correto
		400	-9,977058e-01	0	1,219068e-106	Correto
		800	-9,976494e-01	0	1,219068e-106	Correto
		1600	-9,968147e-01	0	1,219068e-106	Correto
LLG	37	50	-9,997267e-01	5,515292e-01	5,707076e-115	Correto
		100	-9,997503e-01	2,033883e-01	9,742772e-124	Correto
		200	-9,997218e-01	1,230578e-01	1,725708e-124	Correto
		400	-9,996683e-01	1,690072e-01	2,218457e-128	Correto
		800	-9,992777e-01	2,77593e-01	4,796294e-113	Correto
		1600	-9,990643e-01	4,126628e-01	1,805801e-114	Correto
LLN	83	50	-9,989218e-01	3,414416e-03	2,549525e-15	Correto
		100	-9,913086e-01	1,053302e-11	2,549525e-15	Correto
		200	-9,951258e-01	9,960443e-05	2,549525e-15	Correto
		400	-9,924205e-01	2,192579e-12	2,549525e-15	Correto
		800	-9,973315e-01	1,63417e-06	2,549525e-15	Correto
		1600	-9,956878e-01	1,359989e-02	2,549525e-15	Correto
LNG ou NLG	75	50	-9,986059e-01	3,341714e-01	7,857503e-224	Correto
		100	-9,853963e-01	1,413168e-01	5,642883e-76	Correto
		200	-9,983168e-01	6,471666e-01	2,640446e-202	Correto
		400	-9,941575e-01	9,804422e-03	5,38814e-14	Correto
		800	-9,940215e-01	8,286609e-09	5,38814e-14	Correto
		1600	-9,908111e-01	1,308953e-13	5,38814e-14	Correto
LNN ou NLN	191	50	-9,983127e-01	0	4,327318e-33	Correto
		100	-9,643044e-01	0	8,666033e-32	Correto
		200	-8,809681e-01	0	1,557943e-30	Correto
		400	-8,160397e-01	0	4,713855e-28	Correto
		800	-5,275316e-01	0	1,117215e-24	Correto
		1600	-4,747871e-01	0	9,090242e-27	Correto
LNA ou NLA	266	50	-9,983954e-01	0	2,237277e-45	Correto
		100	-9,702513e-01	0	4,484476e-44	Correto
		200	-9,140552e-01	0	8,315181e-43	Correto
		400	-8,662609e-01	0	2,757497e-40	Correto
		800	-6,590607e-01	0	9,051819e-37	Correto
		1600	-6,202826e-01	0	5,346108e-39	Correto

Continua na próxima página.

Tabela 5.7 – Continuação a partir da página anterior.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
AAN	274	50	-9,984972e-01	0	1,097486e-46	Correto
		100	-9,724845e-01	0	2,199983e-45	Correto
		200	-9,155487e-01	0	4,088673e-44	Correto
		400	-8,694689e-01	0	1,377876e-41	Correto
		800	-6,698433e-01	0	4,592678e-38	Correto
		1600	-6,325782e-01	0	2,647414e-40	Correto
AAG	753	50	-9,978832e-01	1,495414e-10	6,466784e-125	Correto
		100	-9,965593e-01	0	6,466784e-125	Correto
		200	-9,978442e-01	0	6,466784e-125	Correto
		400	-9,974488e-01	0	6,466784e-125	Correto
		800	-9,973681e-01	0	6,466784e-125	Correto
		1600	-9,963273e-01	0	6,466784e-125	Correto
AAA	1027	50	-9,98047e-01	0	1,316857e-169	Correto
		100	-9,901362e-01	0	2,64366e-168	Correto
		200	-9,758881e-01	0	5,200232e-167	Correto
		400	-9,633042e-01	0	2,005783e-164	Correto
		800	-9,099856e-01	0	1,272203e-160	Correto
		1600	-8,992803e-01	0	3,977399e-163	Correto

A Tabela 5.8 mostra os resultados alcançados através da comparação entre o estimador combinado Clássico e o método SA. Cabe ressaltar que embora o SA seja considerado simples de implementar, este método pode não ser a melhor opção. Neste contexto, esta tabela é produto da análise estatística realizada entre o estimador combinado Clássico e o método SA, que mostra a superioridade do modelo Clássico sobre o SA.

Tabela 5.8: Sumário do teste comparativo entre o estimador combinado Clássico e o método SA.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
NNG	641	50	5,665208e+02	1,481106e-03	1,219068e-106	Clássico
		100	4,870775e+02	1,401665e-01	1,920725e-298	Clássico
		200	4,610131e+02	1,313739e-01	0	Clássico
		400	4,546338e+02	5,739065e-03	1,219068e-106	Clássico
		800	4,471312e+02	1,665335e-15	1,219068e-106	Clássico
		1600	4,37654e+02	0	1,219068e-106	Clássico
LLG	37	50	1,480303e+08	9,259242e-01	1,313911e-20	Clássico
		100	1,592232e+08	1,600703e-01	3,402387e-22	Clássico
		200	1,534423e+08	2,990009e-01	1,137924e-17	Clássico
		400	1,541636e+08	1,472693e-01	2,880505e-23	Clássico
		800	1,392755e+08	9,239975e-01	6,667966e-25	Clássico
		1600	1,371943e+08	7,804424e-01	1,665578e-30	Clássico
LLN	83	50	1,452251e+08	2,393037e-01	1,457166e-40	Clássico
		100	1,390058e+08	4,655464e-01	5,436763e-42	Clássico
		200	1,347793e+08	2,879137e-01	9,216244e-45	Clássico
		400	1,387705e+08	1,949803e-01	1,826489e-52	Clássico
		800	1,443291e+08	8,298695e-01	6,944107e-56	Clássico
		1600	1,349658e+08	4,852523e-01	1,925078e-53	Clássico
LNG ou NLG	75	50	3,575644e+02	1,451466e-07	2,178909e-13	Clássico
		100	2,061706e+02	1,18156e-06	1,250202e-12	Clássico
		200	1,474345e+02	1,530631e-07	4,34591e-12	Clássico
		400	1,071667e+02	4,520249e-04	5,044162e-12	Clássico
		800	2,876726e+01	1,054879e-04	5,433402e-12	Clássico
		1600	1,990523e+01	2,224301e-03	6,839831e-11	Clássico

Continua na próxima página.

Tabela 5.8 – Continuação a partir da página anterior.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
LNN ou NLN	191	50	5,560788e+07	0	3,263828e-32	Clássico
		100	5,566566e+07	0	9,361668e-32	Clássico
		200	5,311541e+07	0	1,319004e-30	Clássico
		400	5,613184e+07	0	5,276678e-29	Clássico
		800	5,526507e+07	0	4,487117e-29	Clássico
		1600	5,556474e+07	0	2,921591e-27	Clássico
LNA ou NLA	266	50	3,992906e+07	0	5,603416e-44	Clássico
		100	3,997051e+07	0	7,208939e-43	Clássico
		200	3,81393e+07	0	3,459423e-41	Clássico
		400	4,030523e+07	0	2,704795e-39	Clássico
		800	3,968282e+07	0	2,286226e-39	Clássico
		1600	3,9898e+07	0	1,730742e-36	Clássico
AAN	274	50	8,275469e+07	0	4,503253e-46	Clássico
		100	8,091104e+07	0	9,441599e-46	Clássico
		200	7,785301e+07	0	6,134809e-45	Clássico
		400	8,116472e+07	0	8,613986e-44	Clássico
		800	8,224432e+07	0	7,663562e-44	Clássico
		1600	7,961689e+07	0	1,598726e-42	Clássico
AAG	753	50	7,27425e+06	0	7,435432e-125	Clássico
		100	7,82415e+06	0	8,896549e-125	Clássico
		200	7,540069e+06	0	1,014714e-124	Clássico
		400	7,5755e+06	0	1,031019e-124	Clássico
		800	6,843933e+06	0	1,03927e-124	Clássico
		1600	6,741662e+06	0	1,373561e-124	Clássico
AAA	1027	50	2,741217e+07	0	3,046009e-169	Clássico
		100	2,732348e+07	0	5,978258e-169	Clássico
		200	2,629932e+07	0	1,688037e-168	Clássico
		400	2,720885e+07	0	5,569505e-168	Clássico
		800	2,696049e+07	0	5,316186e-168	Clássico
		1600	2,618452e+07	0	3,444154e-167	Clássico

A Tabela 5.9 apresenta os resultados obtidos por meio da comparação entre os estimadores combinados de Cacoullos e Correto. Conforme é mostrado na tabela, quando pelo menos uma das marginais for normal e a cópula geradora dos erros for a Gumbel-Hougaard, parece ser melhor optar pela utilização do estimador combinado Correto ao invés de Cacoullos.

Tabela 5.9: Sumário do teste comparativo entre o estimador combinado Cacoullos e Correto.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
NNG	641	50	-6,31177e-02	0	2,744327e-18	Correto
		100	-1,067695e-01	0	4,565168e-39	Correto
		200	-1,310133e-01	0	1,48382e-63	Correto
		400	-1,428077e-01	0	1,490683e-89	Correto
		800	-1,448866e-01	0	1,744065e-95	Correto
		1600	-1,552066e-01	0	3,841722e-100	Correto
LLG	37	50	4,95284e+05	1,654141e-10	2,671817e-02	Cacoullos
		100	2,646428e+05	1,804579e-10	2,998894e-02	Cacoullos
		200	5,657813e+05	2,645869e-10	1,805728e-02	Cacoullos
		400	6,291731e+05	4,503881e-10	6,860146e-03	Cacoullos
		800	6,664634e+05	4,13249e-10	2,863003e-02	Cacoullos
		1600	7,469809e+05	1,20962e-09	6,860146e-03	Cacoullos

Continua na próxima página.

Tabela 5.9 – Continuação a partir da página anterior.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
LLN	83	50	3,867537e+06	0	5,184099e-02	Equivalentes
		100	6,857343e+06	0	9,05904e-03	Cacoullos
		200	6,474513e+06	0	1,349569e-02	Cacoullos
		400	8,389193e+06	0	4,622783e-03	Cacoullos
		800	9,646193e+06	0	2,196281e-03	Cacoullos
		1600	3,472849e+07	0	5,81328e-03	Cacoullos
LNG ou NLG	75	50	-2,676673e-01	2,45818e-02	5,31443e-06	Correto
		100	-2,4579e-01	1,891733e-05	3,002956e-07	Correto
		200	-1,989513e-01	1,856607e-05	7,23755e-05	Correto
		400	2,399513e-02	2,109424e-15	3,850561e-10	Cacoullos
		800	-2,547299e-02	1,110223e-15	1,890645e-11	Correto
		1600	-6,750397e-02	1,110223e-13	1,557302e-09	Correto
LNN ou NLN	191	50	2,052739e+01	0	3,277621e-02	Cacoullos
		100	1,860659e+02	0	3,212958e-06	Cacoullos
		200	1,719235e+01	0	6,313735e-10	Cacoullos
		400	1,241764e+01	0	6,100976e-10	Cacoullos
		800	3,485058e+01	0	3,800356e-14	Cacoullos
		1600	1,643954e+01	0	4,846967e-09	Cacoullos
LNA ou NLA	266	50	1,466412e+01	0	3,385242e-05	Cacoullos
		100	1,335344e+02	0	2,709868e-11	Cacoullos
		200	1,228878e+01	0	1,196165e-12	Cacoullos
		400	8,92319e+00	0	1,020433e-17	Cacoullos
		800	2,501711e+01	0	1,315192e-23	Cacoullos
		1600	1,17853e+01	0	1,716112e-16	Cacoullos
AAN	274	50	1,171568e+06	0	1,571707e-03	Cacoullos
		100	2,077354e+06	0	1,699745e-08	Cacoullos
		200	1,96127e+06	0	3,312255e-11	Cacoullos
		400	2,54126e+06	0	5,752777e-12	Cacoullos
		800	2,922046e+06	0	9,724255e-16	Cacoullos
		1600	1,051996e+07	0	6,657627e-11	Cacoullos
AAG	753	50	2,433658e+04	0	2,784553e-24	Cacoullos
		100	1,300358e+04	0	1,769557e-46	Cacoullos
		200	2,780054e+04	0	2,522357e-65	Cacoullos
		400	3,091542e+04	0	5,036239e-98	Cacoullos
		800	3,274774e+04	0	4,005814e-104	Cacoullos
		1600	3,67041e+04	0	1,83569e-107	Cacoullos
AAA	1027	50	3,304138e+05	0	2,523779e-25	Cacoullos
		100	5,63765e+05	0	2,778095e-52	Cacoullos
		200	5,436433e+05	0	1,133977e-72	Cacoullos
		400	7,006667e+05	0	1,926756e-101	Cacoullos
		800	8,036025e+05	0	1,878084e-113	Cacoullos
		1600	2,833599e+06	0	1,169498e-108	Cacoullos

A comparação realizada entre o estimador combinado de Cacoullos e o método SA ilustrada na Tabela 5.10 mostra a superioridade do estimador construído via cópula de Cacoullos em relação ao SA. Nesta comparação, o estimador combinado mostra-se melhor em todas as situações, isto ocorre possivelmente pelo fato do estimador combinado ser uma abordagem mais sofisticada, sendo capaz de corrigir o erro das previsões, enquanto o método SA apresentado neste trabalho é uma abordagem simples que mostra não apresenta condições de corrigir os erros das previsões.

Tabela 5.10: Sumário do teste comparativo entre o estimador combinado Cacoullos e o método SA.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
NNG	641	50	4,973836e+02	3,312729e-04	1,965689e-106	Cacoullos
		100	4,238276e+02	2,016802e-08	1,911241e-106	Cacoullos
		200	3,950153e+02	1,598721e-13	1,993491e-106	Cacoullos
		400	3,861723e+02	0	2,251483e-106	Cacoullos
		800	3,809736e+02	0	2,483986e-106	Cacoullos
		1600	3,710793e+02	0	2,664587e-106	Cacoullos
LLG	37	50	8,371832e+08	1,70483e-10	1,455192e-11	Cacoullos
		100	1,404222e+08	2,064e-10	1,455192e-11	Cacoullos
		200	6,7774e+08	2,467678e-10	1,455192e-11	Cacoullos
		400	1,089393e+09	2,821476e-10	1,455192e-11	Cacoullos
		800	5,92405e+08	3,01063e-10	1,455192e-11	Cacoullos
		1600	2,76423e+08	3,640326e-10	1,455192e-11	Cacoullos
LLN	83	50	6,441463e+09	0	2,84408e-15	Cacoullos
		100	3,296711e+10	0	2,949523e-15	Cacoullos
		200	2,845792e+09	0	2,549525e-15	Cacoullos
		400	1,138947e+10	0	2,84408e-15	Cacoullos
		800	1,041464e+10	0	3,172087e-15	Cacoullos
		1600	7,175308e+10	0	3,537275e-15	Cacoullos
LNG ou NLG	75	50	1,952936e+03	4,445507e-06	2,013529e-13	Cacoullos
		100	1,444846e+03	7,516051e-05	2,266532e-13	Cacoullos
		200	9,768401e+02	6,476875e-04	2,013529e-13	Cacoullos
		400	1,065601e+03	5,070376e-04	2,094616e-13	Cacoullos
		800	7,013568e+02	1,723882e-04	1,652138e-13	Cacoullos
		1600	5,209629e+02	5,160358e-02	1,451661e-17	Cacoullos
LNN ou NLN	191	50	2,78019e+09	0	6,024276e-33	Cacoullos
		100	1,202768e+09	0	9,651766e-33	Cacoullos
		200	1,355988e+09	0	1,643616e-32	Cacoullos
		400	1,314764e+09	0	1,543993e-32	Cacoullos
		800	5,617149e+09	0	4,185145e-32	Cacoullos
		1600	3,474575e+09	0	4,250629e-32	Cacoullos
LNA ou NLA	266	50	1,996302e+09	0	7,069937e-45	Cacoullos
		100	8,636424e+08	0	1,434257e-44	Cacoullos
		200	9,736611e+08	0	2,624737e-44	Cacoullos
		400	9,440604e+08	0	2,482091e-44	Cacoullos
		800	4,033366e+09	0	6,262999e-44	Cacoullos
		1600	2,494901e+09	0	8,362013e-44	Cacoullos
AAN	274	50	3,889262e+09	0	1,629012e-46	Cacoullos
		100	1,082481e+10	0	2,637088e-46	Cacoullos
		200	1,807279e+09	0	3,039771e-46	Cacoullos
		400	4,366591e+09	0	3,54207e-46	Cacoullos
		800	7,070402e+09	0	1,007765e-45	Cacoullos
		1600	2,415748e+10	0	1,198997e-45	Cacoullos
AAG	753	50	4,113711e+07	0	1,689599e-124	Cacoullos
		100	6,900399e+06	0	1,689599e-124	Cacoullos
		200	3,33024e+07	0	1,730449e-124	Cacoullos
		400	5,35297e+07	0	2,05353e-124	Cacoullos
		800	2,910927e+07	0	2,128424e-124	Cacoullos
		1600	1,358291e+07	0	2,723972e-124	Cacoullos
AAA	1027	50	1,067803e+09	0	4,945369e-169	Cacoullos
		100	2,893082e+09	0	7,290057e-169	Cacoullos
		200	5,065932e+08	0	8,558544e-169	Cacoullos
		400	1,204239e+09	0	1,125737e-168	Cacoullos
		800	1,907702e+09	0	2,401415e-168	Cacoullos
		1600	6,45509e+09	0	3,366362e-168	Cacoullos

Na comparação entre o estimador combinado Cacoullos e o modelo Clássico, aparentemente nos casos em que as marginais são normais parece ser melhor utilizar o modelo Clássico para combinar as previsões, como mostra a Tabela 5.11. No entanto, quando as marginais não são normais aparenta ser melhor usar o estimador combinado Cacoullos.

Tabela 5.11: Sumário do teste comparativo entre o estimador combinado Cacoullos e Clássico.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
NNG	641	50	-6,930953e-02	0	1,40861e-16	Clássico
		100	-1,087513e-01	0	2,588724e-39	Clássico
		200	-1,358013e-01	0	1,075428e-62	Clássico
		400	-1,457885e-01	0	1,641122e-88	Clássico
		800	-1,490224e-01	0	6,841014e-96	Clássico
		1600	-1,584363e-01	0	3,06152e-100	Clássico
LLG	37	50	4,42622e+00	1,674142e-10	8,349889e-08	Cacoullos
		100	-2,208374e-01	1,917076e-10	8,349889e-08	Clássico
		200	3,241928e+00	2,452156e-10	8,349889e-08	Cacoullos
		400	4,438508e+00	2,167601e-10	8,349889e-08	Cacoullos
		800	2,74761e+00	2,591488e-10	8,349889e-08	Cacoullos
		1600	9,556841e-01	2,951207e-10	8,349889e-08	Cacoullos
LLN	83	50	5,089585e+01	0	7,034528e-07	Cacoullos
		100	2,935109e+02	0	5,069633e-07	Cacoullos
		200	2,249587e+01	0	5,069633e-07	Cacoullos
		400	9,472732e+01	0	1,11366e-07	Cacoullos
		800	8,433328e+01	0	7,930236e-08	Cacoullos
		1600	3,158887e+02	0	7,930236e-08	Cacoullos
LNG ou NLG	75	50	6,976141e+03	9,091146e-10	3,977275e-06	Cacoullos
		100	3,981723e+03	1,53856e-10	7,203233e-07	Cacoullos
		200	2,217126e+03	3,297521e-09	1,279102e-09	Cacoullos
		400	1,076896e+03	1,592666e-08	2,738444e-10	Cacoullos
		800	5,661104e+02	9,289235e-08	1,632454e-12	Cacoullos
		1600	4,132163e+02	2,396591e-06	1,829437e-12	Cacoullos
LNN ou NLN	191	50	1,632003e+04	0	5,523001e-06	Cacoullos
		100	3,095116e+03	0	3,778978e-06	Cacoullos
		200	1,153781e+03	0	2,854864e-08	Cacoullos
		400	5,452696e+02	0	9,730572e-11	Cacoullos
		800	3,959597e+02	0	2,875952e-13	Cacoullos
		1600	3,556054e+02	0	1,755688e-13	Cacoullos
LNA ou NLA	266	50	1,368547e+04	0	4,004693e-11	Cacoullos
		100	3,345099e+03	0	1,750807e-11	Cacoullos
		200	1,453597e+03	0	2,93979e-17	Cacoullos
		400	6,951643e+02	0	3,585243e-20	Cacoullos
		800	4,439345e+02	0	2,709335e-25	Cacoullos
		1600	3,718491e+02	0	2,132504e-25	Cacoullos
AAN	274	50	1,139179e+04	0	7,643122e-01	Equivalentes
		100	2,246455e+03	0	4,087223e-01	Equivalentes
		200	8,110927e+02	0	3,282063e-02	Cacoullos
		400	4,087914e+02	0	1,719527e-03	Cacoullos
		800	3,015619e+02	0	2,465561e-05	Cacoullos
		1600	3,435745e+02	0	9,817422e-06	Cacoullos
AAG	753	50	6,949932e+02	0	1,840212e-11	Cacoullos
		100	3,964825e+02	0	7,279042e-27	Cacoullos
		200	2,20873e+02	0	3,537093e-36	Cacoullos
		400	1,073546e+02	0	1,12703e-51	Cacoullos
		800	5,639365e+01	0	2,972295e-51	Cacoullos
		1600	4,106909e+01	0	2,777422e-54	Cacoullos

Continua na próxima página.

Tabela 5.11 – Continuação a partir da página anterior.

Situações	Amostra	Conjunto de treinamento	\overline{EQM}_R	Teste (KS)	Teste comparativo	Melhor modelo
AAA	1027	50	3,54886e+03	0	2,299447e-08	Cacoullos
		100	8,900487e+02	0	3,271266e-17	Cacoullos
		200	3,783416e+02	0	1,138801e-19	Cacoullos
		400	1,877769e+02	0	3,766034e-25	Cacoullos
		800	1,218037e+02	0	9,628096e-22	Cacoullos
		1600	1,217765e+02	0	4,028628e-23	Cacoullos

5.2 Considerações finais

De modo geral, os resultados apresentados, mostram que é mais interessante utilizar o modelo Correto ao invés da abordagem Clássica para os casos em que as distribuições dos erros dos modelos são diferentes. Já para os casos quando as distribuições dos erros são iguais, o modelo Clássico mostra-se melhor. Na comparação entre o estimador combinado de Cacoullos e Correto foi comprovado que nos casos específicos, isto é, quando pelo menos uma das distribuições for normal e o erro for gerado a partir da cópula Gumbel-Hougaard parece ser melhor optar por utilizar o estimador Correto ao invés de Cacoullos, por outro lado, para as demais situações parece ser melhor usar o estimador combinado Cacoullos.

Entre o estimador combinado de Cacoullos e Clássico, foi possível comprovar que especificamente para os casos, em que, as distribuições dos erros são normais é mais interessante utilizar a abordagem Clássica, porém nos demais casos o estimador combinado Cacoullos mostra-se melhor. Nas comparações realizadas com o método SA, mostraram que os estimadores combinados utilizados nesta dissertação foram superiores em todas as situações simuladas. Cabe ressaltar, que isso pode ocorrer pelo fato da combinação SA não ser capaz de corrigir os erros dos modelos.

Capítulo 6

Conclusões

Foi estudado o problema de previsão de séries temporais, descrevendo suas características básicas. Apresentou-se o estado da arte com relação ao tema, detalhando alguns modelos de predição existentes na literatura. Desenvolveram-se algumas abordagens por meio do formalismo matemático de cópulas para aplicar ao problema de previsão de séries temporais. A partir deste estudo, foram desenvolvidas diversas abordagens, implementadas através de linguagem de programação para combinar previsões. Essas abordagens mostraram-se atrativas, uma vez que foram capazes de produzir previsões melhores que os modelos individuais.

Acredita-se que os resultados apresentados nos capítulos anteriores permitem dizer que os objetivos desta dissertação foram em grande parte alcançados. De modo geral, o estimador combinado para os casos apresentados mostrou-se melhor que os modelos individuais, assim, os experimentos computacionais realizados com os estimadores combinados propostos mostraram-se úteis para tratar o problema previsão de séries temporais. Além disso, no Capítulo 5 é apresentado um caso de estudo envolvendo inúmeras séries temporais e modelos de previsão individuais, utilizados para estudar alguns estimadores combinados, com intuito de investigar via análise estatística qual é o melhor estimador combinado.

Nesta dissertação, outra contribuição dar-se por meio de um caso de estudo que propõe através de simulações, analisar se existe vantagem em estudar a distribuição do erro do modelo. Neste sentido, foram comparados dois modelos combinados de previsão de séries temporais, denominados modelo Correto e Clássico, sendo que, o primeiro é construído

com base no estudo acerca da distribuição dos erros e sobre qual processo gerou os erros. Enquanto o segundo não realiza tal estudo prévio, assumindo que as marginais dos erros são normais e os erros foram gerados por um processo multivariado normal.

A técnica de combinação baseada na Equação de Cacoullos mostrou-se promissora, uma vez que, os experimentos mostraram que o estimador combinado via cópula de Cacoullos obteve melhores resultados quando comparados com os modelos de previsão individuais. Nesse sentido, a combinação de modelos de previsão por meio da cópula de Cacoullos pode fornecer valiosas informações para guiar na tomada de decisão, como previsão de chuva, valor de ações da bolsa de valores futuras, por exemplo.

Ainda no Capítulo 5 é ilustrado a utilidade do estimador combinado via cópula de Gumbel-Hougaard, para combinar dois modelos previsão. Através do experimento computacional realizado, é razoável afirmar que o estimador combinado mostra-se como uma alternativa vantajosa para combinação de modelos.

No entanto, deve-se observar que a dissertação apresentada, de modo algum, esgota a possibilidade de melhoramento dos métodos de previsão.

6.1 Contribuições

Com base nos resultados apresentados nesta dissertação é possível responder os questionamentos introduzidos na Seção 1.3:

i A combinação de preditores através do formalismo de cópulas é capaz de apresentar previsões estatisticamente eficientes e acuradas?

Os resultados apresentados no Capítulo 5 mostram que o formalismo de cópula é capaz de produzir previsões combinadas eficientes. Neste sentido, cópulas pode ser considerada uma alternativa promissora para combinação de modelos de previsão com finalidade de realizar previsões melhoradas.

ii O preditor combinado mostra-se mais eficiente e acurado em relação aos modelos individuais usados no procedimento de combinação?

Os estimadores combinados construídos através de cópulas sugerem melhor desempenho em relação aos preditores utilizados na combinação. Porém, o estimador combinado parece apresentar uma variância um pouco maior em relação a pelo menos um dos modelos individuais.

- iii **Construir um preditor combinado por meio da função multivariada de Cacoullos [6] é capaz de produzir previsões eficientes e acuradas? E ainda, as previsões obtidas são melhores em comparação aos preditores individuais?**

Os resultados alcançados com o estimador combinado ajustado com a função multivariada de Cacoullos, mostrou-se capaz de realizar previsões eficientes e acuradas. A cópula de Cacoullos apresentou ser muito mais acurada em relação aos modelos individuais e SA.

- iv **O ajuste do parâmetro de suavização da função multivariada de Cacoullos [6] conduz a melhores previsões?**

De fato, o ajuste do parâmetro de suavização da cópula de Cacoullos possibilitou uma significativa melhora na acurácia do estimador combinado para os casos que a variação do parâmetro era entre 0,01 e 1.

- v **Existe vantagem em estudar os erros cometidos pelos modelos?**

Constatou-se que o modelo Clássico no qual negligência-se a distribuição do erro mostrou-se equivalente ao modelo Correto para os casos em geral. Porém, tal análise, ilustrou que existe vantagem em estudar a distribuição dos erros dos modelos, uma vez que, nos casos mais específicos, isto é, quando as marginais dos modelos são distintas, os resultados mostram que um modelo mais sofisticado que estude a distribuição dos erros se faz necessário.

6.2 Limitações

Embora a abordagem proposta via cópulas tenha alcançado resultados expressivos, inclusive comparado com outros métodos encontrados na literatura, como descrito na Seção 2.2, ainda existem vários pontos que precisam ser estudados cuidadosamente e aprofundados.

Um destes pontos é descobrir se outra cópula como de Clayton, por exemplo, é mais robusta comparada com o modelo Clássico. Inúmeras análises estatísticas são necessárias para encontrar quais são os quadros que podem ser mais úteis utilizando uma determinada cópula. Maiores estudos são precisos uma vez que estes são desconhecidos.

Outra questão muito importante a ser considerada é o tamanho da amostra utilizado no teste estatístico. Atualmente, a amostra possui cerca de 6 mil séries temporais de diversos tamanhos. Embora a quantidade de séries temporais pareça ser bastante grande, pode não ser suficientemente apropriada para o estudo em questão.

Destaca-se ainda a desvantagem da alta complexidade em implementação da abordagem proposta em relação as combinações lineares como SA, por exemplo. Outra desvantagem consiste no tempo computacional gasto para combinar todas as previsões dos modelos individuais, tal tempo é muito superior quando comparado com a combinação linear SA. Neste sentido, enquanto a combinação linear gasta aproximadamente 30 minutos para realizar todas as combinações vista na Seção 5.1 o estimador combinado via cópula (Gumbel-Hougaard ou Normal) consome cerca de 12 horas e quando a combinação dar-se pela cópula de Cacoullos o consume computacional é ainda maior levando aproximadamente 288 horas.

6.3 Trabalhos futuros

Como trabalhos futuros, aponta-se a necessidade de explorar o potencial de outros tipos de cópulas não mencionadas neste trabalho. Pode-se ainda, replicar os experimentos realizados nos capítulos anteriores, considerando a utilização de séries temporais do mundo real, assim como, os modelos de previsão.

Ainda como trabalhos futuros, sugere-se a implementação de um método de estimação, para estimar o parâmetro de suavização da cópula de Cacoullos. A ideia é estimar o melhor valor para o parâmetro de suavização, bem como, maximizar o desempenho da função de cópulas, e dessa maneira, melhorar a acurácia e eficiência do estimador combinado.

Vislumbra-se, ainda a possibilidade de desenvolver uma técnica de seleção de modelos de predição de séries temporais, para ser aplicada antes da abordagem de combinação. Um

vez que a incorporação de um modelo ruim na combinação pode influencia diretamente na acurácia e eficiência do estimador combinado.

6.4 Trabalhos publicados

Com esta dissertação foram gerados os seguintes trabalhos:

1. Estudando Fundamentos de Modelos de Cópulas [38];
2. Combination of Biased Artificial Neural Network Forecasters [41];
3. Combining Time Series Forecasting Models Via Gumbel-Hougaard Copulas [39].

Referências Bibliográficas

- [1] Alessandra Amendola and Giuseppe Storti. A gmm procedure for combining volatility forecasts. *Computational Statistics & Data Analysis*, 52:3047 – 3060, 2008.
- [2] Torben Gustav Andersen, Richard A. Davis, Jens-Peter Kreiß, and Thomas V. Mikosch. *Handbook of financial time series*. Mathematics and Statistics. Springer-Verlag Berlin Heidelberg, 2009.
- [3] Kazim Azam. Bayesian inference for a semi-parametric copula-based markov chain. Disponível em: <
http://www2.warwick.ac.uk/fac/soc/economics/staff/phd_students/azam/job_market_paper.pdf > Acesso em: 17 mar. 2013.
- [4] Pedro Alberto Barbetta, Marcelo Menezes Reis, and Antonio Cezar Bornia. *Estatística para cursos de engenharia e informática*. Atlas, 2010.
- [5] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time series analysis: forecasting and control*. Forecasting and Control Series. Prentice Hall, 3rd edition, 1994.
- [6] Theophilos Cacoullos. Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18:179–189, 1964.
- [7] Sidia M. Callegari-Jacques. *Bioestatística: princípios e aplicações*. Artmed, 2007.
- [8] Chris Chatfield. *Time-Series Forecasting*. Statistics (Chapman and Hall/CRC). Chapman and Hall/CRC, 1 edition, 2000.
- [9] Robert T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5:559–583, 1989.

- [10] José Aécio Corrêa de Araújo. *Tipificação dos padrões de chuvas intensas em função do posicionamento da intensidade e fatores determinantes para o município de barreiros do estado de Pernambuco*. PhD thesis, Universidade Federal Rural de Pernambuco, 2013.
- [11] Rodrigo Moreira de Assis and Márcio Poletti Laurini. Funções de cópula na precificação de opções. Disponível em: < <http://www.insper.edu.br/working-papers/> > Acesso em: 20 out. 2012.
- [12] Lilian M. de Menezes, Derek W. Bunn, and James W. Taylor. Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, 120:190–204, 2000.
- [13] Luzia Vidal de Souza. *Programação genética e combinação de preditores para previsão de séries temporais*. PhD thesis, Universidade Federal do Paraná, 2006.
- [14] P. Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Acad Roy Belg Bul Cl Sci (5)*, 65:274–292, 1979.
- [15] Rosario Dell'Aquila and Elvezio Ronchetti. Stock and bond return predictability: the discrimination power of model selection criteria. *Computational Statistics & Data Analysis*, 50:1478–1495, March 2006.
- [16] Israel José dos Santos Felipe. Aplicação de modelos arima em séries de preços de soja no norte do paraná. *Tekhne e Logos*, 3:1–17, 2012.
- [17] Luiz Gustavo L. Fernandes, Philippe O. A. Navaux, and Marcelo Savino Portugal. Previsão de séries de tempo: redes neurais artificial e modelos estruturais. In *Encontro Brasileiro de Econometria*, pages 1–18, 1995.
- [18] Tiago A. E. Ferreira. *Uma nova metodologia Híbrida inteligente para a Previsão de séries temporais*. PhD thesis, Universidade Federal de Pernambuco, 2006.
- [19] Tiago A. E. Ferreira, Germano C. Vasconcelos, and Paulo J. L. Adeodato. A new intelligent system methodology for time series forecasting with artificial neural networks. *Neural Processing Letters*, 28:113–129, 2008.

- [20] Carlos Andres Ferreo. Algoritmo knn para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia. Master's thesis, Universidade de São Paulo, 2009.
- [21] Paulo R. Firmino and Enrique López Droguett. Model uncertainty via mixed bayesian networks. In *10th International Probabilistic Safety Assessment & Management Conference.*, 2010.
- [22] Paulo R. A. Firmino, Paulo S.G. Mattos Neto, and Tiago A. E. Ferreira. Correcting and combining time series forecasters. *Neural Networks*, 50:1–11, 2014.
- [23] Christian Genest and Anne catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.
- [24] Simon Haykin. *Redes Neurais: Princípio e prática*. Bookman, 2001.
- [25] Tim Hill, Marcus O'Connor, and William Remus. Neural network models for time series forecasts. *Management science*, 42:1082–1092, 1996.
- [26] Satish G. Iyengar and Thyagaraju Damarla. A parametric copula based framework for multimodal signal processing. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 1893–1896. Taipei, 2009.
- [27] Ashu Jain and Avadhnam Madhav Kumar. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, 7:585–592, 2007.
- [28] Dae Il Jeong and Young-Oh Kim. Combining single-value streamflow forecasts - a review and guidelines for selecting techniques. *Journal of Hydrology*, 377:284–299, oct 2009. Provided by the SAO/NASA Astrophysics Data System.
- [29] Harry Joe and James J. Xu. The estimation method of inference functions for margins for multivariate models. Disponível em: < <http://www.stat.ubc.ca/harry/ifm.pdf> > acesso em: 20 dez. 2013, 2013.
- [30] Shih-Chieh Kao and Rao S. Govindaraju. A copula-based joint deficit index for droughts. *Journal of Hydrology*, 380:121–134, 2010.

-
- [31] Daijin Kim and Chulhyun Kim. Forecasting time series with genetic fuzzy predictor ensemble. *Fuzzy Systems, IEEE Transactions*, 5:523–535, 1997.
- [32] Dora Marina Branco Leal. Aplicação de cópulas ao ramo vida: Risco de resgate e risco de taxa de juro. Master’s thesis, Universidade Técnica de Lisboa, 2010.
- [33] Isabela R. A. Lima and Paulo Renato Alves Firmino. Os reflexos da dependência entre duas variáveis sobre a função de cópulas empíricas. In *Jornada de Ensino, Pesquisa e Extensão da UFRPE*, 2012.
- [34] Thomas Lux and Leonardo Morales-Arias. Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. *Computational Statistics and Data Analysis*, 54:2676 – 2692, 2010.
- [35] João Francisco Galera Monico, Aluir Porfírio Dal Póz, Maurício Galo, Marcelo Carvalho dos Santos, and Leonardo Castro de Oliveira. Acurácia e precisão: revendo os conceitos de forma acurada. *Boletim de Ciências Geodésicas*, 15:1–15, 2009.
- [36] Roger B. Nelsen. *An Introduction to Copulas*, volume 139. Springer, 2 edition, 2006.
- [37] S. P. Neuman. Maximum likelihood bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment*, 17:291–305, 2003.
- [38] Ricardo T. A. Oliveira, Paulo R. A. Firmino, and Tiago A. E. Ferreira. Estudando fundamentos de modelos de cópulas. In *Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria e o Simpósio de Estatística Aplicada à Experimentação Agronômica*, pages 1–21, 2013.
- [39] Ricardo T. A. Oliveira, Thaíze F. Oliveira, Paulo R. Firmino, and Tiago A. E. Ferreira. Combining time series forecasting models via gumbel-hougaard copulas. In *1st BRICS Countries Conference on Computational Intelligence*, pages 1–6, 2013.
- [40] Thaíze F. Oliveira, Paulo Renato Firmino, and Tiago A. E. Ferreira. Estudo da incerteza de modelos via cópula de gumbel. In *58º RBras Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria*, pages 1–13, 2013.

-
- [41] Thaíze. F. Oliveira, Ricardo T. A. Oliveira, P. R. Firmino, Paulo S.G. Mattos Neto, and T. A. E. Ferreira. Combination of biased artificial neural network forecasters. In *1st BRICS Countries Conference on Computational Intelligence*, pages 1–6, 2013.
- [42] Valentyn Panchenko. Goodness-of-fit test for copulas. *Physica A: Statistical Mechanics and its Applications*, 355:176–182, 2005.
- [43] Emanuel Parzen. On estimation of a probability density function and mode. *Annal of Mathematical Statistics*, 33:1065–1076, 1962.
- [44] Benjamin Renard and Michel Lang. Use of a gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, 30:897–912, 2007.
- [45] Maria I. S. Silva, Ednaldo C. Guimarães, and Marcelo Tavares. Previsão da temperatura média mensal de uberlândia, mg, com modelos de séries temporais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 12:480–485, 2008.
- [46] Richard Simard and Pierre L’Ecuyer. Computing the two-sided kolmogorov-smirnov distribution. *Journal of Statistical Software*, 39:1–18, 2011.
- [47] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *l’Institut de Statistique de L’Université de Paris*, 8:229–231, 1959.
- [48] Donald F. Specht. *Generation of polynomial discriminant functions for pattern recognition*. PhD thesis, Stanford University, 1966.
- [49] Donald F. Specht. Probabilistic neural networks. *Neural Networks*, 3:109–118, 1990.
- [50] Zaiyong Tang, Chrys de Almeida, and Paul A. Fishwick. Time series forecasting using neural networks vs. box-jenkins methodology. *Simulation*, 57:303–310, 1991.
- [51] Team. *R: A Language and Environment for Statistical Computing*, 2012.
- [52] Kenneth F. Wallis. Combining forecasts: forty years later. *Applied Financial Economics*, 21:33–41, 2011.

-
- [53] Liane Werner and José Luis Duarte Ribeiro. Previsão de demanda: uma aplicação dos modelos box-jenkins na área de assistência técnica de computadores pessoais. *Gestão & Produção*, 10:47–67, 2003.
- [54] Jorg D. Wichard and Maciej Ogorzalek. Time series prediction with ensemble models. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1625 – 1630, 2004.
- [55] Jun Yan. Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21:1–18, 2007.
- [56] G. Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.
- [57] Hui Zou and Yuhong Yang. Combining time series models for forecasting. *International Journal for Forecasting*, 20:69–84, 2004.

Apêndice A

Cálculos para obter a equação de estimação dos pesos da combinação linear

$$f^*(y_{t,1}, y_{t,2} | \mathbf{u}_t, \Sigma^{-1}) = \frac{1}{(2\pi)^{\frac{k}{2}} \sqrt{\det(\Sigma)}} \exp \left[-\frac{1}{2} (\mathbf{y}_t - \mathbf{u}_t)^T \Sigma^{-1} (\mathbf{y}_t - \mathbf{u}_t) \right] \quad (\text{A.1})$$

onde, \mathbf{y}_t é o vetor de variáveis aleatórias, podendo ser visto neste caso como um modelo de previsão, \mathbf{u}_t é a série temporal de interesse e $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ é a matriz de covariância entre modelos.

$$f^*(y_{t,1}, y_{t,2} | \mathbf{u}_t, \Sigma^{-1}) = c^* \times \exp^{g(\mathbf{u}_t)} \quad (\text{A.2})$$

onde,

$$c^* = \frac{1}{(2\pi)^{\frac{k}{2}} \sqrt{\det(\Sigma)}} \text{ e } g(\mathbf{u}_t) = -\frac{1}{2} (\mathbf{y}_t - \mathbf{u}_t)^T \Sigma^{-1} (\mathbf{y}_t - \mathbf{u}_t) \quad (\text{A.3})$$

$$\frac{df^*}{du} = c^* \times \exp^{g(\mathbf{u}_t)} \times \frac{dg(\mathbf{u}_t)}{du} = 0 \quad (\text{A.4})$$

Logo,

$$\frac{dg(\mathbf{u}_t)}{du} = 0 \quad (\text{A.5})$$

$$\frac{dg(\mathbf{u}_t)}{du} = \frac{d}{du} \left\{ -\frac{1}{2} \begin{bmatrix} y_{t,1} - \mathbf{u}_t & y_{t,2} - \mathbf{u}_t \end{bmatrix} \begin{bmatrix} a^* & b^* \\ c^* & d^* \end{bmatrix} \begin{bmatrix} y_{t,1} - \mathbf{u}_t \\ y_{t,2} - \mathbf{u}_t \end{bmatrix} \right\} \quad (\text{A.6})$$

$$= \frac{d}{du} \left\{ -\frac{1}{2} \begin{bmatrix} a^*(y_{t,1} - \mathbf{u}_t) + c^*(y_{t,2} - \mathbf{u}_t) & b^*(y_{t,1} - \mathbf{u}_t) + d^*(y_{t,2} - \mathbf{u}_t) \end{bmatrix} \begin{bmatrix} y_{t,1} - \mathbf{u}_t \\ y_{t,2} - \mathbf{u}_t \end{bmatrix} \right\} \quad (\text{A.7})$$

$$= \frac{d}{du} \left\{ -\frac{1}{2} \left[a^*(y_{t,1} - \mathbf{u}_t)^2 + c^*(y_{t,1} - \mathbf{u}_t)(y_{t,2} - \mathbf{u}_t) + b^*(y_{t,1} - \mathbf{u}_t)(y_{t,2} - \mathbf{u}_t) + d^*(y_{t,2} - \mathbf{u}_t)^2 \right] \right\} \quad (\text{A.8})$$

$$= \frac{d}{du} \left\{ -\frac{1}{2} \left[a^*(y_{t,1} - \mathbf{u}_t)^2 + (b^* + c^*)[(y_{t,1} - \mathbf{u}_t)(y_{t,2} - \mathbf{u}_t)] + d^*(y_{t,2} - \mathbf{u}_t)^2 \right] \right\} \quad (\text{A.9})$$

$$= \frac{d}{du} \left\{ -\frac{1}{2} \left[a^*(y_{t,1} - \mathbf{u}_t)^2 + (b^* + c^*)[(y_{t,1}y_{t,2} - \mathbf{u}_t y_{t,1} - \mathbf{u}_t y_{t,2} + \mathbf{u}_t^2)] + d^*(y_{t,2} - \mathbf{u}_t)^2 \right] \right\} \quad (\text{A.10})$$

$$\hat{u}_t = \frac{2a^*y_{t,1} + (b^* + c^*)(y_{t,1} + y_{t,2}) + 2d^*y_{t,2}}{2(a^* + b^* + c^* + d^*)} \quad (\text{A.11})$$

$$\hat{u}_t = \left(\frac{2a^* + b^* + c^*}{2(a^* + b^* + c^* + d^*)} \right) y_{t,1} + \left(\frac{b^* + c^* + 2d^*}{2(a^* + b^* + c^* + d^*)} \right) y_{t,2} \quad (\text{A.12})$$

$$\hat{u}_t = w_1 y_{t,1} + w_2 y_{t,2} \quad (\text{A.13})$$

onde,

$$w_1 = \left(\frac{2a^* + b^* + c^*}{2(a^* + b^* + c^* + d^*)} \right) \text{ e } w_2 = \left(\frac{b^* + c^* + 2d^*}{2(a^* + b^* + c^* + d^*)} \right) \quad (\text{A.14})$$

Assim, encontrando a^* , b^* , c^* e d^* , temos:

$$\Sigma^{-1} = \begin{bmatrix} a^* & b^* \\ c^* & d^* \end{bmatrix} \text{ então, } \Sigma^{-1}\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ tomando,}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} s^* & t^* \\ u^* & v^* \end{bmatrix} = \begin{bmatrix} a^* & b^* \\ c^* & d^* \end{bmatrix} \begin{bmatrix} s^* & t^* \\ u^* & v^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a^*s^* + b^*u^* & a^*t^* + b^*v^* \\ c^*s^* + d^*u^* & c^*t^* + d^*v^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(A.15)

Portanto,

$$\begin{aligned} \Sigma^{-1} &= \begin{bmatrix} a^* & b^* \\ c^* & d^* \end{bmatrix} = \begin{bmatrix} \frac{v^*}{t^*u^* - v^*s^*} & \frac{t^*}{t^*u^* - v^*s^*} \\ \frac{u^*}{u^*t^* - v^*s^*} & -\frac{v^*}{t^*u^* - v^*s^*} \end{bmatrix} \times (-1) \\ &= \frac{1}{u^*t^* - v^*s^*} \begin{bmatrix} v^* & -t^* \\ -u^* & s \end{bmatrix} = \frac{1}{\sigma_{21}\sigma_{12} - \sigma_{22}\sigma_{11}} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} \end{aligned} \quad (\text{A.16})$$

Temos,

$$\begin{aligned} w_1 &= \left(\frac{2a^* + b^* + c^*}{2(a^* + b^* + c^* + d^*)} \right) = \frac{\frac{1}{\sigma_{21}\sigma_{12} - \sigma_{22}\sigma_{11}} (2\sigma_{22} - \sigma_{12} - \sigma_{21})}{\frac{2}{\sigma_{21}\sigma_{12} - \sigma_{22}\sigma_{11}} (\sigma_{22} - \sigma_{12} - \sigma_{21} + \sigma_{11})} \\ &= \frac{2\sigma_{22} - \sigma_{12} - \sigma_{21}}{2(\sigma_{22} - \sigma_{12} - \sigma_{21} + \sigma_{11})} = \frac{\sigma_{22} - \sigma_{12}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}} = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \end{aligned} \quad (\text{A.17})$$

e

$$\begin{aligned} w_2 &= \left(\frac{b^* + c^* + 2d^*}{2(a^* + b^* + c^* + d^*)} \right) = \frac{\frac{1}{\sigma_{21}\sigma_{12} - \sigma_{22}\sigma_{11}} (-\sigma_{12} - \sigma_{21} + 2\sigma_{11})}{\frac{2}{\sigma_{21}\sigma_{12} - \sigma_{22}\sigma_{11}} (\sigma_{22} - \sigma_{12} - \sigma_{21} + \sigma_{11})} \\ &= \frac{\sigma_{11} - \sigma_{12}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}} = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \end{aligned} \quad (\text{A.18})$$

Assim,

$$w_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \quad \text{e} \quad w_2 = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

Anexo I

Código-fonte das simulação via cópula de Gumbel-Hougaard

Este anexo apresenta o código-fonte utilizado para obter os resultados apresentados na Seção 5.0.1. Se faz necessário a instalação previa do pacote “copula” para execução do código-fonte presente neste anexo.

```
1 set.seed(0)
2
3 CombiningForecasts = function(mvdc, x1, x2, bounds){
4     l = seq(from=bounds[1], to=bounds[2], length.out=1000)
5     x1 = rep(x1, 1000); x2 = rep(x2, 1000)
6     f1 = dMvdc(cbind(x1-l, x2-l), mvdc)
7     s = which.max(f1)
8     max_l = l[s]
9     max_l
10 }
11
12 generateGraphic = function(data, from, to){
13     min_y = min(autos_data[from:to,])
14     max_y = max(autos_data[from:to,])
15
16     plot_colors <- c("black", "blue", "green", "purple", "red")
17     pch <- c(NA, 22, 8, 2, 20)
18     lty <- c(1, rep(1, 4))
19     lwd <- c(3, rep(1, 4))
20
21     plot(x=(from:to), y=autos_data$u[from:to], type="n", pch=20
22     , col=plot_colors[1], ylim=c(min_y, max_y), axes=FALSE, ann=FALSE, lwd=lwd, lty=1)
23     axis(1, at=from:to, lab=from:to)
24     axis(2, las=1, at=25*-160:max_y)
25     box()
26
27     if(from==1){
28         points(x=c(201:230), y=autos_data$u[201:230], type="l", lty=1, col="black", lwd=lwd+2)
29         lines(x=c(200, 200), c(min_y, max_y), col="black", lwd=1, lty=5)
30     }
31 }
```

```

32     lines(x=(from:to), autos_data$u[from:to], type="l", pch=pch[1], lty=lty[1], col=plot_colors[1], lwd=2)
33     lines(x=(from:to), autos_data$uCombined[from:to], type="o", pch=pch[2], lty=lty[2], col=plot_colors[2])
34     lines(x=(from:to), autos_data$x1[from:to], type="o", pch=pch[3], lty=lty[3], col=plot_colors[3])
35     lines(x=(from:to), autos_data$x2[from:to], type="o", pch=pch[4], lty=lty[4], col=plot_colors[4])
36     lines(x=(from:to), autos_data$SA[from:to], type="o", pch=pch[5], lty=lty[5], col=plot_colors[5])
37
38     title(main="", col.main="red", font.main=4)
39     title(xlab= "Tempo", col.lab=rgb(0,0,0))
40     title(ylab= "S\`{e}ries_␣Temporais", col.lab=rgb(0,0,0))
41
42     legend <- c(expression(u[t]), expression(hat(u)[t]), expression(x[paste(t,1)]), expression(x[paste(t,2)]),
43               expression(SA))
44     legend(x=from+1, y=max_y, legend=legend, cex=1, col=plot_colors, pch=pch, lty=lty)
45 }
46 n = 200
47 m = 30
48
49 u = arima.sim(list(order = c(1,1,1), ar = 0.9, ma=0.9), n = (n+m)); u=u[-1]
50 mvdc = mvdc(copula = gumbelCopula(param=2, dim=2),margins = c("norm", "norm"),paramMargins = list(list(mean= 10, sd=10),
51               list(mean=-20, sd=5)))
52 copulaSample = rMvdc((n+m), mvdc)
53 e1 = copulaSample[,1]
54 e2 = copulaSample[,2]
55 x1 = u+e1; x2 = u+e2;
56
57 min = min(x1,x2,u)
58 max = max(x1,x2,u)
59
60 e1 = x1 - u
61 e2 = x2 - u
62
63 mean_E1=mean(e1)
64 mean_E2=mean(e2)
65
66 sd_E1=sd(e1)
67 sd_E2=sd(e2)
68
69 Fe1=pnorm(e1,mean=mean_E1, sd=sd_E1)
70 Fe2=pnorm(e2,mean=mean_E2, sd=sd_E2)
71
72 mvdc = mvdc(copula = gumbelCopula(param=2, dim=2),margins = c("norm", "norm"), paramMargins = list(list(mean_E1, sd=sd_E1)
73               , list(mean=mean_E2, sd=sd_E2)))
74 F = cbind(Fe1,Fe2)
75 fitIfm = fitCopula(mvdc@copula, F, method="ml")
76 properties_fitIfm = summary(fitIfm)
77 thetaIfm = properties_fitIfm$coefficients[1]
78
79 mvdc = mvdc(copula = gumbelCopula(param=thetaIfm, dim=2),margins = c("norm", "norm"),paramMargins = list(list(mean_E1, sd=
80               sd_E1), list(mean=mean_E2, sd=sd_E2)))
81 uCombined = numeric(length(u))
82 bounds = cbind(min=min, max=max)
83
84 for(t in (n+1):(n+m)){
85     uCombined[t] = CombiningForecasts(mvdc, x1[t], x2[t], bounds)
86 }
87
88 SA = (x1+x2)/2;
89 autos_data = data.frame(u,x1,x2,uCombined,SA)
90 generateGraphic(autos_data,from=201,to=230)
91 }

```