Multivariate Box-Cox symmetric models generated by a normal scale mixture copula

Rodrigo Matheus Rocha de Medeiros

THESIS PRESENTED TO THE
INSTITUTE OF MATHEMATICS AND STATISTICS
OF THE UNIVERSITY OF SÃO PAULO
IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF SCIENCE

Program: Statistics

Advisor: Prof. Ph.D. Silvia Lopes de Paula Ferrari

São Paulo

March, 2024

Multivariate Box-Cox symmetric models generated by a normal scale mixture copula

Rodrigo Matheus Rocha de Medeiros

This version of the thesis includes the corrections and modifications suggested by the Examining Committee during the defense of the original version of the work, which took place on March 13, 2024.

A copy of the original version is available at the Institute of Mathematics and Statistics of the University of São Paulo.

Examining Committee:

Prof. Ph.D. Silvia Lopes de Paula Ferrari (advisor) – IME-USP

Prof. Ph.D. Alexandre Galvão Patriota – IME-USP

Prof. Ph.D. Artur José Lemonte – UFRN

Prof. Ph.D. Marcelo Bourguignon Pereira – UFRN

Prof. Ph.D. Raúl Alejandro Morán-Vásquez – UDEA

Dedicated to my wife, Brendda,
and our son, Pedro Leonardo.

Acknowledgments

I am thankful to God for the patience, determination, and willingness He has given me to complete this work.

To my parents, Isabel and Robson, and my sister, Ruthe, for their support and encouragement. I couldn't have completed this thesis without you.

To my wife, Brendda, for her support, patience, and love, especially when writing the thesis took up most of my time. We went through many important events during this period: our marriage, moving from Natal to São Paulo, the COVID-19 pandemic, moving back to Natal, and the birth of our son.

To my son, Leonardo, for providing additional motivation to complete this work, as he was born just a few days before I finished writing the thesis.

To my advisor, Silvia, for her trust and understanding, which allowed me to complete my doctorate while building my family. Thank you for all the support and availability.

To professors Alejandro Morán Vasquéz, Alexandre Patriota, Artur Lemonte, Júlia Soler, and Marcelo Bourguignon for their comments, corrections, and valuable suggestions that enriched this work.

To my undergraduate and master's advisor, Marcelo Bourguignon, for encouraging and motivating me to pursue an academic career.

To the friends I made during my doctorate, Anny, Diego, Lucas, Paulo, and Javier, who made this period much lighter. In particular, I would like to thank Javier for his loyal friendship, conversations, and motivation to complete this work. To my friends Erika and Lucas for their lasting friendship. To Felipe for hosting me and my wife in São Paulo during my first semester. To all my work colleagues who contributed directly or indirectly.

To Hermínio, for the classes he took with me on basic statistics, which helped me and my wife to stay in São Paulo and contributed to my teaching skills.

To Regiane, for all the availability to solve problems and doubts that arose during my time at IME-USP.

Resumo

Rodrigo Matheus Rocha de Medeiros. **Modelos Box-Cox simétricos multivariados gerados por uma cópula de uma mistura por escala de distribuições normais**. Tese (Doutorado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2024.

A classe das distribuições Box-Cox simétricas oferece um cenário flexível de modelagem para dados contínuos e positivos, alcançando diferentes níveis de assimetria e caudas pesadas. No entanto, essa classe foi pouco explorada na modelagem de dependência. A teoria das cópulas oferece uma abordagem para modelar dependência por meio de uma função que caracteriza as associações entre os elementos de um vetor aleatório com marginais especificadas. Por exemplo, cópulas geradas por uma mistura por escala de distribuições normais possuem propriedades atrativas e descrevem dependência de uma forma similar à distribuição normal multivariada. Esta tese tem como objetivo introduzir uma ampla classe de distribuições de probabilidade multivariadas e modelos de regressão multivariados e marginais associados. A classe possui distribuições marginais Box-Cox simétricas, enquanto que uma cópula de uma mistura por escala de distribuições normais descreve a estrutura de dependência. Os modelos estudados nesta tese formam um cenário geral e flexível para modelar dados contínuos positivos, abrangendo desde dados independentes até dados multivariados com formas gerais de dependência. Derivamos propriedades e fornecemos interpretações sobre a estrutura de dependência induzida pela cópula na classe. Propomos inferência baseada na verossimilhança para a estimação dos parâmetros e discutimos uma estratégia para selecionar os componentes do modelo em aplicações práticas. Apresentamos estudos de simulação para verificar o desempenho dos estimadores de máxima verossimilhança em amostras de tamanho finito e aplicamos os modelos a conjuntos de dados reais com diferentes estruturas.

Palavras-chave: Estrutura de dependência. Dados positivos. Misturas por escala de distribuições normais.

Abstract

Rodrigo Matheus Rocha de Medeiros. **Multivariate Box-Cox symmetric models generated by a normal scale mixture copula**. Thesis (Doctorate). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2024.

The class of the Box-Cox symmetric distributions offers a flexible modeling framework for positive continuous data, reaching different levels of skewness and tail-heaviness. However, this class has been little explored in dependence modeling. The copula theory provides an approach for modeling dependence through a function that characterizes the associations among the elements of a random vector with given marginals. For instance, copulas generated by scale mixtures of normal distributions have attractive properties and describe dependence similarly to the multivariate normal distribution. This thesis aims to introduce a broad class of multivariate probability distributions and associated multivariate and marginal regression models. The class has Box-Cox symmetric marginal distributions, while a copula of a scale mixture of normal distributions describes the dependence structure. The models studied in this thesis constitute a general and flexible framework for modeling positive continuous data, handling from independent data to multivariate data with general forms of dependence. We derive properties and provide interpretations about the copula-induced dependence structure in the class of models. We propose a likelihood-based inference for parameter estimation and discuss a strategy for selecting model components in practical applications. We present simulation studies to verify the performance of the maximum likelihood estimators in finite-sized samples and apply the models to real data sets with different structures.

Keywords: Dependence structure. Positive data. Scale mixture of normal distributions.