UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

CHARLES PEIXOTO MAFALDA

ESSAYS ON UNIT MODELS FOR ANALYSING VOTE PROPORTIONS

Recife
2021

CHARLES PEIXOTO MAFALDA

ESSAYS ON UNIT MODELS FOR ANALYSING VOTE PROPORTIONS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de Mestre em Estatística.

Área de Concentração: Estatística Matemática

Orientador: Gauss Moutinho Cordeiro

Co-Orientador: Renata Rojas Guerra

Recife

2021

Catalogação na fonte
Bibliotecário Cristiano Cosme S. dos Anjos, CRB4-2290

M187e Mafalda, Charles Peixoto
Essays on unit models for analysing vote proportions / Charles Peixoto Mafalda. – 2021.
55 f.: il., fig.,tab.

Orientador: Gauss Moutinho Cordeiro.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2021.
Inclui referências e apêndices.

1. Estatística Matemática. 2. Algoritmo EM. 3. Eleições brasileiras. 4. GAMLSS. I. Cordeiro, Gauss Moutinho (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- CCEN 2021 - 71

CHARLES PEIXOTO MAFALDA

ESSAYS ON UNIT MODELS FOR ANALYSING VOTE PROPORTIONS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 26 de fevereiro de 2021

BANCA EXAMINADORA

Prof. Gauss M. Cordeiro
DE/UFPE

Prof.ª Fernanda De Bastiani
DE/UFPE

Prof. Fábio Mariano Bayer
UFSM

ABSTRACT

Defined as the district vote shares by the total number of valid votes cast in the district, the vote proportion is auseful measure for analyzing election data. Since it is a variable bounded in the standard unit interval, we consider unit distributions for analyzing the probabilistic behavior of vote proportions in Brazilian presidential elections runoff in 2018. The objective of this master thesis is twofold. Firstly, we introduce the two-component unit Weibull mixture model for describing the characteristics of these data, such as the asymmetric behavior, bimodality, and unit interval support. We provide some useful statistical properties of the new model, such as quantile function, moments, and incomplete moments. The Expectation-Maximization (EM) algorithm is derived for maximum-likelihood estimation, and a Monte Carlo study is carried out to evaluate the performance of these estimators on finite samples. The proposed model's superiority is verified when comparing the fit with some usual unit mixing models related in the literature: the two-component beta mixture and the two-component Kumaraswamy mixture models. The second objective is to identify the covariates associated with the elected candidate's vote proportion in the Brazilian municipalities with a population greater than 300.000 inhabitants. Thus, a study on unit regression models is performed using the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) framework. We fitted the beta and simplex regressions considering mean and dispersion sub-models. The Akaike information criterion (AIC), Schwarz's Bayesian criteria (SBC), and pseudo-R^2 statistics are considered as goodness-of-fit measures, and residual analysis is performed for diagnostics. The simplex regression is superior to the beta and is suitable for modeling the variable of interest. The covariates with significant effects are monthly household income per capita, the proportion of evangelicals and the political spectrum of the governors' party elected in 2014 and 2018. We also verify that some Brazilian regions impact the vote proportions' mean and dispersion.

Keywords: Brazilian elections. EM algorithm. GAMLSS. Mixture distributions. Unit models. Unit regression models. Unit Weibull.

RESUMO

Definido como a quantidade de votos do distrito pelo número total de votos válidos lançados no distrito, a proporção de votos é um medida útil para analisar dados eleitorais. Por se tratar de uma variável limitada ao intervalo unitário padrão, é importante considerar distribuições unitárias para analisar o comportamento probabilístico da proporção de votos no segundo turno das eleições presidenciais brasileiras no ano de 2018. Nesse contexto, a presente dissertação possui dois objetivos. Primeiro, propomos o modelo de mistura Weibull unitária de duas componentes para descrever as características desses dados, tais como o comportamento assimétrico, bimodalidade e suporte no intervalo unitário. Fornecemos algumas propriedades estatísticas úteis do novo modelo, como função quantílica, momentos e momentos incompletos. O algoritmo `Expectation-Maximization` (EM) é derivado para estimação de máxima verossimilhança, e um estudo de Monte Carlo é realizado para avaliar o desempenho desses estimadores em amostras finitas. A superioridade do modelo proposto é verificada ao comparar o ajuste com alguns modelos de mistura unitários usuais relacionados na literatura: o modelo de mistura de duas componentes beta e o modelo de mistura de duas componentes Kumaraswamy. O segundo objetivo é identificar as covariáveis associadas à proporção de votos dos candidatos eleitos nos municípios brasileiros com população superior a 300.000 habitantes. Assim, um estudo sobre modelos de regressão unitária é realizado usando a estrutura de Modelos Aditivos Generalizados de Locação, Escala e Forma (GAMLSS). Ajustamos as regressões beta e simplex considerando os submodelos de média e dispersão. O critério de informação de Akaike (AIC), o critério Bayesiano de Schwarz (SBC) e a estatística pseudo-R^2 são considerados como medidas de adequação e a análise residual é realizada para o diagnóstico. O modelo simplex foi superior ao modelo de regressão beta. Verificou-se que a renda familiar mensal per capita, proporção de evangélicos e o espectro político do partido dos governadores eleitos em 2014 e 2018 causaram efeitos significantes sobre a proporção de votos. Além disso, verificou-se que algumas regiões do Brasil apresentaram maiores impactos sobre a média e a dispersão da proporção de votos.

Palavras-chave: Algoritmo EM. Eleições brasileiras. GAMLSS. Mistura de distribuições. Modelos unitários. Modelos de regressão unitário. Weibull unitária.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 PRELIMINARIES

The vote proportion is a double bounded variable defined as the district vote shares by the total number of valid votes cast in the electoral district, which is typically defined as municipalities, states, or regions of a country. One important characteristic in electoral processes is that different candidates (or political parties) may win in distinct electoral districts. Therefore, the distribution of the vote proportions is useful to understand the political landscape during an election. In this work, we consider unit distributions for analyzing the probabilistic behavior of vote proportions in Brazilian presidential elections runoff in 2018 and define the Brazilian municipalities as electoral districts.

The objective of this master thesis is twofold and is presented in two main and independent chapters. Chapter 2 introduces the two-component unit Weibull mixture model and is motivated by the vote proportion's behavior when considering all the Brazilian municipalities. The proposed mixture model is suitable for describing these data since it is supported in the unit interval and accommodates asymmetric and bimodal behaviors. We also provide some useful statistical properties for the new mixture model, such as quantile function, incomplete moments, and moments. For parameter estimation, the maximum likelihood method is considered using the expectation-maximization (EM) algorithm. A Monte Carlo study is carried out to evaluate the performance of the proposed estimators in finite samples. Finally, the vote proportion application shows that the new model is quite competitive with the other two-component unit mixture models listed in the literature.

The second objective is to identify the covariates associated with the elected candidate's vote proportion in the Brazilian municipalities with a population greater than 300.000 inhabitants and is presented in Chapter 3. The study performed considering unit regression models in the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) framework (STASINOPOULOS; RIGBY; BASTIANI, 2018). The GAMLSS allows evaluating the effect of explanatory variables on each parameter indexed to the response distribution and provides a comprehensive framework for incorporating nonlinear, random, and spatial effects. In this work, the beta and simplex regressions considering mean and dispersion sub-models in the GAMLSS framework. In this master thesis, all computational implementations are performed in the `R` programming language (R Core Team, 2020). The notation and terminology are consistent throughout each chapter.

2 TWO-COMPONENT UNIT WEIBULL MIXTURE MODEL TO ANALYZE VOTE PROPORTIONS

2.1 INTRODUCTION

Finite mixture models appeared in a study on the asymmetry of grouped materials not being homogeneous (PEARSON, 1894), being useful in the presence of multimodality, heavy tails, and asymmetry (LACHOS *et al.*, 2017). Many works have appeared in the literature in the context of finite mixtures. For example, Jewell (1982) proposed a model for exponential mixtures. Considering Weibull mixture models, we can cite Jiang and Murthy (1998) for characterizations of the failure rate function and Bučar, Nagode and Fajdiga (2004) for reliability approximations. Recently, Huang and Dong (2019) analyzed individual periods in combined sea waves using parametric mixture models.

In data with limited support, beta mixture models have been studied by several authors. Ji *et al.* (2005) proposed a study on the beta mixture to solve problems related to correlations of gene expression levels, Bouguila, Ziou and Monga (2006) presented a study on Bayesian analysis, and Grün, Kosmidis and Zeileis (2011) studied beta mixture in regression models. The Kumaraswamy mixture model is an alternative to the beta mixture models, Khalid, Aslam and Sindhu (2020) carried out a Bayesian study on the three-component Kumaraswamy mixture.

In this paper, a new two-component mixture model is proposed as an alternative to model population heterogeneities in the unit support. To this aim, we consider that each mixture component follows a unit Weibull (UW) distribution (MAZUCHELI; MENEZES; GHITANY, 2018). Some of the main contributions to the proposition of this new distribution, the so-called Weibull mixture model of the two-component unit (UWUW), are i) parameterization based on quantiles to formulate each component of the mixture; ii) all estimation routines, including simulations and applications, are performed using the expectation-maximization (EM) algorithm, and iii) applicability for electoral data modeling. The advantage of working with reparametrization in terms of quantiles according to Bayes, Bazán and Castro (2017), is its flexibility to model data with heterogeneous conditional distributions. The EM algorithm is a computational method used to calculate the maximum likelihood estimator (MLE) iteratively (DEMPSTER; LAIRD; RUBIN, 1977). It is widely used to estimate the maximum probability for (REDNER; WALKER, 1984) finite mixture models. Finally, the adjustment to electoral data, defined as the district's share of votes by the total number of valid votes cast in the district, the proportions of votes are useful, since the electoral districts can vary considerably in the size of the population (ALEMÁN;

KELLAM, 2017). Also, this measure can analyze other characteristics of the electoral process, such as electoral volatility (POWELL; TUCKER, 2013) and nationalization of electoral change (ALEMÁN; KELLAM, 2017). The data set used refers to the proportions of votes in the Brazilian presidential elections runoff in 2018.

The rest of the work is organized as follows. In Section 2.2, an overview of the presidential elections in Brazil is provided. Section 2.3 presents the unit Weibull distribution and some of its main characteristics. In Section 2.4, the new mixture model is presented. Section 2.5 introduces the EM algorithm to perform maximum likelihood estimation for the UWUW model. In Section 2.7, an application is made with electoral data. The final considerations of this work are presented in Section 2.8.

2.2 MOTIVATING EXAMPLE

The election is one of the most fundamental processes in democratic societies (LYRA *et al.*, 2003), and the vote represents an effective instrument for regular citizens to promote significant changes in their communities (FILHO *et al.*, 2003). Consequently, electoral processes are of interest to many researchers and the general society (CARDOSO *et al.*, 2020). In these processes, not always the same candidate (or political party) wins in all electoral districts. Therefore, the vote proportion obtained in different cities is an important measure to understand the political landscape during an election. It may present several configurations, including asymmetric distribution and multimodality (PAZ; EHLERS; BAZÁN, 2015).

The UWUW model is introduced in order to analyze the probabilistic behavior of vote proportions in electoral processes. The model formulation is motivated by data from the runoff of Brazilian presidential elections in 2018. There were two candidates in the referred dispute - Jair Bolsonaro, from the Liberal Social Party (in Portuguese: Partido Social Liberal, PSL), and Fernando Haddad, from the Workers' Party (in Portuguese: Partido dos Trabalhadores, PT). Bolsonaro was the elected candidate with 55.13% of the valid votes. On the peculiarities of the Brazilian electoral processes, see, for instance, (FILHO *et al.*, 1999; FILHO *et al.*, 2003; CARDOSO *et al.*, 2020).

The results of the Brazilian elections are available on the data repository of the Brazilian superior electoral court (in Portuguese: Tribunal Superior Eleitoral, TSE) (TSE, 2018). It contains information about the number of votes of each candidate in the referred election for all the Brazilian municipalities. Table 1 shows the number and proportion of abstention, blank and null votes, and the votes obtained by each candidate in the runoff dispute. The abstention concerns the voters who do not go to vote on election day. There were a high number of blank and null votes in this election, and the abstention reached is 21.3%. Also known as a protest vote, a blank vote can be interpreted as dissatisfaction with the choice of candidates or refusal of the current political system (LEMONTE; BAZáN, 2016). Adding the percentage of null and blank votes with the percentage of abstention, more than 28% of voters did not choose either of the two candidates for the presidency of the Republic. This percentage of blank and null votes is the highest since the elections held in 1989 (TSE, 2018).

Table 1 – Results of the runoff of the presidential elections in 2018.

Candidate	Number of votes	Vote Proportion of valid (%)	Total of vote proportion (%)
Jair Bolsonaro	57.797.847	55,13%	39,24%
Fernando Haddad	47.040.906	44,87%	31,93%
Blank and null	11.094.698	–	7,53%
Abstention	31.371.704	–	21.3%
Total	147.305.155	100%	100%

Source: Author (2021)

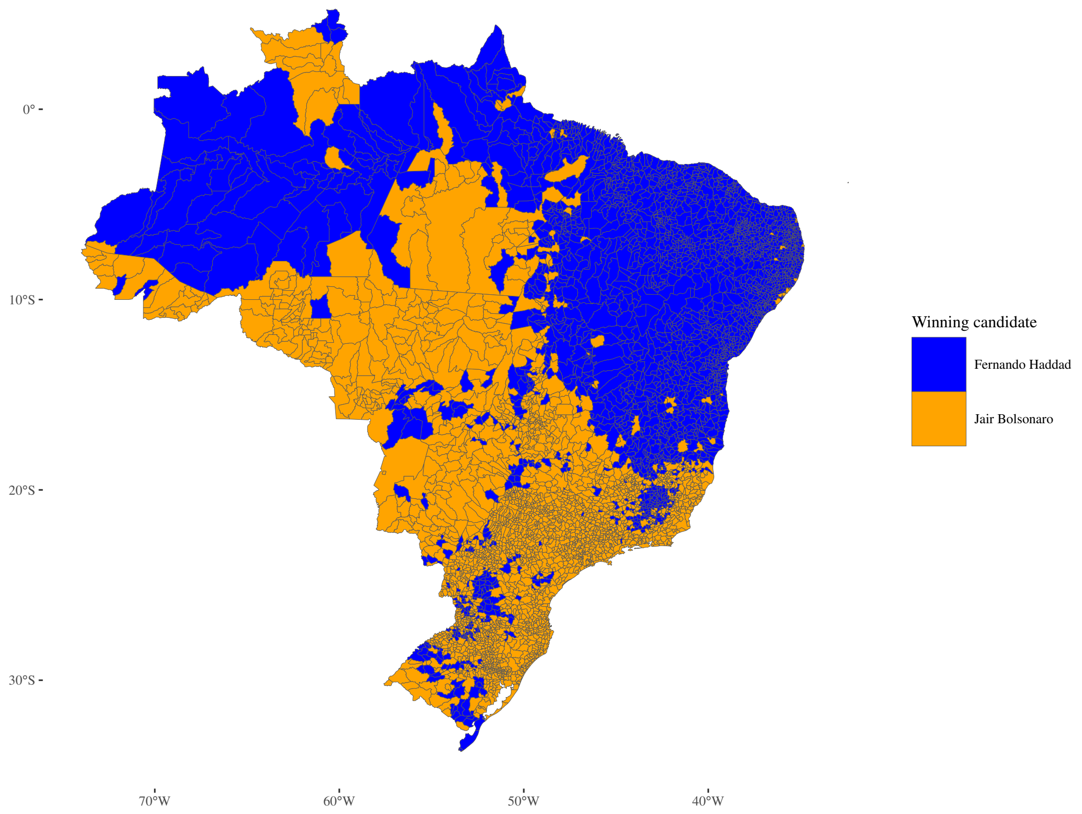
Figure 1 presents the municipalities where each candidate obtained the most votes (more than 50%). The candidate Fernando Haddad won in most of the Northeast's and North's municipalities, while Jair Bolsonaro reached the majority of votes in most of the municipalities located in the other Brazilian regions. Some authors have already carried out studies focusing on the performance of the PT's candidate in the Northeast and North regions. For instance, we can cite Abensur *et al.* (2007) and Junior and Souza (2015), who analyzed the vote proportion in previous elections.

In this paper, we employ these vote proportions for illustrating the two-component unit Weibull mixture model. The introduced model is suitable for describing the characteristics of these data, such as the asymmetric behavior, bimodality, and the unit interval support, see Figure 2.

2.3 THE UNIT WEIBULL DISTRIBUTION

The unit Weibull distribution (UW) was introduced by Mazucheli, Menezes and Ghitany (2018) as an alternative to the classical beta and Kumaraswamy distributions. Since it has closed-form expressions for the cumulative distribution function (cdf) and quantile function (qf), Mazucheli *et al.* (2020) presented a quantile-parametrization, detailed to follow. Let Y be a

Figure 1 – Winning candidate in the Brazilian presidential elections runoff in 2018.



Source: Author (2021)

random variable following a UW distribution. Then, its cdf is

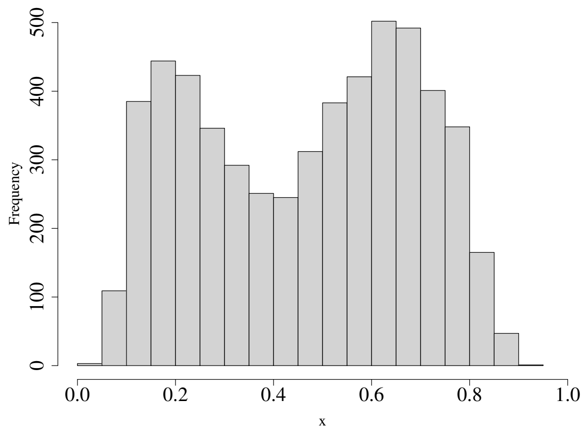
$$F_{\mathrm{UW}}(y|\boldsymbol{\theta}) = \tau^{[\log y/\log \mu]^{\beta}}, \quad 0 < y < 1, \tag{2.1}$$

where $\boldsymbol{\theta} = (\mu, \beta)^{\top}$ is the parameter vector, $\mu \in (0,1)$ is a location parameter that represents the τth quantile of Y, $\beta > 0$ is a shape parameter, and τ is assumed to be known. The corresponding probability density function (pdf) of Y is

$$f_{\mathrm{UW}}(y|\boldsymbol{\theta}) = \frac{\beta}{y}\frac{\log \tau}{\log \mu}\left[\frac{\log y}{\log \mu}\right]^{\beta-1}\tau^{[\log y/\log \mu]^{\beta}}, \quad 0 < y < 1. \tag{2.2}$$

When $\mu = \beta = 1$, the standard uniform distribution (MAZUCHELI; MENEZES; GHITANY, 2018) is obtained as a special case. The power function distribution (MAZUCHELI; MENEZES; GHITANY, 2018) arrises when $\beta = 1$, and the unit Rayleigh (UR) (MAZUCHELI; MENEZES; GHITANY, 2018) when $\beta = 2$. The UW distribution is also part of the extended unit Weibull

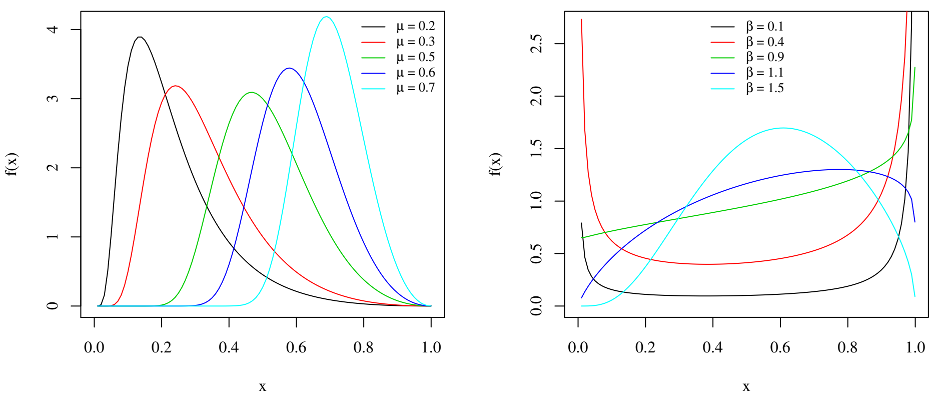
Figure 2 – Histogram of the proportion of candidate Bolsonaro's valid votes in the Brazilian presidential elections runoff in 2018.



Source: Author (2021)

Figure 3 – Plots of the UW densities with different parameter values.

(a) For $\beta = 3$ (b) For $\mu = 0.6$



Source: Author (2021)

family pioneered by Guerra, Peña-Ramírez and Bourguignon (2020). The UW distribution can be very useful in practical situations in several fields, such as hydrology (MAZUCHELI; MENEZES; GHITANY, 2018), biometrics, and economic science (MAZUCHELI *et al.*, 2020).

The UW pdf can take decreasing, increasing, unimodal, and anti-modal shapes. Figure 3 shows some plots of its density for several combinations of $\boldsymbol{\theta}$. The quantile order is fixed at $\tau = 0.5$ since it does not interfere with the density shapes.

The qf of the UW distribution (under this parametrization) is

$$Q_{\mathrm{UW}}(u|\boldsymbol{\theta}) = \mu^{[\log u/\log \tau]^{1/\beta}}. \tag{2.3}$$

For $\beta > 1$, the sth moment can be defined as

$$E(Y^s) = \sum_{k=0}^{\infty} \frac{(s \log \mu)^k}{k!(-\log \tau)^{k/\beta}} \Gamma\left(\frac{k}{\beta} + 1\right), \tag{2.4}$$

where $\Gamma(s) = \int_0^{\infty} \exp\{-t\}\ t^{s-1} \mathrm{d}t$ is the gamma function.

The incomplete moments of a random variable, say Y, play an important role in several fields, such as econometrics, insurance and medicine. The sth incomplete moment of Y is defined as $T_Y(r,s) = \int_0^r y^s f(y) \mathrm{d}y$, where $f(y)$ is the pdf of Y. For example, the Lorenz and Bonferroni curves are applications of the first incomplete moment. For a given probability π, they are defined by $B(\pi) = T_Y(q,1)/(\pi E(Y))$ and $L(\pi) = T_Y(q,1)/(E(Y))$, respectively, where $q = Q_Y(\pi)$ is the πth quantile of Y. Another related measure is the Gini concentration (C_G), which is defined as the area between the curve $L(\pi)$ and the straight line. Hence, $C_G = 1 - 2\int_0^1 L(\pi) \mathrm{d}u$. The closer C_G is to one the more uniqual is the distribution of the variable of interest. Analogously a C_G of zero expresses perfect equality. From the definition of the incomplete moments follows that $E(Y^s) = \lim_{r \to \infty} T_Y(r,s)$ and $F_Y(y) = T_Y(y,0)$.

The following proposition gives a general result on the sth incomplete moment of the UW distribution.

Proposition 2.3.1 *The sth incomplete moment of Y (for $\beta > 1$) is*

$$T_{\mathrm{UW}}(r,s) = \sum_{k=0}^{\infty} \frac{(s \log \mu)^k}{k!(-\log \tau)^{k/\beta}} \Gamma\left(\frac{k}{\beta} + 1\ ,\ \frac{s \log r \log \tau}{(-\log \mu)^{\beta}}\right),$$

where $\Gamma(s,r) = \int_r^{\infty} \exp\{-t\}\ t^{s-1} dt$ is the upper incomplete gamma function. The pdf in (2.2) *can be rewrite as*

$$f_{\mathrm{UW}}(y|\boldsymbol{\theta}) = \frac{\beta}{y} \frac{\log \tau}{\log \mu} \left[\frac{\log y}{\log \mu}\right]^{\beta - 1} \exp\left\{\log \tau \left[\frac{\log y}{\log \mu}\right]^{\beta}\right\}.$$

Using the last expression in the definition of $T_Y(r,s)$, and considering the exponential expansion in $\exp\{u\}$*, with $u = \log \tau (\log y / \log \mu)^{\beta}$, is obtained*

$$T_Y(r,s) = \sum_{k=0}^{\infty} \frac{(s \log \mu)^k}{k!(-\log \tau)^{k/\beta}} \int_{s \log r\ \log \tau/(-\log \mu)^{\beta}}^{\infty} \exp\{-u\}\ u^{[(k/\beta)+1]-1} du.$$

The proof is completed by replacing the upper incomplete gamma function in the above equation.

2.4 THE PROPOSED MODEL

A two-component mixture model is defined as a convex linear combination of two independently distributed random variables. Let Y_i, with $i = 1,2$, be two random variables independently distributed with densities, say $f_1(y;\boldsymbol{\theta}_1)$ and $f_2(y;\boldsymbol{\theta}_2)$, respectively. Then, the pdf of random variable X that defines its two-component mixture is

$$f_X(x;\boldsymbol{\Theta}) = p f_1(x;\boldsymbol{\theta}_1) + (1-p) f_2(x;\boldsymbol{\theta}_2), \tag{2.5}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, p)^\top$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the parameter vectors associated with each mixture component, and $p \in (0,1)$ is the mixing parameter. The corresponding cdf of X is

$$F_X(x;\boldsymbol{\Theta}) = p F_1(x;\boldsymbol{\theta}_1) + (1-p) F_2(x;\boldsymbol{\theta}_2), \tag{2.6}$$

where $F_i(x;\boldsymbol{\theta}_i)$ is the cdf of Y_i $(i = 1,2)$.

In this section, the two-component unit Weibull mixture distribution, so-denoted UWUW, is introduced using the above framework. Some important mathematical and statistical properties of the new model are explored. Thus, let X be a random variable with UWUW distribution. Then, its cdf is obtained replacing (2.1) in each component of (2.6) as

$$\begin{aligned} F_{\mathrm{UWUW}}(x;\boldsymbol{\Theta}) &= p\, F_{UW}(x;\boldsymbol{\theta}_1) + (1-p) F_{UW}(x;\boldsymbol{\theta}_2) \\ &= p\, \tau^{[-\log x/\log \mu_1]^{\beta_1}} + (1-p)\, \tau^{[-\log x/\log \mu_2]^{\beta_2}}, \end{aligned}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, p)^\top$, $\boldsymbol{\theta}_1 = (\mu_1,\beta_1)^\top$, $\boldsymbol{\theta}_2 = (\mu_2,\beta_2)^\top$, μ_1 and $\mu_2 \in (0,1)$ are location parameters associated with the τth quantiles of each component of the mixture, β_1 and $\beta_2 > 0$ are shape parameters, and $\tau \in (0,1)$ is assumed to be known.

Replacing Equation (2.2) in (2.5), the UWUW pdf is obtained as

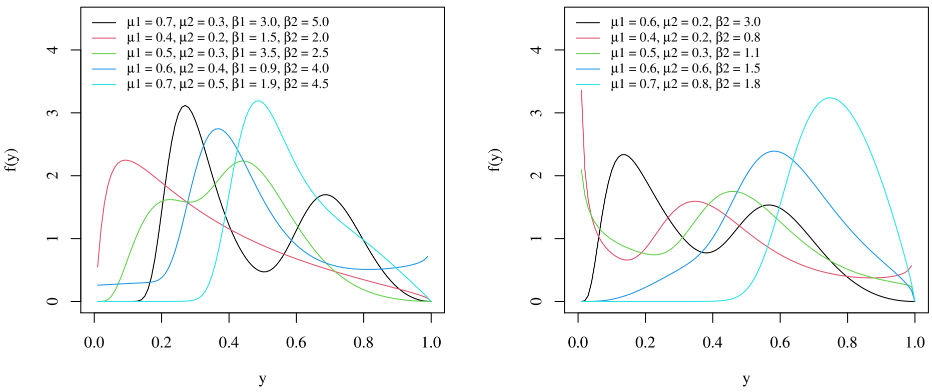
$$\begin{aligned} f_{\mathrm{UWUW}}(x;\boldsymbol{\Theta}) &= p\, f_{\mathrm{UW}}(x;\boldsymbol{\theta}_1) + (1-p)\, f_{\mathrm{UW}}(x;\boldsymbol{\theta}_2) \\ &= p\, \frac{\beta_1 \log \tau}{x \log \mu_1} \left(\frac{\log x}{\log \mu_1}\right)^{\beta_1 - 1} \tau^{(\log x/\log \mu_1)^{\beta_1}} \\ &+ (1-p)\, \frac{\beta_2 \log \tau}{x \log \mu_2} \left(\frac{\log x}{\log \mu_2}\right)^{\beta_2 - 1} \tau^{(\log x/\log \mu_2)^{\beta_2}}. \end{aligned} \tag{2.7}$$

Figure 4 shows some graphs of UWUW pdf with various combinations of parameters and $\tau = 0.5$, which reveals the high flexibility of the new distribution. It accommodates bimodal, unimodal, descending, and bath forms under different asymmetric characteristics. Also, it is possible to identify a bimodal form for different values of p. Hereafter, we denote X as a random variable following a UWUW distribution, this is, $X \sim \mathrm{UWUW}(\boldsymbol{\Theta})$.

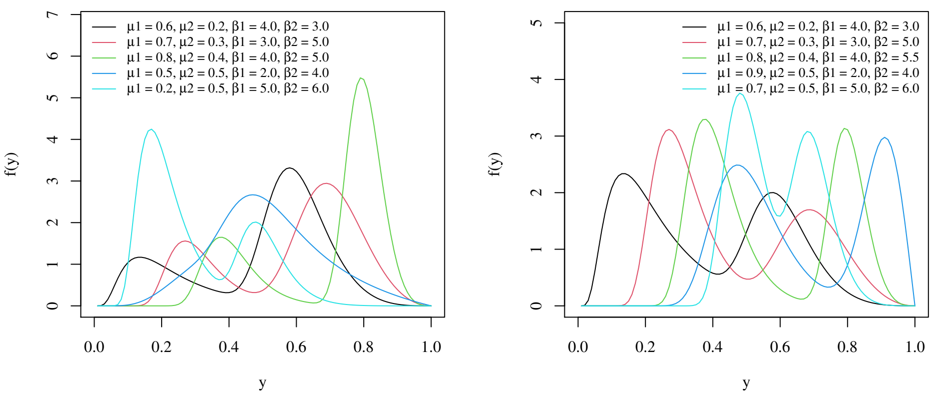
The following result provides the X moments and incomplete moments. It is noteworthy that the incomplete moments and moments are defined for β_1 and $\beta_2 > 1$.

Figure 4 – Plots of the UWUW **density for some parameter values.**

(a) For $p = 0.4$. (b) For $p = 0.6$ and $\beta_1 = 3.0$.



(c) For $p = 0.7$. (d) For $p = 0.1$.



Source: Author (2021)

Proposition 2.4.1 *The sth moment of X (for $\beta_1 > 1$ and $\beta_2 > 1$) is*

$$E(X^s) = p\sum_{k=0}^{\infty}\frac{(s\log\mu_1)^k}{k!(-\log\tau)^{k/\beta_1}}\Gamma\left(\frac{k}{\beta_1}+1\right)$$
$$+(1-p)\sum_{k=0}^{\infty}\frac{(s\log\mu_2)^k}{k!(-\log\tau)^{k/\beta_2}}\Gamma\left(\frac{k}{\beta_2}+1\right).$$

We have that

$$E(X^s) = p\int_0^1 x^s f_{UW}(x;\boldsymbol{\theta}_1)\mathrm{d}x + (1-p)\int_0^1 x^s f_{UW}(x;\boldsymbol{\theta}_2)\mathrm{d}x.$$

Using (2.4)*, the proof is completed.*

From the incomplete moments of the UW distribution, it is possible to obtain the incomplete moments of the UWUW model.

Proposition 2.4.2 *The sth incomplete moment of X (for $\beta_1 > 1$ and $\beta_2 > 1$) is*

$$T_{\mathrm{UWUW}}(r,s) = p\sum_{k=0}^{\infty}\frac{(s\log\mu_1)^k}{k!(-\log\tau)^{k/\beta_1}}\Gamma\left(\frac{k}{\beta_1}+1\,,\ \frac{s\log r\log\tau}{(-\log\mu_1)^{\beta_1}}\right)$$
$$+(1-p)\sum_{k=0}^{\infty}\frac{(s\log\mu_2)^k}{k!(-\log\tau)^{k/\beta_2}}\Gamma\left(\frac{k}{\beta_2}+1\,,\ \frac{s\log r\log\tau}{(-\log\mu_2)^{\beta_2}}\right),$$

We can write

$$T_{\mathrm{UWUW}}(r,s) = p\int_0^r x^s f_{\mathrm{UW}}(x;\boldsymbol{\theta_1})\mathrm{d}x + (1-p)\int_0^r x^s f_{\mathrm{UW}}(x;\boldsymbol{\theta_2})\mathrm{d}x.$$

The proof follows from Proposition 2.3.1.

2.5 PARAMETER ESTIMATION

An approach to the iterative computation of maximum-likelihood estimates (MLE) when the observations can be treated as incomplete data is the well-known expectation-maximization (EM) algorithm. Considering the context of two-component mixture models, let $\boldsymbol{x} = \{x_1,\dots,x_n\}$ be a random sample of size n from a random variable X having pdf (4) with unknown parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, p)^\top$, where $\boldsymbol{\theta}_1 = (\mu_1, \beta_1)^\top$ and $\boldsymbol{\theta}_2 = (\mu_2, \beta_2)^\top$. It is customary to call $\boldsymbol{x}$ of “incomplete data" since it is associated with a second component $\boldsymbol{z} = \{z_1,\dots,z_n\}$ of unobserved values of a latent random variable Z. Each value z_i of Z indicates which component of the mixture belongs to the ith observation x_i, of such that

$$z_i = \begin{cases} 1 & \text{if } x_i \text{ has pdf } f_{\mathrm{UW}}(x|\boldsymbol{\theta_1}),\\ 0 & \text{if } x_i \text{ has pdf } f_{\mathrm{UW}}(x|\boldsymbol{\theta_2}),\end{cases}$$

where $P(Z=1) = p$ and $P(Z=0) = 1-p$. The complete-data specification is determined by the joint density of (X,Z)

$$f_{X,Z}(x_i,z_i;\boldsymbol{\Theta}) = \left[p\,\frac{\beta_1\log\tau}{x_i\log\mu_1}\left(\frac{\log x_i}{\log\mu_1}\right)^{\beta_1-1}\tau^{(\log x_i/\log\mu_1)^{\beta_1}}\right]^{z_i}$$
$$\times\left[(1-p)\,\frac{\beta_2\log\tau}{x\log\mu_2}\left(\frac{\log x}{\log\mu_2}\right)^{\beta_2-1}\tau^{(\log x/\log\mu_2)^{\beta_2}}\right]^{1-z_i},$$

and based on it, the complete log-likelihood function, for the sample of size n, is given by

$$l_c(\boldsymbol{\Theta}) = \sum_{i=1}^{n}\log f_{X,Z}(x_i,z_i;\boldsymbol{\Theta})$$
$$=\sum_{i=1}^{n} z_i\log\left[p\,\frac{\beta_1\log\tau}{x_i\log\mu_1}\left(\frac{\log x_i}{\log\mu_1}\right)^{\beta_1-1}\tau^{(\log x_i/\log\mu_1)^{\beta_1}}\right]$$

$$+\sum_{i=1}^{n}(1-z_i)\log\left[(1-p)\,\frac{\beta_2\log\tau}{x\log\mu_2}\left(\frac{\log x}{\log\mu_2}\right)^{\beta_2-1}\tau^{(\log x/\log\mu_2)^{\beta_2}}\right]. \tag{2.8}$$

The EM algorithm iterates, between two steps, to compute the MLEs of $\boldsymbol{\Theta}$. In the E-step or expectation step, due to (2.8) is unobservable, it is replaced by its conditional expectation with respect to the conditional distribution of Z, given $\boldsymbol{x}$ and the current parameter estimates. More specifically, in the $(k+1)$th iteration, the E-step computes

$$\begin{aligned}
Q(\boldsymbol{\Theta},\boldsymbol{\Theta}^{(k)}) =& E_{\boldsymbol{\Theta}^{(k)}}\,[l_c(\boldsymbol{\Theta})|\boldsymbol{x}]\\
=&\sum_{i=1}^{n}\log f_{X,Z}(x_i,z_i;\boldsymbol{\Theta})\\
=&\sum_{i=1}^{n}\bar{z}_{i1}\,\log\left[p\,\frac{\beta_1\log\tau}{x_i\log\mu_1}\left(\frac{\log x_i}{\log\mu_1}\right)^{\beta_1-1}\tau^{(\log x_i/\log\mu_1)^{\beta_1}}\right]\\
&+\sum_{i=1}^{n}\bar{z}_{i2}\,\log\left[(1-p)\,\frac{\beta_2\log\tau}{x\log\mu_2}\left(\frac{\log x}{\log\mu_2}\right)^{\beta_2-1}\tau^{(\log x/\log\mu_2)^{\beta_2}}\right],
\end{aligned} \tag{2.9}$$

where

$$\bar{z}_{i1}=\frac{p^{(k)}f_{UW}(x;\boldsymbol{\theta}_1^{(k)})}{p^{(k)}f_{UW}(x;\boldsymbol{\theta}_1^{(k)})+(1-p^{(k)})f_{UW}(x;\boldsymbol{\theta}_2^{(k)})},$$

$$\bar{z}_{i2}=\frac{(1-p^{(k)})\,f_{UW}(x;\boldsymbol{\theta}_2^{(k)})}{p^{(k)}f_{UW}(x;\boldsymbol{\theta}_1^{(k)})+(1-p^{(k)})f_{UW}(x;\boldsymbol{\theta}_2^{(k)})},$$

and $\boldsymbol{\Theta}^{(k)}=(\boldsymbol{\theta}_1^{(k)},\boldsymbol{\theta}_2^{(k)},p^{(k)})^\top$ are obtained from the kth iteration.

The M-step or maximization step, requires the maximization of (2.9) with respect to $\boldsymbol{\Theta}$. This is

$$\boldsymbol{\Theta}^{(k+1)}=\underset{\boldsymbol{\Theta}}{\arg\max}\;Q(\boldsymbol{\Theta},\boldsymbol{\Theta}^{(k)}). \tag{2.10}$$

The vector $\boldsymbol{\Theta}^{(k+1)}$ is used to initialize the next iteration. Thus, the EM algorithm is initialized by the starting values $\boldsymbol{\Theta}^{(0)}=(\boldsymbol{\theta}_1^{(0)},\boldsymbol{\theta}_2^{(0)},p^{(0)})^\top$ and the MLEs $\hat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta}$ are obtained by $\hat{\boldsymbol{\Theta}}=\boldsymbol{\Theta}^{(k+1)}$ when a convergence criterion $|\boldsymbol{\Theta}^{(k+1)}-\boldsymbol{\Theta}^{(k)}|<\varepsilon$ is reached (DEMPSTER; LAIRD; RUBIN, 1977). We set $\varepsilon=10000$. It should be noted that it is not possible to obtain analytical results from these expressions. It is necessary to perform this maximization by applying some iterative techniques, for example, Newton Raphson's method (PRESS *et al.*, 2007).

2.6 SIMULATION STUDY

In this section, a Monte Carlo experiment is carried out to evaluate the performance of the EM algorithm to estimate the UWUW parameter vector. The simulation routines are implemented using the `R` programming language (R Core Team, 2020) and the `ContrOptim` function with the `Nelder-Mead` maximization algorithm. The qf of the UWUW model can only be obtained numerically. However, it is possible to generate random numbers using the qf of the UW distribution given in (2.3). Alencar (2018) and Carvalho (2015) carried out simulation studies similar to what we are proposing. Thus, let $\boldsymbol{x} = x_1, \ldots, x_n$ be a random sample of $X \sim \text{UWUW}(\boldsymbol{\Theta})$, then the ith element of $\boldsymbol{x}$, $i = 1, \ldots, n$, can be generated as follows:

1. generate u_1 and u_2 from two independent standard uniform random variables;

2. if $u_1 < p$, where p is the weight of the first component of the mixture, calculate $x_i = Q_{\text{UW}}(u_2|\boldsymbol{\theta_1})$;

3. if $u_1 \geq p$, calculate $x_i = Q_{\text{UW}}(u_2|\boldsymbol{\theta_2})$.

All reported results are based on 10,000 replications, and the sample sizes are set at $n \in \{100, 250, 500\}$. Eight different parameter combinations are considered, and the value of τ is fixed at 0.5. Thus, for all the scenarios generated in this section, the quantile parameters μ_1 and μ_2 represent the median of each mixture component. To assess the performance of the EM algorithm, the percentage relative bias (RB%) and the mean squared error (MSE) are calculated. Further simulations, considering different scenarios and values for τ, are reported in the Appendix.

Table 2 shows the simulation results and allows us to note the RB% and MSE tend to be closer to zero as the sample size increases. Note that RB% of $\hat{\mu}_1$ and $\hat{\mu}_2$ does not exceed 0.9% when n is larger than 100. In general, the median estimates parameters are more accurate than those for the shape parameters β_1 and β_2. Notice $\hat{\beta}_2$ has the highest RB%, and this configuration is observed for all sample sizes.

Table 2 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model.

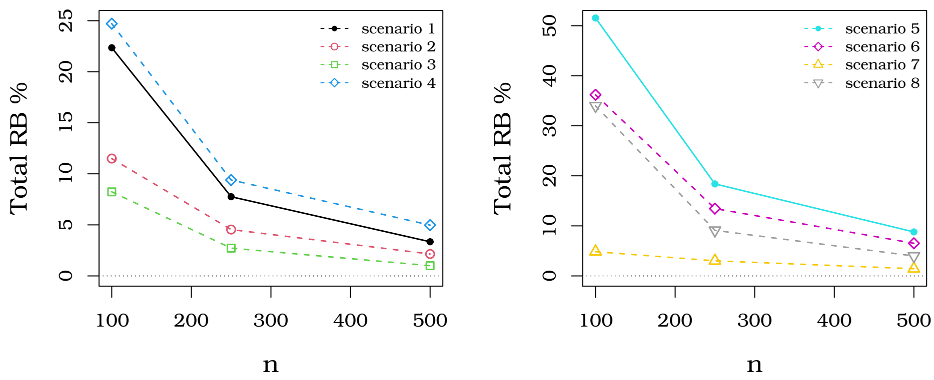
Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$
1	0.3	0.7	2.0	3.0	0.4	100	-0.4189	-0.0881	14.5975	8.2128	0.0479	0.0014	0.0002	0.1647	0.1685	0.0017
						250	-0.1693	-0.0420	4.9168	3.0942	-0.0426	0.0005	0.0001	0.0428	0.0576	0.0006
						500	-0.0930	-0.0143	2.4068	1.0230	0.0241	0.0002	0.0001	0.0184	0.0269	0.0003
2	0.2	0.6	0.9	2.5	0.4	100	0.1731	0.0328	4.3019	6.9431	0.0442	0.0048	0.0005	0.0214	0.1389	0.0013
						250	0.0716	0.0002	1.6727	2.7888	-0.0035	0.0019	0.0002	0.0066	0.0485	0.0005
						500	0.0371	0.0013	0.6670	1.4385	0.0078	0.0009	0.0001	0.0029	0.0234	0.0002
3	0.1	0.8	1.5	0.8	0.5	100	0.6696	-0.6281	5.3045	2.8803	0.0086	0.0014	0.0020	0.0566	0.0187	0.0015
						250	0.2527	-0.2658	1.6845	1.0632	-0.0133	0.0004	0.0006	0.0207	0.0065	0.0006
						500	0.1273	-0.1311	0.6829	0.3130	0.0118	0.0001	0.0003	0.0101	0.0031	0.0003
4	0.7	0.5	0.9	5.0	0.5	100	-0.1970	-0.0549	2.8046	22.1717	-0.0131	0.0022	0.0002	0.0119	0.8689	0.0012
						250	-0.1088	-0.0173	1.1837	8.3364	0.0017	0.0008	0.0001	0.0043	0.2722	0.0005
						500	-0.0286	-0.0144	0.4739	4.5350	0.0099	0.0004	0.0001	0.0019	0.1287	0.0002
5	0.6	0.2	5.0	3.5	0.8	100	-0.0221	-0.0185	8.8737	42.6975	0.0124	0.0001	0.0008	0.2519	1.1632	0.0014
						250	-0.0141	-0.0131	2.9371	15.4892	-0.0127	0.0001	0.0003	0.0911	0.3070	0.0005
						500	-0.0062	0.0022	1.4449	7.3342	0.0158	0.0001	0.0001	0.0444	0.1363	0.0002
6	0.9	0.4	1.9	3.0	0.8	100	-0.0225	-0.0095	2.9770	33.2689	-0.0204	0.0001	0.0012	0.0365	0.7896	0.0015
						250	-0.0079	-0.0297	1.2212	12.2812	-0.0055	0.0001	0.0005	0.0136	0.2170	0.0006
						500	-0.0032	-0.0043	0.5437	5.9671	0.0208	0.0001	0.0002	0.0062	0.0946	0.0003
7	0.5	0.9	2.5	0.4	0.9	100	-0.0818	-3.4265	4.7901	3.5705	-0.0097	0.0003	0.0182	0.0588	4.4945	0.0005
						250	-0.0513	-1.4269	1.7292	2.7833	-0.0173	0.0001	0.0049	0.0214	0.0069	0.0002
						500	-0.0236	-0.6906	0.9218	1.2397	-0.0094	0.0001	0.0020	0.0104	0.0025	0.0001
8	0.8	0.1	3.5	0.6	0.9	100	0.0103	3.8199	3.8603	26.3026	-0.0109	0.0001	0.0281	0.1072	1.1422	0.0007
						250	0.0028	1.5648	1.3948	6.1629	-0.0184	0.0001	0.0107	0.0403	0.0301	0.0002
						500	0.0008	0.8949	0.6650	2.4312	-0.0101	0.0001	0.0051	0.0194	0.0074	0.0001

Source: Author (2021)

Figure 5 illustrates the EM algorithm's convergence using the total RB%, which is defined as the sum of the absolute values of the individual RB%. It is an aggregate measure of the bias, and its outcome confirms that the parameter estimates improve their accuracy as n increases. Another useful aggregate measure is the total MSE, defined as the sum of all the individual MSEs. Figure 6 displays plots of this measure for the eight scenarios considered. Note that when n is larger than 100, the total MSE is less than 1 for all scenarios.

Figure 5 – Total RB% for the estimates of the MLEs of the UWUW model obtained via EM algorithm.

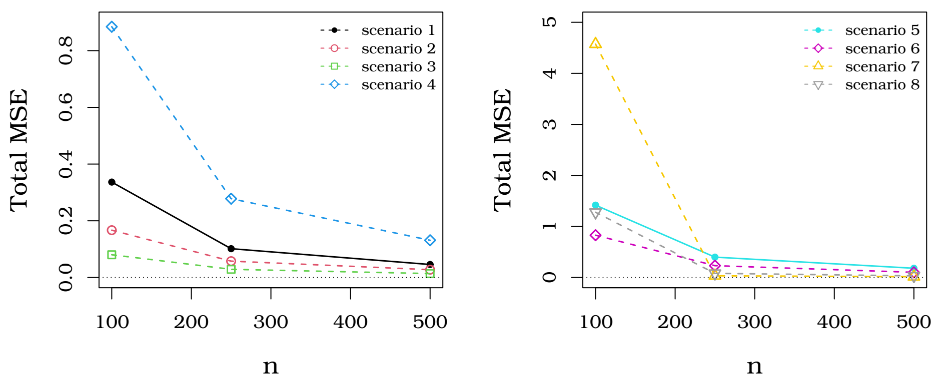
(a) Scenarios 1 to 4. (b) Scenarios 5 to 8.



Source: Author (2021)

Figure 6 – Total MSE for the MLEs of the UWUW model obtained via EM algorithm.

(a) Scenarios 1 to 4. (b) Scenarios 5 to 8.



Source: Author (2021)

2.7 APPLICATION

In what follows, we present a case study that illustrates the suitability of the UWUW distribution for modeling real unit data sets. The database considered is the municipality's vote proportion of the winning candidate in the Brazilian presidential elections runoff in 2018. Since it presents a bimodal shape, see Figure 7, a unimodal distribution would not be appropriate to fit this data set. Therefore, the UWUW distribution is a suitable alternative to model these data. Its performance is compared with other double-bounded component mixtures that have already been studied in the literature: two-component beta mixture (BB) and two-component Kumaraswamy mixture (KWKW) models.

The BB model is obtained by replacing the pdf of the beta distribution in Equation (2.5). In this paper, the parameterization proposed by Ferrari and Cribari-Neto (2004) is considered to define the BB model, which has pdf given by

$$
\begin{aligned}
f(x;\boldsymbol{\Theta}) = p\frac{\Gamma(\mu_1+\beta_1)}{\Gamma(\mu_1)\Gamma(\beta_1)}x^{\mu_1-1}(1-x)^{\beta_1-1}+(1-p) \\
\frac{\Gamma(\mu_2+\beta_2)}{\Gamma(\mu_2)\Gamma(\beta_2)}x^{\mu_2-1}(1-x)^{\beta_2-1}, \quad 0<x<1,
\end{aligned}
$$

where $\boldsymbol{\Theta} = (\mu_1,\mu_2,\beta_1,\beta_2,p)^\top$, μ_1 and $\mu_2 \in (0,1)$ are location parameters associated with the mean of each mixture component, β_1 and $\beta_2 > 0$ are precision parameters, and $p \in (0,1)$ is the parameter that measures the weights of the mixture.

The KWKW model is obtained by replacing the pdf of the Kumaraswamy distribution (KUMARASWAMY, 1980) in Equation (2.5). We consider the parameterization in Mitnik and Baek (2013) to define the KWKW model, which has pdf given by

$$
\begin{aligned}
f(x;\boldsymbol{\Theta}) = p\,\mu_1\beta_1 x^{\mu_1-1}(1-x^{\mu_1})^{\beta_1-1}+(1-p) \\
\mu_2\beta_2 x^{\mu_2-1}(1-x^{\mu_2})^{\beta_2-1}, \quad 0<x<1,
\end{aligned}
$$

where $\boldsymbol{\Theta} = (\mu_1,\mu_2,\beta_1,\beta_2,p)^\top$, μ_1 and $\mu_2 \in (0,1)$ are location parameters associated with the median of each mixture component, β_1 and $\beta_2 > 0$ are precision parameters, and $p \in (0,1)$ is the parameter that measures the weights of the mixture.

For all competitive mixture models, the parameter estimation is carried out using the EM algorithm following the steps described in Section 2.5. The Corrected Anderson-Darling (A^*) (CHEN; BALAKRISHNAN, 1995), Cramér-von Misses (W^*) (DURBIN; KNOTT, 1972), and the Kolmogorov Smirnov (KS) (GOODMAN, 1954) statistics are calculated to assess the quality-of-fit for the three fitted models. The lower their values are, the better is the model fit. All the analysis is performed using the `R` programming language, and the goodness-of-fit measures are computed using the `AdequacyModel` (MARINHO *et al.*, 2019) subroutine.

Table 3 displays the parameter estimates, standard errors, and the model comparison criteria of the three considered models. The results indicate that the UWUW distribution provides the lowest values for all goodness-of-fit statistics. The KWKW presents the worse performance, not being an adequate alternative to fit these data.

Table 3 – Parameter estimates and standard errors (given in parentheses) for the models fitted to Bolsonaro's vote proportion in Brazilian presidential elections in 2018.

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{p}$	W^*	A^*	KS
BB	0.5816	0.1985	9.7510	29.3260	0.7268	1.2937	7.4584	0.0477
	(0.0035)	(0.0026)	(0.3201)	(1.3521)	-			
UWUW	0.2677	0.6491	2.7011	2.9611	0.5368	**0.4119**	**3.6768**	**0.0153**
	(0.0039)	(0.0027)	(0.0545)	(0.0567)	-			
KWKW	0.5475	0.5475	1.0399	1.0398	0.3736	10.4367	60.8235	0.2465
	(0.0103)	(0.0074)	(0.0295)	(0.0214)	-			

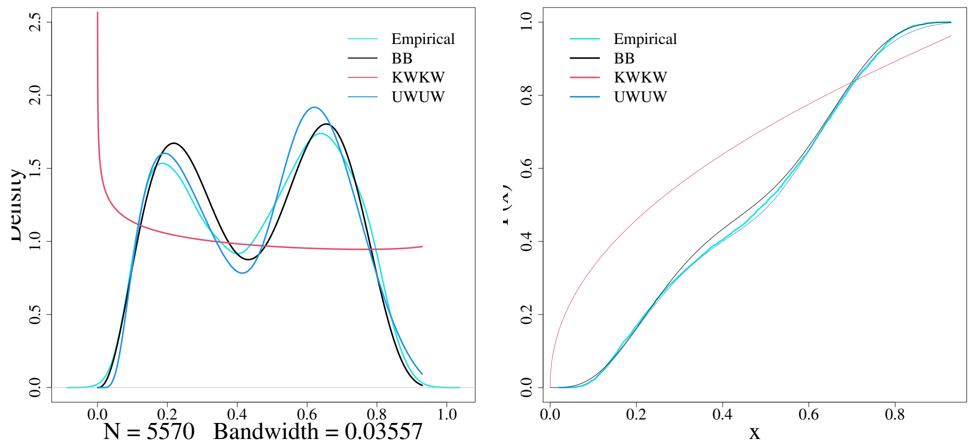
Source: Author (2021)

Figure 7 presents the histogram of the vote proportion data overlaid with the estimated densities of the fitted models. The bimodality of the data is confirmed, and the UWUW model provides the closest fit to the histogram. Clearly, the KWKW model is not adequate to fit these data. Further, Figure 7 gives plots of the empirical and estimated cdf functions. This visual inspection favors the results in Figure 7 and Table 3, indicating that the proposed model is appropriate to fit these data. Thus, it can be an effective alternative to analyze vote proportions, being quite competitive with the BB model and providing consistently better fits than the KWKW model. Therefore, the UWUW provides a useful tool for modeling bimodal data restricted to the unit interval. Also, with the estimates of the mixture parameters, it is possible to identify that more than 50 % of the observations belong to the first mixture component. The estimated median of the first component is $\hat{\mu}_1 = 0.2677$ and the estimated median of the second component is $\hat{\mu}_2 = 0.6649$.

2.8 CONCLUSION

A two-component mixture model was defined to describe the heterogeneities of the population with the limited domain. Named the two-component unit Weibull mixture (UWUW) model, the new distribution is formulated considering that each mixture component follows the unit Weibull distribution. Some of the main properties of UWUW have been presented, such as

Figure 7 – Estimated densities (a) and empirical cdf (b) of the BB, KWKW **and** UWUW **models.**



Source: Author (2021)

ordinary and incomplete moments. The EM algorithm was used to obtain maximum likelihood estimates for the model parameters. To evaluate the performance of the EM algorithm, Monte Carlo simulations were performed. An application to electoral data illustrates the importance and potential of the new model. The motivating data set is about the vote proportions obtained by the winning candidate in the Brazilian presidential runoff elections in 2018. The results indicate that our proposal is adequate to fit this data set since it is suitable to analyze the asymmetric and bimodal behaviors. From the mixing parameter estimate, we can conclude that 53.68% of the observations are from the first component of the mixture with estimated median at $\hat{\mu}_1 = 0.2677$. The estimated median for the municipalites from the second mixture compontent was $\hat{\mu}_2 = 0.6491$. With the application, it was possible to verify that the UWUW performance may overcome other two-component mixture models based on other widely known unit distributions such as the beta and Kumaraswamy.

3 UNIT REGRESSION MODELS TO EXPLAIN VOTE PROPORTIONS IN THE BRAZILIAN PRESIDENTIAL ELECTIONS IN 2018

3.1 INTRODUCTION

Double-bounded variables, such as rates and proportions, usually show asymmetry, and heavy tails. Thus, the assumption of normality becomes inadequate. For these situations, one should look for more flexible models that provide adequate representations for the physiological properties and the empirical distribution of the data (PEREIRA; SOUZA; CRIBARI-NETO, 2014). The beta (FERRARI; CRIBARI-NETO, 2004) and simplex (BARNDORFF-NIELSEN; JØRGENSEN, 1991) regressions are common procedures to accommodate these features. These models resemble the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) (STASINOPOULOS; RIGBY; BASTIANI, 2018) by allowing a regression structure in the mean and the dispersion (or precision) parameter of the beta and simplex distributions, respectively. The GAMLSS allows all parameters of the response distribution to vary with explanatory variables and provides a comprehensive framework for easily incorporating nonlinear, random and spatial effects.

In this chapter, our objective is to explain the mean variations of the vote proportions received by Jair Bolsonaro in the runoff of the 2018 presidential elections. Some works have been used beta regression for previous Brazilian presidential elections. Andrade *et al.* (2013) assessed the impacts of welfare programs and economic growth on the outcome of the 2006 election. Junior and Souza (2015) investigated it in the 2010 election and for the Northeast region. For the 2018 elections, Hunter and Power (2019), and Rennó (2020) carried out some analysis on the political landscape. Yero, Sacco and Nicoletti (2020) analyzed the effects of development indicators in the 2018 elections using machine learning algorithms. They reported that voters residing in less developed regions have left-wing parties as to the preferred choice. However, to our best knowledge, a unit regression analysis modeling the 2018 elections has no been carried out.

In this context, we fit beta and simplex regressions since they are parameterized in terms of the mean and dispersion parameters and are available in the `R` `gamlss` package (RIGBY; STASINOPOULOS, 2005). Alternative unit regressions have been proposed in recent years, but most of them focus on quantile parametrizations. We can cite, for example, Bayes, Bazán and Castro (2017) for quantile regression based on the Kumaraswamy distribution and Lemonte and Bazán (2016) for median regression from the Johnson S_B distribution. Other recent advances are Smithson and Shou (2017), Mazucheli *et al.* (2020), and Pumi, Rauber and Bayer (2020). Those models are not considered in the analysis because our interest lies in the effect of socio-demographic and economic indicators in the mean of the vote proportions.

The rest of this chapter is divided as follows. Section 3.2 presents a theoretical background on the beta and simplex regression models. In Section 3.3, a descriptive analysis and the data preparation is presented. Section 3.4 discusses the fitted regressions and the effects of the explanatory variables in the vote proportions data collected. The concluding remarks are outlined in Section 3.5.

3.2 THEORETICAL BACKGROUND

This section aims to discuss the unit regression models employed in the vote proportion analysis. We present some aspects of parameter estimation, model selection, and diagnostic analysis on these models. Ferrari and Cribari-Neto (2004) proposed a class of beta regression models in which the mean response is related to a linear predictor, which involves covariates and unknown regression parameters, through a link function. The model is also indexed by a precision parameter and assumes that the dependent variable has a beta distribution.

The beta regression model was considered by several authors. For example, Bayer, Tondolo and Müller (2018) proposed beta regression control charts to monitor the tire manufacturing process and the relative humidity in Brasília, Brazil. Ghosh (2019) presented a study on robust inference for the model, with application to health studies. Karlsson, Månsson and Kibria (2020) introduced a Liu estimator for the beta regression model and performed an application to chemical data. Espinheira *et al.* (2019) investigated model selection criteria on beta regression for machine learning.

Assuming that the precision parameter is also related to a linear predictor, Simas, Barreto-Souza and Rocha (2010) formally introduced the varying precision beta regression model. Moreover, an alternative parametrization in terms of a dispersion (not precision) parameter is considered by Cribari-Neto and Souza (2012), Bayer and Cribari-Neto (2017) and Canterle and Bayer (2019). Let Y be a beta random variable indexed by the mean $\mu \in (0,1)$ and the dispersion parameter $\sigma \in (0,1)$, denoted as $Y \sim \text{Beta}(\mu,\sigma)$. Its probability density function is (for $y \in (0,1)$)

$$f(y;\mu,\sigma) = \frac{\Gamma\left(1/\sigma^2-1\right)}{\Gamma\left(\mu(1/\sigma^2-1)\right)\Gamma\left((1-\mu)(1/\sigma^2-1)\right)} y^{\mu(1/\sigma^2-1)-1}(1-y)^{(1-\mu)(1/\sigma^2-1)-1}, \quad (3.1)$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x$ is the gamma function. Under this parameterization the variance of Y is $\mathrm{Var}(Y) = \sigma^2\mu(1-\mu)$. It should be noted that this reparametrization is the one implemented for the beta density in the `R gamlss` package (RIGBY; STASINOPOULOS; VOUDOURIS, 2013; RIGBY; STASINOPOULOS, 2005; STASINOPOULOS *et al.*, 2017).

The simplex distribution was introduced by Barndorff-Nielsen and Jørgensen (1991) as an alternative to the beta distribution. Let $Y \sim S(\mu, \sigma^2)$ be a simplex random variable, which density is (for $y \in (0,1)$)

$$f(y;\mu,\sigma^2) = \{2\pi\sigma^2[y(1-y)]^3\}^{-1/2} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2\mu^2 y(1-y)(1-\mu)^2}\right\}, \quad (3.2)$$

where $0 < \mu < 1$ is the mean of Y and σ^2 is a dispersion parameter. Espinheira and Silva (2019) proposed a general class of simplex regression in which the mean and the disperson parameters can be related to a linear predictor. They also perform residual and influence analysis to the simplex regression.

Several studies have been developed considering simplex regression models. Carrasco and Reid (2019) studied measurement errors. López (2013) considered a Bayesian approach to parameter estimation in a simplex regression model and compared with beta regression. Cordeiro *et al.* (2020) used the beta and simplex regression models to explain homicides in state Brazilian capitals. The simplex regression is also implemented in `R gamlss` package (RIGBY; STASINOPOULOS; VOUDOURIS, 2013; RIGBY; STASINOPOULOS, 2005; STASINOPOULOS *et al.*, 2017).

Let $y_1, \ldots, y_n$ be a set of independent random variables such that each $y_t \sim \mathrm{D}(\mu_t, \sigma_t)$, where $t = 1, \ldots, n$, and $\mathrm{D}(\mu_t, \sigma_t)$ denotes a two-parameter distribution in which both μ_t and σ_t are linear functions of the explanatory variables. We consider the GAMLSS structure (STASINOPOULOS *et al.*, 2017) to define the following systematic components,

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i = \eta_t,$$

$$h(\sigma_t) = \sum_{j=1}^{q} z_{tj}\gamma_j = \upsilon_t, \quad (3.3)$$

where $\beta = (\beta_1, \ldots, \beta_k)^\top \in \mathbb{R}^k$ e $\gamma = (\gamma_1, \ldots, \gamma_q)^\top \in \mathbb{R}^q$ are vectors of unknown parameters, $x_{t1}, \ldots, x_{tk}$ and $z_{t1}, \ldots, z_{tq}$ are k and q covariate observations $(k+q) < n$, respectively, assumed to be fixed and known, $\eta_t = \sum_{i=1}^{k} x_{ti}\beta_i$ and $\upsilon_t = \sum_{j=1}^{q} z_{tj}\gamma_j$ are the linear predictors, and $g(\cdot)$ and $h(\cdot)$ are strictly monotonous and twice differentiable link functions.

Thus, both beta and simplex regression models are defined from Equation (3.3), with $g : (0,1) \to \mathbb{R}$, but differing on the assumption of the random component and the mapping required for the dispersion link function. The beta regression has the random compontent following (3.1), in which $h : (0,1) \to \mathbb{R}$, and the simplex regression has $\mathrm{D}(\mu_t, \sigma_t)$ following (3.2), in which $h : \mathbb{R}^{+} \to \mathbb{R}$.

The link functions that can be considered for the parameters in the unit range are, for instance: i) logit: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$; ii) probit: $g(\mu) = \Phi^{-1}(\mu)$, where $\Phi(\cdot)$ is the cumulative distribution of a standard normal random variable; iii) log-log(loglog): $g(\mu) = -\log[-\log(\mu)]$; iv) complement log-log (cloglog) : $g(\mu) = \log[-\log(1-\mu)]$, and v) Cauchy: $g(\mu) = \tan[\pi(\mu - 0.5)]$. Regarding the dispersion parameter of the simplex regression, we can use the logarithmic function $h(\sigma) = \log\sigma$, and the identity function $h(\sigma) = \sigma$. McCullagh and Nelder (1989) provide a detailed study of some of these link functions.

3.3 DATA PREPARATION AND DESCRIPTIVE ANALYSIS

The variable of interest is the proportion of valid votes received by Jair Bolsonaro in the 2018 presidential elections runoff (Prop_PSL). We consider the data from Brazilian municipalities with a population greater than 300,000 in the 2010 census, totaling 79 observations. Some socio-demographic indicators and the political spectrum of the governors' party are selected as explanatory variables for the vote proportions. The variables were collected using public data from the Brazilian Tribunal Superior Eleitoral (TSE) (TSE, 2018), Instituto Brasileiro de Geografia e Estatística (IBGE) (SIDRA, 2020), and Instituto de Pesquisa Econômica Aplicada (IPEADATA) (BRASIL, 2020). Table 4 lists all variables and their respective sources. Since the variables Region, PG_2014, PG_2018, and Cap_BR are qualitative, they are included in the regression analysis using dummy variables. We define the Midwest region as the reference category for the Region variables. For PG_2014 and PG_2018, the reference is the centre spectrum. Finally, the Cap_BR dummy is defined as a variable that equals one if the municipality is a capital and zero otherwise.

Table 4 – Summary of the variables

Variable	Source	Description
Prop_PSL	TSE	Proportion of valid votes recieved by Jair Bolsonaro in the 2018 presidential elections runoff.
EP	IBGE	Estimated population in the 2010 census, per 100,000.
PE	IBGE	Proportion of evangelicals in the 2010 census.
PEAW	IBGE	Proportion of economically active women (2010 census).
LR	IBGE	Literacy rate in the 2010 census.
MHIC	IBGE	Monthly household income per capita in reals, R$, in the 2010 census.
DD	IBGE	Demographic density in the 2010 census.
Region	-	Brazilian region to which the city belongs (South, Northeast, Southeast, North, and Midwest).
PG_2014	IPEADATA and (SARDINHA; COSTA, 2020)	Political spectrum of the governors' party elected in 2014 (left-wing, centre, right-wing).
PG_2018	IPEADATA and (SARDINHA; COSTA, 2020)	Political spectrum of the governors' party elected in 2018 (left-wing, centre, right-wing).
Cap_BR	-	Brazilian capital to which the city belongs.

Source: Author (2021)

The variable MHIC is calculated as the ratio between total family income-in nominal terms-and the total number of residents. Income from work and other sources for all residents is considered, including those classified as retired, domestic workers, and relatives of domestic workers (IBGE, 2010). We also analyze the governors' party elected in the Brazilian state where the cities belong. Since the Brazilian political system has many parties, the variables PG_2014 and PG_2018 are defined from political spectrum of the governors' party elected in 2014 and 2018 elections, respectively. The political spectrum of the party is obtained from Sardinha and Costa (2020), and the governors' party from (BRASIL, 2020). The variable Region is defined since several authors carried out studies on the impact of the Brazilian regions on presidential elections, see Junior and Souza (2015), Jr (2013), and Jr (2015), for instance.

Table 5 gives some descriptive measures of the response variable and quantitative covariates. The coefficient of variation (CV) indicates that EP has the higher variability and LR has the lower degree of variability. The mean proportion of votes is 0.63, and the amplitude is 0.53. Since the median is 0.66, the elected candidate won in most municipalities considered. We have negative skewness and kurtosis coefficients, indicating that the larger vote proportion values have a fatter tail than the smallest ones. These features can be confirmed through the plots displayed in Figure 8.

From the boxplot of Prop_PSL (Figure 8), we highlight that the cities with the lowest values in relation to the vote proportion are Salvador (Bahia), Caucaia (Ceará), Feira de Santana (Bahia), and Teresina (Piauí), which registered values at 0.314, 0.3667, 0.3715 and 0.3726, respectively. It is noteworthy that all the city are located in the Northeast of Brazil, and Caucaia is also between the three cities with the lowest MHIC. Besides, Blumenau (Santa Catarina) is the city with the highest values for Prop_PSL and LR.

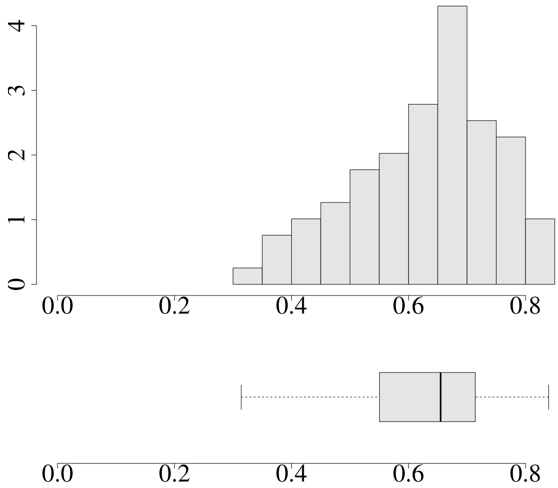
Table 5 – Descriptive statistics for the response variable and quantitative covariates.

Variable	Mean	Median	Skewness	Kurtosis	Min.	Max.	CV
Prop_PSL	0.628	0.655	−0.527	−0.368	0.314	0.840	19.279
EP	9.020	4.720	5.378	32.976	3.005	112.535	160.160
PE	0.267	0.255	−0.024	−0.291	0.117	0.410	24.688
PEAW	0.235	0.235	−0.070	−0.648	0.182	0.284	9.164
MHIC	1,100.113	1,042.840	0.723	−0.172	439.720	2,159.170	36.127
DD	2,515.933	1,315.270	1.722	2.490	12.570	13,024.560	119.971
LR	95.328	96.300	−1.710	2.732	85.500	98.300	2.741

Source: Author (2021)

Figure 9 displays dispersion plots of the response variable versus the quantitative covariates. Since there is no indication of a linear relationship between the variables, we use the Spearman method to compute the correlation matrix (see Table 6). The results indicate that the response

Figure 8 – Histogram and boxplot of the vote proportion data.



Source: Author (2021)

variable is positively correlated with most variables. The highest correlation is with LR, followed by the MHIC. This high correlation shows that covariates are important for the study, as they indicate that LR and MHIC had are associated with the response variable.

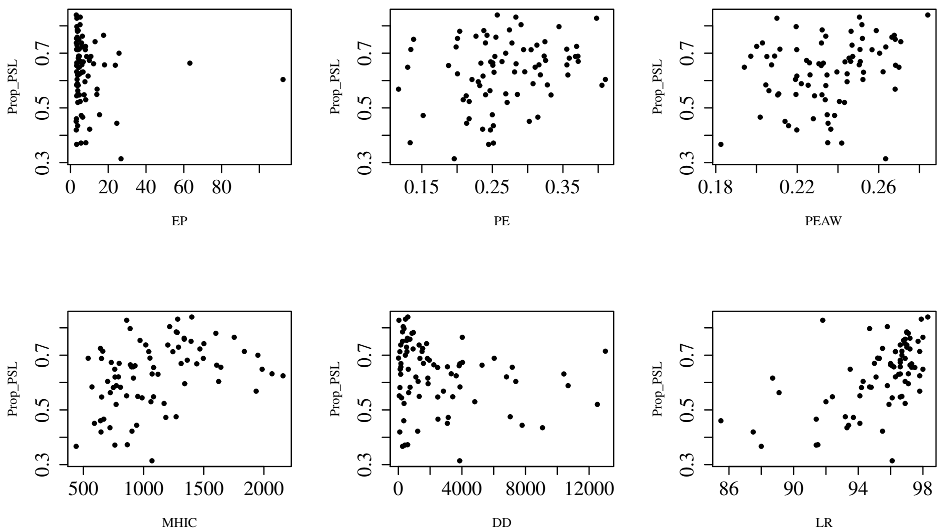
Table 7 presents the frequency distribution for the qualitative variables. About 50% of the municipalities in the study belong to the Southeast region, which is also the most developed. The North and Northeast are the most impoverished and concentrate a relative frequency of 15.19%. In 2014 and 2018, the parties with centre ideology elected more governors. Notice the growing number of right-wing governors from 2014 to 2018. It increased approximately 26.58%.

Table 6 – Spearman's correlation matrix with their respective p-values in parenthesis.

	Prop_PSL	EP	PE	PEAW	MHIC	DD	LR
Prop_PSL							
EP	−0.10						
	(0.3650)						
PE	0.23	−0.09					
	(0.0382)	(0.4403)					
PEAW	0.26	0.24	−0.42				
	(0.0183)	(0.0363)	(0.0001)				
MHIC	0.41	0.30	−0.46	0.79			
	(0.0002)	(0.0072)	(0.0002)	(< 0.0001)			
DD	−0.24	0.34	0.01	0.05	0.02		
	(0.0313)	(0.0022)	(0.9587)	(0.6824)	(0.8582)		
LR	0.56	0.12	−0.24	0.62	0.76	0.11	
	(< 0.0001)	(0.2765)	(0.0316)	(< 0.0001)	(< 0.0001)	(0.3325)	

Source: Author (2021)

Figure 9 – Scatter diagram.



Source: Author (2021)

Table 7 – Frequency distribution for the qualitative variables.

Variable	Frequency	Relative frequency (%)
Region		
Northeast	17	21.5189
South	11	13.9241
Southeast	39	49.3672
Northern	6	7.5949
Midwest	6	7.5949
PG_2014		
Left-wing	32	40.5063
Centre	39	49.3671
Right-wing	8	10.1266
PG_2018		
Left-wing	20	25.3164
Centre	30	37.9747
Right-wing	29	36.7089

Source: Author (2021)

3.4 FITTED REGRESSIONS

This section presents and discusses the fitted regression models considering the Prop_PSL as the dependent variable. For both simplex and beta regressions, parameter estimation is performed by the maximum likelihood method using the Rigby and Stasinopoulos (RS) algorithm (STASINOPOULOS *et al.*, 2017). We use the Akaike information criterion (AIC), Schwarz's Bayesian criteria (SBC), and stepwise technique for model selection. After evaluating the regression subsets and selecting those with the smallest AIC and SBC, we define the systematic components for both classes of regressions and eliminate the non-significant predictors.

Table 8 – Results from the fitted beta and simplex regressions.

	Beta (μ_i, σ_i)			Simplex(μ_i, σ_i)		
Variable	$\hat{\mu}$	Std. Error	p-value	$\hat{\mu}$	Std. Error	p-value
Intercept	−0.7070	0.1695	0.0001	−0.7174	0.1760	0.0001
EP	-	-	-	−0.0030	0.0016	0.0592
MHIC	0.0004	0.0001	< 0.0001	0.0005	0.0001	< 0.0001
PE	3.6723	0.4324	< 0.0001	3.6378	0.4345	< 0.0001
DD	−0.0003	0.0001	0.0026	−0.0002	0.0001	0.0429
South	0.2580	0.1032	0.0149	0.3286	0.0926	0.0007
PG_2014_Left	−0.1996	0.0751	0.0099	−0.2431	0.0689	0.0008
PG_2018_Left	−0.4372	0.0679	< 0.0001	−0.4472	0.0687	< 0.0001
Variable	$\hat{\sigma}$	Std. Error	p-value	$\hat{\sigma}$	Std. Error	p-value
Intercept	−11.1583	3.1705	0.0008	0.8580	0.4838	0.0807
PE	−3.3198	1.5630	0.0374	−5.2280	1.6409	0.0022
LR	0.1126	0.0343	0.0017	-	-	-
North	-	-	-	1.1931	0.3326	0.0006
South	−0.8738	0.3271	0.0095	-	-	-
Southeast	−0.9463	0.2513	0.0004	-	-	-
Cap_BR	−0.9004	0.2795	0.0020	−0.7768	0.2553	0.0033

Source: Author (2021)

Table 9 – Values of AIC, SBC, pseudo-R², and SW p-value for the beta and simplex models.

Modelo	AIC	SBC	R^2_G	SW
Beta (μ_i, σ_i)	−211.5764	−178.3964	0.7939	0.3113
Simplex (μ_i, σ_i)	−219.0181	−188.2081	0.8025	0.5797

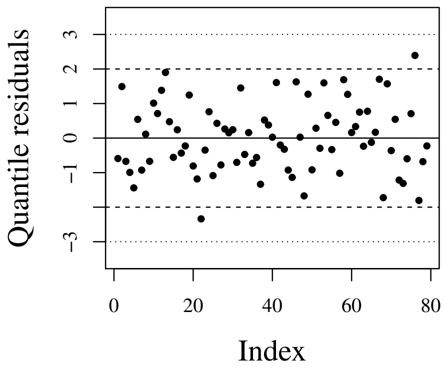
Source: Author (2021)

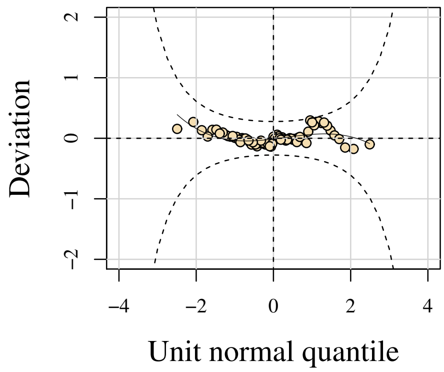
Table 8 lists the parameter estimates, standard errors (Std. Error), and p-values of the final fitted models. Both include the mean and the dispersion submodels. Most of the covariates are significant at 5%, except for EP in the simplex mean submodel, which is significant at 10%. The AIC, SBC and pseudo-R^2 (RAO *et al.*, 1973) statistics are considered as goodness-of-fit measures. Residual analysis is performed through quantile residuals, and we also compute the p-value for the Shapiro-Wilk (SW) test to verify the assumption of normality for the residuals. From reported in Table 9, the null hypothesis that the residual distribution is normal is not rejected at a significance level of 5% for both fitted models but the simplex regression was superior to the beta in all the considered statistics. Thus, there are evidences that the simplex regression is better suited to the current data.

Figure 10 displays plots of the residuals versus the index, wormplot (BUUREN; FREDRIKS, 2001) and quantile-quantile plot (QQ-plot) for the simplex regression quantile residuals. These plots confirm that this model is suitable for modeling the vote proportions.

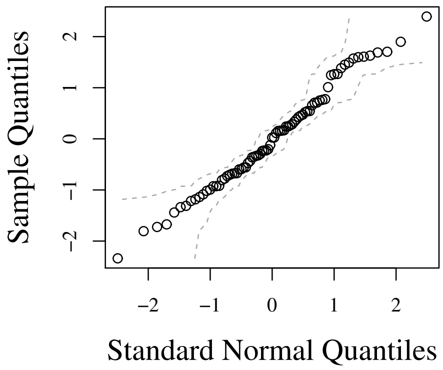
Figure 10 – Residual analysis for the final simplex regression.

(a) Residuals versus index. (b) Wormplot.





(c) QQ-plot.



Source: Author (2021)

Therefore, we select the following simplex regression to explain the variable of interest,

$$\log[\hat{\mu}_t/(1-\hat{\mu}_t)] = -0.7174 - 0.0030\,\text{EP} + 0.0005\,\text{MHIC} + 3.6378\,\text{PE} - 0.0002\,\text{DD} \\ + 0.3286\,\text{South} - 0.2431\,\text{PG\_2014\_Left} - 0.4472\,\text{PG\_2018\_Left}, \qquad (3.4)$$

and

$$\log(\hat{\sigma}_t) = 0.8580 - 5.2280\,\text{PE} + 1.1931\,\text{North} - 0.7768\,\text{Cap\_BR}. \qquad (3.5)$$

Some findings of the vote proportion can follow from Equations (3.4) and (3.5).

- The MHIC and the South region have a positive impact on the vote proportions' mean. These results are consistent with Hunter and Power (2019) which indicate income among the support indicators for Bolsonaro's election. The candidate won among all income groups, except for the poor and very poor.

- The PE is a significant covariate, and the average vote proportions tend to be higher in the municipalities with a more evangelical population. It is also significant about the dispersion, and the tends to decrease as the proportion of evangelicals increases.
- PG_2014_Left and PG_2018_Left had negative influence in the vote proportions' mean. In fact, Bolsonaro adopted a position of reducing state intervention in the economy in his campaign, and right-wing voters in Latin America have traditionally been against state intervention in the economy (RENNÓ, 2020).
- The North region is positively related to the dispersion of the variable of interest, i.e., the municipality on this region present vote proportion more disperse than those from other regions. Finally, the dispersion tends to decrease for the capitals.

3.5 CONCLUSION

In this chapter, we conducted a study on unitary regression models to quantify the effect of explanatory variables on the proportion of votes of Jair Bolsonaro in the second round of the 2018 presidential elections. We considered data from Brazilian municipalities with a population greater than 300,000 and verified that the elected candidate won in most cities considered. The vote proportion distribution is left-skewed with a few outliers in the left tail. All of them are located in the Northeast region. As explanatory variables, we select some socio-demographic indicators and the political spectrum of the governors' party in the 2014 and 2018 elections. We constructed beta and simplex regressions with systematic components for the mean and dispersion parameters using the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) framework. Both fitted regressions are estimated by the maximum likelihood method using the GAMLSS R package. The Akaike information criterion (AIC), Schwarz's Bayesian criteria (SBC), and pseudo-R^2 statistics were considered as goodness-of-fit statistics and the quantile residuals were analyzed as a diagnostic tool. We concluded that the simplex regression is superior to the beta and is suitable for modeling the variable of interest. We evaluated significant effects for the monthly household income per capita, the proportion of evangelicals, and the political spectrum of the governors' party elected in 2014 and 2018. We also verify that some Brazilian regions impact the vote proportions' mean and dispersion.

4 CONCLUDING REMARKS

This masterthesis investigated the unit models for analyzing vote proportions, defined as the district vote shares by the total number of valid votes cast in the district. The motivating data set is vote proportions in Brazilian presidential elections runoff in 2018. The objective of this masterthesis is twofold and is presented in two main independent chapters. In Chapter 2, we defined a two-component mixture model to describe population heterogeneities. The unit Weibull two-component mixture model (UWUW) is a new distribution formulated considering that each mixture's component follows the unit Weibull distribution. Some of the main properties of UWUW are presented, such as ordinary and incomplete moments. The Expectation Maximization algorithm is used to obtain maximum likelihood estimates for the model parameters, and Monte Carlo simulations are performed. An application to electoral data illustrates the importance and potential of the new model. The results indicate that our proposal is adequate to fit this data set since it is adequate to analyze asymmetric and bimodal behaviors. With the application, it was possible to verify that the performance of UWUW can surpass other two-component mixture models based on other widely known unit distributions such as beta and Kumaraswamy.

Chapter 3 carries out a regression study to explain the mean variations of the vote proportions received by Jair Bolsonaro in the runoff of the 2018 presidential elections. We fit beta and simplex regressions since they are parameterized in terms of the mean and dispersion parameters and are available in the `R` programming language. These models resemble the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) by allowing a regression structure in the mean and the dispersion parameter. The simplex regression is superior to the beta and is suitable for modeling the variable of interest. We evaluate significant effects for some socio-demographic indicators, and the political spectrum of the governors' party are selected as explanatory variables for the vote proportions.

In future work, we intend to address the following problems:

- Introduce mixture distributions based on other unit models such as the submodels of the unit extended Weibull family.
- Propose the UWUW quantile-regression model with systematic components for the quantile parameter of each mixture component.
- Analyze the effect of explanatory variables on quantile variations of the vote proportions received by Jair Bolsonaro in the runoff of the 2018 presidential elections.

BIBLIOGRAPHY

ABENSUR, T. C.; CRIBARI-NETO, F.; MENEZES, T. A. *et al.* Impactos do programa bolsa família nos resultados das eleições presidenciais no Brasil em 2006. **Anais do XXXV Encontro Nacional de Economia**, Anpec Recife, v. 51, 2007.

ALEMÁN, E.; KELLAM, M. The nationalization of presidential elections in the americas. **Electoral Studies**, Elsevier, v. 47, p. 125–135, 2017.

ALENCAR, E. R. d. Discriminante não-linear para mistura de distribuições beta. 2018.

ANDRADE, M. V.; NORONHA, K. V. M. d. S.; MENEZES, R. d. M.; SOUZA, M. N.; REIS, C. d. B.; MARTINS, D. R.; GOMES, L. Desigualdade socioeconômica no acesso aos serviços de saúde no Brasil: um estudo comparativo entre as regiões brasileiras em e 2008. **Economia Aplicada**, SciELO Brasil, v. 17, p. 623–645, 2013.

BARNDORFF-NIELSEN, O. E.; JØRGENSEN, B. Some parametric models on the Simplex. **Journal of Multivariate Analysis**, Elsevier, v. 39, p. 106–116, 1991.

BAYER, F. M.; CRIBARI-NETO, F. Model selection criteria in beta regression with varying dispersion. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 46, p. 729–746, 2017.

BAYER, F. M.; TONDOLO, C. M.; MÜLLER, F. M. Beta regression control chart for monitoring fractions and proportions. **Computers & Industrial Engineering**, Elsevier, v. 119, p. 416–426, 2018.

BAYES, C. L.; BAZáN, J. L.; CASTRO, M. D. A quantile parametric mixed regression model for bounded response variables. **Statistics and its Interface**, International Press of Boston, v. 10, p. 483–493, 2017.

BOUGUILA, N.; ZIOU, D.; MONGA, E. Practical Bayesian estimation of a finite beta mixture through Gibbs sampling and its applications. **Statistics and Computing**, Springer, v. 16, p. 215–225, 2006.

BRASIL. Eleições. **EPEADATA. Available in: http://www.ipeadata.gov. br/Default.aspx. Accessed on: Dez**, 2020.

BUČAR, T.; NAGODE, M.; FAJDIGA, M. Reliability approximation using finite Weibull mixture distributions. **Reliability Engineering & System Safety**, Elsevier, v. 84, p. 241–251, 2004.

BUUREN, S. v.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistics in Medicine**, Wiley Online Library, v. 20, p. 1259–1277, 2001.

CANTERLE, D. R.; BAYER, F. M. Variable dispersion beta regressions with parametric link functions. **Statistical Papers**, Springer, v. 60, p. 1541–1567, 2019.

CARDOSO, M.; MENDES, R.; SOUZA, J.; RIBEIRO, H. Gender difference in candidature processes for Brazilian elections. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 537, p. 122–525, 2020.

CARRASCO, J. M.; REID, N. Simplex regression models with measurement error. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 1, p. 1–16, 2019.

CARVALHO, T. M. d. Modelagem da LGD via mistura de distribuições Kumaraswamy. 2015.

CHEN, G.; BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test. **Journal of Quality Technology**, Taylor & Francis, v. 27, p. 154–161, 1995.

CORDEIRO, G. M.; ROCHA, E.; FIGUEIREDO, D.; FERNANDES, A.; ORTEGA, E. M.; PRATAVIERA, F. The beta and simplex regression models to explain homicides in state capitals of Brazil. **Model Assisted Statistics and Applications**, IOS Press, v. 15, p. 215–224, 2020.

CRIBARI-NETO, F.; SOUZA, T. C. Testing inference in variable dispersion beta regressions. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 82, p. 1827–1843, 2012.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 39, p. 1–22, 1977.

DURBIN, J.; KNOTT, M. Components of Cramér-Von Mises statistics. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 34, p. 290–307, 1972.

ESPINHEIRA, P. L.; SILVA, A. d. O. S. Residual and influence analysis to a general class of simplex regression. **TEST**, Springer, v. 1, p. 1–30, 2019.

ESPINHEIRA, P. L.; SILVA, L. C. M. da; SILVA, A. d. O.; OSPINA, R. Model selection criteria on beta regression for machine learning. **Machine Learning and Knowledge Extraction**, Multidisciplinary Digital Publishing Institute, v. 1, p. 427–449, 2019.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 7, p. 799–815, 2004.

FILHO, R. C.; ALMEIDA, M.; ANDRADE, J.; MOREIRA, J. *et al.* Scaling behavior in a proportional voting process. **Physical Review E**, APS, v. 60, p. 1067, 1999.

FILHO, R. C.; ALMEIDA, M.; MOREIRA, J.; JR, J. A. Brazilian elections: voting for a scaling democracy. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 322, p. 698–700, 2003.

GHOSH, A. Robust inference under the beta regression model with application to health care studies. **Statistical Methods in Medical Research**, SAGE Publications Sage UK: London, England, v. 28, p. 871–888, 2019.

GOODMAN, L. A. Kolmogorov-Smirnov tests for psychological research. **Psychological Bulletin**, American Psychological Association, v. 51, p. 160, 1954.

GRÜN, B.; KOSMIDIS, I.; ZEILEIS, A. **Extended beta regression in R: shaken, stirred, mixed, and partitioned.** [S.l.], 2011.

GUERRA, R. R.; PEñA-RAMíREZ, F. A.; BOURGUIGNON, M. The unit extended Weibull families of distributions and its applications. **Journal of Applied Statistics**, Taylor & Francis, v. 1, p. 1–19, 2020.

HUANG, W.; DONG, S. Probability distribution of wave periods in combined sea states with finite mixture models. **Applied Ocean Research**, Elsevier, v. 92, p. 101938, 2019.

HUNTER, W.; POWER, T. J. Bolsonaro and Brazil's illiberal backlash. **Journal of Democracy**, Johns Hopkins University Press, v. 30, p. 68–82, 2019.

IBGE. Censo demográfico 2010. **IBGE: Insituto Brasileiro de Geografia e Estatística**, 2010.

JEWELL, N. P. Mixtures of exponential distributions. **The Annals of Statistics**, JSTOR, p. 479–484, 1982.

JI, Y.; WU, C.; LIU, P.; WANG, J.; COOMBES, K. R. Applications of beta-mixture models in bioinformatics. **Bioinformatics**, Oxford University Press, v. 21, p. 2118–2122, 2005.

JIANG, R.; MURTHY, D. Mixture of Weibull distributions-parametric characterization of failure rate function. **Applied Stochastic Models and Data Analysis**, Wiley Online Library, v. 14, p. 47–65, 1998.

JR, C. Z. When payouts pay off: Conditional cash transfers and voting behavior in brazil 2002–10. **American Journal of Political Science**, Wiley Online Library, v. 57, p. 810–822, 2013.

JR, C. Z. The impacts of conditional cash transfers in four presidential elections (2002–2014). **Brazilian Political Science Review**, SciELO Brasil, v. 9, p. 135–149, 2015.

JUNIOR, P. M. d. A.; SOUZA, T. C. Estimativas de votos da presidente Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do bolsa família. **Ciência e Natura**, Universidade Federal de Santa Maria, v. 37, p. 12–22, 2015.

KARLSSON, P.; MÅNSSON, K.; KIBRIA, B. G. A Liu estimator for the beta regression model and its application to chemical data. **Journal of Chemometrics**, Wiley Online Library, v. 34, p. e3300, 2020.

KHALID, M.; ASLAM, M.; SINDHU, T. N. Bayesian analysis of 3-components Kumaraswamy mixture model: Quadrature method vs. importance sampling. **Alexandria Engineering Journal**, Elsevier, v. 59, p. 2753–2763, 2020.

KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, v. 46, p. 79–88, 1980.

LACHOS, V. H.; MORENO, E. J. L.; CHEN, K.; CABRAL, C. R. B. Finite mixture modeling of censored data using the multivariate Student-t distribution. **Journal of Multivariate Analysis**, Elsevier, v. 159, p. 151–167, 2017.

LEMONTE, A. J.; BAZáN, J. L. New class of Johnson distributions and its associated regression model for rates and proportions. **Biometrical Journal**, Wiley Online Library, v. 58, p. 727–746, 2016.

LÓPEZ, F. O. A Bayesian approach to parameter estimation in simplex regression model: A comparison with beta regression. **Revista Colombiana de Estadística**, Universidad Nacional de Colombia., v. 36, p. 1–21, 2013.

LYRA, M.; COSTA, U.; FILHO, R. C.; JR, J. A. Generalized Zipf's law in proportional voting processes. **EPL (Europhysics Letters)**, IOP Publishing, v. 62, p. 131, 2003.

MARINHO, P. R. D.; SILVA, R. B.; BOURGUIGNON, M.; CORDEIRO, G. M.; NADARAJAH, S. Adequacymodel: An R package for probability distributions and general purpose optimization. **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, p. e0221487, 2019.

MAZUCHELI, J.; MENEZES, A.; GHITANY, M. The unit-Weibull distribution and associated inference. **Journal of Applied Probability and Statistics**, v. 13, p. 1–22, 2018.

MAZUCHELI, J.; MENEZES, A. F. B.; FERNANDES, L. B.; OLIVEIRA, R. P. de; GHITANY, M. E. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. **Journal of Applied Statistics**, Taylor & Francis, v. 47, p. 954–974, 2020.

MCCULLAGH, P.; NELDER, J. A. Generalized linear models 2nd edition chapman and hall. **London, UK**, 1989.

MITNIK, P. A.; BAEK, S. The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation. **Statistical Papers**, v. 54, p. 177–192, 2013.

PAZ, R. F. d.; EHLERS, R. S.; BAZÁN, J. L. A Weibull mixture model for the votes of a Brazilian political party. In: **Interdisciplinary Bayesian Statistics**. [S.l.]: Springer, 2015. p. 229–241.

PEARSON, K. Contributions to the mathematical theory of evolution. **Philosophical Transactions of the Royal Society of London, A.**, JSTOR, v. 185, p. 71–110, 1894.

PEREIRA, T. L.; SOUZA, T. C.; CRIBARI-NETO, F. Uma avaliação da eficiência do gasto público nas regiões do Brasil. **Ciência e Natura**, Universidade Federal de Santa Maria, v. 36, p. 23–36, 2014.

POWELL, E. N.; TUCKER, J. A. Revisiting electoral volatility in post-communist countries: new data, new results and new approaches. **British Journal of Political Science**, v. 44, p. 123–147, 2013.

PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P. **Numerical recipes 3rd edition: The art of scientific computing**. [S.l.]: Cambridge University Press, 2007.

PUMI, G.; RAUBER, C.; BAYER, F. M. Kumaraswamy regression model with Aranda-Ordaz link function. **TEST**, Springer, p. 1–21, 2020.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Available at: <https://www.R-project.org/>.

RAO, C. R.; RAO, C. R.; STATISTIKER, M.; RAO, C. R.; RAO, C. R. **Linear statistical inference and its applications**. [S.l.]: Wiley New York, 1973.

REDNER, R. A.; WALKER, H. F. Mixture densities, maximum likelihood and the EM algorithm. **SIAM Review**, SIAM, v. 26, p. 195–239, 1984.

RENNÓ, L. The Bolsonaro voter: Issue positions and vote choice in the 2018 Brazilian presidential elections. **Latin American Politics and Society**, Cambridge University Press, v. 62, p. 1, 2020.

RIGBY, R.; STASINOPOULOS, D.; VOUDOURIS, V. Discussion: A comparison of gamlss with quantile regression. **Statistical Modelling**, Sage Publications Sage India: New Delhi, India, v. 13, p. 335–348, 2013.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape,(with discussion). **Applied Statistics**, v. 54, p. 507–554, 2005.

SARDINHA, E.; COSTA, S. Direita cresce e engole o centro no congresso mais fragmentado da história. **Congressoemfoco. Available in: https://congressoemfoco.uol.com.br/legislativo/direita-cresce-e-engole-o-centro-no-congresso-mais-fragmentado-da-historia/. Accessed on: Dez**, 2020.

SIDRA. Censo demográfico. **Sistema IBGE de Recuperação Automática - SIDRA. Available in: https://sidra.ibge.gov.br/tabela/3974. Accessed on: Dez**, 2020.

SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 54, p. 348–366, 2010.

SMITHSON, M.; SHOU, Y. Cdf-quantile distributions for modelling random variables on the unit interval. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 70, p. 412–438, 2017.

STASINOPOULOS, M. D.; RIGBY, R. A.; BASTIANI, F. D. Gamlss: a distributional regression approach. **Statistical Modelling**, SAGE Publications Sage India: New Delhi, India, v. 18, p. 248–273, 2018.

STASINOPOULOS, M. D.; RIGBY, R. A.; HELLER, G. Z.; VOUDOURIS, V.; BASTIANI, F. D. **Flexible regression and smoothing: using GAMLSS in R**. [S.l.]: CRC Press, 2017.

TSE. Electoral data repository. **Results. Presents results of Brazilian general elections. Available in: http://english.tse.jus.br/the-brazilian-electoral-system/statistics. Accessed on: May**, 2018.

YERO, E. J. H.; SACCO, N. C.; NICOLETTI, M. do Carmo. Effect of the municipal Human Development index on the results of the 2018 Brazilian Presidential elections. **Expert Systems with Applications**, Elsevier, v. 1, p. 114–305, 2020.

APPENDIX A – Simulation study

In this Section, Monte Carlo simulations are presented to evaluate the performance of the EM algorithm for estimating the parameter vector of the UWUW model. It is noteworthy that the results of simulations are presented for various combinations of parameters. In which the values of τ were varied, considering τ = c (0.25, 0.5, 0.75).

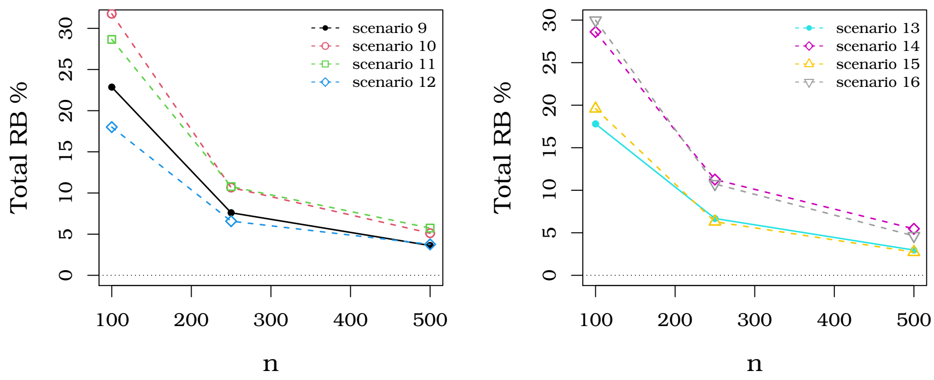
Table 10 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model, $\tau = 0.5$.

Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$
9	0.2	0.1	2.5	0.9	0.4	100	0.3975	0.5491	19.2335	2.6651	0.0161	0.0021	0.0019	0.4928	0.0131	0.0007
						250	0.1689	0.1725	6.2358	1.0401	-0.0172	0.0006	0.0007	0.1353	0.0043	0.0002
						500	0.0712	0.1124	2.8551	0.5806	0.0006	0.0003	0.0003	0.0587	0.0021	0.0001
10	0.4	0.2	5.0	0.8	0.4	100	0.0013	0.4209	29.4588	1.8725	0.0157	0.0003	0.0038	1.3055	0.0081	0.0013
						250	0.0099	0.1079	9.8147	0.7456	-0.0158	0.0001	0.0015	0.3779	0.0029	0.0005
						500	0.0012	0.0907	4.5832	0.4423	0.0011	0.0001	0.0007	0.1796	0.0013	0.0002
11	0.9	0.7	4.7	0.5	0.5	100	-0.0986	-0.0158	2.7373	26.0359	-0.0045	0.0042	0.0001	0.0121	1.1989	0.0013
						250	-0.0558	-0.0073	1.1311	9.7238	-0.0107	0.0016	0.0001	0.0043	0.3864	0.0005
						500	-0.0098	-0.0051	0.4744	5.2583	0.0184	0.0008	0.0001	0.0021	0.1802	0.0002
12	0.4	0.6	1.2	3.2	0.5	100	-0.0707	0.0515	4.1733	13.8448	0.0104	0.0027	0.0005	0.0272	0.3834	0.0011
						250	-0.0477	0.0184	1.7183	4.8896	0.0007	0.0011	0.0002	0.0094	0.1254	0.0004
						500	-0.0077	0.0051	0.6897	3.0752	-0.0028	0.0005	0.0001	0.0043	0.0581	0.0002
13	0.9	0.7	4.7	0.5	0.7	100	-0.0049	-0.8978	17.2693	1.4606	-0.0251	0.0001	0.0075	0.6201	0.0038	0.0018
						250	-0.0018	-0.4241	6.5468	0.5641	-0.0051	0.0001	0.0031	0.2073	0.0013	0.0007
						500	-0.0002	-0.1668	2.8111	0.3263	-0.0026	0.0001	0.0014	0.0961	0.0006	0.0003
14	0.5	0.2	2.3	3.5	0.7	100	-0.0733	0.0426	2.7725	25.8623	0.0153	0.0013	0.0043	0.0083	0.3435	0.0009
						250	0.0332	-0.0917	1.2734	10.0406	0.0068	0.0005	0.0016	0.0031	0.0889	0.0003
						500	0.0224	-0.0036	0.6225	4.8221	0.0068	0.0002	0.0008	0.0014	0.0394	0.0001
15	0.3	0.8	2.4	1.2	0.8	100	0.1928	-0.5945	4.2111	15.8039	-0.0256	0.0004	0.0026	0.0632	0.2177	0.0009
						250	0.0987	-0.3513	1.4691	5.0698	0.0127	0.0001	0.0009	0.0244	0.0525	0.0003
						500	0.0481	-0.1862	0.6871	2.1657	0.0097	0.0001	0.0004	0.0116	0.0221	0.0001
16	0.4	0.9	1.5	1.9	0.8	100	-0.1281	-0.3337	3.6454	26.8186	-0.0245	0.0012	0.0004	0.0265	0.7181	0.0011
						250	-0.0605	-0.1125	1.4978	9.4096	0.0063	0.0004	0.0001	0.0091	0.1364	0.0004
						500	-0.0369	-0.0457	0.7765	3.9336	0.0113	0.0002	0.0001	0.0041	0.0501	0.0002

Source: Author (2021)

Figure 11 – Total RB% for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.5$.

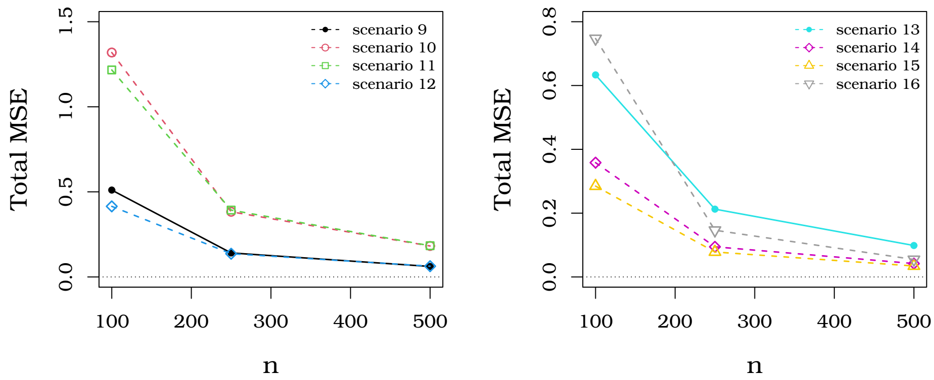
(a) Scenarios 9 to 12. (b) Scenarios 13 to 16.



Source: Author (2021)

Figure 12 – Total MSE for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.5$.

(a) Scenarios 9 to 12. (b) Scenarios 13 to 16.



Source: Author (2021)

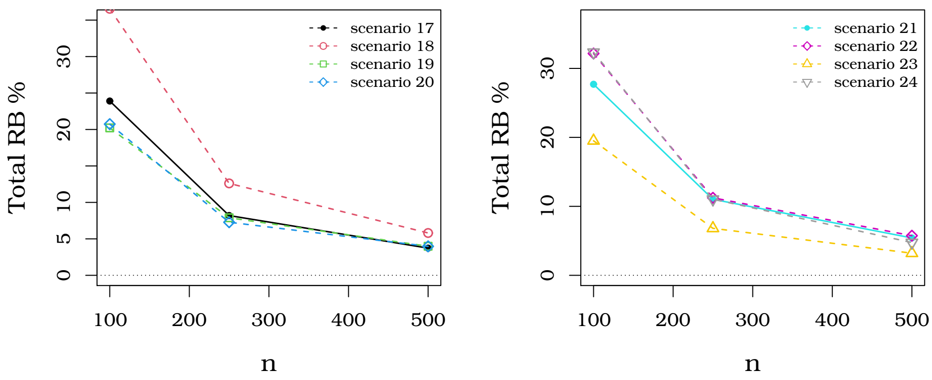
Table 11 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model, para $\tau = 0.25$.

Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{p}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{p}$
17	0.2	0.1	2.5	0.9	0.4	100	0.5151	0.8172	19.2698	3.3123	-0.0111	0.0021	0.0016	0.5305	0.0148	0.0005
						250	0.1616	0.2714	6.6401	1.1325	-0.0082	0.0006	0.0005	0.1527	0.0042	0.0002
						500	0.0541	0.1712	2.8853	0.6107	0.0108	0.0002	0.0002	0.0658	0.0019	0.0001
18	0.4	0.2	5.0	0.8	0.4	100	0.1001	0.7393	33.7887	1.9095	0.0106	0.0003	0.0028	1.4821	0.0079	0.0011
						250	0.0327	0.2411	11.5953	0.7392	-0.0015	0.0001	0.0011	0.4158	0.0028	0.0004
						500	0.0102	0.1796	5.1559	0.4521	0.0077	0.0001	0.0005	0.1863	0.0013	0.0002
19	0.9	0.7	4.7	0.5	0.5	100	0.0102	0.0099	18.8134	1.3925	-0.0076	0.0001	0.0051	0.6508	0.0038	0.0016
						250	0.0051	-0.0495	7.4006	0.5573	-0.0212	0.0001	0.0021	0.2215	0.0013	0.0006
						500	0.0031	0.0386	3.6082	0.3257	0.0103	0.0001	0.0011	0.1035	0.0006	0.0003
20	0.4	0.6	1.2	3.2	0.5	100	0.4011	0.2111	4.7789	15.3331	0.0125	0.0021	0.0005	0.0308	0.4482	0.0007
						250	0.1588	0.0655	1.8713	5.1945	-0.0086	0.0007	0.0001	0.0099	0.1409	0.0003
						500	0.0682	0.0516	0.7881	3.0652	0.0039	0.0004	0.0001	0.0044	0.0661	0.0001
21	0.8	0.3	6.0	2.6	0.7	100	-0.0051	0.0906	12.1058	15.5138	-0.0042	0.0001	0.0011	0.4109	0.2621	0.0021
						250	-0.0011	0.0041	4.8797	6.1073	0.0332	0.0001	0.0004	0.1511	0.0844	0.0008
						500	0.0001	-0.0039	2.0299	3.3756	0.0199	0.0001	0.0002	0.0697	0.0401	0.0004
22	0.5	0.2	2.3	3.5	0.7	100	-0.0658	0.2541	5.6023	26.3541	0.0182	0.0004	0.0003	0.0807	0.7716	0.0013
						250	-0.0097	0.0817	2.1471	9.0033	0.0147	0.0001	0.0001	0.0296	0.1849	0.0005
						500	0.0045	0.0467	1.0381	4.6491	0.0073	0.0001	0.0001	0.0138	0.0827	0.0002
23	0.3	0.8	2.4	1.2	0.8	100	0.1986	-0.1101	3.8889	15.5529	-0.0274	0.0003	0.0029	0.0624	0.2707	0.0011
						250	0.0856	-0.0549	1.4251	5.3471	0.0113	0.0001	0.0011	0.0229	0.0473	0.0004
						500	0.0408	-0.0081	0.7339	2.4304	0.0109	0.0001	0.0004	0.0104	0.0193	0.0002
24	0.4	0.9	1.5	1.9	0.8	100	0.1234	-0.2867	3.4196	29.1013	-0.0238	0.0007	0.0011	0.0261	0.9943	0.0011
						250	0.0467	-0.0635	1.4276	9.6141	0.0071	0.0003	0.0001	0.0089	0.1406	0.0003
						500	0.0209	-0.0257	0.7481	4.0017	0.0101	0.0001	0.0001	0.0041	0.0523	0.0002

Source: Author (2021)

Figure 13 – Total RB% for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.25$.

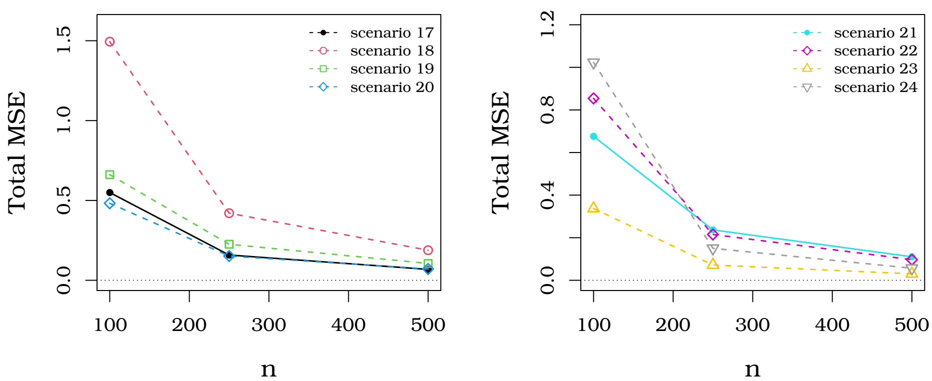
(a) Scenarios 17 to 20. (b) Scenarios 21 to 24.



Source: Author (2021)

Figure 14 – Total MSE for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.25$.

(a) Scenarios 17 to 20. (b) Scenarios 21 to 24.



Source: Author (2021)

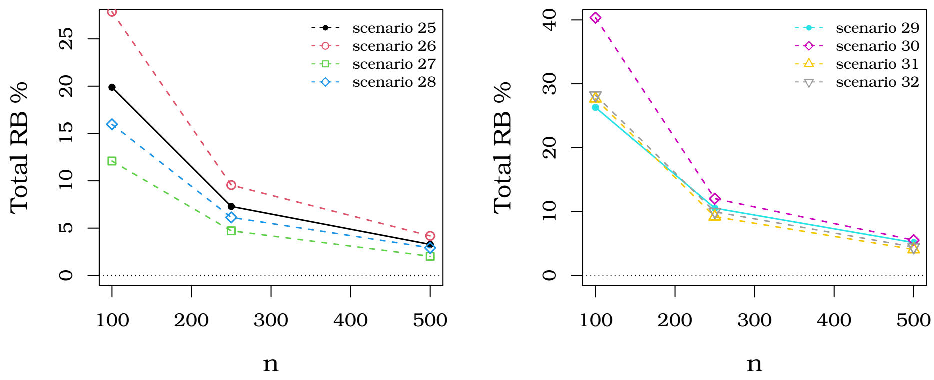
Table 12 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model, para $\tau = 0.75$.

Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$
25	0.2	0.1	2.5	0.9	0.4	100	0.0614	0.6321	16.8864	2.2831	0.0249	0.0026	0.0034	0.4048	0.0113	0.0011
						250	0.0654	0.2027	6.0884	0.9445	-0.0102	0.0011	0.0013	0.1137	0.0041	0.0004
						500	0.0407	0.1089	2.5926	0.5346	0.0041	0.0004	0.0006	0.0491	0.0019	0.0002
26	0.4	0.2	5.0	0.8	0.4	100	-0.0897	0.1097	25.6589	2.1509	0.0295	0.0006	0.0075	1.1182	0.0085	0.0016
						250	-0.0219	-0.0457	8.7453	0.8615	0.0073	0.0002	0.0031	0.3245	0.0031	0.0006
						500	-0.0108	-0.0244	3.7014	0.4897	0.0232	0.0001	0.0015	0.1482	0.0014	0.0003
27	0.9	0.7	4.0	0.6	0.5	100	-0.0375	-2.4091	12.3655	2.1603	0.0128	0.0001	0.0124	0.3989	0.0064	0.0019
						250	-0.0159	-1.0385	4.9285	0.8428	0.0032	0.0001	0.0047	0.1403	0.0022	0.0008
						500	-0.0048	-0.5292	2.0703	0.4681	0.0201	0.0001	0.0022	0.0641	0.0011	0.0004
28	0.4	0.6	1.2	3.2	0.5	100	-0.8291	-0.1121	4.4417	12.4527	0.0194	0.0056	0.0008	0.0271	0.3241	0.0013
						250	-0.3738	-0.0587	1.7871	4.7815	-0.0017	0.0022	0.0003	0.0091	0.1076	0.0005
						500	-0.1477	-0.0289	0.7476	2.3435	0.0073	0.0011	0.0001	0.0041	0.0504	0.0002
29	0.8	0.3	6.0	2.6	0.7	100	-0.0327	-0.1788	12.0966	14.4248	-0.0032	0.0001	0.0021	0.3948	0.2386	0.0021
						250	-0.0126	-0.1266	4.7917	5.8392	0.0395	0.0001	0.0008	0.1464	0.0785	0.0008
						500	-0.0042	-0.0661	2.1076	3.0914	0.0171	0.0001	0.0004	0.0676	0.0377	0.0004
30	0.5	0.2	2.3	3.5	0.7	100	-0.5119	0.6754	7.3478	32.8236	0.0154	0.0011	0.0036	0.1011	1.0829	0.0009
						250	-0.2501	0.5165	2.7293	9.0254	0.0038	0.0004	0.0017	0.0387	0.3449	0.0003
						500	-0.1057	0.2479	1.1895	4.1899	0.0047	0.0001	0.0006	0.0188	0.1662	0.0001
31	0.3	0.8	2.4	1.2	0.8	100	-0.1141	-1.5711	7.3688	22.0105	-0.0223	0.0011	0.0052	0.0745	0.5316	0.0007
						250	-0.0188	-0.6967	3.1843	6.7136	0.0137	0.0004	0.0015	0.0298	0.0632	0.0002
						500	0.0181	-0.3654	1.5101	2.9003	0.0084	0.0002	0.0006	0.0148	0.0275	0.0001
32	0.4	0.9	1.5	1.9	0.8	100	-0.4861	-0.5492	3.8337	25.4473	-0.0256	0.0025	0.0005	0.0269	0.5981	0.0011
						250	-0.2163	-0.2241	1.5794	8.8246	0.0049	0.0011	0.0001	0.0094	0.1229	0.0004
						500	-0.1163	-0.0938	0.8036	3.8307	0.0127	0.0004	0.0001	0.0042	0.0473	0.0002

Source: Author (2021)

Figure 15 – Total RB% for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.75$.

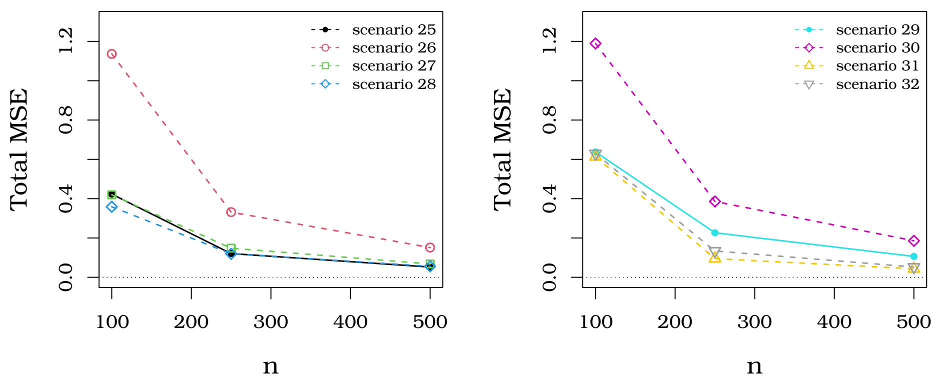
(a) Scenarios 25 to 28. (b) Scenarios 29 to 32.



Source: Author (2021)

Figure 16 – Total MSE for the estimates of the MLEs of the UWUWmodel obtained via EM algorithm, $\tau = 0.75$.

(a) Scenarios 25 to 28. (b) Scenarios 29 to 32.



Source: Author (2021)

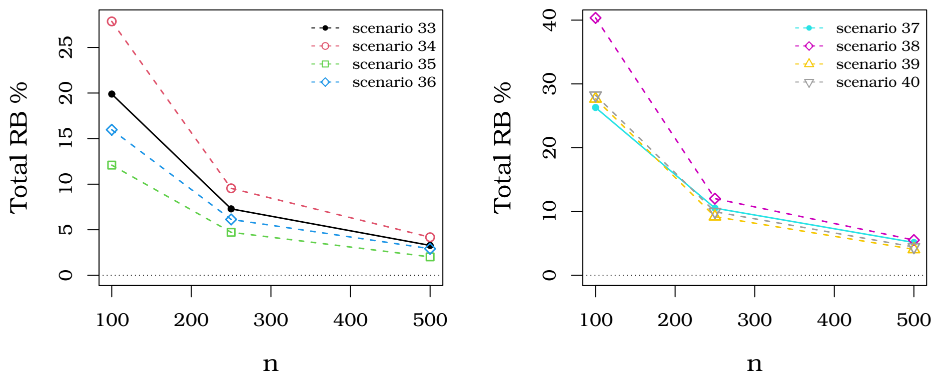
Table 13 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model, para $\tau = 0.5$.

Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$
33	0.1	0.3	0.7	1.3	0.4	100	1.3431	0.8492	5.1813	5.1494	0.0241	0.0072	0.0051	0.0244	0.0469	0.0005
						250	0.5284	0.3362	1.9471	1.8891	-0.0109	0.0025	0.0019	0.0071	0.0166	0.0002
						500	0.2695	0.1891	0.8267	1.0515	0.0056	0.0012	0.0009	0.0029	0.0079	0.0001
34	0.7	0.4	2.9	5.7	0.4	100	-0.0656	-0.0062	14.3154	18.6748	0.0326	0.0004	0.0001	0.2733	0.6269	0.0019
						250	-0.0269	-0.0151	5.3321	7.3435	0.0044	0.0001	0.0001	0.0869	0.2085	0.0008
						500	-0.0091	-0.0003	2.2942	4.0364	0.0129	0.0001	0.0001	0.0393	0.1001	0.0004
35	0.6	0.5	1.1	0.6	0.5	100	-0.2387	-0.6997	7.3386	4.2118	-0.0071	0.0107	0.0179	0.0635	0.0155	0.0003
						250	0.3096	-0.7241	2.1014	1.6881	-0.0084	0.0034	0.0071	0.0214	0.0048	0.0001
						500	0.1159	-0.2658	0.9406	0.7811	0.0078	0.0016	0.0033	0.0095	0.0018	0.0001
36	0.3	0.4	1.7	2.1	0.5	100	1.3058	-0.5612	16.8161	34.5799	0.0041	0.0069	0.0076	0.2003	0.3692	0.0002
						250	0.0242	0.3216	9.5246	12.9866	0.0025	0.0028	0.0032	0.0916	0.0977	0.0001
						500	-0.1507	0.3576	5.1591	6.4477	-0.0006	0.0013	0.0015	0.0469	0.0431	0.0001
37	0.4	0.3	0.9	4.3	0.7	100	0.0951	0.0155	1.9486	50.7801	0.0078	0.0034	0.0012	0.0081	2.7065	0.0007
						250	0.0351	-0.0081	0.7558	17.1542	0.0101	0.0013	0.0003	0.0031	0.6084	0.0003
						500	0.0303	0.0044	0.3515	7.9056	0.0146	0.0006	0.0001	0.0013	0.2662	0.0001
38	0.8	0.9	3.1	2.3	0.7	100	0.1885	-0.4281	6.6299	29.2241	-0.0064	0.0001	0.0004	0.1876	0.7482	0.0008
						250	0.1156	-0.2601	1.3025	8.3721	0.0246	0.0001	0.0001	0.0791	0.1555	0.0003
						500	0.0668	-0.1386	0.1455	3.3921	0.0011	0.0001	0.0001	0.0397	0.0681	0.0001
39	0.9	0.8	1.9	0.6	0.8	100	0.0154	-1.7455	3.3579	6.3849	-0.0078	0.0001	0.0081	0.0468	0.0322	0.0007
						250	0.0074	-0.7837	1.2941	2.3185	-0.0146	0.0001	0.0031	0.0168	0.0072	0.0002
						500	0.0021	-0.3069	0.7142	1.0719	0.0024	0.0001	0.0014	0.0079	0.0031	0.0001
40	0.1	0.4	2.8	1.7	0.8	100	0.1908	-0.4331	8.0864	45.0598	-0.0201	0.0004	0.0084	0.1143	2.8126	0.0006
						250	0.1145	-0.6036	3.3183	11.0698	0.0113	0.0001	0.0039	0.0466	0.1681	0.0002
						500	0.0834	-0.5345	1.2913	4.2572	0.0061	0.0001	0.0021	0.0241	0.0651	0.0001

Source: Author (2021)

Figure 17 – Total RB% for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.5$.

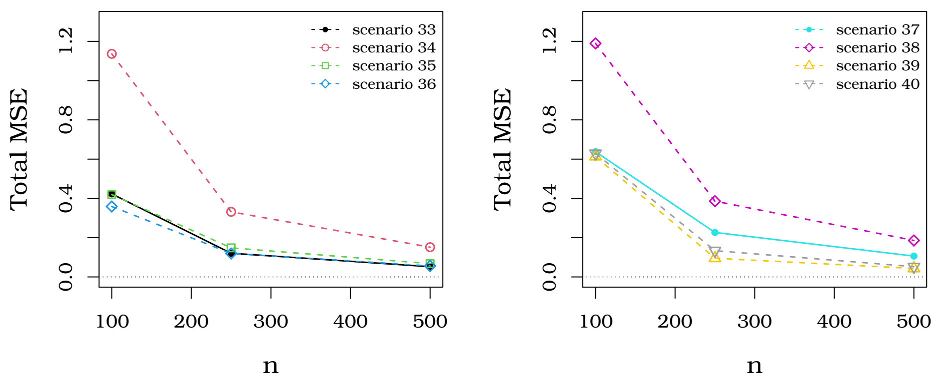
(a) Scenarios 33 to 36. (b) Scenarios 34 to 40.



Source: Author (2021)

Figure 18 – Total MSE for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.5$.

(a) Scenarios 33 to 36. (b) Scenarios 37 to 40.



Source: Author (2021)

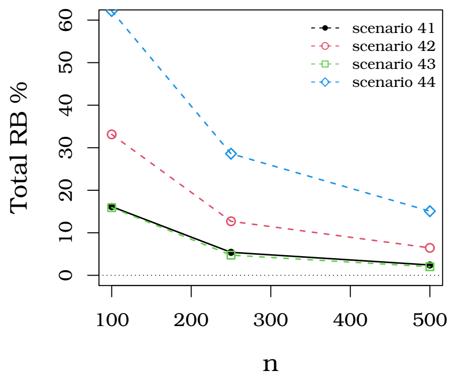
Table 14 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model, para $\tau = 0.25$.

Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$
41	0.1	0.3	0.7	1.3	0.4	100	2.3087	1.2611	7.2043	5.3671	0.0107	0.0092	0.0057	0.0367	0.0531	0.0004
						250	0.6246	0.5861	2.6006	1.6021	-0.0095	0.0018	0.0017	0.0089	0.0193	0.0001
						500	0.2986	0.2698	1.0301	0.8254	0.0055	0.0008	0.0008	0.0032	0.0092	0.0001
42	0.7	0.4	2.9	5.7	0.4	100	0.0309	0.0585	13.2779	19.7381	0.0308	0.0003	0.0001	0.2551	0.6388	0.0021
						250	0.0079	0.0132	4.8609	7.8168	0.0021	0.0001	0.0001	0.0832	0.2192	0.0008
						500	0.0006	0.0163	2.0896	4.3322	0.0141	0.0001	0.0001	0.0383	0.1061	0.0004
43	0.6	0.5	1.1	0.6	0.5	100	-0.2619	1.9723	8.6334	5.5848	-0.0128	0.0112	0.0175	0.0693	0.0201	0.0004
						250	0.1232	0.3801	2.1328	2.1169	-0.0044	0.0035	0.0059	0.0239	0.0056	0.0001
						500	0.0875	0.1816	0.8149	0.9274	0.0096	0.0013	0.0024	0.0111	0.0019	0.0001
44	0.3	0.4	1.7	2.1	0.5	100	2.4869	0.5871	18.7961	40.2805	0.0037	0.0069	0.0071	0.2241	0.4299	0.0001
						250	0.6262	0.9384	11.9024	15.1276	0.0024	0.0024	0.0034	0.1121	0.1118	0.0001
						500	0.0267	0.8458	7.1611	7.0445	-0.0007	0.0009	0.0017	0.0601	0.0503	0.0001
45	0.4	0.3	0.9	4.3	0.7	100	0.5726	0.0498	2.1856	49.7098	-0.0002	0.0026	0.0009	0.0086	2.3576	0.0007
						250	0.2259	0.0053	0.8353	16.9112	0.0222	0.0011	0.0002	0.0029	0.5597	0.0003
						500	0.1404	-0.0026	0.3622	8.8801	-0.0034	0.0005	0.0001	0.0013	0.2471	0.0001
46	0.8	0.9	3.1	2.3	0.7	100	0.1732	-0.3043	5.4412	24.9968	-0.0063	0.0001	0.0004	0.1674	0.7201	0.0009
						250	0.0723	-0.1961	1.3165	6.9376	0.0269	0.0001	0.0001	0.0664	0.1332	0.0003
						500	0.0311	-0.0841	0.5129	3.2706	0.0014	0.0001	0.0001	0.0316	0.0599	0.0001
47	0.9	0.8	1.9	0.6	0.8	100	0.0451	-0.0069	3.4448	7.5111	-0.0217	0.0001	0.0048	0.0488	0.0469	0.0005
						250	0.0153	-0.0754	1.2818	2.4641	-0.0038	0.0001	0.0021	0.0173	0.0073	0.0002
						500	0.0052	0.0308	0.6826	1.1224	0.0161	0.0001	0.0009	0.0083	0.0029	0.0001
48	0.1	0.4	2.8	1.7	0.8	100	0.2174	0.0131	5.3541	31.7771	-0.0228	0.0001	0.0093	0.0997	1.0556	0.0007
						250	0.1076	-0.5425	1.7278	8.3367	0.0115	0.0001	0.0046	0.0399	0.1326	0.0002
						500	0.0564	-0.4514	0.5708	3.2849	0.0074	0.0001	0.0023	0.0201	0.0556	0.0001

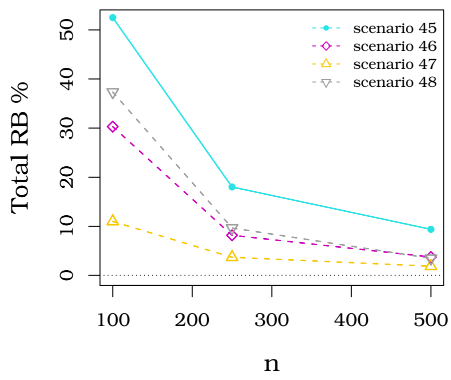
Source: Author (2021)

Figure 19 – Total RB% for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.25$.

(a) Scenarios 41 to 44.



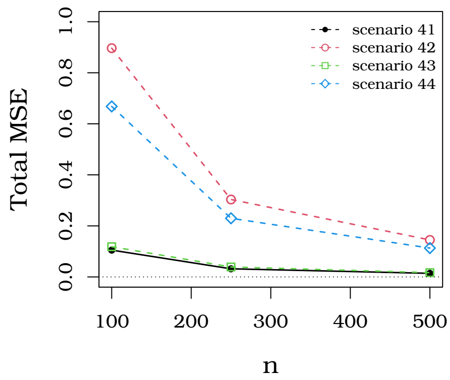
(b) Scenarios 45 to 48.



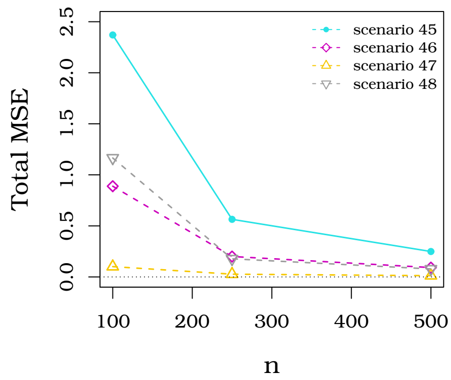
Source: Author (2021)

Figure 20 – Total MSE for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.25$.

(a) Scenarios 41 to 44.



(b) Scenarios 45 to 48.



Source: Author (2021)

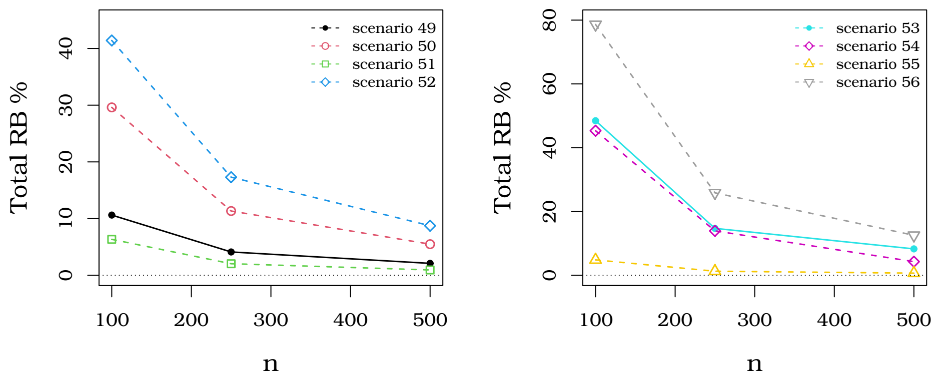
Table 15 – RB% and MSE of the estimates of the MLEs obtained via the EM algorithm of the UWUW model, para $\tau = 0.75$.

Scenario	μ_1	μ_2	β_1	β_2	p	n	RB %					MSE				
							$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\beta_1}$	$\hat{\beta_2}$	$\hat{p}$
49	0.1	0.3	0.7	1.3	0.4	100	1.8442	0.1283	3.7247	4.8947	0.0336	0.0121	0.0063	0.0165	0.0426	0.0008
						250	0.8289	0.0132	1.4405	1.8396	-0.0031	0.0046	0.0025	0.0054	0.0145	0.0003
						500	0.4627	-0.0081	0.5953	1.0576	0.0011	0.0021	0.0012	0.0024	0.0068	0.0001
50	0.7	0.4	2.9	5.7	0.4	100	-0.2511	-0.0269	14.3522	15.5045	0.0337	0.0005	0.0002	0.3013	0.5694	0.0017
						250	-0.0962	-0.0254	5.4704	5.9814	0.0058	0.0002	0.0001	0.0966	0.1883	0.0007
						500	-0.0346	-0.0092	2.1958	3.3271	0.0117	0.0001	0.0001	0.0431	0.0906	0.0003
51	0.6	0.5	1.1	0.6	0.5	100	-0.8314	-2.3712	6.5533	2.9878	0.0073	0.0115	0.0269	0.0559	0.0111	0.0006
						250	-0.2149	-1.1893	2.2669	1.1962	-0.0124	0.0045	0.0118	0.0174	0.0036	0.0002
						500	-0.1053	-0.5362	0.9951	0.5918	0.0021	0.0021	0.0058	0.0077	0.0015	0.0001
52	0.3	0.4	1.7	2.1	0.5	100	0.2557	-1.4483	14.3751	28.2231	0.0048	0.0094	0.0083	0.1762	0.3044	0.0003
						250	-0.2795	-0.2884	7.0162	10.8547	0.0024	0.0043	0.0031	0.0727	0.0824	0.0001
						500	-0.1272	-0.1342	3.3515	5.6571	-0.0004	0.0023	0.0016	0.0357	0.0359	0.0001
53	0.4	0.3	0.9	4.3	0.7	100	-0.3815	-0.0963	1.8176	47.1071	0.0136	0.0065	0.0019	0.0085	2.4788	0.0009
						250	-0.1805	-0.0261	0.7456	14.1534	0.0261	0.0026	0.0006	0.0031	0.5503	0.0003
						500	-0.0554	-0.0404	0.3258	8.0644	-0.0046	0.0013	0.0003	0.0014	0.2442	0.0001
54	0.8	0.9	3.1	2.3	0.7	100	0.1001	-0.7466	10.7375	35.2397	-0.0066	0.0003	0.0008	0.2146	0.9167	0.0007
						250	0.1408	-0.4001	2.5558	11.6171	0.0217	0.0001	0.0002	0.0977	0.1983	0.0002
						500	0.1351	-0.2393	0.0977	4.2931	0.0007	0.0001	0.0001	0.0533	0.0836	0.0001
55	0.9	0.8	1.9	0.6	0.8	100	-0.0398	-6.3071	2.8454	8.3544	-0.0211	0.0001	0.0294	0.0426	0.0557	0.0009
						250	-0.0159	-2.4248	1.1844	2.5477	-0.0101	0.0001	0.0078	0.0161	0.0076	0.0003
						500	-0.0067	-1.0405	0.5931	1.1345	-0.0002	0.0001	0.0032	0.0076	0.0031	0.0002
56	0.1	0.4	2.8	1.7	0.8	100	-0.1207	-1.3397	14.5299	65.5843	-0.0162	0.0005	0.0128	0.1453	4.2333	0.0004
						250	-0.1128	-0.5662	7.8871	18.6456	0.0107	0.0002	0.0051	0.0593	0.2724	0.0001
						500	-0.0635	-0.2924	4.6242	8.2711	0.0044	0.0001	0.0025	0.0299	0.0826	0.0001

Source: Author (2021)

Figure 21 – Total RB% for the estimates of the MLEs of the UWUWmodel obtained via EM algorithm, $\tau = 0.75$.

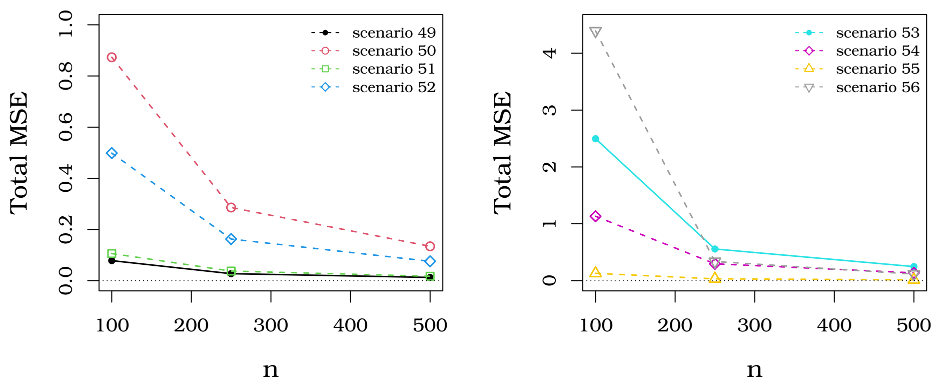
(a) Scenarios 49 to 52. (b) Scenarios 52 to 56.



Source: Author (2021)

Figure 22 – Total MSE for the estimates of the MLEs of the UWUW model obtained via EM algorithm, $\tau = 0.75$.

(a) Scenarios 49 to 52. (b) Scenarios 53 to 56.



Source: Author (2021)