ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 14, No.1
February 2016

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 845 40 84
    Web site: http://www.ine.pt
    Customer Support Service
    (National network) : 808 201 808
    Other networks: + 351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

ACKNOWLEDGEMENT

Maria José Carrilho was a leading senior statistician at the *Department of Social Studies* of *INE—Instituto Nacional de Estatística*, and the Editor-in--Chief of the *Demographic Studies Review*, where she published definitive studies on Portuguese demography. Among other activities, she also served as *Councilor for Gender Equality* and as the *Coordinator of the INE's Data Base on Gender*. She further was member of the *Working Group on Indicators of Human Rights*.

Ten years ago Maria José Carrilho has been appointed the *Executive Editor* (EE) of *REVSTAT—STATISTICAL JOURNAL*. Her management of the journal has been exemplar. She has indeed devotedly served *REVSTAT* as EE, from 2005 until 2015, and *INE*, *REVSTAT* and I wish to acknowledge our debt to her outstanding performance, wishing her a very happy and fruitful retirement.

Maria Ivette Gomes
(Editor-in-Chief)

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 1–28

GENERAL MULTIVARIATE DEPENDENCE USING ASSOCIATED COPULAS

Author: Yuri Salazar Flores
– Centre for Financial Risk, Macquarie University, Sydney, Australia
yuri.salazar@mq.edu.au

Received: April 2013 Revised: January 2014 Accepted: September 2014

Abstract:

- This paper studies the general multivariate dependence and tail dependence of a random vector. We analyse the dependence of variables going up or down, covering the 2^d orthants of dimension d and accounting for non-positive dependence. We extend definitions and results from positive to general dependence using the associated copulas. We study several properties of these copulas and present general versions of the tail dependence functions and tail dependence coefficients. We analyse the perfect dependence models, elliptical copulas and Archimedean copulas. We introduce the monotonic copulas and prove that the multivariate Student's t copula accounts for all types of tail dependence simultaneously while Archimedean copulas with strict generators can only account for positive tail dependence.

Key-Words:

- *non-positive dependence; tail dependence; copula theory; perfect dependence models; elliptical copulas; Archimedean copulas.*

AMS Subject Classification:

- 62H20, 60G70.

1. INTRODUCTION

A great deal of literature has been written on the analysis of the dependence structure between random variables. There is an increasing interest in the understanding of the dependencies between extreme values in what is known as tail dependence. However, the analysis of multivariate tail dependence in copula models has been exclusively focused on the positive case. Only the lower and upper tail dependence have been considered, leaving a void in the analysis of dependence structure implied by the use of these models. In this paper we tackle this issue by considering the dependence in the 2^d different orthants of dimension d for a random vector.

The use of the tail dependence coefficient (TDC) and the tail dependence function comes as a response to the inability of other measures when it comes to tail dependence (see [22, 13] and [20, Chapter 5]). This includes the Pearson's correlation coefficient and copula measures such as the Spearman's ρ, Kendall's τ and the Blomqvist's β.

The analysis of lower tail dependence has been derived using the copula, C, see e.g. [13, 22, 23]. In the context of nonparametric statistics, it is possible to measure upper tail dependence by using negative transformations or rotations. However, presenting a formal definition of upper tail dependence in the multivariate case and analysing it in copula models can not be achieved by the use of such methods. Also, trying to define it in terms of C becomes cumbersome in higher dimensions. By using the survival copula, the results and analysis of lower tail dependence have been generalised to upper tail dependence. For more on the analysis of the use of the survival copula for upper tail dependence, see [10, 23, 14, 15, 20, 27]. The study of non-positive tail dependence is also relevant when dealing with empirical data and in copula models analysis, see e.g. [32, 4]. In the case of copula models, the study of tail dependence helps in the understanding of the underlying assumptions implied by the use of these models. For example, the Student's t copula is often used to model data with only positive tail dependence. However, although this model accounts for the positive tail dependence, it also assumes the existence of negative tail dependence. Table 1 illustrates positive and negative tail dependence in the bivariate case which we generalise to the multivariate one.

Table 1: Tail dependence in the four different orthants of dimension two for variables X and Y.

	Lower Tail of X	Upper Tail of X
Lower Tail of Y	classical lower tail dependence	upper-lower tail dependence
Upper Tail of Y	lower-upper tail dependence	classical upper tail dependence

Although much has been written on the need to understand multivariate non-positive tail dependence, no formal definition has been presented. In this work we define the necessary concepts to study non-positive tail dependence in multivariate copula models. We use a copula approach and base our study on the associated copulas (see [13, p. 15]). If a copula is the distribution of $\mathbf{U} = (U_1, ..., U_d)$, the associated copulas are the distribution functions of vectors of the form $(U_1, 1-U_2, U_3, ..., 1-U_{d-1}, 1-U_d)$. The use of copulas of transformations for non-positive dependence is also suggested in [5, 30].

The reasoning behind the use of associated copulas is the same as for the use of the survival copula for upper tail dependence analysis. Similarly to that case, the definition and study of non-positive tail dependence is simplified by the use of these copulas. They enable us to present a unified definition of multivariate general tail dependence. This definition is consistent with generalisations from dimension 2 to d of positive tail dependence. The study of the associated copulas to analyse non-positive tail dependence is then a generalisation of the use of the copula and the survival copula for lower and upper tail dependence respectively.

The reminder of this work is divided in three sections: In the second section we present the concepts we use to study dependence in all the orthants. This includes general definitions of dependence and probability functions. We present a version of Sklar's theorem that proves that the copulas that link these general probability functions and its marginals are the associated copulas. We then present four propositions regarding these copulas. At the end of this section we present general definitions of the tail dependence functions and TDCs. In the third section we use the results obtained in Section 2 to study the perfect dependence models, elliptical copulas and Archimedean copulas. We present the copulas of the perfect dependence cases, which include non-positive perfect dependence. We call these copulas the monotonic copulas. We then characterise the associated elliptical copulas and obtain an expression for the associated tail dependence functions of the Student's t copula model. This model accounts for all 2^d types of tail dependence simultaneously. After that, we prove that, by construction, Archimedean copulas with strict generators can not account for non-positive tail dependence. We then present three examples with non-strict generators which account for negative tail dependence. At the end of this section we discuss a method for modelling arbitrary tail dependence using copula models. Finally, in the fourth section, we conclude and discuss future lines of research for general dependence.

Unless we specifically state it, all the definitions and results presented regarding general dependence are a contribution of this work.

2. ASSOCIATED COPULAS, TAIL DEPENDENCE FUNCTIONS AND TAIL DEPENDENCE COEFFICIENTS

In this section we analyse the dependence structure among random variables using copulas. Given a random vector $\mathbf{X} = (X_1, ..., X_d)$, we use the corresponding copula C and its associated copulas to analyse its dependence structure. For this we introduce a general type of dependence $\mathbf{D}$, one for each of the 2^d different orthants. This corresponds to the lower and upper movements of the different variables.

To analyse different dependencies, we introduce the $\mathbf{D}$-probability function and present a version of Sklar's theorem that states that an associated copula is the copula that links this function and its marginals. We present a formula to link all associated copulas and three results on monotone functions and associated copulas. We then introduce the associated tail dependence function and the associated tail dependence coefficient for the type of dependence $\mathbf{D}$. These functions generalise the positive (lower and upper) cases (extensively studied in [12, 13, 23]). With the concepts studied in this section, we aim to provide the tools to analyse the whole dependence structure among random variables, including non-positive dependence.

2.1. Copulas and dependence

The concept of copula was first introduced by [29], and is now a cornerstone topic in multivariate dependence analysis (see [13, 22, 20]). We now present the concepts of copula, general dependence and associated copulas that are fundamental for the rest of this work.

Definition 2.1. A multivariate copula $C(u_1, ..., u_d)$ is a distribution function on the d-dimensional-square $[0, 1]^d$ with standard uniform marginal distributions.

If C is the distribution function of $\mathbf{U} = (U_1, ..., U_d)$, we denote as $\widehat{C}$ the distribution function of $(1-U_1, ..., 1-U_d)$. C is used to link distribution functions with their corresponding marginals, accordingly we refer to C as the distributional copula. On the other hand, $\widehat{C}$ is used to link multivariate survival functions with their marginal survival functions, this copula is known as the survival copula.[1] Let $\mathbf{X} = (X_1, ..., X_d)$ be a random vector with joint distribution function F, joint survival function $\overline{F}$, marginals F_i and marginal survival functions $\overline{F}_i$, for $i \in$

[1] We use the term distributional for C, to distinguish it from the other associated copulas. The notation for the survival copula corresponds to the one used in the seminal work of [13].

$\{1, ...d\}$. Two versions of Sklar's theorem guarantees the existence and uniqueness of a copulas C and $\widehat{C}$ which satisfy

$$F(x_1, ..., x_d) = C\big(F_1(x_1), ..., F_d(x_d)\big) , \tag{2.1}$$

$$\overline{F}(x_1, ..., x_d) = \widehat{C}\big(\overline{F}_1(x_1), ..., \overline{F}_d(x_d)\big) , \tag{2.2}$$

see [13, 22]. In the next section we generalise these equations using the concept of general dependence, which we now define.

Definition 2.2. In d dimensions, we call the vector $\mathbf{D} = (D_1, ..., D_d)$ a type of dependence if each D_i is a boolean variable, whose value is either L (lower) or U (upper) for $i \in \{1, ...d\}$. We denote by Δ the set of all 2^d types of dependence.

Each type of dependence corresponds to the variables going up or down simultaneously. Tail dependence, which we define later, refers to the case when the variables go extremely up or down simultaneously. Two well known types of dependence are lower and upper dependence. Lower dependence refers to the case when all variables go down at the same time ($D_i = L$ for $i \in \{1, ..., d\}$) and upper dependence to the case when they all go up at the same time ($D_i = U$ for $i \in \{1, ..., d\}$). These two cases are examples of positive dependence and they have been extensively studied for tail dependence analysis, see e.g. [13, 22]. In the bivariate case the dependencies $\mathbf{D} = (L, U)$ and $\mathbf{D} = (U, L)$ correspond to one variable going up while the other one goes down. These are examples of negative dependence. Negative tail dependence is often present in financial time series, see [32, 4, 14]. Hence, in dimension 2 there are four types of dependence that correspond to the four quadrants. Note that, in dimension d, for each of the 2^d orthants we define a dependence $\mathbf{D}$.

Using the concept of dependence, we now present the associated copulas, see [13, Chapter 1, p. 15].

Definition 2.3. Let $\mathbf{X} = (X_1, ..., X_d)$ be a random vector with corresponding copula C, which is the distribution function of the vector $(U_1, ..., U_d)$ with uniform marginals. Let Δ denote the set of all types of dependencies of Definition 2.2. For $\mathbf{D} = (D_1, ..., D_d) \in \Delta$, let $\mathbf{V_D} = (V_{D_1,1}, ..., V_{D_d,d})$ with

$$V_{D_i,i} = \begin{cases} U_i & \text{if } D_i = L \\ 1 - U_i & \text{if } D_i = U \end{cases} .$$

Note that $\mathbf{V_D}$ also has uniform marginals. We call the distribution function of $\mathbf{V_D}$, which is a copula, the associated $\mathbf{D}$-copula and denote it $C_{\mathbf{D}}$. We denote $\mathcal{A}_{\mathbf{X}} = \{C_{\mathbf{D}} |\ \mathbf{D} \in \Delta\}$, the set of 2^d associated copulas of the random vector $\mathbf{X}$. Also, for any $\emptyset \neq S \subseteq I$, let $\mathbf{D}(S)$ denote the corresponding $|S|$-dimensional marginal dependence of $\mathbf{D}$. Then the copula $C_{\mathbf{D}(S)}$, the distribution of the $|S|$-dimensional marginal vector $(V_{D_i,i} |\ i \in S)$, is known as a marginal copula of $C_{\mathbf{D}}$.

Note that the distributional and the survival copula are $C = C_{(L,...,L)}$ and $\widehat{C} = C_{(U,...,U)}$ respectively.

2.1.1. The **D**-probability function and its associated **D**-copula

The distributional copula C and the survival copula $\widehat{C}$ are used to explain the lower and upper dependence structure of a random vector respectively. We use the associated **D**-copula to explain the **D**-dependence structure of a random vector. For this, we first present the **D**-probability functions, which generalise the joint distribution and survival functions.

Definition 2.4. Let $\mathbf{X} = (X_1, ..., X_d)$ be a random vector with marginal distributions F_i for $i \in \{1, ...d\}$ and $\mathbf{D} = (D_1, ..., D_d)$ a type of dependence according to Definition 2.2. Define the event $\mathcal{B}_i(x_i)$ in the following way

$$\mathcal{B}_i(x_i) = \begin{cases} \{X_i \leq x_i\} & \text{if } D_i = L \\ \{X_i > x_i\} & \text{if } D_i = U \end{cases}.$$

Then the corresponding **D**-probability function is

$$F_{\mathbf{D}}(x_1, ..., x_d) = P\left(\bigcap_{i=1}^{d} \mathcal{B}_i(x_i)\right).$$

We refer to

$$F_{D_i,i} = \begin{cases} F_i & \text{if } D_i = L \\ \overline{F}_i & \text{if } D_i = U \end{cases},$$

for $i \in \{1, ...d\}$ as the marginal functions of $F_{\mathbf{D}}$ (note that the marginals are either univariate distribution or survival functions).

In the bivariate case for example, there are four **D**-probability functions: $F(x_1, x_2)$, $\overline{F}(x_1, x_2)$, $F_{LU}(x_1, x_2) = P(X_1 \leq x_1, X_2 > x_2)$ and $F_{UL}(x_1, x_2) = P(X_1 > x_1, X_2 \leq x_2)$. In general, these functions complement the use of the joint distribution and survival functions in our analysis of dependence in the 2^d orthants.

The following theorem presents the associated copula $C_{\mathbf{D}}$ in terms of the $F_{\mathbf{D}}$ and its marginals. It is because of this theorem that we can use the associated copula $C_{\mathbf{D}}$ to analyse **D**-dependence. We restrict the proof to the continuous case (for Sklar's theorem for distribution functions see [20, 13, 22]).

Theorem 2.1. ***Sklar's theorem for* D*-probability functions and associated copulas.***

Let $\mathbf{X} = (X_1, ..., X_d)$ *be a random vector,* $\mathbf{D} = (D_1, ..., D_d)$ *a type of dependence,* $F_{\mathbf{D}}$ *its* **D***-probability function and* $F_{D_i,i}$ *for* $i \in \{1, ...d\}$ *the marginal functions of* $F_{\mathbf{D}}$ *as in Definition 2.4. Let the marginal functions of* $F_{\mathbf{D}}$ *be continuous and* $F^{\leftarrow}$ *denote the generalised inverse of* F*, defined as* $F^{\leftarrow}(u) := \inf\{x \in \mathbb{R} | \ F(x) \geq u\}$*. Then the associated copula* $C_{\mathbf{D}} \colon [0,1]^d \to [0,1]$*, satisfies, for all* $x_1, ..., x_2$ *in* $[-\infty, \infty]$*,*

$$F_{\mathbf{D}}(x_1, ..., x_d) \ = \ C_{\mathbf{D}}\big(F_{D_1,1}(x_1), ..., F_{D_d,d}(x_d)\big) \ , \tag{2.3}$$

which is equivalent to

$$C_{\mathbf{D}}(u_1, ..., u_d) \ = \ F_{\mathbf{D}}\big(F^{\leftarrow}_{D_1,1}(u_1), ..., F^{\leftarrow}_{D_d,d}(u_d)\big) \ . \tag{2.4}$$

Conversely, let $\mathbf{D} = (D_1, ..., D_d)$ *be a dependence and* $F_{D_i,i}$ *a univariate distribution, if* $D_i = L$*, or a survival function, if* $D_i = U$*, for* $i \in \{1, ...d\}$*, then:*

- (**a**) *If* $C_{\mathbf{D}}$ *is a copula, then* $F_{\mathbf{D}}$ *in (2.3) defines a* **D***-probability function with marginals* $F_{D_i,i}$*,* $i \in \{1, ...d\}$*.*
- (**b**) *If* $F_{\mathbf{D}}$ *is any* **D***-probability function, then* $C_{\mathbf{D}}$ *in (2.4) is a copula.*

Proof: The proof of this theorem is analogous to the proof of Sklar's theorem for distribution functions. When two random variables have the same probability functions, we say they are equivalent in probability and denote it as $\overset{P}{\sim}$. In this general version of the theorem, we have that for the distribution function F_i, the events $\{X_i \leq x_i\} \overset{P}{\sim} \{F_i(X_i) \leq F_i(x_i)\}$ and $\{X_i > x_i\} \overset{P}{\sim} \{\overline{F}_i(X_i) \leq \overline{F}_i(x_i)\}$, for $i \in \{1, ..., d\}$ and $x_i \in [-\infty, \infty]$. This implies

$$P\big(\mathcal{B}_i(x_i)\big) \ = \ P\big(F_{D_i,i}(X_i) \leq F_{D_i,i}(x_i)\big) \ , \tag{2.5}$$

for $i \in \{1, ..., d\}$.

Considering equation (2.5) and Definition 2.4, we have that for any $x_1, ..., x_d$ in $[-\infty, \infty]$

$$F_{\mathbf{D}}(x_1, ..., x_d) \ = \ P\Big(F_{D_1,1}(X_1) \leq F_{D_1,1}(x_1), \ ..., \ F_{D_d,d}(X_d) \leq F_{D_d,d}(x_d)\Big) \ . \tag{2.6}$$

Using the continuity of F_i, we have that $F_i(X_i)$ is uniformly distributed (see [20, Proposition 5.2 (2)]). Hence, if we define $\mathbf{U} = (F_1(X_1), ..., F_d(X_d))$, its distribution function is a copula C. Note that in this case $\mathbf{V}_{\mathbf{D}}$, defined as in Definition 2.3, is equal to $(F_{D_1,1}(X_1), ..., F_{D_d,d}(X_d))$. It follows that the distribution function of $(F_{D_1,1}(X_1), ..., F_{D_d,d}(X_d))$ is the associated copula $C_{\mathbf{D}}$, in which case equation (2.5) implies

$$C_{\mathbf{D}}\big(F_{D_1,1}(x_1), ..., F_{D_d,d}(x_d)\big) = P\Big(F_{D_1,1}(X_1) \leq F_{D_1,1}(x_1), ..., F_{D_d,d}(X_d) \leq F_{D_d,d}(x_d)\Big) \ ,$$

and equation (2.3) follows.

Now, one of the properties of the generalised inverse is that, when T is continuous, $T \circ T^{\leftarrow}(x) = x$ (see [20, Proposition A.3]). Hence, if we evaluate $F_{\mathbf{D}}$ in $(F^{\leftarrow}_{D_1,1}(u_1), ..., F^{\leftarrow}_{D_d,d}(u_d))$, using equation (2.3), we get equation (2.4). This equation explicitly represents $C_{\mathbf{D}}$ in terms of $F_{\mathbf{D}}$ and its marginals implying its uniqueness.

For the converse statement of the theorem, we have

(**a**) Let $\mathbf{U} = (U_1, ..., U_d)$ be the random vector with distribution function C. Define $\mathbf{X} = (X_1, ..., X_d) = (F^{\leftarrow}_{D_1,1}(U_1), ..., F^{\leftarrow}_{D_d,d}(U_d))$ and

$$\mathcal{B}_i(x_i) = \begin{cases} \{X_i \leq x_i\} & \text{if } D_i = L \\ \{X_i > x_i\} & \text{if } D_i = U \end{cases},$$

for $i \in \{1, ...d\}$. Considering that $F(x) \leq y \iff x \leq F^{\leftarrow}(y)$, we have $\overline{F}^{\leftarrow}(x) \leq y \iff x \geq \overline{F}(y)$. Using these properties, we get

$$\big\{U_i \leq F_{D_i,i}(x_i)\big\} \overset{P}{\sim} \mathcal{B}_i(x_i)\,,$$

for $i \in \{1, ...d\}$. Using this, the $\mathbf{D}$-probability function of $\mathbf{X}$ is

$$P\left(\bigcap_{i=1}^{d} \mathcal{B}_i(x_i)\right) = C\big(F_{D_1,1}(x_1), ..., F_{D_d,d}(x_d)\big)\,.$$

This implies that $F_{\mathbf{D}}$ defined by (2.3) is the $\mathbf{D}$-probability function of $\mathbf{X}$ with marginals

$$P\big(\mathcal{B}_i(x_i)\big) = P\big(U_i \leq F_{D_i,i}(x_i)\big) = F_{D_i,i}(x_i)\,,$$

for $i \in \{1, ...d\}$.

(**b**) Similarly, let $(X_1, ..., X_d)$ be the random vector with $\mathbf{D}$-probability function $F_{\mathbf{D}}$. Define $\mathbf{U} = (U_1, ..., U_d) = (F_{D_1,1}(X_1), ..., F_{D_d,d}(X_d))$ (note that the vector is uniformly distributed). Again, using the properties of the generalised inverse, we have

$$\big\{U_i \leq u_i\big\} \overset{P}{\sim} \mathcal{B}_i\big(F^{\leftarrow}_{D_i,i}(u_i)\big)\,,$$

for $i \in \{1, ...d\}$. Hence the distribution function of $\mathbf{U}$ is $F_{\mathbf{D}}\big(F^{\leftarrow}_{D_1,1}(u_1), ..., F^{\leftarrow}_{D_d,d}(u_d)\big)$, which implies that the function is a copula.

For the properties of the generalised inverse function used in this proof, see [20, Proposition A.3]. □

For this theorem we referred to generalised inverse functions as they are more general than inverse functions. However, whenever we are not proving a general property, we assume distribution functions have inverse functions.

Note that this theorem implies that in the continuous case $C_{\mathbf{D}}$ is the $\mathbf{D}$-probability function of $(F_{D_1,1}(X_1), ..., F_{D_d,d}(X_d))$ characterised in (2.3). This

theorem implies the importance of the associated copulas to analyse dependencies. It also implies the Fréchet bounds for the **D**-probability functions of Definition 2.4. The bounds can also be obtained similarly to [13, Theorems 3.1 and 3.5],

$$\begin{aligned}\max\Big\{0,\, F_{D_1,1}(x_1)+\cdots+F_{D_d,d}(x_d)-(d-1)\Big\} \;&\leq\; F_{\mathbf{D}}(x_1,...,x_d)\\ &\leq\; \min\Big\{F_{D_1,1}(x_1),...,F_{D_d,d}(x_d)\Big\}\,.\end{aligned}\tag{2.7}$$

2.1.2. Properties of the associated copulas

In the bivariate case, [13, Chapter 1], and [22, Chapter 2], presented the expressions to link the associated copulas with the distributional copula C. In the multivariate case [14, Equation 8.1] and [10, Theorem 3], presented the expression between the distributional and the survival copula and [5, Theorem 2.7] proved that is possible to express the associated copulas in terms of the distributional copula C. We now present a general equation for the relationship between any two associated copulas $C_{\mathbf{D}^*}$ and $C_{\mathbf{D}^+}$ in the multivariate case. The equation is based on all the subsets of the indices where the $\mathbf{D}^*$ and $\mathbf{D}^+$ are different.

Proposition 2.1. *Let $\mathbf{X}=(X_1,...,X_d)$ be a random vector with associated copulas $\mathcal{A}_{\mathbf{X}}$ and $\mathbf{D}^*=(D_1^*,...,D_d^*)$ and $\mathbf{D}^+=(D_1^+,...,D_d^+)$ any two types of dependence. Consider the following sets and notations: $I=\{1,...,d\}$; $I_1=\{i\in I\,|\,D_i^*=D_i^+\}$ and $I_2=\{i\in I\,|\,D_i^*\neq D_i^+\}$; $d_1=|I_1|$ and $d_2=|I_2|$; $S_j=\{$the subsets of size j of $I_2\}$ and $S_{j,k}=\{$The k-th element of $S_j\}$ for $j\in\{1,...,d_2\}$ and $k\in\big\{1,...,\binom{d_2}{j}\big\}$. We define $S_0=\emptyset$ and $S_{0,1}=\emptyset$; for each $S_{j,k}$ define $\mathbf{W}_{j,k}=(W_{j,k.1},...,W_{j,k,d})$ with*

$$W_{j,k,i}=\begin{cases}u_i & \text{if } i\in I_1\\ 1-u_i & \text{if } i\in S_{j,k}\\ 1 & \text{if } i\notin I_1\cup S_{j,k}\end{cases},$$

for $i\in\{1,...d\}$, $j\in\{0,...,d_2\}$ and $k\in\Big\{1,...,\binom{d_2}{j}\Big\}$.

Then the associated $\mathbf{D}^*$-copula $C_{\mathbf{D}^*}$ is expressed in terms of the $\mathbf{D}^+$-copula $C_{\mathbf{D}^+}$ according to the following equation

$$C_{\mathbf{D}^*}(u_1,...,u_d)\;=\sum_{j=0}^{d_2}(-1)^j\sum_{k=1}^{\binom{d_2}{j}}C_{\mathbf{D}^+}(\mathbf{W}_{j,k})\,.\tag{2.8}$$

Note that in the cases when at least a 1 appears in $\mathbf{W}_{j,k}$, $C_{\mathbf{D}^+}(\mathbf{W}_{j,k})$ becomes a marginal copula of $C_{\mathbf{D}^+}$.

Proof: Throughout this proof, it must be borne in mind that $C_{\mathbf{D}^*}$ is the distribution function of the random vector $\mathbf{V}_{\mathbf{D}^*}$ and $C_{\mathbf{D}^+}$ of $\mathbf{V}_{\mathbf{D}^+}$, defined according to Definition 2.3. Note that, for $i \in I_2$, $V_{D_i^*,i} = 1 - V_{D_i^+ i}$ and they are equal otherwise.

In the case $d_2 = 0$, we have $\mathbf{D}^* = \mathbf{D}^+$, $j \in \{0\}$ and $k \in \{1\}$ ([2]), hence (2.8) holds. We prove (2.8) by induction on d, the dimension; it can also be proven by induction on d_2, the number of elements in which $D_i^* \neq D_i^+$. Note that in dimension $d = 1$, a copula becomes the identity function. If $D_1^* \neq D_1^+$, the expression becomes $u_1 = 1 - (1 - u_1)$; the case $D_1^* = D_1^+$ has already been covered in $d_2 = 0$, and expression (2.8) holds.

Now, in dimension d, we prove the formula works if it works in dimension $d-1$. We obtain an expression for $C_{\mathbf{D}^*}(u_1, ..., u_d)$ using the induction hypothesis. Consider the dependencies, on the $(d-1)$-dimension, $\mathbf{F}^* = (D_1^*, ..., D_{d-1}^*)$ and $\mathbf{F}^+ = (D_1^+, ..., D_{d-1}^+)$. We use an apostrophe on the sets and notations of $\mathbf{F}^*$ and $\mathbf{F}^+$ to differentiate them from those of $\mathbf{D}^*$ and $\mathbf{D}^+$. It follows that $d' = d-1$ and $I' = I - \{d\}$. By the induction hypothesis, equation (2.8) holds to express $C_{\mathbf{F}^*}$ in terms of $C_{\mathbf{F}^+}$. In terms of probabilities this is equivalent to

$$\begin{aligned} &P\Big(V_{D_i^*,1} \leq u_1,\, ...,\, V_{D_{d-1}^*,d-1} \leq u_{d-1}\Big) = \\ &\quad = \sum_{j=0}^{d_2-1} (-1)^j \sum_{k=1}^{\binom{d_2-1}{j}} P\Big(V_{D_1^+,1} \leq W'_{j,k,1},\, ...,\, V_{D_{d-1}^+,d-1} \leq W'_{j,k,d-1}\Big)\,. \end{aligned} \tag{2.9}$$

There are two cases to consider depending on whether D_d^* is equal to D_d^+ or not.

Case 1. $D_d^* = D_d^+$.

In this case, it follows that, $I'_1 = I_1 - \{d\}$, $I'_2 = I_2$, $d'_2 = d_2$ *and* $V_{D_d^*,d} = V_{D_d^+,d}$. *If we intersect the events in equation (2.9) with the event* $\{V_{D_d^* d} \leq u_d\}$ *we get*

$$\begin{aligned} &P\Big(V_{D_i^*,1} \leq u_1,\, ...,\, V_{D_{d-1}^*,d-1} \leq u_{d-1},\, V_{D_d^* d} \leq u_d\Big) = \\ &= \sum_{j=0}^{d_2} (-1)^j \sum_{k=1}^{\binom{d_2}{j}} P\Big(V_{D_1^+,1} \leq W'_{j,k,1},\, ...,\, V_{D_{d-1}^+,d-1} \leq W'_{j,k,d-1},\, V_{D_d^+,d} \leq u_d\Big)\,. \end{aligned} \tag{2.10}$$

Because $I'_2 = I_2$, *in this case, for* $j \in \{1, ..., d_2\}$ *and* $k \in \big\{1, ..., \binom{d_2}{j}\big\}$, *the events* $S'_{j,k}$ *are equal to* $S_{j,k}$. *Considering this, and* $I'_1 = I_1 - \{d\}$, *we have*

$$\big(\mathbf{W}'_{j,k}, u_d\big)_i = W_{j,k,i}$$

for $i \in \{1, ..., d\}$, *so* $(\mathbf{W}'_{j,k}, u_d) = \mathbf{W}_{j,k}$ *for* $j \in \{1, ..., d_2\}$ *and* $k \in \big\{1, ..., \binom{d_2}{j}\big\}$. *Equation (2.10) then implies:*

$$C_{D^*}(u_1, ..., u_d) = \sum_{j=0}^{d_2} (-1)^j \sum_{k=1}^{\binom{d_2}{j}} C_{D^+}(\mathbf{W}_{j,k})\,.$$

[2]Note that we are using the convention $0! = 1$

Case 2. $D_d^* \neq D_d^+$.

In this case, it holds that, $I_1' = I_1$, $I_2' = I_2 - \{d\}$, $d_2' = d_2 - 1$. *To obtain an expression for* $C_{\mathbf{D}^*}(u_1, ..., u_d) = P(V_{D_i^*,1} \leq u_1, ..., V_{D_d^*,d} \leq u_d)$, *we use the induction hypothesis. Considering* $P(A) = P(A \cap B) + P(A \cap B^c)$, *we have*

$$\begin{aligned} P\Big(V_{D_i^*,1} \leq u_1,\, ...,\, V_{D_{d-1}^*,d-1} \leq u_{d-1}\Big) \,=& \\ =\, P\Big(V_{D_i^*,1} \leq u_1,\, ...,\, V_{D_{d-1}^*,d-1} \leq u_{d-1},\, V_{D_d^*,d} \leq u_d\Big)& \\ +\, P\Big(V_{D_i^*,1} \leq u_1,\, ...,\, V_{D_{d-1}^*,d-1} \leq u_{d-1},\, V_{D_d^*,d} \geq u_d\Big)&, \end{aligned}$$

which implies

$$\begin{aligned} C_{\mathbf{D}^*}(u_1, ..., u_d) \,=\, & P\Big(V_{D_1^*,1} \leq u_1,\, ...,\, V_{d-1}^* \leq u_{d-1}\Big) \\ & - P\Big(V_{D_1^*,1} \leq u_1,\, ...,\, V_{d-1}^* \leq u_{d-1},\, V_{D_d^*,d} \geq u_d\Big)\,. \end{aligned} \tag{2.11}$$

Note that, in this case $V_{D_d^*,d} = 1 - V_{D_d^+,d}$. *This implies that the event* $\{V_{D_d^*,d} \geq u_d\}$ *is equivalent to* $\{V_{D_d^+,d} \leq 1 - u_d\}$. *If we intersect the events involved in equation (2.9) with the event* $\{V_{D_d^*,d} \geq u_d\}$ *we get*

$$\begin{aligned} & P\Big(V_{D_1^*,1} \leq u_1,\, ...,\, V_{D_{d-1}^*,d-1} \leq u_{d-1},\, V_{D_d^*,d} \geq u_d\Big) \,= \\ & = \sum_{j=0}^{d_2-1} (-1)^j \sum_{k=1}^{\binom{d_2-1}{j}} P\Big(V_{D_1^+,1} \leq W'_{j,k,1},\, ...,\, V_{D_{d-1}^+,d-1} \leq W'_{j,k,d-1},\, V_{D_d^+,d} \leq 1 - u_d\Big)\,. \end{aligned} \tag{2.12}$$

Combining equations (2.9), (2.11) and (2.12), we obtain

$$C_{\mathbf{D}^*}(u_1, ..., u_d) = \sum_{j=0}^{d_2-1} (-1)^j \sum_{k=1}^{\binom{d_2-1}{j}} C_{\mathbf{D}^+}(\mathbf{W}'_{j,k}, 1) - \sum_{j=0}^{d_2-1} (-1)^j \sum_{k=1}^{\binom{d_2-1}{j}} C_{\mathbf{D}^+}(\mathbf{W}'_{j,k}, 1 - u_d)\,. \tag{2.13}$$

Note that, in this case, the sets I_2 *and* I_2' *satisfy* $I_2 = I_2' \cup \{d\}$.

The rest of the proof is based on the fact that for $j \in \{1, ..., d-1\}$ the elements of size j of I_2 are the elements of size j of I_2' plus the elements of size $j-1$ of I_2' attaching them $\{d\}$. Considering our notation, this means

$$S_j \,=\, S'_j \cup S''_{j-1}\,, \tag{2.14}$$

with $S''_{j-1} = \big\{S''_{j-1,k} = S'_{j-1,k} \cup \{d\} \;\big|\; k \in \big\{1, ..., \binom{d_2}{j}\big\}\big\}$ for $j \in \{1, ..., d-1\}$. Further to this, by definition of $\mathbf{W}_{j,k}$ we have the following three equalities:

$$\big(\mathbf{W}'_{j,k}, 1\big)_i = \begin{cases} u_i & \text{if } i \in I_1 \\ 1 - u_i & \text{if } i \in S'_{j,k} \\ 1 & \text{if } i \notin I_1 \cup S'_{j,k} \end{cases}, \qquad W_{j,k,i} = \begin{cases} u_i & \text{if } i \in I_1 \\ 1 - u_i & \text{if } i \in S_{j-1,k} \\ 1 & \text{if } i \notin I_1 \cup S_{j,k} \end{cases}$$

$$\text{and} \quad \left(\mathbf{W}'_{j-1,k}, 1-u_d\right)_i = \begin{cases} u_i & \text{if } i \in I_1 \\ 1-u_i & \text{if } i \in S''_{j-1,k} \\ 1 & \text{if } i \notin I_1 \cup S''_{j-1,k} \end{cases},$$

for $i \in \{1,...d\}$, $j \in \{1,...,d-1\}$ and $k \in \left\{1,...,\binom{d_2}{j}\right\}$. These three equalities and equation (2.14) imply that, for a fixed j, if we sum $C_{\mathbf{D}^+}$ evaluated in all of the $(\mathbf{W}'_{j,k}, 1)$ and $(\mathbf{W}'_{j,k}, 1-u_d)$ for different k, we get the sum of $C_{\mathbf{D}^+}$ evaluated on $\mathbf{W}_{j,k}$ for different k, that is:

$$\sum_{k=1}^{\binom{d_2-1}{j}} C_{\mathbf{D}^+}(\mathbf{W}'_{j,k}, 1) + \sum_{k=1}^{\binom{d_2-1}{j-1}} C_{\mathbf{D}^+}(\mathbf{W}'_{j-1,k}, 1-u_d) = \sum_{k=1}^{\binom{d_2}{j}} C_{\mathbf{D}^+}(\mathbf{W}_{j,k}), \tag{2.15}$$

for $j \in \{1,...,d-1\}$. Also, the equalities

$$\left(\mathbf{W}'_{0,1}, 1\right)_i = W_{0,1,i} \qquad \text{and} \qquad \left(\mathbf{W}'_{d-1,1}, 1-u_d\right)_i = W_{d,1,i}$$

hold for $i \in \{1,...d\}$; the result is implied by these two equalities and equations (2.13) and (2.15). □

Note that this expression is reflexible, meaning that it yields the same formula to express $C_{\mathbf{D}^+}$ in terms of $C_{\mathbf{D}^*}$. As a particular case, equation (2.8) can be used to express any associated copula in terms of the distributional copula C, which is the expression found in literature for copula models. A copula is said to be exchangeable if for every permutation $P\colon i \to p_i$ of $I = \{1,...,d\}$, we have $C(u_1,...,u_d) = C(u_{p_1},...,u_{p_d})$. In order to analyse the symmetry and exchangeability of copula models, we use the following definition.

Definition 2.5. Let $\mathbf{D} = (D_1,...,D_d)$ be a type of dependence, the complement dependence is defined as $\mathbf{D}^{\complement} = (D_1^{\complement},...,D_d^{\complement})$, with

$$D_i^{\complement} = \begin{cases} U & \text{if } D_i = L \\ L & \text{if } D_i = U \end{cases},$$

for $i \in \{1,...,d\}$. We say that the random vector $\mathbf{X}$, with associated copulas $\mathcal{A}_{\mathbf{X}}$, is complement (reflection or radial) symmetric, if there exists $\mathbf{D}^* \in \Delta$, such that $C_{\mathbf{D}^*} = C_{\mathbf{D}^{*\complement}}$.

Note that $\mathbf{X}$ is symmetric if there exists one dependence which satisfies $C_{\mathbf{D}^*} = C_{\mathbf{D}^{*\complement}}$. Along with other important properties, in the following proposition we prove that, if it holds for one dependence, it holds for them all.

Proposition 2.2. *Let $\mathbf{X}$ be a vector with corresponding associated copulas $\mathcal{A}_{\mathbf{X}}$, and let $\mathbf{D}^*$, $\mathbf{D}^+$, $\mathbf{D}^\circ$ and $\mathbf{D}^\times$ be types of dependencies. Denote as $I_1(\mathbf{D}^1, \mathbf{D}^2)$ and $I_2(\mathbf{D}^1, \mathbf{D}^2)$ the elements where the corresponding dependencies are equal or different respectively. Then the following equivalences hold:*

(**i**) *If* $C_{\mathbf{D}^*} \equiv C_{\mathbf{D}^+}$ *and* $I_2(\mathbf{D}^*, \mathbf{D}^+) = I_2(\mathbf{D}^\times, \mathbf{D}^\circ)$ *then* $C_{\mathbf{D}^\times} \equiv C_{\mathbf{D}^\circ}$. *In particular,* $C_{\mathbf{D}^*} \equiv C_{\mathbf{D}^{*\complement}}$, *for some* $\mathbf{D}^*$, *implies* $C_{\mathbf{D}} \equiv C_{\mathbf{D}^\complement}$ *for all* $\mathbf{D} \in \Delta$.

(**ii**) *If* $C_{\mathbf{D}^\circ}$ *is exchangeable, then* $C_{\mathbf{D}^*}$ *is exchangeable over the elements of* $I_1(\mathbf{D}^*, \mathbf{D}^\circ)$ *and over the elements of* $I_2(\mathbf{D}^*, \mathbf{D}^\circ)$. *In particular, if* $C_{\mathbf{D}^\circ}$ *is exchangeable, then* $C_{\mathbf{D}^{\circ\complement}}$ *is exchangeable.*

Proof: (**i**) This follows from the fact $I_2(\mathbf{D}^*, \mathbf{D}^+) = I_2(\mathbf{D}^\times, \mathbf{D}^\circ) \implies I_2(\mathbf{D}^\times, \mathbf{D}^*) = I_2(\mathbf{D}^\circ, \mathbf{D}^+)$, which is easily verified considering the different cases. From Proposition 2.1, we have that the vectors $\mathbf{W}_{j,k}$ are the same in both cases, which implies

$$\begin{aligned} C_{\mathbf{D}^\times}(u_1, ..., u_d) &= \sum_{j=0}^{d_2} (-1)^j \sum_{k=1}^{\binom{d_2}{j}} C_{\mathbf{D}^*}(\mathbf{W}_{j,k}) \\ &= \sum_{j=0}^{d_2} (-1)^j \sum_{k=1}^{\binom{d_2}{j}} C_{\mathbf{D}^+}(\mathbf{W}_{j,k}) \\ &= C_{\mathbf{D}^\circ}(u_1, ..., u_d) \, . \end{aligned}$$

In particular, note that $I_2(\mathbf{D}^*, \mathbf{D}^{*\complement}) = I_2(\mathbf{D}, \mathbf{D}^\complement) = \{1, ..., d\}$ for every $\mathbf{D} \in \Delta$. Then, $C_{\mathbf{D}^*} \equiv C_{\mathbf{D}^{*\complement}}$ implies $C_{\mathbf{D}} \equiv C_{\mathbf{D}^\complement}$ for every $\mathbf{D} \in \Delta$.

(**ii**) From Proposition 2.1 we have

$$C_{\mathbf{D}^*}(u_1, ..., u_d) = \sum_{j=0}^{d_2} (-1)^j \sum_{k=1}^{\binom{d_2}{j}} C_{\mathbf{D}^\circ}(\mathbf{W}_{j,k}) \, . \tag{2.16}$$

Consider $j \in \{0, ..., d_2\}$ and $k \in \left\{1, ..., \binom{d_2}{j}\right\}$, from the way it is defined, $W_{j,k,i} = u_i$ for every $i \in I_1(\mathbf{D}^*, \mathbf{D}^\circ)$. The exchangeability of $C_{\mathbf{D}^\circ}$ implies that $C_{\mathbf{D}^\circ}(\mathbf{W}_{j,k})$ is exchangeable over $I_1(\mathbf{D}^*, \mathbf{D}^\circ)$. Hence, equation (2.16) implies that $C_{\mathbf{D}^*}$ is exchangeable over $I_1(\mathbf{D}^*, \mathbf{D}^\circ)$. Now, let $j \in \{0, ..., d_2\}$ be fixed, note that each $\mathbf{W}_{j,k}$, $k \in \left\{1, ..., \binom{d_2}{j}\right\}$, is based on a different subset of size j of $I_2(\mathbf{D}^*, \mathbf{D}^\circ)$. Consider the sum $\sum_{k=1}^{\binom{d_2}{j}} C_{\mathbf{D}^\circ}(\mathbf{W}_{j,k})$ as a function, given that $C_{\mathbf{D}^\circ}$ is exchangeable and that the sum considers all the subsets of size j of $I_2(\mathbf{D}^*, \mathbf{D}^\circ)$, it follows that this function is exchangeable over $I_2(\mathbf{D}^*, \mathbf{D}^\circ)$. Equation (2.16) then implies that $C_{\mathbf{D}^*}$ is exchangeable over $I_2(\mathbf{D}^*, \mathbf{D}^\circ)$. In particular $C_{\mathbf{D}^{\circ\complement}}$ is exchangeable over $I_2(\mathbf{D}^\circ, \mathbf{D}^{\circ\complement}) = \{1, ..., d\}$. □

It is well known that elliptical copulas satisfy $C = \widehat{C}$. Hence, it follows that in the bivariate case, $C_{LU} = C_{UL}$ and in three dimensions, for instance, $C_{ULU} = C_{LUL}$. Also, from (ii), it follows that the survival copulas of Archimedean families are exchangeable in all dimensions. These examples illustrate some of the applications of this proposition.

In the following proposition we prove that, same as the distributional copula, all associated copulas are invariant under strictly increasing transformations.

Proposition 2.3. *Let T_1,..., T_d be strictly increasing functions and $\mathbf{X} = (X_1, ..., X_d)$ a random vector with corresponding distribution function and marginals, $\mathbf{D}$ a type of dependence and $\mathbf{D}$-copula $C_{\mathbf{D}}$. Then, in the continuous case,*

$$\widetilde{\mathbf{X}} = \big(T_1(X_1), ..., T_d(X_d)\big)$$

also has the same corresponding $\mathbf{D}$-copula $C_{\mathbf{D}}$.

Proof: This result follows straightforwardly from the fact that the distributional copula is invariant under strictly increasing transformations (see [20, Proposition 5.6]) as all associated copulas are implied by this copula using Proposition 2.1. □

In the bivariate case, [22, Theorem 2.4.4] and [5, Theorem 2.7], characterised the copula after the use of strictly monotone functions on random variables. In the multivariate case, this can be done using the associated copulas as we show in the following proposition.

Proposition 2.4. *Let T_1,..., T_d be strictly monotone functions and $\mathbf{X} = (X_1, ..., X_d)$ a random vector with corresponding distributional copula C. Then the distributional copula of $\widetilde{\mathbf{X}} = (T_1(X_1), ..., T_d(X_d))$ is the associated $\mathbf{D}$-copula $C_{\mathbf{D}}$ of $\mathbf{X}$, with*

$$D_i = \begin{cases} L & \text{if } T_i \text{ is strictly increasing} \\ U & \text{if } T_i \text{ is strictly decreasing} \end{cases},$$

for $i \in \{1, .., d\}$, whose expression is given by Proposition 2.1.

Proof: By using the inverse functions of T_i and F_i, $i \in \{1, ..., d\}$ we have:

$$T_i(X_i) \leq \big(\widetilde{F}_i^{\leftarrow}(u_i)\big) \overset{P}{\sim} \mathcal{B}_i\big(F_{D_i,i}^{\leftarrow}(u_i)\big) ,$$

for $i \in \{1, ..., d\}$, with $\mathcal{B}_i$ as in Definition 2.4, which implies that the distributional copula of $\widetilde{\mathbf{X}}$ is $C_{\mathbf{D}}$. □

2.2. Associated tail dependence functions and tail dependence coefficients

Considering the results obtained so far, it is possible to introduce a general definition of tail dependence function and tail dependence coefficients considering the dependence $\mathbf{D}$. For the analysis of the conditions of the existence of the tail dependence function see [21]. The general expression of the tail dependence function is the following (for the positive case, see [23])

Definition 2.6. Let $I = \{1, ..., d\}$, $\mathbf{X} = (X_1, ..., X_d)$ be a random vector with copula C, $\mathbf{D}$ a type of dependence and $C_{\mathbf{D}}$ the corresponding associated copula. For any $\emptyset \neq S \subseteq I$, let $C_{\mathbf{D}(S)}$ denote the corresponding marginal copula. Define the associated $\mathbf{D}(S)$-tail dependence functions $b_{\mathbf{D}(S)}$ of $C_{\mathbf{D}}$, $\emptyset \neq S \subseteq I$ as

$$b_{\mathbf{D}(S)}(w_i, i \in S) = \lim_{u \downarrow 0} \frac{C_{\mathbf{D}(S)}(u w_i,\, i \in S)}{u} \,, \qquad \forall w = (w_1, ..., w_d) \in \mathbb{R}_+^d \,.$$

Given that these functions come from the associated copulas, we call the set of all $\mathbf{D}$-tail dependence functions the associated tail dependence functions. When $S = \{1, ..., d\}$ we omit such subindex.

In particular, the corresponding TDCs are presented in the following definition (for the positive TDCs, see [23, 12]).

Definition 2.7. Consider the same conditions of Definition 2.6. Define the associated $\mathbf{D}(S)$-tail dependence coefficients $\lambda_{\mathbf{D}(S)}$ of $C_{\mathbf{D}}$, $\emptyset \neq S \subseteq I$ as

$$\lambda_{\mathbf{D}(S)} = \lim_{u \downarrow 0} \frac{C_{\mathbf{D}(S)}(u, ..., u)}{u} \,.$$

We say that $\mathbf{D}(S)$-tail dependence exists whenever $\lambda_{\mathbf{D}(S)} > 0$.

Note that

$$C_{\mathbf{D}(S)}(u, ..., u) = C_{\mathbf{D}}(u_1, ..., u_d) \geq C_{\mathbf{D}}(u, ..., u) \,,$$

with $u_i = \begin{cases} u & \text{if } i \in S \\ 1 & \text{if } i \notin S \end{cases}$, $i \in \{1, ..., d\}$. Because of this, $\lambda_{\mathbf{D}(S)} \geq \lambda_{\mathbf{D}}$, so $\mathbf{D}$-tail dependence implies $\mathbf{D}(S)$-tail dependence for all $\emptyset \neq S \subseteq I$.

3. MODELLING GENERAL DEPENDENCE

In this section we analyse general dependence and tail dependence in three examples of copula models. To this end we use the definitions and results obtained on the previous section. We first analyse the perfect dependence cases and obtain their corresponding copulas, this includes perfect non-positive dependence. We then study the elliptical copulas for which we characterise the associated copulas. Using this characterisation, we obtain an expression for the associated tail dependence functions of the Student's t copula, which accounts for all types of tail dependence simultaneously. After that we study the Archimedean copulas, we prove that they can only account for non-positive tail dependence when their generator is non-strict and present three examples when they do. At the end of the section we discuss a method for modelling general tail dependence using copula models. The analysis of general dependence presented in this section complements the analysis of positive tail dependence for these models.

3.1. Perfect dependence cases

We now analyse the most basic examples of copula models. They correspond to all the variables being either independent or perfectly dependent.

For the independence case, let $\mathbf{U} = (U_1, ..., U_d)$ be a random vector with $\{U_i\}_{i=1}^d$ independent uniform random variables. The distribution function of U is the copula $C(u_1, ..., u_d) = \prod_{i=1}^{d} u_i$, which is known as the independence copula. It follows that the associated copula are also equal to the independence copula. This is the copula of any random vector formed by independent variables.

Our analysis of perfect dependence corresponds to the distribution of vectors of the form $(W, -W, -W, ..., W, -W)$ with W a uniform random variable. From Definition 2.3 and Proposition 2.4 it follows that the distribution of a vector of this form is an associated copula of the vector $\mathbf{W} = (W, ..., W)$. The distributional copula of $\mathbf{W}$ is

$$C(u_1, ..., u_d) = \min\{u_i\}_{i=1}^d \,. \tag{3.1}$$

Given that $1 - W$ is also uniform it follows that this is also the survival copula, so the vector is symmetric. This copula is the comonotonic copula. Now, let $\mathbf{D}$ be a type of dependence and $I = \{1, ..., d\}$. Define $I_L = \{i \in I \mid D_i = L\}$ and $I_U = \{i \in I \mid D_i = L\}$. Let us assume that neither I_L nor I_U are empty. That is, we assume perfect non-positive dependence (the case of perfect positive dependence is covered in equation (3.1)). Then the associated $\mathbf{D}$-copula is

$$C_{\mathbf{D}}(u_1, ..., u_d) = P\Big(\big(W \leq \min\{u_i\}_{i\in I_L}\big) \cap \big(W \geq \max\{1 - u_i\}_{i\in I_U}\big)\Big) \,.$$

It follows that, for $\min\{u_i\}_{i\in I_L} > \max\{1-u_i\}_{i\in I_U}$, this probability is equal to zero; therefore, a general expression is

$$C_{\mathbf{D}}(u_1, ..., u_d) = \max\Big\{0, \min\{u_i\}_{i\in I_L} + \min\{u_i\}_{i\in I_U} - 1\Big\} \,. \tag{3.2}$$

In the bivariate case the associated (L, U)-copula C_{LU} is equal to the Fréchet lower bound for copulas, also known as the countermonotonic copula. Copulas of this form appear in perfect non-positive dependence, see [20, Example 5.22]. In the following proposition we prove that, in d dimensions, the copulas of (3.1) and (3.2) correspond not only to vectors of the form $(W, -W, W, ..., W, -W)$, but to the use of strictly monotone transformations on a random variable. Because of this, we call these copulas the monotonic copulas.

Proposition 3.1. *Let Z be a random variable, and let $\{T_i\}_{i=1}^d$ be strictly monotone functions, then the distributional copula of the vector $X =$*

$(T_1(Z), ..., T_d(Z))$ is one of the monotonic copulas of equations (3.1) or (3.2) with $\mathbf{D} = (D_1, ..., D_d)$,

$$D_i = \begin{cases} L & \text{if } T_i \text{ is strictly increasing} \\ U & \text{if } T_i \text{ is strictly decreasing} \end{cases} .$$

Conversely, consider a random vector $\mathbf{X} = (X_1, ..., X_d)$ whose distributional copula is a monotonic copula of equation (3.1) or (3.2) for certain $\mathbf{D}$. Then there exist monotone functions $\{T_i\}_{i=1}^d$ and a random variable Z such that

$$(X_1, ..., X_d) \overset{d}{=} \big(T_1(Z), ..., T_d(Z)\big) , \tag{3.3}$$

the $\{T_i\}_{i=1}^d$ satisfy that T_i is strictly increasing if $D_i = L$ and strictly decreasing if $D_i = U$ for $i \in \{1, ..., d\}$. In both cases the vector $\mathbf{X}$ is complement symmetric.

Proof: Let F be the distribution function of Z. Considering the uniform random variable $F(Z)$ it is clear that the copula of the d-dimensional vector $(Z, ..., Z)$ is the Fréchet upper bound copula $\min\{u_i\}_{i=1}^d$ of equation (3.1). The result is then implied by Proposition 2.4.

The converse statement is a generalisation of [5, Theorem 3.1]. We have that the distributional copula of $\mathbf{X}$ is a monotonic copula for certain $\mathbf{D}$. Note that the associated $\mathbf{D}$-copula of $\mathbf{X}$ is the Fréchet upper bound copula. Let $\{\alpha_i\}_{i=1}^d$ be any invertible monotone functions that satisfy α_i is strictly increasing if $D_i = L$ and strictly decreasing if $D_i = U$ for $i \in \{1, ..., d\}$. Proposition 2.4 implies that the copula of $\mathbf{A} = (\alpha_1(X_1), ..., \alpha_d(X_d))$ is the Fréchet upper bound copula. According to [9, 6], there exists a random variable Z and strictly increasing $\{\beta_i\}_{i=1}^d$ such that

$$\big(\alpha_1(X_1), ..., \alpha_d(X_d)\big) \overset{d}{=} \big(\beta_1(Z), ..., \beta_d(Z)\big) .$$

By defining $T_i = \alpha_i^{-1} \circ \beta_i$ for $i \in \{1, ..., d\}$ we get the result.

In both cases the associated copulas of $\mathbf{X}$ are the monotonic copulas implying that the vector is complement symmetric. □

Regarding tail dependence, suppose the vector $\mathbf{X}$ has distributional copula C^* equal to a monotonic copula $C_{\mathbf{D}}$ of equations (3.1) or (3.2) for certain $\mathbf{D}$. Considering Definition 2.3 of the associated copulas, this implies that $C^*_{\mathbf{D}}$ is the comonotonic copula. It follows that the $\mathbf{D}$ and $\mathbf{D}^{\complement}$ tail dependence functions of the vector $\mathbf{X}$ are

$$b^*_{\mathbf{D}}(w_1, ..., w_d) = b^*_{\mathbf{D}^{\complement}}(w_1, ..., w_d) = \min\{w_1, ..., w_d\} .$$

The other associated copulas satisfy equation (3.2) for some $\mathbf{D}^0$. It follows that the corresponding tail dependence functions are equal to zero.

3.2. Elliptically contoured copulas

We now analyse the dependence structure of elliptically contoured copulas. We present the definition of this model, a result for its corresponding associated copulas and the associated tail dependence functions of the Student's t copula.

Elliptical distributions, were introduced by [17] and have been analysed by several authors (see e.g. [8, 11]). They have the following form.

Definition 3.1. The random vector $\mathbf{X} = (X_1, ..., X_d)$ has a multivariate elliptical distribution, denoted as $\mathbf{X} \sim El_d(\mu, \Sigma, \psi)$, if for $\mathbf{x} = (x_1, ..., x_d)'$ its characteristic function has the form

$$\varphi(\mathbf{x}; \mu, \Sigma) \;=\; \exp(i\,\mathbf{x}'\mu)\;\psi_d\!\left(\frac{1}{2}\,\mathbf{x}'\Sigma\,\mathbf{x}\right),$$

with μ a vector, $\Sigma = (\sigma_{ij})_{1\leq i,j\leq d}$ a symmetric positive-definite matrix and $\psi_d(t)$ a function called the characteristic generator.

Elliptical contoured distributions include a large number of distributions (see [31, Appendix]). In the case when the joint density exists, several results have been obtained (see [11, 2, 19]). The corresponding copula is referred to as elliptical copula. This copula has also been subject to numerous analysis (see [7, 1, 5, 3]). Note that the process of standardising the marginal distributions of a vector uses strictly increasing transformations. From Proposition 2.3, we have that the copulas associated to $\mathbf{X} \sim El_d(\mu, \Sigma, \psi)$ are the same as the copulas associated to $\mathbf{X}^* \sim El_d(0, R, \psi)$. Here $R = \left(\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{iI}\sigma_{jJ}}}\right)_{1\leq i,j\leq d}$ is the corresponding "correlation" matrix implied by $\Sigma = (\sigma_{ij})_{1\leq i,j\leq d}$ (see [5, Theorem 5.2] or [7, 3]). Hence, we always assume $\mathbf{X} \sim El_d(R, \psi)$ with $R = (\rho_{ij})_{1\leq i,j\leq d}$.

In general, there is no closed-form expression for elliptical copulas but they can be expressed as multivariate integrals of the joint density. In the following proposition we prove an identity for the associated copulas of the elliptical copula.

Proposition 3.2. *Let $\mathbf{X} \sim El_d(R, \psi)$ as in Definition 3.1, with correlation matrix $R = (\rho_{ij})_{1\leq i,j\leq d}$, and let $\mathbf{D}$ be a type of dependence. Then the associated $\mathbf{D}$-copula of $\mathbf{X}$ is the same as the distributional copula of $\mathbf{X}^+ \sim El_d(\wp_{\mathbf{D}} R \wp'_{\mathbf{D}}, \psi)$, with $\wp_{\mathbf{D}}$ a diagonal matrix (all values in it are zero except for the values in its diagonal) $\wp_{\mathbf{D}} \in M_{d\times d}$, whose diagonal is $\mathbf{p} = (p_1, ..., p_d)$ with*

$$p_i \;=\; \begin{cases} 1 & \text{if } D_i = L \\ -1 & \text{if } D_i = U \end{cases},$$

for $i \in \{1, ..., d\}$.

Proof: The vector $\wp_{\mathbf{D}}\mathbf{X}$ is equal to $(T_1(X_1), ..., T_d(X_d))$ with $T_i(x) = p_i x$, $i \in \{1, ..., d\}$. Using Proposition 2.4, the distributional copula of $\wp_{\mathbf{D}}\mathbf{X}$ is the associated $\mathbf{D}$-copula of $\mathbf{X}$. From the stochastic representation of $\mathbf{X}$ (see [8]), it follows that $\wp_{\mathbf{D}}\mathbf{X} \sim El_d(\wp_{\mathbf{D}} R \wp'_{\mathbf{D}}, \psi)$ (see [5, Theorem 5.2]). □

Given that $C = \widehat{C}$ in elliptical copulas, we have that these copulas are symmetric. This can be easily verified considering that $\wp_{\mathbf{D}^{\complement}} = -\wp_{\mathbf{D}}$, for every dependence $\mathbf{D}$. This implies $\wp_{\mathbf{D}^{\complement}} \cdot R \cdot \wp'_{\mathbf{D}^{\complement}} = \wp_{\mathbf{D}} \cdot R \cdot \wp'_{\mathbf{D}}$. Hence, both $C_{\mathbf{D}}$ and $C_{\mathbf{D}^{\complement}}$ are equal to the distributional copula of $\mathbf{X}^+ \sim El_d(\wp_{\mathbf{D}} R \wp'_{\mathbf{D}}, \psi)$.

Proposition 3.2 makes it possible to use the results of elliptical copulas in associated copulas. This includes the analysis of tail dependence. In the bivariate case [18, 26] studied positive tail dependence in elliptical copulas under regular variation conditions. The Gaussian copula does not account for positive tail dependence, Proposition 3.2 implies that it does not account for tail dependence for any $\mathbf{D}$. In contrast the Student's t copula does account for tail dependence (see e.g. [14, 23, 3, 20]). The Student's t copula with ν degrees of freedom and correlation matrix R is expressed in terms of integrals and density $t_{\nu,R}$ as

$$C(\mathbf{u}) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_d)} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d\,|R|}} \left(1 + \frac{\mathbf{x}' R^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}} d\mathbf{x}\,,$$

with $\mathbf{u} = (u_1, ..., u_d)$ and $\mathbf{x} = (x_1, ..., x_d)'$. [23] analysed in detail the extreme value properties of this copula and obtained an expression for the lower and upper tail dependence functions among other results. More recently, in the bivariate case, [14] obtained an expression for the $\mathbf{D} = (L, U)$ and the $\mathbf{D} = (U, L)$ tail dependence coefficients proving that this copula accounts for negative tail dependence. We now present the expression for the associated $\mathbf{D}$-tail dependence function of the multivariate Student's t copula. This result follows from [23, Theorem 2.3] and Proposition 3.2.

Proposition 3.3. *Let $\mathbf{X} = (X_1, ..., X_d)$ have multivariate t distribution with ν degrees of freedom, and correlation matrix $R = (\rho_{ij})_{1 \le i,j \le d}$, that is $\mathbf{X} \sim T_{d,\nu,R}$. Let $\mathbf{D} = (D_1, ..., D_d)$ be a type of dependence. Then the associated $\mathbf{D}$-tail dependence function $b_{\mathbf{D}}$ is given by*

$$b_{\mathbf{D}}(w) = \sum_{j=1}^{d} w_j\, T_{d-1,\nu+1,R_j^*}\left(\sqrt{\frac{\nu+1}{1-\rho_{ij}^2}}\left[-\left(\frac{w_i}{w_j}\right)^{-\frac{1}{\nu}} + p_i p_j \rho_{ij}\right],\ i \in I_j\right),$$

with

$$R_j^* = \begin{pmatrix} 1 & \cdots & \rho^*_{1,j-1;j} & \rho^*_{1,j+1;j} & \cdots & \rho^*_{1,d;j} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \rho^*_{j-1,1;j} & \cdots & 1 & \rho^*_{j-1,j+1;j} & \cdots & \rho^*_{j-1,d;j} \\ \rho^*_{j+1,1;j} & \cdots & \rho^*_{j+1,j-1;j} & 1 & \cdots & \rho^*_{j+1,j-1;j} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^*_{d,1;j} & \cdots & \rho^*_{d,j-1;j} & \rho^*_{d,j+1;j} & \cdots & 1 \end{pmatrix};$$

$\rho^*_{i,k;j} = p_i p_k \frac{\rho_{ik}-\rho_{ij}\,\rho_{kj}}{\sqrt{1-\rho^2_{ij}}\sqrt{1-\rho^2_{kj}}}$, *the modified partial correlations;* $I_j = I - \{j\}$ *and*

$$p_j = \begin{cases} 1 & \text{if } D_j = L \\ -1 & \text{if } D_j = U \end{cases},$$

for $j \in \{1, ..., d\}$.

Proof: Proposition 3.2 implies that the associated **D**-tail dependence function of the random vector $\mathbf{X} \sim T_{d,\nu,R}$ is the lower tail dependence function of the vector $\mathbf{X}^+ \sim T_{d,\nu,\wp_{\mathbf{D}} R \wp'_{\mathbf{D}}}$. The modified correlation matrix is $\wp_{\mathbf{D}} R \wp'_{\mathbf{D}} = R^* = (\rho^*_{ij})_{1\le i,j\le d}$, it follows that

$$(\rho^*_{ij})_{1\le i,j\le d} = (p_i p_j \rho_{ij})_{1\le i,j\le d}\,.$$

Hence $(\rho^*_{ij})^2 = p_i^2 p_j^2 \rho_{ij}^2 = 1 \cdot 1 \cdot \rho_{ij}^2 = \rho_{ij}^2$. Under this change, the partial correlations are modified as follows:

$$\rho^*_{i,k;j} = p_i p_k \frac{\rho_{ik} - \rho_{ij}\,\rho_{kj}}{\sqrt{1-\rho^2_{ij}}\sqrt{1-\rho^2_{kj}}}\,.$$

The result is then implied by [23, Theorem 2.3]. □

This proposition implies that the Student's t copula accounts for all 2^d dependencies simultaneously. It can happen that we have negative dependence and positive tail dependence. In that case, the variables might generally exhibit negative dependence but, when it comes to extreme values, they can also be positively dependent.

3.3. Archimedean copulas

Now we analyse the dependence structure of Archimedean copulas. We present the bivariate and multivariate definition of these copulas. We then prove that, when the generator is strict, they can only account for positive tail dependence. Finally, we present three examples with non-strict generators that account for negative tail dependence. For the analysis of positive tail dependence in these copulas we refer to [15, Propositions 2.5 and 3.3], [13, Theorems 4.12 and 4.15] and [22, Corollary 5.4.3]

Much has been written on Archimedean copulas and their applications to different areas of statistics. [28] provide an excellent monography of their history. For further references on their analysis we refer to the seminal works of [13, 22]. [13] analyses several examples with strict generators and [22] extends the analysis

to non-strict generators. In order to consider both cases, we follow the notation used in [22].

A bivariate Archimedean copula is defined in terms of a generator, which we denote φ, in the following way:

$$C(u_1, u_2) \;=\; \varphi^{[-1]}\big(\varphi(u_1) + \varphi(u_2)\big)\;, \tag{3.4}$$

where $\varphi^{[-1]}(u) = \begin{cases} \varphi^{-1}(u) & \text{if } 0 \leq u \leq \varphi(0) \\ 0 & \text{if } \varphi(0) \leq u \leq \infty \end{cases}$, is the pseudo-inverse of φ. In order for this function to be a copula, the generator must satisfy the following properties:

i) $\varphi\colon [0,1] \to R^+ \cup \infty$,

ii) φ is continuous, strictly decreasing and convex,

iii) $\varphi(1) = 0$.

φ is called a strict generator when $\varphi(0) = \infty$. Note that, when φ is strict, $\varphi^{[-1]} = \varphi^{-1}$. ([3])

[16] proved that a strict generator gives a copula in any dimension d if and only if the generator inverse φ^{-1} is completely monotonic. In that case, the multivariate Archimedean copula is defined as

$$C(u_1, ..., u_d) \;=\; \varphi^{-1}\left(\sum_{i=1}^{d} \varphi(u_i)\right), \tag{3.5}$$

In the next proposition we prove that, by construction, Archimedean copulas with strict generators, do not account for any non-positive tail dependence.

Proposition 3.4. *Let C be an Archimedean copula with differentiable strict generator φ and let $\mathbf{D}$ be a non-positive type of dependence. Then, if the corresponding tail dependence function $b_{\mathbf{D}}$ exists, it is equal to zero.*

Proof: Let C be a bivariate Archimedean copula with strict generator φ. As we pointed out before, given that φ is strict, $\varphi^{[-1]} = \varphi^{-1}$. We begin this proof with the bivariate case and prove that $\lambda_{LU} = 0$.

Let $G(h) = \dfrac{\varphi^{-1}\big(\varphi(h) + \varphi(1-h)\big)}{h}$, by definition

$$\begin{aligned} \lambda_{LU} \;&=\; \lim_{h \to 0} \frac{C_{LU}(h,h)}{h} \\ &=\; \lim_{h \to 0} \frac{h - C(h, 1-h)}{h} \\ &=\; 1 - \lim_{h \to 0} G(h)\;. \end{aligned} \tag{3.6}$$

[3] In [13], the construction of Archimedean copulas covers the strict generator case when φ^{-1} is a Laplace transform, they denote such Laplace transform as ϕ.

Along with the three properties of the generator φ mentioned above, in this case it is strict and differentiable. This implies the following for φ^{-1}:

i) φ^{-1} is differentiable,
ii) φ^{-1} is strictly decreasing and convex,
iii) $\lim_{s\to\infty} \varphi^{-1}(s) = 0$.

Note that property iii) is only satisfied when the generator is strict, the behaviour of φ^{-1} around ∞ is fundamental in this proof. If we visualise the graphic of a function with such three features, it is intuitively straightforward that the slope of its tangent will tend to zero as $s \to \infty$, that is $\lim_{s\to\infty} (\varphi^{-1})'(s) = 0$. To prove this, note that, from ii), $(\varphi^{-1})'$ is always negative and increasing. This implies $(\varphi^{-1})'(s)$ converges, as $s \to \infty$, to $c \leq 0$. Suppose $c < 0$, this would imply that φ^{-1} crosses the x-axis. So it follows that $\lim_{s\to\infty} (\varphi^{-1})'(s) = 0$. Hence, we have

$$\begin{aligned} \lim_{s\to\infty} (\varphi^{-1})'(s) &= \lim_{x\to\infty} \lim_{y\to 0} \frac{\varphi^{-1}(x+y) - \varphi^{-1}(x)}{y} \\ &= 0\,. \end{aligned} \tag{3.7}$$

Also, φ is differentiable, strictly decreasing and $\varphi(1) = 0$, hence we have

$$-\infty < \varphi'(1) < 0\,. \tag{3.8}$$

If we take $x(h) = \varphi(h)$ and $y(h) = \varphi(1-h)$ in equation (3.7), we get:

$$\begin{aligned} 0 &= \lim_{h\to 0} \frac{\varphi^{-1}\big(\varphi(h) + \varphi(1-h)\big) - \varphi^{-1}\big(\varphi(h)\big)}{\varphi(1-h)} \\ &= \lim_{h\to 0} \frac{h\,G(h) - h}{\varphi(1-h) - \varphi(1)} \\ &= \frac{\lim_{h\to 0} 1 - G(h)}{\varphi'(1)}\,. \end{aligned}$$

From equation (3.6) and inequality (3.8), this this implies $\lambda_{LU} = 0$. Analogously, we get $\lambda_{UL} = 0$. The multivariate extension is straightforward: let C be a multivariate Archimedean copula and $\mathbf{D}$ a non-positive dependence. Then, there exist $i_1 < i_2$ such that $D_{i_1} \neq D_{i_2}$. Let $C_{(i_1,i_2)}$ be the bivariate marginal copula of C. Hence $\lambda_{(i_1,i_2),(D_{i_1},D_{i_2})} \geq \lambda_{\mathbf{D}}$ and, given that $C_{(i_1,i_2)}$ is also Archimedean, it satisfies $\lambda_{(i_1,i_2),(L,U)} = \lambda_{(i_1,i_2),(U,L)} = 0$. Then $\lambda_{\mathbf{D}} = 0$ follows. □

The same holds for other multivariate constructions based on nesting of Archimedean copulas, such as the ones described in [13, Section 4.2].

When the generator is non-strict, Archimedean copulas can account for non-positive tail dependence. This is the case in the three bivariate examples

presented in Table 2. These examples can be found in [22, Section 4.2]. The first two examples are the one-parameter copulas 4.2.7 and 4.2.8 in [22]. The third example is a two-parameter family of copulas known as the rational Archimedean copulas. The construction of these copulas can be found in [22, Subsection 4.5.2]. The expression is equation (4.5.9) and the generator is studied in p. 149 therein.

Table 2: Examples of Archimedean copulas with non-strict generators that account for negative tail dependence.

Generator $\varphi(s)$	Copula	b_{LU} and b_{UL}
$-\ln(\theta s+1-\theta)$, $0<\theta\leq 1$	$\max\Big\{\theta u_1u_2+(1-\theta)(u_1+u_2-1),\,0\Big\}$	$\min\Big\{w_1,\,(1-\theta)w_2\Big\}$, $\min\Big\{(1-\theta)w_1,\,w_2\Big\}$
$\frac{1-s}{1+(\theta-1)s}$, $\theta\geq 1$	$\max\Big\{\frac{\theta^2u_1u_2-(1-u_1)(1-u_2)}{\theta^2-(\theta-1)^2(1-u_1)(1-u_2)},\,0\Big\}$	$\min\Big\{w_1,\,\frac{w_2}{\theta^2}\Big\}$, $\min\Big\{\frac{w_1}{\theta^2},w_2\Big\}$
see [22, p. 149], $0\leq\beta\leq 1-\lvert\alpha\rvert$	$\max\Big\{\frac{u_1u_2-\beta(1-u_1)(1-u_2)}{1-\alpha(1-u_1)(1-u_2)},\,0\Big\}$	$\min\Big\{w_1,\,\beta w_2\Big\}$, $\min\Big\{\beta w_1,\,w_2\Big\}$

3.4. Use of rotations to model general tail dependence

We now discuss a method to model an arbitrary type of tail dependence using a copula model. The condition on the copula model is to account for, at least, one type of tail dependence. Similar procedures have been suggested in [25, Section 2.4] and [14, Example 8.1]. To illustrate how this procedure works, consider the bivariate Generalised Clayton copula, C^{GC} ([4]). This Archimedean copula accounts for upper tail dependence. Suppose that we are trying to model data that exhibits lower-upper tail dependence with a model C^* and want to use C^{GC} and the fact that it accounts for upper tail dependence. The use of this procedure implies defining $C^*_{LU}=\widehat{C}^{GC}$. And it holds that C^* accounts for lower-upper tail dependence. Using Proposition 2.1, $C^*(u_1,u_2)=u_1-C^{GC}(1-u_1,u_2)$. Note that the fact that C^{GC} also accounts for lower tail dependence implies that C^* accounts for upper-lower tail dependence. So, before using this technique, the whole dependence structure of the model and the data must be analysed.

We generalise this idea to model arbitrary $\mathbf{D}^\circ$-tail dependence using a copula model C that accounts for $\mathbf{D}^+$-tail dependence. Let $\mathcal{A}_{\mathbf{X}}=\{C_{\mathbf{D}}\,|\,\mathbf{D}\in\Delta\}$ be the associated copulas of model C, we know that $\lim\limits_{h\to 0}\frac{C_{\mathbf{D}^+}(h,\ldots,h)}{h}>0$. Now, define a $\mathbf{D}^\circ$-associated copula as $C^*_{\mathbf{D}^\circ}=C_{\mathbf{D}^+}$. By construction, as in the example, this

[4] $C^{GC}_{\theta,\delta}(u,v)=\Big\{\big[(u^{-\theta}-1)^\delta+(v^{-\theta}-1)^\delta\big]^{\frac{1}{\delta}}+1\Big\}^{-\frac{1}{\theta}}$.

copula model accounts for $\mathbf{D}^\circ$-tail dependence. The associated copulas, $\mathcal{A}^*_{\mathbf{X}} = \{C^*_{\mathbf{D}} \mid \mathbf{D} \in \Delta\}$, of this model can be obtained from $C^*_{\mathbf{D}^\circ}$, using Proposition 2.1. Note that the set $\mathcal{A}^*_{\mathbf{X}}$ is the same as $\mathcal{A}_{\mathbf{X}}$, but with rotated dependencies. The whole dependence structure of model C^* is implied by C.

4. CONCLUSIONS AND FUTURE WORK

In this section we discuss the main findings of this work and some future lines of research. In Section 2 we introduce the concepts to analyse, in the multivariate case, the whole dependence structure among random variables. We consider the 2^d different orthants of dimension d. We first introduce general dependence, the $\mathbf{D}$-probability functions and the associated copulas. We then present a version of Sklar's theorem that proves that the associated copulas link the $\mathbf{D}$-probability functions with their marginals. It is through this result that we are able to generalise the use of the distributional and survival copulas for positive dependence. In this generalisation we use the associated copulas to cover general dependence. We introduce an expression for the relationship among all associated copulas and present a proposition regarding symmetry and exchangeability. After that, we prove that they are invariant under strictly increasing transformations and characterise the copula of a vector after using monotone transformations. At the end of this section, we introduce the associated tail dependence functions and associated tail dependence coefficients of a random vector. With them we can analyse tail dependence in the different orthants.

In Section 3 we use the concepts and results obtained in Section 2 to analyse three examples of copula models. The first example corresponds to the perfect dependence models. We begin this analysis with the independence case and then consider perfect dependence, including perfect non-positive dependence. We find and expression for their copulas, which are a generalisation of the Fréchet copula bounds of the bivariate case. Given that they correspond to the use of strictly monotone transformations on a random variable, we call them the monotonic copulas. The second example corresponds to the elliptical copulas. In this case, we characterise the corresponding associated copulas. We then present an expression for the associated tail dependence function of the Student's t copula. This result proves that this copula model accounts for tail dependence in all orthants. The third example corresponds to Archimedean copulas. In this case, we prove that, if their generator is strict, they can only account for positive tail dependence. We then present three examples of Archimedean copulas with non-strict generators that account for negative tail dependence. After that we discuss a method for modelling arbitrary tail dependence using copula models.

There are several areas where future research regarding general dependence is worth being pursued. For instance, the use of $\mathbf{D}$-probability functions is not

restricted to copula theory. The analysis of probabilities in the multivariate case has sometimes been centered in distribution functions, but, just like survival functions, **D**-probability functions can serve different purposes in dependence analysis. Another possibility is the use of nonparametric estimators to measure non-positive tail dependence, as the use of these estimators has been restricted to the lower and upper cases. The results obtained in this work are useful in the understanding of the dependence structure implied by different copula models. As we have seen, without analysing general dependence, the analysis of these models is incomplete. Therefore, it is relevant to extend this analysis to models such as the hierarchical Archimedean copulas and vine copulas. The use of vine copulas has proven to provide a flexible approach to tail dependence and account for asymmetric positive tail dependence (see e.g. [24, 15]).

ACKNOWLEDGMENTS

This paper is based on results from the first chapters of my doctorate thesis supervised by Dr. Wing Lon Ng. I would like to thank Mexico's CONACYT, for the funding during my studies. I would also like to thank my examiners Dr. Aristidis Nikoloulopoulos and Dr. Nick Constantinou for the corrections, suggestions and comments that have made this work possible. I am also very grateful to the editor and an anonymous referee for the helpful comments and suggestions that led to an improvement of the paper.

REFERENCES

[1] Abdous, B.; Genest, C. and Rémillard, B. (2005). *Dependence properties of meta-elliptical distributions*. In "Statistical Modelling and Analysis for Complex Data Problems" (P. Duchesne and B. Rémillard, Eds.), Kluwer, Dordrecht, 1–15.

[2] Das Gupta, S.; Eaton, M.L.; Olkin, I.; Perlman, M.; Savage, L.J. and Sobel, M. (1972). *Inequalities on the probability content of convex regions for elliptically contoured distributions*. In "Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 2" (L. Le Cam, J. Neyman and E. Scott, Eds.), University of California Press, Berkeley, 241–264.

[3] Demarta, S. and McNeil, A.J. (2005). The t copula and related copulas, *International Statistical Review*, **73**(1), 111–129.

[4] Embrechts, P.; Lambrigger, D. and Wuthrich, M.V. (2009). Multivariate extremes and aggregation of dependent risks: examples and counter-examples, *Extremes*, **12**(1), 107–127.

[5] Embrechts, P.; Lindskog, F. and McNeil, A. (2001). *Modelling dependencies with copulas and applications to risk management*. In "Handbook of Heavy

Tailed Distributions in Finance, Chapter 8" (S.T. Rachev, Ed.), Elsevier, North Holland, 329–384.

[6] Embrechts, P.; McNeil, A.J. and Straumann, D. (2002). *Correlation and dependency in risk management: properties and pitfalls*. In "Risk Management: Value at Risk and Beyond" (M. Dempster, Ed.), Cambridge University Press, Cambridge, 176–223.

[7] Fang, H.B.; Fang, K.T. and Kotz, S. (2002). The meta-elliptical distributions with given marginals, *Journal of Multivariate Analysis*, **82**(1), 1–16.

[8] Fang, K.T.; Kotz, S. and Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

[9] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données, *Annales de l'Université de Lyon, Sciences Mathématiques et Astronomié*, **3**(14), 53–77.

[10] Georges, P.; Lamy, A.G.; Nicolas, E.; Quibel, G. and Roncalli, T. (2001). Multivariate survival modelling: a unified approach with copulas, *Groupe de Recherche Opérationnelle Crédit Lyonnais France*, Unpublished results.

[11] Gupta, A.K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics*, Kluwer Academic Publishers, Netherlands.

[12] Joe, H. (1993). Parametric families of multivariate distributions with given margins, *Journal of Multivariate Analysis*, **46**(2), 262–282.

[13] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

[14] Joe, H. (2011). *Tail dependence in vine copulae*. In "Dependence Modeling: Vine Copula Handbook, Chapter 8" (D. Kurowicka and H. Joe, Eds.), World Scientific, Singapore, 165–189.

[15] Joe, H.; Li, H. and Nikoloulopoulos, A.K. (2010). Tail dependence functions and vine copulas, *Journal of Multivariate Analysis*, **101**(1), 252–270.

[16] Kimberling, C.H. (1974). A probabilistic interpretation of complete monotonicity, *Aequationes Mathematicae*, **10**(2-3), 152–164.

[17] Kelker, D. (1970). Distribution theory of spherical distributions and location-scale parameter generalization, *Sankhya: The Indian Journal of Statistics, Series A*, **32**(4), 419–430.

[18] Klüppelberg, C.; Kuhn, G. and Peng, L. (2008). Semi-parametric models for the multivariate tail dependence function - the asymptotically dependent case, *Scandinavian Journal of Statistics*, **35**(4), 701–718.

[19] Landsman, Z. and Valdez, E.A. (2003). Tail conditional expectations for elliptical distributions, *North American Actuarial Journal*, **7**(4), 55–71.

[20] McNeil, A.; Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance, New Jersey.

[21] Mikosch, T. (2006). Copulas: tales and facts, *Extremes*, **9**(1), 3–20.

[22] Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd edn, Springer, New York.

[23] Nikoloulopoulos, A.K.; Joe, H. and Li, H. (2009). Extreme value properties of multivariate t copulas, *Extremes*, **12**(2), 129–148.

[24] Nikoloulopoulos, A.K.; Joe, H. and Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data, *Computational Statistics and Data Analysis*, **56**(11), 3659–3673.

[25] Nikoloulopoulos, A.K. and Karlis, D. (2010). Modeling multivariate count data using copulas, *Communications in Statistics-Simulation and Computation*, **39**(1), 172-187.

[26] Schmidt, R. (2002). Tail dependence for elliptically contoured distributions, *Mathematical Methods of Operations Research*, **55**(2), 301–327.

[27] Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence, *Scandinavian Journal of Statistics*, **33**(2), 307–335.

[28] Schweizer B. (1991). *Thirty years of copulas*. In "Advances in Probability Distributions with Given Marginals" (G. Dall'Aglio, S. Kotz and G. Salinetti, Eds.), Kluwer, Dordrecht, 13–50.

[29] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs margés, *Publications de l'Institut de Statistique de L'Université de Paris*, **8**(1), 229–231.

[30] Taylor, M.D. (2007). Multivariate measures of concordance, *Annals of the Institute of Statistical Mathematics*, **59**(4), 789–806.

[31] Valdez, E.A. and Chernih, A. (2003). Wang's capital allocation formula for elliptically-contoured distributions, *Insurance: Mathematics and Economics*, **33**(3), 517–532.

[32] Zhang, M.H. (2007). Modelling total tail dependence along diagonals, *Insurance: Mathematics and Economics*, **42**(1), 77–80.

REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 29–48

GOODNESS OF FIT TESTS AND POWER COMPARISONS FOR WEIGHTED GAMMA DISTRIBUTION

Authors: Neetu Singla
– Department of Statistics, Panjab University, Chandigarh-160014, India
neetu.singla12@gmail.com

Kanchan Jain
– Department of Statistics, Panjab University, Chandigarh-160014, India
jaink14@gmail.com

Suresh Kumar Sharma
– Department of Statistics, Panjab University, Chandigarh-160014, India
ssharma643@yahoo.co.in

Received: December 2013 Revised: July 2014 Accepted: September 2014

Abstract:

- In this paper, a weighted version of Gamma distribution known as Weighted Gamma (WG) distribution has been considered. Various tests of goodness of fit viz Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling have been applied to this family. Monte Carlo simulations have been carried out for power calculations. The powers of these tests have been compared which helps in ranking of these goodness of fit tests.

Key-Words:

- *Weighted Gamma; Kolmogorov–Smirnov; Cramér–von Mises and Anderson–Darling; power.*

AMS Subject Classification:

- 62E15, 62F03.

1. INTRODUCTION

Goodness of fit tests (GOFTs) validate the closeness of the theoretical distribution function to the empirical distribution function. They are also known as empirical distribution function tests. These tests determine how well the distribution under study fits to a data set. They are used to test simple hypothesis which completely specifies the model and also the composite hypotheses where only the name of the model/distribution is stated but not its parameters. In the latter case, the parameters are estimated from the data. The common GOFTs are Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling.

In literature, many authors have studied the goodness of fit tests. Nikulin [21,22] studied Chi-squared test for continuous distributions. Rao and Robson [25] studied Chi-squared statistic for exponential family. Power of a series of goodness of fit tests for simple and complex hypotheses have been analyzed by Lemeshko *et al.* [14,15]. Lemeshko *et al.* [16] analyzed the goodness of fit test for Inverse Gaussian family. Goodness of fit tests for testing composite hypotheses, using maximum likelihood estimators (MLEs) of double exponential distribution, have been given in Lemeshko and Lemeshko [17].

The idea of weighted distributions was conceptualized by Fisher [6] and studied by Rao [24] in a unified manner who pointed out that in many situations, the recorded observations cannot be considered as a random sample from the original distribution. This can be due to one or the other reason viz non-observability of some events, damage caused to original observations and adoption of unequal probability sampling. In observational studies for human, wild-life, insect, plant or fish population, it is not possible to select sampling units with equal probabilities. In such cases, there are no well-defined sampling frames and recorded observations are biased. These observations do not follow the original distribution and hence their modelling uses the theory of weighted distributions. It is, therefore, important to study the stochastic orderings and ageing properties of the weighted random variables with respect to the original random variables.

For a non-negative random variable X with pdf $f(x)$, the weighted random variable X^w has the pdf given by

$$f^w(x) = \frac{w(x)\, f(x)}{E\big[w(X)\big]}\ , \tag{1.1}$$

where $w(x)$ is a non-negative weight function such that $E[w(X)]$ is non-zero and finite. The distribution of X^w is called the weighted distribution corresponding to X.

The weighted distribution with $w(x) = x$ is called the length-biased (size-biased) distribution which finds various applications in biomedical areas such as early detection of a disease. Rao [24] used this distribution in the study of human

families and wild-life populations. Various other important weighted distributions and their properties have been discussed by Mahfoud and Patil [19], Jain *et al.* [12], Gupta and Kirmani [10], Nanda and Jain [20], Patil [23] and Gupta and Kundu [11].

A brief discussion of weighted version of Gamma distribution labelled as Weighted Gamma (WG) distribution is provided in Section 2. This distribution has been introduced by Jain *et al.* [13]. The Weighted Gamma (WG) distribution has Weighted exponential, Gamma and Exponential distributions as its submodels. This distribution can also be interpreted as a hidden upper truncation model as in case of skew-normal distribution (Arnold and Beaver [2]). The pdf of WG distribution is also expressible as a linear combination of two Gamma pdfs. This distribution accommodates increasing and upside-down bathtub shaped failure rate function and hence has wider applicability in reliability and survival analysis.

The motive of this study is to carry out goodness of fit tests viz Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling and to compare their powers for Weighted Gamma and some competing distributions namely Weighted Weibull, Weighted Exponential and Gamma distributions. Using the calculated powers of these goodness of fit tests, we can determine the sample size at which these various closely related distributions can be distinguished from each other.

The paper is organized as follows. In Section 2, we provide a brief description of Weighted Gamma (WG) distribution. Various goodness of fit tests have been described in Section 3. Testing of simple and composite hypotheses for WG versus Weighted Weibull (WW), Weighted Exponential (WE) and Gamma is presented in Section 4. This section also consists of results and power studies based on simulations and real data set analysis. Section 5 includes the concluding remarks.

2. WEIGHTED GAMMA DISTRIBUTION

The random variable X is said to follow Weighted Gamma distribution with scale parameter λ and shape parameters α and β if the probability density function (pdf) of X is given by

$$f_X(x;\alpha,\beta,\lambda) = k\,\frac{(1-\mathrm{e}^{-\alpha\lambda x})\,\lambda^{\beta}\,x^{\beta-1}\,\mathrm{e}^{-\lambda x}}{\Gamma(\beta)}\,, \qquad x>0\,,\quad \alpha,\beta,\lambda>0\,, \tag{2.1}$$

where $k^{-1} = 1-\left(\frac{1}{1+\alpha}\right)^{\beta}$.

If X is a random variable with pdf given in (2.1), we use the notation $X \sim WG(\alpha,\beta,\lambda)$.

The distribution function of X can be written as

$$F(x) = \left[\frac{(1+\alpha)^{\beta}}{(1+\alpha)^{\beta}-1}\right]\left[G(x;\beta,\lambda) - \frac{1}{(1+\alpha)^{\beta}}\,G(x;\beta,\lambda(1+\alpha))\right], \tag{2.2}$$

where $G(x;a,b) = \dfrac{b^a \int_0^x \mathrm{e}^{-bt}t^{a-1}dt}{\Gamma(a)}$ is the cumulative distribution function of Gamma distribution with shape parameter a and scale parameter b.

Remark 2.1. (2.1) is the weighted version of the Gamma pdf with weight function

$$w(x) = 1 - \mathrm{e}^{-\alpha\lambda x}\ , \qquad \alpha,\lambda > 0\ .$$

The choice of the weight function has been made so that Weighted Exponential (Gupta and Kundu [11]), Gamma and Exponential distributions can be obtained as special cases of WG distribution for particular values of parameters. The special cases are:

- Weighted Exponential (WE) distribution obtained by putting $\beta = 1$,
- the Gamma distribution when $\alpha \to \infty$.
- For $\alpha \to \infty$ and $\beta = 1$, Exponential distribution can be obtained.

Suppose X follows WG distribution and let $\boldsymbol{\theta} = (\alpha,\beta,\lambda)^T$ be the parameter vector. The log likelihood based on the observed sample $(x_1, x_2, ..., x_n)$ is

$$\begin{aligned} l &= l\,(\alpha,\beta,\lambda) \\ &= n\left\{\log(1+\alpha)^{\beta} - \log\{(1+\alpha)^{\beta}-1\}\right\} + \sum_{i=1}^{n}\log\left(1-\mathrm{e}^{-\alpha\lambda x_i}\right) + n\,\beta\log\lambda \\ &\quad + (\beta-1)\sum_{i=1}^{n}\log x_i - \lambda\sum_{i=1}^{n}x_i - n\log\{\Gamma(\beta)\}\ . \end{aligned} \tag{2.3}$$

The first derivative of the log likelihood function is called Fisher's score function and is written as

$$\boldsymbol{u}(\boldsymbol{\theta}) = \frac{\partial l}{\partial \boldsymbol{\theta}}\ .$$

Score is a vector of first partial derivatives, one for each element of $\boldsymbol{\theta}$. If the log likelihood is concave, then MLEs can be obtained by solving the system of equations

$$\boldsymbol{u}(\boldsymbol{\theta}) = \mathbf{0}\ ,$$

where elements of $\boldsymbol{u}(\boldsymbol{\theta})$ are given by

$$\frac{\partial l}{\partial \alpha} = -n\,\beta(1+\alpha)^{\beta-1}\left\{\frac{1}{\{(1+\alpha)^{\beta}-1\}(1+\alpha)^{\beta}}\right\} + \lambda\sum_{i=1}^{n}\frac{x_i\,\mathrm{e}^{-\alpha\lambda x_i}}{1-\mathrm{e}^{-\alpha\lambda x_i}}\ , \tag{2.4}$$

$$\frac{\partial l}{\partial \beta} = n \log(1+\alpha) - \frac{n(1+\alpha)^{\beta} \log(1+\alpha)}{\left\{(1+\alpha)^{\beta} - 1\right\}} + n \log \lambda + \sum_{i=1}^{n} \log x_i - n\, \psi(\beta) \,, \tag{2.5}$$

$$\frac{\partial l}{\partial \lambda} = \alpha \sum_{i=1}^{n} \frac{x_i \mathrm{e}^{-\alpha\lambda x_i}}{1 - \mathrm{e}^{-\alpha\lambda x_i}} + \frac{n\,\beta}{\lambda} - \sum_{i=1}^{n} x_i \,, \tag{2.6}$$

where $\psi(\cdot)$ denotes the digamma function, the logarithmic derivative of the gamma function.

As these equations are difficult to be solved, Newton–Raphson method can be used for finding ML estimates. Using this method, the score function is evaluated at the MLE $\hat{\boldsymbol{\theta}}$ around an initial value $\boldsymbol{\theta}_0$, using a first order Taylor series which gives

$$\boldsymbol{u}(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{u}(\boldsymbol{\theta}_0) + \frac{\partial \boldsymbol{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \,. \tag{2.7}$$

Equating (2.7) to zero and solving for $\hat{\boldsymbol{\theta}}$ leads to first approximation:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 - \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\, \boldsymbol{u}(\boldsymbol{\theta}_0) \,, \tag{2.8}$$

where

$$\boldsymbol{H}(\boldsymbol{\theta}) = \frac{\partial^2 l}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} = \frac{\partial \boldsymbol{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

denotes the Hessian matrix.

Given a trial value, (2.8) is employed for obtaining an improved estimate and the process is repeated until the differences between successive estimates are sufficiently close to zero. The estimates obtained are considered as maxima if the Hessian matrix is negative definite, that is, all its eigenvalues are negative.

As sometimes, it is computationally difficult to invert the Hessian matrix, hence we use the quasi Newton method in R for finding the ML estimates as this method usually generates an estimate of H^{-1} directly. The results have been included in Table 2 of Section 4.

3. GOODNESS OF FIT TESTS

For a random sample of size n, let $x_{(1)}, ..., x_{(n)}$ be ordered observations. The empirical distribution function (edf) $F_n(x)$ is a step function with a step of height $\frac{1}{n}$ at each ordered sample observation. Empirical Distribution Function (EDF) tests measuring the distance between the edf and theoretical cdf are described by Dufour *et al.* [6]. Arshad *et al.* [3] and Seier [26] claimed that the widely used EDF tests are Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling tests.

For a random variable X, we let $F(x)$ to be the theoretical cumulative distribution function (cdf). $F(x,\theta)$ denotes the cdf for a particular distribution with parameter θ. The focus shall be on testing the following types of null hypotheses:

- **Simple null hypothesis:**

  H_0: $F(x) = F(x,\theta)$,
  where the form of $F(x,\theta)$ is completely specified;

- **Composite null hypothesis:**

  H_0: $F(x) \in \big\{F(x,\theta),\ \theta \in \Theta\big\}$,
  where Θ is the domain of unknown parameter θ which is replaced by its estimator.

We will use the tests explained in the subsequent discussion.

Kolmogorov–Smirnov Test:

This test is based upon the largest vertical distance between empirical distribution function $F_n(x)$ and theoretical distribution function $F(x,\theta)$. The statistic is

$$D_n = \sup_{|n|<\infty} \big|F_n(x) - F(x,\theta)\big| , \qquad \theta \in \Theta . \tag{3.1}$$

If the value of KS statistic is greater than critical point, we reject the null hypothesis (Gibbons and Chakraborti [9]).

Cramér–von Mises and Anderson–Darling statistics belong to the class of quadratic EDF statistics (Stephens [28]) defined as

$$n \int_{-\infty}^{\infty} \big(F_n(x) - F(x)\big)^2 \, w(x) \, dF(x) , \tag{3.2}$$

where $w(x)$ is a weighting function.

Cramér–von Mises Test:

For $w(x) = 1$, (3.2) gives n times the Cramér–von Mises (CVM) statistic. This statistic can be computed using the sum of squared differences between the empirical distribution function (EDF) and theoretical CDF (Anderson and Darling [1]) and is defined as

$$CVM = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_i,\theta) - \frac{2i-1}{2n} \right)^2 . \tag{3.3}$$

If the value of CVM test statistic is greater than the critical point, we reject the null hypothesis. According to Conover [5], CVM is more powerful than KS test because it uses more sample data.

Anderson–Darling test:

It is a modification of the CVM Test. It gives more weightage to the tails of the distribution (Farrel and Stewart [7]).

By taking $w(x) = [F(x)\,(1-F(x))]^{-1}$ in (3.2), Anderson–Darling (AD) test statistic (Anderson and Darling [1]) is obtained as

$$n \int_{-\infty}^{\infty} \frac{\big(F_n(x) - F(x)\big)^2}{\big[F(x)\,\big(1-F(x)\big)\big]}\, dF(x)\ .$$

It can also be written as

$$AD \;=\; -n - 2\sum_{i=1}^{n} \left\{ \frac{2i-1}{2n} \ln F(x_i,\theta) + \left(1 - \frac{2i-1}{2n}\right) \ln\big(1-F(x_i,\theta)\big) \right\} \tag{3.4}$$

(Lewis [18]).

If the value of AD test statistic is greater than critical point, we reject the null hypothesis.

The critical points (C.P.) of these tests have been calculated by generating random samples from the distribution under null hypothesis, calculating value of test statistics and arranging values of test statistic in increasing order. $(1-\alpha)^{\text{th}}$ largest order test statistic gives the critical point corresponding to α level of significance. These values have been calculated for sample sizes $n = 50, 100, 200,$ $500, 1000$ and 2000 at $\alpha = .20, .15, .10$ and $.05$ and are shown in Table 1.

Table 1: Critical points for Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling tests.

n	**Kolmogorov–Smirnov**				**Cramér–von Mises**				**Anderson–Darling**			
	Level of significance				Level of significance				Level of significance			
	.20	.15	.10	.05	.20	.15	.10	.05	.20	.15	.10	.05
50	.151	.161	.172	.192	.241	.281	.344	.455	1.427	1.619	1.900	2.422
100	.107	.114	.122	.136	.244	.286	.361	.475	1.388	1.603	1.909	2.412
200	.076	.081	.086	.096	.237	.282	.339	.453	1.39	1.59	1.92	2.49
500	.048	.051	.055	.061	.232	.275	.334	.444	1.405	1.609	1.932	2.500
800	.038	.040	.043	.048	.241	.286	.347	.469	1.410	1.617	1.904	2.399
1000	.034	.036	.039	.043	.245	.286	.347	.449	1.423	1.638	1.945	2.514
2000	.024	.026	.027	.030	.241	.287	.349	.476	1.395	1.588	1.914	2.438

4. APPLICATION

4.1. Simulations for estimation and applying GOFTs

Weighted Exponential and Gamma distributions are considered as competing distributions for WG. The Weighted Weibull (WW) distribution with three parameters α, β and λ (Shahbaz *et al.* [27]) has also been considered as one of the competing distributions for WG. The cdf and pdf of WW are

$$F(x;\alpha,\beta,\lambda) = \frac{(1+\alpha)\left[1-\mathrm{e}^{-\lambda x^{\beta}} - \frac{\left(1-\mathrm{e}^{-(1+\alpha)\lambda x^{\beta}}\right)}{1+\alpha}\right]}{\alpha}$$

and

$$f(x;\alpha,\beta,\lambda) = \frac{(1+\alpha)\,\lambda\,\beta\,x^{\beta-1}\,\mathrm{e}^{-\lambda x^{\beta}}\left(1-\mathrm{e}^{-\alpha\lambda x^{\beta}}\right)}{\alpha}\,.$$

A random sample of size 200 from Weighted Gamma (WG) distribution with parameters $\alpha = 5$, $\beta = 2.5$ and $\lambda = 2$ is generated. The empirical cumulative distribution function (ecdf) based on the data and the theoretical cdf of WG distribution are plotted in Figure 1. This figure depicts that ecdf and exact cdf of WG distribution are quite close to each other.

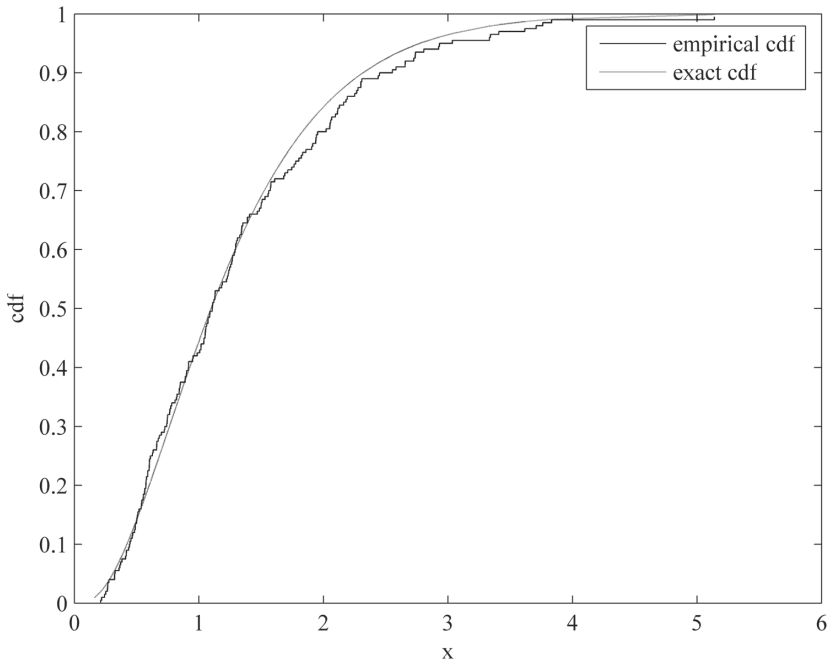


Figure 1: Plots for ecdf and exact cdf of WG distribution.

For the generated data set, maximum likelihood estimates (MLEs) of parameters of WG, WW, WE and Gamma distributions and the corresponding AIC and AICc values are given in Table 2. In quasi Newton algorithm in R, the Broyden–Fletcher–Goldfarb Shanno (BFGS) method has been used by applying optim routine. Hessian matrices have been checked for all the distributions and found to be negative definite as all the eigenvalues of each Hessian matrix come out to be negative. This implies that the estimates obtained are maximum likelihood estimates.

Table 2: Estimates of the parameters and AIC and AICc values for different distributions.

Distribution	MLE			AIC	AICc
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$		
WG	3.0230	2.1688	1.7399	432.9696	433.092
WW	53.0360	1.6025	0.5560	437.9989	438.121
WE	0.0006	1	1.5444	438.4732	438.534
Gamma	—	2.4871	1.9212	433.4271	433.488

From the above table, we can conclude that:

a) Since AIC and AICc values are the lowest for WG distribution, it can be considered to be the best fit.

b) Since AIC values of WG and Gamma distributions are close, hence a large sample size shall be required to distinguish between WG and Gamma distributions.

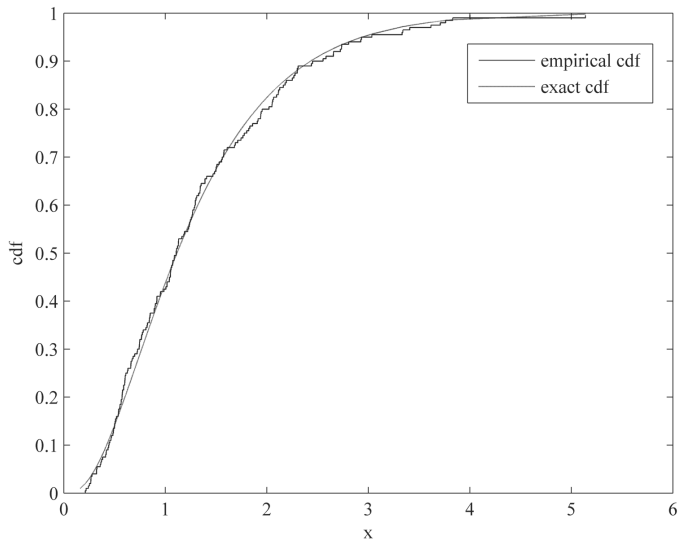


Figure 2: Plots of ecdf and estimated cdf of WG distribution.

Figure 2 displays the plots of empirical cdf and theoretical cdf using estimates of parameters of WG distribution for generated data set.

Weighted Weibull (WW), Weighted Exponential (WE) and Gamma distributions are taken up as the competing distributions for WG distribution. The estimates of parameters for all these distributions are found for generated data set.

To check whether the generated data set fits well to WG distribution (with assumed and estimated parameters), WW, WE and Gamma distributions, the simple and composite hypotheses have been tested in the sequel.

Testing of simple hypothesis:

The aim is to test the simple hypothesis

H_{01}: WG (α=5, β=2.5, λ=2) distribution fits well to the generated data set

versus

H_{11}: It does not fit well.

The values of KS, CVM and AD test statistics and critical points (extracted from Table 1) are given in Table 3.

Table 3: Values of test statistic and critical points for testing H_{01} versus H_{11}.

Test	Statistic values	C.P. at 0.05 level of significance
Kolmogorov–Smirnov	0.0447	.096
Cramér–von Mises	0.0779	.453
Anderson–Darling	0.5654	2.49

It is observed that for all the tests, the null hypothesis is not rejected at 0.05 level of significance implying that WG distribution fits well to the generated data set under all testing procedures.

Testing of composite hypotheses:

We consider testing of composite hypotheses

H_{02}: WG ($\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\lambda}$) distribution fits the generated data well

versus

H_{12}: It does not fit the data well,

where $\widehat{\alpha} = 3.0230$, $\widehat{\beta} = 2.1688$ and $\widehat{\lambda} = 1.7399$.

The following table gives calculated values of test statistics.

Table 4: Values of test statistic and critical points for testing H_{02} versus H_{12}.

Test	Statistic values	C.P. at 0.05 level of significance
Kolmogorov–Smirnov	0.0422	.096
Cramér–von Mises	0.0546	.453
Anderson–Darling	0.3480	2.49

It is observed that values of test statistics for KS, CVM and AD are less than C.Ps. at 0.05 level of significance for $n = 200$. This means that WG distribution with estimated parameters fits the data well.

Next, we consider testing of composite hypotheses:

i) H_{03}: Generated data set is fitted well by WW ($\widehat{\alpha} = 53.0360$, $\widehat{\beta} = 1.6025$, $\widehat{\lambda} = .5566$)
versus
H_{13}: compliment to H_{03}, that is, data is not fitted well.

ii) H_{04}: WE ($\widehat{\alpha} = .0006$, $\widehat{\lambda} = 1.5444$) distribution fits the generated data well
versus
H_{14}: WE ($\widehat{\alpha} = .0006$, $\widehat{\lambda} = 1.5444$) distribution does not fit the data well.

iii) H_{05}: Gamma ($\widehat{\beta}$=2.4871, $\widehat{\lambda}$=1.9212) distribution fits the generated data well
versus
H_{15}: Gamma distribution with estimated parameters does not fit the data well.

Tables 5–7 display the values of test statistics for KS, CVM and AD tests and corresponding critical points at 0.05 level of significance for testing the composite hypotheses.

Table 5: Values of test statistic and critical points for testing H_{03} versus H_{13}.

Test	Statistic values	C.P. at 0.05 level of significance
Kolmogorov–Smirnov	0.0576	.096
Cramér–von Mises	0.1489	.453
Anderson–Darling	0.8541	2.49

Table 6: Values of test statistic and critical points for testing H_{04} versus H_{14}.

Test	Statistic values	C.P. at 0.05 level of significance
Kolmogorov–Smirnov	0.0540	.096
Cramér–von Mises	0.0786	.453
Anderson–Darling	0.8263	2.49

Table 7: Values of test statistic and critical points for testing H_{05} versus H_{15}.

Test	Statistic values	C.P. at 0.05 level of significance
Kolmogorov–Smirnov	0.0464	.096
Cramér–von Mises	0.0822	.453
Anderson–Darling	0.5203	2.49

The values in Tables 3–7 help us to conclude that:

a) All the distributions fit well to the given data set at 0.05 level of significance because the values of test statistics are less than critical points.

b) WG distribution fits best to the data set because the values of test statistics are lowest in case of WG distribution.

In the next subsection, we find the powers of goodness of fit tests viz KS, CVM and AD for comparing WG distribution with WW, WE and Gamma distributions. The values of power for GOFTs help us in differentiating among the distributions under consideration and also in determining the optimal sample size for differentiation.

4.1.1. Powers of goodness of fit tests for WG

To differentiate among different distributions, we carry out the power study for testing of hypotheses about belonging of the sample to WG distribution, considering WW, WE and Gamma distributions as competing distributions.

For power analysis, we use the technique of Bootstrapping to generate the samples. We generate 10,000 copies of random sample under alternative hypotheses. The values of the test statistics have been calculated using estimates of parameters for different distributions. The power analysis has been carried out for sample sizes $n = 50, 100, 200, 500, 1000, 2000$ at .20, .15, .10, .05 levels of significance.

Using estimated parameters, Tables 8–10 give the power of KS, CVM and AD tests for testing about belonging of the samples to WG distribution against that sample is from WW, WE and Gamma distributions.

Power of Anderson–Darling test is more than those of Cramér–von Mises and Kolmogorov–Smirnov tests in all cases. Hence, AD is the most powerful and KS is the least powerful test.

From Tables 8–10, it is observed that at 0.10 level of significance to obtain low probability of type II error (less than or equal to 0.1):

a) A sample size greater than or equal to 2000 is required to differentiate WG distribution from WW distribution, since the power of AD test is .9630 implying that probability of type II error is .0370;

b) A sample of at least 2000 observations is required to distinguish WG distribution from WE and Gamma distributions.

Table 8: Power of tests for testing goodness of fit of WG versus WW with estimated parameters.

Level of significance	Sample size					
	$n=50$	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
Power of Anderson–Darling						
.20	.4993	.5834	.6253	.7351	.9005	.9899
.15	.4321	.4995	.5535	.7032	.8622	.9869
.10	**.2557**	**.3072**	**.4993**	**.5938**	**.8123**	**.9630**
.05	.1540	.2505	.3857	.4887	.7945	.8756
Power of Cramér–von Mises						
.20	.4286	.4993	.5547	.5790	.8750	.9666
.15	.4274	.4740	.5038	.5732	.8443	.9311
.10	.1571	.2946	.4586	.5606	.7801	.8959
.05	.1243	.2815	.2783	.4043	.7278	.8322
Power of Kolmogorov–Smirnov						
.20	.4078	.4551	.5013	.5485	.8539	.9521
.15	.3451	.4738	.5008	.5308	.8123	.9222
.10	.1526	.2574	.4165	.5243	.6959	.7898
.05	.1182	.2299	.2439	.2858	.5557	.6345

Table 9: Power of tests for testing goodness of fit of WG versus WE with estimated parameters.

Level of significance	Sample size					
	$n=50$	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
Power of Anderson–Darling						
.20	.5947	.6543	.7686	.8504	.9404	.9969
.15	.4928	.5689	.7038	.8153	.8935	.9851
.10	**.3853**	**.4537**	**.6583**	**.7328**	**.8589**	**.9708**
.05	.1549	.3839	.5841	.6685	.8040	.9146
Power of Cramér–von Mises						
.20	.4899	.5251	.6493	.8039	.8991	.9784
.15	.4518	.5103	.5552	.6751	.8599	.9485
.10	.2538	.3840	.4993	.5998	.8328	.9113
.05	.1959	.3014	.3547	.4853	.7943	.8993
Power of Kolmogorov–Smirnov						
.20	.4286	.4865	.5878	.6438	.8689	.9663
.15	.3945	.4793	.5584	.5991	.8402	.9365
.10	.2090	.2940	.4738	.5344	.7556	.8734
.05	.1547	.2591	.2973	.3905	.6938	.7488

Table 10: Power of tests for testing goodness of fit of WG versus Gamma with estimated parameters.

Level of significance	Sample size					
	$n=50$	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
Power of Anderson–Darling						
.20	.4037	.4270	.4408	.6728	.8875	.8993
.15	.3018	.3363	.3401	.5556	.8543	.8775
.10	**.2134**	**.2627**	**.2800**	**.3959**	**.8024**	**.8543**
.05	.1172	.1498	.2268	.3463	.7738	.8345
Power of Cramér–von Mises						
.20	.3535	.3889	.4229	.5389	.8198	.8856
.15	.3008	.3232	.4113	.4458	.7993	.8691
.10	.1418	.2138	.2542	.3304	.7583	.8434
.05	.1004	.1184	.2034	.3183	.7234	.8138
Power of Kolmogorov–Smirnov						
.20	.3038	.3359	.3947	.5126	.8057	.8535
.15	.2857	.3015	.3998	.4032	.7328	.8119
.10	.1218	.2028	.2238	.3123	.6888	.7735
.05	.0926	.1039	.1727	.2485	.5311	.6188

Further, it can also be concluded on the basis of Tables 8–10 that:

a) Power in case of testing goodness of fit of WG versus WE distribution is more than in other cases. Hence, the tests are detecting the gap between WG and WE distributions with high power and hence a small sample is sufficient to differentiate WG from WE.

b) The power of all GOFTs for all sample sizes and levels of significance is least when comparing WG and Gamma distributions. This means that the GOFTs are not detecting the difference between these two distributions as efficiently as in other cases. It implies that these distributions are quite close to each other. So, large sample sizes are required to differentiate these distributions.

4.2. Real data set illustration

We consider a data set consisting of survival times of guinea pigs injected with different amount of tubercle bacilli and studied by Bjerkedal [4]. The observations in the data set are: 12 15 22 24 24 32 32 33 34 38 38 43 44 48 52 53 54 54 55 56 57 58 58 59 60 60 60 60 61 62 63 65 65 67 68 70 70 72 73 75 76 76 81 83 84 85 87 91 95 96 98 99 109 110 121 127 129 131 143 146 146 175 175 211 233 258 258 263 297 341 341 376.

This data set was also considered by Gupta and Kundu [9] for fitting of Weighted Exponential (WE) distribution. The estimates of parameters, AIC and AICc values for above considered data set are reported in Table 11.

Table 11: Estimates of the parameters, AIC and AICc values for different distributions.

Distribution	MLE			AIC	AICc
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$		
WG	2.274	1.513	.0172	791.438	791.784
WW	139.4	1.39	.0014	799.271	799.624
WE	1.624	1	.0138	791.138	791.312
Gamma	—	2.081	.0208	792.495	792.669

From Table 11, it is seen that there is not a significant difference in AIC and AICc values for WG and WE models, hence both the models can be considered for fitting to this real data set. As WG provides generalization to many existing distributions viz WE, Gamma and Exponential distributions, hence it can be considered as a better choice for this data set.

4.2.1. Powers of goodness of fit tests for real data set

For power calculation, we generate random samples of sizes 100, 200, 500, 1000 and 2000 under alternative hypothesis. Test statistics are calculated using the estimates of parameters. By comparing these values with critical points, we either reject or do not reject the null hypothesis. Repeating this process 10,000 times and dividing the total number of rejections by 10,000, gives power.

Powers for goodness of fit tests for the following hypotheses have been reported in Tables 12, 13 and 14 respectively:

H_{06}: WG ($\widehat{\alpha}$=2.274, $\widehat{\beta}$=1.513, $\widehat{\lambda}$=.0172) fits the data set well

versus

i) $H_{16}^{'}$: WW ($\widehat{\alpha}$=139.4, $\widehat{\beta}$=1.39, $\widehat{\lambda}$=.0014) fits the data set well,

ii) $H_{16}^{''}$: WE ($\widehat{\alpha}$=1.624, $\widehat{\lambda}$=.0138) distribution fits the data well,

iii) $H_{16}^{'''}$: Gamma ($\widehat{\beta}$=2.081, $\widehat{\lambda}$=.0208) distribution fits the data well.

From the Tables 12–14, it can be concluded that:

a) Anderson–Darling (AD) is the most powerful and Kolmogorov–Smirnov (KS) is the least powerful test.

b) Power for testing GOF of WG versus WW is more than for testing in other cases.

c) Power is least when comparing WG distribution versus WE distribution. This means that for the considered data set, the GOFT's are not detecting the difference between these two models.

Table 12: Power of tests for testing goodness of fit of WG versus WW with estimated parameters.

Level of significance	Sample size				
	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
Power of Anderson–Darling					
.20	.4829	.6238	.7812	.8524	.9423
.15	.4458	.5744	.7123	.8047	.8850
.10	.3943	.5209	.6838	.7773	.8595
.05	.3451	.4753	.6552	.7239	.8391
Power of Cramér–von Mises					
.20	.4467	.5924	.6874	.7955	.8620
.15	.4139	.5251	.6193	.7338	.8354
.10	.3533	.4435	.5366	.6940	.7889
.05	.2669	.3981	.4921	.6569	.7495
Power of Kolmogorov–Smirnov					
.20	.3999	.5099	.6434	.7809	.8345
.15	.3458	.4875	.5701	.7051	.8003
.10	.3049	.4223	.4959	.6532	.7448
.05	.2225	.3801	.4153	.6034	.7115

Table 13: Power of tests for testing goodness of fit of WG versus WE with estimated parameters.

Level of significance	Sample size				
	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
Power of Anderson–Darling					
.20	.4145	.5125	.6720	.7518	.8498
.15	.3509	.4809	.6548	.7285	.8156
.10	.2877	.4053	.5740	.6893	.7632
.05	.2329	.3069	.4169	.6673	.7253
Power of Cramér–von Mises					
.20	.3595	.4430	.5407	.6863	.8002
.15	.3250	.4018	.4933	.6545	.7803
.10	.2589	.2944	.4356	.6187	.7234
.05	.2055	.2882	.3204	.5522	.6868
Power of Kolmogorov–Smirnov					
.20	.3486	.3997	.4407	.6562	.7259
.15	.3058	.3449	.4113	.5328	.6885
.10	.2137	.2507	.3876	.4935	.6138
.05	.1851	.2187	.3092	.4580	.5609

Table 14: Power of tests for testing goodness of fit of WG against Gamma with estimated parameters.

Level of significance	Sample size				
	$n=100$	$n=200$	$n=500$	$n=1000$	$n=2000$
Power of Anderson–Darling					
.20	.4277	.5459	.7032	.8089	.8927
.15	.3639	.4994	.6633	.7746	.8558
.10	.3073	.4227	.6118	.7268	.7982
.05	.2857	.3998	.5844	.6934	.7639
Power of Cramér–von Mises					
.20	.4008	.4935	.6632	.7604	.8239
.15	.3401	.4349	.6110	.7093	.7994
.10	.2831	.3970	.5256	.6859	.7530
.05	.2217	.3239	.5012	.6221	.7126
Power of Kolmogorov–Smirnov					
.20	.3603	.4158	.5728	.6953	.7649
.15	.3041	.3945	.4592	.5889	.7325
.10	.2859	.3567	.4182	.5234	.7049
.05	.2130	.3018	.3993	.5008	.6532

ACKNOWLEDGMENTS

The authors are grateful to the referees for their suggestions that helped in improving the paper. The first author is also thankful to University Grants Commission, Government of India, for providing financial support for this work.

REFERENCES

[1] Anderson, T.W. and Darling, D.A. (1954). A Test of Goodness of Fit, *Journal of American Statistical Association*, **29**, 765–769.

[2] Arnold, B.C. and Beaver, R.J. (2000). The Skew Cauchy distribution, *Statistics and Probability Letters*, **49**, 285–290.

[3] Arshad, M.; Rasool, M.T. and Ahmad, M.I. (2003). Anderson–Darling and Modified Anderson–Darling Tests for Generalized Pareto Distribution, *Pakistan Journal of Applied Sciences*, **3**(2), 85–88.

[4] Bjerkedal, T. (2003). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli, *Amer. J. Hyg.*, **72**, 130–148.

[5] Conover, W.J. (1999). *Practical Nonparametric Statistics*, John Wiley & Sons, New York.

[6] Dufour, J.M.; Farhat, A.; Gardiol, L. and Khalaf, L. (1998). Simulation-based Finite Sample Normality Tests in Linear Regressions, *Econometrics Journal*, **1**, 154–173.

[7] Farrel, P.J. and Stewart, K.R. (2006). Comprehensive Study Of Tests For Normality And Symmetry: Extending The Spiegelhalter Test, *Journal of Statistical Computation and Simulation*, **76**(9), 803–816.

[8] Fisher, R.A. (1934). The effects of methods of ascertainment upon the estimation of frequencies, *Annals of Eugenics*, **6**, 13–25.

[9] Gibbons J.D. and Chakraborti S. (2011). *Nonparametric Statistical Inference*, Chapman and Hall / CRC, Florida.

[10] Gupta, R.C. and Kirmani, S.N.U.A. (1990). The role of weighted distributions in stochastic modelling, *Communications in Statistics – Theory and Methods*, **19**(9), 3147–3162.

[11] Gupta, R.D. and Kundu, D. (2009). A new class of weighted exponential distributions, *Statistics*, **43**(6), 621–634.

[12] Jain, K.; Singh, H. and Bagai, I. (1989). Relations for Reliability measures of weighted distribution, *Communications in Statistics – Theory and Methods*, **18**(12), 4393–4412.

[13] Jain, K.; Singla, N. and Gupta, R.D. (2014). A weighted version of Gamma distribution, *Discussiones Mathematicae Probability and Statistics*, **34**, 89–111.

[14] Lemeshko, B.Yu.; Lemeshko, S.B. and Postovalov, S.N. (2009). Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses, *Journal of Applied and Industrial Mathematics*, **3**(4), 462–475.

[15] Lemeshko, B.Yu.; Lemeshko, S.B. and Postovalov, S.N. (2010a). Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. II. Verification of Complex Hypotheses, *Journal of Applied and Industrial Mathematics*, **4**(1), 79–93.

[16] Lemeshko, B.Yu.; Lemeshko, S.B.; Akushkina, K.A.; Nikulin, M.S. and Saaidia, N. (2010b). *Inverse Gaussian Model and Its Applications in Reliability and Survival Analysis*. In "Mathematical and Statistical Models and Methods in Reliability: Applications to Medicine, Finance and Quality Control" (V. Rykov, N. Balakrishanan and M. Nikulin, Eds.), Birkhäuser, Boston, 433–453.

[17] Lemeshko, B.Yu. and Lemeshko, S.B. (2011). Models of Statistic Distributions of Nonparametric Goodness-of-Fit Tests in Composite Hypotheses Testing for Double Exponential Law Cases, *Communications in Statistics – Theory and Methods*, **40**, 2879–2892.

[18] Lewis, P.A.W. (1961). Distribution of the Anderson–Darling Statistic, *The Annals of Mathematical Statistics*, **32**(4), 1118–1124.

[19] Mahfoud, M. and Patil, G.P. (1982). *On weighted distributions*. In "Statistics and Probability: Essays in Honor of C.R. RAO" (G. Kallianpur, P.R. Krishnaiah and J.K. Ghosh, Eds.), North-Holland, Amsterdam, 479–492.

[20] Nanda, A.K. and K. Jain (1999). Some weighted distribution results on univariate and bivariate cases, *Journal of Statistical Planning and Inference*, **77**(2), 169–180.

[21] Nikulin, M.S. (1973). On a chi-squared test for continuous distributions, *Theory of Probability and Its Applications*, **18**(3), 638–639.

[22] Nikulin, M.S. (1973). Chi-squared test for continuous distributions with shift and scale parameters, *Theory of Probability and Its Applications*, **18**(3), 559–568.

[23] Patil, G.P. (2002). *Weighted Distributions*. In "Encyclopaedia of Environmetrics" (Abdel H. El-Shaarawi and Walter W. Piegorsch, Eds.), John Wiley & Sons, Ltd, Chichester, 2369–2377.

[24] Rao, C.R. (1965). *On discrete distributions arising out of methods of ascertainment*. In "Classical and Contagious Discrete Distributions" (G.P. Patil, Eds.), Pergamon Press, Oxford and Statistical Publishing Society, Calcutta, 320–332.

[25] Rao, K.C. and Robson, D.S. (1974). A Chi-squared statistic for goodness-of-fit tests within the exponential family, *Communication in Statistics*, **3**, 1139–1153.

[26] Seier, E. (2002). Comparison of Tests for Univariate Normality, *Interstat Statistical Journal*, **1**, 1–17.

[27] Shahbaz, S.; Shahbaz, M.Q. and Butt, N.S. (2010). A class of Weighted Weibull Distributions, *Pakistan Journal of Statistical Operation Research*, **6**(1), 53–59.

[28] Stephens, M.A. (1986). *Tests based on EDF statistics*. In "Goodness-of-Fit Techniques" (R.B. D'Agostino and M.A. Stephens, Eds.), Marcel Dekker, New York, 97–193.

REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 49–73

HIERARCHICAL DYNAMIC BETA MODEL

Authors: Cibele Queiroz Da-Silva
– Departmento de Estatística, Universidade de Brasília, Brasília, Brazil
cibeleqs@unb.br

Helio S. Migon
– Departmento de Estatística, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
migon@im.ufrj.br

Received: August 2013 Revised: August 2014 Accepted: September 2014

Abstract:

- We develop a hierarchical dynamic Bayesian beta model for modelling a set of time series of rates or proportions. The proposed methodology enables to combine the information contained in different time series so that we can describe a common underlying system, which is though flexible enough to allow the incorporation of random deviations, related to the individual series, not only through time but also across series. That allows to fit the case in which the observed series may present some degree of level shift. Additionally, the proposed model is adaptive in the sense that it incorporates precision parameters that can be heterogeneous no only over time but also across the series. Our methodology was applied to both real and simulated data. The real data sets used in this article include three time series of Brazilian monthly unemployment rates, observed in the cities of Recife, São Paulo and Porto Alegre, in the period from March 2002 to March 2012. A new parametrization of the precision parameter makes possible the use of the same type of link function for both the mean and the precision parameters, which are then expressed in the $(0, 1)$ interval, providing a more meaningful interpretation in terms of the magnitude of the scale.

Key-Words:

- *dynamic models; beta distribution; hierarchical models; Bayesian analysis.*

AMS Subject Classification:

- 91B84, 62M10.

1. INTRODUCTION

The beta regression models, proposed by Ferrari and Cribari-Neto (2004), have attracted the attention of many researchers. Those models are useful in situations where the response is restricted to the standard unit interval. In this seminal work the authors developed generalized linear models (GLM) theory for dealing with the situation where only the parameter related to the mean of the beta distribution was allowed to vary.

In the context of GLM's Nelder and Lee (1991) and Smyth and Verbyla (1999) describe a class of joint generalized linear models which allow both the mean and the dispersion parameters in the GLM model to vary with the response.

Nelder and Lee (1991) argue that it is necessary to use two GLM's when both mean and dispersion are to be modeled, i.e., we would have the so called *mean process* and the *dispersion process*. Pregibon (1984) was the first to suggest this kind of specification. Other articles related to such perspective, in which the dispersion parameter of the beta model is allowed to vary, include Cuervo-Cepeda and Gamerman (2004), Smithson and Verkuilen (2006), Espinheira (2007), Simas *et al.* (2010) and Bayer (2011). These works emphasize the need of correctly modelling the dispersion parameter of the beta regression in order to achieve efficient estimation.

Based on the class of beta regressions introduced by Ferrari and Cribari-Neto (2004), Rocha and Cribari-Neto (2009) proposed a dynamic model for continuous random variates whose range is described by the standard unit interval (0,1). The proposed frequentist βARMA model includes both autoregressive and moving average dynamics, and also includes a set of regressors. Da-Silva *et al.* (2011) proposed a dynamic Bayesian beta model for modelling and forecasting single time series of rates or proportions. In such work only the mean parameter of the beta model was allowed to vary with time.

In the present work we build upon the dynamic Bayesian beta model introduced by Da-Silva *et al.* (2011) and upon the class of conditionally Gaussian dynamic models (see Cargnoni *et al.*, 1997; Gamerman and Migon, 1993) to propose a hierarchical dynamic Bayesian beta model in which both the mean and the dispersion parameters of the beta model can vary with time. Since the proposed model is hierarchical, the parameters in the model are related both through time and hierarchically across several series, which supposedly share a common underlying trend.

Even though it is possible to individually fit time series that share common features, gains are obtained when those series are analyzed jointly (Gamerman and Migon, 1993). Naturally, by disregarding existing common features shared

by a given set of time series (e.g. trends, seasonal behavior, etc) one could end up with poorer analyses and forecasts.

We also would like to stress the fact that Cargnoni *et al.* (1997) and Gamerman and Migon (1993) do not deal with the situation of fitting the *dispersion process*, a feature that we introduce in our present model formulation. Thus, in this paper we address the issues of formulating a hierarchical dynamic beta model that allows dealing with a set of related time series, each one following related beta models that may present time-evolving mean and precision parameters.

We motivate our study with the problem of forecasting monthly Brazilian unemployment rates in different cities. The Brazilian Institute of Geography and Statistics (IBGE) implemented the Monthly Unemployment Survey (PME) in 1980, but since 2002 a new survey methodology has been adopted.

The PME is a monthly survey about workforce and income. The most important metropolitan regions in Brazil are included in such survey: São Paulo, Rio de Janeiro, Belo Horizonte, Porto Alegre, Recife and Salvador. The data can be found at `http://www.ibge.gov.br/`.

In Figure 1 we present the PME data for the cities of Recife, São Paulo and Porto Alegre. As we can observe, the three series have similar underlying trends but distinct levels and, possibly, distinct dispersions, specially in the case of the city of Recife.

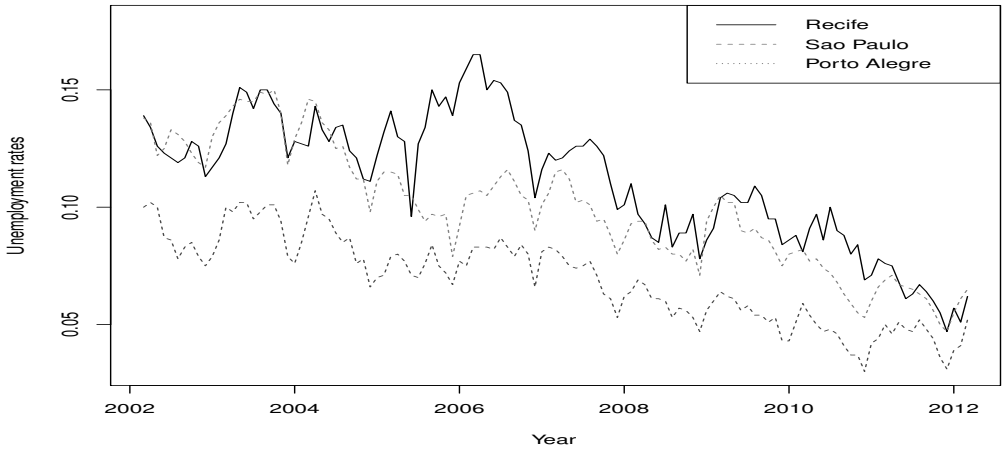


Figure 1: Observed unemployment rates in the cities of Recife, São Paulo and Porto Alegre — Brazil.

This article is organized as follows. In Section 2 we introduce the hierarchical dynamic beta model. In Section 3 we describe a fully Bayesian methodology to analyze data from a hierarchical dynamic beta process. In Sections 4 to 6 we apply the methods to simulated and real data.

2. THE HIERARCHICAL DYNAMIC BETA MODEL

In this section we present a methodology for modelling a set of I time series of rates or proportions, y_{it}, $i = 1, ..., I$, which share certain characteristics which allows us to treat them in the class of the hierarchical models.

Da-Silva *et al.* (2011) used the parametrization of the beta distribution given by Ferrari and Cribari-Neto (2004) to describe a dynamic beta model in which the precision parameter ζ was considered fixed. However, a more general model can be described by considering both the mean and the precision parameters varying with time. In such case, the *observation equation* of the dynamic model is given by

$$(2.1) \quad p\big(y_{it} \mid \mu_{it}, \zeta_{it}\big) = \frac{\Gamma(\zeta_{it})}{\Gamma(\zeta_{it}\mu_{it})\,\Gamma\big(\zeta_{it}(1-\mu_{it})\big)}\; y_{it}^{\zeta_{it}\mu_{it}-1}(1-y_{it})^{\zeta_{it}(1-\mu_{it})-1}\,,$$

and we have $E(y_{it} \mid \mu_{it}, \zeta_{it}) = \mu_{it}$ and $V(y_{it} \mid \mu_{it}, \zeta_{it}) = \mu_{it}(1-\mu_{it})/(1+\zeta_{it})$, with $0 \leq \mu_{it} \leq 1$ and $\zeta_{it} > 0$, $t = 1, ..., N$ and $i = 1, ..., I$.

Another parametrization for ζ, proposed by Bayer (2011), can be used in our context, since it allows us to use link functions for the transformed ζ which are easier to interpret than, say, a log link function, whose the upper limit is unbounded.

In equation (2.1), let $\phi_{it} = \frac{1}{1+\zeta_{it}}$ so that $\zeta_{it} = \frac{1-\phi_{it}}{\phi_{it}}$. Thus, $0 < \phi_{it} < 1$, and the *observation equation* of the model is now written as

Observation equation: Let

$$(2.2) \qquad p\big(y_{it} \mid \mu_{it}, \phi_{it}\big) = \frac{y_{it}^{\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right)-1}(1-y_{it})^{(1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)-1}}{B\left(\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right), (1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right)}\,,$$

with $i = 1, ..., I$, $t = 1, ..., N$ and $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, be the *observation equation* of the dynamic model. Let $y = (y_1, ..., y_N)$ with $y_t = (y_{1t}, ..., y_{It})'$, $t = 1, ..., N$.

The model structure is such that we have I time series in study, in which $(y_{it} \mid \mu_{it}, \phi_{it})$ is independent of $(y_{jt} \mid \mu_{jt}, \phi_{jt})$ for $i \neq j$. Equation (2.2) incorporates heterogeneity in the precision parameter that may occur both over time or across the series.

Other components which are essential in the description of our hierarchical dynamic beta model include

(**i**) the definition of real transformations applied to μ_{it} and ϕ_{it}, allowing the use of some simplifying Gaussian properties;

(**ii**) the description of **structural equations** represented in terms of linear models relating the transformed parameters and the latent states and

(**iii**) the representation of the **system equation** of the dynamic model in which the state parameters are related to surrogate observation equations described by the structural equations.

In order to describe the structural equation, two link function, $h_1(\cdot)$ and $h_2(\cdot)$, associated to, respectively, the *mean process* and the *dispersion process*, should be defined. These are real valued transformations and are useful in the model construction since some of the nice properties of the Gaussian dynamic linear models (DLM's) follow from that.

Take $\eta_{1it} = h_1(\mu_{it})$ and $\eta_{2it} = h_2(\phi_{it})$ with $\eta_{it} = (\eta_{1it}, \eta_{2it})'$ such that η_{it} is a real valued vector. Let $\eta_t = (\eta_{1t}, ..., \eta_{It})'$, i.e., η_t is a $2I \times 1$ vector of *structural parameters* for all the I series at time t with $\eta_{it} = (\eta_{1it}, \eta_{2it})'$, thus $\eta = (\eta_1, ..., \eta_N)$.

Now, y_{it} is parametrized by η_{it}, i.e., $(y_{it}\,|\,\eta_{it}) \sim \text{Beta}(y_{it}\,|\,\eta_{it})$.

Structural equations: Let

$$\eta_t = F_t\theta_t + v_t \ , \qquad v_t \sim N(\mathbf{0}, V) \ , \tag{2.3}$$

with $\eta_{it} = F_{it}\theta_t + v_{it}$ be the *structural equation* in our model formulation. The error term v_{it} in the structural equation is assumed to follow a Gaussian distribution with zero mean vector and covariance matrix V_i, i.e., $v_{it} \sim N(\mathbf{0}, V_i)$, with $t = 1, ..., N$ and $i \in \{1, ..., I\}$.

In equation (2.3) the term θ_t, representing the *state parameter* of the dynamic model at time t, is a real valued s-dimensional vector of latent states. Besides, $F_t = (F_{1t}, ..., F_{It})'$ is the $2I \times s$ design matrix for all the I series at time t, $v_t = (v_{1t}, ..., v_{It})'$ is the $2I \times 1$ vector of errors for the structural equations and $V = \text{block-diag}(V_1, ..., V_I)$ is a $(2I \times 2I)$ block diagonal matrix.

System Equation: Let

$$\theta_t = H_t\theta_{t-1} + w_t \ , \qquad w_t \sim (\mathbf{0}, W) \ , \tag{2.4}$$

with $t = 1, ..., N$, be the *system equation* of the dynamic model.

The error term w_t in the system equation is assumed to follow a Gaussian distribution with zero mean vector and covariance matrix W, i.e., $w_t \sim N(\mathbf{0}, W)$, with $t = 1, ..., N$. Additionally, we assume that the error terms w_t and v_{it} are all mutually independent.

The s-dimensional covariance matrix W (for the s-dimensional vector of latent states, θ_t), is assumed to be block-diagonal including k blocks, with $k \leq s$.

Those blocks are associated to the effects included in the latent states. Thus, $W = \text{block-diag}(W_1, ..., W_k)$. The matrix H_t is a specified $s \times s$ state evolution matrix.

The hierarchical dynamic beta model (HDBM) requires the specification of a $(2I \times 2I)$ covariance matrix V in the structural equations and another covariance matrix W for the state vector. That might become complicated for large matrix dimensions. In many applications it may be sufficient to model simpler dependences, in particular to allow individual random effects. That is why in our proposed model both V and W are block-diagonal matrices.

Notice that equations (2.3) and (2.4) represent a standard dynamic linear model for the state vector θ_t. Additionally, θ is conditionally independent of y given η. These combined features imply a substantial simplification in the posterior computations of the parameters η and θ, as described in Cargnoni *et al.* (1997).

3. MODELLING THE LATENT COMPONENTS OF THE HDBM

In this section we set up the hierarchical beta model for a hypothetical case in which y_{it} represents a given rate or proportion at region i and time t, $i = 1, ..., I$ and $t = 1, ..., N$. We take the logit transformation of both μ_{it} and ϕ_{it} and, to η_{1it} and η_{2it}, we fit dynamic models considering, respectively, a second-order polynomial trend seasonal effects and a second-order polynomial trend effects. The formulation of the structural equations is given below:

$$
\begin{aligned}
\eta_{1it} &= \log\left(\frac{\mu_{it}}{1-\mu_{it}}\right) = F_{i1t}\theta_t + v_{i1t} \ , \qquad v_{i1t} \sim N(0, V_{i1}) \ , \\
\eta_{2it} &= \log\left(\frac{\phi_{it}}{1-\phi_{it}}\right) = F_{i2t}\theta_t + v_{i2t} \ , \qquad v_{i2t} \sim N(0, V_{i2}) \ ,
\end{aligned} \tag{3.1}
$$

with $V_i = \text{diag}(V_{i1}, V_{i2})$.

In equation (3.1) the term $F_{i1t}\theta_t$, on the right-hand side of η_{i1t}, is the linear predictor of the logit transformed expected value of the beta model for time t and region i. We use a second-order polynomial trend seasonal effects model with offset term in order to describe η_{i1t}, that is

$$
\eta_{1it} = \beta_t + \lambda_{t0} + \gamma_{it} + v_{i1t} \ . \tag{3.2}
$$

The DLM representation of the model for η_{1it} is

Second-order polynomial effects for the level with respect to μ_{it}:

$$
\begin{aligned}
\beta_t &= \beta_{t-1} + \delta_{t-1} + w_{\beta_t} \ , \\
\delta_t &= \delta_{t-1} + w_{\delta_t} \ ,
\end{aligned} \tag{3.3}
$$

Free-form Seasonal effects:

$$\begin{aligned} \lambda_{tr} &= \lambda_{t-1,r+1} + w_{tr} \ , \qquad r = 0, ..., p-2 \ , \\ \lambda_{t,p-1} &= \lambda_{t-1,0} + w_{t,p-1} \ , \end{aligned} \tag{3.4}$$

First-order polynomial effects for the offset term:

$$\gamma_{it} = \gamma_{i,t-1} + w_{\gamma_{it}} \ , \tag{3.5}$$

where

- β_t represents an underlying level at time t, with respect to $h_1(\mu_{it})$, that is common to the I series;
- δ_t is the incremental growth;
- λ_{t0} represents a seasonal effect that is common to the I series. We denote the size of the seasonal cycle as p.
- γ_{it} is an offset parameter representing deviations of the observed rate in region i at time t with respect to the average β_t;
- v_{i1t} represents the region i series-specific stochastic deviation.

In equation (3.1) the term $F_{i2t}\theta_t$, on the right-hand side of η_{2it}, is the linear predictor of the logit transformed term related to the precision of the beta model for time t and region i. We use a second-order polynomial effects model with offset term in order to describe η_{2it}, that is

$$\eta_{2it} = \psi_t + \alpha_{it} + v_{i2t} \ . \tag{3.6}$$

The DLM representation of the model for η_{2it} is

Second-order polynomial effects for the level with respect to ϕ_{it}:

$$\begin{aligned} \psi_t &= \psi_{t-1} + \xi_{t-1} + w_{\psi_t} \ , \\ \xi_t &= \xi_{t-1} + w_{\xi_t} \ , \end{aligned} \tag{3.7}$$

First-order polynomial effects for the offset term:

$$\alpha_{it} = \alpha_{i,t-1} + w_{\alpha_{it}} \ , \tag{3.8}$$

where

- ψ_t represents an underlying level at time t, with respect to $h_2(\phi_{it})$, that is common to the I series;
- ξ_t is the incremental growth;
- α_{it} is an offset parameter representing deviations of the observed rate in region i at time t with respect to the average ψ_t;
- v_{i2t} represents the region i series-specific stochastic deviation.

Identifiability restrictions:

$$\lambda_{t,p-1} = -\sum_{r=0}^{p-2} \lambda_{tr} \ , \qquad \gamma_{It} = -\sum_{i=1}^{I-1} \gamma_{it} \ , \qquad \alpha_{It} = -\sum_{i=1}^{I-1} \alpha_{it} \ .$$

In order to exemplify the construction of the model, we consider $I = 3$ regions where the rates are measured over time. Thus, the vector $(\eta_{1it}, \eta_{2it})'$ is described by

$$\begin{pmatrix} \eta_{1it} \\ \eta_{2it} \end{pmatrix} = \begin{pmatrix} \beta_t + \lambda_{t0} + \gamma_{it} \\ \psi_t + \alpha_{it} \end{pmatrix} + \begin{pmatrix} v_{i1t} \\ v_{i2t} \end{pmatrix} , \qquad i = 1, 2, 3 \ .$$

That is,

$$\eta_{it} = F_{it}\theta_t + v_{it} \ , \qquad i = 1, 2, 3 \ ,$$

where $\gamma_{3t} = -(\gamma_{1t} + \gamma_{2t})$, $\alpha_{3t} = -(\alpha_{1t} + \alpha_{2t})$. For example, for seasonal cycles of size $p = 4$ (quarters), then $\lambda_{t3} = -(\lambda_{t0} + \lambda_{t1} + \lambda_{t2})$.

The state vector θ_t for generic-sized p cycles is represented by

$$\theta_t \ = \ \Big(\beta_t, \, \delta_t, \, \lambda_{t0}, \, \lambda_{t1}, \, ..., \, \lambda_{t,p-2}, \, \psi_t, \, \xi_t, \, \gamma_{1t}, \, \gamma_{2t}, \, \alpha_{1t}, \, \alpha_{2t}\Big) \, .$$

Consider the following design matrices:

$$J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad P = \begin{pmatrix} -\mathbf{1}'_{p-2} & -1 \\ \mathbf{I}_{p-2} & \mathbf{0} \end{pmatrix} .$$

Matrices J and P are essential in the description of our dynamic model. Suppose a DLM such that the observation equation is $y_t = \beta_t + \epsilon_t$ and the system equation is given by the pair of equations in expression (3.3). Such model is called a *linear growth model* and it includes a time-varying slope β_t. If we define $\theta_t = (\beta_t, \delta_t)'$ and $F = (1, 0)'$, then the observation equation can be represented by $y_t = F'\theta_t + \epsilon_t$, while the system equation, by $\theta_t = J\theta_{t-1} + (w_{\beta_t}, w_{\delta_t})'$.

Matrix J allows us to write a *linear growth model* such the *permutation matrix* P is *p-cyclic*, so that $P^{np} = I_p$ and $P^{h+np} = P^h$, for $h = 1, ..., p$, and any integer $n \geq 0$. For example, suppose, for simplicity, a DLM model with $y_t = F\theta_t + \epsilon_t$ describing the observation equation and $\theta_t = \theta_{t-1} + w_t$, the system equation. Additionally, suppose a purely seasonal series and quarterly data y_t, $t = 1, 2, ...$, so that when y_{t-1} refers to the first quarter of the year, y_t refers to the second one.

Due to the restriction $\sum_{i=1}^{4} \alpha_i = 0$, the series might be described by seasonal deviations from the zero. Thus assume that $y_{t-1} = \alpha_1 + \epsilon_{t-1}$, $y_t = \alpha_2 + \epsilon_t$, and so on, so that to $(y_{t-1}, \, y_t, \, y_{t+1}, \, y_{t+2}, \, y_{t+3}, \, y_{t+4}, \, y_{t+5}, \, y_{t+6})$ are associated the respective seasonal deviations from zero, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$. Consider now that $\theta_{t-1} = (\alpha_1, \alpha_4, \alpha_3, \alpha_2)$ and that $F' = (1, 0, 0, 0)$. Then, the successive application of matrix P makes possible to formulate the desired quarterly seasonal pattern.

Considering $I = 3$ sub-populations or regions, the design matrices associated to the hierarchical beta dynamic model given by expressions (2.3), (2.4) and (3.1) to (3.8) are given by

$$H = \text{block-diag}\big(J, P, J, \mathbf{I}_2, \mathbf{I}_2\big)\,, \tag{3.9}$$

$$F_1 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}_{1\times(p-2)} & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$F_2 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}_{1\times(p-2)} & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$F_3 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}_{1\times(p-2)} & 1 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}.$$

The incorporation of seasonal components in the model can also be done by using the Fourier Representation Theorem (see Pole, West and Harrison, 1994, pp. 49) in which any cyclical function of period p defined by a set of p effects $\psi_1, ..., \psi_p$, can be expressed as a linear combination of sine and cosine terms. Let $\omega = 2\pi/p$, then there exist $(p-1)$ real numbers $a_1, ..., a_h$; $b_1, ..., b_{h-1}$ such that, for $j = 1, ..., p$,

$$\psi_j = a_h \cos(\pi j) + \sum_{r=1}^{h-1} \big[a_r \cos(\omega r j) + b_r \sin(\omega r j)\big]\,, \tag{3.10}$$

where $p = 2h$ if p is even, and $p = 2h - 1$ with $a_h = 0$ if p is odd. The Fourier coefficients a_r and b_r are known quantities and we usually set $a_h = 0$. Thus equation (3.10) can be written as $\psi_j = \sum_{r=1}^{h} S_r(j)$, where

$$S_r(j) = a_r \cos(\omega r j) + b_r \sin(\omega r j) = A_r \cos(\omega r j + \gamma_r)\,,$$
$$A_r = (a_r^2 + b_r^2)^{1/2} \qquad \text{and} \qquad \gamma_r = \arctan(-b_r/a_r)\,.$$

The terms $S_r(j)$ is called the r-th harmonic. A_r, ωr and γ_r describe, respectively, the *amplitude*, the *frequency* and the *phase* of $S_r(j)$.

For seasonal cycles of even size p (say quarters), we replace matrix P by G where $G = \text{block-diag}\big(J_2(1,\omega), J_2(1,2\omega), ..., J_2(1,(p/2-1)), -1\big)$, with $G^p = G$ and

$$J_2(1,\omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}.$$

For a second-order polynomial trend two harmonic model, the design matrices are given by

$$G = \text{block-diag}\big(J_2(1,\omega), J_2(1,2\omega)\big)\,,$$

$$H = \text{block-diag}\big(J, G, J, \mathbf{I}_2, \mathbf{I}_2\big)\,, \tag{3.11}$$

$$F_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$F_2 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$F_3 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}.$$

3.1. Estimated proportions and forecasting

The estimated proportions are calculated using the following procedure:

(**1**) The inverse transformations $\mu_{it} = \frac{\exp(\eta_{1it})}{1+\exp(\eta_{1it})}$ and $\phi_{it} = \frac{\exp(\eta_{2it})}{1+\exp(\eta_{2it})}$ are evaluated at the estimated values (posterior means) of η_{1it} and η_{2it}, for $i = 1, ..., I$ and $t = 1, ..., N$.

(**2**) For $i = 1, ..., I$ and $t = 1, ..., N$ we simulate n (say, $n = 1{,}000$) samples from a beta distribution $\text{Beta}\Big(\mu_{it}\Big(\frac{1-\phi_{it}}{\phi_{it}}\Big),\, (1-\mu_{it})\,\Big(\frac{1-\phi_{it}}{\phi_{it}}\Big)\Big)$ and then we take the average value of those draws.

(**3**) For the confidence bands we repeat steps (1) and (2) for calculating the 2.5% and 97.5% percentiles of the posterior distribution of η_{it}.

The k-step-ahead forecasts for the states are obtained by the repeated application of the system equation (see expression (2.3)), that is,

$$\theta_{t+k} = HH_{t+k}(k)\,\theta_t + \sum_{r=1}^{k} HH_{t+k}(k-r)\,w_{t+r} \ ,$$

where $HH_{t+k}(r) = H_{t+k}H_{t+k-1} \times \cdots \times H_{t+k-r+1}$ for all t and integer $r \leq k$, with $HH_{t+k}(0) = I$. Thus, by linearity and independence and also taking into account the Bayesian linear estimation method,

$$\theta_{t+k} \sim \big(a_t(k), R_t(k)\big) \ ,$$

with $a_t(k) = H_{t+k}\, a_t(k-1)$ and $R_t(k) = H_{t+k}\, R_{t-k}\, H'_{t+k} + W_{t+k}$, and $a_t(0) = m_t$ and $R_t(0) = C_t$. Therefore the "future" θ_t values are obtained by successively sampling from the system equation followed by the evaluation of the structural equation (see expression (2.4)). The forecast rates are then obtained by running steps (1) to (3) given above.

4. BAYESIAN ANALYSIS

In the prior specification for θ_0, V and W we assume that θ_0, $V_1, ..., V_I$ and $W_1, ..., W_k$ are mutually independent, with $\theta_0 \sim N(m_0, C_0)$, V_i, $i = 1, ..., I$ have a common inverse Wishart prior and W be block-diagonal with an inverse Wishart prior for each block.

It is more convenient to work with the precision matrices instead of with the covariances matrices. Let $\Phi_{0i} = V_i^{-1}$, $i = 1, ..., I$, $\Phi_l = W_l^{-1}$, $l = 1, ..., k$, $\Phi_0 = \text{block-diag}(\Phi_{01}, ..., \Phi_{0I})$ and $\Phi = \text{block-diag}(\Phi_1, ..., \Phi_k)$. Suppose that Φ_{0i}, $i = 1, ..., I$, follow independent Wishart distributions such that $\Phi_{0i} \sim W(\nu_{0i}, S_{0i})$, where S_{0i} is a symmetric positive definite matrix of dimensions $p_i \times p_i$. Similarly, $\Phi_l \sim W(\varsigma_l, Z_l)$, follows independent prior distributions for $l = 1, ..., k$, where Z_l is a symmetric positive definite matrix of dimensions $q_l \times q_l$.

The joint posterior distribution is given by

$$p\big(\eta, \theta, \Phi_0, \Phi \mid y\big) \propto \left[\prod_{t=1}^{N}\left(\prod_{i=1}^{I} \text{Beta}(y_{it} \,|\, \eta_{it})\, N\big(\eta_{it}; F_{it}\theta_t, \Phi_{0i}^{-1}\big)\right) N\big(\theta_t; H_t\theta_{t-1}, \Phi^{-1}\big)\right]$$

$$\times\; N(\theta_0; m_0, C_0) \prod_{i=1}^{I} W(\Phi_{0i}; \nu_{0i}, S_{0i}) \prod_{l=1}^{k} W(\Phi_l; \varsigma_l, Z_l)\ . \tag{4.1}$$

The Markov chain Monte Carlo (MCMC) procedure used for the inferential processes involves sampling from the full conditional posteriors $p(\eta \mid \theta, \Phi_0, \Phi, y)$, $p(\theta \mid \eta, \Phi_0, \Phi, y)$ and $p(\Phi_0, \Phi \mid \eta, \theta, y)$.

4.1. Sampling from $p\big(\theta \mid \eta, \Phi_0, \Phi, y\big)$

As mentioned before, the equations (2.3) and (2.4) represent a standard dynamic linear model for the state vector θ_t. In such setting, the fact that θ is conditionally independent of y given η implies that $p(\theta \mid \eta, \Phi_0, \Phi, y) = p(\theta \mid \eta, \Phi_0, \Phi)$. Then, in a regular DLM, η has the same rule as y, so that in the sequential updating formulations of the DLM, y will be replaced by η.

The representation of the full conditional posterior distribution of $p(\theta \mid \eta, \Phi_0, \Phi)$, considering the conditional independence structure of the DLM as well

the Bayes theorem is given by

$$\begin{aligned} p\big(\theta \mid \eta, \Phi_0, \Phi\big) &= p\big(\theta_N \mid \eta, \Phi_0, \Phi\big) \prod_{t=0}^{N-1} p\big(\theta_t \mid \theta_{t+1}, ..., \theta_N, \eta, \Phi_0, \Phi\big) \\ &= p\big(\theta_N \mid \eta, \Phi_0, \Phi\big) \prod_{t=0}^{N-1} p\big(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi\big) \\ &\propto p\big(\theta_N \mid \eta, \Phi_0, \Phi\big) \prod_{t=0}^{N-1} p\big(\theta_{t+1} \mid \theta_t, \eta, \Phi_0, \Phi\big)\, p\big(\theta_t \mid \eta, \Phi_0, \Phi\big) \ . \end{aligned} \tag{4.2}$$

Thus, all the state vectors can be sampled from $p(\theta \mid \eta, \Phi_0, \Phi)$ using the FFBS (Forward-filtering, backward-sampling) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). Conditionally on the "observed values" of η, the algorithm below allows us to draw a sample $\theta_N, \theta_{N-1}, ..., \theta_0$ from $p(\theta \mid \eta, \Phi_0, \Phi)$ as follows:

(**1**) **Filtering**

Using the Kalman filter (de Jong, 1991), compute the moments m_t and C_t of the joint posterior $p(\theta_t \mid \eta, \Phi_0, \Phi)$, $t = 1, ..., N$, by applying the standard DLM sequential updating formulae with y replaced by η. For more details see West and Harrison (1997).

- $m_t = a_t + A_t e_t, \quad C_t = R_t - A_t Q_t A_t'$;
- $A_t = R_t F_t Q_t^{-1}, \quad e_t = \eta_t - f_t$;
- $a_t = H_t m_{t-1}, \quad R_t = H_t C_{t-1} H_t' + \Phi^{-1}$;
- $f_t = F_t a_t, \quad Q_t = F_t' R_t F_t + \Phi_0^{-1}$.

(**2**) **Smoothing**

At time $t = N$, sample the vector state θ_N from $p(\theta_N \mid \eta, \Phi_0, \Phi)$, i.e., sample θ_N from $(\theta_N \mid \eta, \Phi_0, \Phi) \sim N(m_N, C_N)$. For times $t = N-1, ..., 0$, sample θ_t from $p(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi)$ conditionally on the just sampled value θ_{t+1}. That is performed by sampling θ_t from $(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi) \sim N(u_t, U_t)$, where

- $u_t = m_t + B_t(\theta_{t+1} - a_{t+1})$;
- $U_t = C_t - B_t R_{t+1} B_t'$;
- $B_t = C_t H_t R_{t+1}^{-1}$.

4.2. Sampling from $p(\eta \mid \theta, \Phi_0, \Phi, y)$

Given θ, Φ_0 and Φ, the η_{it}'s are mutually independent. That implies that a sample from the conditional posterior of $(\eta \mid \theta, \Phi_0, \Phi, y)$ is obtained through $I \times N$ independent samples from the respective distributions given by

$$p\big(\eta_{it} \mid \theta_t, \Phi_{0i}, \Phi, y_{it}\big) \ \propto \ p\big(y_{it} \mid \eta_{it}\big)\, p\big(\eta_{it} \mid \theta_t, \Phi_{0i}\big) \ . \tag{4.3}$$

The second term on the right-hand side of the full conditional (4.3) is the normal prior $\eta_{it} \sim N(F_{it}\theta_t, \Phi_{0i})$, while the first term is given by the beta model described by expression (2.2), such that $\eta_{1it} = h_1(\mu_{it})$ and $\eta_{2it} = h_2(\phi_{it})$.

Since the distribution $p(\eta_{it} \mid \theta_t, \Phi_{i0}, y_{it})$ does not have a closed form, it is necessary to use the Metropolis–Hastings algorithm (M-H) (Metropolis *et al.*, 1953; Hastings, 1970) in order to draw samples from such distribution. Let m represent the m-th MCMC draw. We use the following M-H random-walk with symmetric normal proposal for η_{it}:

(**a**) Draw $\eta^*_{1it} \sim q_1(\eta^{m-1}_{1it}, \eta^*_{1it}) \stackrel{d}{=} N(\eta^{m-1}_{1it}, \Phi^{-1}_{1i0})$ and $\eta^*_{2it} \sim q_2(\eta^{m-1}_{2it}, \eta^*_{2it}) \stackrel{d}{=} N(\eta^{m-1}_{2it}, \Phi^{-1}_{2i0})$.

(**b**) Calculate the acceptance probability $\alpha(\eta^{m-1}_{it}, \eta^*_{it}) = \min\{1, R_{\eta_{it}}\}$, where

$$R_{\eta_{it}} \;=\; \frac{\pi\big(\eta^*_{it}\,|\cdot\big)}{\pi\big(\eta^{m-1}_{it}\,|\cdot\big)}\,\frac{q\big(\eta^*_{it}, \eta^{m-1}_{it}\big)}{q\big(\eta^{m-1}_{it}, \eta^*_{it}\big)} \;=\; \frac{\pi\big(\eta^*_{it}\,|\cdot\big)}{\pi\big(\eta^{m-1}_{it}\,|\cdot\big)}\;,$$

with $\pi(\eta^*_{it}|\cdot) = p(y_{it}|\eta^*_{it})\,p(\eta^*_{it}|\theta_t, \Phi_{0i})$, $\pi(\eta^{m-1}_{it}|\cdot) = p(y_{it}|\eta^{m-1}_{it})\,p(\eta^{m-1}_{it}|\theta_t, \Phi_{0i})$, and $q(\eta^*_{it}, \eta^{m-1}_{it}) = q_1(\eta^{m-1}_{1it}, \eta^*_{1it})\,q_2(\eta^{m-1}_{2it}, \eta^*_{2it})$.

(**c**) Set

$$\eta^m_{it} \;=\; \begin{cases} \eta^*_{it} & \text{with probability } \alpha\big(\eta^{m-1}_{it}, \eta^*_{it}\big)\,, \\ \eta^{m-1}_{it} & \text{otherwise}\,. \end{cases}$$

4.3. Sampling from $p(\Phi_0, \Phi \mid \eta, \theta, y)$

Considering that $\Phi_0 = \text{block-diag}(\Phi_{01},...,\Phi_{0I})$ and $\Phi = \text{block-diag}(\Phi_1,...,\Phi_k)$ where $\Phi_{0i} = V_i^{-1}$, $i = 1, ..., I$ and $\Phi_l = W_l^{-1}$, $l = 1, ..., k$, with $\Phi_{0i} \sim W(\nu_{0i}, S_{0i})$ and $\Phi_l \sim W(\varsigma_l, Z_l)$, $l = 1, ..., k$, the full conditional distribution of Φ_l is given by

$$\begin{aligned} p\big(\Phi_l \mid \eta, \theta, \Phi_0, y\big) \;&\propto\; \left[\prod_{t=1}^{N}\prod_{m=1}^{k} |\Phi_m|^{1/2}\,\exp\left\{-\frac{1}{2}(\theta_t - H_t\theta_{t-1})^T\,\Phi(\theta_t - H_t\theta_{t-1})\right\}\right] \\ &\quad\times\; |\Phi_l|^{\varsigma_l-(p_l+1)/2}\,\exp\big\{-\operatorname{tr}(Z_l\Phi_l)\big\} \\ &\propto\; |\Phi_l|^{N/2+\varsigma_l-(p_l+1)/2}\,\exp\left\{-\operatorname{tr}\left(\frac{1}{2}\sum_{t=1}^{N} ZZ_{ll,t}\Phi_l\right) - \operatorname{tr}(Z_l\Phi_l)\right\} \\ &\propto\; |\Phi_l|^{N/2+\varsigma_l-(p_l+1)/2}\,\exp\left\{-\operatorname{tr}\left(\left(\frac{1}{2}\,ZZ_{l\cdot} + Z_l\right)\Phi_l\right)\right\}, \end{aligned} \tag{4.4}$$

with $ZZ_t = (\theta_t - H_t\theta_{t-1})\,(\theta_t - H_t\theta_{t-1})^T$ and $ZZ_{l\cdot} = \sum_{t=1}^{N} ZZ_{ll,t}$. Thus,

$$\big(\Phi_l \mid \eta, \theta, \Phi_0, y\big) \;\sim\; \text{Wishart}\left(\frac{N}{2} + \varsigma_l,\; \frac{1}{2}\,ZZ_{l\cdot} + Z_l\right), \qquad l = 1, ..., k\;.$$

The full conditional distribution of Φ_{i0} is given by

$$
\begin{aligned}
p\big(\Phi_{i0} \mid \eta, \theta, \Phi_0, y\big) \;&\propto\; \left[\prod_{t=1}^{N} N\big(\eta_{it}; F_{it}\theta_t, \Phi_{0i}^{-1}\big)\right] W(\Phi_{0i}; \nu_{0i}, S_{0i}) \\
&\propto\; \left[\prod_{t=1}^{N} |\Phi_{0i}|^{1/2} \exp\left\{-\frac{1}{2}\left(\eta_{it} - F_{it}\theta_t\right)^T \Phi_{0i}(\eta_{it} - F_{it}\theta_t)\right\}\right] \\
&\quad\times\; |\Phi_{0i}|^{\nu_{0i}-(p_{0i}+1)/2} \exp\big\{-\operatorname{tr}(S_{0i}\Phi_{0i})\big\} \\
&\propto\; |\Phi_{0i}|^{N/2+\nu_{0i}-(p_{0i}+1)/2} \exp\left\{-\operatorname{tr}\left(\left(\frac{1}{2}SS_{\eta_i} + S_{0i}\right)\Phi_{0i}\right)\right\},
\end{aligned}
\tag{4.5}
$$

with $SS_{\eta_i} = (\eta_{it} - F_{it}\theta_t)\,(\eta_{it} - F_{it}\theta_t)^T$. Thus,

$$
\big(\Phi_{0i} \mid \eta, \theta, \Phi, y\big) \;\sim\; \text{Wishart}\left(\frac{N}{2} + \nu_{0i},\; \frac{1}{2}SS_{\eta_i} + S_{0i}\right), \qquad i = 1, ..., I\ .
$$

4.4. The case of static dispersion parameters

It is also possible to describe a beta hierarchical model such that the dispersion process does not vary with time, i.e., the precision parameters are static. In such case, the vector η_t on the left-hand side of the structural equation (2.3) will only include the term related to the mean process and $\eta_{it} = \eta_{1it} = h_1(\mu_{it})$. However, we can still associate a link function to the precision parameters, and we will denote it by $\eta_{2i} = h_2(\phi_i)$, $i = 1, ..., I$.

The observation equation for such case is then $(y_{it} \mid \eta_{it}, \eta_{2i}) \sim \text{Beta}(y_{it} \mid \eta_{it}, \eta_{2i})$, $i = 1, ..., I$. The Bayesian analysis for such situation can be adapted from the one we just described in the previous sections.

The MCMC developments for η_{it} are largely the same described in Section 4.2, but now they will be conditioned upon the current values of η_{2i}. Additionally, for a given prior distribution for η_{2i}, the corresponding full conditional distribution is

$$
p\big(\eta_{2i} \mid \eta, \theta, \Phi_0, \Phi, y\big) \;\propto\; \left[\prod_{t=1}^{N} \text{Beta}(y_{it} \mid \eta_{it}, \eta_{2i})\right] p(\eta_{2i})\ .
\tag{4.6}
$$

In this work the prior $p(\eta_{2i})$ was set to be a Gaussian distribution, with parameters chosen as the average mean and average variance of the initial estimated values of η_{2i}, $i = 1, ..., I$, that were obtained from separate MCMC runs for each of the individual time series.

5. A SIMULATION STUDY

We applied the model described in Section 3 to simulated data in which we considered $N = 72$ time points (say, six years), $I = 3$ sub-populations and cycles of size $p = 4$.

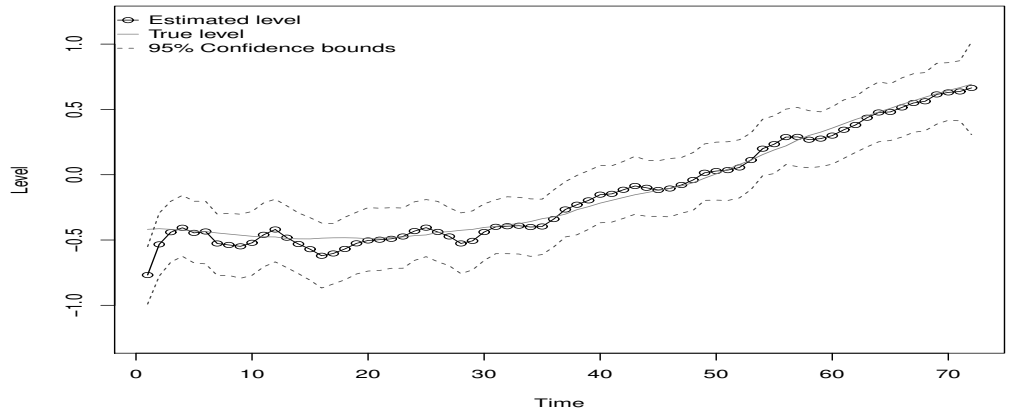
In order to obtain initial values for the MCMC procedure, we estimated the parameters involved by running separate DLM models (described by equations (2.3) and (2.4)) for each of the sub-populations. All the routines were written using the `R` language (`http://www.r-project.org/`). We also made extensive use of the excellent `dlm R` library by Petris (2010).

In such DLM setting the η_{it}'s have the same rule as the observed data. Thus, in order to run those initial models we estimated η_{1it} by $\log\left(\frac{y_{it}}{1-y_{it}}\right)$ and η_{2it} by $\log\left(\frac{\tilde{\sigma}^2_{it}}{1-\tilde{\sigma}^2_{it}}\right)$ with $\tilde{\sigma}^2_{it} = \text{var}(y_i)/(y_{it}(1-y_{it}))$ (see properties of expression (2.1)).

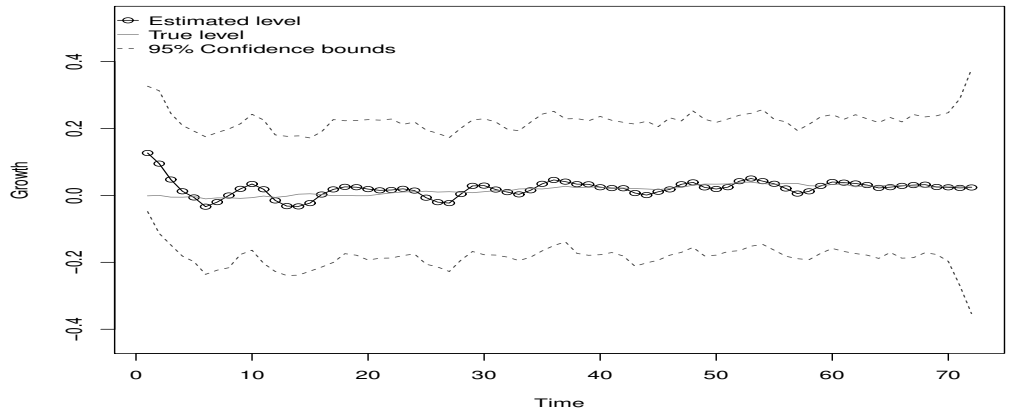
For the simulated data we considered a hierarchical dynamic beta model in which a second-order polynomial trend seasonal effects were fitted to the parameters related to the mean, μ_{it}, and a second-order polynomial effects was fitted to the parameters related to the precision, ϕ_{it}. We run chains of size 50,000 with burn-in period of 20,000. The autocorrelations could be significantly controlled by using gaps of size 30.

Figure 2 shows the true values (in red) used in the simulations, the estimated values of the parameters involved in expressions (3.2), and the respective confidence bands for the main effects of level, growth and seasonality. Figure 3 shows the estimated proportions for each of the sub-populations and their corresponding confidence bands. As we can observe all the effects and probabilities are well estimated.

(a)



(b)



(c)

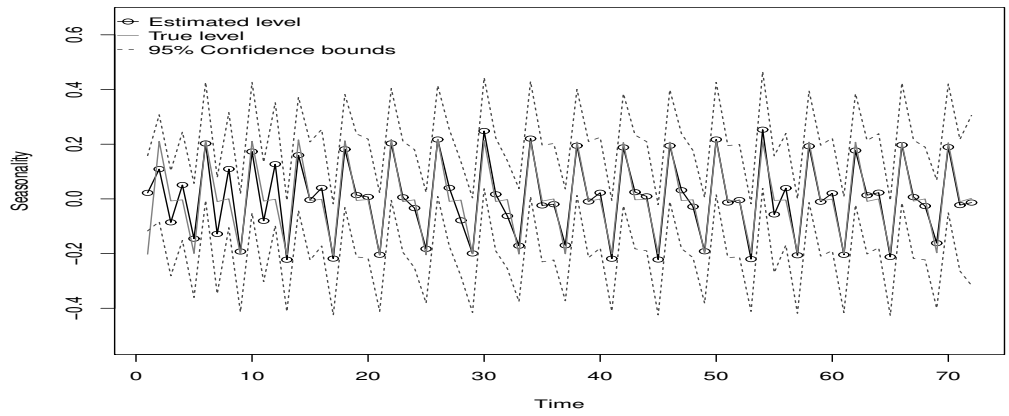
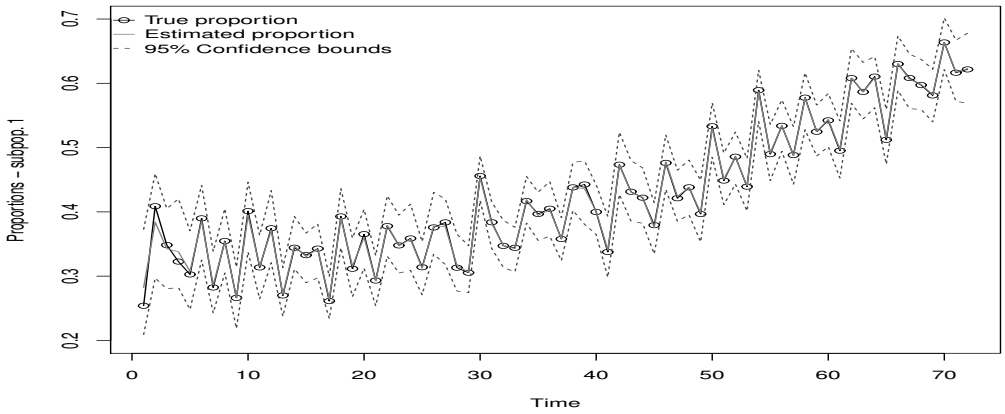
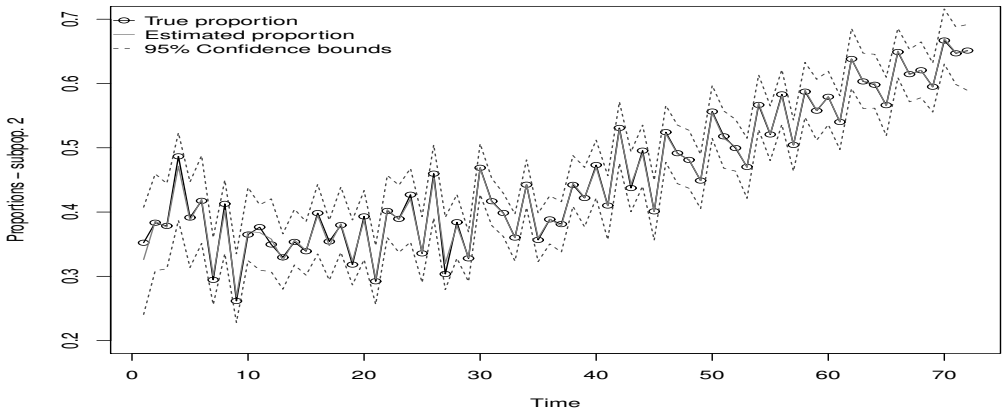


Figure 2: Simulated data — estimated values and 95% credibility bounds for the components of η_{1it}: (a) Level (β_t), (b) Growth (δ_t), (c) Seasonality (λ_t).

(a)



(b)



(c)

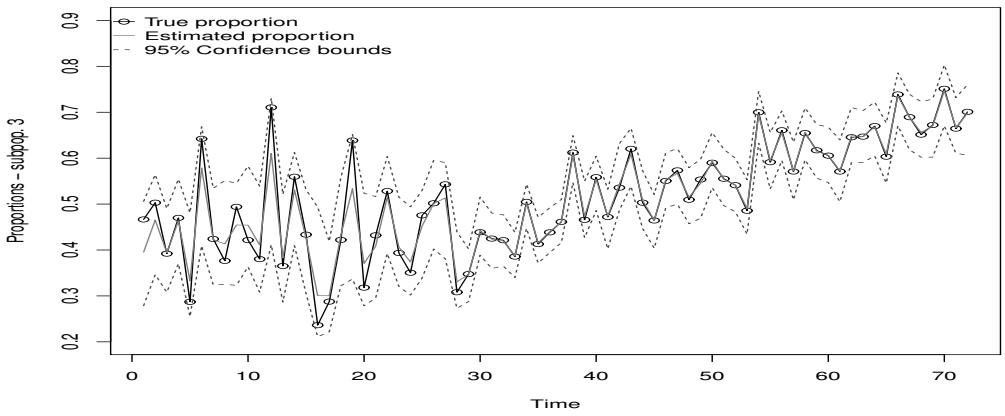


Figure 3: Simulated data — estimated proportions and 95% credibility bounds for the three sub-populations.

6. APPLYING THE HIERARCHICAL DYNAMIC BETA MODEL TO BRAZILIAN UNEMPLOYMENT RATES

In this section we apply our methods to fit the three time series of Brazilian monthly unemployment rates that were described in the Introduction (see Section 1). The three sub-populations in our the analysis are Recife, São Paulo and Porto Alegre, i.e., $I = 3$. We analyze monthly unemployment rates (MUR) based on PME data in the period from March 2002 to July 2011 ($N = 113$ observations). As a procedure for checking the performance of the model, forecast rates are also provided. We used MUR data for the months of August 2011 to March 2012.

The Brazilian Institute of Geography and Statistics acknowledges the necessity of incorporating seasonal components in any analyses based on MUR data (`http://www.ibge.gov.br`). In fact the MURs are affected by yearly cycles caused by factors such as climatic changes, Christmas festivities and school vacations.

To the *mean process*, we considered a hierarchical dynamic beta model in which a second-order polynomial trend seasonal effects model (with cycles of $p = 12$ months) was fitted to the parameters related to the mean, μ_{it} (see expressions (3.1) to (3.5)). The model structure was similar to the one used in Da-Silva *et al.* (2011), and we opted for using a free-form seasonal effects model (see equation (3.9)) instead of using harmonic analysis via Fourier representation (see equation (3.10)), just to be consistent with Da-Silva *et al.* (2011).

To the *dispersion process* we used two models: **Model 1** describes a static hierarchical model with respect to the precision parameter (see Section 4.4). **Model 2** adds dynamics to the precision parameter of the beta model. For that purpose we use a second-order polynomial trend effects model (see expressions (3.6) to (3.8)).

Considering Model 2 and the parameters of both mean and dispersion processes, the design matrices H and $F = (F_1', F_2', F_3')'$ are the ones defined by expression (3.9). For Model 1 those matrices are:

$$H = \text{block-diag}(J, P, \mathbf{I}_2) \ ,$$

$$F_1 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 1 & 0 \end{pmatrix} \ , \qquad F_2 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 0 & 1 \end{pmatrix} \quad \text{and}$$

$$F_3 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & -1 & -1 \end{pmatrix} \ .$$

We used the mean absolute deviation (MAD), the mean square error (MSE) and the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) to compare the forecasting accuracies of Model 1 and Model 2. The MAD and MSE are defined, respectively, by the following formulae: $MAD = \frac{1}{n}\sum_{t=1}^{n} |e_t|$ and $MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$, where $e_t = Y_t - E(Y_t \,|\, y_{1:t-1})$ (see Section 4.1 for details).

According to Spiegelhalter *et al.* (2002), the DIC is a measure of fit based on a trade-off between the fit of the data to the model and the corresponding complexity of the model: DIC = goodness of fit + complexity. The fit for model M_i is measured in terms of the posterior distribution of the deviance statistic, $D(\theta_i) = -2\log L(Y\,|\,\theta_i)$, while complexity is measured by an estimate of the effective number of parameters:

$$d_i \;=\; \overline{D}_i - D(\bar{\theta}_i) \;=\; E\Big(D(\theta_i \mid Y, M_i) - D(E(\theta_i \mid Y, M_i)\Big)\,,$$

i.e., the posterior mean deviance minus deviance evaluated at the posterior mean of the parameters. The DIC is defined as:

$$DIC(M_i) \;=\; D(\bar{\theta}_i) + 2\,d_i\ .$$

The DIC generalizes the AIC (Akaike, 1973) in the sense that it explicitly applies to non nested non IID problems. Besides that, DIC can be approximated via MCMC samples from the posterior density. For the reasons exposed so far, in this work we used DIC instead of either AIC or BIC (Schwarz, 1978).

Models with smaller DIC are better supported by the data. The DIC is a positive number, in general. However, it can be negative but such occurrence does not pose any difficulty in terms of model comparison, since the focus is in the difference between two values and not in the DIC value itself.

In general, the d_i component is a positive value. However, it can be negative in cases where the likelihood function is not log-concave, when there is a conflict between the prior and the likelihood or when the posterior distribution of the parameters is too skewed or symmetric and multi-modal, so that the posterior mean/median are poor measures of central tendency. In those cases, the use of the posterior mode can be a fair alternative.

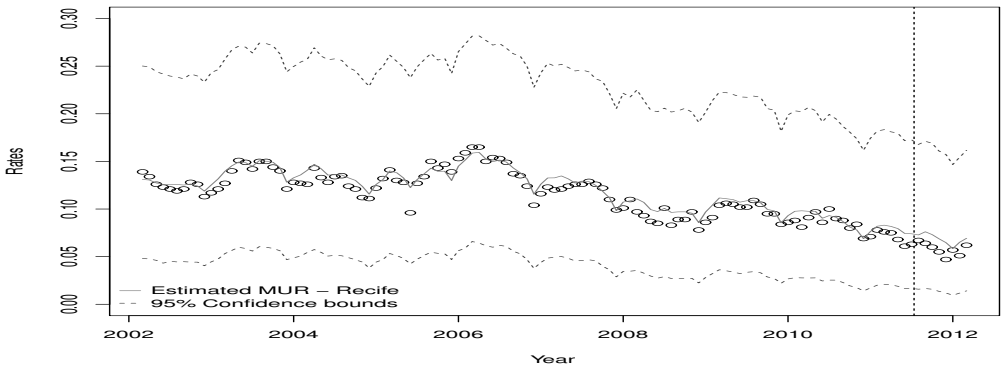
Table 1 displays the MAD, MSE, the effective number of parameters, d_i, and DIC values for Model 1 and Model 2. As we observe, the total MAD and total MSE values for Model 2 are somewhat smaller than those values for Model 1. However, the DIC for Model 2 is much smaller than the DIC for Model 1, giving strong indication that Model 2 provides a superior fit compared to Model 1. Additionally, the effective number of parameters, d_i, for both, Model 1 and Model 2, are positive. Thus besides of the limitations of the DIC, in our applications it seems to be performing properly.

Table 1: MSE, MAD, d_i and DIC values for Models 1 and 2.

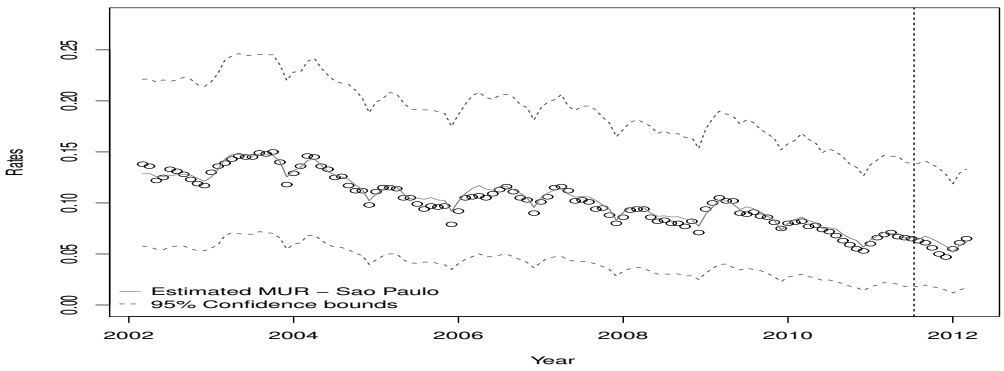
Model	**MSE**	**MAD**	d_i	**DIC**
Model 1	0.00158	0.03934	36.602	−1510.622
Model 2	0.00150	0.03559	229.819	−2445.007

In order to gain a better perspective of the real advantages of using Model 2 as opposed to Model 1, we present Figures 4 and 5 which display the estimated proportions or rates for each of the sub-populations and their corresponding confidence bands.

(a)



(b)



(c)

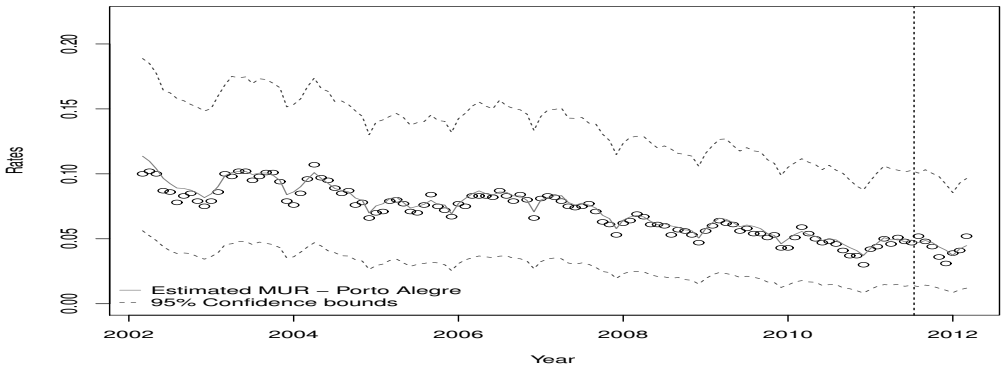
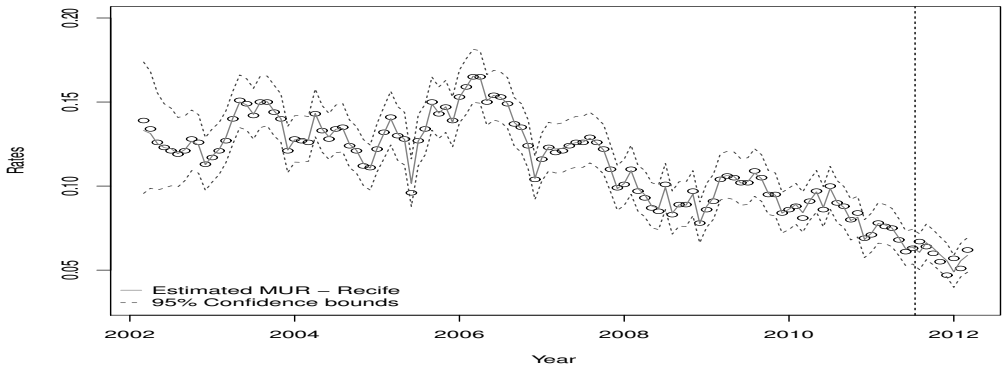


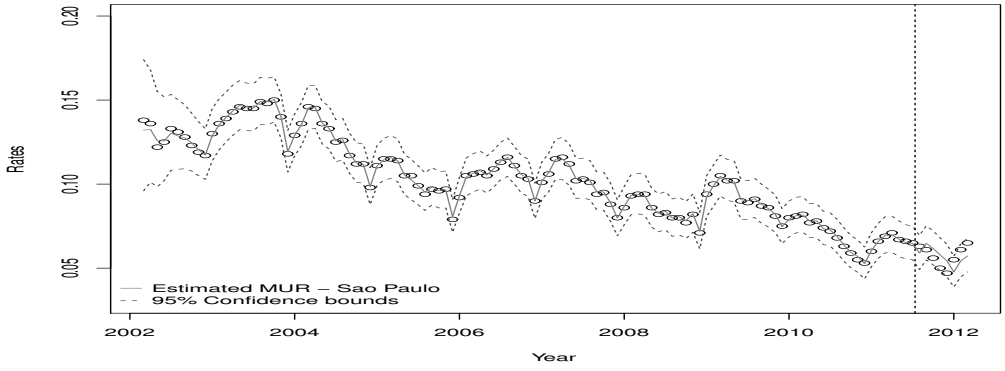
Figure 4: MUR data (Model 1 with static precision parameter) — estimated proportions and 95% credibility bounds for the three sub-populations: (a) Recife, (b) São Paulo and (c) Porto Alegre.
The forecast rates are presented after the dotted vertical lines.

(a)



[3mm]

(b)



[3mm]

(c)

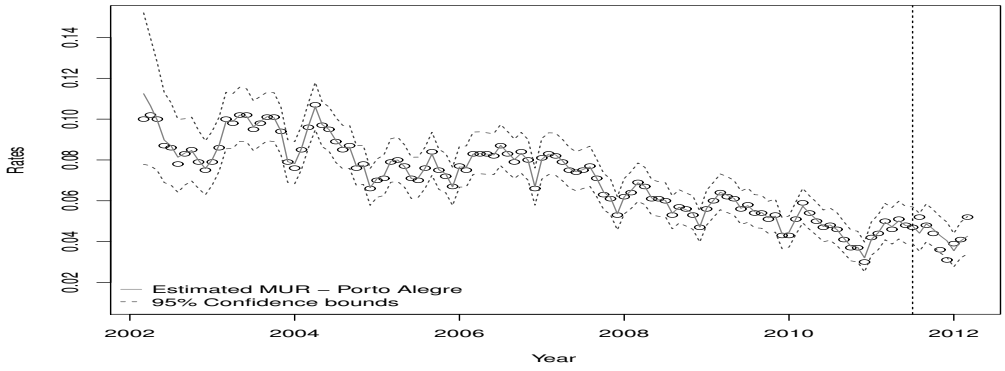


Figure 5: MUR data (Model 2 with dynamic precision parameter) — estimated proportions and 95% credibility bounds for the three sub-populations: (a) Recife, (b) São Paulo and (c) Porto Alegre.
The forecast rates are presented after the dotted vertical lines.

It is really reassuring the superiority of Model 2 compared to Model 1 in terms of both the precision of the credibility intervals, and how well Model 2 is able of describing the observed proportions for each of the sub-populations.

7. DISCUSSION

In this article we propose an extension to the Bayesian beta dynamic model developed by Da-Silva *et al.* (2011). We develop a hierarchical dynamic Bayesian beta model for modelling a set of time series of rates or proportions. The proposed methodology enables to combine the information contained in different time series so that we can describe a common underlying system, which is though flexible enough to allow the incorporation of random deviations, related to the individual series, not only through time but also across series. That allows to fit the case in which the observed series may present some degree of level shift. Additionally, the proposed model is adaptive in the sense that it incorporates precision parameters that can be heterogeneous no only over time but also across the series. The use of two link functions, one for the *mean process* and another to the *dispersion process*, makes such extension possible. Additionally, the choice of the matrices F_t and H_t allow for a multiplicity of ways of specifying the model, even allowing for the inclusion of covariates.

Missing observations can be easily accommodated: if the observation at time t is missing, then $y_t = NA$ and y_t does not carry any information. Then, we set $p(\theta_t \,|\, D_t) = p(\theta_t \,|\, D_{t-1})$.

Our methodology was applied to both real and simulated data. The real data set used are three time series of Brazilian monthly unemployment rates, observed in the cities of Recife, São Paulo and Porto Alegre, in the period from March 2002 to March 2012. We used a second-order polynomial trend seasonal effects to the parameters related to the mean, μ_{it}, and a second-order polynomial effects to the parameters related to the precision, ϕ_{it}. The very good features of the proposed model can be appreciated by the inspection of the graphs presented. The new parametrization of the precision parameter that was proposed by Bayer (2011) was used in the model formulation. It is very convenient since both, the link functions for μ_{it} and ϕ_{it}, are expressed in the $(0, 1)$ interval, which gives us a more meaningful interpretation in terms of the magnitude of the scale.

For future work we envision the possibility of extending the current model to enable the inclusion of different type of regimes for both, the level of the mean process and the level of the dispersion process.

ACKNOWLEDGMENTS

The authors are deeply indebted to the Office to Improve University Research (CAPES-Brazil), for financial support via Project PROCAD-NF 2008. Cibele Q. Da-Silva was supported by CNPq-Brazil, BPPesq. Helio S. Migon was supported by the National Research Council (CNPq-Brazil, BPPesq) and the Rio de Janeiro State Research Foundation (Faperj) together with CAPES via the Pronex project. We also want to thank the anonymous referees for their careful reading of this manuscript and also for their suggestions and criticisms which lead to very important improvements.

REFERENCES

[1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In: "Proceedings of the Second International Symposium on Information Theory" (B.N. Petrov and F. Csáski, Eds.), Budapest, Akadémiai Kiadó, pp. 267–281.

[2] Bayer, F.B. (2011). *Inferência e Modelagem em Regressão Beta*, Doctoral Dissertation, Statistics Department, Federal University of Pernambuco, Recife, Brazil.

[3] Cargnoni, C.; Müller, P. and West, M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models, *Journal of the American Statistical Association*, **92**, 438, 640–647.

[4] Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state-space models, *Biometrika*, **81**, 541–553.

[5] Cuervo-Cepeda, E. and Gamerman, D. (2004). Bayesian modeling of joint regressions for the mean and covariance matrix, *Biometrical Journal*, **46**(4), 430–440.

[6] Da-Silva, C.Q.; Migon, H.S. and Correia, L.T. (2011). Dynamic Bayesian beta models, *Computational Statistics & Data Analysis*, **55**, 2074–2089.

[7] Espinheira, P.L. (2007). *Regressão Beta*, Doctoral Dissertation, Statistics Department, Institute of Mathematics and Statistics, University of São Paulo (USP).

[8] Ferrari, S.L.P. and Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions, *Journal of Applied Statistics*, **31**, 799–815.

[9] Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of the Royal Statistical Society, Ser. B*, **55**, 629–642.

[10] Gamerman, D. and Migon, H. (1993). Dynamic hierarchical models, *Journal of the Royal Statistical Society, Series B*, **55**, 629–642.

[11] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.

[12] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemistry and Physics*, **21**, 1087–1091.

[13] Nelder, J.A. and Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments, *Appl. Stochastic Models Data Anal.*, **7**, 107–120.

[14] Petris, G. (2010). An R package for dynamic linear models, *Journal of Statistical Software*, **36**(12), 1–16. http://www.jstatsoft.org/v36/i12/

[15] Pole, A.; Mike West, M. and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*, CRC Press, USA.

[16] Pregibon, D. (1984). Review of generalized linear methods by McCullagh and Nelder, *Ann. Statist.*, **12**, 1589–1596.

[17] Rocha, A.V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models, *Test*, **18**(3), 529–545.

[18] Simas, A.B.; Barreto-Souza, W. and Rocha, A.V. (2010). Improved estimators for a general class of beta regression models, *Computational Statistics & Data Analysis*, **54**, 348–366.

[19] Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, *Psychological Methods*, **11**(1), 54–71.

[20] Smyth, G.K. and Verbyla, A.P. (1999). Adjusted likelihood methods for modeling dispersion in generalized linear models, *Environmetrics*, **10**(6), 695–709.

[21] Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Linde, A. Van der (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

[22] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 497–511.

[23] West, M. and Harrison, P.J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd edition, Springer, New York.

REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 75–87

ON THE BOUNDS FOR DIAGONAL AND OFF-DIAGONAL ELEMENTS OF THE HAT MATRIX IN THE LINEAR REGRESSION MODEL

Author: Mohammad Mohammadi
– Department of Statistics, University of Isfahan,
Isfahan, 81746-73441, Iran
m.mohammadi@sci.ui.ac.ir

Received: April 2014 Revised: September 2014 Accepted: November 2014

Abstract:

- In the least squares analysis, an appropriate criterion to detect potentially influential observations, either individually or jointly, deals with the values of corresponding Hat matrix elements. Hence, some conditions for which these elements give the extreme values are interesting in the model sensitivity analysis. In this article, we find a new and sharper lower bound for off-diagonal elements of the Hat matrix in the intercept model, which is shorter than those for the no-intercept model. We give necessary and sufficient conditions on the space of design matrix, under which the corresponding Hat matrix elements get desired extreme values.

Key-Words:

- *Hat matrix; high-leverage; influential observations; linear regression model.*

AMS Subject Classification:

- 62J05.

1. INTRODUCTION

In the least squares approach, any sensitivity analysis is essentially related to how points are observed, so reflected on the elements of the Hat matrix. As the most widely used concepts in regression diagnostics, influential observations and outliers are identified by the size of these quantities. Consider the general linear regression model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \ , \qquad (i = 1, 2, ..., n) \ , \tag{1.1}$$

where y_i is the i-th observed response, $\mathbf{x}_i$ is a $p \times 1$ deterministic vector, $\boldsymbol{\beta} \in R^p$ is an unknown $p \times 1$ vector of parameters, and the ε_i's are uncorrelated errors with mean zero and variance σ^2. Writing $\mathbf{y} = (y_1, ..., y_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)'$, and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)'$, model (1.1) can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \ . \tag{1.2}$$

The matrix $\mathbf{X}$ is called *design matrix*, which contains the column one in the intercept model. We assume throughout that $\mathbf{X}$ is full-rank matrix, so $\mathbf{X}'\mathbf{X}$ is nonsingular. In this case the ordinary least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \ . \tag{1.3}$$

The $n \times 1$ vector of ordinary predicted values of the response variable is $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$, where the $n \times n$ prediction or Hat matrix, $\mathbf{H}$, is given by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \ . \tag{1.4}$$

The residual vector is given by $\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ with the variance-covariance matrix $\mathbf{V} = (\mathbf{I}_n - \mathbf{H})\sigma^2$, where $\mathbf{I}_n$ is the identity matrix of order n. The matrix $\mathbf{H}$ plays an important role in the linear regression analysis. Let h_{ij} indicate the (i, j)-th element of $\mathbf{H}$. Hence,

$$h_{ij} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j \ , \qquad (i, j = 1, 2, ..., n) \ . \tag{1.5}$$

The diagonal element h_{ii} is so-called the leverage of the i-th data point and measures how far the observation $\mathbf{x}_i$ is from the rest of points in the $\mathbf{X}$-space. Any point with large values of h_{ii} tends to be an influential observation. Such a point is called *high-leverage*. Cook and Weisberg (1982, p. 13) point to the following conditions, to h_{ii} be large:

- $\mathbf{x}_i'\mathbf{x}_i$ is large relative to the square of the norm $\mathbf{x}_j'\mathbf{x}_j$ of the vectors $\mathbf{x}_j$; i.e. $\mathbf{x}_i$ is far removed from the bulk of other points in the data set, or
- $\mathbf{x}_i'\mathbf{x}_i$ is substantially in the direction of an eigenvector corresponding to a small eigenvalue of $\mathbf{X}'\mathbf{X}$.

The various criteria are suggested for the size of h_{ii} to $\mathbf{x}_i$ being high-leverage (see Chatterjee and Hadi, 1988, p. 100–101).

On the other hand, off-diagonal elements of the Hat matrix may be regarded as another criterion in the regression analysis. Ignoring the constant σ^2, these elements are covariances of any pair of the estimated residuals, so can be useful to check the independency assumption. From theoretical point of view, there may exist situations in which observations are jointly but not individually influential (Chatterjee and Hadi, 1988, p. 185). Huber (1975) mentions that large values of h_{ij} typically correspond to outlying design points. Hadi (1990) proposed two graphical displays of the elements of $\mathbf{H}$, that are useful in the detection of potentially influential subsets of observations.

In this paper we discuss the necessary and sufficient conditions for the design matrix to have some extreme values of Hat matrix elements, in the intercept and no-intercept linear regression models. We obtain a sharper lower bound for off-diagonal elements of the Hat matrix in the with intercept linear model, which is shorter than those for no-intercept model by $1/n$.

Repeated application of the following first lemma is made. Part (a) of this lemma is due to Chipman (1964).

Lemma 1.1. *Let* $\mathbf{A}$ *be a matrix of* $n\times p$ *with rank* $p - m_1$, $(m_1 > 0)$.

(**a**) *If* $\mathbf{B}$, *of order* $m_1 \times p$ *and full row rank, has it's rows LIN (linearly independent) of those of* $\mathbf{A}$, *then*

$$\mathbf{A}(\mathbf{A}'\mathbf{A} + \mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' = \mathbf{0}_{n\times m_1} \quad \textit{and} \quad \mathbf{B}(\mathbf{A}'\mathbf{A} + \mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' = \mathbf{I}_{m_1} .$$

(**b**) *If* $\mathbf{R}$, *of order* $m_2 \times p$; *(*$m_2 \leq m_1$*) and rank 1, has the first row* $\mathbf{r}'$ *of the form* $\mathbf{R} = \boldsymbol{\delta}\mathbf{r}'$, *where* $\boldsymbol{\delta} = (1, \delta_2, ..., \delta_{m_2})'$, *and* $\mathbf{r}$ *be LIN of rows of* $\mathbf{A}$, *then*

$$\mathbf{R}(\mathbf{A}'\mathbf{A} + \mathbf{R}'\mathbf{R})^{-1}\mathbf{R}' = \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{\|\boldsymbol{\delta}\|^2} .$$

Lemma 1.2. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n\times p$ *matrices. Then,* $\text{rank}(\mathbf{A} - \mathbf{B}) = \text{rank}(\mathbf{A}) - \text{rank}(\mathbf{B})$, *if and only if* $\mathbf{A}\mathbf{A}^-\mathbf{B} = \mathbf{B}\mathbf{A}\mathbf{A}^- = \mathbf{B}\mathbf{A}^-\mathbf{B} = \mathbf{B}$, *where* $\mathbf{A}^-$ *is a generalized inverse of* $\mathbf{A}$ *satisfying* $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ *(see Seber, 2007).*

Throughout this paper we use the notation (i) written as a subscript to a quantity to indicate the omission of the i-th observation. For example, $\mathbf{X}_{(i)}$ and $\mathbf{X}_{(ij)}$ are matrix $\mathbf{X}$ with the i-th row and (i, j)-th rows omitted, respectively. The vector $\bar{\mathbf{x}}$ denotes the mean of $\mathbf{X}$'s rows and $\mathbf{J}_p$ is a $p\times p$ matrix of ones.

2. BOUNDS FOR DIAGONAL ELEMENTS OF THE HAT MATRIX

This section is allotted to determine the lower and upper bounds of h_{ii}, along with necessary and sufficient conditions for observation matrix $\mathbf{X}$ to take those values. These conditions are fundamentally on the basis of some special forms of $\mathbf{x}_i$ and $\mathbf{X}_{(i)}$. We consider two customary full rank linear regression models; without and with intercept.

Lemma 2.1. *Let $\mathbf{X}_{n\times k}$ be full column rank matrix without column one. Then,*

- **(i)** $0 \le h_{ii} \le 1$.
- **(ii)** h_{ii}*=0, if and only if* $\mathbf{x}_i = \mathbf{0}$.
- **(iii)** h_{ii}*=1, if and only if* $\operatorname{rank}(\mathbf{X}_{(i)}) = k-1$.

Proof: Part (i) is immediately proved since $\mathbf{H}$ and $\mathbf{I}_n - \mathbf{H}$ are positive semi-definite (p.s.d.) matrices. Similarly part (ii) is obtained since $(\mathbf{X}'\mathbf{X})^{-1}$ is a p.d. matrix. To verify part (iii), without loss of generality, suppose that $\mathbf{x}_i$ is the last row of $\mathbf{X}$, i.e. $\mathbf{X}' = \begin{bmatrix} \mathbf{X}'_{(i)} & \mathbf{x}_i \end{bmatrix}$. If h_{ii}=1, then

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)} & \mathbf{0}_{(n-1)\times 1} \\ \mathbf{0}_{1\times(n-1)} & 1 \end{bmatrix}. \tag{2.1}$$

Since $\mathbf{H}$ is an idempotent matrix, $\mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}$ is also idempotent. Hence,

$$\operatorname{rank}(\mathbf{X}_{(i)}) = \operatorname{rank}\big(\mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\big) = \operatorname{trace}\big(\mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\big) = k-1 .$$

Conversely, let $\operatorname{rank}(\mathbf{X}_{(i)}) = k-1$. Since, $\operatorname{rank}\begin{bmatrix} \mathbf{X}'_{(i)} & \mathbf{x}_i \end{bmatrix} = \operatorname{rank}(\mathbf{X}') = k$, it follows that $\mathbf{x}_i$ is LIN from the rows of $\mathbf{X}'_{(i)}$. Using part (a) of Lemma 1.1,

$$\mathbf{x}'_i \left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)} + \mathbf{x}_i\mathbf{x}'_i\right)^{-1} \mathbf{x}_i (= h_{ii}) = 1 ,$$

and proof is completed. □

Lemma 2.2. *If the full column rank matrix $\mathbf{X}_{n\times(k+1)}$ contains column one, then*

- **(i)** $\frac{1}{n} \le h_{ii} \le 1$.
- **(ii)** $h_{ii} = \frac{1}{n}$*, if and only if* $\mathbf{x}_i = \bar{\mathbf{x}}$.
- **(iii)** h_{ii}*=1, if and only if* $\operatorname{rank}(\mathbf{X}_{(i)}) = k$.

Proof: In this case $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ and $\mathbf{I}_n - \mathbf{H}$ are both p.s.d. matrices, so part (i) holds. To verify part (ii) note that in the with intercept model, we have:

$$(\mathbf{X}'\mathbf{X})^{-1}\bar{\mathbf{x}} = \frac{1}{n}\begin{bmatrix} 1 \\ \mathbf{0}_{k\times 1} \end{bmatrix}. \tag{2.2}$$

The sufficient condition is established by noting that

$$\mathbf{x}_i'\,(\mathbf{X}'\mathbf{X})^{-1}\,\bar{\mathbf{x}} \;=\; \bar{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\,\bar{\mathbf{x}} \;=\; \frac{1}{n}\;. \tag{2.3}$$

Conversely, if $h_{ii} = 1/n$, we have:

$$(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \;=\; 0$$

which satisfies $\mathbf{x}_i = \bar{\mathbf{x}}$. Part (iii) is verified similar to part (iii) of Lemma 2.1. □

Example 2.1. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with usual assumption. In this case,

$$h_{ii} \;=\; \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2}\;.$$

It is clear that $x_i = \bar{x}$ satisfies $h_{ii} = 1/n$. Also, if for all $k \neq i$, we have $x_k = c\,(\neq x_i)$, then $\bar{x} = c + (x_i - c)/n$ and $h_{ii} = 1$. Figures 1 and 2 show two examples of these situations. In Figure 1, the i-th observation gives minimum possible value h_{ii}, and the fitted slope is not affected by this observation. Conversely, Figure 2 shows an example with maximum possible value for h_{ii}. In this case, the slope of fitted line is determined by y_i, and deleting such observation changes $\mathbf{X}'_{(i)}\mathbf{X}_{(i)}$ to a singular matrix.

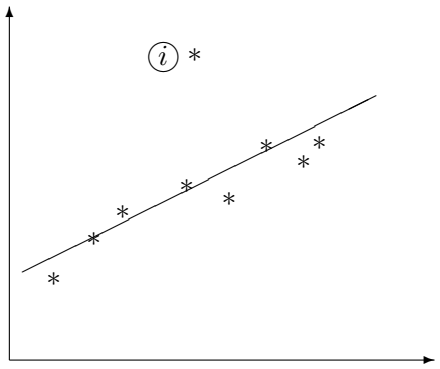


Figure 1: A simple linear regression model with intercept for which $h_{ii} = \frac{1}{n}$.

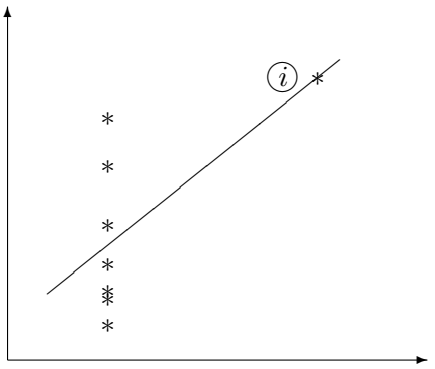


Figure 2: A simple linear regression model with intercept for which $h_{ii} = 1$.

3. BOUNDS FOR OFF-DIAGONAL ELEMENTS OF THE HAT MATRIX

In this case we assume two situations with and without intercept term in the linear regression model. Part (i) of the following lemma is shown by Chatterjee and Hadi (1988, p. 18). (They have appreciated Professor J. Brian Gray for bringing part (i) of this lemma).

Lemma 3.1. *Let* $\mathbf{X}_{n\times k}$ *be full column rank matrix without column one. Then,*

(i) $-\frac{1}{2} \le h_{ij} \le \frac{1}{2}$.

(ii) $h_{ij} = -\frac{1}{2}$, *if and only if* $\mathbf{x}_i = -\mathbf{x}_j$ *and* $\text{rank}(\mathbf{X}_{(ij)}) = k-1$.

(iii) $h_{ij} = \frac{1}{2}$, *if and only if* $\mathbf{x}_i = \mathbf{x}_j$ *and* $\text{rank}(\mathbf{X}_{(ij)}) = k-1$.

Proof: Since $\mathbf{H}$ is idempotent, we have:

$$h_{ii} \;=\; \sum_{i=1}^{n} h_{ik}^2 \;=\; h_{ii}^2 + h_{ij}^2 + \sum_{k\neq(i,j)} h_{ik}^2 \;, \tag{3.1}$$

which implies that $h_{ij}^2 = h_{ii}(1-h_{ii}) + \sum_{k\neq(i,j)} h_{ik}^2$. Since $0 \le h_{ii} \le 1$, part (i) is obtained by conditions $h_{ii} = h_{jj} = 1/2$ and $h_{ik} = h_{jk} = 0$ for all $k\,(\neq i,j) = 1,2,...,n$. To verify sufficient condition of part (ii), let $h_{ij} = -1/2$. From (3.1) we have $h_{ii} = h_{jj} = 1/2$, so

$$(\mathbf{x}_i + \mathbf{x}_j)'\,(\mathbf{X}'\mathbf{X})^{-1}\,(\mathbf{x}_i + \mathbf{x}_j) \;=\; 0 \;,$$

which holds only if $\mathbf{x}_i = -\mathbf{x}_j$. Again, if $\mathbf{X}' = \big[\, \mathbf{X}'_{(ij)} \;\; \mathbf{x}_i \;\; \mathbf{x}_j \,\big]$, then

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}_{(ij)}\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)} & \mathbf{0}_{(n-2)\times 2} \\ \mathbf{0}_{2\times(n-2)} & \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix}. \tag{3.2}$$

Since $\mathbf{H}$ is idempotent, it follows from equation (3.2) that $\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}$ is also idempotent. Hence,

$$\text{rank}\big(\mathbf{X}_{(ij)}\big) \;=\; \text{rank}\big(\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}\big) \;=\; \text{trace}\big(\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}\big) \;=\; k-1 \;.$$

Conversely, if $\mathbf{x}_i = -\mathbf{x}_j$ and $\text{rank}(\mathbf{X}_{(ij)}) = k-1$, since $\text{rank}\big[\, \mathbf{X}'_{(ij)} \;\; \mathbf{x}_i \;\; \mathbf{x}_j \,\big] = \text{rank}(\mathbf{X}) = k$, it follows that $\mathbf{x}_i$ is LIN from the rows of $\mathbf{X}_{(ij)}$. Applying part (b) of Lemma 1.1 with replacing $\mathbf{A}$ and $\mathbf{R}$ by $\mathbf{X}_{(ij)}$ and $\big[\, \mathbf{x}_i \;\; -\mathbf{x}_i \,\big]$, with $\boldsymbol{\delta} = (1,-1)$ gives

$$\mathbf{R}(\mathbf{A}'\mathbf{A} + \mathbf{C}'\mathbf{C})^{-1}\mathbf{R}' \;=\; \begin{bmatrix} h_{ii} & h_{ij} \\ h_{ij} & h_{jj} \end{bmatrix} \;=\; \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

Part (iii) is proved similarly by multiplying $\mathbf{x}_j$ by -1. □

The following lemma gives the boundary of h_{ij} in the with intercept model. We will find its upper bound similar to the case of the no-intercept model, whereas its lower bound has sharpened by the constant $1/n$.

Lemma 3.2. *If* $\mathbf{X}_{n\times(k+1)}$ *is full column rank matrix with column one, then*

(i) $\frac{1}{n} - \frac{1}{2} \le h_{ij} \le \frac{1}{2}$.

(ii) $h_{ij} = \frac{1}{n} - \frac{1}{2}$, *if and only if* $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$ *and* $\text{rank}(\mathbf{X}_{(ij)}) = k$.

(iii) $h_{ij} = \frac{1}{2}$, *if and only if* $\mathbf{x}_i = \mathbf{x}_j$ *and* $\text{rank}(\mathbf{X}_{(ij)}) = k$.

Proof: In this case $\mathbf{H}$ is idempotent and has the property of a transition probability matrix, i.e. $\mathbf{H1}=\mathbf{1}$. Thus, we should minimize h_{ij} with restriction (3.1) along with

$$\sum_{i=1}^{n} h_{ik} = 1 \ . \tag{3.3}$$

Using λ as a Lagrangian multiplier, we minimize

$$h_{ij} = 1 - h_{ii} - \sum_{k\neq(i,j)} h_{ik} + \lambda \left[h_{ii}(1-h_{ii}) - h_{ij}^2 - \sum_{k\neq i,j} h_{ik}^2 \right], \tag{3.4}$$

with respect to the λ and elements h_{ik} for $k \neq j = 1, 2, ..., n$. Clearly $\partial h_{ij}/\partial\lambda = 0$ gives (3.1), and $\partial h_{ij}/\partial h_{ii} = 0$ gives $h_{ii} = \frac{1}{2} - \frac{1}{2\lambda}$. On the other hand setting $\partial h_{ij}/\partial h_{ik} = 0$ results to $h_{ik} = -\frac{1}{2\lambda}$. Substituting in (3.1) gives:

$$h_{ij}^2 = \frac{1}{4}\left(1 - \frac{n-1}{\lambda^2}\right), \tag{3.5}$$

and so (3.3) yields

$$h_{ij} = \frac{1}{2}\left(1 + \frac{n-1}{\lambda}\right). \tag{3.6}$$

Solving equations (3.5) and (3.6) with respect to λ gives the boundary of h_{ij} as

$$\frac{1}{n} - \frac{1}{2} \leq h_{ij} \leq \frac{1}{2} \ .$$

In order to prove part (ii), note that $h_{ij} = 1/n - 1/2$ produces all h_{ik}, $(k \neq i, j)$ be equal to $1/n$, which leads to $h_{ii} = h_{jj} = 1/n + 1/2$. Hence,

$$(\mathbf{x}_i + \mathbf{x}_j - 2\bar{\mathbf{x}})' (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{x}_i + \mathbf{x}_j - 2\bar{\mathbf{x}}) = 0 \ ,$$

which holds only if $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$. Furthermore, we have

$$\mathbf{H} - \frac{1}{n}\mathbf{J}_n = \begin{bmatrix} \mathbf{X}_{(ij)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2} & \mathbf{0}_{(n-2)\times 2} \\ \mathbf{0}_{2\times(n-2)} & \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix}. \tag{3.7}$$

Since $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ is idempotent, equation (3.7) results to $\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2}$ is idempotent, also. Hence,

$$\begin{aligned} k - 1 &= \operatorname{trace}\left(\mathbf{X}_{(ij)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2} \right) \\ &= \operatorname{rank}\left(\mathbf{X}_{(ij)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2} \right). \end{aligned}$$

We now show that the last rank of difference matrix is equal to the difference of corresponding rank of matrices. Let $\mathbf{A} = \mathbf{X}_{(ij)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_{(ij)}$ and $\mathbf{B} = \frac{1}{n}\mathbf{J}_{n-2}$.

Since $\mathbf{A}$ is symmetric, we have $\mathbf{A}\mathbf{A}^- = \mathbf{A}^-\mathbf{A}$, resulting $\mathbf{A}^2\mathbf{A}^- = \mathbf{A}$. Using equation (3.7) and noting that $(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{1} = \mathbf{0}$, we have

$$\mathbf{AB} = \mathbf{BA} = \mathbf{B}^2 = \left(\frac{n-2}{n}\right)\mathbf{B} \qquad \text{and} \qquad \mathbf{A}^2 = \mathbf{A} - \frac{2}{n}\mathbf{B}\ .$$

Therefore,

$$\left(\mathbf{A} - \frac{2}{n}\mathbf{B}\right)\mathbf{A}^- \ = \ \mathbf{A}\ . \tag{3.8}$$

Multiplying (3.8) by $\mathbf{A}$ from the left side, we find $\mathbf{ABA}^- = \mathbf{B}$. Similarly, the equality $\mathbf{A}^-\mathbf{BA} = \mathbf{B}$ is verified. It remains to show that $\mathbf{BA}^-\mathbf{B} = \mathbf{B}$. Multiplying (3.8) by $\mathbf{B}$ to the right hand side and noting that $\mathbf{A}^-\mathbf{B}$ is symmetric, we have

$$\mathbf{BA}^-\mathbf{B} \ = \ \frac{n}{2}\big(\mathbf{AA}^-\mathbf{B} - \mathbf{AB}\big) \ = \ \frac{n}{2}\left[\mathbf{B} - \left(\frac{n-2}{n}\right)\mathbf{B}\right] = \ \mathbf{B}\ .$$

Using Lemma 1.2, we have $\operatorname{rank}\big(\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}\big) - \operatorname{rank}(\mathbf{J}_{n-2}) = k-1$, and thus $\operatorname{rank}\big(\mathbf{X}_{(ij)}\big) = k$.

Conversely, suppose $\mathbf{X}_{(ij)}$ of order $(n-2)\times(k+1)$ has rank k and $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}} = \bar{\mathbf{x}}_{(ij)}$, the row means of $\mathbf{X}_{(ij)}$. In this case $h_{ij} = 2/n - h_{ii}$. Now, since $\mathbf{X}_{(i)}$ is full column rank, then $\mathbf{x}_j$ is LIN from the rows of $\mathbf{X}_{(ij)}$. Using part (a) of the Lemma 1.1, we have:

$$\begin{aligned}
\mathbf{x}_i'\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i \ &= \ (2\bar{\mathbf{x}} - \mathbf{x}_j)'\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}(2\bar{\mathbf{x}} - \mathbf{x}_j) \\
&= \ 4\bar{\mathbf{x}}'_{(ij)}\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\bar{\mathbf{x}}_{(ij)} \\
&\quad - \ \frac{4}{n-2}\mathbf{1}'\mathbf{X}_{(ij)}\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\mathbf{x}_j \\
&\quad + \ \mathbf{x}_j'\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\mathbf{x}_j \\
&= \ 4\bar{\mathbf{x}}'_{(ij)}\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\bar{\mathbf{x}}_{(ij)} \ + 1 \\
&= \ 4\left(\frac{(n-1)\bar{\mathbf{x}}_{(i)} + \mathbf{x}_i}{n}\right)'\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\left(\frac{(n-1)\bar{\mathbf{x}}'_{(i)} + \mathbf{x}_i}{n}\right) + 1 \\
&= \ 4\left[\mathbf{x}_i'\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}n + 1\right] + 1\ .
\end{aligned}$$

Hence, $\mathbf{x}_i'(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}\mathbf{x}_i = \frac{h_{ii}}{1-h_{ii}} = \frac{n+2}{n-2}$, which implies $h_{ij} = 1/n - 1/2$. Proof of part (iii) is analogous to part (iii) of Lemma 3.1. □

Example 3.1. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with usual assumptions. In this model,

$$h_{ij} \ = \ \frac{1}{n} + \frac{(x_i - \bar{x})\,(x_j - \bar{x})}{\sum_{k=1}^{n}(x_k - \bar{x})^2}\ .$$

Now if $x_i = x_j = c$ and $x_k = d \neq c$ (for every $k \neq i, j$) then $\bar{x} = d + 2(c-d)/n$. It is easy to show that $h_{ij} = 1/2$. On the other hand, if $x_i \neq x_j$, $x_k = d \neq x_i, x_j$ (for every $k \neq i, j$) and $x_i + x_j = 2\bar{x} = 2d$, then $h_{ij} = 1/n - 1/2$. Figures 3 and 4 show two examples of mentioned situations, in the case when h_{ij} gives its maximum and minimum possible values.

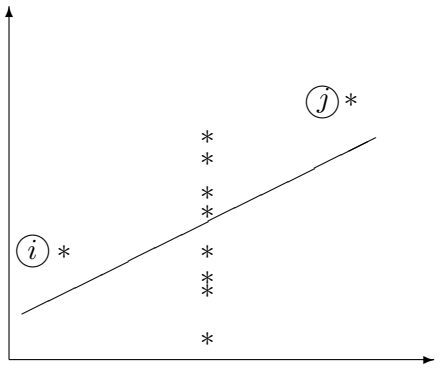


Figure 3: A simple linear regression model with intercept for which $h_{ij} = \frac{1}{n} - \frac{1}{2}$.

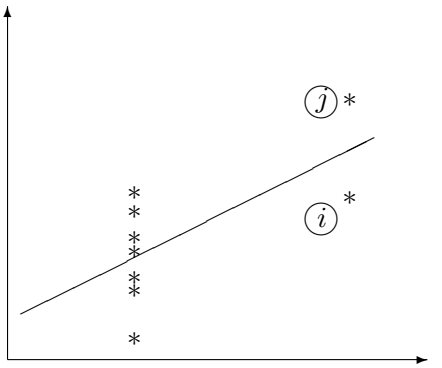


Figure 4: A simple linear regression model with intercept for which $h_{ij} = \frac{1}{n} + \frac{1}{2}$.

Example 3.2. Suppose the multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$ with design matrix $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & 8 & 4 \\ 1 & 1 & 6 & 6 \\ 1 & 3 & 5 & 8 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 14 & 2 \\ 1 & 4 & 9 & 11 \\ 1 & 1 & 7 & 2 \\ 1 & 2 & 6 & 5 \end{bmatrix}.$$

Hat matrix is

$$\mathbf{H} = \begin{bmatrix} 0.625 & -0.375 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 \\ -0.375 & 0.625 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 \\ 0.125 & 0.125 & 0.298 & 0.133 & -0.161 & 0.335 & -0.003 & 0.148 \\ 0.125 & 0.125 & 0.133 & 0.618 & -0.196 & -0.235 & 0.242 & 0.188 \\ 0.125 & 0.125 & -0.161 & -0.196 & 0.791 & 0.008 & 0.296 & 0.049 \\ 0.125 & 0.125 & 0.335 & -0.235 & 0.008 & 0.658 & -0.123 & 0.106 \\ 0.125 & 0.125 & -0.003 & 0.242 & 0.260 & -0.123 & 0.250 & 0.124 \\ 0.125 & 0.125 & 0.148 & 0.188 & 0.049 & 0.106 & 0.124 & 0.126 \end{bmatrix}.$$

It is observed that $h_{12} = -0.375 = 1/n - 1/2$, and this is because of $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$ and for any $i \geq 3$: $x_{i3} = 3x_{i1} - 1$.

4. CONCLUDING REMARKS

A large number of statistical measures, such as Mahalanobis distance, weighted square standardized distance, PRESS, etc, have been proposed in the literatures of diagnosing influential observations, which are typically based on h_{ij}'s. Removing the i-th point or (i, j)-th points jointly may be useful to detect the leverage in regression diagnostics. The following outcomes are obtained from the previous lemmas in sections 2 and 3:

- $h_{ii} = 0$ (or $h_{ii} = 1/n$ in the intercept model). In this case the i-th observation potentially is an outlier, recognized by large distance between y_i and $\bar{y}$. This point has no effect on the estimation of unknown parameter $\boldsymbol{\beta}$, except constant term in the with intercept model (see Figure 1). In this situation, y_i has minimum effect to determine $\hat{y}_i$.
- $h_{ii} = 1$. Presence of such point obviates full collinearity of some columns of $\mathbf{X}$, so it is likely to be an influential observation. This point is capable to elongate the regression line itself. In other words, the fitted regression line passes through other data points to place of the i-th observation. In this case we see $e_i = 0$ (see Figure 2).
- $h_{ij} = -1/2$ (or $h_{ij} = 1/n - 1/2$ in the intercept model). This case may be declared as a competition between i-th and j-th observations. Using Lemma 3.1 and Lemma 3.2, it can be shown that if any of these points removed, then other point has the maximum value 1 of diagonal element of corresponding Hat matrix constructed based on the remaining $n - 1$ observations, so will be an influential observation. In this case, $e_i = e_j$, so $\rho(\hat{y}_i, \hat{y}_j) = -1$. This situation occurs when (i, j)-th points are at the different sides of the bulk of other points (see Figure 3).
- $h_{ij} = 1/2$. Contrary to the previous case, in this case the i-th and the j-th observations are at the same side of the bulk of other points. It can be shown that predicted values of these observations are at the same direction, i.e. $\rho(\hat{y}_i, \hat{y}_j) = 1$ (see Figure 4).

APPENDIX: Proof of Lemma 1.1

(a): Without loss of generality, let the first $p-m_1$ rows $\mathbf{A}_1$ of $\mathbf{A}$ be full row rank; then the last $(n+m_1-p)$ rows $\mathbf{A}_2$ of $\mathbf{A}$ may be written as $\mathbf{A}_2 = \mathbf{N}\mathbf{A}_1$, where $\mathbf{N}$ is $(n+m_1-p)\times(p-m_1)$. Since $\mathbf{B}$ has its rows LIN of those of $\mathbf{A}$, we may define:

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{D} \end{bmatrix},$$

where $\mathbf{C}_1$ and $\mathbf{D}$ are $p\times(p-m_1)$ and $p\times m_1$ matrices, respectively. Then,

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1\mathbf{C}_1 & \mathbf{A}_1\mathbf{D} \\ \mathbf{B}\mathbf{C}_1 & \mathbf{B}\mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{p-m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_1} \end{bmatrix}.$$

Now define the $p\times(n+m_1-p)$ matrix $\mathbf{C}_2$ as $\mathbf{C}_2 = \mathbf{C}_1\mathbf{N}'$. So, we have

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{p-m_1} \\ \mathbf{N} \end{bmatrix} \mathbf{A}_1 \ ,$$

and

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix} = \mathbf{C}_1 \begin{bmatrix} \mathbf{I}_{p-m_1} & \mathbf{N}' \end{bmatrix}.$$

From the solutions we obtain

$$\mathbf{AC} = \begin{bmatrix} \mathbf{I}_{p-m_1} & \mathbf{N}' \\ \mathbf{N} & \mathbf{N}\mathbf{N}' \end{bmatrix}, \qquad \mathbf{AD} = \mathbf{0}_{n\times m_1} = (\mathbf{BC})', \qquad \mathbf{BD} = \mathbf{I}_{m_1} \ ,$$

where $\operatorname{rank}(\mathbf{AC}) = \operatorname{rank}(\mathbf{A}) = p-m_1$. Now since $\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}$ has rank p, $\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B}$ is positive definite and therefore invertible. From above we have expressions

$$(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})\,\mathbf{D} \;=\; \mathbf{B}' \ .$$

Premultiplying by $(\mathbf{A}(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1})$, and then by $(\mathbf{B}(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1})$ we obtain

$$\mathbf{A}\,(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1}\,\mathbf{B}' \;=\; \mathbf{AD} \;=\; \mathbf{0}_{n\times m_1}$$

and

$$\mathbf{B}\,(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1}\,\mathbf{B}' \;=\; \mathbf{BD} \;=\; \mathbf{I}_{m_1} \ .$$

(b): Equalizing $\mathbf{R} = \boldsymbol{\delta}\mathbf{r}'$ results to:

$$\mathbf{R}\,(\mathbf{A}'\mathbf{A}+\mathbf{R}'\mathbf{R})^{-1}\,\mathbf{R}' \;=\; \boldsymbol{\delta}\mathbf{r}'\,(\mathbf{A}'\mathbf{A}+\mathbf{r}\,\boldsymbol{\delta}\boldsymbol{\delta}'\mathbf{r}')^{-1}\,\mathbf{r}\boldsymbol{\delta}' \ .$$

Now using part (a) and substituting $\mathbf{B}$ by $\sqrt{\boldsymbol{\delta}\boldsymbol{\delta}'}\mathbf{r}'$ give

$$\mathbf{R}\,(\mathbf{A}'\mathbf{A}+\mathbf{R}'\mathbf{R})^{-1}\,\mathbf{R}' \;=\; \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{\|\boldsymbol{\delta}\|^2} \ .$$

REFERENCES

[1] Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*, New York: John Wiley & Sons.

[2] Chipman, J.S. (1964). On least squares with insufficient observations, *Journal of the American Statistical Association*, **59**, 1078–1111.

[3] Cook, R.D. and Weisberg, S. (1982). *Residual and Influence in Regression*, London: Chapman & Hall.

[4] Hadi, A.S. (1990). Two graphical displays for the detection of potentially influential subsets in regression, *Journal of Applied Statistics*, **17**, 313–327.

[5] Hoaglin, D.C. and Welsch, R.E. (1978). The Hat matrix in regression and ANOVA *The American Statistician*, **32**, 17–22.

[6] Huber, P. (1975). *Robustness and designs*. In "A Survey of Statistical Design and Linear Models" (Jagdish N. Srivastava, Ed.), Amsterdam, North-Holland Publishing Co.

[7] Seber, S.R. (2007). *Matrix handbook for Statisticians*, New York, John Wiley & Sons.

REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 89–100

SHAPIRO–WILK TEST WITH KNOWN MEAN

Authors: Zofia Hanusz
– Department of Applied Mathematics and Computer Science, University of Life Sciences in Lublin, Poland
zofia.hanusz@up.lublin.pl

Joanna Tarasinska
– Department of Applied Mathematics and Computer Science, University of Life Sciences in Lublin, Poland
joanna.tarasinska@up.lublin.pl

Wojciech Zielinski
– Department of Econometrics and Statistics, Warsaw University of Life Sciences, Poland
wojtek.zielinski@statystyka.info

Received: August 2014 Revised: October 2014 Accepted: November 2014

Abstract:

- An adaptation of the Shapiro–Wilk W test to the case of normality with a known mean is considered. The table of critical values for different sample sizes and several significance levels is given. The power of the test is investigated and compared with the Kolmogorov test and the two-step procedure consisting of the Shapiro–Wilk W and t tests. Additionally, the normalizing coefficients for the test statistic are given.

Key-Words:

- *normality; Shapiro–Wilk test; Kolmogorov test; Student t test; power.*

AMS Subject Classification:

- 62F03; 62E20.

1. INTRODUCTION AND MOTIVATION

One of the common problems in applications is to check whether the mean value of an investigated phenomena equals a given number, i.e. testing the hypothesis $H_0\colon \mu = \mu_0$. For example, for econometrical applications see [5], for biological applications [8], for engineering [11], for medical applications [7]. See also [1], [2], [4], [6].

To test the hypothesis H_0, the classical t test is used. However, this test requires the assumption of normality of the phenomena, so it is advised (see statistical packages such as SAS, Statistica, Statgraphics) to check normality first, for example with the Shapiro–Wilk W test. If normality is rejected, tests other than t are recommended (e.g. the sign test). So, the procedure of testing the hypothesis $H_0\colon \mu = \mu_0$ becomes a little complicated, and should be conducted in two steps:

1. check normality with the W test,

2. if normality is not rejected then use the t test else use the sign test.

In this paper we propose a modification of the Shapiro–Wilk W test, dedicated to checking normality with known mean value μ_0, i.e. to testing the hypothesis $H_0\colon X \sim N\left(\mu_0, \sigma^2\right)$, where X is the random variable of interest. This test could have very wide applications. For example, when we apply the paired t-test, the differences are assumed to be normally distributed with a given mean value $\mu_0 = \mu_1 - \mu_2$. The other application can be measurement errors which should be distributed as $N(0, \sigma^2)$, i.e. a measurement should be unbiased and normally distributed. Also, dimensions or weight of manufactured products should be normally distributed with given mean value. Another application is in the analysis of linear models, where one has to verify that residuals are normally distributed with null mean.

The modification of the W test and its properties are described in Section 2. The simulation results on its power are given in Section 3. Some concluding remarks are given in Section 4.

2. DERIVATION OF THE W_0 STATISTIC AND ITS PROPERTIES

Suppose that a random variable X is observed and we are interested in testing the hypothesis

$$H_0\colon\ X \sim N(\mu, \sigma^2)\ .$$

Shapiro and Wilk ([12]) proposed the W test based on the statistic

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} , \tag{2.1}$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the ordered values of a sample $X_1, X_2, ..., X_n$, and a_i are tabulated coefficients. A lower tail of W indicates nonnormality.

Now, let us assume that the expected value μ, say μ_0, is known. Thus it is of interest to test the null hypothesis

$$H_0 \colon X \sim N(\mu_0, \sigma^2) . \tag{2.2}$$

Application of Shapiro and Wilk's technique to the problem of testing (2.2) gives the statistic

$$W_0 = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} \left(X_i - \mu_0\right)^2} .$$

The null hypothesis (2.2) is rejected when $W_0 < W_0(\alpha, n)$, where $W_0(\alpha, n)$ is the critical value at significance level α.

The statistic W_0 has properties similar to the W statistic, namely, W_0 is scale invariant and the maximum value of W_0 is one. As it is known, the minimum value of W is $\varepsilon = \dfrac{n\, a_1^2}{n-1}$ ([12]).

Lemma 2.1. *The minimum value of W_0 is zero.*

Proof: Since W_0 is scale invariant it suffices to consider the maximization of $\sum_{i=1}^{n} \left(X_i - \mu_0\right)^2$ subject to $\sum_{i=1}^{n} a_i X_{(i)} = 1$. The lemma follows from the fact that $\sum_{i=1}^{n} \left(X_i - \mu_0\right)^2$ may be arbitrarily large. □

Shapiro and Wilk ([12]) gave an analytic form of the probability density function for the W statistic in the case of sample size $n = 3$. It is

$$g(w) = \frac{3}{\pi} \left(1 - w\right)^{-\frac{1}{2}} w^{-\frac{1}{2}} \qquad \text{for} \quad \frac{3}{4} \leq w \leq 1 . \tag{2.3}$$

They also establish that W is statistically independent of $\bar{X}$ and of $\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$ for samples from a normal distribution.

Thus, it is easy to obtain the probability density function of W_0 for samples of size $n = 3$. Let us notice that $W_0 = W \cdot C$, where

$$C = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{\sum_{i=1}^{n} \left(X_i - \mu_0\right)^2} = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu_0\right)^2}$$

is a random variable distributed as $Beta\left(\frac{n-1}{2}, \frac{1}{2}\right)$, independent of W. Thus in the case of $n = 3$, under H_0, we have the probability density function of C, namely,

$$f(c) = \frac{1}{2}\left(1 - c\right)^{-\frac{1}{2}} \qquad \text{for} \quad 0 \le c \le 1 \ .$$

Taking the new variable $W_0 = W \cdot C$ in the joint probability density function $g(w)\, f(c)$ and integrating this function over c, we get the probability density function for W_0 in the following form

$$\varphi(w_0) = \begin{cases} \dfrac{3}{2\pi} \cdot w_0^{-\frac{1}{2}} \cdot \displaystyle\int_{w_0}^{\frac{4}{3}w_0} (1-c)^{-\frac{1}{2}} \left(c - w_0\right)^{-\frac{1}{2}} \, dc & \text{for} \quad 0 \le w_0 \le \frac{3}{4} \ , \\ \dfrac{3}{2\pi} \cdot w_0^{-\frac{1}{2}} \cdot \displaystyle\int_{w_0}^{1} (1-c)^{-\frac{1}{2}} \left(c - w_0\right)^{-\frac{1}{2}} \, dc & \text{for} \quad \frac{3}{4} \le w_0 \le 1 \ . \end{cases}$$

Finally, after integrating, we get

$$\varphi\left(w_0\right) = \begin{cases} \dfrac{3}{2\pi} \cdot w_0^{-\frac{1}{2}} \cdot \left(\arcsin \dfrac{5\,w_0 - 3}{3\left(1 - w_0\right)} + \dfrac{\pi}{2}\right) & \text{for} \quad 0 \le w_0 \le \frac{3}{4} \ , \\ \dfrac{3}{2} \cdot w_0^{-\frac{1}{2}} & \text{for} \quad \frac{3}{4} \le w_0 \le 1 \ . \end{cases}$$

The plot of $\varphi\left(w_0\right)$ is shown in Figure 1.

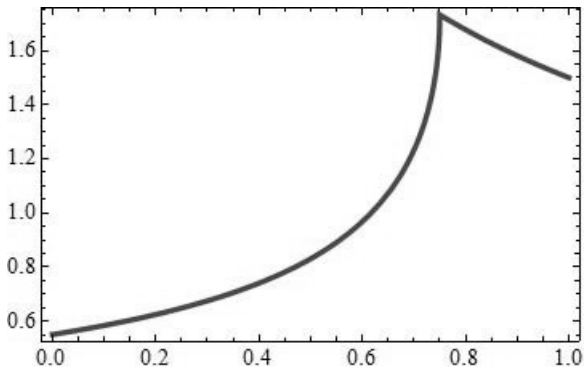


Figure 1: Plot of probability density function of W_0 for $n = 3$.

For sample size $n > 3$ the analytical form of the null distribution of W_0 is not available. Hence, to obtain any information about the distribution, Monte Carlo simulations were performed. In simulations, for each $n = 3, 4, ..., 50$, $N =$ 1,000,000 samples from the distribution $N(0, 1)$ were drawn and for each sample the value of W_0 was calculated, thus the sample $w_1, w_2, ..., w_N$ of values of the W_0 statistic was obtained. The critical value $W_0(\alpha, n)$ was taken as the α-th quantile of $w_1, w_2, ..., w_N$. All calculations were done in the R program ([9]) using the procedure *shapiro.test* in which Royston's procedure is used ([10]). The same calculations were also done independently in Mathematica. The results are given in Table 1.

Table 1: Critical values of W_0 statistic for sample sizes n and significance level α.

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
3	0.0184	0.0881	0.1714	27	0.7379	0.8232	0.8601
4	0.0721	0.2037	0.3127	28	0.7463	0.8287	0.8645
5	0.1419	0.3086	0.4190	29	0.7539	0.8340	0.8688
6	0.2090	0.3867	0.4952	30	0.7611	0.8394	0.8730
7	0.2742	0.4525	0.5543	31	0.7677	0.8437	0.8765
8	0.3299	0.5051	0.5998	32	0.7746	0.8482	0.8800
9	0.3785	0.5493	0.6374	33	0.7804	0.8524	0.8834
10	0.4233	0.5852	0.6682	34	0.7871	0.8565	0.8863
11	0.4606	0.6165	0.6935	35	0.7917	0.8602	0.8894
12	0.4940	0.6431	0.7154	36	0.7969	0.8634	0.8921
13	0.5246	0.6661	0.7346	37	0.8008	0.8670	0.8947
14	0.5494	0.6862	0.7504	38	0.8063	0.8701	0.8972
15	0.5739	0.7038	0.7651	39	0.8109	0.8731	0.8996
16	0.5954	0.7196	0.7778	40	0.8145	0.8760	0.9018
17	0.6126	0.7337	0.7890	41	0.8194	0.8787	0.9040
18	0.6319	0.7476	0.7998	42	0.8227	0.8816	0.9061
19	0.6478	0.7590	0.8088	43	0.8271	0.8839	0.9081
20	0.6626	0.7696	0.8176	44	0.8301	0.8862	0.9100
21	0.6761	0.7792	0.8250	45	0.8343	0.8887	0.9120
22	0.6876	0.7875	0.8319	46	0.8374	0.8911	0.9138
23	0.7008	0.7965	0.8390	47	0.8403	0.8931	0.9154
24	0.7104	0.8034	0.8446	48	0.8433	0.8951	0.9169
25	0.7205	0.8103	0.8501	49	0.8470	0.8974	0.9187
26	0.7296	0.8170	0.8553	50	0.8491	0.8989	0.9200

Shapiro and Wilk ([13]) approximated the distribution of the W statistic by a Johnson curve. For each n they made the least squares regression of the empirical sampling value of

$$u(p) = \ln \frac{W(p) - \varepsilon}{1 - W(p)}$$

on the p-th quantile of the standard normal distribution z_p, where ε was the minimum value of the W statistic and $W(p)$ was the p-th empirical sampling quantile. They took the following values of p:

$$p = 0.01,\ 0.02,\ 0.05\ (0.05)\ 0.25,\ 0.5,\ 0.75\ (0.05)\ 0.95,\ 0.98,\ 0.99\ ,$$

and gave the tables for ε, γ and δ such that $z = \gamma + \delta \ln \dfrac{W - \varepsilon}{1 - W}$ has approximately standard normal distribution.

In our study, a similar approach was applied for the W_0 statistic for sample sizes $n = 3, 4, ..., 50$. As $\varepsilon = 0$ (see Lemma 2.1), the least squares regression of $\ln \dfrac{W_0(p)}{1 - W_0(p)}$ on z_p was based on 1,000,000 pseudorandom samples from $N(0, 1)$. The values of δ and γ, such that $Z = \gamma + \delta \ln \dfrac{W_0}{1 - W_0}$ has approximately standard normal distribution are listed in Table 2. The lower tail of Z's indicates nonnormality.

Table 2: The normalizing constants for W_0 for sample sizes n.

n	γ	δ	n	γ	δ	n	γ	δ
3	−0.3137	0.5551	19	−3.2563	1.3698	35	−4.4593	1.5241
4	−0.6479	0.7282	20	−3.3584	1.3847	36	−4.5088	1.5272
5	−0.9586	0.8510	21	−3.4511	1.3983	37	−4.5621	1.5336
6	−1.2299	0.9384	22	−3.5365	1.4095	38	−4.6152	1.5382
7	−1.4778	1.0092	23	−3.6320	1.4236	39	−4.6749	1.5467
8	−1.6950	1.0671	24	−3.7067	1.4319	40	−4.7186	1.5495
9	−1.8960	1.1157	25	−3.7869	1.4431	41	−4.7771	1.5574
10	−2.0790	1.1573	26	−3.8624	1.4520	42	−4.8195	1.5597
11	−2.2470	1.1929	27	−3.9346	1.4606	43	−4.8711	1.5659
12	−2.4039	1.2238	28	−4.0077	1.4703	44	−4.9137	1.5693
13	−2.5513	1.2517	29	−4.0770	1.4783	45	−4.9706	1.5769
14	−2.6821	1.2755	30	−4.1538	1.4891	46	−5.0118	1.5797
15	−2.8104	1.2979	31	−4.2084	1.4935	47	−5.0512	1.5826
16	−2.9320	1.3181	32	−4.2782	1.5030	48	−5.0908	1.5858
17	−3.0400	1.3350	33	−4.3354	1.5086	49	−5.1470	1.5935
18	−3.1553	1.3542	34	−4.4017	1.5172	50	−5.1795	1.5954

To check the goodness of approximation, another $N = 1{,}000{,}000$ pseudorandom samples from $N(0, 1)$ were generated. For each of them W_{0i} and $Z_i = \gamma + \delta \ln \dfrac{W_{0i}}{1 - W_{0i}}$ were calculated $(i = 1, 2, ..., N)$. The ratios $\dfrac{\#\{Z_i : Z_i \le z_p\}}{N}$ with $p = 0.01,\ 0.02,\ 0.05,\ 0.1,\ 0.5,\ 0.9,\ 0.95,\ 0.98,\ 0.99$ are given in Table 3.

Table 3: The simulated probabilities $P\left(\gamma + \delta \ln \frac{W_0}{1-W_0} \leq z_p\right)$ for sample sizes n.

n	Probability								
	0.01	0.02	0.05	0.10	0.5	0.90	0.95	0.98	0.99
3	0.015	0.023	0.047	0.06	0.458	0.919	0.957	0.979	0.987
4	0.014	0.024	0.049	0.091	0.453	0.912	0.957	0.981	0.989
5	0.014	0.024	0.051	0.094	0.453	0.908	0.955	0.982	0.990
6	0.013	0.024	0.051	0.095	0.454	0.906	0.956	0.983	0.991
7	0.013	0.024	0.052	0.096	0.456	0.905	0.956	0.983	0.991
8	0.013	0.024	0.053	0.097	0.457	0.903	0.955	0.983	0.992
9	0.013	0.024	0.052	0.097	0.457	0.902	0.955	0.983	0.992
10	0.013	0.024	0.053	0.098	0.457	0.90	0.955	0.983	0.992
11	0.013	0.024	0.054	0.099	0.456	0.900	0.954	0.983	0.992
12	0.013	0.024	0.054	0.099	0.458	0.900	0.954	0.984	0.992
13	0.013	0.024	0.054	0.100	0.459	0.900	0.954	0.984	0.993
14	0.013	0.024	0.054	0.099	0.458	0.899	0.954	0.984	0.992
15	0.013	0.024	0.053	0.099	0.456	0.898	0.954	0.984	0.993
16	0.013	0.024	0.054	0.100	0.458	0.899	0.954	0.984	0.993
17	0.013	0.024	0.054	0.099	0.457	0.898	0.954	0.984	0.993
18	0.013	0.024	0.054	0.099	0.457	0.897	0.953	0.984	0.993
19	0.013	0.024	0.054	0.100	0.457	0.897	0.953	0.984	0.993
20	0.013	0.024	0.054	0.100	0.458	0.897	0.953	0.984	0.993
21	0.013	0.024	0.054	0.100	0.457	0.897	0.953	0.984	0.993
22	0.013	0.024	0.054	0.100	0.457	0.897	0.953	0.984	0.993
23	0.013	0.024	0.055	0.100	0.458	0.897	0.953	0.984	0.993
24	0.013	0.024	0.054	0.101	0.459	0.897	0.954	0.984	0.993
25	0.013	0.024	0.054	0.100	0.458	0.897	0.954	0.984	0.993
26	0.013	0.024	0.054	0.100	0.459	0.898	0.953	0.984	0.993
27	0.013	0.024	0.055	0.100	0.458	0.897	0.954	0.985	0.993
28	0.013	0.024	0.054	0.100	0.457	0.897	0.953	0.984	0.993
29	0.013	0.024	0.054	0.100	0.458	0.898	0.954	0.984	0.993
30	0.013	0.024	0.055	0.101	0.458	0.897	0.953	0.984	0.993
31	0.013	0.024	0.054	0.100	0.458	0.897	0.954	0.985	0.993
32	0.013	0.024	0.055	0.101	0.459	0.900	0.953	0.984	0.993
33	0.013	0.024	0.054	0.100	0.457	0.897	0.953	0.984	0.993
34	0.013	0.024	0.055	0.101	0.459	0.897	0.954	0.984	0.993
35	0.013	0.024	0.054	0.101	0.458	0.896	0.953	0.984	0.993
36	0.013	0.024	0.054	0.100	0.458	0.897	0.954	0.985	0.993
37	0.013	0.024	0.054	0.100	0.457	0.896	0.953	0.984	0.993
38	0.013	0.024	0.054	0.101	0.458	0.897	0.953	0.985	0.993
39	0.013	0.024	0.054	0.100	0.456	0.896	0.953	0.984	0.993
40	0.013	0.024	0.055	0.100	0.457	0.897	0.954	0.985	0.993
41	0.013	0.024	0.054	0.100	0.456	0.896	0.953	0.984	0.993
42	0.013	0.024	0.055	0.101	0.458	0.897	0.954	0.985	0.993
43	0.013	0.024	0.054	0.100	0.457	0.896	0.954	0.985	0.993
44	0.013	0.024	0.054	0.099	0.456	0.896	0.953	0.984	0.993
45	0.013	0.024	0.055	0.100	0.457	0.896	0.953	0.984	0.993
46	0.013	0.024	0.055	0.100	0.458	0.897	0.954	0.984	0.993
47	0.013	0.024	0.055	0.101	0.458	0.897	0.954	0.985	0.994
48	0.013	0.024	0.054	0.100	0.457	0.897	0.954	0.985	0.994
49	0.013	0.025	0.055	0.101	0.458	0.896	0.954	0.985	0.994
50	0.013	0.024	0.054	0.100	0.457	0.896	0.953	0.984	0.993

3. POWER COMPARISONS

Suppose that the hypothesis H_0: $X \sim N(\mu_0, \sigma^2)$ is verified using the W_0 test. Three kinds of alternative hypothesis are considered:

a) $X \sim N\left(\mu, \sigma^2\right)$ with $\mu \neq \mu_0$;

b) X is not normal with $\mu = \mu_0$;

c) X is not normal with $\mu \neq \mu_0$.

We focus on the power of the W_0 test. The Shapiro–Wilk W test was investigated against different nonnormal alternatives. Very exhaustive research was done by Shapiro *et al.* ([14]) and Chen ([3]). It was showed that the W test is very powerful in comparison to other normality tests such as Kolmogorov, chi-square, β_1, β_2 and against very different distributions including Student's t, Gamma, Beta or Uniform.

As the construction of W_0 is similar to the W test, it may be expected that the W_0 test will also be powerful against alternatives of kind b) and c). Hence, in our study we confine ourselves to the a) alternative, i.e. when the true distribution is normal with a mean other than μ_0. The W_0 test is compared with two other procedures. The first one is the Kolmogorov test (modified to the case of known mean). The test statistic of the Kolmogorov test is given by

$$\max_{1\leq i\leq n}\left\{\left|F\big(X_{(i)}\big)-\frac{i-1}{n}\right|,\ \left|F\big(X_{(i)}\big)-\frac{i}{n}\right|\right\},$$

where $F\big(X_{(i)}\big)=\Phi\left(\frac{X_{(i)}-\mu_0}{S}\right)$, $S=\dfrac{1}{n}\sqrt{\sum\limits_{i=1}^{n}\left(X_i-\mu_0\right)^2}$ and Φ is the CDF of the standard normal distribution.

The second procedure, denoted by $W+t$, is a two-step one. In the first step the normality is verified by the classical W test. If normality is not rejected, then the hypothesis of equality of the mean to a given number μ_0 is verified by the t test.

All three tests are calculated at the significance level α. In the case of the $W+t$ test we need to apply two significance levels α_W and α_t for both tests. Those numbers were chosen in such a way that the overall significance level is α, i.e.

$$P_{H_0}\Big\{W \text{ accepts normality and } t \text{ accepts mean } \mu_0\Big\} \geq 1-(\alpha_W+\alpha_t) = 1-\alpha\ .$$

Because there are no preferences to the W or t test, $\alpha_W=\alpha_t=\frac{\alpha}{2}$ were taken. The power comparison of the three tests was performed by the Monte Carlo method. A sample of size n from the standard normal distribution was generated and this sample was used in all tests. The sample was then shifted to different values of μ and then each of the tests was applied to the shifted sample.

This procedure was repeated 10,000 times. The number of rejections of the hypothesis $H_0 : X \sim N(\mu_0, \sigma^2)$ was calculated. It should be added that critical points for Kolmogorov test were also determined by Monte Carlo method. In the simulations the hypothesis H_0 was verified for samples of sizes $10, 20, 30, 40, 50$ and significance levels $\alpha = 0.01, 0.05, 0.1$. The simulated powers are given in Table 4.

Table 4: Power of W_0, Kolmogorov and $W + t$ tests.

n	μ	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
		W_0	K	$W+t$	W_0	K	$W+t$	W_0	K	$W+t$
10	0.0	0.011	0.010	0.010	0.050	0.049	0.048	0.101	0.100	0.098
	0.3	0.034	0.030	0.023	0.129	0.112	0.097	0.216	0.198	0.176
	0.6	0.148	0.115	0.097	0.386	0.326	0.287	0.526	0.464	0.422
	0.9	0.392	0.310	0.287	0.705	0.610	0.585	0.829	0.746	0.727
	1.2	0.687	0.565	0.560	0.916	0.845	0.847	0.964	0.925	0.924
	1.5	0.892	0.789	0.813	0.987	0.957	0.965	0.996	0.985	0.989
	1.8	0.974	0.920	0.943	0.998	0.991	0.995	1.000	0.998	0.999
	2.1	0.996	0.978	0.990	1.000	0.999	1.000	1.000	1.000	1.000
	2.4	1.000	0.993	0.997	1.000	1.000	1.000	1.000	1.000	1.000
20	0.0	0.011	0.011	0.010	0.049	0.050	0.051	0.096	0.099	0.096
	0.2	0.036	0.031	0.027	0.131	0.113	0.106	0.214	0.189	0.179
	0.4	0.165	0.123	0.121	0.383	0.313	0.297	0.515	0.442	0.424
	0.6	0.435	0.329	0.344	0.709	0.602	0.610	0.808	0.732	0.733
	0.8	0.741	0.601	0.656	0.916	0.842	0.861	0.960	0.913	0.955
	1.0	0.924	0.835	0.879	0.988	0.959	0.976	0.995	0.984	0.991
	1.2	0.988	0.950	0.978	0.999	0.994	0.997	1.000	0.998	0.999
	1.4	0.999	0.991	0.997	1.000	1.000	1.000	0.999	0.991	1.000
	1.6	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.999	1.000
	1.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
30	0.00	0.012	0.009	0.012	0.054	0.053	0.051	0.105	0.103	0.097
	0.15	0.039	0.030	0.028	0.124	0.107	0.099	0.201	0.182	0.165
	0.30	0.150	0.111	0.106	0.350	0.285	0.267	0.478	0.408	0.388
	0.45	0.398	0.290	0.312	0.657	0.550	0.565	0.762	0.679	0.683
	0.60	0.693	0.540	0.608	0.876	0.793	0.817	0.930	0.878	0.889
	0.75	0.891	0.777	0.844	0.972	0.932	0.952	0.987	0.968	0.977
	0.90	0.975	0.922	0.959	0.996	0.985	0.992	0.999	0.994	0.997
	1.05	0.997	0.981	0.994	1.000	0.998	1.000	1.000	0.999	1.000
	1.20	1.000	0.997	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	1.35	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.00	0.011	0.009	0.010	0.051	0.051	0.050	0.098	0.101	0.101
	0.15	0.042	0.034	0.033	0.144	0.125	0.114	0.236	0.206	0.196
	0.30	0.213	0.152	0.159	0.440	0.364	0.357	0.565	0.486	0.486
	0.45	0.539	0.402	0.459	0.776	0.672	0.702	0.861	0.789	0.802
	0.60	0.837	0.706	0.785	0.953	0.900	0.924	0.978	0.951	0.963
	0.75	0.972	0.910	0.955	0.995	0.981	0.991	0.998	0.993	0.996
	0.90	0.997	0.981	0.995	1.000	0.998	1.000	1.000	1.000	1.000
	1.05	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	0.00	0.010	0.008	0.011	0.050	0.050	0.050	0.099	0.100	0.099
	0.15	0.064	0.048	0.044	0.179	0.150	0.141	0.272	0.237	0.225
	0.30	0.301	0.212	0.224	0.538	0.444	0.452	0.661	0.579	0.578
	0.45	0.691	0.542	0.609	0.872	0.784	0.817	0.927	0.872	0.891
	0.60	0.937	0.840	0.905	0.983	0.956	0.973	0.993	0.978	0.986
	0.75	0.994	0.968	0.988	0.999	0.996	0.999	1.000	0.999	1.000
	0.90	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The relative powers of W_0 with respect to the Kolmogorov and $W+t$ tests are shown in Figure 2. On the x-axis there are values of $\mu \geq 0$ and on the y-axis there are values of

$$\frac{\text{power of } W_0 \text{ test}}{\text{power of Kolmogorov test}} \text{ (solid line)} \quad \text{and} \quad \frac{\text{power of } W_0 \text{ test}}{\text{power of } W+t \text{ test}} \text{ (dashed line)}.$$

One can see that generally the lines are above 1, which shows that W_0 is more powerful than the other two tests.

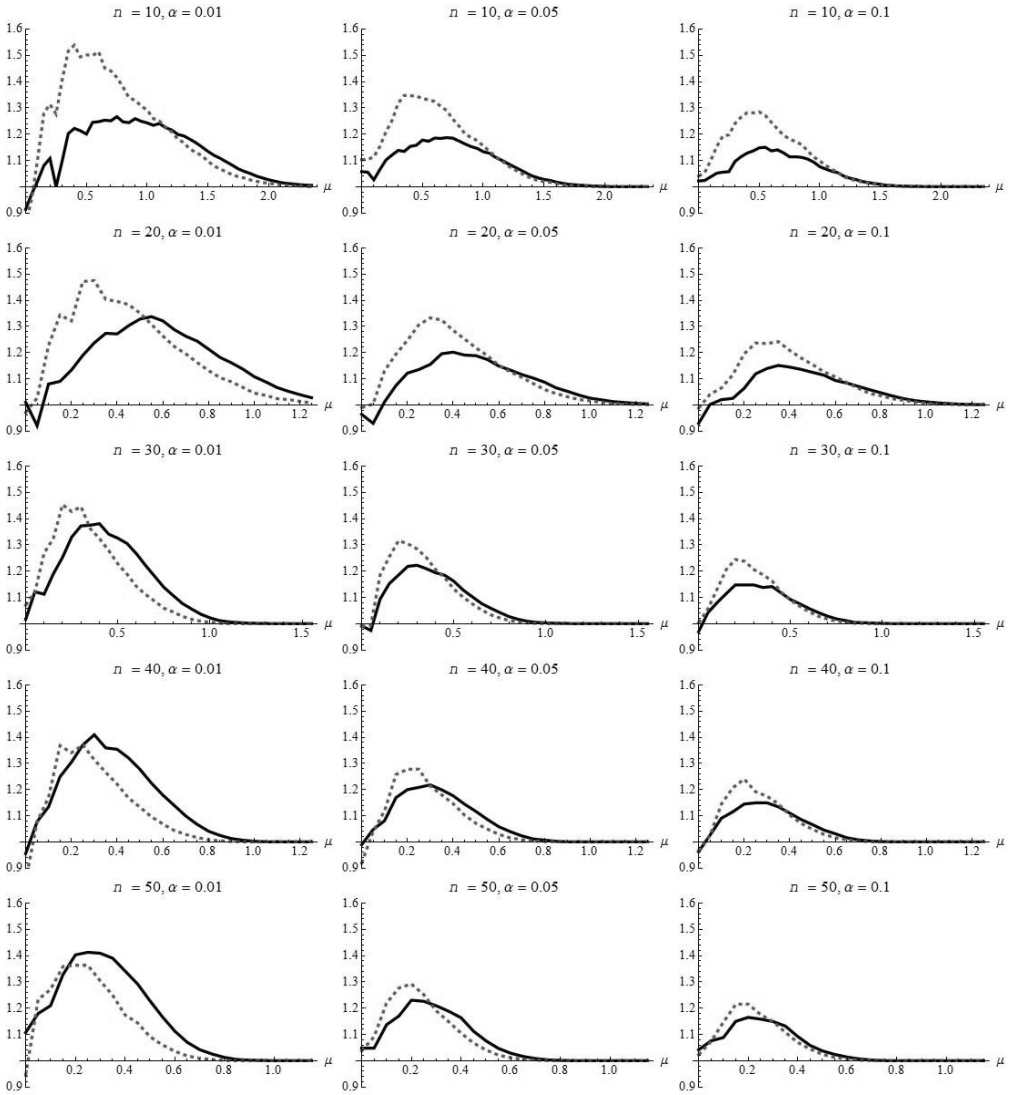


Figure 2: Relative power of W_0 with respect to Kolmogorov and $W+t$ tests.

4. CONCLUDING REMARKS

In many statistical models it is assumed that random variables are normally distributed with known mean. Thus the W_0 test is more adequate and should be used instead of the classical Shapiro–Wilk W test.

In the paper it is shown via a simulation study that the W_0 test is generally more powerful than the Kolmogorov, and W and Student t tests combined.

REFERENCES

[1] Bartoszynski, R. and Niewiadomska-Bugaj, M. (1996). *Probability and Statistical Inference*, Wiley Series in Probability and Statistics, John Wiley and Sons, New York.

[2] Bickel, P.J. and Doksum, K.A. (1980). *Mathematical Statistics*, Holden-Day, Inc., San Francisco.

[3] Chen, E.H. (1971). The power of the Shapiro–Wilk W test for normality in samples from contaminated normal distributions, *Journal of the American Statistical Association*, **66**, 760–762.

[4] Chow, G.C. (1983). *Econometrics*, Mc-Graw-Hill Book Company, New York.

[5] Greene, W.H. (2000). *Econometric Analysis*, Prentice Hall Inc., New Jersey.

[6] Knight, K. (2000). *Mathematical Statistics*, Chapman and Hall/CRC.

[7] Petrie, A. and Sabin, C. (2009). *Medical Statistics at a Glance*, Hoboken, Wiley-Blackwell, New Jersey.

[8] Quinn, G.P. and Keough, M.J. (2002). *Experimental Design and Data Analysis for Biologists*, Cambridge University Press.

[9] R Development Core Team (2008). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, `http://www.R-project.org`.

[10] Royston, P. (1992). Approximating the Shapiro–Wilk W-test for non-normality, *Statistics and Computing*, **2**, 117–119.

[11] Ryan, T.P. (2007). *Modern Engineering Statistics*, Wiley-Interscience, Hoboken, New Jersey.

[12] Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611.

[13] Shapiro, S.S. and Wilk, M.B. (1968). Approximations for the null distribution of the W statistic, *Technometrics*, **10**, 861–866.

[14] Shapiro, S.S.; Wilk, M.B. and Chen, H.J. (1968). A comparative study of various tests for normality, *Journal of the American Statistical Association*, **63**, 1343–1372.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in February, one in April, one in June and the other in October.

— On average, five articles will be published per issue

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.