

On a new NBUE property in multivariate sense: an application

*Original*

On a new NBUE property in multivariate sense: an application / Fernandez Ponce, J. M.; Pellerrey, Franco; Rodriguez Grinolo, R.. - In: COMPUTATIONAL STATISTICS & DATA ANALYSIS. - ISSN 0167-9473. - STAMPA. - 55:12(2011), pp. 3283-3294. [10.1016/j.csda.2011.06.010]

*Availability:*

This version is available at: 11583/2424776 since: 2015-12-18T10:04:30Z

*Publisher:*

Elsevier

*Published*

DOI:10.1016/j.csda.2011.06.010

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

# On a new NBUE property in multivariate sense: an application.

## Author's version

*boxed* Published in Computational Statistics and Data Analysis 55 (2011) 32833294  
doi:10.1016/j.csda.2011.06.010  
<http://www.sciencedirect.com/science/article/pii/S0167947311002131>

J.M. Fernández-Ponce<sup>1,\*</sup>

*Departamento de Estadística e I.O.  
C/ Tarfia s/n  
Universidad de Sevilla  
41012-Sevilla, SPAIN*

F. Pellerey\*

*Dipartimento di Matematica  
Politecnico di Torino  
Corso Duca degli Abruzzi 24  
I-10129 Torino, ITALY*

M.R. Rodríguez-Griñolo,\*

*Departamento Economía, Métodos Cuantitativos e Historia Económica  
Universidad Pablo de Olavide  
Ctra. de Utrera, Km 1  
41013 Sevilla, SPAIN.*

---

## Abstract

Multivariate lifetime data frequently arise so it is important to consider different multivariate distributions that could be used to model aging concepts. The generalization of univariate versions has two main problems. One of them is the concept of the multivariate quantile and the other one is that the survival function valued at a multivariate quantile depends on the underlying distributions. For this reason, a new multivariate aging concept based on the upper corrected orthant is developed. This new multivariate aging is a generalization of the *New Better than Used in Expectation* univariate version and a characterization by using the multivariate excess wealth function is provided. Finally, data from the Comprehensive Cohort Study performed by the German Breast Cancer Study Group are analysed in order to estimate the parameters in a non-deterministic model which describes the tumor growth and to realize how this new multivariate ageing notion influences in the final results.

*Keywords:* Multivariate aging, Excess wealth function, Multivariate  $\mathbf{u}$ -quantiles, Upper corrected orthants, Tumor growth.

---

\*Corresponding author. Tel: +34 954557946; Fax: +34 954622800.

## 1. Introduction

Roughly speaking, reliability theory is concerned with determining the probability that a system, possibly consisting of many components, will work. Therefore, the problem of determining the reliability of a system is essentially the problem of determining the life distribution of the system. To do this, the usual statistical method is to fit a life distribution function based on a set of the observed lifetime data. On the other hand, because of the aging effect, most systems degenerate or improve overtime so that they may have life distributions with some aging property. Multivariate lifetime data frequently arise, and in these situations it is important to consider different multivariate distributions that could be used to model such aging concept. Some multivariate versions of aging have widely been studied in the literature (see Block and Savits, 1981, Johnson and Kotz, 1975 and Marshall and Shaked, 1986 among others). Shaked and Shanthikumar (1991) provided new extensions, not just technical but from a dynamic point of view. In this paper, our starting point is different in this topic since the classical concept of quadrant is not used to give a new notion of aging.

The organization of the paper is as follows. In Section 2 the multivariate excess wealth concept and some important properties obtained of this concept are developed. In Section 3, the multivariate New Better than Used in Expectation (NBUE) concept is defined and characterized throughout the excess wealth function. The model of tumor growth and an application to real data set is shown in Section 4.

Some notation is given here which will be used throughout the paper. Fundamentally, random vectors will be dealt with which take on values in  $\mathbb{R}$ . The space  $\mathbb{R}^n$  is endowed with the usual componentwise partial order, which is defined as follows. Let  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$  be two vectors in  $\mathbb{R}^n$ ; and therefore  $\mathbf{x} \leq \mathbf{y}$  if  $x_i \leq y_i$  for  $i = 1, \dots, n$ . Throughout the paper “increasing” means “non-decreasing” and “decreasing” means “non-increasing”. Particularly, a function  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is said to be an increasing function when  $\phi(\mathbf{x}) \leq \phi(\mathbf{y})$  for  $\mathbf{x} \leq \mathbf{y}$ . The notation  $\sim_{st}$  stands for equality in law. The vector of ones will be denoted by  $\mathbf{1}$ , i.e.  $\mathbf{1} = (1, \dots, 1)$  and the corresponding of zeros by  $\mathbf{0}$ . And the dimension of  $\mathbf{1}$  will be clear from the expression in which it appears. The multiple integral  $\int_A F(t_1, \dots, t_n) dt_1 \dots dt_n$  will be denoted as  $\int_A F(\mathbf{t}) d\mathbf{t}$ . The dimension of a random vector is clear from the context and unless otherwise stated it is assumed that it is  $n$ .

## 2. The multivariate excess wealth concept

Let  $\mathbf{X}$  be a random vector in  $\mathbb{R}^n$  with distribution function (cdf)  $F(\cdot)$ . Let  $\mathbf{u}_n = (u_1, \dots, u_n)$  be a vector in  $[0, 1]^n$ , the multivariate  $\mathbf{u}$ -quantile or the regression representation for  $\mathbf{X}$ , denoted by  $\hat{\mathbf{x}}(\mathbf{u}_n)$ , is defined as

$$\hat{x}_1(u_1) = F_{X_1}^-(u_1), \quad \hat{x}_2(\mathbf{u}_2) = F_{X_2|X_1=\hat{x}_1(u_1)}^-(u_2), \dots, \quad \hat{x}_n(\mathbf{u}_n) = F_{X_n|\mathbf{X}_{n-1}=\hat{\mathbf{x}}(\mathbf{u}_{n-1})}^-(u_n),$$

where  $F^-(u) = \inf\{x : F(x) \geq u\}$ ,  $\mathbf{u}_i = (u_1, \dots, u_i)$  and  $\mathbf{X}_i = (X_1, \dots, X_i)$  for all  $i = 1, \dots, n$ .

The interest in such representation is in simulation theory as well as in queue theory and other areas. In particular, O'Brien (1975) studied a comparison method for stochastic processes by using the regression representation and Arjas and Lehtonen (1978) used it to approximate many server queues by means of single server queues. Later, Rüschemdorf (1981) ordered distributions in stochastic sense using the multivariate  $\mathbf{u}$ -quantile.

More recently, Fernández-Ponce and Suárez-LLoréns (2003) also defined several concepts for a multivariate random vector. The first concept is the multivariate  $\mathbf{x}$ -rate vector, denoted by  $\mathbf{x}^*(\mathbf{x})$  and defined as  $\mathbf{x}^*(x_1) = P(X_1 \leq x_1), \dots, \mathbf{x}^*(x_n) = P(X_n \leq x_n | \bigcap_{j=1}^{n-1} X_j = x_j)$ . The second concept is the right-upper

---

*E-mail addresses:* ferpon@us.es

orthant at a point  $\mathbf{z}$ , defined as  $C(\mathbf{z}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{z} \leq \mathbf{x}\}$ . At the end, the upper-corrected orthant at a point  $\mathbf{z}$  for the random variable  $\mathbf{X}$ , denoted as  $R_{\mathbf{X}}(\mathbf{z})$ , is defined as

$$R_{\mathbf{X}}(\mathbf{z}) = \{\mathbf{x} \in \mathbb{R}^n : x_1 \geq F_{X_1}^-(x_1^*(z_1)), \dots, x_n \geq F_{X_n | \bigcap_{j=1}^{n-1} X_j=x_j}^-(x_n^*(z_n))\}.$$

In fact, Fernández-Ponce and Suárez-Lloréns (2003) did not consider vectors of non-negative variables having limited supports on the right. Since vectors of lifetimes with limited supports will also be considered in this paper, a slight modification of this notion must be defined as a generalization of the previous one. By taking into account that the support of a random vector  $\mathbf{X}$  is defined as  $Supp(\mathbf{X}) = \{\mathbf{x} \in \mathbb{R}^n : P[\mathbf{X} \in B_{\mathbf{x}}(\varepsilon)] > 0 \text{ for all } \varepsilon > 0\}$  where  $B_{\mathbf{x}}(\varepsilon)$  is the centered ball at  $\mathbf{x}$  with radius  $\varepsilon$ , then the following definition can be given.

**Definition 2.1.** *Let  $\mathbf{X}$  be a random vector. The upper corrected orthant associated to  $\mathbf{X}$  at  $\mathbf{z} \in Supp(\mathbf{X})$  is defined as*

$$R_{\mathbf{X}}(\mathbf{z}) = \{\mathbf{x} \in Supp(\mathbf{X}) : x_1 \geq F_{X_1}^-(x_1^*(z_1)), \dots, x_n \geq F_{X_n | \bigcap_{j=1}^{n-1} X_j=x_j}^-(x_n^*(z_n))\}.$$

It is easily shown that if  $\mathbf{X}$  is a random vector with independent components then  $R_{\mathbf{X}}(\mathbf{z}) = C(\mathbf{z}) \cap Supp(\mathbf{X})$ .

**Remark 2.1.** *Note that if  $\mathbf{t}_1 \leq \mathbf{t}_2$  is verified then it could not be held that  $R_{\mathbf{X}}(\mathbf{t}_2) \subset R_{\mathbf{X}}(\mathbf{t}_1)$ . For example, let  $\mathbf{X}$  be a bivariate random vector with joint density function given by*

$$f_{\mathbf{X}}(\mathbf{t}) = \begin{cases} 2/3 & \text{if } \mathbf{t} \in T_1; \\ 4/3 & \text{if } \mathbf{t} \in T_2; \\ 0 & \text{otherwise} \end{cases}$$

where  $T_1$  is in the triangle with vertices  $(0,0)$ ,  $(0,1)$  and  $(1,1)$ , and  $T_2$  is in the triangle with vertices  $(0,0)$ ,  $(1,0)$  and  $(1,1)$ .

Let  $\mathbf{t}_1 = (\frac{1}{2}, \frac{1}{2})$  be a point in a border of  $T_1$ . By straightforward computations, it is easily verified that

$$R_{\mathbf{X}}(\mathbf{t}_1) = \left\{ x \geq \frac{1}{2}, y \geq \frac{x+1}{3} \right\}.$$

Now, by considering  $\mathbf{t}_2 = (\frac{2}{3}, \frac{1}{2})$ , it is held that  $\mathbf{t}_1 \leq \mathbf{t}_2$ , and  $\mathbf{t}_2 \in R_{\mathbf{X}}(\mathbf{t}_2)$ . But,  $\mathbf{t}_2 \notin R_{\mathbf{X}}(\mathbf{t}_1)$ . Thus  $R_{\mathbf{X}}(\mathbf{t}_2) \not\subset R_{\mathbf{X}}(\mathbf{t}_1)$ .

From now on, throughout the paper, assume that the following regularity conditions (RC) are verified by every cdf  $F$ .

1.  $F$  is a continuous function.
2. The function  $\hat{\mathbf{x}} : \mathbf{u} \rightarrow \hat{\mathbf{x}}(\mathbf{u})$  is differentiable at each component.
3. The conditional distribution of  $X_i$  to  $X_1, \dots, X_{i-1}$  ( $F_{i|1, \dots, i-1}$ ) is a continuous and strictly increasing function for  $i = 1, \dots, n$ . For convenience  $F_{1|0} = F_1$ .
4.  $F_{X_i | X_1, \dots, X_{i-1}}^-(0) < \infty$  for all  $i = 1, \dots, n$ .

It is easy to verify that under the regularity conditions above there is a one to one correspondence between vectors  $\mathbf{u}_n \in [0, 1]^n$  and the points  $\mathbf{x} \in Supp(\mathbf{X})$ .

The next result is the main reason of interest in the notion of upper corrected orthant. The proof can be seen in Proposition 2.1 in Fernández-Ponce and Suárez-Lloréns (2003).

**Proposition 2.1.** *Let  $\mathbf{X}$  be a random vector. Then*

$$P\{\mathbf{X} \in R_{\mathbf{X}}[\hat{\mathbf{x}}(\mathbf{u}_n)]\} = \prod_{j=1}^n (1 - u_j) \text{ for all } \mathbf{u}_n \in [0, 1]^n.$$

This result means that the upper corrected orthant at the  $\mathbf{u}_n$ -quantile accumulate the same probability for two random vectors with distribution functions under the regularity conditions. By using this property, Fernández-Ponce and Suárez-LLoréns (2003) defined a new dispersive ordering based on conditional quantiles more widely separated.

Furthermore, let  $\mathbf{t} = (t_1, t_2)$  be a point in  $Supp(\mathbf{X})$ . Then it is easily shown that

$$\begin{aligned} P(\mathbf{X} \in R_{\mathbf{X}}(\mathbf{t})) &= \int_{t_1}^{+\infty} \int_{F_{2|w_1}^-(F_{2|t_1}(t_2))}^{+\infty} f(w_1, w_2) dw_2 dw_1 \\ &= \int_{t_1}^{+\infty} f_{X_1}(w_1) \int_{F_{2|w_1}^-(F_{2|t_1}(t_2))}^{+\infty} f_{X_2|X_1=w_1}(w_2) dw_2 dw_1. \end{aligned}$$

Consequently, it is obtained that

$$P(\mathbf{X} \in R_{\mathbf{X}}(\mathbf{t})) = P(X_1 > t_1)P(X_2 > t_2|X_1 = t_1). \quad (1)$$

This equality can easily be generalized for the  $n$ -dimensional case by induction arguments.

Note that, in particular, from Proposition 2.1 easily follows that  $R_{\mathbf{X}}[\hat{\mathbf{x}}(\mathbf{0})] = Supp(\mathbf{X})$ .

Now, the multivariate excess wealth function is defined by generalizing the univariate version. Recall that the univariate excess wealth function of a random variable  $X$ , which was defined and widely studied by Fernández-Ponce *et al.* (1996), is defined as

$$S_X^+(u) = \mathbf{E}[(X - F_X^-(u))^+] = \mathbf{E}[\max\{X - F_X^-(u), 0\}] = \int_{F_X^-(u)}^{\infty} \bar{F}_X(t) dt \text{ for } u \in [0, 1].$$

**Definition 2.2.** Let  $\mathbf{X}$  be a nonnegative random vector. The multivariate excess wealth function associated to  $\mathbf{X}$ , when it exists, is defined as

$$S_{\mathbf{X}}^+(\mathbf{u}) = \int_{R_{\mathbf{X}}[\hat{\mathbf{x}}(\mathbf{u})]} P[\mathbf{X} \in R_{\mathbf{X}}(\mathbf{t})] dt \text{ for all } \mathbf{u} \in [0, 1]^n. \quad (2)$$

Note that if  $U$  is a uniform  $[0, 1]$  random variable then  $F_X^-(U) =_{st} X$ . In multivariate settings, this property is generalized by the regression representation (see Shaked and Shanthikumar, 2007, pag 267). That is,  $\hat{\mathbf{x}}(\mathbf{U}) =_{st} \mathbf{X}$  where  $\mathbf{U} = (U_1, \dots, U_n)$  with  $U_1, \dots, U_n$  are independent uniform  $[0, 1]$  random variables. Consequently,  $\hat{\mathbf{x}}(\mathbf{u})$  plays the same role in the multivariate case as  $F_X^-(u)$  in the univariate case and  $R_{\mathbf{X}}[\hat{\mathbf{x}}(\mathbf{u})]$  as  $(F_X^-(u), +\infty)$ . This similarity justifies the integration domain in (2). As it can be seen in Proposition 2.1, the corresponding probability in the upper corrected orthant associated to  $\mathbf{X}$  at  $\hat{\mathbf{x}}(\mathbf{u})$  is the multivariate version of the probability in the  $u$ -quantile upper orthant for univariate distributions, being  $P[X \in R_X(\hat{x}(u))] = P[X \geq F_X^-(u)] = 1 - u$  in the univariate case. For this reason, it is used in the integrand in (2).

Several interesting properties for the multivariate excess wealth function can be shown as in the univariate case. For example, it is easily verified that  $S_{\mathbf{X}}^+(\mathbf{u})$  is a decreasing function in  $\mathbf{u}$  and invariant with respect to changes in a location parameter. Also, in a similar way to the univariate case,  $0 \leq S_{\mathbf{X}}^+(\mathbf{u}) \leq S_{\mathbf{X}}^+(\mathbf{0})$  for all  $\mathbf{u}$  in  $[0, 1]^n$  holds. Furthermore, in the case of independent random variables, the multivariate excess-wealth function can be decomposed as the product of the corresponding univariate excess-wealth functions.

Particularly, it is interesting to obtain an expression easier to hand for the bivariate excess wealth function. Let  $\mathbf{X} = (X_1, X_2)$  be a bidimensional random vector with density function  $f_{\mathbf{X}}(x_1, x_2)$ . Let  $(u_1, u_2)$  in  $[0, 1]^2$  and let  $\mathbf{x} = (x_1, x_2)$  be a point in  $\mathbb{R}^2$  such that  $x_1 \geq F_{X_1}^-(u_1)$  and  $x_2 \geq F_{X_2|X_1=x_1}^-(u_2)$  so that it is held  $\mathbf{x} \in R_{\mathbf{X}}(\hat{\mathbf{x}}(u_1, u_2))$ .

By using (1) the bivariate excess wealth function can be expressed as

$$S_{\mathbf{X}}^+(u_1, u_2) = \int_{F_{X_1}^-(u_1)}^{\infty} \bar{F}_{X_1}(t_1) \cdot S_{X_2|X_1=t_1}^+(u_2) dt_1. \quad (3)$$

Fernández-Ponce *et al.* (1996) showed that if  $X$  is a univariate random variable with support on  $[0, \infty)$  and finite mean  $\mu$  then  $S_X^+(0) = \mu$ . It is interesting to study a possible generalization of this property. For this reason, the following definition is given.

**Definition 2.3.** Let  $\mathbf{X}$  be a nonnegative random vector. Then the multiple expectation associated to  $\mathbf{X}$ , when it exists, is defined as the real value

$$\bar{\mu}_{\mathbf{X}} = \int_{\text{Supp}(\mathbf{X})} P[\mathbf{X} \in R_{\mathbf{X}}(\mathbf{t})] d\mathbf{t}.$$

If  $X$  is a univariate random variable with support on  $(a, +\infty)$ , where  $a$  is a nonnegative value, then it is held that

$$\int_a^{+\infty} \bar{F}(t) dt = \mathbf{E}(X) - a.$$

Obviously, this equality does not exist for the multivariate case. However, a similar condition can be obtained. That is, if  $\mathbf{X} = (X_1, X_2)$  is a bivariate random variable, then under straightforward calculations is easily obtained by using (1) that

$$\bar{\mu}_{\mathbf{X}} = \int_{F_{X_1}^-(0)}^{\infty} \bar{F}_1(t) \mathbf{E}[X_2|X_1 = t] dt - \int_{F_{X_1}^-(0)}^{+\infty} \bar{F}_1(t) F_{X_2|X_1=t}^-(0) dt. \quad (4)$$

Particularly, if  $\mathbf{X}$  represents a non-negative bivariate random lifetime with  $\text{Supp}(\mathbf{X}) = [0, +\infty)^2$ , then  $\bar{\mu}_{\mathbf{X}} = \int_{F_{X_1}^-(0)}^{\infty} \bar{F}_1(t) \mathbf{E}[X_2|X_1 = t] dt$  is verified. For this reason, the corrected multiple expectation associated to  $\mathbf{X}$  is defined as

$$\nu_{\mathbf{X}} = \int_{F_{X_1}^-(0)}^{\infty} \bar{F}_1(t) \mathbf{E}[X_2|X_1 = t] dt, \quad (5)$$

when it exists.

Furthermore, if the components are independent random variables, then the multiple expectation coincides with the product of the marginal expectations.

Also, it is interesting to note that in general it is held that

$$S_{\mathbf{X}}^+(0, 0) = \bar{\mu}_{\mathbf{X}} \leq \nu_{\mathbf{X}}.$$

**Remark 2.2.** The corrected multiple expectation,  $\nu_{\mathbf{X}}$ , can be defined for higher dimension. An method to compute  $\bar{\mu}_{\mathbf{X}}$  and  $\nu_{\mathbf{X}}$  for any dimensions is given in Appendix A.

### 3. The NBUE property: a new multivariate version

Different aging concepts have been defined and studied in reliability literature (see Barlow and Proschan, 1975, Fernández-Ponce *et al.*, 1996, and Shaked and Shanthikumar, 2007, Gupta and Peng, 2009, among others). In particular, the New Better than Used in Expectation (NBUE) notion has been defined to describe the property satisfied by an item whose residual lifetime decrease in the expectation along time. Formally,

assume that  $X$  represents the lifetime of an item with absolutely continuous distribution  $F_X(\cdot)$ , survival function  $\bar{F}_X(t) = P[X > t]$  and mean residual life function  $\mu_X(t) = \mathbf{E}[X - t | X > t] = \frac{\int_t^\infty \bar{F}_X(u) du}{\bar{F}_X(t)}$ , for  $t \geq 0$ . Then,

$$X \text{ is NBUE if } \mu_X(0) \geq \mu_X(t) \text{ for all } t \geq 0.$$

An exhaustive list of applications and properties of this notion may be found, for example, in Barlow and Proschan (1975), Hollander and Proschan (1984) or Kochar and Wiens (1987). NBUE random variables have been also characterized by mean of different stochastic orderings (see Shaked and Shanthikumar, 2007).

Now, a new multivariate version of the mean residual life function is given.

**Definition 3.1.** Let  $\mathbf{X}$  be an  $n$ -dimensional random variable. The total expected residual life of  $\mathbf{X}$  in  $\mathbf{x}$  is defined, when it exists, as the following real value:

$$\bar{\mu}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{P\{\mathbf{X} \in R_{\mathbf{X}}(\mathbf{x})\}} \int_{R_{\mathbf{X}}(\mathbf{x})} P[\mathbf{X} \in R_{\mathbf{X}}(\mathbf{t})] dt \quad (6)$$

for all  $\mathbf{x} \in \text{Supp}(\mathbf{X})$ .

Note there exists a closed relationship between the total expected residual life function and the multivariate excess wealth function. In fact, it is held that

$$S_{\mathbf{X}}^+(\mathbf{u}) = \bar{\mu}_{\mathbf{X}}(\hat{\mathbf{x}}(\mathbf{u})) \prod_{i=1}^n (1 - u_i). \quad (7)$$

By taking into account the last comments in the above section and the definition of the univariate NBUE property, the following definition is given.

**Definition 3.2.** Let  $\mathbf{X}$  be a non-negative random vector with finite translate total expectation  $\nu_{\mathbf{X}}$ .  $\mathbf{X}$  is said to have a Corrected Multivariate New [Worse] Better than Used in Expectation ( $\mathbf{X} \in \mathcal{F}_{CMNBUE[CMNWUE]}$ ) distribution if

$$\nu_{\mathbf{X}} \geq [\leq] \bar{\mu}_{\mathbf{X}}(\mathbf{x})$$

for all  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n) \in \text{Supp}(\mathbf{X})$ .

**Example 3.1.** Let  $\mathbf{X} = (X_1, X_2)$  be a bivariate vector with  $\text{Supp}(\mathbf{X}) = A \cup B$  and a joint density function given by

$$f(x_1, x_2) = \begin{cases} \frac{1}{2} & \text{if } (x_1, x_2) \in A, \\ 1 & \text{if } (x_1, x_2) \in B, \\ 0 & \text{otherwise} \end{cases}$$

where

$$A = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\} \text{ and } B = \{(x_1, x_2) \in \mathbb{R}^2 : 1 \leq x_1 \leq 2, 0 \leq x_2 \leq 2 - x_1\}.$$

The total expected residual life of  $\mathbf{X}$  is given by

$$\bar{\mu}_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{(x_2-1)(2x_1^2-8x_1+7)}{8(x_1-2)} & \text{if } (x_1, x_2) \in A, \\ \frac{1}{8}(x_1-2)(x_1+x_2-2) & \text{if } (x_1, x_2) \in B. \end{cases}$$

In particular,  $\nu_{\mathbf{X}} = \bar{\mu}_{\mathbf{X}} = 7/16$ . Moreover, the total expected residual life is decreasing in  $\mathbf{x}$ . In fact, if  $(x_1, x_2) \in A$ ,

$$\frac{\partial}{\partial x_1} \bar{\mu}_{\mathbf{X}}(\mathbf{x}) = \frac{(x_2-1)(2x_1^2-8x_1+9)}{8(x_1-2)^2} \quad \text{and} \quad \frac{\partial}{\partial x_2} \bar{\mu}_{\mathbf{X}}(\mathbf{x}) = \frac{2x_1^2-8x_1+7}{8(x_1-2)}$$

which are negative for all  $(x_1, x_2) \in A$  and if  $(x_1, x_2) \in B$ ,

$$\frac{\partial}{\partial x_1} \bar{\mu}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{8}(2x_1 + x_2 - 4) \quad \text{and} \quad \frac{\partial}{\partial x_2} \bar{\mu}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{8}(x_1 - 2)$$

which are also negative for all  $(x_1, x_2) \in B$ . Therefore, given that  $\bar{\mu}_{\mathbf{X}}(\mathbf{x})$  is decreasing, for all  $\mathbf{x} \geq 0$ ,  $\nu_{\mathbf{X}} = \bar{\mu}_{\mathbf{X}} \geq \bar{\mu}_{\mathbf{X}}(\mathbf{x})$ , that is,  $\mathbf{X}$  has a CMNBUE distribution.

Now, it is proved that if  $X_1$  and  $X_2$  have NBUE distributions and  $\varphi$  is a nonnegative function, the random vector  $(X_1, \varphi(X_1) + X_2)$  has a CMNBUE distribution.

**Theorem 3.1.** *Let  $X_1$  and  $X_2$  be two independent and NBUE univariate distributions and let  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  be a nonnegative function. Hence  $\mathbf{Z} = (Z_1, Z_2) = (X_1, \varphi(X_1) + X_2)$  has a CMNBUE distribution.*

*Proof.* Let  $F_1$  and  $F_2$  be the corresponding distribution functions of the random variables  $X_1$  and  $X_2$ , respectively. It is easily shown that

$$\begin{aligned} \nu_{\mathbf{Z}} &= \int_0^{\infty} \bar{F}_1(t) \mathbf{E}(\varphi(X_1) + X_2 | X_1 = t) dt \\ &= \int_0^{\infty} \varphi(t) \bar{F}_1(t) dt + \mathbf{E}(X_1) \mathbf{E}(X_2). \end{aligned} \quad (8)$$

Obviously, it is held that  $\text{Supp}(\mathbf{Z}) = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 > 0, z_2 - \varphi(z_1) > 0\}$ .

Then, using (6), if  $\mathbf{z} = (z_1, z_2) \in \text{Supp}(\mathbf{Z})$ , it is obtained that

$$\begin{aligned} \bar{\mu}_{\mathbf{Z}}(\mathbf{z}) &= \frac{1}{\bar{F}_1(z_1) \bar{F}_2(z_2 - \varphi(z_1))} \int_{z_1}^{\infty} \bar{F}_1(t_1) dt_1 \int_{z_2 - \varphi(z_1)}^{\infty} \bar{F}_2(t_2) dt_2 \\ &= \mathbf{E}(X_1 - z_1 | X_1 > z_1) \mathbf{E}[X_2 - (z_2 - \varphi(z_1)) | X_2 > (z_2 - \varphi(z_1))] \end{aligned}$$

for all  $z_2 > \varphi(z_1)$ . Furthermore, by using the fact that  $X_1$  and  $X_2$  have NBUE distributions, the following inequality is verified

$$\mathbf{E}(X_1 - z_1 | X_1 > z_1) \mathbf{E}[X_2 - (z_2 - \varphi(z_1)) | X_2 > (z_2 - \varphi(z_1))] \leq \mathbf{E}(X_1) \mathbf{E}(X_2).$$

Consequently, from (8) and using the fact that  $\varphi(\cdot)$  is a nonnegative function, it is easily deduced that

$$\nu_{\mathbf{Z}} > \mathbf{E}(X_1) \mathbf{E}(X_2) \geq \bar{\mu}_{\mathbf{Z}}(\mathbf{z}) \text{ for all } \mathbf{z} \text{ in } \text{Supp}(\mathbf{Z}).$$

Thus, the random vector  $\mathbf{Z} = (X_1, \varphi(X_1) + X_2)$  is CMNBUE by using Definition 3.2.  $\square$

The following result gives a necessary and sufficient condition for the CMNBUE property of  $\mathbf{X}$ , based on the multivariate excess wealth function. It can be considered as a generalization of Corollary 3.1. (c) in Fernández-Ponce *et al.* (1998).

**Theorem 3.2.** *Let  $\mathbf{X}$  be a random vector verifying the regularity conditions (RC).  $\mathbf{X}$  is a CMNBUE distribution if and only if*

$$S_{\mathbf{X}}^+(\mathbf{u}) \leq \nu_{\mathbf{X}} \prod_{j=1}^n (1 - u_j) \text{ for all } \mathbf{u} \in [0, 1]^n.$$

*Proof.* Firstly, assume that  $\mathbf{X}$  has a CMNBUE distribution. By Definition 3.2,  $\nu_{\mathbf{X}} \geq \bar{\mu}_{\mathbf{X}}(\mathbf{x})$  is held for all  $\mathbf{x} \in \text{Supp}(\mathbf{X})$ . From (RC) defined in Section 2, it is also known that for each  $\mathbf{x} \in \text{Supp}(\mathbf{X})$ , only one vector  $\mathbf{u} \in [0, 1]^n$  exists such that  $\mathbf{x} = \hat{\mathbf{x}}(\mathbf{u})$ . Now, by using the equality (7) and the Proposition 2.1, the result is held. Conversely, if  $S_{\mathbf{X}}^+(\mathbf{u}) \leq \nu_{\mathbf{X}} \prod_{j=1}^n (1 - u_j)$  for all  $\mathbf{u} \in [0, 1]^n$ , and  $\mathbf{X}$  verifies the regularity conditions, then it is easily seen that  $\mathbf{X}$  has a CMNBUE distribution.  $\square$

## 4. An application to real dataset

### 4.1. The model

Albert *et al.* (1978a) and Albert *et al.* (1978b) studied the age and tumor size at detection. They assumed that the models for tumor detectability can be synthesized by first modeling the behavior of tumor growth over time and superimposing a model for detection probability as a function of tumor size. Hanin and Boucher (1999) used this model to approach the problem of optimal cancer screening.

Let  $T$  be the age at tumor onset and  $W$  be the time of spontaneous detection of the tumor measured from the onset of disease. Define a random variable  $S$  to represent tumor size (the number of cells in a tumor) at spontaneous detection. Hanin and Boucher (1999) supposed that the law of tumor growth is described by a deterministic function  $f : [0, \infty) \rightarrow [1, \infty)$  with  $f(0) = 1$ , such that  $S = f(W)$ . It is assumed that:

- (a) random variables  $T$  and  $W$  are absolutely continuous and independent.
- (b) function  $f$  is differentiable and  $f' > 0$ .
- (c) the hazard rate for spontaneous detection of the tumor is proportional to the current tumor size with a non-negative coefficient. That is,  $r_W(w) = \alpha S(w) = \alpha f(w)$ , where  $\alpha$  is a non-negative constant.

Sample values of the random vector  $\mathbf{Y} = (S, T + W)$  with components interpreted as tumor size at spontaneous detection and age, respectively, can be observed. In the particular case of exponential tumor growth, the joint density function for the random variable  $\mathbf{Y}$  is easily obtained (see Bartoszyński *et al.*, 2001).

Under these considerations, we study estimations of the parameters in a non-deterministic model which relates the variables  $S$  and  $W$ . Also, we analyze the unknown random variable  $T$  plus a unknown random delay, which will be denoted by  $\Delta$ .

For our model, the following hypothesis are considered.

- C1. The unknown variables  $W$  and  $T$  have NBUE distributions. Note that this property is an intuitive survival property for the random variable  $W$ , as well as for the random variable  $T$ . In fact,  $W$  has an NBUE distribution if  $\mathbf{E}(W) \geq \mathbf{E}(W - w | W > w)$  is verified for all  $w$  in  $\mathbb{R}$ . This inequality indicates that the mean time of spontaneous detection of tumor measured from the onset of disease is greater than or equal to the mean residual time of spontaneous tumor detection by assuming that this time is greater than  $w$ . Similarly, if  $T$  has an NBUE distribution, then the mean time from the birth of an individual to the appearance of the first tumor cell is greater than, or equal to, the mean residual time to the appearance of the first tumor cell by assuming that this time is greater than  $t$ .
- C2. The distribution of tumor size  $S$  is unknown but it is assumed that the logarithm of  $S$  has an NBUE distribution.
- C3. A non-deterministic exponential tumor growth is assumed. So, the relationship between  $S$  and  $W$  can be modeled as

$$W = \alpha + \beta \ln S + \Delta \tag{9}$$

where  $\beta > 0$  and  $\Delta$  is a random delay which is independent of  $S$ .

- C4. It is assumed that  $T + \Delta$  has an NBUE distribution.

Consequently, we have a random sample of size  $n$  from a homogeneous population:

$$\mathbf{y}_i = (s_i, v_i) \quad i = 1, \dots, n \text{ with } v_i = t_i + w_i \text{ and } w_i = \alpha + \beta \ln s_i + \delta_i$$

and where  $s_i$  and  $v_i$  represent the tumor size and the age at detection for the  $i$ -th patient, respectively. Note that the values of  $t_i$  and  $w_i$  are completely unknown.

From the relationship  $v_i = t_i + \alpha + \beta \ln s_i + \delta_i$  for all  $i$ , it follows that

$$t_i + \delta_i = v_i - \alpha - \beta \ln s_i \text{ for } i. \quad (10)$$

For simplicity, the value  $t_i + \delta_i$  will be denoted by  $\eta_i$  for  $i = 1, \dots, n$ , as sample values of the variable  $\eta = T + \Delta$ , where  $T$  is the time at onset and  $\Delta$  is the nonnegative random time in (9).

It is easy to show, by using Theorem 3.1 and taking  $X_1 = \ln S$ ,  $X_2 = T + \alpha + \Delta$  and  $\phi(x) = \beta x$ , that the random vector  $(\ln S, V) = (\ln S, T + \alpha + \beta \ln S + \Delta)$  has a CMNBUE distribution. That means that the multiple expectation associated to the vector with component logarithm of tumor size and age at detections is greater than or equal to the total expected residual life of this vector in any particular value of the vector  $(\ln S, V)$ . This property enables a condition to be established in the problem of estimating the parameters  $\alpha$  and  $\beta$  in the model given in (9).

Therefore, from a conservative point of view, to estimate the above parameters is equivalent to estimating  $\alpha$  and  $\beta$  such that the sum  $\sum_{i=1}^n \eta_i$  is minimized under the constraints  $\eta_i \geq 0$  for all  $i$  and  $(\ln S, T + \alpha + \beta \ln S + \Delta) \in \mathcal{F}_{CMNBUE}$ . This problem can be formulated as

$$\begin{aligned} \min \quad & \sum_{i=1}^n \eta_i \\ \text{s.t.} \quad & \\ & \eta_i \geq 0 \quad \text{for all } i, \\ & (\ln S, T + \alpha + \beta \ln S + \Delta) \in \mathcal{F}_{CMNBUE}. \end{aligned} \quad (11)$$

$$(12)$$

In order to break down the model more easily, the restriction (12) must be expressed in a different way. In Appendix B, it is shown that, by using Theorem 3.2, it is verified that

$(\ln S, T + \alpha + \beta \ln S + \Delta) \in \mathcal{F}_{CMNBUE}$  if, and only if,

$$\frac{S_{\ln S}^+(u_1)}{\mathbf{E}(\ln S)(1 - u_1)} \cdot \frac{S_{T+\Delta}^+(u_2)}{\mathbf{E}(T + \Delta)(1 - u_2)} \leq 1 + \frac{\alpha}{\mathbf{E}(T + \Delta)} + \beta \frac{\mathbf{E}(\ln^2 S)}{\mathbf{E}(\ln S)\mathbf{E}(T + \Delta)} \quad (13)$$

for all  $(u_1, u_2) \in [0, 1]^2$ .

In particular, the inequality (13) is obtained for  $(0, 0)$ . Moreover, by assumption, the variables  $\ln S$  and  $T + \Delta$  have NBUE distributions, so their corresponding univariate excess wealth functions  $S_{\ln S}^+(u_1)$  and  $S_{T+\Delta}^+(u_2)$  are decreasing for all  $u_1$  and  $u_2$ . Moreover, it is verified that (see Fernández-Ponce *et al.*, 1998)

$$S_{\ln S}^+(u_1) \leq \mathbf{E}(\ln S)(1 - u_1) \text{ and } S_{\ln S}^+(0) = \mathbf{E}(\ln S) \text{ for all } u_1 \in [0, 1),$$

$$S_{T+\Delta}^+(u_2) \leq \mathbf{E}(T + \Delta)(1 - u_2) \text{ and } S_{T+\Delta}^+(0) = \mathbf{E}(T + \Delta) \text{ for all } u_2 \in [0, 1).$$

Therefore, by taking into account that  $\mathbf{E}(T + \Delta) > 0$ , the inequality (13) is equivalent to

$$0 \leq \alpha + \beta \frac{\mathbf{E}(\ln^2 S)}{\mathbf{E}(\ln S)}.$$

Thus, the problem for estimating  $\alpha$  and  $\beta$  is the following linear programming problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^n \eta_i \\ \text{s.t.} \quad & \eta_i \geq 0 \text{ for all } i, \\ & 0 \leq \alpha + \beta \frac{\mathbf{E}(\ln^2 S)}{\mathbf{E}(\ln S)}. \end{aligned}$$

Given that the distribution of  $\ln S$  is unknown, the values  $\mathbf{E}(\ln^2 S)$  and  $\mathbf{E}(\ln S)$  are also unknown but finite. Considering that  $T_n = \frac{\sum_{i=1}^n \ln^2 s_i}{\sum_{i=1}^n \ln s_i}$  converges almost sure to  $\frac{\mathbf{E}(\ln^2 S)}{\mathbf{E}(\ln S)}$ , this problem can be solved by replacing the last constraint with the following inequality

$$0 \leq \alpha + \beta T_n. \quad (14)$$

Note that our model is not exactly a parametric regression model since the ‘residual’ random variable is non-negative. Neither a non-regular regression can be used (see Smith, 1994), because of the hypothesis on the tail of the distribution function for the errors, which cannot be verified for NBUE distributions in general. Consequently, this problem must be solved by using a different tool. In particular, the quantile regression model is used. For more details about quantile regression and the iterative algorithm which must be used to estimate the parameters see Koenker and Basset (1978) and Koenker and D’Orey (1987).

Now, note that the values of  $\alpha$  and  $\beta$  that minimize the sum of  $\eta_i$  for all  $\eta_i \geq 0$  can be obtained by solving the problem of the smallest regression quantile when the values  $\eta_i = v_i - \alpha - \beta \ln s_i$  for  $i = 1, \dots, n$  are considered as the values of the residual variable.

Thus, it can be concluded that, from a conservative point of view, estimating the parameters in our model is equivalent to solving a particular problem of quantile regression. Moreover, it is possible to analyse the unknown variable  $T + \Delta$  by means of the residuals obtained in this quantile regression.

## 4.2. Dataset analysis

### 4.2.1. Materials and Methods

From July 1984 to December 1989, the German Breast Cancer Study Group initially recruited 720 patients with primary node positive breast cancer into the Comprehensive Cohort Study (Schmoor *et al.*, 1996). Some of the variables considered for each patient were: date of diagnosis, patient’s age at diagnosis, tumor size (tumor diameter in mm), tumor grade and number of nodes involved. The study itself which was realized by this group is not of interest for the present purpose. Our attention is focused on two variables: age at detection and tumor size. The 686 patients who completed the data for the standard factors of age and tumor size are analysed in this study.

Tumor size, initially given in mm, was transformed into the number of tumor cells per  $cm^3$ . For this, it is assumed the tumor is a symmetric ball in  $\mathbb{R}^3$  and it is considered that approximately  $10^{12}$  tumor cells exist per  $cm^3$ , (see Spanish Society of Medical website, <http://www.seom.org/>). A descriptive analysis was realized for tumor size as well as patient’s age at detection. The NBUE property for both variables was confirmed by using a statistical test given by Fernández-Ponce *et al.* (1996). The plots of their univariate excess-wealth functions were also used to recognize this property.

The parameters in the model that relate the variables sojourn time ( $W$ ) and tumor size ( $S$ ) were estimated by using the model proposed in the above section. The residuals of this regression are considered as the estimation of the age at tumor onset plus a random delay for each patient.

All statistical analyses were performed using **R** software. In particular, the *quantreg* package was used to solve the problems of quantile regression (see <http://cran.r-project.org/> and Appendix A in Koenker, 2005).

### 4.2.2. Results

A descriptive analysis for the variables age, tumor size and the logarithm of tumor size is given in Table 1. The  $p$ -values for the Shapiro-Wilk normality test was 0.00 for both cases, therefore the normality hypothesis is rejected for both variables.

The plots of the empirical excess-wealth (ew) functions of the age and logarithm of tumor size are given in Figure 1. It can be observed that the empirical ew function of the age variable is under the line from the

Table 1: A descriptive analysis for the age, tumor size and the logarithm of tumor size .

	Mean	Median	Interquartile interval	Standard deviation
Age	53	53.2	46 – 61	10.12
Tumor size	$2.53 \cdot 10^{16}$	$8.18 \cdot 10^{12}$	$4.18 \cdot 10^{15} - 22.4 \cdot 10^{15}$	$5.96 \cdot 10^{16}$
Logarithm of tumor size	29.89	29.732	20.06 – 30.74	1.37

Figure 1: (a) Empirical ew function of age (blue line) and empirical ew function for exponential with parameter equal to the inverse of mean age; (b) Empirical ew function of logarithm tumor size (blue line) and empirical ew function for exponential with parameter equal to the inverse of mean logarithm tumor size; (c) Empirical ew function of tumor size (blue line) and empirical ew function for exponential with parameter equal to the inverse of mean tumor size.

point  $(1, 0)$  to point  $(0, \mu_V)$ , where  $\mu_V$  is the sample mean age. Note that this line corresponds to the ew function of an exponential distribution with parameter  $\lambda = \mu_V^{-1}$ . Similarly, this fact occurs for the logarithm of tumor size. From a theoretical viewpoint, this property is always verified for all variables having an NBUE distribution. Moreover, the NBUE property of age and the logarithm of tumor size were checked by means of the test for NBUE alternatives given by the Fernández-Ponce *et al.* (1996). The statistic for this test,  $\Psi(F_n(t))$ , is based on the empirical ew function and it has really interesting asymptotic properties. For patient's age is obtained that  $\Psi(F_n(t)) = 0.4741$  and, for the logarithm of tumor size, the corresponding statistic is obtained with a value  $\Psi(F_n(t)) = 0.4942$ . Therefore, for both random variables we can assume the NBUE property (see critical values for NBUE alternatives in Table 2). In contrast with these variables, the tumor size variable  $S$  does not have this ageing property, but it has the dual property NWUE (New Worst than Used in Expectation). In Figure 1 c), it can be seen that the empirical ew function is on the line from the point  $(1,0)$  to point  $(0, \mu_S)$ , where  $\mu_S$  is the sample mean of tumor size. When the constraint  $(\ln S, T + \alpha + \beta \ln S + \Delta) \in \mathcal{F}_{CMNBUE}$  was not included in the quantile regression problem, the estimates intercept and slope of the smallest quantile regression line were  $\hat{\alpha} = -36.04$  and  $\hat{\beta} = 2.02$ , respectively. It should be added that the smallest regression quantile is obtained for  $\theta = 0.0001$  (since for values of  $\theta$  upper than 0.0001 some residuals are negative and then they would not verify the constraint (11), and for values of  $\theta$  lower than 0.0001, it is obtained the same residuals and the same quantile regression line).

The mean and the standard deviation for the residuals obtained from the quantile regression were 28.56 and 10.47, respectively. A 95 percent confidence interval for this mean was (27.78; 29.35) and the standard error was 0.40. Recall that, from a conservative point of view, the  $i$ th residual in the quantile regression is considered as the estimation of the value  $t_i + \delta_i$ . Therefore, it would be said that the mean age at appearance of the first tumor cell plus a random delay was approximately 28.56 years, when the CMNBUE

Table 2: Critical value for NBUE alternative with different significance levels and sample size.

$n$	$\alpha$		
	0.1	0.05	0.01
400	0.0549	0.0732	0.0925
450	0.0544	0.0715	0.0910
475	0.0543	0.0661	0.0876
500	0.0532	0.0654	0.0818
600	0.0434	0.0570	0.0737
650	0.0425	0.0535	0.0736
700	0.0401	0.0532	0.0721

Table 3: Quantiles of the estimated variable  $T + \Delta$  when the CMNBUE property of vector  $(\ln S, T + \alpha + \beta \ln S + \Delta)$  is not considered (case 1) and is considered (case 2) as a constraint in the quantile regression problem.

	$Q(0.05)$	$Q(0.25)$	$Q(0.5)$	$Q(0.75)$	$Q(0.95)$
Case 1	10.81	20.89	28.54	36.90	45.00
Case 2	11.96	21.85	29.38	37.56	45.35

Table 4: Approximate age at detection for different values of tumor size.

Tumor size (mm)	$\hat{V}$	$\hat{V}_l$	$\hat{V}_u$	Tumor size (mm)	$\hat{V}$	$\hat{V}_l$	$\hat{V}_u$
3	42.63	41.86	43.41	16	50.61	49.84	51.39
4	44.00	43.23	44.78	18	51.17	50.40	51.95
5	45.07	44.29	45.84	20	51.67	50.90	52.45
6	45.94	45.16	46.71	25	52.74	51.96	53.51
7	46.81	46.03	47.58	30	53.61	52.83	54.38
8	47.31	46.53	48.08	40	54.98	54.20	55.75
9	47.87	47.09	48.64	50	56.04	55.27	56.82
10	48.37	47.60	49.15	60	56.91	56.14	57.69
12	49.24	48.46	50.01	70	57.64	56.87	58.42
14	49.98	49.21	50.76	80	58.28	57.51	59.06

constraint is not considered. It was also proved that the variable  $T + \Delta$  has an NBUE distribution, given that the value of statistic for the test for NBUE alternatives was  $\Psi(F_n(t)) = 0.4305$ .

On the other hand, if the property  $(\ln S, T + \alpha + \beta \ln S + \Delta) \in \mathcal{F}_{CMNBUE}$  is included as a constraint in the quantile regression problem, the estimated intercept and slope are  $\hat{\alpha} = -23.80$  and  $\hat{\beta} = 1.59$ , respectively. The mean and the standard deviation for the residual variable obtained from the quantile regression were 29.30 and 10.33, respectively. A 95 percent confidence interval for this mean was (28.53; 30.08) and the standard error was 0.39. Therefore, if the multivariate aging property is considered, it could be concluded that the mean age at the first tumor cell appearance plus a random delay is approximately 29.30 years.

Table 3 shows different estimated quantiles for the variable  $T + \Delta$  when the CMNBUE property is and is not considered in the quantile regression problem.

Considering the estimations of the parameters  $\alpha$  and  $\beta$  when it is included the multivariate aging property in the linear program problem, the patients's age at detection is approximated for different tumor sizes by using the relationship  $\hat{V} = \hat{\alpha} + \hat{\beta} \ln S + \overline{T + \Delta}$ , where  $\overline{T + \Delta}$  is the estimation of the mean of the variable  $T + \Delta$  obtained as the mean of the residuals in the quantile regression. Table 4 shows the results. In the third and fourth columns, the age at detection is also given when the variable  $T + \Delta$  is estimated by the lower and upper values of the confidence interval for  $\overline{T + \Delta}$ , respectively.

#### 4.2.3. Conclusions

Data from the Comprehensive Cohort Study performed by the German Breast Cancer Study Group are analysed in order to estimate the parameters in the non-deterministic model which describes the tumor growth when the patient's age at detection and the tumor size are known. Several descriptive statistics of the time onset variable plus a random delay were also obtained. For this propose, the model studied in Section 4.2 has been applied. In order to realize how the multivariate aging CMNBUE property influences in our results, the linear programming problem is solved in two different cases. In the first case the CMNBUE property is not included as a constraint in the problem whereas it is included in the second case.

The hypothesis of the model are actually confirmed. The initially assumed NBUE property of the detection

age and the logarithm of tumor size are checked as well as the NBUE property of the estimated variable  $T + \Delta$ .

Under a non-deterministic exponential tumor growth, and from a conservative point of view, the estimations of parameters in the model (9) were  $\hat{\alpha} = -36.04$  and  $\hat{\beta} = 2.02$  when the multivariate property was not included as a constraint and  $\hat{\alpha} = -23.80$  and  $\hat{\beta} = 1.59$  when this property was included. Estimations of the values  $t_i + \delta_i$ , defined as in (10), were obtained as the residuals of the smallest quantile regression of the response variable  $V$  and the independent variable  $\ln S$ . A paired t-test is applied for these residuals (with and without the inequality (14)) obtaining a  $p$ -value  $< 2.2 \times 10^{-16}$ . Consequently, the residuals are significantly different.

Given that the CMNBUE property appears as an inherent property of the vector  $(\ln S, V)$ , attention was focused on the approximate age at detection, when this constraint is considered. Some interesting conclusions about the age at detection can be obtained from Table 4. Mammography is a well-known test commonly used in breast cancer screening. The main advantage of mammography is that it can find tumors that are too small to palpated and allows an early diagnosis to be obtained. However, the mammography has some disadvantages. False-negatives test results frequently occur in young women because, in this group, the breast tissue is more dense. The same occurs with the false-positives. They are more common in younger women, women with a family history of breast cancer or women who have had previous breast biopsies. In addition, mammography exposes the breast to radiation which is considered as a risk factor for breast cancer. For women younger than 50 years of age, the risks from a radiation exposure or a false-positive test result may be greater than the benefits provided by annual mammogram screening. In light of these disadvantages of mammography in breast cancer screening, the test is recommended every two years in women between the ages of 50 and 74 years. (See, for example, the National Cancer Institute (USA) website <http://www.cancer.gov>, and the Spanish Society of Medical Oncology website <http://www.seom.org/>). However, the results in Table 4 show that tumors of a smaller size (tumor diameter between 3 mm and 14 mm) are detected at ages of less than 50 years. It is worth noting that the 40 percent of women in this study were younger than 50 years of age. Clear evidence is shown for the need to consider breast cancer screening in women younger than 50 years of age. However, given the disadvantages presented by mammography, other tests should be considered for younger women. This fact would justify the time and capital investment needed in research for new techniques to detect breast cancer and to consider younger women for breast cancer screening.

## Appendix A. A method to compute $\mu_{\mathbf{X}}$ and $\nu_{\mathbf{X}}$

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a random vector in  $\mathbb{R}^n$ . A recursive method to compute  $\mu_{\mathbf{X}}$  and  $\nu_{\mathbf{X}}$  is developed. Let  $\mathbf{t}_i = (t_1, \dots, t_i)$  be a point in  $\mathbb{R}^i$  and  $\mathbf{X}_i = (X_1, \dots, X_i)$  be a random vector in  $\mathbb{R}^i$ , for  $i = 1, \dots, n-1$ .

### Algorithm.

1. By using (4) and (5), it is obtained that

$$\bar{\mu}(X_{n-1}, X_n) | \mathbf{X}_{n-2} = \mathbf{t}_{n-2} = \nu(X_{n-1}, X_n) | \mathbf{X}_{n-2} = \mathbf{t}_{n-2} - \int_{F_{X_{n-1} | \mathbf{X}_{n-2} = \mathbf{t}_{n-2}}^-(0)}^{\infty} \bar{F}_{X_{n-1} | \mathbf{X}_{n-2} = \mathbf{t}}(w) F_{X_n | (\mathbf{X}_{n-2} = \mathbf{t}_{n-2}, X_{n-1} = w)}^-(0) dw.$$

and

$$\nu(X_{n-1}, X_n) | \mathbf{X}_{n-2} = \mathbf{t}_{n-2} = \int_{F_{X_{n-1} | \mathbf{X}_{n-2} = \mathbf{t}_{n-2}}^-(0)}^{\infty} \bar{F}_{X_{n-1} | \mathbf{X}_{n-2} = \mathbf{t}}(w) E[X_n | \mathbf{X}_{n-2} = \mathbf{t}, X_{n-1} = w] dw.$$

2. The following steps are

$$\bar{\mu}(X_i, \dots, X_n) | \mathbf{X}_{i-1} = \mathbf{t}_{i-1} = \int_{F_{X_i | \mathbf{X}_{i-1} = \mathbf{t}_{i-1}}^-(0)}^{\infty} \bar{F}_{X_i | \mathbf{X}_{i-1} = \mathbf{t}_{i-1}}(w) \bar{\mu}(X_{i+1}, \dots, X_n) | (\mathbf{X}_{i-1} = \mathbf{t}_{i-1}, X_i = w) dw.$$

for  $i = 2, \dots, n-2$ .

3. The last computation is

$$\bar{\mu}_{\mathbf{X}} = \int_{F_{X_1}^-(0)}^{\infty} \bar{F}_{X_1}(w) \bar{\mu}(X_2, \dots, X_n) | X_1 = w dw$$

and

$$\nu_{\mathbf{X}} = \bar{\mu}_{\mathbf{X}} + \int_{\text{Supp}(\mathbf{X}_{n-1})} P(\mathbf{X}_{n-1} \in R_{\mathbf{X}_{n-1}}(\mathbf{t}_{n-1})) F_{X_n | \mathbf{X}_{n-1} = \mathbf{t}_{n-1}}^-(0) d\mathbf{t}_{n-1}.$$

## Appendix B. Proof of the inequality (19)

Denote  $L = \ln S$ , and consider  $\mathbf{H} = (L, \alpha + \beta L + T + \Delta)$ . By Definition 2.2, it is obtained that

$$S_{\mathbf{H}}^+(u_1, u_2) = \int_{F_L^-(u_1)}^{\infty} \bar{F}_L(t_1) \cdot S_{\alpha + \beta L + T + \Delta | L = t_1}^+(u_2) dt_1, \quad (\text{B.1})$$

and we know that

$$S_{\alpha + \beta L + T + \Delta | L = t_1}^+(u_2) = \int_{F_{\alpha + \beta L + T + \Delta | L = t_1}^-(u_2)}^{\infty} \bar{F}_{\alpha + \beta L + T + \Delta | L = t_1}(t_2) dt_2. \quad (\text{B.2})$$

Then, by using the following trivial equality

$$\bar{F}_{\alpha + \beta L + T + \Delta | L = t_1}(t_2) = \bar{F}_{T + \Delta}(t_2 - \alpha - \beta t_1) \quad (\text{B.3})$$

and replacing (B.3) in (B.2), it follows that

$$\begin{aligned}
S_{\alpha+\beta L+T+\Delta|L=t_1}^+(u_2) &= \int_{F_{\alpha+\beta L+T+\Delta|L=t_1}^-(u_2)}^{\infty} \bar{F}_{T+\Delta}(t_2 - \alpha - \beta t_1) dt_2 \\
&= \int_{F_{T+\Delta}^-(u_2) + \alpha + \beta t_1}^{\infty} \bar{F}_{T+\Delta}(t_2 - \alpha - \beta t_1) dt_2 \\
&= \int_{F_{T+\Delta}^-(u_2)}^{\infty} \bar{F}_{T+\Delta}(r) dr \\
&= S_{T+\Delta}^+(u_2).
\end{aligned} \tag{B.4}$$

Thus, by replacing (B.4) in (B.1), it is held that

$$\begin{aligned}
S_{\mathbf{H}}^+(u_1, u_2) &= \int_{F_L^-(u_1)}^{\infty} \bar{F}_L(t_1) \cdot S_{T+\Delta}^+(u_2) dt_1 \\
&= S_L^+(u_1) S_{T+\Delta}^+(u_2).
\end{aligned}$$

Now, the value of  $\nu_{\mathbf{H}}$  is obtained. By (5), we have that

$$\begin{aligned}
\nu_{(L, \alpha+\beta L+T+\Delta)} &= \int_0^{+\infty} \bar{F}_L(t_1) \mathbf{E}[\alpha + \beta L + T + \Delta | L = t_1] dt_1 \\
&= \alpha \mathbf{E}[L] + \beta \int_0^{+\infty} \bar{F}_L(t_1) t_1 dt_1 + \mathbf{E}[T + \Delta] \mathbf{E}[L]
\end{aligned} \tag{B.5}$$

where

$$\begin{aligned}
\int_0^{+\infty} \bar{F}_L(t_1) t_1 dt_1 &= \int_0^{+\infty} \int_{t_1}^{+\infty} t_1 dF_L(y) dt_1 \\
&= \frac{1}{2} \mathbf{E}[L^2].
\end{aligned} \tag{B.6}$$

Finally, if (B.6) is replaced in (B.5), it follows that

$$\begin{aligned}
\nu_{(L, \alpha+\beta L+T+\Delta)} &= \alpha \mathbf{E}[L] + \beta \frac{\mathbf{E}[L^2]}{2} + \mathbf{E}[T + \Delta] \mathbf{E}[L] \\
&= \mathbf{E}[L] \mathbf{E}[T + \Delta] \left\{ 1 + \frac{\alpha}{\mathbf{E}[T + \Delta]} + \beta \frac{\mathbf{E}[L^2]}{2 \mathbf{E}[L] \mathbf{E}[T + \Delta]} \right\}.
\end{aligned}$$

Therefore, from Theorem 3.2,

$\mathbf{H} = (\ln S, T + \alpha + \beta \ln S + \Delta)$  is CMNBUE if and only if

$$S_{\mathbf{H}}^+(u_1, u_2) \leq \nu_{\mathbf{H}}(1 - u_1)(1 - u_2) \text{ for all } (u_1, u_2) \in (0, 1)^2$$

or, equivalently, if

$$\frac{S_{\ln S}^+(u_1)}{(1 - u_1) \mathbf{E}[\ln S]} \frac{S_{T+\Delta}^+(u_2)}{(1 - u_2) \mathbf{E}[T + \Delta]} \leq 1 + \frac{\alpha}{\mathbf{E}[T + \Delta]} + \beta \frac{\mathbf{E}[\ln^2 S]}{\mathbf{E}[\ln S] \mathbf{E}[T + \Delta]}$$

for all  $(u_1, u_2) \in [0, 1]^2$ , and the inequality (13) is obtained.

## Acknowledgements

The authors thank the Associated Editor and two anonymous referees for valuable suggestions that helped to improve our work.

Albert, A., Gertman, P.M. and Louis, T.A. (1978 a). Screening for the early detection of cancer. I. The temporal natural history of a progressive disease state, *Mathematical Biosciences*, **40**, 1–59.

Albert, A., Gertman, P.M., Louis, T.A. and Liu, S.I. (1978 b). Screening for the early detection of cancer. II. The impact of the screening on the natural history of the disease, *Mathematical Biosciences*, **40**, 61–109.

Arjas, E. and Lehtonen, T. (1978). Approximating many server queues by means of single server queues, *Math. Oper. Res.*, **3**, 205–223.

Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*, (Holt, Rinehart and Winston.) NewYork.

Bartoszyński, R., Edler, L., Hanin, L., Kopp-Schneider, A., Pavlova, L., Tsodikov, A., Zorin, A. and Yakolev, A.Y. (2001). Modeling cancer detection: tumor size as a source of information on unoboservable stages of carcinogenesis, *Mathematical Biosciences*, **171**, 2, 113–142.

Block, H.W. and Savits, T.H. (1981). Multivariate classes of life distributions in reliability theory, *Math. Oper. Res.*, **6**, 453–461.

Fernández-Ponce, J.M., Infante-Macías, R. and Muñoz-Pérez, J. (1996). Characterization of lifetime distributions based on a quantile dispersion measure, *Computational Statistics & Data Analysis*, **21**, 5, 547–561.

Fernández-Ponce, J.M., Kochar, S.C. and Muñoz-Pérez, J. (1998). Partial orderings of distributions based on right-spread functions, *Journal of Applied Probability*, **35**, 1, 221–228.

Fernández-Ponce, J.M. and Suárez-Lloréns, A. (2003). A multivariate dispersion ordering based on quantiles more widely separated, *Journal of Multivariate Analysis*, **85**, 1, 40–53.

Gupta, R.C. and Peng, Ch. (2009) Estimating reliability in proportional odds ratio models Original Research Article *Computational Statistics Data Analysis*, **53**, 4, 1495–1510.

Hanin, L.G. and Boucher, K.M. (1999). Identifiability of parameters in the Yakovlev-Polig model of carcinogenesis, *Math. Biosci.*, **160**, 1, 1–24.

Hollander, M. and Proschan, F. (1984) Nonparametric concepts and methods in reliability. *Nonparametric methods. Handbook of Statist. 4*, 613655. North-Holland, Amsterdam.

Johnson, N.L. and Kotz, S. (1975). A vector multivariate hazard rate, *Journal of Multivariate Analysis*, **5**, 53–66.

Kochar, S.C. and Wiens, D.P. (1987). Partial orderings of life distributions with respect to their aging properties, *Naval Research Logistics*, **34**, 6, 823–829.

Koenker, R W (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press. New York .

Koenker, R W and Bassett, Jr. G. (1978). Regression quantiles, *Econometrica*, **46**, (1), 33–50.

Koenker, R W and D'Orey, V. (1987). Computing regression quantiles, *Applied Statistics*, **36**, (3), 383–393.

Marshall, A.W. and Shaked, M. (1986). Multivariate new better than used distributions: a survey, *Scandinavian J. Statist.*, **13**, 277–290.

O'Brien, G. (1975). The comparison method for stochastic processes, *Ann. Prob.*, **3**, 1, 80–88.

Rüschendorf, L. (1981). Stochastically ordered distributions and monotonicity of the OC-function of sequential probability ratio tests. *Mathematischen. Operationsforsch. Statist.* **12**, 327–338.

Schmoor, C., Olschewski, M. and Schumacher, M. (1996). Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies, *Statistics in Medicine*, **15**, 3, 263–271.

Shaked, M. and Shanthikumar, J.G. (1991). Dynamic multivariate aging notions in reliability theory, *Stochastic Process. Appl.*, **38**, 85–97.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer Series in Statistics, Springer.

Smith, R. L. (1994). Non-regular regression. *Biometrika*, **81**,1 173–183.

Spanish Society of Medical Oncology, <http://www.seom.org/>

The R Project for Statistical Computing , <http://www.r-project.org/>