ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 17, No.4
October 2019

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
Trimestral. - 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
    + 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**
  - *140 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

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 17, Number 4, October 2019, 429–448

CONFIDENCE INTERVALS AND REGIONS FOR THE GENERALIZED INVERTED EXPONENTIAL DISTRIBUTION BASED ON PROGRESSIVELY CENSORED AND UPPER RECORDS DATA

Authors: İsmail Kinaci
– Department of Statistics, Faculty of Science, Selcuk University, Konya 42250, Turkey
ikinaci@selcuk.edu.tr

Shuo-Jye Wu
– Department of Statistics, Tamkang University, Tamsui, New Taipei City 25137, Taiwan
shuo@stat.tku.edu.tw

Coşkun Kuş
– Department of Statistics, Faculty of Science, Selcuk University, Konya 42250, Turkey
coskun@selcuk.edu.tr

Received: February 2016 Revised: February 2017 Accepted: June 2017

Abstract:

- In this paper, we study the estimation problems for the generalized inverted exponential distribution based on progressively type-II censored order statistics and record values. We establish some theorems to construct the exact confidence intervals and regions for the parameters. Monte Carlo simulation studies are used to assess the performance of our proposed methods. Simulation results show that the coverage probabilities of the exact confidence interval and the exact confidence region are all close to the desired level. Finally, two numerical examples are presented to illustrate the methods developed here.

Key-Words:

- *confidence interval; joint confidence region; pivot; progressive type-II censoring; record values.*

AMS Subject Classification:

- 62F25, 62N01, 62N05.

1. INTRODUCTION

The exponential distribution was the first widely discussed lifetime distribution in the literature. This is because of its simplicity and mathematical feasibility. If the random variable T has an exponential distribution, then the random variable $Y = 1/X$ has an inverted exponential distribution. The exponential distribution was generalized, by introducing a shape parameter, and discussed by several researchers such as Gupta and Kundu [11, 12] and Raqab and Madi [19]. By introducing a shape parameter in the inverted exponential distribution, Abouammoh and Alshingiti [1] proposed a generalized inverted exponential (GIE) distribution. The probability density function and cumulative distribution function of the generalized inverted exponential distribution are given, respectively, by

$$f(x;\beta,\lambda) = \frac{\lambda\beta}{x^2}\exp(-\lambda/x)\left(1-\exp(-\lambda/x)\right)^{(\beta-1)}, \quad x>0,$$

and

$$F(x;\beta,\lambda) = 1-\left(1-\exp(-\lambda/x)\right)^{\beta}, \quad x>0,$$

where $\beta > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter.

The properties and inferences for the GIE distribution were investigated by several authors. Abouammoh and Alshingiti [1] derived some distributional properties and reliability characteristics as well as maximum likelihood estimators (MLEs) based on complete sample. Krishna and Kumar [14] obtained the MLEs and least squares estimators of the parameters of the GIE distribution under progressively type-II censored sample. Dey and Dey [8] discussed the necessary and sufficient conditions for existence, uniqueness and finiteness of the MLEs of the parameters based on progressively type-II censored sample data. Recently, Dey and Pradhan [9] made Bayesian inference for the GIE parameters under hybrid random censoring. Ghitany *et al.* [10] established the existence and uniqueness of the MLEs of the parameters for a general class of inverse exponentiated distributions based on complete as well as progressively type-I and type-II censored data.

In this study, statistical inference for both progressive type-II right censored sample and record values from the GIE distribution are investigated. Dey and Dey [8] obtained approximate confidence intervals for the GIE parameters based on progressive censored sample. However, if the sample size is small, the approximate confidence interval may not be adequate. Thus, exact confidence intervals and regions become important when the sample size is small. The method of pivotal quantity are used to construct the confidence intervals and regions for the model parameters. The rest of this paper is organized as follows. In Section 2, an exact confidence interval and an exact confidence region for the parameters are constructed based on progressive type-II right censored sample. In Section 3, two theorems are proposed to obtain the exact confidence interval and region for the parameters based on upper record values. Two numerical examples are presented in Section 4. Some conclusions are made in Section 5.

2. INTERVAL ESTIMATION UNDER PROGRESSIVE TYPE-II CENSORING

Progressive type-II right censoring is of importance in the field of reliability and life testing. Suppose n identical units are placed on a lifetime test. At the time of the i-th failure, r_i surviving units are randomly withdrawn from the experiment, $1 \leq i \leq m$. Thus, if m failures are observed then $r_1 + \cdots + r_m$ units are progressively censored; hence, $n = m + r_1 + \cdots + r_m$. Let $X^{\mathbf{r}}_{1:m:n} < X^{\mathbf{r}}_{2:m:n} < \cdots < X^{\mathbf{r}}_{m:m:n}$ be the progressively censored failure times, where $\mathbf{r} = (r_1, ..., r_m)$ denotes the censoring scheme. As a special case, if $\mathbf{r} = (0, ..., 0)$ where no withdrawals are made, we obtain the ordinary order statistics (Bairamov and Eryılmaz [5]). If $\mathbf{r} = (0, ..., 0, n - m)$, the progressive type-II censoring becomes type-II censoring. For more details see Balakrishnan and Aggarwala [6].

In this section, we will construct the exact confidence interval and region for model parameters by using pivotal quantity method. We will also conduct a simulation study to assess the performance of proposed interval and region.

2.1. Exact confidence interval and region

Suppose that $X^{\mathbf{r}}_{1:m:n} < X^{\mathbf{r}}_{2:m:n} < \cdots < X^{\mathbf{r}}_{m:m:n}$ denote progressively type-II right censored order statistics from a GIE distribution. Let

$$Y^{\mathbf{r}}_{i:m:n} = -\beta \log\left(1 - \exp\left(-\lambda / X^{\mathbf{r}}_{i:m:n}\right)\right), \quad i = 1, 2, ..., m.$$

It can be seen that $Y^{\mathbf{r}}_{1:m:n} < Y^{\mathbf{r}}_{2:m:n} < \cdots < Y^{\mathbf{r}}_{m:m:n}$ are progressively type-II right censored order statistics from a standard exponential distribution. It is well known that, from Thomas and Wilson [21],

$$\begin{aligned}
\pi_1 &= nY^{\mathbf{r}}_{1:m:n} \\
\pi_2 &= (n - r_1 - 1)(Y^{\mathbf{r}}_{2:m:n} - Y^{\mathbf{r}}_{1:m:n}) \\
&\vdots \\
\pi_m &= (n - r_1 - \cdots - r_{m-1} - m + 1)\left(Y^{\mathbf{r}}_{m:m:n} - Y^{\mathbf{r}}_{m-1:m:n}\right)
\end{aligned}$$

are independent and identically distributed as a standard exponential distribution. Hence,

$$\kappa_1 = 2\pi_1 = 2nY^{\mathbf{r}}_{1:m:n}$$

has a chi-squared distribution with 2 degrees of freedom and

$$\varepsilon_1 = 2\sum_{i=2}^{m} \pi_i = 2\left\{\sum_{i=1}^{m} (r_i + 1)\, Y^{\mathbf{r}}_{i:m:n} - nY^{\mathbf{r}}_{1:m:n}\right\}$$

has a chi-squared distribution with $2m-2$ degrees of freedom. It is also clear that ε_1 and κ_1 are independent random variables. Let

$$\xi_1 = \frac{\varepsilon_1}{(m-1)\,\kappa_1} = \frac{\sum_{i=1}^{m}(r_i+1)\,Y^{\mathbf{r}}_{i:m:n} - nY^{\mathbf{r}}_{1:m:n}}{n\,(m-1)\,Y^{\mathbf{r}}_{1:m:n}} \tag{2.1}$$

and

$$\eta_1 = \varepsilon_1 + \kappa_1 = 2\sum_{i=1}^{m}(r_i+1)\,Y^{\mathbf{r}}_{i:m:n}. \tag{2.2}$$

It is easy to show that ξ_1 has an F distribution with $2m-2$ and 2 degrees of freedom and η_1 has a chi-squared distribution with $2m$ degrees of freedom. Furthermore, ξ_1 and η_1 are independent (see Johnson *et al.* [13]).

The following lemma helps us to construct the exact confidence interval for λ and exact joint confidence region for (λ, β).

Lemma 2.1. *Suppose that* $0 < a_1 < a_2 < \cdots < a_m$. *Let*

$$\xi_1(\lambda) = \frac{1}{n\,(m-1)}\sum_{i=1}^{m}(r_i+1)\,\frac{\log\left(1-\exp\left(-\lambda/a_i\right)\right)}{\log\left(1-\exp\left(-\lambda/a_1\right)\right)} - \frac{1}{m-1},$$

where $r_i \geq 0$, $i = 1, 2, ..., m$, *and* $\sum_{i=1}^{m} r_i = n-m$. *Then,* $\xi_1(\lambda)$ *is strictly increasing in* λ *for any* $\lambda > 0$.

Proof: To prove $\xi_1(\lambda)$ is strictly increasing, it suffices to show that the function

$$g(\lambda) = \frac{\log\left(1-\exp\left(-\lambda/a_i\right)\right)}{\log\left(1-\exp\left(-\lambda/a_1\right)\right)}$$

is strictly increasing in λ. The derivative of $g(\lambda)$ is given by

$$g'(\lambda) = \left(\frac{h_1(a_1)}{h_2(a_i)} - \frac{h_1(a_i)}{h_2(a_1)}\right)\left(\frac{1}{h_1(a_1)}\right)^2,$$

where

$$h_1(x) = \log\left(1-\exp\left(-\lambda/x\right)\right)$$

and

$$h_2(x) = x\left(\exp\left(\lambda/x\right)-1\right).$$

If both $h_1(x)$ and $h_2(x)$ are decreasing, it can be said that $\left(\frac{h_1(a_1)}{h_2(a_i)} - \frac{h_1(a_i)}{h_2(a_1)}\right) > 0$ for $a_i > a_1$ and hence $g'(\lambda) > 0$.

It is clear that $h_1(x)$ is strictly decreasing in x. From the second order Taylor polynomial of $\exp(a)$ at $a = 0$, one has the following inequality, for $a < 0$,

$$\exp(a) > a + 1. \tag{2.3}$$

Let $a = -\lambda/x$. Equation (2.3) can be written as

$$1 - \lambda/x - \exp(-\lambda/x) < 0, \quad \text{for} \quad x > 0. \tag{2.4}$$

Note that the first derivative of $h_2(x)$ is

$$h_2'(x) = \exp(\lambda/x)\left[1 - \lambda/x - \exp(-\lambda/x)\right].$$

From Equation (2.4), it is easy to see that $h_2'(x) < 0$ for $x > 0$. That is, $h_2(x)$ is strictly decreasing in x. Hence, $g'(\lambda)$ is positive. This completes the proof. □

Let $F_{\alpha(\delta_1,\delta_2)}$ be the upper α percentile of F distribution with δ_1 and δ_2 degrees of freedom. The following theorem gives an exact confidence interval for the parameter λ.

Theorem 2.1. *Suppose that $X^{\mathbf{r}}_{1:m:n} < X^{\mathbf{r}}_{2:m:n} < \cdots < X^{\mathbf{r}}_{m:m:n}$ is a progressively type-II censored sample from the GIE distribution. Then, for any $0 < \alpha < 1$,*

$$\Big(\varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{1-\alpha/2;2m-2,2}\right), \varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{\alpha/2;2m-2,2}\right)\Big)$$

is a $100(1-\alpha)\%$ confidence interval for λ, where $\varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, t\right)$ is the solution of λ for the equation

$$\frac{1}{n(m-1)}\sum_{i=1}^{m}(r_i+1)\frac{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)}{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{1:m:n}\right)\right)} - \frac{1}{m-1} = t. \tag{2.5}$$

Proof: From Equation (2.1), we know that the pivot

$$\begin{aligned}\xi_1(\lambda) &= \frac{\sum_{i=1}^{m}(r_i+1)Y^{\mathbf{r}}_{i:m:n} - nY^{\mathbf{r}}_{1:m:n}}{n(m-1)Y^{\mathbf{r}}_{1:m:n}} \\ &= \frac{1}{n(m-1)}\sum_{i=1}^{m}(r_i+1)\frac{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)}{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{1:m:n}\right)\right)} - \frac{1}{m-1}\end{aligned}$$

has an F distribution with $2m-2$ and 2 degrees of freedom. By Lemma 2.1, $\xi_1(\lambda)$ is strictly increasing function of λ, and hence, $\xi_1(\lambda) = t$ has a unique solution for any $\lambda > 0$. Thus, for $0 < \alpha < 1$, the event

$$F_{1-\alpha/2;2m-2,2} < \frac{1}{n(m-1)}\sum_{i=1}^{m}(r_i+1)\frac{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)}{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{1:m:n}\right)\right)} - \frac{1}{m-1} < F_{\alpha/2;2m-2,2}$$

is equivalent to the event

$$\varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{1-\alpha/2;2m-2,2}\right) < \lambda \\ < \varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{\alpha/2;2m-2,2}\right).$$

Then, the proof follows. □

Let us now discuss the joint confidence region for (λ, β). Let $\chi^2_{\alpha;\delta}$ denote the upper α percentile of a chi-squared distribution with δ degrees of freedom. An exact joint confidence region for (λ, β) is given in the following theorem.

Theorem 2.2. *Suppose that $X^{\mathbf{r}}_{i:m:n}$, $i = 1, 2, ..., m$, are progressive type-II right censored order statistics from the GIE distribution with censoring scheme* $\mathbf{r}$. *Then for any $0 < \alpha < 1$, a $100(1-\alpha)\%$ joint confidence region for (λ, β) is determined by the following inequalities:*

$$\begin{cases} \varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{(1+\sqrt{1-\alpha})/2;2m-2,2}\right) < \lambda \\ \qquad < \varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{(1-\sqrt{1-\alpha})/2;2m-2,2}\right) \\ -\dfrac{\chi^2_{(1+\sqrt{1-\alpha})/2;2m}}{2\sum_{i=1}^{m}(r_i+1)\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)} < \beta \\ \qquad < -\dfrac{\chi^2_{(1-\sqrt{1-\alpha})/2;2m}}{2\sum_{i=1}^{m}(r_i+1)\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)}, \end{cases}$$

where $\varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, t\right)$ is defined in Equation (2.5).

Proof: From Equation (2.1), we know that the pivot

$$\begin{aligned} \xi_1(\lambda) &= \frac{\sum_{i=1}^{m}(r_i+1)Y^{\mathbf{r}}_{i:m:n} - nY^{\mathbf{r}}_{1:m:n}}{n(m-1)Y^{\mathbf{r}}_{1:m:n}} \\ &= \frac{1}{n(m-1)}\sum_{i=1}^{m}(r_i+1)\frac{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)}{\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{1:m:n}\right)\right)} - \frac{1}{m-1} \end{aligned}$$

has an F distribution with $2m-2$ and 2 degrees of freedom. From Equation (2.2), we also know that

$$\eta_1 = 2\sum_{i=1}^{m}(r_i+1)Y^{\mathbf{r}}_{i:m:n} = -2\beta\sum_{i=1}^{m}(r_i+1)\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)$$

has a chi-squared distribution with $2m$ degrees of freedom, and it is independent

of $\xi_1(\lambda)$. Thus, for $0 < \alpha < 1$, we have

$$\begin{aligned}
&P\Bigg\{\varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{(1+\sqrt{1-\alpha})/2;2m-2,2}\right) < \lambda \\
&\qquad < \varphi_1\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, F_{(1-\sqrt{1-\alpha})/2;2m-2,2}\right), \\
&\qquad -\frac{\chi^2_{(1+\sqrt{1-\alpha})/2;2m}}{2\sum_{i=1}^{m}(r_i+1)\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)} < \beta \\
&\qquad < -\frac{\chi^2_{(1-\sqrt{1-\alpha})/2;2m}}{2\sum_{i=1}^{m}(r_i+1)\log\left(1-\exp\left(-\lambda/X^{\mathbf{r}}_{i:m:n}\right)\right)}\Bigg\} \\
&= P\left(F_{(1+\sqrt{1-\alpha})/2;2m-2,2} < \xi_1 < F_{(1-\sqrt{1-\alpha})/2;2m-2,2}\right) \\
&\qquad P\left(\chi^2_{(1+\sqrt{1-\alpha})/2;2m} < \eta_1 < \chi^2_{(1-\sqrt{1-\alpha})/2;2m}\right) \\
&= \sqrt{1-\alpha}\sqrt{1-\alpha} \\
&= 1-\alpha.
\end{aligned}$$

The proof is completed. □

2.2. Simulation study

The simulation study is performed with 5000 trials to investigate the performance of exact and approximate confidence intervals and confidence regions under progressive censoring. We consider the values of parameters $(\lambda, \beta) = (2, 0.5), (0.5, 2)$ and different combinations of n, m, and censoring schemes $\mathbf{r}$. The approximate intervals are considered as in Dey and Dey [8]. The nominal confidence level is chosen as 95%. The results are given in Table 1 and Table 2. From these tables, one can conclude that both the coverage probabilities of approximate and exact confidence intervals are close to the desired level. The coverage probabilities of exact confidence regions are also close to the nominal level. However, the coverage probabilities of the approximate confidence regions are lower than the nominal level. When the sample size increases, the coverage probability of approximate confidence region reaches to nominal level 95%. During simulation, the authors observed that the MLEs of parameters are not obtained uniquely for different initials values. However, this problem disappeared for the large sample size. In this regards, coverage probability of approximate confidence region works for only large sample. As a conclusion, exact confidence region should be used for the small sample size.

Table 1: Coverage probabilities for the proposed methods and the approximations under progressive censoring when $(\lambda, \beta) = (2, 0.5)$.

n	m	**r**	λ		(λ, β)	
			approx.	exact	approx.	exact
20	10	(1,1,1,1,1,1,1,1,1,1)	0.9482	0.9494	0.8122	0.9484
		(5,0,0,0,0,0,0,0,0,5)	0.9532	0.9500	0.8966	0.9468
		(5,5,0,0,0,0,0,0,0,0)	0.9476	0.9500	0.8966	0.9468
		(0,0,0,0,0,0,0,0,5,5)	0.9540	0.9480	0.8640	0.9420
		(0,0,0,0,5,5,0,0,0,0)	0.9422	0.9492	0.9130	0.9474
		(2,2,1,0,0,0,0,1,2,2)	0.9498	0.9456	0.9504	0.9456
40	20	(1,1,1,1,1,...,1,1)	0.9462	0.9526	0.9294	0.9536
		(10,0,0,0,...,0,10)	0.9556	0.9518	0.9344	0.9460
		(10,10,0,...,0,0)	0.9480	0.9510	0.9380	0.9530
		(0,0,0,0,...,10,10)	0.9586	0.9512	0.9156	0.9534
		(0,...,0,10,10,0,...,0,0)	0.9508	0.9550	0.9432	0.9526
		(2,2,2,2,2,0,...,0,2,2,2,2,2)	0.9520	0.9568	0.9584	0.9562
100	50	(1,1,1,1,1,...,1,1)	0,9506	0.9552	0.8416	0.9574
		(25,0,0,0,...,0,25)	0,9544	0.9456	0.9528	0.9496
		(25,25,0,0,...,0,0)	0,9530	0.9516	0.9524	0.9488
		(0,0,0,0,,...,25,25)	0,9508	0.9528	0.9404	0.9540
		(0,...,0,25,25,0,...,0,0)	0,9464	0.9496	0.9484	0.9500
		(2,...,2,1,0,...,0,1,2,...,2)	0.9484	0.9512	0.9594	0.9534
10	5	(1,1,1,1,1)	0.9452	0.9468	0.8356	0.9446
		(2,1,0,0,2)	0.9486	0.9484	0.9202	0.9468
		(2,2,1,0,0)	0.9436	0.9510	0.9352	0.9484
		(0,0,1,2,2)	0.9530	0.9486	0.9016	0.9458
		(0,2,1,2,0)	0.9384	0.9462	0.9168	0.9468

Table 2: Coverage probabilities for the proposed methods and the approximations under progressive censoring when $(\lambda, \beta) = (0.5, 2)$.

n	m	**r**	λ		(λ, β)	
			approx.	exact	approx.	exact
20	10	(1,1,1,1,1,1,1,1,1,1)	0.9538	0.9556	0.8904	0.9522
		(5,0,0,0,0,0,0,0,0,5)	0.9534	0.9502	0.9570	0.9514
		(5,5,0,0,0,0,0,0,0,0)	0.9540	0.9588	0.9570	0.9514
		(0,0,0,0,0,0,0,0,5,5)	0.9526	0.9530	0.9396	0.9510
		(0,0,0,0,5,5,0,0,0,0)	0.9474	0.9474	0.9530	0.9534
		(2,2,1,0,0,0,0,1,2,2)	0.9452	0.9482	0.9440	0.9482
40	20	(1,1,1,1,1,...,1,1)	0.9460	0.9546	0.8946	0.9482
		(10,0,0,0,...,0,10)	0.9534	0.9502	0.9570	0.9514
		(10,10,0,...,0,0)	0.9540	0.9488	0.9576	0.9538
		(0,0,0,0,...,10,10)	0.9526	0.9472	0.9522	0.9486
		(0,...,0,10,10,0,...,0,0)	0.9504	0.9534	0.9534	0.9508
		(2,2,2,2,2,0,...,0,2,2,2,2,2)	0.9468	0.9422	0.9330	0.9478
100	50	(1,1,1,1,1,...,1,1)	0.9486	0.9488	0.8988	0.9488
		(25,0,0,0,...,0,25)	0.9490	0.9538	0.9530	0.9508
		(25,25,0,0,...,0,0)	0.9510	0.9500	0.9514	0.9470
		(0,0,0,0,,...,25,25)	0.9486	0.9504	0.9560	0.9494
		(0,...,0,25,25,0,...,0,0)	0.9510	0.9492	0.9534	0.9490
		(2,...,2,1,0,...,0,1,2,...,2)	0.9496	0.9514	0.9418	0.9516
10	5	(1,1,1,1,1)	0.9466	0.9524	0.8194	0.9528
		(2,1,0,0,2)	0.9510	0.9474	0.8292	0.9482
		(2,2,1,0,0)	0.9420	0.9538	0.8694	0.9552
		(0,0,1,2,2)	0.9656	0.9544	0.7712	0.9516
		(0,2,1,2,0)	0.9390	0.9504	0.8226	0.9464

3. INTERVAL ESTIMATION UNDER RECORD VALUES

Record values were first introduced by Chandler [7]. A record value is either the largest or the smallest value obtained from a sequence of random variables. Ahsanullah and Nevzorov [3] pointed out that records are very popular because they arise naturally in many fields of studies such as climatology, sports, medicine, traffic, industry and so on. In reliability studies, Lee *et al.* [16] indicated that there are some situations in lifetime testing experiments in which a failure time of a product is recorded if it exceeds all preceding failure times. These recorded failure times are the upper record value sequence. An account on record values can be found in the books by Ahsanullah [2] and Arnold *et al.* [4].

In this section, we will establish the exact confidence interval and region for model parameters based on pivotal quantity method. A simulation study is also conducted to investigate the performance of proposed interval and region.

3.1. Exact confidence interval and region

Let $X_{U(1)} < X_{U(2)} < \cdots < X_{U(m)}$ be the first m upper record values from the GIE distribution. Set

$$W_i = -\beta \log\left(1 - \exp\left(-\lambda / X_{U(i)}\right)\right), \quad i = 1, 2, ..., m.$$

Then, it is easily seen that $W_1 < W_2 < \cdots < W_m$ are the first m upper record values from a standard exponential distribution. Moreover, Arnold *et al.* [4] showed that

$$\begin{aligned} \rho_1 &= W_1 \\ \rho_2 &= W_2 - W_1 \\ &\vdots \\ \rho_n &= W_m - W_{m-1} \end{aligned}$$

are independent and identically distributed random variables from a standard exponential distribution. Hence,

$$\kappa_2 = 2\rho_1 = 2W_1$$

has a chi-squared distribution with 2 degrees of freedom and

$$\varepsilon_2 = 2\sum_{i=2}^{m} \rho_i = 2\left(W_m - W_1\right)$$

has a chi-squared distribution with $2m-2$ degrees of freedom. We can also find that ε_2 and κ_2 independent. Let

$$\xi_2 = \frac{\varepsilon_2}{(m-1)\,\kappa_2} = \frac{1}{m-1}\frac{W_m - W_1}{W_1} \tag{3.1}$$

and

$$\eta_2 = \varepsilon_2 + \kappa_2 = 2W_m. \tag{3.2}$$

It is easy to show that ξ_2 has an F distribution with $2m-2$ and 2 degrees of freedom and η_2 has a chi-squared distribution with $2m$ degrees of freedom. Furthermore, ξ_2 and η_2 are independent.

Lemma 3.1. *Suppose that $0 < a_1 < a_2 < \cdots < a_m$. Let*

$$\begin{aligned}\xi_2(\lambda) &= \frac{1}{m-1}\frac{W_m - W_1}{W_1} \\ &= \frac{1}{m-1}\left(\frac{\log\left(1-\exp\left(-\lambda/a_m\right)\right)}{\log\left(1-\exp\left(-\lambda/a_1\right)\right)} - 1\right).\end{aligned}$$

Then, $\xi_2(\lambda)$ is strictly increasing in λ for any $\lambda > 0$.

Proof: The proof is analogous to that of Lemma 2.1. □

To construct the exact confidence interval for λ based on record values, we have the following theorem.

Theorem 3.1. *Suppose that $X_{U(1)} < X_{U(2)} < \cdots < X_{U(m)}$ are first m upper record values from the GIE distribution. Then, for any $0 < \alpha < 1$,*

$$\Big(\varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{1-\alpha/2;2m-2,2}\right), \varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{\alpha/2;2m-2,2}\right)\Big)$$

is a $100(1-\alpha)\%$ confidence interval for λ, where $\varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, t\right)$ is the solution of λ for the equation

$$\frac{1}{m-1}\left(\frac{\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)}{\log\left(1-\exp\left(-\lambda/X_{U(1)}\right)\right)} - 1\right) = t. \tag{3.3}$$

Proof: From Equation (3.1), we know that the pivot

$$\begin{aligned}\xi_2(\lambda) &= \frac{1}{m-1}\frac{W_m - W_1}{W_1} \\ &= \frac{1}{m-1}\left(\frac{\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)}{\log\left(1-\exp\left(-\lambda/X_{U(1)}\right)\right)} - 1\right)\end{aligned}$$

has an F distribution with $2m-2$ and 2 degrees of freedom. By Lemma 3.1, $\xi_2(\lambda)$ is strictly increasing function of λ, and hence, $\xi_2(\lambda) = t$ has a unique solution for any $\lambda > 0$. Thus, for $0 < \alpha < 1$, the event

$$F_{1-\alpha/2;2m-2,2} < \frac{1}{m-1}\left(\frac{\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)}{\log\left(1-\exp\left(-\lambda/X_{U(1)}\right)\right)} - 1\right) < F_{\alpha/2;2m-2,2}$$

is equivalent to the event

$$\varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{1-\alpha/2;2m-2,2}\right) < \lambda < \varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{\alpha/2;2m-2,2}\right).$$

Then, the proof follows. □

For the joint confidence region for (λ, β) based on record values, we have the following result.

Theorem 3.2. *Suppose that $X_{U(i)}$, $i = 1, 2, ..., m$ are first i-th upper record values from the GIE distribution. Then, for any $0 < \alpha < 1$, a $100(1-\alpha)\%$ joint confidence region for (λ, β) is determined by the following inequalities:*

$$\begin{cases} \varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{(1+\sqrt{1-\alpha})/2;2m-2,2}\right) < \lambda \\ \qquad < \varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{(1-\sqrt{1-\alpha})/2;2m-2,2}\right) \\ -\dfrac{\chi^2_{(1+\sqrt{1-\alpha})/2;2m}}{2\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)} < \beta \\ \qquad < -\dfrac{\chi^2_{(1-\sqrt{1-\alpha})/2;2m}}{2\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)}, \end{cases}$$

where $\varphi_2\left(X^{\mathbf{r}}_{1:m:n}, X^{\mathbf{r}}_{2:m:n}, ..., X^{\mathbf{r}}_{m:m:n}, t\right)$ is defined in Equation (3.3).

Proof: From Equation (3.1), we know that the pivot

$$\xi_2(\lambda) = \frac{1}{m-1}\left(\frac{\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)}{\log\left(1-\exp\left(-\lambda/X_{U(1)}\right)\right)} - 1\right)$$

has an F distribution with $2m-2$ and 2 degrees of freedom. From Equation (3.2), we know that

$$\eta_2 = -2\beta\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right).$$

has a chi-square distribution with $2m$ degrees of freedom, and it is independent of $\xi_2(\lambda)$. For $0 < \alpha < 1$, we have

$$\begin{aligned}
&P\Bigg\{\varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{(1+\sqrt{1-\alpha})/2;2m-2,2}\right) < \lambda \\
&\qquad\qquad < \varphi_2\left(X_{U(1)}, X_{U(2)}, ..., X_{U(m)}, F_{(1-\sqrt{1-\alpha})/2;2m-2,2}\right), \\
&\qquad -\frac{\chi^2_{(1+\sqrt{1-\alpha})/2;2m}}{2\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)} < \beta < -\frac{\chi^2_{(1-\sqrt{1-\alpha})/2;2m}}{2\log\left(1-\exp\left(-\lambda/X_{U(m)}\right)\right)}\Bigg\} \\
=&P\left(F_{(1+\sqrt{1-\alpha})/2;2m-2,2} < \xi_2 < F_{(1-\sqrt{1-\alpha})/2;2m-2,2}\right) \\
&\qquad P\left(\chi^2_{(1+\sqrt{1-\alpha})/2;2m} < \eta_2 < \chi^2_{(1-\sqrt{1-\alpha})/2;2m}\right) \\
=&\sqrt{1-\alpha}\sqrt{1-\alpha} \\
=&1-\alpha.
\end{aligned}$$

□

3.2. Simulation study

It is important to examine how well our proposed method works for constructing confidence interval and region. We consider the values of parameters $(\lambda, \beta) = (2, 0.5), (0.5, 2)$ and different values of m. For each case, we simulated 5000 upper record samples from the GIE distribution. The nominal confidence level is chosen as 95%. The results are given in Table 3. From this table, one can see that the exact confidence intervals and regions have desired coverage probability for small and large sample sizes. As a conclusion, the proposed methods work well.

Table 3: Coverage probability of exact confidence interval and confidence region based on upper record values when $(\lambda, \beta) = (2, 0.5), (0.5, 2)$.

m	$(\lambda,\beta)=(2,0.5)$		$(\lambda,\beta)=(0.5,2)$	
	λ	(λ,β)	λ	(λ,β)
2	0.9502	0.9520	0.9566	0.9540
3	0.9502	0.9488	0.9466	0.9446
4	0.9474	0.9546	0.9548	0.9504
5	0.9510	0.9500	0.9454	0.9498
6	0.9476	0.9526	0.9546	0.9528
7	0.9548	0.9606	0.9502	0.9512
8	0.9522	0.9606	0.9540	0.9548
9	0.9518	0.9604	0.9514	0.9498
10	0.9498	0.9578	0.9512	0.9516
11	0.9476	0.9570	0.9522	0.9526
12	0.9532	0.9600	0.9478	0.9488
13	0.9478	0.9560	0.9472	0.9468
14	0.9494	0.9524	0.9488	0.9452
15	0.9498	0.9490	0.9488	0.9520

4. ILLUSTRATIVE EXAMPLES

To illustrate the use of our proposed estimation method, the following two examples are discussed.

Example 4.1 (*Progressively Type-II Censored Data*)**.** We apply the proposed interval estimation methods to the polished window strengths data set presented in Abouammoh and Alshingiti [1]. Dey and Dey [8] indicated that the GIE distribution is acceptable for these data. For the purposes of illustrating the estimation methods discussed in this paper, we adopt the progressively type-II censored sample with $n = 31$ and $m = 11$ which was generated from this data set by Dey and Dey [8]. The progressively censored data are reported in Table 4.

To obtain a 95% confidence interval for λ, we need the percentiles

$$F_{0.025;22,2} = 39.4479 \qquad \text{and} \qquad F_{0.975;22,2} = 0.2242.$$

Then, we can solve Equation (2.5) and get the following values

$$\varphi_1\left(x^{\mathbf{r}}_{1:m:n}, x^{\mathbf{r}}_{2:m:n}, ..., x^{\mathbf{r}}_{m:m:n}, F_{0.975;22,2}\right) = 81.8086,$$

and

$$\varphi_1\left(x^{\mathbf{r}}_{1:m:n}, x^{\mathbf{r}}_{2:m:n}, ..., x^{\mathbf{r}}_{m:m:n}, F_{0.025;22,2}\right) = 401.0639.$$

By Theorem 2.1, the 95% confidence interval for λ is obtained as (81.8086, 401.0639).

Table 4: Progressively type-II censored data based on window strength data.

i	1	2	3	4	5	6
r_i	0	0	0	0	0	0
$x^{\mathbf{r}}_{i:m:n}$	18.83	20.8	21.657	23.03	23.23	24.05
i	7	8	9	10	11	
r_i	0	0	0	0	20	
$x^{\mathbf{r}}_{i:m:n}$	24.321	25.5	25.52	25.8	26.69	

Furthermore, to obtain a 95% joint confidence region for (λ, β), we need the percentiles

$$F_{0.9873;22,2} = 0.1825, \qquad F_{0.0127;22,2} = 78.4361,$$

$$\chi^2_{.9873;24} = 9.8824, \qquad \text{and} \qquad \chi^2_{.0127;24} = 39.4099.$$

By Theorem 2.2, the 95% confidence region for (λ, β) is determined by the following two inequalities:

$$71.9165 < \lambda < 458.4111$$

and

$$-\frac{9.8824}{2\sum_{i=1}^{11}(r_i+1)\log\left(1-\exp\left(-\lambda/x^{\mathbf{r}}_{i:m:n}\right)\right)} < \beta < -\frac{39.4099}{2\sum_{i=1}^{11}(r_i+1)\log\left(1-\exp\left(-\lambda/x^{\mathbf{r}}_{i:m:n}\right)\right)}.$$

Figure 1 shows the 95% joint confidence region for (λ, β) based on progressively type-II censored data given in Table 1. It can be seen that the region is large when λ is large.

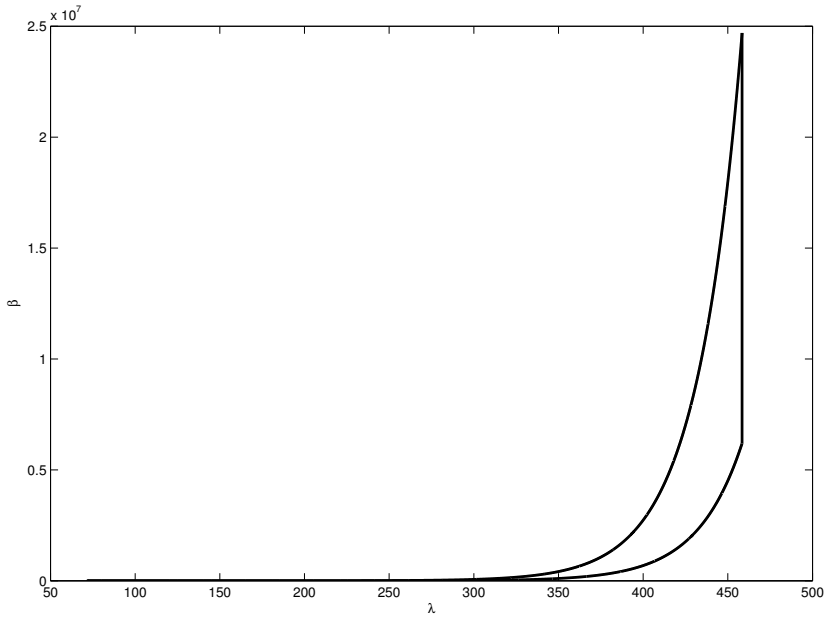

Figure 1: A 95% joint confidence region for (λ, β) based on progressively type-II censored data given in Table 4.

Example 4.2 (*Record Value Data*)**.** To illustrate the use of the interval estimation based on records, we analyze one real data set. Lawless [15, p.3] presented 11 times to breakdown of electrical insulating fluid subjected to 30 kilovolts. The data, under a logarithm transformation, is 2.836, 3.120, 3.045, 5.169, 4.934, 4.970, 3.018, 3.770, 5.272, 3.856, 2.046. Luckett [18] extracted the $m = 4$ upper record values from this data set and indicated that the GIE distribution is acceptable for this data set. The upper record value data are presented in Table 5.

Table 5: Upper record values based on breakdown of electrical insulating fluid data.

i	1	2	3	4
$x_{u(i)}$	2.836	3.120	5.169	5.272

To obtain a 95% confidence interval for λ, we need the percentiles

$$F_{0.025;6,2} = 39.3315 \qquad \text{and} \qquad F_{0.975;6,2} = 0.1377.$$

By Theorem 3.1, we have the following results.

$$\varphi_2\left(x_{u(1)}, x_{u(2)}, ..., x_{u(10)}, F_{0.975;6,2}\right) = 0.8644,$$

and

$$\varphi_2\left(x_{u(1)}, x_{u(2)}, ..., x_{u(10)}, F_{0.025;6,2}\right) = 29.3207.$$

That is, the 95% confidence interval for λ is $(0.8644, 29.3207)$.

To obtain a 95% joint confidence region for (λ, β), we need the percentiles

$$F_{0.9873;6,2} = 0.1013, \qquad F_{0.0127;6,2} = 78.3196,$$

$$\chi^2_{.9873;8} = 1.7670, \qquad \text{and} \qquad \chi^2_{.0127;8} = 19.4433.$$

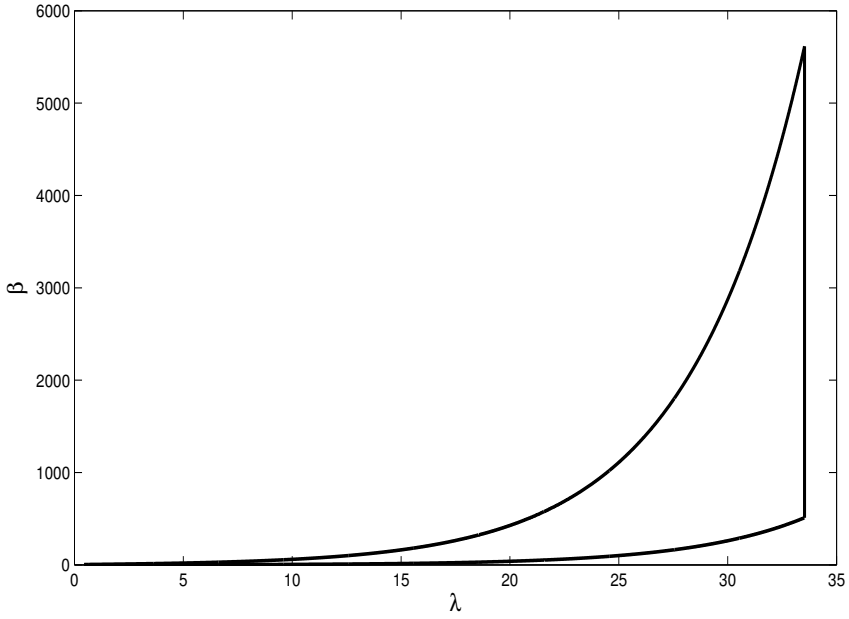

Figure 2: A 95% joint confidence region for (λ, β) based on record values given in Table 5.

By Theorem 3.2, a 95% confidence region for (λ, β) is determined by the following two inequalities:

$$0.4484 < \lambda < 33.5289$$

and

$$-\frac{1.7670}{2\log\left(1-\exp\left(-\lambda/5.272\right)\right)} < \beta < -\frac{19.4433}{2\log\left(1-\exp\left(-\lambda/5.272\right)\right)}.$$

Figure 2 shows the 95% joint confidence region for (λ, β) based on record data given in Table 5. It is easy to see that the region is large when λ is large.

5. CONCLUSIONS

Progressive censoring and record values have received attention in the past few decades. The GIE distribution is a new lifetime distribution and can be widely used in reliability applications. The main purpose of this study is to investigate the interval estimation of parameters of the GIE distribution based on progressive type-II censored sample and record values, respectively. We provide four theorems based on the method of pivotal quantity to construct the exact confidence intervals and regions for the parameters. The simulation results show that the proposed methods perform well. Two numerical examples are used to illustrate the proposed methods.

ACKNOWLEDGMENTS

This work is funded by BAP office of Selcuk University. We also acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Abouammoh, A.M. and Alshingiti, A.M. (2009). Reliability estimation of generalized inverted exponential distribution, *Journal of Statistical Computation and Simulation*, **79**, 1301–1315.

[2] Ahsanullah, M. (1995). *Introduction to Record Statistics*, NOVA Science Publishers Inc., Huntington, New York.

[3] Ahsanullah, M. and Nevzorov, V.B. (2011). *Record Statistics*. In: "International Encyclopedia of Statistical Science" (M. Lovric, Ed.), Springer, Berlin.

[4] Arnold, B.C.; Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*, Wiley, New York.

[5] Bairamov, I. and Eryilmaz, S. (2006). Spacings, exceedances and concomitants in progressive type II censoring scheme, *Journal of Statistical Planning and Inference*, **136**, 527–536.

[6] Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods and Applications*, Birkhäuser, Boston.

[7] Chandler, K.N. (1952). The distribution and frequency of record values, *Journal of the Royal Statistical Society: Series B*, **14**, 220–228.

[8] Dey, S. and Dey, T. (2014). On progressively censored generalized inverted exponential distribution, *Journal of Applied Statistics*, **41**, 2557–2576.

[9] Dey, S. and Pradhan, B. (2014). Generalized inverted exponential distribution under hybrid censoring, *Statistical Methodology*, **18**, 101–114.

[10] Ghitany, M.E.; Tuan, V.K. and Balakrishnan, N. (2014). Likelihood estimation for a general class of inverse exponentiated distributions based on complete and progressively censored data, *Journal of Statistical Computation and Simulation*, **84**, 96–106.

[11] Gupta, R.D. and Kundu, D. (1999). Generalized exponential distribution, *Australian and New Zealand Journal of Statistics*, **41**, 173–183.

[12] Gupta, R.D. and Kundu, D. (2001). Generalized exponential distribution: Different methods of estimations, *Journal of Statistical Computation and Simulation*, **69**, 315–338.

[13] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1, 2nd edition, Wiley, New York.

[14] Krishna, H. and Kumar, H. (2013). Reliability estimation in generalized inverted exponential distribution with progressively type II censored sample, *Journal of Statistical Computation and Simulation*, **83**, 1007–1019.

[15] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

[16] Lee, H.M.; Lee, W.C.; Lei, C.L. and Wu, J.W. (2011). Computational procedure of assessing lifetime performance index of Weibull lifetime products with the upper record values, *Mathematics and Computers in Simulation*, **81**, 1177–1189.

[17] Leiblein, J. and Zelen, M. (1956). Statistical investigation of the fatigue life of deep-groove ball bearings, *Journal of Research of the National Bureau of Standards*, **57**, 273–316.

[18] Luckett, D.J. (2013). *Statistical Inference Based on Upper Record Values*, Honors & Senior Thesis, The College of William and Mary, Department of Mathematics.

[19] Raqab, M.Z. and Madi, M.T. (2005). Bayesian inference for the generalized exponential distribution, *Journal of Statistical Computation and Simulation*, **75**, 841–852.

[20] Singh, S.K.; Singh, U. and Kumar, M. (2013). Estimation of parameters of generalized inverted exponential distribution for progressive type-II censored sample with binomial removals, *Journal of Probability and Statistics*, **2013**, Article ID 183652, DOI: 10.1155/2013/183652.

[21] Thomas, D.R. and Wilson, W.M. (1972). Linear order statistics estimation for the two parameter Weibull and extreme value distributions from type-II progressively censored samples, *Technometrics*, **14**, 679–691.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 449–474

A NEW ROBUST PARTIAL LEAST SQUARES REGRESSION METHOD BASED ON A ROBUST AND AN EFFICIENT ADAPTIVE REWEIGHTED ESTIMATOR OF COVARIANCE

Authors: Esra Polat
– Department of Statistics, Hacettepe University, Turkey
espolat@hacettepe.edu.tr

Suleyman Gunay
– Department of Statistics, Hacettepe University, Turkey
sgunay@hacettepe.edu.tr

Received: November 2016 Revised: July 2017 Accepted: July 2017

Abstract:

- Partial Least Squares Regression (PLSR) is a linear regression technique developed as an incomplete or "partial" version of the least squares estimator of regression, applicable when high or perfect multicollinearity is present in the predictor variables. Robust methods are introduced to reduce or remove the effects of outlying data points. In the previous studies it has been showed that if the sample covariance matrix is properly robustified further robustification of the linear regression steps of the PLS1 algorithm (PLSR with univariate response variable) becomes unnecessary. Therefore, we propose a new robust PLSR method based on robustification of the covariance matrix used in classical PLS1 algorithm. We select a reweighted estimator of covariance, in which the Minimum Covariance Determinant as initial estimator is used, with weights adaptively computed from the data. We compare this new robust PLSR method with classical PLSR and four other well-known robust PLSR methods. Both simulation results and the analysis of a real data set show the effectiveness and robustness of the new proposed robust PLSR method.

Key-Words:

- *efficient estimation; Minimum Covariance Determinant (MCD); partial least squares regression; robust covariance matrix; robust estimation.*

AMS Subject Classification:

- 62F35, 62H12, 62J05.

1. INTRODUCTION

Classical PLSR is a well-established technique in multivariate data analysis. It is used to model the linear relation between a set of regressors and a set of response variables, which can then be used to predict the value of the response variables for a new sample. A typical example is multivariate calibration where the x-variables are spectra and the y-variables are the concentrations of certain constituents. Since classical PLSR is known to be severely affected by the presence of outliers in the data or deviations from normality, several PLSR methods with robust behaviour towards data contamination have been proposed (Hubert and Vanden Branden, 2003; Liebmann *et al.*, 2010). NIPALS and SIMPLS are the popular algorithms for PLSR and they are very sensitive to outliers in the dataset. For univariate or multivariate response variable several robustified versions of these algorithms have already been proposed (González *et al.*, 2009).

The two main strategies in the literature for robust PLSR are (1) the downweighting of outliers and (2) robust estimation of the covariance matrix. The early approaches for robust regression by downweighting of outliers are considered semi-robust: they had, for instance, non-robust initial weights or the weights were not resistant to leverage points (Hubert and Vanden Branden, 2003). Based on the first strategy, for example, Wakeling and Macfie (1992) worked with the PLS with multivariate response variables (which will be called PLS2) and their idea was to replace the set of regressions involved in the standard PLS2 algorithm by M estimates based on weighted regressions. Griep *et al.* (1995) compared least median of squares (LMS), Siegel's repeated median (RM) and iterative reweighted least squares (IRLS) for PLS with univariate response variable (PLS1 algorithm), but these methods are not resistant to high leverage outliers (González *et al.*, 2009). Based on the second strategy, a robust covariance estimation, the robust PLSR methods provide resistance to all types of outliers including leverage points (Hubert and Vanden Branden, 2003). For instance, Gil and Romera (1998) proposed a robust PLSR method based on statistical procedures for covariance matrix robustification for PLS1 algorithm. They selected the well-known Stahel–Donoho estimator (SDE) (Gil and Romera, 1998). Since SIMPLS is based on the empirical cross-covariance matrix between the y-variables and the x-variables and on linear Least Squares (LS) regression, the results are affected by outliers in the data set. Hence, Hubert and Vanden Branden (2003) have been suggested a robust version of this method called RSIMPLS that it is used in case of both univariate and multivariate response variables. A robust method RSIMPLS starts by applying ROBPCA on the x- and y-variables in order to replace the covariance matrices S_{xy} and S_x by robust estimates and then proceeds analogously to the SIMPLS algorithm. A robust regression method (ROBPCA regression) is performed in the second stage. ROBPCA is a robust PCA method which combines projection pursuit ideas with Minimum Covariance Determinant (MCD) covariance estimation in lower dimensions (Engelen *et al.*, 2004; Hubert and Vanden Branden, 2003).

Serneels *et al.* (2005) proposed a method called as Partial Robust M (PRM) regression that it is conceptually different from the other robust PLSR methods: instead of robust partial least squares, a partial robust regression estimator was proposed. This method uses SIMPLS algorithm and it could be used in case of univariate response. In this method, with an appropriately chosen weighting scheme, both vertical outliers and leverage points were downweighted (Serneels *et al.*, 2005). As the name suggests, it is a partial version of the robust M-regression. In an iterative scheme, weights ranging between zero and one are calculated to reduce the influence of deviating observations in the y space as well as in the space of the regressor variables. PRM is very efficient in terms of computational cost and statistical properties (Liebmann *et al.*, 2010). González *et al.* (2009) also concentrated in the case of univariate response (PLS1) and showed that if the sample covariance matrix is properly robustified the PLS1 algorithm will be robust and, therefore, further robustification of the linear regression steps of the PLS1 algorithm is unnecessary (González *et al.*, 2009).

In this paper, we concentrate in the case of univariate response (PLS1) and we present a procedure which applies the standard PLS1 algorithm to a robust covariance matrix similar to Gil and Romera (1998) and González *et al.* (2009) studies. In our study, we estimate the covariance matrix used in PLS1 algorithm robustly by using '*an adaptive reweighted estimator of covariance using Minimum Covariance Determinant (MCD) estimators in the first step as robust initial estimators of location and covariance*'.

The rest of the paper is organized as follows. Section 2 reviews briefly the PLS1 algorithm (PLS with univariate response variable). Section 3 presents the new proposed robust PLSR method 'PLS-ARWMCD'. Section 4 contains a simulation study where the performance of the new robust PLSR method is compared to classical PLSR method and other four robust PLSR methods existing in robust PLSR literature. Section 5 illustrates the performance of the new proposed robust PLSR method 'PLS-ARWMCD' in a well known set of real data in robust PLSR literature. Finally, Section 6 collects some conclusions.

2. THE CLASSICAL PLS1 ALGORITHM

It is supposed that we have a sample of size n of a $1+p$ dimensional vector $\boldsymbol{z} = (\boldsymbol{y}, \boldsymbol{X})'$ which could be decomposed as a set of p independent variables, x and a univariate response variable y. Throughout this paper, matrices are denoted by bold capital letters and vectors are denoted by bold lowercase letters. Let $\boldsymbol{S_z}$, be the sample covariance matrix of $\boldsymbol{z}$, consisting of the elements $\boldsymbol{S_z} = \begin{bmatrix} \boldsymbol{s_y^2} & \boldsymbol{s'_{y,X}} \\ \boldsymbol{s_{y,X}} & \boldsymbol{S_X} \end{bmatrix}$, where $\boldsymbol{s_{y,X}}$ is the $p \times 1$ vector of covariances between y and the x variables.

The aim of this study is to estimate the linear regression $\hat{\boldsymbol{y}} = \hat{\boldsymbol{\beta}}'\boldsymbol{x}$, and it is assumed that the response variable can be linearly explained by a set of a components $\boldsymbol{t}_1, ..., \boldsymbol{t}_k$ with $k << p$, which are linear functions of the x variables. Hence, calling $\boldsymbol{X}$ the $n \times p$ data matrix of the independent variables, and $\boldsymbol{x}_i'$ to its ith row, the following model showed by (2.1) and (2.2) holds (González *et al.*, 2009):

$$\boldsymbol{x}_i = \boldsymbol{P}\boldsymbol{t}_i + \boldsymbol{\varepsilon}_i \,, \tag{2.1}$$

$$\boldsymbol{y}_i = \boldsymbol{q}'\boldsymbol{t}_i + \boldsymbol{\eta}_i \,. \tag{2.2}$$

Here, $\boldsymbol{P}$ is the $p \times k$ matrix of the loadings of the vector $\boldsymbol{t}_i = (t_{i1}, ..., t_{ik})'$ and $\boldsymbol{q}$ is the k-dimensional vector of the y-loadings. The vectors $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\eta}_i$ have zero mean, follow normal distributions and are uncorrelated. The component matrix $\boldsymbol{T} = (\boldsymbol{t}_1, ..., \boldsymbol{t}_k)'$ is not directly observed and should be estimated. Then, it can be shown that the maximum likelihood estimation of the $\boldsymbol{T}$ matrix is given as in (2.3) (González *et al.*, 2009):

$$\boldsymbol{T} = \boldsymbol{X}\boldsymbol{W}_k \,. \tag{2.3}$$

Here, the loading matrix $\boldsymbol{W}_k = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_k]$ is the $p \times k$ matrix of coefficients and the vectors $\boldsymbol{w}_i$, $1 \leq i < k$ are the solution of (2.4) under the constraint in (2.5) with $\boldsymbol{w_1} \boldsymbol{\alpha} \boldsymbol{s_{y,x}}$. Consequently, we can conclude that components $(\boldsymbol{t}_1, ..., \boldsymbol{t}_k)$ are orthogonal (González *et al.*, 2009):

$$\boldsymbol{w}_i = \arg\max_{w} cov^2\left(\boldsymbol{X}\boldsymbol{w}, \boldsymbol{y}\right) \,, \tag{2.4}$$

$$\boldsymbol{w}'\boldsymbol{w} = 1 \quad \texttt{and} \quad \boldsymbol{w}_i'\boldsymbol{S_x}\boldsymbol{w}_j = 0 \;\text{ for }\; 1 \leq j < i \,. \tag{2.5}$$

It can be shown that vectors $\boldsymbol{w}_i$ are found as the eigenvectors linked to the largest eigenvalues of the matrix is given as in (2.6):

$$\left(\boldsymbol{I} - \boldsymbol{P}_x(i)\right) \boldsymbol{s_{y,x}} \boldsymbol{s_{y,x}}' \,. \tag{2.6}$$

$\boldsymbol{P}_x(i)$ is the projection matrix on the space spanned by $\boldsymbol{S_X}\boldsymbol{W}_i$, given by $\boldsymbol{P}_x(i) = (\boldsymbol{S_x}\boldsymbol{W}_i)\left[(\boldsymbol{S_x}\boldsymbol{W}_i)'(\boldsymbol{S_x}\boldsymbol{W}_i)\right]^{-1}(\boldsymbol{S_x}\boldsymbol{W}_i)'$. From these results it is easy to see that the vectors $\boldsymbol{w}_i$ can be computed recursively as in below:

$$\boldsymbol{w_1} \boldsymbol{\alpha} \boldsymbol{s_{y,x}} \,, \tag{2.7}$$

$$\boldsymbol{w_{i+1}} \boldsymbol{\alpha} \boldsymbol{s_{y,x}} - \boldsymbol{S_x}\boldsymbol{W}_i \left(\boldsymbol{W}_i'\boldsymbol{S_x}\boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i'\boldsymbol{s_{y,x}}, \quad 1 \leq i < k \,. \tag{2.8}$$

It could be mentioned that by using the expressions given by (2.7) and (2.8), it is not necessary to calculate the PLS components $\boldsymbol{t}_i$. In each step of the

algorithm, $\boldsymbol{w}_{i+1}$ only depends on the value of the i previous vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_i$, on $\boldsymbol{S}_x$ and on $\boldsymbol{s_{y,x}}$. Moreover, as $\boldsymbol{w}_1$ only depends on $\boldsymbol{s_{y,x}}$, the calculation of $\boldsymbol{W}$ is completely fixed by the values of $\boldsymbol{S}_x$ and $\boldsymbol{s_{y,x}}$. Finally, as the regression coefficients in (2.2) are uncorrelated, due to the uncorrelation of the t variables, it is easy to see that the regression coefficients $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{PLS}}$ are given by (2.9) (González *et al.*, 2009):

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{PLS}} = \boldsymbol{W}_k \left(\boldsymbol{W}_k' \boldsymbol{S_x} \boldsymbol{W}_k\right)^{-1} \boldsymbol{W}_k' \boldsymbol{s_{y,x}} \,. \tag{2.9}$$

The application of this algorithm can be seen as a two step procedure: (1) the weights $\boldsymbol{w}_i$, defining the new orthogonal regressor $\boldsymbol{t}_i$, are computed with (2.7) and (2.8) by using the covariance matrix of the observations; (2) the y-loadings $\boldsymbol{q}_i$ are computed by regressing y on individual regressor $\boldsymbol{t}_i$. As it is shown in (2.9) these two steps depend only on the covariance matrix of the observations and it may be thought that if this matrix is properly robustified the procedure will be robust (González *et al.*, 2009).

3. THE NEW PROPOSED ROBUST PLSR METHOD

In this section, the new robust PLSR method, which we proposed based on '*an adaptive reweighted estimator of covariance using MCD estimators in the first step as robust initial estimators of location and covariance*', will be introduced. This adaptive reweighted estimator of covariance will be used in order to robustify the sample covariance matrix, $\boldsymbol{S_z}$, in the PLS1 algorithm. Hence, while defining this estimator, the equations are examined on $\boldsymbol{z}_i = (y_i, \boldsymbol{x}_i),\ i = 1, ..., n \in \Re^{p'}$, here, $p' = p + 1$. In this method, the MCD estimator is calculated by well-known 'FAST-MCD' algorithm. Hence, in this section, firstly, information about MCD estimator and operation of the FAST-MCD algorithm will be given.

Besides high outlier resistance, if robust multivariate estimators are to be of practical use in statistical inference they should offer a reasonable efficiency under the normal model and a manageable asymptotic distribution. However, Minimum Volume Ellipsoid (MVE) and MCD estimators are not in this category. Gervini (2003) stated that by taking care of both robustness and efficiency considerations, the best choice seems to be a two-stage procedure. In this procedure, firstly, a highly robust but perhaps inefficient estimator is computed, which is used for detecting outliers and computing the sample mean and covariance of the 'cleaned' data set as in Rousseeuw and Van Zomeren (1990). This procedure consists of discarding those observations whose Mahalanobis distances exceed a certain fix threshold value. In the previous studies, the MVE was commonly used as initial estimator for these procedures. However, Rousseeuw and Van Driessen (1999) have proposed an algorithm for calculating MCD estimator,

although this algorithm does not guarantee that the exact estimator is found, it is faster and more accurate than previously existing algorithms even for very large data sets ($n >> p' = p + 1$). This fact, added to its $1/\sqrt{n}$ rate of convergence, seems to point to the MCD method using the FAST-MCD algorithm as the current best choice in comparison to MVE for initial estimator of a two-step procedure (Gervini, 2003).

MCD method, proposed by Rousseeuw (1984), is searching for those h data points for which the determinant of the classical covariance matrix is minimal. Hence, the MCD estimators of location and covariance will be the mean and covariance matrix of these h data points, respectively. The calculation of MCD estimation is not simple. Let $\boldsymbol{z}_i' = (y_i, \boldsymbol{x}_i)'$, $i = 1, ..., n$ be an unified data set. The MCD estimator can only be applied to data sets where the number of observations is larger than the number of variables ($n > p' = p + 1$). The reason is that if $p' > n$ then also $p' > h$, and the covariance matrix of any h data points will always be singular, leading to a determinant of zero. Thus, each subset of h data points would lead to the smallest possible determinant, resulting in a non-unique solution (Filzmoser *et al.*, 2009; Polat, 2014).

FAST-MCD algorithm could deal with a sample size n in the tens of thousands. FAST-MCD finds the exact solution for small data sets and it is faster and more accurate than previously existing algorithms, even for very large data sets. Rousseeuw and Van Driessen (1999) suggested to use FAST-MCD algorithm in order to estimate location and covariance as considering the its statistical efficiency and fastness in computation (Rousseeuw and Van Driessen, 1999). In FAST-MCD algorithm as the raw MCD estimators of location and covariance are reweighted in order to improve the finite sample efficiency, they are called as Reweighted Minimum Covariance Determinant (RMCD) estimators (Hubert and Vanden Branden, 2003; Moller *et al.*, 2005).

3.1. Construction of the FAST-MCD algorithm

3.1.1. Basic theorem and the C-step for the FAST-MCD algorithm

A key step of the FAST-MCD algorithm is the fact that starting from any approximation to the MCD, it is possible to compute another approximation with an even lower determinant. 'C-step' procedure, which is used in FAST-MCD algorithm, given in following Theorem 3.1 (Rousseeuw and Van Driessen, 1999).

Theorem 3.1. *Since* $\boldsymbol{z}_i' = (y_i, \boldsymbol{x}_i)'$, $i = 1, ..., n$ *consider a data set* $\boldsymbol{Z}_n = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$ *of* $p' = p + 1$*-variate observations. Let a set of observations that de-*

fined as $H_1 \subset \{1, ..., n\}$ *with* $|H_1| = h$. *Here,* H_1 *shows the subset of* h *observations having the lowest determinant. Hence, as the location and covariance for subset of* h *observations* $\hat{\boldsymbol{\mu}}_1 := (1/h) \sum_{i \in H_1} \boldsymbol{z}_i$ *and* $\hat{\boldsymbol{\Sigma}}_1 := (1/h) \sum_{i \in H_1} (\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_1) \cdot (\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_1)'$, *respectively, if* $det\left(\hat{\boldsymbol{\Sigma}}_1\right) \neq 0$ *then the relative distances are defined as* $d_1(i) := \sqrt{(\boldsymbol{z}_1 - \hat{\boldsymbol{\mu}}_1)' \hat{\boldsymbol{\Sigma}}_1^{-1} (\boldsymbol{z}_1 - \hat{\boldsymbol{\mu}}_1)}$, $i = 1, ..., n$. *Then, a set of observations* H_2 *is taken such that* $\{d_1(i); i \in H_2\} := \{(d_1)_{1:n}, ..., (d_1)_{h:n}\}$ *where* $(d_1)_{1:n} \leq (d_1)_{2:n} \leq \cdots \leq (d_1)_{n:n}$ *are the ordered distances, and* $\hat{\boldsymbol{\mu}}_2$ *and* $\hat{\boldsymbol{\Sigma}}_2$ *are computed based on* H_2. *Then,* $det\left(\hat{\boldsymbol{\Sigma}}_2\right) \leq det\left(\hat{\boldsymbol{\Sigma}}_1\right)$ *with equality if and only if* $\hat{\boldsymbol{\mu}}_2 = \hat{\boldsymbol{\mu}}_1$ *and* $\hat{\boldsymbol{\Sigma}}_2 = \hat{\boldsymbol{\Sigma}}_1$ *(Polat, 2014; Rousseeuw and Van Driessen, 1999).*

If $det\left(\hat{\boldsymbol{\Sigma}}_1\right) > 0$, applying the Theorem 3.1 yields $\hat{\boldsymbol{\Sigma}}_2$ with $det\left(\hat{\boldsymbol{\Sigma}}_2\right) \leq det\left(\hat{\boldsymbol{\Sigma}}_1\right)$. In FAST-MCD algorithm the construction in Theorem 3.1 is referred to as 'C-step', where 'C' can be taken to stand for 'covariance' since $\hat{\boldsymbol{\Sigma}}_2$ is the covariance matrix of H_2, or for 'concentration' since we concentrate on the h observations with smallest distances, and $\hat{\boldsymbol{\Sigma}}_2$ is more concentrated (has a lower determinant) than $\hat{\boldsymbol{\Sigma}}_1$ (Rousseeuw and Van Driessen, 1999).

Repeating C-steps yields an iteration process. If $det\left(\hat{\boldsymbol{\Sigma}}_2\right) = 0$ or $det\left(\hat{\boldsymbol{\Sigma}}_2\right) = det\left(\hat{\boldsymbol{\Sigma}}_1\right)$ we stop; otherwise we run another C-step yielding $det\left(\hat{\boldsymbol{\Sigma}}_3\right)$, and so on. The sequence $det\left(\hat{\boldsymbol{\Sigma}}_1\right) \geq det\left(\hat{\boldsymbol{\Sigma}}_2\right) \geq det\left(\hat{\boldsymbol{\Sigma}}_3\right) \geq ...$ is nonnegative and hence must converge. In fact, since there are only finitely many h subsets there must be an index m such that $det\left(\hat{\boldsymbol{\Sigma}}_m\right) = 0$ or $det\left(\hat{\boldsymbol{\Sigma}}_m\right) = det\left(\hat{\boldsymbol{\Sigma}}_{m-1}\right)$, hence convergence is reached. In practice, m is often below 10. Afterwards, running the C-step on $\left(\hat{\boldsymbol{\mu}}_m, \hat{\boldsymbol{\Sigma}}_m\right)$ no longer reduces the determinant. This is not sufficient for $det\left(\hat{\boldsymbol{\Sigma}}_m\right)$ to be the global minimum of the MCD objective function, but it is a necessary condition (Rousseeuw and Van Driessen, 1999). Thus, Theorem 3.1 provides a partial idea for an algorithm: '*Take many initial choices of* H_1 *and apply C-steps to each until convergence, and keep the solution with lowest determinant*'. However, several things must be decided to make this idea operational: how to generate sets H_1 to begin with, how many H_1 are needed, how to avoid duplication of work since several H_1 may yield the same solution, can't we do with fewer C-steps, what about large sample sizes, and so on. These matters will be discussed in the next sections.

3.1.2. Creating initial subsets H_1

In order to apply the algorithmic idea given in the previous section, it must be decided how to construct the initial subsets H_1. For this purpose, first of all, a

random $(p'+1)$-subset J must be drawn according to method given in Rousseeuw and Van Driessen (1999) study and then $\hat{\boldsymbol{\mu}}_0 := ave(J)$ and $\hat{\boldsymbol{\Sigma}}_0 := cov(J)$ must be computed. If $det\left(\hat{\boldsymbol{\Sigma}}_0\right) = 0$ then extend J by adding another random observation, and continue adding observations until $det\left(\hat{\boldsymbol{\Sigma}}_0\right) > 0$. Then compute the distances $d_0^2(i) := (\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_0)' \hat{\boldsymbol{\Sigma}}_0^{-1} (\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_0)$ for $i = 1, ..., n$. Sort them into $d_0(\pi(1)) \leq ... \leq d_0(\pi(n))$ and put $H_1 := \{\pi(1), ..., \pi(h)\}$. Rousseeuw and Van Driessen (1999) mentioned that it would be useless to draw fewer than $p'+1$ points, since then $\hat{\boldsymbol{\Sigma}}_0$ is always singular (Polat, 2014; Rousseeuw and Van Driessen, 1999).

3.1.3. Selective iteration

Each C-step calculates a covariance matrix, its determinant, and all relative distances. Therefore, reducing the number of C-steps would improve the speed. Rousseeuw and Van Driessen (1999) mentioned that often the distinction between good (robust) solutions and bad solutions already becomes visible after two or three C-steps. Moreover, they proposed to take only two C-steps from each initial subsample, select the 10 different subsets with the lowest determinants, and only for these 10 to continue taking C-steps until convergence (Rousseeuw and Van Driessen, 1999).

3.1.4. Nested extensions

For a small sample size n, the above algorithm, which was mentioned in Section 3.1.1, does not take much time. But when n grows, the computation time increases, mainly due to the n distances that needed to be calculated each time. To avoid doing all the computations in the entire data set, Rousseeuw and Van Driessen (1999) considered a special structure. When $n > 1500$, the algorithm generates a nested system of subsets which looks like in Figure 1, where the arrows mean 'is a subset of'.

In Figure 1 the five subsets of size 300 do not overlap, and together they form the merged set of size 1500, which in turn is a proper subset of the data set of size n. Since the method showed in Figure 1 work with two stages, 'nested' name is used. To construct the Figure 1 the algorithm draws 1500 observations, one by one, without replacement. The first 300 observations, that it encounters, are put in the first subset, and so on. Because of this mechanism each subset of size 300 is roughly representative for the data set, and the merged set with 1500 cases even more representative. When $n < 600$ the algorithm operates as in the

previous Section 3.1.1. However, when $n \geq 1500$ Figure 1 is used (Rousseeuw and Van Driessen, 1999).

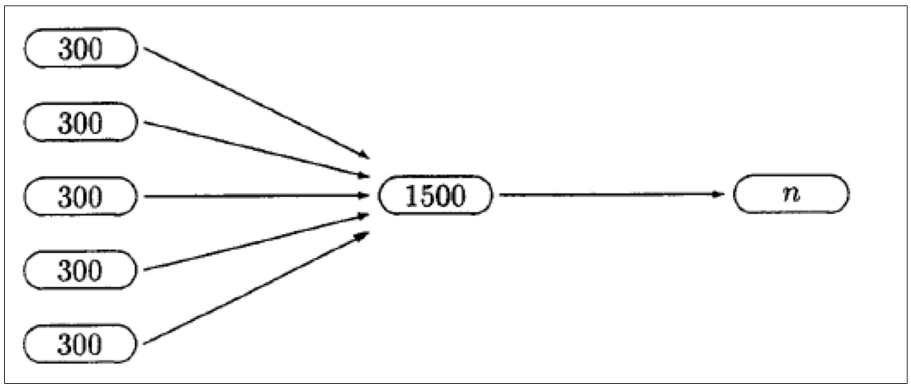

Figure 1: Nested system of subsets generated by the FAST-MCD algorithm.

3.2. The implementation of the FAST-MCD algorithm

Combining all the components of the preceding sections yields the FAST-MCD algorithm. The steps of the algorithm for $p' = p + 1$ dimensional unified vector $\boldsymbol{z}_i' = (y_i, \boldsymbol{x}_i)'$, $i = 1, ..., n$ are given as in below (Polat, 2014; Rousseeuw and Van Driessen, 1999).

Step 1: The MCD estimates can resist $(n - h)$ outliers, hence the number h (or equivalently the proportion $\alpha = h/n$) determines the robustness of the estimator. The default h value is $[(n + p' + 1)/2]$ in FAST-MCD algorithm and the highest resistance towards contamination is achieved by taking this value. However, the user may choose any integer h with $[(n + p' + 1)/2] \leq h < n$. When a large proportion of contamination is presumed in data set, h should thus be chosen $h = [0.5n]$ with $\alpha = 0.5$. Otherwise if it is exact that the data contains less than 25% of contamination, which is usually the case, a good compromise between breakdown value and statistical efficiency is obtained by putting $h = [0.75n]$ (Polat, 2014; Rousseeuw and Van Driessen, 1999).

Step 2: From here on $h < n$ and $p' \geq 2$. If n is small (say, $n < 600$) then,

- repeat (say) 500 times:
  - construct an initial h-subset H_1 using method in Section 3.1.2, i.e. starting from a random $(p' + 1)$-subset,
  - carry out two C-steps described in Section 3.1.1;

- for the 10 results with lowest $det\left(\hat{\boldsymbol{\Sigma}}_3\right)$:
  - carry out C-steps until convergence;
- report the solution $\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$ with the lowest $det\left(\hat{\boldsymbol{\Sigma}}\right)$.

Step 3: If n is larger (say, $n \geq 600$) then,

- construct up to five disjoint random subsets of size n_{sub} according to Section 3.1.4 (say, subsets of size $n_{sub} = 300$);
- inside each subset, repeat $500/5 = 100$ times:
  - construct an initial subset H_1 of size $h_{sub} = [n_{sub}\,(h/n)]$,
  - carry out two C-steps, using n_{sub} and h_{sub},
  - keep the 10 best results $\left(\hat{\boldsymbol{\mu}}_{sub}, \hat{\boldsymbol{\Sigma}}_{sub}\right)$;
- pool the subsets, yielding the merged set (say, of size $n_{merged} = 1500$);
- in the merged set, repeat for each of the 50 solutions $\left(\hat{\boldsymbol{\mu}}_{sub}, \hat{\boldsymbol{\Sigma}}_{sub}\right)$:
  - carry out two C-steps, using n_{merged} and $h_{merged} = [n_{merged}\,(h/n)]$,
  - keep the 10 best results $\left(\hat{\boldsymbol{\mu}}_{merged}, \hat{\boldsymbol{\Sigma}}_{merged}\right)$;
- in the full data set, repeat for the m_{full} best results:
  - take several C-steps, using n and h,
  - keep the best final result $\left(\hat{\boldsymbol{\mu}}_{full}, \hat{\boldsymbol{\Sigma}}_{full}\right)$.

Here, m_{full} and the number of C-steps (preferably, until convergence) depend on how large the data set is (Polat, 2014; Rousseeuw and Van Driessen, 1999).

This algorithm called as FAST-MCD. It is affine equivariant: when the data are translated or subjected to a linear transformation, the resulting $\left(\hat{\boldsymbol{\mu}}_{full}, \hat{\boldsymbol{\Sigma}}_{full}\right)$ will transform accordingly. For convenience, the computer program contains two more steps (Rousseeuw and Van Driessen, 1999):

Step 4: In order to obtain consistency when the data come from a multivariate normal distribution, $\hat{\boldsymbol{\mu}}_{\text{MCD}} = \hat{\boldsymbol{\mu}}_{full}$ and $\hat{\boldsymbol{\Sigma}}_{\text{MCD}} = \dfrac{\text{med}_i\; d^2_{\left(\hat{\boldsymbol{\mu}}_{full}, \hat{\boldsymbol{\Sigma}}_{full}\right)}(i)}{\chi^2_{p',0.5}} \hat{\boldsymbol{\Sigma}}_{full}$ are putted.

Step 5: In order to obtain 'one-step reweighted' estimates, each observation is reweighted as in (3.1). Hence, by using these weights, the RMCD estimators are obtained as in (3.2):

$$w_i = \begin{cases} 1, \text{ if } \left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}}\right)' \hat{\boldsymbol{\Sigma}}^{-1}_{\text{MCD}} \left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}}\right) \leq \chi^2_{p',0.975}\,, \\ 0, \text{ otherwise}\,. \end{cases} \tag{3.1}$$

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\mathrm{RMCD}} &= \frac{\sum_{i=1}^{n} w_i \boldsymbol{z}_i}{\sum_{i=1}^{n} w_i}, \\ \hat{\boldsymbol{\Sigma}}_{\mathrm{RMCD}} &= \frac{\sum_{i=1}^{n} w_i \left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\mathrm{RMCD}}\right)\left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\mathrm{RMCD}}\right)'}{\sum_{i=1}^{n} w_i}. \end{aligned} \tag{3.2}$$

The FAST-MCD algorithm code named as '*mcdcov*' could be found in MATLAB LIBRA Toolbox which is written by Verboven and Hubert (2005). The implementation of mcdcov function could be given briefly as in below (Polat, 2014; Verboven and Hubert, 2005):

- The data set contains n observations and $p' = p + 1$ variables. When $n < 600$, the algorithm analyzes the data set as a whole. When the data set is analyzed as a whole, a subsample of $p' + 1$ observations is taken, of which of them the mean and covariance matrix are calculated. The h observations with smallest relative distances are used to calculate the next mean and covariance matrix, and this cycle is repeated two C-step times. FAST-MCD algorithm is a resampling algorithm. 500 subsets of size $p' + 1$ out of n are drawn randomly. Afterwards, the 10 best solutions (means and corresponding covariance matrices) are used as starting values for the final iteration. The number of the subsets is chosen as '500' to ensure a high probability of sampling at least one clean subset. These iterations stop when two subsequent determinants become equal. At most three C-step iteration are done. The solution with smallest determinant (location and covariance) is retained.

- However, when $n \geq 600$ (whether $n < 1500$ or not), the algorithm does part of the calculations on (at most) 5 non-overlapping subsets of (roughly) 1500 observations. In this case, the algorithm functions in three stages.
  - Stage 1: For each H_1 subsample in each subset, two C-steps iterations are carried out in that subset. In this stage, 5 subsets and 500 subsamples are chosen. For each subset, the 10 best solutions (location and covariance) are stored.
  - Stage 2: Then the subsets are pooled, yielding a merged set with at most 1500 observations. If n is large, the merged set is a proper subset of the entire data set. In this merged set, each of these (at most 50) best solutions $\left(\hat{\boldsymbol{\mu}}_{sub}, \hat{\boldsymbol{\Sigma}}_{sub}\right)$ of Stage 1 are used as starting values for C-step iterations. In this stage, starting from each $\left(\hat{\boldsymbol{\mu}}_{sub}, \hat{\boldsymbol{\Sigma}}_{sub}\right)$, it is continued taking C-steps by using all 1500 observations in the merged set. Also here, the 10 best solutions $\left(\hat{\boldsymbol{\mu}}_{merged}, \hat{\boldsymbol{\Sigma}}_{merged}\right)$ are stored.
  - Stage 3: This stage depends on n, the total number of observations in the data set. Finally, each of these 10 solutions is extended to the

full data set in the same way and the best $\left(\hat{\boldsymbol{\mu}}_{full}, \hat{\boldsymbol{\Sigma}}_{full}\right)$ solution is obtained. Since the final computations are carried out in the entire data set, they take more time when n increases. Rousseeuw and Van Driessen (1999) mentioned that the number of initial solutions $\left(\hat{\boldsymbol{\mu}}_{merged}, \hat{\boldsymbol{\Sigma}}_{merged}\right)$ and/or the number of C-steps in the full data set could be limited in order to speed up the algorithm as n becomes large (Rousseeuw and Van Driessen, 1999; Verboven and Hubert, 2005). Therefore, the default values of '*mcdcov*' function are: If $n \leq 5000$, all 10 preliminary solutions are iterated. If $n > 5000$, only the best preliminary solution is iterated. The number of iterations decreases to 1 according to $n \times p$. If $n \times p \leq 100000$, the number of C-steps take on the full data set in the Stage 3 iterate three times, whereas for $n \times p > 1000000$ only one iteration step is taken.

In the next section, information about 'a robust and efficient adaptive reweighted covariance estimator', which was proposed in Gervini (2003), will be given. This robust covariance estimator is constructed by using MCD estimators in the first step as robust initial estimators of location and covariance.

3.3. A robust and efficient adaptive reweighted estimator of covariance

In the context of linear regression, many estimators have been proposed that aim to reconcile high efficiency and robustness. Overall, if one wants to take care of both robustness and efficiency considerations, the best choice seems to be a two-stage procedure. Gervini (2003) proposed essentially an improvement over Rousseeuw and Van Zomeren (1990). It consists of a reweighted one-step estimator that uses adaptive threshold values. This adaptive reweighting scheme is able to maintain the outlier resistance of the initial estimator in breakdown and bias and, at the same time, attain 100% efficiency at the normal distribution. This kind of adaptive reweighting was first proposed in Gervini (2002) for the linear regression model. In Gervini (2003), this idea is extended and an adaptive method is proposed for multivariate location and covariance estimation.

Given a sample $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$ in $\Re^{p'}$ with $p' = p+1$ and initial robust estimators of location and covariance $\left(\hat{\boldsymbol{\mu}}_{0n}, \hat{\boldsymbol{\Sigma}}_{0n}\right)$ consider the Mahalanobis distances given in (3.3) (Gervini, 2003; Polat, 2014):

$$d_i := d\left(\boldsymbol{z}_i, \hat{\boldsymbol{\mu}}_{0n}, \hat{\boldsymbol{\Sigma}}_{0n}\right) = \left\{\left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{0n}\right)' \hat{\boldsymbol{\Sigma}}_{0n}^{-1} \left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{0n}\right)\right\}^{1/2} . \tag{3.3}$$

An outlier will typically have a larger Mahalanobis distance than a 'good' observation. If one assumes a normal distribution, d_i^2 is approximately $\chi_{p'}^2$ distributed and it is reasonable to suspect of those observations with, for instance,

$d_i^2 \geq \chi^2_{p',0.975}$. What Rousseeuw and Van Zomeren (1990) propose is to skip those outlying observations and compute the sample mean and covariance matrix of the rest of the data, obtaining in this way new estimators $\left(\hat{\boldsymbol{\mu}}_{1n}, \hat{\boldsymbol{\Sigma}}_{1n}\right)$ (Gervini, 2003; Polat, 2014).

Since the MCD method calculated by FAST-MCD algorithm is improved as a good alternative to MVE method, Gervini (2003) stated that MCD estimators could be used as the initial robust estimators of location and covariance in the 'adaptive reweighted' method. Hence, in this study, in 'adaptive reweighted' method using the MCD estimators $\left(\hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}}\right)$ as initial robust estimators of location and covariance $\left(\hat{\boldsymbol{\mu}}_{0n}, \hat{\boldsymbol{\Sigma}}_{0n}\right)$, the obtained robust location and covariance estimators $\left(\hat{\boldsymbol{\mu}}_{1n}, \hat{\boldsymbol{\Sigma}}_{1n}\right)$ are called as 'Adaptive Reweighted Minimum Covariance Determinant/ARWMCD' estimators $\left(\hat{\boldsymbol{\mu}}_{\text{ARWMCD}}, \hat{\boldsymbol{\Sigma}}_{\text{ARWMCD}}\right)$(Gervini, 2003; Polat, 2014).

This reweighting step given in Gervini (2003) is known to improve the efficiency of the initial estimator while retaining (most of) its robustness. However, the threshold value $\chi^2_{p',0.975}$ is an arbitrary number. For large data sets a considerable number of observations have to be discarded from the analysis even if they follow the normal model. One way to avoid this problem is to increase the threshold value to another arbitrary fix number, however, in this case the bias of the reweighted estimator will be affected. Hence, a better alternative is to use 'an adaptive threshold value' that increases with n if the data is 'clean' but remains bounded if there are outliers in the sample. Gervini (2003), proposed a method of constructing such adaptive threshold values. Let (3.4) be the empirical distribution of the squared Mahalanobis distances (Gervini, 2003; Polat, 2014):

$$G_n(u) = \frac{1}{n}\sum_{i=1}^{n} I\left(d^2\left(\boldsymbol{z}_i, \hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}}\right) \leq u\right) . \tag{3.4}$$

Let $G_{p'}(u)$ be the $\chi^2_{p'}$ distribution function. For a normally distributed sample it is expected to G_n to converge to $G_{p'}$. Therefore, a way to detect outliers is to compare the tails of G_n with the tails of $G_{p'}$. If $\eta = \chi^2_{p',1-\alpha}$ for a certain small α, say $\alpha = 0.025$, (3.5) is defined (Gervini, 2003; Polat, 2014)

$$\alpha_n = \sup_{u \geq \eta} \left\{G_{p'}(u) - G_n(u)\right\}^+ , \tag{3.5}$$

where $\{\cdot\}^+$ indicates the positive part. This α_n can be regarded as a measure of outliers in the sample. Since a negative difference would not indicate presence of outliers, it is only taken into account positive differences in (3.5). If $d^2_{(i)}$ denotes the ith order statistic of the squared Mahalanobis distances and $i_0 = max\left\{i : d^2_{(i)} < \eta\right\}$, then (3.5) comes down to as in (3.6) (Gervini, 2003; Po-

lat, 2014):

$$\alpha_n = \max_{i>i_0} \left\{ G_{p'}(d^2_{(i)}) - \frac{i-1}{n} \right\}^+ . \tag{3.6}$$

Those observations corresponding to the largest $\lfloor \alpha_n n \rfloor$ distances are considered as outliers and eliminated in the reweighting step. Here $\lfloor a \rfloor$, is the largest integer that is less than or equal to a. The cut-off value is then defined as in (3.7) where as usual $G_n^{-1}(u) = \min\{s : G_n(s) \geq u\}$. Note that $c_n = d^2_{(i_n)}$ with $i_n = n - \lfloor \alpha_n n \rfloor$ and that $i_n > i_0$ as a consequence of the definition of α_n. Therefore, $c_n > \eta$ (Gervini, 2003; Polat, 2014):

$$c_n = G_n^{-1}(1 - \alpha_n) . \tag{3.7}$$

To define the reweighted estimator, weights of the form in (3.8) are used (Gervini, 2003; Polat, 2014):

$$w_{in} = w\left(\frac{d^2\left(\boldsymbol{z}_i, \hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}}\right)}{c_n} \right) . \tag{3.8}$$

Here, the weight function that satisfies $(\boldsymbol{W})$ $w : [0, \infty) \to [0, 1]$ is non-increasing, $w(0) = 1$, $w(u) > 0$ for $u \in [0, 1)$ and $w(u) = 0$ for $u \in [1, \infty)$. The simplest choice among those functions satisfying $(\boldsymbol{W})$ is the hard-rejection function $w(u) = I(u < 1)$ which is the one most commonly used in the practice.

Once weights in (3.8) are computed, the one-step reweighted estimators $\left(\hat{\boldsymbol{\mu}}_{\text{ARWMCD}}, \hat{\boldsymbol{\Sigma}}_{\text{ARWMCD}}\right)$ are defined as in (3.9) and (3.10) (Gervini, 2003; Polat, 2014):

$$\hat{\boldsymbol{\mu}}_{\text{ARWMCD}} = \frac{\sum_{i=1}^n w_{in} \boldsymbol{z}_i}{\sum_{i=1}^n w_{in}} , \tag{3.9}$$

$$\hat{\boldsymbol{\Sigma}}_{\text{ARWMCD}} = \frac{\sum_{i=1}^n w_{in} \left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\text{ARWMCD}}\right)\left(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_{\text{ARWMCD}}\right)'}{\sum_{i=1}^n w_{in}} . \tag{3.10}$$

It is clear that under appropriate conditions, the threshold values in (3.7) will tend to infinity under the multivariate normal model and then (3.9) and (3.10) will be asymptotically equivalent to the common sample mean and covariance, and thus attain full asymptotic efficiency (Gervini, 2003).

Finally, in this study, first of all, by using robust covariance estimator $\hat{\boldsymbol{\Sigma}}_{\text{ARWMCD}}$ that it is given in (3.10), the robust covariance estimator $\hat{\boldsymbol{S}}_z$ of $\boldsymbol{S}_z = \begin{bmatrix} \boldsymbol{s}_{\boldsymbol{y}}^2 & \boldsymbol{s}'_{\boldsymbol{y},\boldsymbol{X}} \\ \boldsymbol{s}_{\boldsymbol{y},\boldsymbol{X}} & \boldsymbol{S}_{\boldsymbol{X}} \end{bmatrix}$ is obtained. Then, by using robust covariance estimator $\hat{\boldsymbol{S}}_z$ in

the alternative definition of PLS1 algorithm given between (2.7)–(2.9), a new robust PLSR method called 'PLS-ARWMCD' is proposed. The steps of the PLS-ARWMCD algorithm could be given as in (3.11) (Polat, 2014):

$$
\begin{aligned}
& \boldsymbol{w_1}\boldsymbol{\alpha}\hat{\boldsymbol{s}}_{y,x}\,, \\
& \boldsymbol{w}_{i+1}\boldsymbol{\alpha}\hat{\boldsymbol{s}}_{y,x} - \hat{\boldsymbol{S}}_x\boldsymbol{W}_i\left(\boldsymbol{W}_i'\hat{\boldsymbol{S}}_x\boldsymbol{W}_i\right)^{-1}\boldsymbol{W}_i'\hat{\boldsymbol{s}}_{y,x}, \quad 1 \le i < k\,, \\
& \hat{\boldsymbol{\beta}}_k^{\mathbf{PLS-ARWMCD}} = \boldsymbol{W}_k\left(\boldsymbol{W}_k'\hat{\boldsymbol{S}}_x\boldsymbol{W}_k\right)^{-1}\boldsymbol{W}_k'\hat{\boldsymbol{s}}_{y,x}\,.
\end{aligned}
\tag{3.11}
$$

Here, the robust covariance estimations $\hat{\boldsymbol{s}}_{y,x}$ and $\hat{\boldsymbol{S}}_x$ are obtained by decomposing the robust covariance estimation of unified data set $\boldsymbol{z}_i' = (y_i, \boldsymbol{x}_i)'$, $i = 1, ..., n$, which is calculated by ARWMCD estimator, as in $\hat{\boldsymbol{S}}_z = \begin{bmatrix} \hat{\boldsymbol{s}}_{\boldsymbol{y}}^2 & \hat{\boldsymbol{s}}_{\boldsymbol{y},\boldsymbol{X}}' \\ \hat{\boldsymbol{s}}_{\boldsymbol{y},\boldsymbol{X}} & \hat{\boldsymbol{S}}_{\boldsymbol{X}} \end{bmatrix}$ (Polat, 2014).

4. SIMULATION STUDY

In this section, the new proposed robust PLS-ARWMCD method is compared with other four robust PLSR methods RSIMPLS (Hubert and Vanden Branden, 2003), PRM (Serneels *et al.*, 2005), PLS-SD (Gil and Romera, 1998), PLS-KurSD (González *et al.*, 2009) and the classical PLSR method in order to validate the good properties of the new PLSR robustification. The new proposed robust PLS-ARWMCD method and the other five methods (including the classical method) are compared in terms of efficiency, goodness-of-fit (GOF) and predictive ability by performing a simulation study on uncontaminated and contaminated data sets.

According to the initial models given in (2.1) and (2.2), and following a simulation design similar as the one described in González *et al.* (2009), we have generated the data sets as in (4.1):

$$
\begin{aligned}
& \boldsymbol{T} \sim N_2\left(\mathbf{0}_2, \boldsymbol{\Sigma}_t\right)\,, \\
& \boldsymbol{X} = \boldsymbol{T}\boldsymbol{I}_{2,p} + N_p\left(\mathbf{0}_p, 0.1\boldsymbol{I}_p\right)\,, \\
& \boldsymbol{y} = \boldsymbol{T}\boldsymbol{A}_{2,1} + N(0,1)\,.
\end{aligned}
\tag{4.1}
$$

Here, $(\boldsymbol{I}_{k,p})_{i,j} = 1$, for $i = j$ and $(\boldsymbol{I}_{k,p})_{i,j} = 0$, otherwise; $\boldsymbol{I}_p$ is $p \times p$ dimensional identity matrix; $\mathbf{0}_2 = (0,0)'$ is a two-dimensional vector of zeros and $\boldsymbol{A}_{2,1} = (1,1)'$ is a two-dimensional vector of ones and $\boldsymbol{T}$ is the $n \times 2$ dimensional component matrix. Furthermore, we select $n = 200$, $p = 5$, $k = 2$ and we set $\boldsymbol{\Sigma}_t = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$.

In this simulation study, the performance of the new proposed robust PLS-ARWMCD method is compared with other four robust PLSR methods existing in the literature and the classical method in the presence of five types of contamination.

1. Bad leverage points, which occurs when an observation is far away from the regression hyperplane while its projection onto the regression hyperplane falls outside the large majority of the projected observations (good observations):

$$\text{Bad Leverage Points}: \begin{array}{l} \boldsymbol{T}_\epsilon \sim N_2\left(\mathbf{10}_2, \boldsymbol{\Sigma}_t\right), \\ \boldsymbol{X}_\epsilon = \boldsymbol{T}_\epsilon \boldsymbol{I}_{2,p} + N_p\left(\mathbf{0}_p, 0.1\boldsymbol{I}_p\right). \end{array}$$

2. Vertical outliers, which are observations with large distance from the hyperplane but with projections within the large majority of the projected observations:

$$\text{Vertical outliers}: \ \boldsymbol{y}_\epsilon = \boldsymbol{T}\boldsymbol{A}_{2,1} + N(10, 0.1).$$

3. Good leverage points, which are observations located in the vicinity of the hyperplane but far away from the cluster of the large majority of the observations:

$$\text{Good Leverage Points}: \begin{array}{l} \boldsymbol{T}_\epsilon \sim N_2\left(\mathbf{10}_2, \boldsymbol{\Sigma}_t\right), \\ \boldsymbol{X}_\epsilon = \boldsymbol{T}_\epsilon \boldsymbol{I}_{2,p} + N_p\left(\left(\mathbf{0}_2, \mathbf{10}_{p-2}\right), 0.1\boldsymbol{I}_p\right). \end{array}$$

4. Concentrated Outliers, which are clusters of bad leverage points:

$$\text{Concentrated Outliers}: \begin{array}{l} \boldsymbol{T}_\epsilon \sim N_2\left(\mathbf{10}_2, \boldsymbol{\Sigma}_t\right), \\ \boldsymbol{X}_\epsilon = \boldsymbol{T}_\epsilon \boldsymbol{I}_{2,p} + N_p\left(\mathbf{10}_p, 0.001\boldsymbol{I}_p\right). \end{array}$$

5. Orthogonal outliers, which were first used by Hubert and Vanden Branden (2003). They have the property that they lie far from the t-space, but they become regular observations after projection in the t-space. Hence they will not badly influence the computation of the regression parameters, but they might influence the loadings:

$$\text{Orthogonal outliers}: \ \boldsymbol{X}_\epsilon = \boldsymbol{T}\boldsymbol{I}_{2,p} + N_p\left(\left(\mathbf{0}_2, \mathbf{10}_{p-2}\right), 0.1\boldsymbol{I}_p\right).$$

For each situation, $m = 1000$ data sets were generated. The efficiency of the considered methods is evaluated by means of the MSE of the estimated regression parameters $\hat{\boldsymbol{\beta}}$ that is defined as in (4.2). Moreover, it is clear that the true parameter vector is determined as $\boldsymbol{\beta}_{p,1} = \boldsymbol{I}'_{p,2}\boldsymbol{A}_{2,1}$. Here, $\hat{\boldsymbol{\beta}}_k^{(l)}$ denotes the estimated parameter based on k components in the lth simulation. The MSE indicates to what extent the slope and intercept are correctly estimated. Therefore, the aim is to obtain a MSE value close to zero (Engelen *et al.*, 2004):

$$\text{MSE}_k\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{m}\sum_{l=1}^{m}\left\|\hat{\boldsymbol{\beta}}_k^{(l)} - \boldsymbol{\beta}\right\|^2. \tag{4.2}$$

Furthermore, we are interested on how well the methods fit the regular data points. Because of the simulation settings, we know exactly their indices as we store in the set G_r. Then, the GOF criterion is defined as in (4.3). Here $r_{i,k}$ is the residual of the ith observation when k components are computed. The objective is to obtain a GOF value close to 1 (Engelen *et al.*, 2004):

$$\text{GOF}_k = 1 - \frac{\underset{i\in G_r}{var}\,(r_i,k)}{\underset{i\in G_r}{var}\,(y_i)}\,. \tag{4.3}$$

The predictive ability of the methods could be measured by means of the Root Mean Squared Error (RMSE). First a test set G_t of uncontaminated data points with size $n_t = 100$ is generated and then (4.4) is computed. Here, $\hat{y}_{i,k}$ is the predicted y-value of observation i from the test set when the regression parameter estimates are based on the training set (X, Y) of size n and k components are retained in the model (Engelen *et al.*, 2004):

$$\text{RMSE}_k = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}(y_i - \hat{y}_{i,k})^2}\,. \tag{4.4}$$

After $m = 1000$ replications, the mean angle (denoted by mean(angle)) between the estimated slope $\hat{\boldsymbol{\beta}}_{[\boldsymbol{y}_\epsilon,\boldsymbol{X}_\epsilon],\boldsymbol{k}}$ and the true slope β are also evaluated and included in the simulation results (González *et al.*, 2009; Hubert and Vanden Branden, 2003).

The results obtained according to simulation settings given in above for the data sets uncontaminated and contaminated by replacing first 10% and 20% of the observations by different types of outliers: bad leverage points, vertical outliers, good leverage points, concentrated outliers and orthogonal outliers. The simulation results for the $n = 200$, $p = 5$, $k = 2$ when the proportion of outliers is 10% given in Table 1. The simulation results for the same simulation setting when the proportion of outliers is 20% given in Table 2.

Table 1 shows that in case of no contamination is added the new proposed robust PLS-ARWMCD method and the four robust PLSR methods existing in the literature (RSIMPLS, PRM, PLS-SD, PLS-KurSD) have nearly close performance to classical PLSR method in terms of efficiency, fitting to data and predictive ability. However, when the data set is contaminated by different types of outliers, the four robust PLSR methods existing in literature and the new proposed robust PLSR method outperform the classical PLSR method especially in terms of efficiency and predictive ability. Especially when the data contain bad leverage points or concentrated outliers, the performance of classical PLSR method in terms of efficiency, fitting to data and predictive ability is much lower than the new proposed robust PLS-ARWMCD method. The mean angle values

between the estimated slope and the true slope for the classical PLSR method are also higher than the new proposed robust PLS-ARWMCD method for these two types of outliers.

Table 1: The sample size is $n = 200$, $p = 5$ and $k = 2$, the proportion of outliers is 10%.

	PLSR	RSIMPLS	PRM	PLS- SD	PLS-KurSD	**PLS-ARWMCD**
No Contamination						
MSE	0.0092	0.0111	0.0101	0.0104	0.01	0.0105
GOF	0.8312	0.8308	0.8308	0.8308	0.8309	0.8308
RMSE	1.0961	1.0974	1.0969	1.0973	1.0968	1.0974
Mean(angle)	0.0446	0.0519	0.0462	0.0492	0.0477	0.0491
Bad Leverage Points						
MSE	1.7184	0.0115	0.0688	0.0969	0.0109	0.0104
GOF	0.2585	0.8306	0.8177	0.8098	0.8307	0.8309
RMSE	2.2892	1.0996	1.1413	1.1654	1.0996	1.0989
Mean(angle)	1.1403	0.0515	0.0796	0.0943	0.0496	0.0478
Vertical Outliers						
MSE	0.0489	0.0107	0.0121	0.0118	0.0113	0.0106
GOF	0.817	0.8295	0.8294	0.8296	0.8299	0.83
RMSE	1.1384	1.0989	1.0998	1.0998	1.0987	1.0981
Mean(angle)	0.113	0.0467	0.0516	0.0526	0.0507	0.0485
Good Leverage Points						
MSE	1.0282	0.0118	1.0346	0.0162	0.0109	0.0103
GOF	0.6988	0.8307	0.7721	0.8305	0.8307	0.8309
RMSE	1.4658	1.0996	1.2789	1.1002	1.0996	1.0988
Mean(angle)	0.768	0.053	0.7027	0.0583	0.0496	0.0476
Concentrated Outliers						
MSE	1.9646	0.0118	1.6318	0.03	0.0109	0.0104
GOF	0.5093	0.8307	0.7503	0.8281	0.8307	0.8309
RMSE	1.8671	1.0996	1.3228	1.1078	1.0996	1.0989
Mean(angle)	1.1031	0.0529	0.6964	0.0707	0.0496	0.0478
Orthogonal Outliers						
MSE	0.1815	0.0137	0.1341	0.0107	0.0109	0.0103
GOF	0.7847	0.8295	0.7988	0.8298	0.8298	0.83
RMSE	1.2316	1.1002	1.1917	1.0996	1.0997	1.099
Mean(angle)	0.2821	0.0575	0.2323	0.0494	0.0503	0.0488

Table 1 shows that there are no big differences between the classical method and the robust PLSR methods (including the new proposed robust PLS-ARWMCD method) in terms of fitting to data for the contaminated data sets with the exception of good leverage points, bad leverage points and concentrated outliers. It could be mentioned that for all the types of outliers the new proposed robust PLS-ARWMCD method comes to the forefront with robust RSIMPLS and PLS-KurSD methods existing in the literature especially in terms of efficiency. Overall, for all the types of outliers the new proposed robust PLS-ARWMCD method with more or less differences gives better results than robust PRM method in terms

of efficiency, fitting to data and predictive ability. Furthermore, for all types of outliers but especially when the data contain bad leverage points or concentrated outliers, the new proposed robust PLS-ARWMCD method outperforms robust PLS-SD method in terms of efficiency, fitting to data and predictive ability. The mean angle values between the estimated slope and the true slope for the PLS-AWMCD method is also lower than the classical method (as expected) and all the other four robust PLSR methods for all types of outliers with the exception of vertical outliers. Because when the data contain vertical outliers RSIMPLS gives somewhat lower mean(angle) value than the PLS-ARWMCD method.

Table 2 shows that for all the types of outliers with the exception of vertical outliers when the proportion of outliers increases, it is seen that the performance of robust PRM method decreases especially in terms of efficiency and predictive ability, moreover, the mean angle values between the estimated slope and the true slope for this robust method is also higher than the other four robust PLSR methods (including the new proposed robust PLS-ARWMCD method). Especially when the proportion of concentrated outliers or orthogonal outliers is 20% in the data set, PRM method performs worse even than classical PLSR method in terms of MSE, GOF, RMSE and mean(angle) criterions. Furthermore, when there is 20% proportion of good leverage points PRM performs worse than classical PLSR method in terms of efficiency.

It is clear that when there is 20% proportion of bad leverage points or vertical outliers in the data set, the new proposed robust PLS-ARWMCD method, robust PLS-KurSD and RSIMPLS methods existing in the literature are the three forefront methods in terms of efficiency and predictive ability. Moreover, the mean angle values between the estimated slope and the true slope of these three robust methods are also lower than the robust PRM and PLS-SD methods for these two types of outliers. The concentrated outliers are the hardest type of outliers to cope with. It is seen that when there is 20% proportion of bad leverage points or concentrated outliers in the data set, the new proposed robust PLS-ARWMCD method performs better than both robust PLS-SD and PRM methods existing in the literature in terms of efficiency, fitting to data and predictive ability. Furthermore, PLS-ARWMCD method's mean angle values are also lower than these two robust methods for these two types of outliers. It could be mentioned that when the proportion of outliers in the data set gets a high-level as 20%, the new proposed robust PLS-ARWMCD method still gives better results than classical PLSR method for all the types of outliers in terms of efficiency, fitting to data and predictive ability.

Overall, both of from Table 1 and Table 2, it could be concluded that the new proposed robust PLS-ARWMCD method outperforms especially its two robust competitors (PRM and PLS-SD) existing in the literature with more or less differences in terms of efficiency, fitting to data and predictive ability for five different types of outliers.

Table 2: The sample size is $n = 200$, $p = 5$ and $k = 2$, the proportion of outliers is 20%.

	PLSR	RSIMPLS	PRM	PLS- SD	PLS-KurSD	**PLS-ARWMCD**
No Contamination						
MSE	0.0092	0.0111	0.0101	0.0104	0.01	0.0105
GOF	0.8312	0.8308	0.8308	0.8308	0.8309	0.8308
RMSE	1.0961	1.0974	1.0969	1.0973	1.0968	1.0974
Mean(angle)	0.0446	0.0519	0.0462	0.0493	0.0477	0.0491
Bad Leverage Points						
MSE	1.8946	0.0122	1.7726	0.4134	0.0121	0.0109
GOF	0.1858	0.8309	0.2395	0.7143	0.831	0.8313
RMSE	2.4002	1.1012	2.3205	1.4282	1.1011	1.0998
Mean(angle)	1.3018	0.0537	1.1833	0.2467	0.054	0.05
Vertical Outliers						
MSE	0.0791	0.0115	0.0174	0.0176	0.0126	0.0112
GOF	0.8057	0.8278	0.8265	0.8267	0.8282	0.8286
RMSE	1.1681	1.1002	1.106	1.1063	1.1003	1.0989
Mean(angle)	0.1437	0.0471	0.0632	0.0656	0.054	0.0503
Good Leverage Points						
MSE	0.9975	0.0128	1.0568	0.044	0.0121	0.0109
GOF	0.6741	0.831	0.6817	0.8282	0.831	0.8313
RMSE	1.5213	1.1011	1.5049	1.1102	1.1011	1.0998
Mean(angle)	0.7739	0.057	0.7813	0.1165	0.0539	0.05
Concentrated Outliers						
MSE	1.8527	0.0128	1.926	0.1628	0.0121	0.0109
GOF	0.4929	0.831	0.485	0.8107	0.831	0.8313
RMSE	1.8946	1.1012	1.9104	1.1648	1.1011	1.0998
Mean(angle)	1.1091	0.0569	1.1119	0.2307	0.0539	0.05
Orthogonal Outliers						
MSE	0.1987	0.0176	0.2332	0.0108	0.0115	0.0104
GOF	0.7806	0.831	0.7718	0.8319	0.8319	0.8322
RMSE	1.2488	1.1026	1.2739	1.1007	1.101	1.0999
Mean(angle)	0.2982	0.066	0.3247	0.0504	0.0519	0.0491

5. APPLICATION TO FISH DATA

In this section, the new proposed robust PLSR method and four robust PLSR methods, existing in the literature, will be compared on a real data including outliers in terms of goodness-of-fit and predictive ability by using (4.3) and (4.4). For this purpose, the fish data which was given in Naes (1985) will be used. The fish data comprise 45 observations and the last 7 are outliers (in the words of Naes, 'abnormal samples'). In this example, fat concentration (percentage, %) of 45 fish samples (rainbow trout) and independent variables of the absorbance at 9 Near Infrared Reflectance (NIR) wavelengths measured after sample homogenisation. The aim of the analysis made on this data set is to model the relationships between the fat concentration (one response variable) and these nine spectrums (independent variables). In this study, the data set is divided into two parts.

The first 20 observations are the test set and the other remained 25 samples are the training set (Gil and Romera, 1998; Hardy *et al.*, 1996; Naes, 1985).

Firstly, similar to the our simulation studies, while computing the GOF values 7 outliers are removed from training set that occurs of 25 samples. However, while computing the RMSE values the models are constituted using the training set including the 7 outliers. Then, by using the regression coefficients obtained from these models, the predictions are made from clean test set that occurs of 20 samples. Hence, the predictive ability of the new robust PLSR method 'PLS-ARWMCD' is examined especially against the classical PLSR method and the other four robust methods.

The GOF or RMSE values could be considered while selecting the number components that will be retained in the model. The optimal number of components could be selected as the k for which the GOF values are no more change. However, as it is mentioned before in Engelen *et al.* (2004), it is more convenient to consider the RMSE values while selecting the optimal number of components. The significant point while selecting the optimal number of components retaining in the model is that adding one more component whether cause an important decrease or not in RMSE value. Hence, both the aim of data reduction is not deviated and an unnecessary component is not added to model. In Figure 2, the figure of RMSE values against the number of components in the model is drawn.

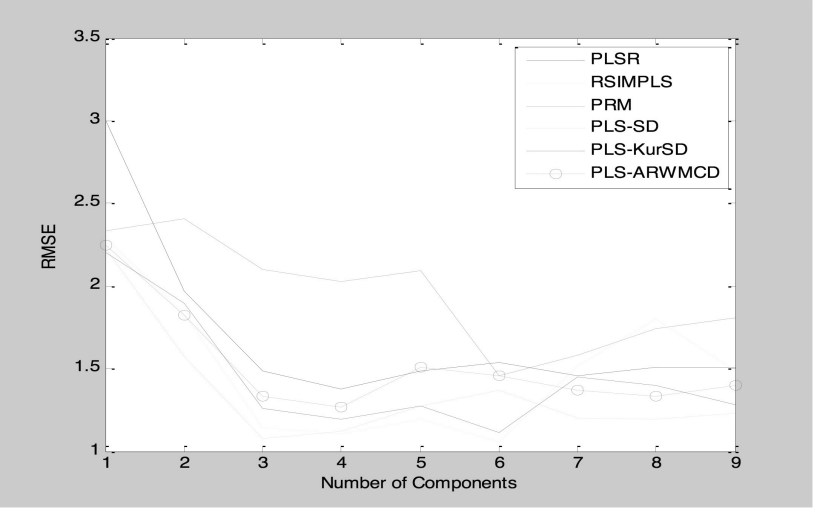

Figure 2: The RMSE values against the number of components in the model for fish data with the training set of 25 samples and the test set of 20 samples.

When Figure 2 is examined, it is seen that it is right to select the number of components retaining in the model as three for this data set. Because from the figure it is seen that adding the third component to the model causes a

significant decrease in the RMSE values of all the methods. It could be seen also much clearly from Table 3 that the optimal number of components should be selected as $k_{opt} = 3$, as adding the third component to the model cause an important decrease in the RMSE values for all the robust methods. Furthermore, it is clear that the fitting to data also improves for all the methods after adding the third component to the model. Table 3 shows that the new proposed robust PLS-ARWMCD method fitting to the data better and it has a higher predictive ability than both classical PLSR method and robust PRM method for $k_{opt} = 3$.

Table 3: The GOF and RMSE values for fish data in case of the first 20 observations are the test set and the other 25 samples are the training set.

Number of Components		**PLSR**	**RSIMPLS**	**PRM**	**PLS-SD**	**PLS-KurSD**	**PLS-ARWMCD**
k=1	GOF	0.2912	0.4335	0.3777	0.4417	0.4444	0.4397
	RMSE	3.0001	2.2937	2.3307	2.2274	2.2029	2.2487
k=2	GOF	0.6927	0.7421	0.2713	0.7853	0.6948	0.7605
	RMSE	1.9715	1.8293	2.4072	1.573	1.8935	1.8234
k=3	GOF	0.882	0.9687	0.6166	0.9665	0.9594	0.9579
	RMSE	1.4861	1.1401	2.0993	1.0797	1.259	1.3322
k=4	GOF	0.8987	0.971	0.6277	0.9737	0.9447	0.9662
	RMSE	1.3742	1.1089	2.0237	1.1198	1.1924	1.2646
k=5	GOF	0.9113	0.979	0.6782	0.9716	0.9777	0.9713
	RMSE	1.4874	1.1918	2.0921	1.2705	1.2708	1.5054
k=6	GOF	0.9231	0.9825	0.7705	0.9796	0.9854	0.9816
	RMSE	1.5348	1.0543	1.4578	1.3727	1.1129	1.4545
k=7	GOF	0.9299	0.9829	0.7806	0.9714	0.9865	0.9862
	RMSE	1.4553	1.519	1.5835	1.2033	1.4528	1.3679
k=8	GOF	0.9463	0.9768	0.8063	0.9769	0.9868	0.9861
	RMSE	1.5056	1.7989	1.7409	1.1925	1.4019	1.33
k=9	GOF	0.9463	0.9851	0.8087	0.9812	0.9798	0.987
	RMSE	1.5052	1.4874	1.8095	1.2338	1.2843	1.399

6. CONCLUSIONS

In this study, we propose a new robust PLSR method for the linear regression model with one response variable, PLS-ARWMCD, in order to obtain robust predictions in case of outliers present in the data set.

In the simulation study, the new proposed robust PLSR method is compared with classical PLSR method and four robust PLSR methods existing in the literature in terms of efficiency, fitting to data and predictive ability on a clean data set and on contaminated data sets with bad leverage points, vertical outliers, good leverage points, concentrated outliers or orthogonal outliers.

The optimal number of components is selected as $k = 2$ at the beginning of the simulation study. 10% and 20% proportions of this data set are replaced by outliers, respectively. Thus, the increment in the proportion of outliers how affects on performances of the new proposed robust PLSR method and four robust PLSR methods (existing in the literature) is examined. When the 10% proportion of the data set is contaminated by different types of outliers, both the new proposed robust PLS-ARWMCD method and the four robust PLSR methods existing in the literature outperform classical PLSR method in terms of efficiency and predictive ability (exception of PRM method that performs not better than classical PLSR method in terms of efficiency in case of good leverage points existence). The PLS-ARWMCD method comes to the forefront as a good alternative method against robust PRM and PLS-SD methods in terms of efficiency, fitting to data and predictive ability for all the types of outliers. Moreover, PLS-ARWMCD method shows a close performance with robust RSIMPLS and PLS-KurSD methods in terms of efficiency, fitting to data, predictive ability and mean angle measures. When the proportion of outliers in the data set is reached to a high level as 20%, robust PRM method shows a lower performance than other robust methods in terms of efficiency, fitting to data and predictive ability for all the types of outliers except that vertical outliers. Furthermore, if there is 20% proportion of concentrated outliers or orthogonal outliers in the data set, robust PRM method looses its performance completely against classical PLSR method. When there is high proportion of bad leverage points or concentrated outliers in the data set, robust PLS-SD method is less efficient and it has a lower predictive ability than the other robust RSIMPLS, PLS-KurSD methods and new proposed robust PLS-ARWMCD method.

The results obtained from real data analysis show that the optimal number of components is selected as $k_{opt} = 3$, as adding the third component to the model causes a considerably decrease in the RMSE values of robust methods. It is clear from the results of the model containing $k = 3$ components that GOF values of the new proposed robust PLS-ARWMCD method are higher than both classical PLSR method and robust PRM method. Moreover, when $k_{opt} = 3$ is selected, the RMSE value for PLS-ARWMCD is lower than both classical PLSR method and robust PRM method. Generally, whatever the optimal number of the components in the model for the fish data set, the new proposed robust PLS-ARWMCD method gives better models than both classical PLSR method and robust PRM method in terms of fitting to data and predictive ability.

Consequently, it is seen that the new proposed robust PLS-ARWMCD method gives more efficient results than especially classical PLSR method in data sets contaminated by a reasonable amount of outliers. The simulation study shows that when the data contain 10% or 20% proportion of bad leverage points, the new robust PLS-ARWMCD method outperforms both of the robust PRM and PLS-SD methods in terms of efficiency and predictive ability.

When the data contain 10% proportion of vertical outliers, the new robust PLS-ARWMCD method shows a close performance to the other four robust PLSR methods existing in literature. However, when there is 20% proportion of vertical outliers in the data set; the new robust PLS-ARWMCD method, robust RSIMPLS and PLS-KurSD methods are the forefront methods in terms of efficiency and predictive ability. When the data contain 10% or 20% proportion of good leverage points; the new robust PLS-ARWMCD method has a better performance than robust PRM method both in terms of efficiency and predictive ability, however, it is only more efficient than robust PLS-SD method. When there is 10% proportion of concentrated outliers; the new robust PLS-ARWMCD method is both more efficient and it has a higher predictive ability than robust PRM method, however, it is only more efficient than robust PLS-SD method. When there is 20% proportion of concentrated outliers in the data set, the new robust PLS-ARWMCD method is both more efficient and it has a higher predictive ability than both robust PRM and PLS-SD methods. When the data contain 10% or 20% proportion of orthogonal outliers; the new robust PLS-ARWMCD method has a better performance than robust PRM method in terms of efficiency, fitting to data and predictive ability. Overall, it could be concluded that the new proposed robust PLS-ARWMCD could cope with different types and proportions of outliers efficiently and it give robust predictions.

REFERENCES

[1] Engelen S.; Hubert, M.; Vanden Branden, K. and Verboven, S. (2004). *Robust PCR and Robust PLSR: A comparative study.* In "Theory and Applications of Recent Robust Methods" (M. Hubert, G. Pison, A. Struyf and S.V. Aelst, Eds.), Birkhäuser, Basel, 105–117.

[2] Filzmoser, P.; Serneels, S.; Maronna, R. and Van Espen, P.J. (2009). *Robust multivariate methods in chemometrics.* In "Comprehensive Chemometrics" (B. Walczak, R.T. Ferre and S. Brown, Eds.), 681–722.

[3] Gervini, D. (2003). A robust and efficient adaptive reweighted estimator of multivariate location and scatter, *Journal of Multivariate Analysis*, **84**, 116–144.

[4] Gil, J.A. and Romera, R. (1998). On robust partial least squares (PLS) methods, *Journal of Chemometrics*, **12**, 365–378.

[5] González, J.; Peña, D. and Romera, R. (2009). A robust partial least squares regression method with applications, *Journal of Chemometrics*, **23**, 78–90.

[6] Hardy, A.J.; Maclaurin, P.; Haswell, S.J.; De Jong, S. and Vandeginste, B.G.M. (1996). Double-case diagnostic for outliers identification, *Chemometrics and Intelligent Laboratory Systems*, **34**, 117–129.

[7] Hubert M. and Vanden Branden, K. (2003). Robust methods for Partial Least Squares Regression, *Journal of Chemometrics*, **17**, 537–549.

[8] Liebmann, B.; Filzmoser, P. and Varmuza, K. (2010). Robust and Classical PLS Regression Compared, *Journal of Chemometrics*, **24**(3–4), 111–120.

[9] Moller, S.F.; Von Frese, J. and Bro, R. (2005). Robust Methods for Multivariate Data Analysis, *Journal of Chemometrics*, **19**(10), 549–563.

[10] Naes, T. (1985). Multivariate calibration when the error covariance matrix is structured, *Technometrics*, **27**(3), 301–311.

[11] Polat, E. (2014). *New Approaches in Robust Partial Least Squares Regression Analysis*, Ph.D. Turkish diss., Hacettepe University Department of Statistics, Ankara, Turkey.

[12] Rousseeuw, P.J. (1984). Least median of squares regression, *J. Amer. Statist. Assoc.*, **79**, 871–880.

[13] Rousseeuw, P.J. and Van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points, *J. Amer. Statist. Assoc.*, **85**, 633–639.

[14] Rousseeuw, P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, **41**, 212–224.

[15] Serneels, S.; Croux, C.; Filzmoser, P. and Van Espen, P.J. (2005). Partial Robust M-regression, *Chemometrics and Intelligent Laboratory Systems*, **79**, 55–64.

[16] Verboven, S. and Hubert, M. (2005). LIBRA: a MATLAB library for robust analysis, *Chemometrics and Intelligent Laboratory Systems*, **75**, 127–136.

[17] Wakeling, I.N. and Macfie, H.J.H. (1992). A Robust PLS Procedure, *Journal of Chemometrics*, **6**, 189–198.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 475–492

RELIABILITY ASPECTS OF PROPORTIONAL MEAN RESIDUAL LIFE MODEL USING QUANTILE FUNCTIONS

Authors: N. Unnikrishnan Nair
– Department of Statistics, Cochin University of Science and Technology, Cochin, Kerala, India
unnikrishnannair4@gmail.com

P.G. Sankaran
– Department of Statistics, Cochin University of Science and Technology, Cochin, Kerala, India
sankaran.pg@gmail.com

S.M. Sunoj
– Department of Statistics, Cochin University of Science and Technology, Cochin, Kerala, India
smsunoj@gmail.com

Received: March 2017 Revised: July 2017 Accepted: July 2017

Abstract:

- In the present work we propose a study of the proportional mean residual life model using quantile functions instead of distribution functions employed in the traditional approach. Through various quantile-based reliability concepts ageing properties of the model are discussed. We present characterizations of the proportional mean residual life model by properties of measures of uncertainty in the residual lives of the variables. Some stochastic order properties relating to the model are also derived.

Key-Words:

- *proportional mean residual life model; quantile function; divergence measures; characterization; stochastic orders.*

AMS Subject Classification:

- 62N05, 62E15.

1. INTRODUCTION

The proportional mean residual life model (PMRLM) was introduced by [11] and [14] as an alternative to the well-known proportional hazards model (PHM). These two papers explain the relevance of PMRLM and its advantages over the PHM. Let X and Y be two non-negative absolutely continuous random variables with finite expectation and survival functions $\bar{F}_X(\cdot)$ and $\bar{F}_Y(\cdot)$ respectively. Then the PMRLM is represented by

$$m_Y(x) = \theta m_X(x)\,, \qquad \theta > 0\,, \tag{1.1}$$

where $m_X(x) = \frac{1}{\bar{F}_X(x)} \int_x^\infty \bar{F}_X(t)dt$ and $m_Y(x) = \frac{1}{\bar{F}_Y(x)} \int_x^\infty \bar{F}_Y(t)dt$ are the mean residual life functions of X and Y. Unlike the PHM, (1.1) may not be valid for all $\theta > 0$. If $m_X(x)$ is increasing then $\theta > 0$ while for a decreasing $m_X(x), 0 < \theta \leq \theta_0$, where $\theta_0^{-1} = max\,(0, -min(m_X(x)))$. The relationship between PMRLM and PHM, the ageing properties and certain bounds on residual moments and residual variance of the former in the context of reliability analysis were studied in [4]. In the same direction [9] discussed the closure properties of the ageing classes related to PMRLM and preservation of certain stochastic orders. The reliability aspects of a dynamic version of (1.1) obtained by replacing the constant θ in (1.1) by a non-negative function of x are also investigated in [10]. All the works mentioned above make use of the identity (1.1), the relationship between the survival functions derived therefrom and the properties of the mean residual life function.

An associated concept is the percentile residual life discussed in several papers like [13], [2] and their references. Instead of the distribution function, a life distribution and its desired characteristics can also be represented through the quantile function

$$Q_X(u) = \inf\{x : F_X(x) \geq u\}\,, \qquad 0 \leq u \leq 1\,,$$

and various reliability functions evaluated from $Q_X(\cdot)$. The relevance and advantages of using $Q_X(\cdot)$ over $\bar{F}_X(\cdot)$ in various forms of statistical analysis are well documented in [3] and the associated methodology for reliability analysis in [8]. The present article focuses attention on studying the reliability implications of PMRLM using quantile function and the associated reliability concepts.

The factors that motivated the present work are as follows. There are many quantile functions that have simple forms capable of representing a wide variety of lifetime data. These are discussed extensively in [8]. Our work enables the induction of such quantile functions as lifetime models in the analysis of PMRLM. Many of the flexible quantile functions in literature have no tractable distribution functions to make use of them in the conventional analysis.

Further, the quantile analogue of (1.1) reads

$$m_Y\left(Q_X(u)\right) = \theta m_X\left(Q_X(u)\right)\,, \quad \theta > 0\,, \tag{1.2}$$

in which the right side is the mean residual quantile function of X while the left side is not the mean residual quantile function of Y. Thus the analogues of the mean residual life in the quantile-based analysis are not proportional as in (1.1). This points out to the possibility of properties of PMRLM that are different from the conventional one. Because of the special properties of quantile functions we can obtain results that are difficult to obtain by the distribution function approach. Various measures of uncertainty in the residual life of a device and their association with reliability concepts are of recent interest on the premise that increase in uncertainty implies that the device becomes more unreliable. We propose two characterizations of PMRLM based on the Kullback–Leibler divergence and its cumulative form using quantile functions.

The rest of the paper contains four sections. In Section 2 we present some preliminary results required for the deliberations in the sequence. This is followed in Section 3 with discussion on the ageing concepts of Y in relation to those of X. The characterizations of the PMRLM are presented in Section 4. Finally in Section 5 some quantile-based stochastic orders associated with PMRLM are discussed.

2. THE PROPORTIONAL MEAN RESIDUAL LIFE MODEL

Let X and Y be as defined in the previous section with strictly decreasing survival functions, hazard rate functions $h_X(\cdot)$ and $h_Y(\cdot)$ and quantile functions $Q_X(\cdot)$ and $Q_Y(\cdot)$. From [4],

$$\bar{F}_Y(x) = \left[\bar{F}_X(x)\right]^{\frac{1}{\theta}} \left(\frac{m_X(x)}{\mu_X}\right)^{\frac{1}{\theta}-1}, \quad \mu_X = E(X)\,, \tag{2.1}$$

and

$$h_Y(x) - h_X(x) = \frac{1-\theta}{\theta m_X(x)}\,. \tag{2.2}$$

The quantile-based reliability functions of X and Y are the hazard quantile functions

$$H_X(u) = h_X\left(Q_X(u)\right) = \frac{1}{(1-u)q_X(u)}$$

and

$$H_Y(u) = \frac{1}{(1-u)q_Y(u)},$$

and the mean residual quantile functions

$$M_X(u) = m_X\left(Q_X(u)\right) = \frac{1}{(1-u)}\int_u^1 (1-p)q_X(p)dp$$

and

$$M_Y(u) = \frac{1}{(1-u)}\int_u^1 (1-p)q_Y(p)dp\,,$$

where $q_X(u) = \frac{dQ_X(u)}{du}$ and $q_Y(u) = \frac{dQ_Y(u)}{du}$ are the quantile density functions of X and Y. For the definitions, interpretations of these functions and interrelationships between them, we refer to [7].

Setting $x = Q_X(u)$ in (2.1) we see that

$$\bar{F}_Y\left(Q_X(u)\right) = (1-u)^{\frac{1}{\theta}}\left(\frac{M_X(u)}{\mu_X}\right)^{\frac{1}{\theta}-1}$$

or

$$Q_X(u) = Q_Y\left[1-(1-u)^{\frac{1}{\theta}}\left(\frac{M_X(u)}{\mu_X}\right)^{\frac{1}{\theta}-1}\right]. \tag{2.3}$$

Writing $A(u) = 1-(1-u)^{\frac{1}{\theta}}\left(\frac{M_X(u)}{\mu_X}\right)^{\frac{1}{\theta}-1}$, (2.3) becomes

$$Q_X(u) = Q_Y\left(A(u)\right). \tag{2.4}$$

It is not difficult to see that when (2.3) is satisfied (1.2) also holds. Notice that $A(\cdot)$ is a distribution function on $[0,1]$ so that $A(u)$ is increasing in u with $A(0) = 0$ and $A(1) = 1$. One can work with the identity (1.1) and obtain another relationship involving $\bar{F}_X(\cdot)$ and $\bar{F}_Y(\cdot)$ in the form

$$\bar{F}_X(x) = \left[\bar{F}_Y(x)\right]^\theta\left(\frac{m_Y(x)}{\mu_Y}\right)^{\theta-1}, \quad \mu_Y = E(Y)\,,$$

or equivalently

$$Q_Y(u) = Q_X\left(B(u)\right), \tag{2.5}$$

where $B(u) = 1-(1-u)^\theta\left(\frac{M_Y(u)}{\mu_Y}\right)^{\theta-1}$. Since Q_X is an increasing function, it is easy to see that $B(u) = A^{-1}(u)$ and $A(u) = B^{-1}(u)$.

From (2.3) by differentiation,

$$q_X(u) = q_Y\left(A(u)\right)A'(u)\,,$$

so that

$$(1-u)q_X(u) = \frac{(1-u)A'(u)}{1-A(u)}\left(1-A(u)\right)q_Y\left(A(u)\right),$$

giving

$$H_X(u) = \frac{1 - A(u)}{(1-u)A'(u)} H_Y\left(A(u)\right) \tag{2.6}$$

and from (1.2)

$$\theta M_X(u) = m_Y\left[Q_Y A(u)\right] = M_Y\left(A(u)\right) = m_Y\left(Q_X(u)\right). \tag{2.7}$$

Equation (2.7) suggests that in general $M_X(\cdot)$ and $M_Y(\cdot)$ need not be proportional. Therefore unlike the distribution function approach wherein properties of $m_Y(x)$ can be directly obtained from equation (1.1), the quantile analysis does not directly provide characteristics of mean residual quantile function of Y from that of X. We give an example of the use of quantile function in the analysis of PMRLM and in modelling real data.

Remark 2.1. When the mean residual quantile function of Y is proportional to that of X,

$$\frac{1}{1-u}\int_u^1 (1-p) q_Y(p)dp = \frac{1}{1-u}\int_u^1 (1-p) q_X(p)dp$$

which is equivalent to $Q_Y(u) = \theta Q_X(u) = Q_X(\theta u)$. This is the case when Y is obtained as a change of scale in X.

Example 2.1. Let the distribution of X be represented by the quantile function

$$Q_X(u) = -(\alpha+\mu)\log(1-u) - 2\alpha u\,, \quad \mu > 0\,,\ -\mu < \alpha < \mu\,. \tag{2.8}$$

Equation (2.8) specifies a family of flexible distributions that includes exponential and uniform distributions as special cases and approximates well distributions like Weibull, gamma, beta and half-normal. A detailed discussion of (2.8) is available in [5]. Notice that the general form of the distribution does not admit a closed form for its distribution function, except that the distribution function $F_X(\cdot)$ and density function $f_X(\cdot)$ are related through

$$f_X(x) = \frac{1 - F_X(x)}{2\alpha F_X(x) + \mu - \alpha}\,.$$

Thus it becomes difficult to work with $m_X(\cdot)$ and more so with $m_Y(\cdot)$ and conclude their general properties using (1.1). From a quantile perspective we have

$$M_X(u) = \mu + \alpha u\,,$$

$$F_Y\left(Q_X(u)\right) = 1 - (1-u)^{\frac{1}{\theta}}\left(1 + \frac{\alpha u}{\mu}\right)^{\frac{1}{\theta}-1},$$

$$m_Y\left(Q_X(u)\right) = \theta(\mu + \alpha u)$$

and

$$Q_Y(u) = Q_X\left[A^{-1}(u)\right], \quad A(u) = 1 - (1-u)^{\frac{1}{\theta}}\left(1 + \frac{\alpha u}{\mu}\right)^{\frac{1}{\theta}-1}. \tag{2.9}$$

To illustrate the application of quantile functions in modelling proportional mean residual life using the above model we consider the data on the time to first failure of 20 electric carts given in [15].

The methodology used here is described as follows. Let $x_1 < x_2 < ... < x_n$ be the distinct observations in the sample. We estimate the sample distribution function as

$$\hat{F}_X(x_r) = u_r = \frac{r - 0.5}{n}, \quad r = 1, 2, ..., n, \quad u_r \le x < u_{r+1},$$

by dividing the interval $(0, 1)$ with equal parts and using their midpoints to symmetrically place the u values. This gives

$$\hat{Q}_X(u_r) = x_r, \quad u_{r-1} < u < u_r, \quad r = 1, 2, ..., n, \quad u_0 = 0,$$

and

$$\begin{aligned}\hat{F}_Y\hat{Q}_X(u_r) &= \hat{F}_Y(x_r) \\ &= 1 - (1-u_r)^{\frac{1}{\theta}}\left(\frac{M_X(u_r)}{\mu}\right)^{\frac{1}{\theta}-1}.\end{aligned}$$

Recall $G_Y(\cdot) = F_Y\left(Q_X(\cdot)\right)$ is a distribution function over $(0, 1)$. If $g_Y(\cdot)$ is the probability density function of $G_Y(\cdot)$, then

$$g_Y(u) = f_Y\left(Q_X(u)\right) q_X(u) = \frac{f_Y Q_X(u)}{f_X Q_X(u)}.$$

We estimate the parameter of the PMRLM by minimizing

$$E = -\int_0^1 g_Y(u) \log g_Y(u) du.$$

For the given data $g_Y(\cdot)$ is replaced by its estimated value

$$\hat{g}_Y(u) = \frac{\left(\hat{F}_Y(x_r) - \hat{F}_Y(x_{r-1})\right)}{u_r - u_{r-1}}$$

and the minimization of E is carried out. The estimates obtained are

$$\hat{\alpha} = 0.6078, \quad \hat{\mu} = 1.054 \quad \text{and} \quad \hat{\theta} = 0.8438.$$

Thus

$$\hat{m}_Y\left(\hat{Q}_X(u)\right) = 0.8894 + 0.5129u.$$

for the minimum of E obtained as 1.54626×10^{-10} which shows the closeness of the fit. Further analysis of the data can be accomplished based on $G_Y(\cdot)$ or $m_Y\left(Q_X(\cdot)\right)$ obtained above.

For all quantile functions equation, (1.2) may not be satisfied for $\theta > 0$. We present a necessary and sufficient condition for the existence of the PMRLM for a distribution.

Theorem 2.1. *A quantile function $Q_X(\cdot)$ admits a PMRLM if and only if θ satisfies*

$$q_X(u) + \theta M'_X(u) \geq 0$$

where the prime denotes differentiation with respect to u.

Proof: The relationship (2.4) is satisfied for some $Q_Y(\cdot)$ if and only if there exists a quantile function $Q_X(\cdot)$ for which $q_X(\cdot) > 0$ since $Q_X(\cdot)$ must be an increasing function. Now

$$\begin{aligned} q_X(u) \;=\; & \frac{q_Y\left(A(u)\right)(1-u)^{\frac{1}{\theta}-1}}{\theta}\left(\frac{M_X(u)}{\mu_X}\right)^{\frac{1}{\theta}-1} \\ & -\; q_Y\left(A(u)\right)(1-u)^{\frac{1}{\theta}}\left(\frac{1}{\theta}-1\right)\left(\frac{M_X(u)}{q_X(u)}\right)^{\frac{1}{\theta}-2}\frac{M'_X(u)}{\mu_X} \;\geq\; 0 \end{aligned}$$

$$\begin{aligned} & \iff \quad \frac{M_X(u)}{\theta} - \left(\frac{1}{\theta}-1\right)M'_X(u) > 0 \\ & \iff \quad M_X(u) - (1-u)M'_X(u) + (1-u)\theta M'_X(u) > 0 \\ & \iff \quad \frac{1}{H_X(u)} + (1-u)\theta M'_X(u) > 0 \\ & \iff \quad q_X(u) + \theta M'_X(u) > 0\,. \end{aligned}$$

□

Remark 2.2. In view of Theorem 1, we see that

(**i**) if $M'_X(\cdot) \geq 0$ then PMRLM holds for all $\theta > 0$
and

(**ii**) if $M'_X(\cdot) < 0$, X admits PMRLM only when the range of θ is limited to $[0, \theta_0]$, $\theta_0 = \max\left(0, -\min \frac{q_X(u)}{M'_X(u)}\right)$.

3. AGEING PROPERTIES

There are situations when the distribution of Y specified by $Q_Y(\cdot)$ may not have tractable form to study the ageing properties of Y analytically. For example see $Q_Y(\cdot)$ given in (2.9). This does not pose any problems to data analysis since

$F_Y\left(Q_X(\cdot)\right)$ can be employed for inferential purposes. Generally, the baseline distribution is one for which the ageing criteria is known or can be evaluated and therefore results that enable the inference of ageing characteristics of Y in terms of those of X become useful. In this section we prove some theorems in this direction. For this we need the following definitions. The definitions and results are given only for positive ageing concepts as it is easy to deduce their negative ageing counterparts by reversing the monotonicity or the inequality in each case. The random variable X is said to be (i) increasing hazard rate (IHR) if $H_X(\cdot)$ is increasing (ii) new better than used in hazard rate (NBUHR) if $H_X(0) \leq H_X(u)$ for all u and new better than used in hazard rate average (NBUHRA) if $-\frac{\log(1-u)}{Q_X(u)} \geq H_X(u)$ (iii) increasing hazard rate average (IHRA) if $\frac{Q_X(u)}{-\log(1-u)}$ is decreasing in u (iv) decreasing mean residual life (DMRL) if $M_X(\cdot)$ is decreasing (v) decreasing mean residual life in harmonic average (DMRLHA) if $\frac{1}{Q_X(u)}\int_0^u \frac{q_X(p)}{M_X(p)}dp$ is decreasing in u, (vi) new better than used (NBU) if $Q_X(u+v-uv) \leq Q_X(u)+Q_X(v)$ for $0 \leq u < v < 1$ and (vii) new better than used in expectation (NBUE) if $M_X(u) \leq \mu_X$. It may be noticed the definitions of the above concepts in the distribution function approach and the quantile function approaches are equivalent. However the results pertaining to PMRLM in the two differ at least in some cases. For a detailed discussion of the characteristics of various quantile ageing classes, see [8].

Theorem 3.1. *If X is IHR, $\theta > 1$ and $M_X(\cdot)$ is logconvex then Y is IHR.*

Proof: Recall that

$$H_Y\left(A(u)\right) = \frac{(1-u)A'(u)}{1-A(u)} H_X(u)\,.$$

Also

$$T(u) = \frac{(1-u)A'(u)}{1-A(u)} = \frac{1}{\theta} - \left(\frac{1}{\theta}-1\right)(1-u)\frac{d\log M_X(u)}{du}\,,$$

gives

$$T'(u) = \left(1-\frac{1}{\theta}\right)(1-u)\frac{d^2\log M_X(u)}{du^2} + \left(\frac{1}{\theta}-1\right)\frac{d\log M_X(u)}{du}\,.$$

When X is IHR, it is also DMRL. Under the conditions of the theorem $T(\cdot)$ is increasing and so is $H_X(\cdot)$. Thus $H_Y\left(A(u)\right)$ is an increasing function of $A(u)$ and hence of u, showing that $H_Y(u)$ is increasing. □

Remark 3.1. It can be shown that if Y is IHR, $\theta < 1$ and $M_Y(\cdot)$ is logconcave, then X is IHR. To prove this we use (2.5) and work with $B(u)$ in the same manner as with $A(u)$. Various results concerning other ageing properties proved below can also have parallel results relating Y with X. Because of similarity they are not pursued further.

We give two examples, one demonstrating the usefulness of Theorem 2 and the other to show that the conditions imposed on the reliability functions of X are essential.

Example 3.1. Consider the linear hazard quantile function distribution with quantile function

$$Q_X(u) = \frac{1}{a+b} \log\left(\frac{a+bu}{a(1-u)}\right), \quad a > 0\,, \quad b > 0\,. \tag{3.1}$$

whose properties and applications are studied in [6]. The hazard and the mean residual quantile functions of (3.1) are

$$H_X(u) = a + bu$$

and

$$M_X(u) = \frac{1}{b(1-u)} \log\left(\frac{a+b}{a+bu}\right).$$

Obviously X is IHR and

$$\frac{d \log M_X(u)}{du} = \frac{1}{1-u} - \frac{b}{(a+bu)\left[\log(a+b) - \log(a+bu)\right]}$$

is increasing and hence by Theorem 2, Y is IHR. This method looks easier than evaluating the reliability aspects directly from

$$Q_Y(u) = \frac{1}{a+b} \log\left(\frac{a + bA^{-1}(u)}{a(1 - A^{-1}(u))}\right)$$

with

$$A(u) = 1 - (1-u)^{\frac{1}{\theta}} \left[\frac{\log(a+b) - \log(a+bu)}{(1-u)\left(\log(a+b) - \log a\right)}\right]^{\frac{1}{\theta}-1}$$

derived from (2.4).

Example 3.2. The quantile function

$$Q_X(u) = \frac{3\alpha\beta u^2}{2} + \alpha u(2-\beta)\,, \quad \alpha > 0\,, \quad \beta = \frac{1}{2}\,,$$

has

$$H_X(u) = \left[\alpha(1-u)\left(3\beta u + 2 - \beta\right)\right]^{-1}$$

and

$$M_X(u) = \alpha(1-u)(1+\beta u)\,.$$

Differentiating $H_X(u)$, the sign of $H'_X(u)$ depends on the sign of $3\beta u - 2\beta + 1$ which is positive for $\beta = \frac{1}{2}$. Hence X is IHR. Further

$$M'_X(u) = \alpha(\beta - 1) - \alpha\beta u$$

so that $M_X(\cdot)$ is decreasing and concave. Thus the conditions of the theorem are not satisfied. From (2.6) and

$$A(u) = 1 - (1-u)^{\frac{1}{\theta}} \left[(1-u)(1+\beta u)\right]^{\frac{1}{\theta}-1},$$

and

$$H_Y\left(A(u)\right) = \frac{\left(\frac{2}{\theta}-1\right)(1+\beta u) - \left(\frac{1}{\theta}-1\right)\beta(1-u)}{\alpha(1+\beta u)(1-u)(3\beta u + 2 - \beta)}, \qquad \beta = \frac{1}{2},$$

splitting up the terms we see that $H_Y(u)$ is not increasing for all u.

Theorem 3.2.

(**i**) *If X is IHRA, $\theta > 1$ and* $\left(\frac{\log\left(\frac{M_X(u)}{\mu_X}\right)}{\log(1-u)}\right)$ *is decreasing then Y is IHRA.*

(**ii**) *If X is IHR and $\theta > 1$, then Y is NBUHR.*

(**iii**) *IF X is NBUHRA, $\theta > 1$ and* $\left(\frac{\log\frac{M_X(u)}{\mu_X}}{\log(1-u)} + \frac{M_X'(0)}{\mu_X}\right) \leq 0$, *then Y is NBUHRA.*

Proof: (i) First we note that

$$\frac{Q_Y(u)}{-\log(1-u)} = \frac{Q_X\left(A^{-1}(u)\right)}{-\log(1-u)} = \frac{Q_X(u)}{-\log\left(1-A(u)\right)} = \frac{Q_X(u)}{-\log(1-u)} \frac{\log(1-u)}{\log\left(1-A(u)\right)}. \tag{3.2}$$

The sign of $\frac{\log(1-u)}{\log(1-A(u))}$ depends on

$$\begin{aligned} D(u) &= \frac{A"(u)\log(1-u)}{1-A(u)} - \log\frac{1-A(u)}{1-u} \\ &= \frac{\left(\frac{1}{\theta}-1\right)}{1-u}\left[-\log\frac{M_X(u)}{\mu_X} - (1-u)\log(1-u)\frac{M_X'(u)}{M_X(u)}\right] \\ &= -\left(\frac{1}{\theta}-1\right)(\log(1-u))^2 \frac{d}{du}\left(\frac{\log\left(\frac{M_X(u)}{\mu_X}\right)}{\log(1-u)}\right). \end{aligned}$$

When X is IHRA, the first term on the right of (3.2) increases and the second term increases when $\theta > 1$ and $\frac{\log\left(\frac{M_x(u)}{\mu_X}\right)}{\log(1-u)}$ increases. Hence Y is IHRA.

(ii) From (2.2)

$$H_Y\left(A(u)\right) - H_X(u) = \left(\frac{1}{\theta}-1\right)\left[M_X(u)\right]^{-1}.$$

This gives

$$H_Y\left(A(u)\right) - H_Y\left(A(0)\right) = H_X(u) - H_X(0) + \left(\frac{1}{\theta} - 1\right)\left(\frac{1}{M_X(u)} - \frac{1}{\mu_X}\right).$$

Since IHR implies NBUHR and NBUE $H_X(u) \geq H_X(0)$ and $\frac{1}{M_X(u)} \geq \frac{1}{\mu_X}$. Hence $H_Y\left(A(u)\right) \geq H_Y\left(A(0)\right) = H_Y(0)$ for all $A(u)$ implies that $H_Y(u) \geq H_Y(0)$, $0 \leq u \leq 1$ and Y is NBUHR.

(iii) From (3.2)

$$\frac{Q_Y(u)}{-\log(1-u)} = \frac{Q_Y(u)}{-\log(1-u)} \frac{\log(1-u)}{\log\left(1-A(u)\right)}$$

and

$$\frac{-\log(1-u)}{Q_Y(u)} - H_Y(u) = \frac{-\log(1-u)}{Q_X(u)} \frac{\log\left(1-A(u)\right)}{\log(1-u)} - H_Y(0)\,. \tag{3.3}$$

Also (2.6) leads to

$$\begin{aligned} H_X(0) &= \frac{H_Y(0)}{A'(0)} \quad (\text{since } A(0) = 0) \\ &= \frac{1}{\theta} - \left(\frac{1}{\theta} - 1\right)\frac{M_X'(0)}{\mu_X} \end{aligned} \tag{3.4}$$

and

$$\log\left(1 - A(u)\right) = \frac{1}{\theta}\log(1-u) + \left(\frac{1}{\theta} - 1\right)\log\frac{M_X(u)}{\mu_X}\,. \tag{3.5}$$

Using (3.4) and (3.5) in (3.3),

$$\begin{aligned} &\frac{-\log(1-u)}{Q_Y(u)} - H_Y(u) = \\ &\qquad = \frac{-\log(1-u)}{Q_X(u)}\left[\frac{\frac{1}{\theta} + \left(\frac{1}{\theta} - 1\right)\log\frac{M_X(u)}{\mu_X}}{\log(1-u)}\right] - A'(0)H_X(0) \\ &\qquad \geq H_X(0)\left[\frac{\frac{1}{\theta} + \left(\frac{1}{\theta} - 1\right)\log\frac{M_X(u)}{\mu_X}}{\log(1-u)} - \left(\frac{1}{\theta} - \left(\frac{1}{\theta} - 1\right)\right)\frac{M_X'(0)}{\mu_X}\right] \\ &\qquad = H_X(0)\left(\frac{1}{\theta} - 1\right)\left[\frac{\log\left(\frac{M_X(u)}{\mu_X}\right)}{\log(1-u)} + \frac{M_X'(0)}{\mu_X}\right]. \end{aligned}$$

Under the conditions assumed in Theorem 3, Y is NBUHRA. □

Theorem 3.3. *If X is DMRL and $\theta < 1$, then Y is DMRL.*

Proof: First we notice that

$$A'(u) = \frac{(1-u)^{\frac{1}{\theta}-1}\left(M_X(u)\right)^{\frac{1}{\theta}-2}}{(\mu_X)^{\frac{1}{\theta}-1}}\left[\frac{1}{\theta}M_X(u) - (1-u)\left(\frac{1}{\theta}-1\right)M'_X(u)\right]$$

implies that $A(u)$ is increasing and X is DMRL only when $\theta < 1$. Now

$$M'_X(u) = \theta M'_Y\left(A(u)\right)A'(u)$$

provides $M'_Y\left(A(u)\right) \leq 0$. When $M'_Y\left(A(u)\right)$ is decreasing so does $M_Y(u)$ and hence Y is DMRL. □

Remark 3.2. In the distribution function approach Y is DMRL if and only if X is DMRL ([9]) irrespective of the value of θ. In our case the restriction on θ cannot be dropped. For example, when X is beta with $\bar{F}_X(x) = (1-x)^2, 0 \leq x \leq 1$ and $\theta = 4$

$$M_X(u) = \frac{1}{3}(1-u)^{\frac{1}{2}}$$

which is decreasing, while

$$M_Y(u) = \frac{4}{3}(1-u)^{-4}$$

is increasing.

Theorem 3.4. *If X is DMRLHA if and only if Y is DMRLHA.*

Proof: X is DMRLHA $\iff$ $\dfrac{1}{Q_X(u)}\displaystyle\int_0^u \frac{q_X(p)dp}{M_X(p)}$ is decreasing in u

$\iff$ $\dfrac{1}{Q_Y\left(A(u)\right)}\displaystyle\int_0^u \frac{q_Y\left(A(p)\right)A'(p)dp}{M_Y\left(A(p)\right)}$ is decreasing in u

$\iff$ $\dfrac{1}{Q_Y(u)}\displaystyle\int_0^u \frac{q_Y(p)dp}{M_Y(p)}$ is decreasing in u

$\iff$ Y is DMRLHA. □

Theorem 3.5. *If X is NBU and $\theta < 1$ and $-\log\frac{M_X(u)}{\mu_X}$ is super additive, then Y is NBU.*

Proof:

$$-\log\frac{(1-u)M_X(u)}{\mu_X} = -\log(1-u) - \log\frac{M_X(u)}{\mu_X}\,.$$

Since X is NBU, $-\log\frac{(1-u)M_X(u)}{\mu_X}$ is super additive.

$$S_Y\left(Q_X(u)\right) = (1-u)\left[\frac{(1-u)M_X(u)}{\mu_X}\right]^{\frac{1}{\theta}-1}.$$

The right side is the product of two survival functions each of which is NBU and therefore $S_Y(\cdot)$ is NBU, which proves the result. □

4. CHARACTERIZATIONS

In this section we attempt two characterization theorems of the PMRLM by properties of measures of uncertainty in the residual lives of X and Y. The first measure is the Kullback–Leibler divergence between the residual life distributions of X and Y given by

$$i(t) = \frac{1}{\bar{F}_X(x)}\int_x^\infty \left(\log\frac{f_X(t)}{f_Y(t)}\right) f_X(t)dt + \log\frac{\bar{F}_Y(x)}{\bar{F}_X(x)}\,. \tag{4.1}$$

The quantile version of (4.1),

$$i\left(Q_X(u)\right) = I(u) = \log\left(\frac{\bar{F}_Y\left(Q_X(u)\right)}{1-u}\right) + \frac{1}{1-u}\int_u^1 \log\frac{d\bar{F}_Y\left(Q_X(p)\right)}{dp}dp\,, \tag{4.2}$$

was studied by [12] and several properties including characterization of PHM were obtained by them. In the following theorem we investigate the distributions satisfying PMRLM for which $I(u) = C$, a constant.

Theorem 4.1. *Let X and Y be continuous non-negative random variables as defined in Section 1 satisfying PMRLM. Then $I(u) = C$, a constant if and only if the distribution of X is either exponential with quantile function*

$$Q_E(u) = \lambda^{-1}\left(-\log(1-u)\right), \quad \lambda > 0\,,$$

or Pareto with

$$Q_P(u) = \alpha\left[(1-u)^{-\frac{1}{c}} - 1\right], \quad c > 1, \;\; \alpha > 0\,,$$

or beta having quantile function

$$Q_B(u) = \beta\left[1 - (1-u)^{\frac{1}{a}}\right], \quad a > 0, \;\; \beta > 0\,.$$

Proof: In the case of PMRLM, the divergence measure (4.2) reduces to

$$I(u) = \log \frac{1 - A(u)}{1 - u} - \frac{1}{1 - u} \int_u^1 \log A'(p) dp \,.$$

When X is exponential,

$$A(u) = 1 - (1 - u)^{\frac{1}{\theta}}$$

and hence

$$I(u) = \theta - \log \theta - 1 \,.$$

In the case of $Q_P(u)$, $A(u) = 1 - (1 - u)^{\frac{c+\theta-1}{c\theta}}$ and

$$I(u) = \frac{c + \theta - 1}{c\theta} - \log \frac{c + \theta - 1}{c\theta} - 1$$

and similarly for the beta distribution $A(u) = 1 - (1 - u)^{\frac{a-\theta+1}{a\theta}}$ gives

$$I(u) = \frac{a - \theta + 1}{a\theta} - \log \frac{a - \theta + 1}{a\theta} - 1 \,.$$

Thus $I(u)$ is a constant for all the three distributions. Conversely, when $I(u) = C$,

$$\log \frac{1 - A(u)}{1 - u} - \frac{1}{1 - u} \int_u^1 \log A'(p) dp = C$$

takes the form

$$\int_u^1 \log A'(p) dp = (1 - u) \left[\log \frac{1 - A(u)}{1 - u} - C \right].$$

Differentiating with respect to u and simplifying

$$P(u) = C + 1 + \log P(u) \tag{4.3}$$

where $P(u) = \frac{(1-u)A'(u)}{1-A(u)}$. Differentiating (4.3)

$$P'(u) \left[1 - \frac{1}{P(u)} \right] = 0$$

which leaves two solutions $P(u) = K$, a constant or $P(u) = 1$. Of these $P(u) = 1$ leads to $\frac{M'_X(u)}{M_X(u)} = \frac{1}{1-u}$ or $M_X(u) = \frac{K}{1-u}$ which cannot be mean residual quantile function of a proper distribution. The second solution $P(u) = K$, simplifies to $M_X(u) = K(1 - u)^{-b}$ which is the mean residual quantile function of the exponential or Pareto or beta distribution according as $b = 0$ or $b > 0$ or $b < 0$. This completes the proof. □

In the second theorem the choice of the uncertainty measure is the cumulative Kullback–Leibler divergence proposed by [1] for the residual lives of X and Y as

$$j(x) = \left(\log \frac{\bar{F}_Y(x)}{\bar{F}_X(x)} - 1 \right) m_X(x) + \frac{1}{\bar{F}_X(x)} \int_x^\infty \log \frac{\bar{F}_X(t)}{\bar{F}_Y(t)} \bar{F}_X(t) dt + m_Y(x) \,. \tag{4.4}$$

In terms of quantile functions, we can write (4.4) in the form

$$J(u) = j\left(Q_X(u)\right) = M_Y(u) - M_X(u) + \frac{1}{1-u}\int_u^1 \left(1 - \frac{H_Y\left(A(p)\right)}{H_X(p)}\right) M_X(p)dp\,.$$

Here $Q_X(\cdot)$ is taken as representing the true distribution and $Q_Y(\cdot)$ an arbitrary reference model. The measure $J(\cdot)$ provides the relative amount of uncertainty in the residual life of Y in comparison with that of X. We prove a theorem that identifies the class of distributions for which this relative entropy is a constant.

Theorem 4.2. *The cumulative divergence measure $J(u) = C$ for all u in $[0,1]$ if and only if the quantile function of Y admits the representation*

$$Q_Y(u) = Q_1(u) + Q_2(u)$$

where $Q_1(\cdot)$ is the quantile function of the exponential distribution with mean C and $Q_2(\cdot)$ is the quantile function of $\frac{X}{\theta}$.

Proof: When $J(u) = C$, we have

$$M_Y(u) - M_X(u) + \frac{1}{1-u}\int_u^1 \left(1 - \frac{h_Y\left(Q_X(p)\right)}{H_X(p)}\right) M_X(p)dp = C\,. \tag{4.5}$$

From (2.2)

$$h_Y\left(Q_X(u)\right) = H_X(u) + \frac{1-\theta}{\theta M_X(u)}$$

and so $1 - \frac{h_Y(Q_X(u))}{H_X(u)} = \frac{\theta-1}{\theta}\frac{1}{M_X(u)H_X(u)}$. Hence

$$\begin{aligned}\int_u^1 \left(1 - \frac{h_Y\left(Q_X(p)\right)}{H_X(p)}\right) M_X(p)dp &= \frac{\theta-1}{\theta}\int_u^1 \frac{1}{H_X(p)}dp \\ &= \frac{\theta-1}{\theta}\int_u^1 -\frac{d}{dp}(1-p)M_X(p)dp \\ &= \frac{\theta-1}{\theta}(1-u)M_X(u).\end{aligned} \tag{4.6}$$

Inserting (4.6) in (4.5) and simplifying

$$M_Y(u) = C + \frac{M_X(u)}{\theta}\,, \tag{4.7}$$

or

$$\int_u^1 (1-p)q_Y(p)dp = C(1-u) + \frac{1}{\theta}\int_u^1 (1-p)q_X(p)dp\,,$$

giving

$$q_Y(u) = \frac{C}{1-u} + \frac{1}{\theta}q_X(u)\,.$$

Integrating from 0 to u

$$Q_Y(u) = -C\log(1-u) + \frac{1}{\theta}Q_X(u) \tag{4.8}$$
$$= Q_1(u) + Q_2(u)\,,$$

as stated. Conversely assuming (4.8) we have (4.7) and substituting this in the expression on the left side of (4.5) we have the result stated and this completes the proof. □

Corollary 4.1.

1. *When X is exponential (λ), Y is also exponential with parameter $\frac{1+C\lambda\theta}{\lambda\theta}$.*
2. *When X has linear mean residual quantile distribution ([5]) with*

$$Q_X(u) = -(\alpha+\mu)\log(1-u) - 2\alpha u\,, \quad \mu > 0, \quad -\mu < \alpha < \mu\,,$$

and

$$M_X(u) = \alpha u + \mu\,,$$

Y also has the same form of distribution with linear mean residual quantile function

$$M_X(u) = \frac{\alpha}{\theta}u + \left(C + \frac{\mu}{\theta}\right).$$

It is noted that (4.8) gives a class of distributions many of which do not possess a closed form distribution functions, so that it is difficult to arrive such forms using the distribution function approach.

We conclude this work by noting that here we have proposed an alternative approach in analysing PMRLM through quantile functions. This brings in some new results and models that are sometimes difficult to arrive at by using the traditional approach.

ACKNOWLEDGMENTS

The authors wish to thank the editor and referees for their constructive comments. The second and third authors wish to thank the support of the University Grants Commission, India, under the Special Assistance Programme.

REFERENCES

[1] Chamany, A. and Baratpour, S. (2014). A dynamic discrimination information based on cumulative residual entropy and its properties, *Communications in Statistics — Theory and Methods*, **43**, 1041–1049.

[2] Franco-Pereira, A.M.; Lillo, R.E. and Shaked, M. (2012). The decreasing percentile residual ageing notion, *Statistics*, **46**, 587–603.

[3] Gilchrist, W.G. (2000). *Statistical Modelling with Quantile Functions*, Chapman and Hall/CRC Press, Boca Raton.

[4] Gupta, R.C. and Kirmani, S.N.U.A. (1998). On the proportional mean residual life model and its implications, *Statistics*, **32**, 175–187.

[5] Midhu, N.N.; Sankaran, P.G. and Nair, N.U. (2013). A class of distributions with linear mean residual quantile function and its generalizations, *Statistical Methodology*, **15**, 1–24.

[6] Midhu, N.N.; Sankaran, P.G. and Nair, N.U. (2014). A class of distributions with linear hazard quantile functions, *Communications in Statistics — Theory and Methods*, **43**, 3674–3689.

[7] Nair, N.U. and Sankaran, P.G. (2009). Quantile-based reliability analysis, *Communications in Statistics — Theory and Methods*, **38**(2), 222–232.

[8] Nair, N.U.; Sankaran, P.G. and Balakrishnan, N. (2013). *Quantile-based Reliability Analysis*, Birkhäuser, Springer, New York.

[9] Nanda, A.K.; Bhattacharjee, S. and Alam, S.S. (2006). Properties of proportional mean residual life model, *Statistics and Probability Letters*, **76**(9), 880–890.

[10] Nanda, A.K.; Das, S. and Balakrishnan, N. (2013). On dynamic proportional mean residual life model, *Probability in the Engineering and Informational Sciences*, **27**(4), 553–588.

[11] Oakes, D. and Dasu, T. (1990). A note on residual life, *Biometrika*, **77**(2), 409–410.

[12] Sankaran, P.G.; Sunoj, S.M. and Nair, N.U. (2016). Kullback–Leibler divergence: A quantile approach, *Statistics and Probability Letters*, **111**, 72–79.

[13] Song, J.K. and Cho, G.Y. (1995). A note on percentile residual life, *Sankhya*, **57**, 333–335.

[14] Zahedi, H. (1991). Proportional mean remaining life model, *Journal of Statistical Planning and Inference*, **29**(1–2), 221–228.

[15] Zimmer, W.; Keats, J.B. and Wang, F.K. (1998). The Burr XII distribution in reliability analysis, *Journal of Quality Technology*, **30**, 386–394.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 493–521

GENERALIZED ESTIMATORS OF STATIONARY RANDOM-COEFFICIENTS PANEL DATA MODELS: ASYMPTOTIC AND SMALL SAMPLE PROPERTIES

Author: Mohamed Reda Abonazel
– Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt
mabonazel@hotmail.com

Received: April 2016 Revised: July 2017 Accepted: July 2017

Abstract:

- This article provides generalized estimators for the random-coefficients panel data (RCPD) model where the errors are cross-sectional heteroskedastic and contemporaneously correlated as well as with the first-order autocorrelation of the time series errors. Of course, under the new assumptions of the error, the conventional estimators are not suitable for RCPD model. Therefore, the suitable estimator for this model and other alternative estimators have been provided and examined in this article. Furthermore, the efficiency comparisons for these estimators have been carried out in small samples and also we examine the asymptotic distributions of them. The Monte Carlo simulation study indicates that the new estimators are more efficient than the conventional estimators, especially in small samples.

Key-Words:

- *classical pooling estimation; contemporaneous covariance; first-order autocorrelation; heteroskedasticity; mean group estimation; random coefficient regression.*

AMS Subject Classification:

- 91G70, 97K80.

1. INTRODUCTION

The econometrics literature reveals a type of data called "panel data", which refers to the pooling of observations on a cross-section of households, countries, and firms over several time periods. Pooling this data achieves a deep analysis of the data and gives a richer source of variation which allows for more efficient estimation of the parameters. With additional, more informative data, one can get more reliable estimates and test more sophisticated behavioral models with less restrictive assumptions. Also, panel data sets are more effective in identifying and estimating effects that are simply not detectable in pure cross-sectional or pure time series data. In particular, panel data sets are more effective in studying complex issues of dynamic behavior. Some of the benefits and limitations of using panel data sets are listed in Baltagi (2013) and Hsiao (2014).

The pooled least squares (classical pooling) estimator for pooled cross-sectional and time series data (panel data) models is the best linear unbiased estimator (BLUE) under the classical assumptions as in the general linear regression model.[1] An important assumption for panel data models is that the individuals in our database are drawn from a population with a common regression coefficient vector. In other words, the coefficients of a panel data model must be fixed. In fact, this assumption is not satisfied in most economic models, see, e.g., Livingston *et al.* (2010) and Alcacer *et al.* (2013). In this article, the panel data models are studied when this assumption is relaxed. In this case, the model is called "random-coefficients panel data (RCPD) model". The RCPD model has been examined by Swamy in several publications (Swamy 1970, 1973, and 1974), Rao (1982), Dielman (1992a, b), Beck and Katz (2007), Youssef and Abonazel (2009), and Mousa *et al.* (2011). Some statistical and econometric publications refer to this model as Swamy's model or as the random coefficient regression (RCR) model, see, e.g., Poi (2003), Abonazel (2009), and Elhorst (2014, ch.3). In RCR model, Swamy assumes that the individuals in our panel data are drawn from a population with a common regression parameter, which is a fixed component, and a random component, that will allow the coefficients to differ from unit to unit. This model has been developed by many researchers, see, e.g., Beran and Millar (1994), Chelliah (1998), Anh and Chelliah (1999), Murtazashvili and Wooldridge (2008), Cheng *et al.* (2013), Fu and Fu (2015), Elster and Wübbeler (2017), and Horváth and Trapani (2016).

The random-coefficients models have been applied in different fields and they constitute a unifying setup for many statistical problems. Moreover, several applications of Swamy's model have appeared in the literature of finance and

[1] Dielman (1983, 1989) discussed these assumptions. In the next section in this article, we will discuss different types of classical pooling estimators under different assumptions.

economics.[2] Boot and Frankfurter (1972) used the RCR model to examine the optimal mix of short and long-term debt for firms. Feige and Swamy (1974) applied this model to estimate demand equations for liquid assets, while Boness and Frankfurter (1977) used it to examine the concept of risk-classes in finance. Recently, Westerlund and Narayan (2015) used the random-coefficients approach to predictăthe stock returns at the New York Stock Exchange. Swamy *et al.* (2015) applied a random-coefficient framework to deal with two problems frequently encountered in applied work; these problems are correcting for misspecifications in a small area level model and resolving Simpson's paradox.

Dziechciarz (1989) and Hsiao and Pesaran (2008) classified the random-coefficients models into two categories (stationary and non-stationary models), depending on the type of assumption about the coefficient variation. Stationary random-coefficients models regard the coefficients as having constant means and variance-covariances, like Swamy's (1970) model. On the other hand, the coefficients in non-stationary random-coefficients models do not have a constant mean and/or variance and can vary systematically; these models are relevant mainly for modeling the systematic structural variation in time, like the Cooley–Prescott (1973) model.[3]

The main objective of this article is to provide the researchers with general and more efficient estimators for the stationary RCPD models. To achieve this objective, we propose and examine alternative estimators of these models under an assumption that the errors are cross-sectional heteroskedastic and contemporaneously correlated as well as with the first-order autocorrelation of the time series errors.

The rest of the article is organized as follows. Section 2 presents the classical pooling (CP) estimators of fixed-coefficients models. Section 3 provides generalized least squares (GLS) estimators of the different random-coefficients models. In section 4, we examine the efficiency of these estimators, theoretically. In section 5, we discuss alternative estimators for these models. The Monte Carlo comparisons between various estimators have been carried out in section 6. Finally, section 7 offers the concluding remarks.

[2]The RCR model has been applied also in different sciences fields, see, e.g., Bodhlyera *et al.* (2014).

[3]Cooley and Prescott (1973) suggested a model where coefficients vary from one time period to another on the basis of a non-stationary process. Similar models have been considered by Sant (1977) and Rausser *et al.* (1982).

2. FIXED-COEFFICIENTS MODELS

Suppose the variable y for the ith cross-sectional unit at time period t is specified as a linear function of K strictly exogenous variables, x_{kit}, in the following form:

$$(2.1) \qquad y_{it} = \sum_{k=1}^{K} \alpha_{ki} x_{kit} + u_{it} = \mathrm{x}_{it}\alpha_i + u_{it}\,, \qquad i = 1, 2, ..., N; \quad t = 1, 2, ..., T\,,$$

where u_{it} denotes the random error term, x_{it} is a $1 \times K$ vector of exogenous variables, and α_i is the $K \times 1$ vector of coefficients. Stacking equation (2.1) over time, we obtain:

$$(2.2) \qquad y_i = X_i\alpha_i + u_i\,,$$

where $y_i = (y_{i1},\ ...,\ y_{iT})'$, $X_i = (\mathrm{x}'_{i1}, ...,\ \mathrm{x}'_{iT})'$, $\alpha_i = (\alpha_{i1},\ ..., \alpha_{iK})'$, and $u_i = (u_{i1},\ ..., u_{iT})'$.

When the performance of one individual from the database is of interest, separate equation regressions can be estimated for each individual unit using the ordinary least squares (OLS) method. The OLS estimator of α_i, is given by:

$$(2.3) \qquad \hat{\alpha}_i = \left(X_i'X_i\right)^{-1} X_i'y_i\,.$$

Under the following assumptions, $\hat{\alpha}_i$ is a BLUE of α_i:

Assumption 1: The errors have zero mean, i.e., $E(u_i) = 0; \forall\ i = 1, 2, ..., N$.

Assumption 2: The errors have the same variance for each individual:

$$E\left(u_i u_j'\right) = \begin{cases} \sigma_u^2 I_T & \text{if } \ i = j \\ 0 & \text{if } \ i \neq j \end{cases} \qquad i, j = 1, 2, ..., N\,.$$

Assumption 3: The exogenous variables are non-stochastic, i.e., fixed in repeated samples, and hence, not correlated with the errors. Also, $rank\,(X_i) = K < T;\ \forall\ i = 1, 2, ..., N$.

These conditions are sufficient but not necessary for the optimality of the OLS estimator.[4] When OLS is not optimal, estimation can still proceed equation by equation in many cases. For example, if variance of u_i is not constant, the errors are either heteroskedastic and/or serially correlated, and the GLS method will provide relatively more efficient estimates than OLS, even if GLS was applied to each equation separately as in OLS.

[4]For more information about the optimality of the OLS estimators, see, e.g., Rao and Mitra (1971, ch. 8) and Srivastava and Giles (1987, pp. 17–21).

Another case, If the covariances between u_i and u_j $(i, j = 1, 2, ..., N)$ do not equal to zero as in assumption (2) above, then contemporaneous correlation is present, and we have what Zellner (1962) termed as seemingly unrelated regression (SUR) equations, where the equations are related through cross-equation correlation of errors. If the $X_i\,(i = 1, 2, ..., N)$ matrices do not span the same column space and contemporaneous correlation exists, a relatively more efficient estimator of α_i than equation by equation OLS is the GLS estimator applied to the entire equation system, as shown in Zellner (1962).

With either separate equation estimation or the SUR methodology, we obtain parameter estimates for each individual unit in the database. Now suppose it is necessary to summarize individual relationships and to draw inferences about certain population parameters. Alternatively, the process may be viewed as building a single model to describe the entire group of individuals rather than building a separate model for each. Again, assume that assumptions 1–3 are satisfied and add the following assumption:

Assumption 4: The individuals in the database are drawn from a population with a common regression parameter vector $\bar{\alpha}$, i.e., $\alpha_1 = \alpha_2 = \cdots = \alpha_N = \bar{\alpha}$.

Under this assumption, the observations for each individual can be pooled, and a single regression performed to obtain an efficient estimator of $\bar{\alpha}$. Now, the equation system is written as:

$$Y = X\bar{\alpha} + u\,, \tag{2.4}$$

where $Y = (y_1', ..., y_N')'$, $X = (X_1', ..., X_N')'$, $u = (u_1', ..., u_N')'$, and $\bar{\alpha} = (\bar{\alpha}_1, ..., \bar{\alpha}_K)'$ is a vector of fixed coefficients which to be estimated. We will differentiate between two cases to estimate $\bar{\alpha}$ in (2.4) based on the variance-covariance structure of u. In the first case, the errors have the same variance for each individual as given in assumption 2. In this case, the efficient and unbiased estimator of $\bar{\alpha}$ under assumptions 1–4 is:

$$\widehat{\bar{\alpha}}_{CP-OLS} = \left(X'X\right)^{-1}X'Y.$$

This estimator has been termed the classical pooling-ordinary least squares (CP-OLS) estimator. In the second case, which the errors have different variances along individuals and are contemporaneously correlated as in the SUR framework:

Assumption 5: $E\left(u_i u_j'\right) = \begin{cases} \sigma_{ii} I_T & \text{if } i = j \\ \sigma_{ij} I_T & \text{if } i \neq j \end{cases} \quad i, j = 1, 2, ..., N.$

Under assumptions 1, 3, 4 and 5, the efficient and unbiased CP estimator of $\bar{\alpha}$ is:

$$\widehat{\bar{\alpha}}_{CP-SUR} = \left[X'(\Sigma_{sur} \otimes I_T)^{-1} X\right]^{-1} \left[X'(\Sigma_{sur} \otimes I_T)^{-1} Y\right],$$

where

$$\Sigma_{sur} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{pmatrix}.$$

To make this estimator ($\widehat{\bar{\alpha}}_{CP-SUR}$) a feasible, the σ_{ij} can be replaced with the following unbiased and consistent estimator:

$$\hat{\sigma}_{ij} = \frac{\hat{u}_i' \hat{u}_j}{T - K}; \quad \forall \ i, j = 1, 2, ..., N, \tag{2.5}$$

where $\hat{u}_i = y_i - X_i \hat{\alpha}_i$ is the residuals vector obtained from applying OLS to equation number i.[5]

3. RANDOM-COEFFICIENTS MODELS

This section reviews the standard random-coefficients model proposed by Swamy (1970), and presents the random-coefficients model in the general case, where the errors are allowed to be cross-sectional heteroskedastic and contemporaneously correlated as well as with the first-order autocorrelation of the time series errors.

3.1. RCR model

Suppose that each regression coefficient in (2.2) is now viewed as a random variable; that is the coefficients, α_i, are viewed as invariant over time, but varying from one unit to another:

Assumption 6 (for the stationary random-coefficients approach): The coefficient vector α_i is specified as:[6] $\alpha_i = \bar{\alpha} + \mu_i$, where $\bar{\alpha}$ is a $K \times 1$ vector of

[5]The $\hat{\sigma}_{ij}$ in (2.5) are unbiased estimators because, as assumed, the number of exogenous variables of each equation is equal, i.e., $K_i = K$ for $i = 1, 2, ..., N$. However, in the general case, $K_i \neq K_j$, the unbiased estimator is $\hat{u}_i' \hat{u}_j / \left[T - K_i - K_j + tr\left(P_{xx}\right)\right]$, where $P_{xx} = X_i (X_i' X_i)^{-1} X_i' X_j \left(X_j' X_j\right)^{-1} X_j'$. See Srivastava and Giles (1987, pp. 13–17) and Baltagi (2011, pp. 243–244).

[6]This means that the individuals in our database are drown from a population with a common regression parameter $\bar{\alpha}$, which is a fixed component, and a random component μ_i, allowed to differ from unit to unit.

constants, and μ_i is a $K \times 1$ vector of stationary random variables with zero means and constant variance-covariances:

$$E(\mu_i) = 0 \text{ and } E(\mu_i \mu_j') = \begin{cases} \Psi & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad i, j = 1, 2, ..., N,$$

where $\Psi = diag\{\psi_k^2\}$; for $k = 1, 2, ..., K$, where $K < N$. Furthermore, $E(\mu_i \mathrm{x}_{jt}) = 0$ and $E(\mu_i u_{jt}) = 0 \; \forall \; i$ and j.

Also, Swamy (1970) assumed that the errors have different variances along individuals:

Assumption 7: $E\left(u_i u_j'\right) = \begin{cases} \sigma_{ii} I_T & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad i, j = 1, 2, ..., N.$

Under the assumption 6, the model in equation (2.2) can be rewritten as:

$$Y = X\bar{\alpha} + e; \;\; e = D\mu + u, \tag{3.1}$$

where Y, X, u, and $\bar{\alpha}$ are defined in (2.4), while $\mu = (\mu_1', \; ..., \; \mu_N')'$, and $D = diag\{X_i\}$; for $i = 1, 2, ..., N$.

The model in (3.1), under assumptions 1, 3, 6 and 7, called the 'RCR model', which was examined by Swamy (1970, 1971, 1973, and 1974), Youssef and Abonazel (2009), and Mousa *et al.* (2011). We will refer to assumptions 1, 3, 6 and 7 as RCR assumptions. Under these assumptions, the BLUE of $\bar{\alpha}$ in equation (3.1) is:

$$\widehat{\bar{\alpha}}_{RCR} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y,$$

where Ω is the variance-covariance matrix of e:

$$\Omega = (\Sigma_{rcr} \otimes I_T) + D(I_N \otimes \Psi) D',$$

where $\Sigma_{rcr} = diag\{\sigma_{ii}\}$; for $i = 1, 2, ..., N$. Swamy (1970) showed that the $\widehat{\bar{\alpha}}_{RCR}$ estimator can be rewritten as:

$$\widehat{\bar{\alpha}}_{RCR} = \left[\sum_{i=1}^{N} X_i'\left(X_i \Psi X_i' + \sigma_{ii} I_T\right)^{-1} X_i\right]^{-1} \sum_{i=1}^{N} X_i'\left(X_i \Psi X_i' + \sigma_{ii} I_T\right)^{-1} y_i.$$

The variance-covariance matrix of $\widehat{\bar{\alpha}}_{RCR}$ under RCR assumptions is:

$$var\left(\widehat{\bar{\alpha}}_{RCR}\right) = \left(X'\Omega^{-1}X\right)^{-1} = \left\{\sum_{i=1}^{N}\left[\Psi + \sigma_{ii}\left(X_i'X_i\right)^{-1}\right]^{-1}\right\}^{-1}.$$

To make the $\widehat{\bar{\alpha}}_{RCR}$ estimator feasible, Swamy (1971) suggested using the estimator in (2.5) as an unbiased and consistent estimator of σ_{ii}, and the following unbiased estimator for Ψ:

$$\hat{\Psi} = \left[\frac{1}{N-1}\left(\sum_{i=1}^{N} \hat{\alpha}_i \, \hat{\alpha}_i' - \frac{1}{N}\sum_{i=1}^{N} \hat{\alpha}_i \sum_{i=1}^{N} \hat{\alpha}_i'\right)\right] - \left[\frac{1}{N}\sum_{i=1}^{N} \hat{\sigma}_{ii}\left(X_i'X_i\right)^{-1}\right]. \tag{3.2}$$

Swamy (1973, 1974) showed that the estimator $\widehat{\bar{\alpha}}_{RCR}$ is consistent as both $N, T \to \infty$ and is asymptotically efficient as $T \to \infty$.[7]

It is worth noting that, just as in the error-components model, the estimator (3.2) is not necessarily non-negative definite. Mousa *et al.* (2011) explained that it is possible to obtain negative estimates of Swamy's estimator in (3.2) in case of small samples and if some/all coefficients are fixed. But in medium and large samples, the negative variance estimates does not appear even if all coefficients are fixed. To solve this problem, Swamy has suggested replacing (3.2) by:[8]

$$\hat{\Psi}^{+} = \frac{1}{N-1}\left(\sum_{i=1}^{N} \hat{\alpha}_i \, \hat{\alpha}_i' - \frac{1}{N}\sum_{i=1}^{N} \hat{\alpha}_i \, \sum_{i=1}^{N} \hat{\alpha}_i'\right).$$

This estimator, although biased, is non-negative definite and consistent when $T \to \infty$. See Judge *et al.* (1985, p. 542).

3.2. Generalized RCR model

To generalize RCR model so that it would be more suitable for most economic models, we assume that the errors are cross-sectional heteroskedastic and contemporaneously correlated, as in assumption 5, as well as with the first-order autocorrelation of the time series errors. Therefore, we add the following assumption to assumption 5:

Assumption 8: $u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it}$; $|\rho_i| < 1$, where ρ_i $(i = 1, 2, ..., N)$ are fixed first-order autocorrelation coefficients. Assume that: $E(\varepsilon_{it}) = 0$, $E(u_{i,t-1}\varepsilon_{jt}) = 0$; $\forall\, i$ and j, and

$$E\left(\varepsilon_i \varepsilon_j'\right) = \begin{cases} \sigma_{\varepsilon_{ii}} I_T & \text{if } i = j \\ \sigma_{\varepsilon_{ij}} I_T & \text{if } i \neq j \end{cases} \qquad i, j = 1, 2, ..., N.$$

This means that the initial time period of the errors have the same properties as in subsequent periods, i.e., $E\left(u_{i0}^2\right) = \sigma_{\varepsilon_{ii}} / \left(1 - \rho_i^2\right)$ and $E(u_{i0}u_{j0}) = \sigma_{\varepsilon_{ij}} / (1 - \rho_i \rho_j)$ $\forall\, i$ and j.

We will refer to assumptions 1, 3, 5, 6, and 8 as the general RCR assumptions. Under these assumptions, the BLUE of $\bar{\alpha}$ is:

$$\widehat{\bar{\alpha}}_{GRCR} = \left(X'\Omega^{*-1}X\right)^{-1} X'\Omega^{*-1}Y,$$

[7] The statistical properties of $\widehat{\bar{\alpha}}_{RCR}$ have been examined by Swamy (1971), of course, under RCR assumptions.

[8] This suggestion has been used by Stata program, specifically in xtrchh and xtrchh2 Stata's commands. See Poi (2003).

where

(3.3)
$$\Omega^* = \begin{pmatrix} X_1\Psi X_1' + \sigma_{\varepsilon_{11}}\omega_{11} & \sigma_{\varepsilon_{12}}\omega_{12} & \cdots & \sigma_{\varepsilon_{1N}}\omega_{1N} \\ \sigma_{\varepsilon_{21}}\omega_{21} & X_2\Psi X_2' + \sigma_{\varepsilon_{22}}\omega_{22} & \cdots & \sigma_{\varepsilon_{2N}}\omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\varepsilon_{N1}}\omega_{N1} & \sigma_{\varepsilon_{N2}}\omega_{N2} & \cdots & X_N\Psi X_N' + \sigma_{\varepsilon_{NN}}\omega_{NN} \end{pmatrix},$$

with

(3.4)
$$\omega_{ij} = \frac{1}{1-\rho_i\rho_j}\begin{pmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{T-1} \\ \rho_j & 1 & \rho_i & \cdots & \rho_i^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T-1} & \rho_j^{T-2} & \rho_j^{T-3} & \cdots & 1 \end{pmatrix}.$$

Since the elements of Ω^* are usually unknown, we develop a feasible Aitken estimator of $\bar{\alpha}$ based on consistent estimators of the elements of Ω^*:

(3.5)
$$\hat{\rho}_i = \frac{\sum_{t=2}^{T}\hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{t=2}^{T}\hat{u}_{i,t-1}^2},$$

where $\hat{u}_i = (\hat{u}_{i1},\ ...,\hat{u}_{iT})'$ is the residuals vector obtained from applying OLS to equation number i,

$$\hat{\sigma}_{\varepsilon_{ij}} = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{T-K},$$

where $\hat{\varepsilon}_i = (\hat{\varepsilon}_{i1},...,\hat{\varepsilon}_{iT})'$; $\hat{\varepsilon}_{i1} = \hat{u}_{i1}\sqrt{1-\hat{\rho}_i^2}$, and $\hat{\varepsilon}_{it} = \hat{u}_{it} - \hat{\rho}_i\hat{u}_{i,t-1}$ for $t = 2,...,T$.

Replacing ρ_i by $\hat{\rho}_i$ in (3.4), yields consistent estimators of ω_{ij}, say $\hat{\omega}_{ij}$, which leads together with $\hat{\sigma}_{\varepsilon_{ij}}$ and $\hat{\omega}_{ij}$ to a consistent estimator of Ψ:[9]

(3.6)
$$\begin{aligned}\hat{\Psi}^* = {} & \tfrac{1}{N-1}\left(\sum_{i=1}^{N}\hat{\alpha}_i^*\hat{\alpha}_i^{*\prime} - \tfrac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i^*\sum_{i=1}^{N}\hat{\alpha}_i^{*\prime}\right) - \tfrac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon_{ii}}\left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1} \\ & + \begin{bmatrix} \frac{1}{N(N-1)} \sum\limits_{\substack{i\neq j \\ i,j=1}}^{N} \hat{\sigma}_{\varepsilon_{ij}}\left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1} \\ X_i'\hat{\omega}_{ii}^{-1}\hat{\omega}_{ij}\hat{\omega}_{jj}^{-1}X_j\left(X_j'\hat{\omega}_{jj}^{-1}X_j\right)^{-1} \end{bmatrix},\end{aligned}$$

where

(3.7)
$$\hat{\alpha}_i^* = \left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1}X_i'\hat{\omega}_{ii}^{-1}y_i.$$

By using the consistent estimators ($\hat{\sigma}_{\varepsilon_{ij}}$, $\hat{\omega}_{ij}$, and $\hat{\Psi}^*$) in (3.3), and proceed a consistent estimator of Ω^* is obtained, say $\hat{\Omega}^*$, that leads to get the generalized

[9]The estimator of ρ_i in (3.5) is consistent, but it is not unbiased. See Srivastava and Giles (1987, p. 211) for other suitable consistent estimators of ρ_i that are often used in practice.

RCR (GRCR) estimator of $\bar{\alpha}$:

$$\widehat{\bar{\alpha}}_{GRCR} = \left(X'\hat{\Omega}^{*-1}X\right)^{-1}X'\hat{\Omega}^{*-1}Y.$$

The estimated variance-covariance matrix of $\widehat{\bar{\alpha}}_{GRCR}$ is:

$$\widehat{var}\left(\widehat{\bar{\alpha}}_{GRCR}\right) = \left(X'\hat{\Omega}^{*-1}X\right)^{-1}. \tag{3.8}$$

4. EFFICIENCY GAINS

In this section, we examine the efficiency gains from the use of GRCR estimator. Under the general RCR assumptions, It is easy to verify that the classical pooling estimators ($\widehat{\bar{\alpha}}_{CP-OLS}$ and $\widehat{\bar{\alpha}}_{CP-SUR}$) and Swamy's estimator ($\widehat{\bar{\alpha}}_{RCR}$) are unbiased for $\bar{\alpha}$ and with variance-covariance matrices:

$$\begin{aligned} var\left(\widehat{\bar{\alpha}}_{CP-OLS}\right) &= G_1\Omega^*G_1'; \\ var\left(\widehat{\bar{\alpha}}_{CP-SUR}\right) &= G_2\Omega^*G_2'; \\ var\left(\widehat{\bar{\alpha}}_{RCR}\right) &= G_3\Omega^*G_3', \end{aligned}$$

where

$$\begin{aligned} G_1 &= \left(X'X\right)^{-1}X'; \\ G_2 &= \left[X'\left(\Sigma_{sur}^{-1}\otimes I_T\right)X\right]^{-1}X'\left(\Sigma_{sur}^{-1}\otimes I_T\right); \\ G_3 &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}. \end{aligned} \tag{4.1}$$

The efficiency gains, from the use of GRCR estimator, can be summarized in the following equation:

$$EG_\gamma = var\left(\widehat{\bar{\alpha}}_\gamma\right) - var\left(\widehat{\bar{\alpha}}_{GRCR}\right) = \left(G_h - G_0\right)\Omega^*(G_h - G_0)'; \text{ for } h = 1, 2, 3,$$

where the subscript γ indicates the estimator that is used (CP-OLS, CP-SUR, or RCR), $G_0 = \left(X'\Omega^{*-1}X\right)^{-1}X'\Omega^{*-1}$, and G_h (for $h = 1, 2, 3$) matrices are defined in (4.1).

Since Ω^*, Σ_{rcr}, Σ_{sur} and Ω are positive definite matrices, then EG_γ matrices are positive semi-definite matrices. In other words, the GRCR estimator is more efficient than CP-OLS, CP-SUR, and RCR estimators. These efficiency gains increase when $|\rho_i|$, $\sigma_{\varepsilon_{ij}}$, and ψ_k^2 increase. However, it is not clear to what extent these efficiency gains hold in small samples. Therefore, this will be examined in a simulation study.

5. ALTERNATIVE ESTIMATORS

A consistent estimator of $\bar{\alpha}$ can also be obtained under more general assumptions concerning α_i and the regressors. One such possible estimator is the mean group (MG) estimator, proposed by Pesaran and Smith (1995) for estimation of dynamic panel data (DPD) models with random coefficients.[10] The MG estimator is defined as the simple average of the OLS estimators:

$$\widehat{\bar{\alpha}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i. \tag{5.1}$$

Even though the MG estimator has been used in DPD models with random coefficients, it will be used here as one of alternative estimators of static panel data models with random coefficients. Note that the simple MG estimator in (5.1) is more suitable for the RCR Model. But to make it suitable for the GRCR model, we suggest a general mean group (GMG) estimator as:

$$\widehat{\bar{\alpha}}_{GMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i^*, \tag{5.2}$$

where $\hat{\alpha}_i^*$ is defined in (3.7).

Lemma 5.1. *If the general RCR assumptions are satisfied, then $\widehat{\bar{\alpha}}_{MG}$ and $\widehat{\bar{\alpha}}_{GMG}$ are unbiased estimators of $\bar{\alpha}$, with the estimated variance-covariance matrices of $\widehat{\bar{\alpha}}_{MG}$ and $\widehat{\bar{\alpha}}_{GMG}$ are:*

$$\begin{aligned}\widehat{var}\left(\widehat{\bar{\alpha}}_{MG}\right) = \; & \frac{1}{N}\hat{\Psi}^* + \frac{1}{N^2}\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon_{ii}}\left(X_i'X_i\right)^{-1}X_i'\hat{\omega}_{ii}X_i\left(X_i'X_i\right)^{-1} \\ & + \frac{1}{N^2}\sum_{\substack{i\neq j \\ i,j=1}}^{N}\hat{\sigma}_{\varepsilon_{ij}}\left(X_i'X_i\right)^{-1}X_i'\hat{\omega}_{ij}X_j\left(X_j'X_j\right)^{-1},\end{aligned} \tag{5.3}$$

$$\widehat{var}\left(\widehat{\bar{\alpha}}_{GMG}\right) = \frac{1}{N(N-1)}\begin{bmatrix} \sum_{i=1}^{N}\hat{\alpha}_i^*\hat{\alpha}_i^{*\prime} - \frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i^*\sum_{i=1}^{N}\hat{\alpha}_i^{*\prime} \\ + \sum_{\substack{i\neq j \\ i,j=1}}^{N}\hat{\sigma}_{\varepsilon_{ij}}\left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1} \\ X_i'\hat{\omega}_{ii}^{-1}\hat{\omega}_{ij}\hat{\omega}_{jj}^{-1}X_j\left(X_j'\hat{\omega}_{jj}^{-1}X_j\right)^{-1} \end{bmatrix}. \tag{5.4}$$

[10] For more information about the estimation methods for DPD models, see, e.g., Baltagi (2013), Abonazel (2014, 2017), Youssef *et al.* (2014a,b), and Youssef and Abonazel (2017).

Proof of Lemma 5.1:

A. Unbiasedness property of MG and GMG estimators:

Proof: By substituting (3.7) and (2.2) into (5.2):

$$\begin{aligned}\widehat{\bar{\alpha}}_{GMG} &= \frac{1}{N}\sum_{i=1}^{N}\left(X_i'\omega_{ii}^{-1}X_i\right)^{-1}X_i'\omega_{ii}^{-1}\left(X_i\alpha_i+u_i\right)\\ &= \frac{1}{N}\sum_{i=1}^{N}\alpha_i+\left(X_i'\omega_{ii}^{-1}X_i\right)^{-1}X_i'\omega_{ii}^{-1}u_i.\end{aligned} \tag{5.5}$$

Similarly, we can rewrite $\widehat{\bar{\alpha}}_{MG}$ in (5.1) as:

$$\widehat{\bar{\alpha}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i+\left(X_i'X_i\right)^{-1}X_i'u_i. \tag{5.6}$$

Taking the expectation for (5.5) and (5.6), and using assumptions 1 and 6:

$$E\left(\widehat{\bar{\alpha}}_{GMG}\right) = E\left(\widehat{\bar{\alpha}}_{MG}\right) = \frac{1}{N}\sum_{i=1}^{N}\bar{\alpha} = \bar{\alpha}. \qquad \square$$

B. Derive the variance-covariance matrix of GMG:

Proof: Note first that under assumption 6, $\alpha_i = \bar{\alpha} + \mu_i$. Add $\hat{\alpha}_i^*$ to the both sides:

$$\alpha_i + \hat{\alpha}_i^* = \bar{\alpha} + \mu_i + \hat{\alpha}_i^*,$$

$$\hat{\alpha}_i^* = \bar{\alpha} + \mu_i + \hat{\alpha}_i^* - \alpha_i = \bar{\alpha} + \mu_i + \tau_i, \tag{5.7}$$

where $\tau_i = \hat{\alpha}_i^* - \alpha_i = \left(X_i'\omega_{ii}^{-1}X_i\right)^{-1}X_i'\omega_{ii}^{-1}u_i$. From (5.7):

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i^* = \bar{\alpha} + \frac{1}{N}\sum_{i=1}^{N}\mu_i + \frac{1}{N}\sum_{i=1}^{N}\tau_i,$$

which means that

$$\widehat{\bar{\alpha}}_{GMG} = \bar{\alpha} + \bar{\mu} + \bar{\tau}, \tag{5.8}$$

where $\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mu_i$ and $\bar{\tau} = \frac{1}{N}\sum_{i=1}^{N}\tau_i$. From (5.8) and using the general RCR

assumptions:

$$\begin{aligned} var\left(\widehat{\bar{\alpha}}_{GMG}\right) &= var\left(\bar{\mu}\right) + var\left(\bar{\tau}\right) \\ &= \frac{1}{N}\Psi + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{\varepsilon_{ii}}\left(X_i'\omega_{ii}^{-1}X_i\right)^{-1} \\ &+ \frac{1}{N^2}\sum_{\substack{i\neq j \\ i,j=1}}^{N}\sigma_{\varepsilon_{ij}}\left(X_i'\omega_{ii}^{-1}X_i\right)^{-1}X_i'\omega_{ii}^{-1}\omega_{ij}\omega_{jj}^{-1}X_j\left(X_j'\omega_{jj}^{-1}X_j\right)^{-1}. \end{aligned}$$

Using the consistent estimators of Ψ, $\sigma_{\varepsilon_{ij}}$, and ω_{ij} defined above, then we get the formula of $\widehat{var}\left(\widehat{\bar{\alpha}}_{GMG}\right)$ as in equation (5.4). □

C. Derive the variance-covariance matrix of MG:

Proof: As above, equation (2.3) can be rewritten as follows:

$$\hat{\alpha}_i = \bar{\alpha} + \mu_i + \lambda_i, \tag{5.9}$$

where $\lambda_i = \hat{\alpha}_i - \alpha_i = (X_i'X_i)^{-1}X_i'u_i$. From (5.9):

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i = \bar{\alpha} + \frac{1}{N}\sum_{i=1}^{N}\mu_i + \frac{1}{N}\sum_{i=1}^{N}\lambda_i,$$

which means that

$$\widehat{\bar{\alpha}}_{MG} = \bar{\alpha} + \bar{\mu} + \bar{\lambda}, \tag{5.10}$$

where $\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mu_i$, and $\bar{\lambda} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i$. From (5.10) and using the general RCR assumptions:

$$\begin{aligned} var\left(\widehat{\bar{\alpha}}_{MG}\right) &= var\left(\bar{\mu}\right) + var\left(\bar{\lambda}\right) \\ &= \frac{1}{N}\Psi + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_{\varepsilon_{ii}}\left(X_i'X_i\right)^{-1}X_i'\omega_{ii}X_i\left(X_i'X_i\right)^{-1} \\ &+ \frac{1}{N^2}\sum_{\substack{i\neq j \\ i,j=1}}^{N}\sigma_{\varepsilon_{ij}}\left(X_i'X_i\right)^{-1}X_i'\omega_{ij}X_j\left(X_j'X_j\right)^{-1}. \end{aligned}$$

As in the GMG estimator, and by using the consistent estimators of Ψ, $\sigma_{\varepsilon_{ij}}$, and ω_{ij}, then we get the formula of $\widehat{var}\left(\widehat{\bar{\alpha}}_{GM}\right)$ as in equation (5.3). □

It is noted from lemma 1 that the variance of the GMG estimator is less than the variance of the MG estimator when the general RCR assumptions are

satisfied. In other words, the GMG estimator is more efficient than the MG estimator. But under RCR assumptions, we have:

$$var\left(\widehat{\bar{\alpha}}_{MG}\right) = var\left(\widehat{\bar{\alpha}}_{GMG}\right) = \frac{1}{N\left(N-1\right)}\left(\sum_{i=1}^{N}\alpha_i\alpha_i' - \frac{1}{N}\sum_{i=1}^{N}\alpha_i\sum_{i=1}^{N}\alpha_i'\right) = \frac{1}{N}\Psi^{+}.$$

The next lemma explains the asymptotic variances (as $T \to \infty$ with N fixed) properties of GRCR, RCR, GMG, and MG estimators. In order to justify the derivation of the asymptotic variances, we must assume the following:

Assumption 9: $\underset{T\to\infty}{\text{plim}}\, T^{-1}X_i'X_i$ and $\underset{T\to\infty}{\text{plim}}\, T^{-1}X_i'\hat{\omega}_{ii}^{-1}X_i$ are finite and positive definite for all i and for $|\rho_i| < 1$.

Lemma 5.2. *If the general RCR assumptions and assumption 9 are satisfied, then the estimated asymptotic variance-covariance matrices of GRCR, RCR, GMG, and MG estimators are equal:*

$$\begin{aligned}\underset{T\to\infty}{\text{plim}}\,\widehat{var}\left(\widehat{\bar{\alpha}}_{GRCR}\right) &= \underset{T\to\infty}{\text{plim}}\,\widehat{var}\left(\widehat{\bar{\alpha}}_{RCR}\right) = \underset{T\to\infty}{\text{plim}}\,\widehat{var}\left(\widehat{\bar{\alpha}}_{GMG}\right)\\ &= \underset{T\to\infty}{\text{plim}}\,\widehat{var}\left(\widehat{\bar{\alpha}}_{MG}\right) = N^{-1}\Psi^{+}.\end{aligned}$$

Proof of Lemma 5.2:

Following the same argument as in Parks (1967) and utilizing assumption 9, we can show that:

$$\begin{aligned}&\underset{T\to\infty}{\text{plim}}\,\hat{\alpha}_i = \underset{T\to\infty}{\text{plim}}\,\hat{\alpha}_i^{*} = \alpha_i, \quad \underset{T\to\infty}{\text{plim}}\,\hat{\rho}_{ij} = \rho_{ij},\\ &\underset{T\to\infty}{\text{plim}}\,\hat{\sigma}_{\varepsilon_{ij}} = \sigma_{\varepsilon_{ij}}, \text{and } \underset{T\to\infty}{\text{plim}}\,\hat{\omega}_{ij} = \omega_{ij},\end{aligned} \tag{5.11}$$

and then

$$\begin{aligned}\underset{T\to\infty}{\text{plim}}\,\frac{1}{T}\hat{\sigma}_{\varepsilon_{ii}}T\left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1} &= \underset{T\to\infty}{\text{plim}}\,\frac{1}{T}\hat{\sigma}_{\varepsilon_{ii}}T\left(X_i'X_i\right)^{-1}X_i'\hat{\omega}_{ii}X_i\left(X_i'X_i\right)^{-1}\\ &= \underset{T\to\infty}{\text{plim}}\,\frac{1}{T}\hat{\sigma}_{\varepsilon_{ij}}T\left(X_i'X_i\right)^{-1}X_i'\hat{\omega}_{ij}X_j\left(X_j'X_j\right)^{-1}\\ &= \underset{T\to\infty}{\text{plim}}\,\frac{1}{T}\hat{\sigma}_{\varepsilon_{ij}}T\left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1}X_i'\hat{\omega}_{ii}^{-1}\hat{\omega}_{ij}\hat{\omega}_{jj}^{-1}X_j\\ &\quad\left(X_j'\hat{\omega}_{jj}^{-1}X_j\right)^{-1} = 0.\end{aligned} \tag{5.12}$$

Substituting (5.11) and (5.12) in (3.6):

$$\underset{T\to\infty}{\text{plim}}\,\hat{\Psi}^{*} = \frac{1}{N-1}\left(\sum_{i=1}^{N}\alpha_i\alpha_i' - \frac{1}{N}\sum_{i=1}^{N}\alpha_i\sum_{i=1}^{N}\alpha_i'\right) = \Psi^{+}. \tag{5.13}$$

By substituting (5.11)–(5.13) into (5.3), (5.4), and (3.8):

(5.14)
$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{MG}\right) = & \frac{1}{N}\operatorname*{plim}_{T\to\infty}\hat{\Psi}^{*} \\ & + \frac{1}{N^2}\sum_{i=1}^{N}\operatorname*{plim}_{T\to\infty}\frac{1}{T}\hat{\sigma}_{\varepsilon_{ii}}T\left(X_i'X_i\right)^{-1}X_i'\hat{\omega}_{ii}X_i\left(X_i'X_i\right)^{-1} \\ & + \frac{1}{N^2}\sum_{\substack{i\neq j\\ i,j=1}}^{N}\operatorname*{plim}_{T\to\infty}\frac{1}{T}\hat{\sigma}_{\varepsilon_{ij}}T\left(X_i'X_i\right)^{-1}X_i'\hat{\omega}_{ij}X_j\left(X_j'X_j\right)^{-1} \\ & = \frac{1}{N}\Psi^{+},\end{aligned}$$

(5.15)
$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{GMG}\right) = & \tfrac{1}{N(N-1)}\operatorname*{plim}_{T\to\infty}\left(\sum_{i=1}^{N}\hat{\alpha}_i^{*}\hat{\alpha}_i^{*\prime} - \tfrac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i^{*}\sum_{i=1}^{N}\hat{\alpha}_i^{*\prime}\right) \\ & + \tfrac{1}{N(N-1)}\sum_{\substack{i\neq j\\ i,j=1}}^{N}\begin{bmatrix}\operatorname*{plim}_{T\to\infty}\frac{1}{T}\hat{\sigma}_{\varepsilon_{ij}}T\left(X_i'\hat{\omega}_{ii}^{-1}X_i\right)^{-1} \\ X_i'\hat{\omega}_{ii}^{-1}\hat{\omega}_{ij}\hat{\omega}_{jj}^{-1}X_j\left(X_j'\hat{\omega}_{jj}^{-1}X_j\right)^{-1}\end{bmatrix} = \tfrac{1}{N}\Psi^{+},\end{aligned}$$

(5.16)
$$\operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{GRCR}\right) = \operatorname*{plim}_{T\to\infty}\left(X'\hat{\Omega}^{*-1}X\right)^{-1} = \left[\sum_{i=1}^{N}\Psi^{+-1}\right]^{-1} = \frac{1}{N}\Psi^{+}.$$

Similarly, we will use the results in (5.11)–(5.13) in case of RCR estimator:

(5.17)
$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{RCR}\right) &= \operatorname*{plim}_{T\to\infty}\left[\left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}\hat{\Omega}^{*}\,\hat{\Omega}^{-1}X\left(X'\hat{\Omega}^{-1}X\right)^{-1}\right] \\ &= \frac{1}{N}\Psi^{+}.\end{aligned}$$

From (5.14)–(5.17), we can conclude that:

$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{GRCR}\right) &= \operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{RCR}\right) \\ &= \operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{GMG}\right) = \operatorname*{plim}_{T\to\infty}\widehat{var}\left(\widehat{\bar{\alpha}}_{MG}\right) = \frac{1}{N}\Psi^{+}. \qquad \square\end{aligned}$$

From Lemma 5.2, we can conclude that the means and the variance-covariance matrices of the limiting distributions of $\widehat{\bar{\alpha}}_{GRCR}$, $\widehat{\bar{\alpha}}_{RCR}$, $\widehat{\bar{\alpha}}_{GMG}$, and $\widehat{\bar{\alpha}}_{MG}$ are the same and are equal to $\bar{\alpha}$ and $N^{-1}\Psi$ respectively even if the errors are correlated as in assumption 8. it is not expected to increase the asymptotic efficiency of $\widehat{\bar{\alpha}}_{GRCR}$, $\widehat{\bar{\alpha}}_{RCR}$, $\widehat{\bar{\alpha}}_{GMG}$, and $\widehat{\bar{\alpha}}_{MG}$. This does not mean that the GRCR estimator cannot be more efficient than RCR, GMG, and MG in small samples when the errors are correlated as in assumption 8. This will be examined in our simulation study.

6. MONTE CARLO SIMULATION

In this section, the Monte Carlo simulation has been used for making comparisons between the behavior of the classical pooling estimators (CP-OLS and CP-SUR), random-coefficients estimators (RCR and GRCR), and mean group estimators (MG and GMG) in small and moderate samples. The program to set up the Monte Carlo simulation, written in the R language, is available upon request. Monte Carlo experiments were carried out based on the following data generating process:

$$y_{it} = \sum_{k=1}^{3} \alpha_{ki} x_{kit} + u_{it}, \qquad i = 1, 2, ..., N;\ t = 1, 2, ..., T. \tag{6.1}$$

To perform the simulation under the general RCR assumptions, the model in (6.1) was generated as follows:

1. The independent variables, $(x_{kit}; k = 1,\ 2,\ 3)$, were generated as independent standard normally distributed random variables. The values of x_{kit} were allowed to differ for each cross-sectional unit. However, once generated for all N cross-sectional units the values were held fixed over all Monte Carlo trials.

2. The errors, u_{it}, were generated as in assumption 8: $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$, where the values of $\varepsilon_i = (\varepsilon_{i1},\ ..., \varepsilon_{iT})' \ \forall\ i = 1, 2, ..., N$ were generated as multivariate normally distributed with means zeros and variance-covariance matrix:

$$\begin{pmatrix} \sigma_{\varepsilon_{ii}} & \sigma_{\varepsilon_{ij}} & \cdots & \sigma_{\varepsilon_{ij}} \\ \sigma_{\varepsilon_{ij}} & \sigma_{\varepsilon_{ii}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{\varepsilon_{ij}} \\ \sigma_{\varepsilon_{ij}} & \cdots & \sigma_{\varepsilon_{ij}} & \sigma_{\varepsilon_{ii}} \end{pmatrix},$$

where the values of $\sigma_{\varepsilon_{ii}}$, $\sigma_{\varepsilon_{ij}}$, and ρ were chosen to be: $\sigma_{\varepsilon_{ii}} = 1$ or 100; $\sigma_{\varepsilon_{ij}} = 0, 0.75$, or 0.95, and $\rho = 0, 0.55$, or 0.85, where the values of $\sigma_{\varepsilon_{ii}}$, $\sigma_{\varepsilon_{ij}}$, and ρ are constants for all $i,\ j = 1, 2, ..., N$ in each Monte Carlo trial. The initial values of u_{it} are generated as $u_{i1} = \varepsilon_{i1}/\sqrt{1-\rho^2}\ \forall\ i = 1, 2, ..., N$. The values of errors were allowed to differ for each cross-sectional unit on a given Monte Carlo trial and were allowed to differ between trials. The errors are independent with all independent variables.

3. The coefficients, α_{ki}, were generated as in assumption 6: $\alpha_i = \bar{\alpha} + \mu_i$, where $\bar{\alpha} = (1, 1, 1)'$, and μ_i were generated from two distributions. First, multivariate normal distribution with means zeros and variance-covariance matrix $\Psi = diag\left\{\psi_k^2\right\}; k = 1, 2, 3$. The values of Ψ_k^2 were chosen to be fixed for all k and equal to 5 or 25. Second, multivariate student's t distribution with degree of freedom (df): $df = 1$ or 5.

To include the case of fixed-coefficients models in our simulation study, we assume that $\mu_i = 0$.

4. The values of N and T were chosen to be 5, 8, 10, 12, 15, and 20 to represent small and moderate samples for the number of individuals and the time dimension. To compare the small and moderate samples performance for the different estimators, three different samplings schemes have been designed in our simulation, where each design contains four pairs of N and T. The first two represent small samples while the moderate samples are represented by the second two pairs. These designs have been created as follows: First, case of $N < T$, the pairs of N and T were chosen to be $(N, T) = (5, 8), (5, 12), (10, 15)$, or (10, 20). Second, case of $N = T$, the pairs are $(N, T) = (5, 5), (10, 10), (15, 15)$, or (20, 20). Third, case of $N > T$, the pairs are $(N, T) = (8, 5), (12, 5), (15, 10)$, or (20, 10).

5. All Monte Carlo experiments involved 1000 replications and all the results of all separate experiments are obtained by precisely the same series of random numbers. To raise the efficiency of the comparison between these estimators, we calculate the average of total standard errors (ATSE) for each estimator by:

$$\text{ATSE} = \frac{1}{1000} \sum_{l=1}^{1000} \left\{ trace \left[\widehat{var} \left(\widehat{\bar{\alpha}}_l \right) \right]^{0.5} \right\},$$

where $\widehat{\bar{\alpha}}_l$ is the estimated vector of $\bar{\alpha}$ in (6.1), and $\widehat{var}\left(\widehat{\bar{\alpha}}_l\right)$ is the estimated variance-covariance matrix of the estimator.

The Monte Carlo results are given in Tables 1–6. Specifically, Tables 1–3 present the ATSE values of the estimators when $\sigma_{\varepsilon_{ii}} = 1$, and in cases of $N < T, N = T$, and$N > T$, respectively. While case of $\sigma_{\varepsilon_{ii}} = 100$ is presented in Tables 4–6 in the same cases of N and T. In our simulation study, the main factors that have an effect on the ATSE values of the estimators are N, T, $\sigma_{\varepsilon_{ii}}$, $\sigma_{\varepsilon_{ij}}$, ρ, ψ_k^2 (for normal distribution), and *df* (for student's t distribution). From Tables 1–6, we can summarize some effects for all estimators in the following points:

- When the values of N and T are increased, the values of ATSE are decreasing for all simulation situations.
- When the value of $\sigma_{\varepsilon_{ii}}$ is increased, the values of ATSE are increasing in most situations.
- When the values of $(\rho, \sigma_{\varepsilon_{ij}})$ are increased, the values of ATSE are increasing in most situations.
- When the value of ψ_k^2 is increased, the values of ATSE are increasing for all situations.
- When the value of *df* is increased, the values of ATSE are decreasing for all situations.

Table 1: ATSE for various estimators when $\sigma_{\varepsilon_{ii}} = 1$ and $N < T$.

$(\rho,\ \sigma_{\varepsilon_{ij}})$	$(0, 0)$				$(0.55, 0.75)$				$(0.85, 0.95)$			
(N, T)	$(5, 8)$	$(5, 12)$	$(10, 15)$	$(10, 20)$	$(5, 8)$	$(5, 12)$	$(10, 15)$	$(10, 20)$	$(5, 8)$	$(5, 12)$	$(10, 15)$	$(10, 20)$
$\mu_i = 0$												
CP-OLS	0.920	0.746	0.440	0.436	0.857	0.888	0.409	0.450	1.107	1.496	0.607	0.641
CP-SUR	0.958	0.767	0.419	0.417	0.829	0.880	0.381	0.384	0.947	1.469	0.453	0.532
MG	0.947	0.765	0.470	0.469	0.886	0.910	0.442	0.468	1.133	1.475	0.608	0.636
GMG	0.702	0.556	0.369	0.375	0.638	0.662	0.289	0.305	0.644	1.098	0.302	0.291
RCR	1.012	30.746	0.517	0.497	1.064	1.130	2.241	0.726	1.365	5.960	0.856	1.326
GRCR	0.754	0.624	0.352	0.357	0.634	0.703	0.302	0.295	0.735	1.141	0.324	0.388
$\mu_i \sim N(0, 5)$												
CP-OLS	4.933	4.682	2.320	2.742	2.588	2.902	2.598	2.130	3.627	5.079	2.165	2.935
CP-SUR	5.870	5.738	2.852	3.411	3.143	3.456	3.212	2.592	4.011	5.906	2.668	3.549
MG	4.057	4.112	2.086	2.494	2.173	2.478	2.352	1.888	3.094	4.040	1.938	2.626
GMG	4.057	4.110	2.084	2.494	2.176	2.479	2.348	1.879	3.052	4.024	1.908	2.606
RCR	4.053	4.114	2.083	2.493	2.632	3.304	2.352	1.888	3.287	6.422	2.052	2.648
GRCR	4.030	4.092	2.067	2.480	2.104	2.413	2.331	1.855	2.969	3.905	1.865	2.578
$\mu_i \sim N(0, 25)$												
CP-OLS	7.528	7.680	7.147	6.341	8.293	8.156	6.321	6.739	7.942	7.214	4.691	6.423
CP-SUR	8.866	9.439	8.935	8.046	10.104	9.880	8.028	8.402	9.074	8.482	5.739	7.937
MG	6.272	6.549	6.324	5.597	6.879	6.650	5.541	5.917	6.442	6.083	4.118	5.672
GMG	6.271	6.548	6.324	5.597	6.881	6.650	5.538	5.913	6.422	6.078	4.103	5.662
RCR	6.271	6.548	6.324	5.597	6.885	6.657	5.541	5.917	7.546	6.098	4.122	5.686
GRCR	6.251	6.539	6.319	5.590	6.857	6.626	5.530	5.906	6.389	6.010	4.082	5.649
$\mu_i \sim t(5)$												
CP-OLS	2.253	1.983	1.562	1.544	1.479	1.977	1.060	1.223	2.115	3.301	1.470	1.439
CP-SUR	2.626	2.419	1.925	1.912	1.694	2.266	1.275	1.454	2.403	3.903	1.717	1.643
MG	1.859	1.776	1.410	1.401	1.324	1.722	0.984	1.078	1.923	2.707	1.335	1.260
GMG	1.856	1.771	1.408	1.400	1.316	1.718	0.970	1.064	1.826	2.666	1.284	1.215
RCR	2.002	1.768	1.452	1.396	2.020	3.260	1.017	1.087	12.328	6.655	2.035	2.650
GRCR	1.788	1.727	1.377	1.375	1.215	1.655	0.926	1.019	1.786	2.552	1.221	1.155
$\mu_i \sim t(1)$												
CP-OLS	16.112	4.096	2.732	10.189	12.490	24.982	6.424	2.837	6.685	5.668	12.763	1.786
CP-SUR	19.483	5.046	3.365	12.976	14.940	29.854	8.009	3.555	7.807	7.043	15.947	2.126
MG	11.751	3.427	2.432	9.094	9.811	19.875	5.742	2.306	5.568	4.365	11.473	1.620
GMG	11.751	3.423	2.431	9.094	9.811	19.875	5.740	2.298	5.540	4.352	11.468	1.583
RCR	11.751	3.423	2.431	9.094	9.813	19.877	5.742	2.304	5.591	7.730	11.475	1.829
GRCR	11.739	3.403	2.417	9.090	9.795	19.868	5.733	2.271	5.498	4.228	11.462	1.530

Table 2: ATSE for various estimators when $\sigma_{\varepsilon_{ii}} = 1$ and $N = T$.

$(\rho,\ \sigma_{\varepsilon_{ij}})$	$(0, 0)$				$(0.55, 0.75)$				$(0.85, 0.95)$			
(N, T)	$(5,5)$	$(10,10)$	$(15,15)$	$(20,20)$	$(5,5)$	$(10,10)$	$(15,15)$	$(20,20)$	$(5,5)$	$(10,10)$	$(15,15)$	$(20,20)$
$\mu_i = 0$												
CP-OLS	1.671	0.461	0.259	0.174	2.081	0.424	0.274	0.207	3.351	0.678	0.394	0.276
CP-SUR	2.387	0.550	0.299	0.178	3.340	0.478	0.291	0.182	4.301	0.716	0.293	0.192
MG	1.686	0.486	0.280	0.183	2.058	0.474	0.300	0.210	3.093	0.668	0.377	0.255
GMG	1.174	0.395	0.234	0.159	1.669	0.363	0.209	0.149	2.028	0.370	0.190	0.115
RCR	1.905	0.557	0.314	0.179	1.997	0.953	0.411	0.502	3.249	1.982	0.471	0.458
GRCR	1.294	0.320	0.173	0.102	1.678	0.264	0.151	0.093	2.480	0.380	0.145	0.094
$\mu_i \sim N(0, 5)$												
CP-OLS	4.119	3.404	1.982	1.651	4.593	2.002	1.517	1.474	5.023	2.926	1.847	1.740
CP-SUR	6.478	5.521	3.511	3.097	8.141	3.313	2.735	2.737	7.176	4.951	3.313	3.368
MG	3.480	2.750	1.744	1.520	4.015	1.671	1.295	1.341	4.284	2.531	1.633	1.608
GMG	3.481	2.750	1.743	1.520	4.008	1.664	1.289	1.337	4.034	2.515	1.615	1.599
RCR	5.955	2.749	1.743	1.520	4.232	1.666	1.295	1.342	12.312	2.574	1.651	1.617
GRCR	3.400	2.727	1.730	1.513	3.826	1.622	1.266	1.328	3.913	2.463	1.591	1.590
$\mu_i \sim N(0, 25)$												
CP-OLS	8.056	6.265	4.022	3.637	7.976	5.496	4.240	3.968	10.264	6.615	4.558	3.733
CP-SUR	12.776	10.403	7.168	6.869	14.233	9.622	7.606	7.540	15.004	11.368	8.361	7.229
MG	6.474	5.145	3.558	3.348	6.491	4.599	3.692	3.623	6.798	5.597	4.042	3.464
GMG	6.476	5.145	3.558	3.348	6.498	4.596	3.690	3.622	6.822	5.589	4.036	3.460
RCR	6.469	5.145	3.558	3.348	6.457	4.597	3.692	3.624	10.576	5.614	4.050	3.468
GRCR	6.412	5.134	3.552	3.345	6.399	4.581	3.683	3.618	6.534	5.566	4.027	3.456
$\mu_i \sim t(5)$												
CP-OLS	2.017	1.444	1.054	0.818	2.719	2.306	1.452	1.202	3.512	1.374	1.130	0.866
CP-SUR	2.952	2.278	1.848	1.499	4.581	4.002	2.602	2.251	4.784	2.113	1.960	1.584
MG	1.900	1.215	0.933	0.759	2.435	1.892	1.228	1.113	3.241	1.209	1.017	0.800
GMG	1.752	1.214	0.933	0.759	2.369	1.886	1.221	1.108	2.635	1.177	0.989	0.780
RCR	2.987	1.209	0.931	0.758	2.862	1.886	1.229	1.114	11.891	1.760	1.527	0.815
GRCR	1.628	1.165	0.908	0.744	2.193	1.848	1.199	1.097	2.727	1.073	0.951	0.762
$\mu_i \sim t(1)$												
CP-OLS	2.946	4.082	36.296	32.249	170.833	4.983	7.221	5.545	5.447	14.094	27.076	2.245
CP-SUR	4.663	6.691	70.583	64.229	291.169	8.653	13.554	10.472	7.942	25.514	54.690	4.290
MG	2.569	3.337	23.288	26.932	92.236	4.064	5.831	5.069	4.403	11.428	20.763	2.085
GMG	2.565	3.337	23.288	26.932	92.238	4.060	5.829	5.068	4.362	11.420	20.759	2.078
RCR	5.160	3.337	23.288	26.932	92.238	4.061	5.831	5.069	7.663	11.440	20.767	2.091
GRCR	2.433	3.320	23.280	26.931	92.226	4.042	5.823	5.065	4.024	11.401	20.753	2.072

Table 3: ATSE for various estimators when $\sigma_{\varepsilon_{ii}} = 1$ and $N > T$.

$(\rho,\ \sigma_{\varepsilon_{ij}})$	$(0, 0)$				$(0.55, 0.75)$				$(0.85, 0.95)$			
(N, T)	$(8, 5)$	$(12, 5)$	$(15, 10)$	$(20, 10)$	$(8, 5)$	$(12, 5)$	$(15, 10)$	$(20, 10)$	$(8, 5)$	$(12, 5)$	$(15, 10)$	$(20, 10)$
$\mu_i = 0$												
CP-OLS	1.763	3.198	0.510	0.438	1.254	1.399	0.436	0.536	1.218	1.350	0.688	0.591
CP-SUR	2.504	4.585	0.635	0.518	1.748	1.963	0.497	0.607	1.637	1.808	0.780	0.655
MG	1.856	2.927	0.576	0.475	1.434	1.455	0.501	0.618	1.528	1.523	0.830	0.631
GMG	1.288	1.767	0.452	0.391	1.017	0.995	0.350	0.417	1.014	0.982	0.468	0.433
RCR	7.356	2.702	0.567	0.573	1.353	1.333	0.693	1.625	1.490	1.468	2.432	1.605
GRCR	1.289	2.277	0.342	0.267	0.937	1.010	0.248	0.306	0.865	0.856	0.413	0.312
$\mu_i \sim N(0, 5)$												
CP-OLS	3.136	4.014	2.525	2.017	3.677	3.352	2.477	3.105	2.146	3.501	1.927	2.415
CP-SUR	4.590	5.845	3.576	2.888	5.279	4.824	3.485	4.396	3.080	4.935	2.687	3.393
MG	2.753	3.418	2.153	1.685	2.972	2.643	2.113	2.628	2.191	2.813	1.724	2.156
GMG	2.665	3.425	2.152	1.684	2.951	2.660	2.106	2.617	2.097	2.748	1.679	2.142
RCR	3.611	3.306	2.146	1.681	2.897	3.034	2.109	2.621	61.169	137.429	2.187	2.147
GRCR	2.400	2.982	2.103	1.636	2.774	2.399	2.066	2.572	1.852	2.550	1.532	2.075
$\mu_i \sim N(0, 25)$												
CP-OLS	6.919	6.434	6.179	5.259	6.442	5.639	4.972	4.460	6.279	7.428	5.480	5.366
CP-SUR	10.250	9.292	8.750	7.682	9.200	8.224	7.123	6.378	9.507	10.544	7.791	7.698
MG	5.090	5.029	5.092	4.381	4.987	4.505	4.167	3.688	5.353	5.689	4.545	4.756
GMG	5.046	5.031	5.092	4.380	4.971	4.512	4.163	3.680	5.316	5.677	4.530	4.749
RCR	4.986	4.735	5.091	4.380	4.939	4.466	4.165	3.683	5.303	6.219	4.538	4.753
GRCR	4.898	4.588	5.071	4.362	4.874	4.408	4.142	3.645	5.189	5.559	4.479	4.720
$\mu_i \sim t(5)$												
CP-OLS	1.779	2.367	1.151	1.080	1.780	2.464	1.986	1.308	2.157	2.848	1.473	1.283
CP-SUR	2.541	3.365	1.604	1.493	2.596	3.711	2.929	1.745	3.137	4.179	1.987	1.730
MG	1.839	1.989	1.010	0.943	1.647	2.276	1.603	1.074	2.109	2.401	1.260	1.467
GMG	1.577	1.974	1.008	0.942	1.563	2.245	1.586	1.076	1.730	2.362	1.235	1.255
RCR	2.573	2.327	0.991	0.960	2.785	2.945	1.591	1.097	3.523	3.020	3.322	3.509
GRCR	1.336	1.738	0.924	0.837	1.529	1.893	1.525	0.982	1.652	2.120	1.124	1.049
$\mu_i \sim t(1)$												
CP-OLS	23.572	9.953	1.708	9.638	9.612	3.030	5.400	4.609	6.932	8.340	25.666	4.259
CP-SUR	35.133	13.767	2.466	14.035	15.207	4.429	8.027	6.816	9.309	12.412	39.880	6.199
MG	17.304	6.568	1.410	6.014	7.568	2.654	4.164	3.451	4.802	6.004	16.848	3.318
GMG	17.295	6.563	1.409	6.014	7.580	2.629	4.155	3.452	4.781	5.991	16.840	3.267
RCR	17.295	6.535	1.398	6.012	7.546	2.499	4.158	3.456	6.130	5.997	16.849	4.158
GRCR	17.263	6.483	1.345	5.979	7.492	2.345	4.128	3.407	4.593	5.877	16.779	3.081

Table 4: ATSE for various estimators when $\sigma_{\varepsilon_{ii}} = 100$ and $N < T$.

$(\rho,\ \sigma_{\varepsilon_{ij}})$	$(0, 0)$				$(0.55, 0.75)$				$(0.85, 0.95)$			
(N, T)	$(5, 8)$	$(5, 12)$	$(10, 15)$	$(10, 20)$	$(5, 8)$	$(5, 12)$	$(10, 15)$	$(10, 20)$	$(5, 8)$	$(5, 12)$	$(10, 15)$	$(10, 20)$
$\mu_i = 0$												
CP-OLS	2.908	2.357	1.389	1.379	2.756	2.863	1.414	1.395	3.798	5.179	2.042	2.208
CP-SUR	3.028	2.422	1.323	1.316	2.806	2.997	1.335	1.302	3.520	5.316	1.692	1.989
MG	2.993	2.419	1.486	1.483	2.830	2.984	1.492	1.503	3.850	4.907	2.010	2.292
GMG	2.221	1.759	1.168	1.187	1.975	2.180	1.027	1.004	2.132	3.466	1.022	1.191
RCR	3.199	97.225	1.634	1.570	3.205	6.691	2.576	2.846	4.711	7.169	2.708	3.170
GRCR	2.381	1.970	1.111	1.128	2.188	2.399	1.061	1.029	2.667	3.872	1.220	1.429
$\mu_i \sim N(0, 5)$												
CP-OLS	5.096	4.872	2.481	2.890	3.298	3.570	2.732	2.260	4.432	6.390	2.479	3.180
CP-SUR	5.787	5.751	2.856	3.437	3.573	3.960	3.305	2.557	4.449	6.946	2.463	3.524
MG	4.533	4.450	2.361	2.737	3.193	3.448	2.575	2.172	4.327	5.642	2.363	3.076
GMG	4.507	4.427	2.349	2.734	2.869	3.165	2.539	2.101	3.695	5.110	2.150	2.849
RCR	11.579	5.572	2.500	2.702	3.871	8.045	3.278	3.489	7.748	9.539	5.301	22.220
GRCR	4.179	4.294	2.166	2.576	2.755	3.026	2.378	1.911	3.456	5.004	1.879	2.560
$\mu_i \sim N(0, 25)$												
CP-OLS	7.670	7.803	7.209	6.407	8.362	8.314	6.380	6.781	7.971	7.887	4.852	6.554
CP-SUR	8.833	9.460	8.952	8.050	10.073	10.032	8.245	8.508	9.153	9.160	5.890	8.277
MG	6.570	6.760	6.431	5.714	7.118	7.016	5.653	6.018	6.812	7.017	4.338	5.913
GMG	6.556	6.749	6.426	5.713	7.116	7.013	5.625	5.991	6.658	6.996	4.240	5.795
RCR	10.949	6.908	6.423	5.706	7.103	7.629	5.647	6.008	11.120	16.814	9.260	6.478
GRCR	6.400	6.633	6.370	5.646	6.945	6.826	5.558	5.932	6.286	6.595	4.057	5.661
$\mu_i \sim t(5)$												
CP-OLS	3.227	2.672	1.820	1.804	2.894	3.067	1.534	1.558	4.052	5.630	2.112	2.299
CP-SUR	3.432	2.879	1.975	1.959	3.045	3.327	1.529	1.560	3.998	6.065	1.838	2.099
MG	3.186	2.654	1.829	1.810	2.924	3.097	1.588	1.617	4.042	5.146	2.071	2.318
GMG	2.816	2.405	1.799	1.782	2.296	2.690	1.394	1.435	2.792	4.288	1.603	1.692
RCR	3.665	3.442	2.592	2.462	4.922	4.147	3.057	4.985	9.667	14.064	3.871	6.113
GRCR	2.666	2.317	1.625	1.543	2.374	2.662	1.232	1.233	3.045	4.365	1.456	1.604
$\mu_i \sim t(1)$												
CP-OLS	16.193	4.345	2.882	10.228	12.527	25.028	6.481	2.957	6.842	6.962	12.819	2.363
CP-SUR	19.488	5.071	3.383	12.975	14.929	30.583	8.213	3.571	7.803	7.838	16.626	2.317
MG	11.990	3.871	2.673	9.164	9.996	19.985	5.841	2.595	6.095	5.929	11.548	2.434
GMG	11.990	3.832	2.665	9.163	9.979	19.993	5.819	2.524	5.898	5.591	11.512	1.988
RCR	11.965	4.529	2.625	9.162	9.966	19.996	5.839	3.527	13.705	59.015	11.574	14.464
GRCR	11.840	3.650	2.507	9.122	9.862	19.940	5.762	2.360	5.434	5.506	11.460	1.773

Table 5: ATSE for various estimators when $\sigma_{\varepsilon_{ii}} = 100$ and $N = T$.

$(\rho,\ \sigma_{\varepsilon_{ij}})$	$(0, 0)$				$(0.55, 0.75)$				$(0.85, 0.95)$			
(N, T)	$(5,5)$	$(10,10)$	$(15,15)$	$(20,20)$	$(5,5)$	$(10,10)$	$(15,15)$	$(20,20)$	$(5,5)$	$(10,10)$	$(15,15)$	$(20,20)$
$\mu_i = 0$												
CP-OLS	5.284	1.456	0.818	0.548	6.920	1.339	0.904	0.629	11.353	2.314	1.215	0.871
CP-SUR	7.548	1.737	0.942	0.559	10.528	1.580	0.977	0.589	15.654	2.573	0.987	0.625
MG	5.331	1.537	0.886	0.577	6.606	1.417	0.998	0.658	10.554	2.362	1.238	0.839
GMG	3.712	1.250	0.741	0.503	5.470	1.105	0.693	0.466	6.959	1.419	0.602	0.410
RCR	6.023	1.759	0.990	0.564	8.315	2.026	2.034	1.388	10.978	3.817	2.088	1.241
GRCR	4.090	1.007	0.545	0.318	5.497	0.907	0.527	0.318	8.037	1.363	0.525	0.325
$\mu_i \sim N(0, 5)$												
CP-OLS	5.580	3.519	2.061	1.705	7.429	2.182	1.629	1.543	10.993	3.155	1.991	1.859
CP-SUR	8.237	5.479	3.497	3.091	11.726	3.255	2.651	2.742	15.414	4.585	3.080	3.221
MG	5.622	2.996	1.876	1.592	6.993	1.987	1.522	1.438	10.338	3.017	1.864	1.733
GMG	4.959	2.994	1.876	1.591	6.571	1.968	1.459	1.406	7.682	2.893	1.712	1.649
RCR	8.572	3.064	1.861	1.588	8.773	2.645	2.696	1.435	10.818	6.531	3.172	1.779
GRCR	4.679	2.764	1.747	1.520	6.313	1.727	1.249	1.322	8.234	2.397	1.489	1.558
$\mu_i \sim N(0, 25)$												
CP-OLS	8.220	6.333	4.056	3.661	9.384	5.567	4.285	3.991	12.808	6.724	4.618	3.788
CP-SUR	12.685	10.388	7.152	6.865	15.219	9.557	7.574	7.573	18.954	11.401	8.194	7.215
MG	7.404	5.282	3.620	3.380	8.388	4.740	3.779	3.657	11.236	5.845	4.138	3.523
GMG	7.257	5.281	3.620	3.380	8.438	4.728	3.754	3.645	9.858	5.787	4.073	3.482
RCR	12.035	5.272	3.618	3.380	9.526	4.731	3.774	3.658	12.921	6.137	4.153	3.545
GRCR	6.703	5.166	3.556	3.347	7.863	4.608	3.688	3.613	9.475	5.537	3.995	3.440
$\mu_i \sim t(5)$												
CP-OLS	5.268	1.758	1.205	0.930	6.905	2.466	1.566	1.289	11.183	2.322	1.363	1.078
CP-SUR	7.487	2.302	1.826	1.505	10.462	3.902	2.518	2.232	15.445	2.648	1.486	1.354
MG	5.301	1.734	1.173	0.901	6.588	2.197	1.457	1.231	10.371	2.363	1.359	1.024
GMG	3.914	1.688	1.171	0.900	5.741	2.170	1.392	1.193	7.036	1.810	1.138	0.874
RCR	6.313	2.356	1.226	0.885	8.980	4.088	1.806	1.224	10.384	6.372	4.418	4.574
GRCR	4.238	1.313	0.937	0.764	5.796	1.894	1.179	1.094	8.124	1.489	0.823	0.688
$\mu_i \sim t(1)$												
CP-OLS	5.492	4.176	36.310	32.254	170.969	5.046	7.246	5.564	11.208	14.166	27.093	2.332
CP-SUR	8.085	6.670	70.596	64.232	277.362	8.718	13.502	10.390	15.450	26.068	54.457	4.185
MG	5.469	3.529	23.379	26.943	92.536	4.228	5.898	5.095	10.448	11.655	20.834	2.180
GMG	4.346	3.528	23.378	26.943	92.558	4.213	5.878	5.086	7.748	11.603	20.786	2.114
RCR	7.220	3.503	23.365	26.943	92.513	4.383	5.895	5.096	13.141	12.397	20.840	2.210
GRCR	4.471	3.354	23.296	26.932	92.445	4.050	5.822	5.064	8.345	11.384	20.731	2.046

Table 6: ATSE for various estimators when $\sigma_{\varepsilon_{ii}} = 100$ and $N > T$.

$(\rho,\ \sigma_{\varepsilon_{ij}})$	$(0, 0)$				$(0.55, 0.75)$				$(0.85, 0.95)$			
(N, T)	$(8, 5)$	$(12, 5)$	$(15, 10)$	$(20, 10)$	$(8, 5)$	$(12, 5)$	$(15, 10)$	$(20, 10)$	$(8, 5)$	$(12, 5)$	$(15, 10)$	$(20, 10)$
$\mu_i = 0$												
CP-OLS	5.574	3.501	1.511	1.493	5.616	4.178	1.764	1.546	8.088	9.255	2.325	2.474
CP-SUR	7.919	4.835	1.798	1.840	7.780	5.841	2.229	1.813	11.886	12.804	2.723	2.975
MG	5.868	3.453	1.659	1.676	5.678	4.306	1.908	1.629	9.127	8.473	2.678	2.773
GMG	4.073	2.490	1.349	1.337	3.643	3.717	1.515	1.219	5.788	7.373	1.382	1.581
RCR	23.253	3.498	1.759	1.808	5.403	6.417	5.387	2.286	8.172	11.799	2.744	4.156
GRCR	4.072	2.397	0.931	0.972	3.998	3.241	1.142	0.872	5.937	6.519	1.267	1.352
$\mu_i \sim N(0, 5)$												
CP-OLS	5.574	4.258	2.867	2.692	5.221	5.014	2.744	2.396	8.256	9.261	2.333	3.037
CP-SUR	7.899	5.954	3.858	3.725	7.202	7.096	3.802	3.166	12.049	12.885	2.782	4.092
MG	5.793	3.775	2.616	2.509	5.407	4.904	2.622	2.241	9.299	8.462	2.682	3.135
GMG	4.753	3.635	2.615	2.503	4.022	4.657	2.663	2.226	6.423	7.531	2.230	2.815
RCR	7.585	5.340	2.525	2.569	25.633	6.314	8.404	2.808	10.171	10.268	15.344	8.355
GRCR	4.220	3.123	2.206	2.063	3.901	3.925	2.101	1.771	6.533	6.464	1.443	2.026
$\mu_i \sim N(0, 25)$												
CP-OLS	7.383	6.000	5.791	4.700	6.808	7.512	4.220	6.284	7.648	11.202	4.729	4.463
CP-SUR	10.777	8.636	8.118	6.667	9.409	11.012	5.987	8.667	11.213	16.010	6.596	6.367
MG	6.876	4.940	4.816	4.146	6.287	6.642	3.722	5.162	8.635	9.623	4.346	4.168
GMG	6.442	4.902	4.815	4.143	6.205	6.532	3.765	5.156	7.205	9.360	4.171	3.961
RCR	11.741	5.730	4.792	4.090	11.299	7.379	3.776	5.160	12.146	12.980	13.643	7.505
GRCR	5.510	4.310	4.615	3.915	5.288	5.902	3.379	4.983	6.356	8.403	3.669	3.352
$\mu_i \sim t(5)$												
CP-OLS	5.373	3.666	1.719	1.726	5.575	4.294	1.789	1.805	8.085	9.347	2.373	2.455
CP-SUR	7.646	5.136	2.115	2.217	7.757	5.989	2.248	2.223	11.901	13.041	2.803	2.974
MG	5.706	3.482	1.779	1.837	5.623	4.394	1.926	1.802	9.133	8.456	2.695	2.784
GMG	4.249	3.082	1.722	1.759	3.683	3.907	1.647	1.727	5.933	7.429	1.691	1.879
RCR	9.861	5.223	2.501	2.758	5.421	5.238	3.195	3.158	13.392	14.875	4.908	6.298
GRCR	3.915	2.670	1.150	1.268	4.044	3.334	1.188	1.170	6.032	6.570	1.342	1.415
$\mu_i \sim t(1)$												
CP-OLS	5.821	3.703	4.328	6.252	6.016	5.931	31.442	4.149	11.344	10.999	5.576	3.013
CP-SUR	8.533	5.188	6.188	9.132	8.500	8.555	47.659	5.806	17.261	15.893	8.562	3.969
MG	5.986	3.550	3.544	5.182	5.876	5.420	21.165	3.416	11.058	9.507	4.826	3.140
GMG	4.941	3.242	3.537	5.179	5.579	5.219	21.177	3.402	8.986	9.203	4.557	2.831
RCR	8.791	13.034	13.254	5.140	7.133	6.561	21.171	3.896	13.086	12.317	10.078	10.717
GRCR	4.403	2.740	3.115	4.987	4.936	4.559	21.041	3.093	8.697	7.876	3.877	2.021

For more deeps in simulation results, we can conclude the following results:

1. Generally, the performance of all estimators in cases of $N \leqslant T$ is better than their performance in case of $N > T$. Similarly, their performance in cases of $\sigma_{\varepsilon_{ii}} = 1$ is better than the performance in case of $\sigma_{\varepsilon_{ii}} = 100$, but not as significantly better as in N and T.

2. When $\sigma_{\varepsilon_{ij}} = \rho = \mu_i = 0$, the ATSE values of the classical pooling estimators (CP-OLS and CP-SUR) are approximately equivalent, especially when the sample size is moderate and/or $N \leqslant T$. However, the ATSE values of GMG and GRCR estimators are smaller than those of the classical pooling estimators in this situation ($\sigma_{\varepsilon_{ij}} = \rho = \mu_i = 0$) and other simulation situations (case of $\sigma_{\varepsilon_{ii}}$, $\sigma_{\varepsilon_{ij}}, \rho, \psi_k^2$ are increasing, and *df* is decreasing). In other words, GMG and GRCR are more efficient than CP-OLS and CP-SUR whether the regression coefficients are fixed or random.

3. If $T \geq 15$, the values of ATSE for the MG and GMG estimators are approximately equivalent. This result is consistent with Lemma 5.2. According to our study, this case ($T \geq 15$) is achieved when the sample size is moderate in Tables 1, 2, 4, and 5. Moreover, convergence slows down if $\sigma_{\varepsilon_{ii}}$, $\sigma_{\varepsilon_{ij}}$, and ρ are increased. But the situation for the RCR and GRCR estimators is different; the convergence between them is very slow even if $T = 20$. So the MG and GMG estimators are more efficient than RCR in all simulation situations.

4. When the coefficients are random (whether they are distributed as normal or student's t), the values of ATSE for GMG and GRCR are smaller than those of MG and RCR in all simulation situations (for any N, T, $\sigma_{\varepsilon_{ii}}$, $\sigma_{\varepsilon_{ij}}$, and ρ). However, the ATSE values of GRCR are smaller than those of GMG estimator in most situations, especially when the sample size is moderate. In other words, the GRCR estimator performs better than all other estimators as long as the sample size is moderate regardless of other simulation factors.

7. CONCLUSION

In this article, the classical pooling (CP-OLS and CP-SUR), random-coefficients (RCR and GRCR), and mean group (MG and GMG) estimators of stationary RCPD models were examined in different sample sizes for the case where the errors are cross-sectionally and serially correlated. Analytical efficiency comparisons for these estimators indicate that the mean group and random-coefficients estimators are equivalent when T is sufficiently large. Furthermore, the Monte Carlo simulation results show that the classical pooling estimators are absolutely not suitable for random-coefficients models. And, the MG and GMG estimators are more efficient than the RCR estimator for random- and fixed-coefficients models, especially when T is small ($T \leq 12$). But when $T \geq 20$, the MG, GMG, and GRCR estimators are approximately equivalent. However, the GRCR estimator performs better than the MG and GMG estimators in most situations, especially in moderate samples. Therefore, we conclude that the GRCR estimator is suitable to stationary RCPD models whether the coefficients are random or fixed.

REFERENCES

[1] Abonazel, M.R. (2009). *Some Properties of Random Coefficients Regression Estimators*, MSc thesis, Institute of Statistical Studies and Research, Cairo University, Egypt.

[2] Abonazel, M.R. (2014). *Some Estimation Methods for Dynamic Panel Data Models*, PhD thesis, Institute of Statistical Studies and Research, Cairo University, Egypt.

[3] Abonazel, M.R. (2017). Bias correction methods for dynamic panel data models with fixed effects, *International Journal of Applied Mathematical Research*, **6**(2), 58–66.

[4] Alcacer, J.; Chung, W.; Hawk, A. and Pacheco-de-Almeida, G. (2013). *Applying random coefficient models to strategy research: testing for firm heterogeneity, predicting firm-specific coefficients, and estimating Strategy Trade-Offs*, Working Paper, No. 14-022, Harvard Business School Strategy Unit.

[5] Anh, V. and Chelliah, T. (1999). Estimated generalized least squares for random coefficient regression models, *Scandinavian Journal of Statistics*, **26**(1), 31–46.

[6] Baltagi, B.H. (2011). *Econometrics*, 5th ed., Berlin, Springer-Verlag Berlin–Heidelberg.

[7] Baltagi, B.H. (2013). *Econometric Analysis of Panel Data*, 5th ed., Chichester, John Wiley and Sons.

[8] Beck, N. and Katz, J. (2007). Random coefficient models for time-series-cross-section data: Monte Carlo experiments, *Political Analysis*, **15**(2), 182–195.

[9] Beran, R. and Millar, P. (1994). Minimum distance estimation in random coefficient regression models, *The Annals of Statistics*, **22**(4), 1976–1992.

[10] Bodhlyera, O.; Zewotir, T. and Ramroop, S. (2014). Random coefficient model for changes in viscosity in dissolving pulp, *Wood Research*, **59**(4), 571–582.

[11] Boness, A. and Frankfurter, G. (1977). Evidence of non-homogeneity of capital costs within "risk-classes", *The Journal of Finance*, **32**(3), 775–787.

[12] Boot, J. and Frankfurter, G. (1972). The dynamics of corporate debt management, decision rules, and some empirical evidence, *Journal of Financial and Quantitative Analysis*, **7**(4), 1957–1965.

[13] Chelliah, N. (1998). A new covariance estimator in random coefficient regression model, *The Indian Journal of Statistics*, Series B, **60**(3), 433–436.

[14] Cheng, J.; Yue, R.X. and Liu, X. (2013). Optimal designs for random coefficient regression models with heteroscedastic errors, *Communications in Statistics — Theory and Methods* **42**(15), 2798–2809.

[15] Cooley, T. and Prescott, E. (1973). Systematic (non-random) variation models: varying parameter regression: a theory and some applications, *Annals of Economic and Social Measurement*, **2**(4), 463–473.

[16] Dielman, T. (1983). Pooled cross-sectional and time series data: a survey of current statistical methodology, *The American Statistician*, **37**(2), 111–122.

[17] Dielman, T. (1989). *Pooled Cross-Sectional and Time Series Data Analysis*, Marcel Dekker, New York.

[18] Dielman, T. (1992a). Misspecification in random coefficient regression models: a Monte Carlo simulation, *Statistical Papers*, **33**(1), 241–260.

[19] Dielman, T. (1992b). Small sample properties of random coefficient regression estimators: a Monte Carlo simulation, *Communications in Statistics — Simulation and Computation*, **21**(1), 103–132.

[20] Dziechciarz, J. (1989). *Changing and random coefficient models. A survey.* In: "Statistical Analysis and Forecasting of Economic Structural Change" (P. Hackl, Ed.), Springer Berlin Heidelberg, Berlin, 217–251.

[21] Elhorst, J. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer, Berlin.

[22] Elster, C. and Wübbeler, G. (2017). Bayesian inference using a noninformative prior for linear Gaussian random coefficient regression with inhomogeneous within-class variances, *Computational Statistics*, **32**(1), 51–69.

[23] Feige, E. and Swamy, P. (1974). A random coefficient model of the demand for liquid assets, *Journal of Money, Credit and Banking*, **6**(2), 241–252.

[24] Fu, K. and Fu, X. (2015). Asymptotics for the random coefficient first-order autoregressive model with possibly heavy-tailed innovations, *Journal of Computational and Applied Mathematics*, **285**, 116–124.

[25] Horváth, L. and Trapani, L. (2016). Statistical inference in a random coefficient panel model, *Journal of Econometrics*, **193**(1), 54–75.

[26] Hsiao, C. (2014). *Analysis of Panel Data*, 3rd ed., Cambridge University Press, Cambridge.

[27] Hsiao, C. and Pesaran, M. (2008). *Random coefficient models*. In: "The Econometrics of Panel Data", Vol. 46 (L. Matyas, and P. Sevestre, Eds.), Springer Berlin Heidelberg, Berlin, 185–213.

[28] Judge, G.; Griffiths, W.; Hill, R.; Lütkepohl, H. and Lee, T. (1985). *The Theory and Practice of Econometrics*, 2nd ed., Wiley, New York.

[29] Livingston, M.; Erickson, K. and Mishra, A. (2010). *Standard and Bayesian random coefficient model estimation of US Corn–Soybean farmer risk attitudes*. In: "The Economic Impact of Public Support to Agriculture", (V. Ball, R. Fanfani and L. Gutierrez, Eds.), Springer, New York, 329–343.

[30] Mousa, A.; Youssef, A. and Abonazel, M. (2011). *A Monte Carlo study for Swamy's estimate of random coefficient panel data model*, Working paper, No. 49768, University Library of Munich, Germany.

[31] Murtazashvili, I. and Wooldridge, J. (2008). Fixed effects instrumental variables estimation in correlated random coefficient panel data models, *Journal of Econometrics*, **142**, 539–552.

[32] Parks, R. (1967). Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated, *Journal of the American Statistical Association*, **62**, 500–509.

[33] Pesaran, M. and Smith, R. (1995). Estimation of long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics*, **68**, 79–114.

[34] Poi, B. (2003). From the help desk: Swamy's random-coefficients model, *The Stata Journal*, **3**(3), 302–308.

[35] Rao, C. and Mitra, S. (1971). *Generalized Inverse of Matrices and Its Applications*, John Wiley and Sons Ltd, New York.

[36] Rao, U. (1982). A note on the unbiasedness of Swamy's estimator for the random coefficient regression model, *Journal of Econometrics*, **18**(3), 395–401.

[37] Rausser, G.; Mundlak, Y. and Johnson, S. (1982). *Structural change, updating, and forecasting*. In: "New Directions in Econometric Modeling and Forecasting US Agriculture" (G. Rausser, Ed.), North-Holland, Amsterdam.

[38] Sant, D. (1977). Generalized least squares applied to time-varying parameter models, *Annals of Economic and Social Measurement*, **6**(3), 301–314.

[39] Srivastava, V. and Giles, D. (1987). *Seemingly Unrelated Regression Equations Models: Estimation and Inference*, Marcel Dekker, New York.

[40] Swamy, P. (1970). Efficient inference in a random coefficient regression model, *Econometrica*, **38**, 311–323.

[41] Swamy, P. (1971). *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, New York.

[42] Swamy, P. (1973). Criteria, constraints, and multicollinearity in random coefficient regression model, *Annals of Economic and Social Measurement*, **2**(4), 429–450.

[43] Swamy, P. (1974). *Linear models with random coefficients*. In: "Frontiers in Econometrics" (P. Zarembka, Ed.), Academic Press, New York.

[44] Swamy, P.; Mehta, J.; Tavlas, G. and Hall, S. (2015). Two applications of the random coefficient procedure: Correcting for misspecifications in a small area level model and resolving Simpson's paradox, *Economic Modelling*, **45**, 93–98.

[45] Westerlund, J. and Narayan, P. (2015). A random coefficient approach to the predictability of stock returns in panels, *Journal of Financial Econometrics*, **13**(3), 605–664.

[46] Youssef, A. and Abonazel, M. (2009). *A comparative study for estimation parameters in panel data model*, Working paper, No. 49713, University Library of Munich, Germany.

[47] Youssef, A. and Abonazel, M. (2017). Alternative GMM estimators for first-order autoregressive panel model: an improving efficiency approach, *Communications in Statistics — Simulation and Computation*, **46**(4), 3112–3128.

[48] Youssef, A.; El-sheikh, A. and Abonazel, M. (2014a). Improving the efficiency of GMM estimators for dynamic panel models, *Far East Journal of Theoretical Statistics*, **47**, 171–189.

[49] Youssef, A.; El-sheikh, A. and Abonazel, M. (2014b). New GMM estimators for dynamic panel data models, *International Journal of Innovative Research in Science, Engineering and Technology*, **3**, 16414–16425.

[50] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, *Journal of the American Statistical Association*, **57**, 348–368.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 523–549

REPEATED MEASURES ANALYSIS FOR FUNCTIONAL DATA USING BOX-TYPE APPROXIMATION — WITH APPLICATIONS

Author: Łukasz Smaga
– Faculty of Mathematics and Computer Science,
Adam Mickiewicz University in Poznań,
Uniwersytetu Poznańskiego 4, 61-614 Poznań, Poland
ls@amu.edu.pl

Received: November 2016 Revised: March 2017 Accepted: August 2017

Abstract:

- The repeated measures analysis for functional data is investigated. In the literature, the test statistic for the two-sample problem when data are from the same subject is considered. Unfortunately, the known permutation and bootstrap approximations for distribution of it may be time-consuming. To avoid this drawback, a Box-type approximation for asymptotic null distribution of that test statistic is proposed. This approximation results in a new testing procedure, which is more efficient from a computational point of view than the known ones. Root-n consistency of the new method is also proved. Via intensive simulation studies, it is found that in terms of size control and power, the new test is comparable with the known tests. An illustrative example of the use of tests in practice is also given.

Key-Words:

- *Box-type approximation; χ^2-type mixture; functional data; repeated measures analysis; two-cumulant approximation; two-sample test.*

AMS Subject Classification:

- 62H15, 62M99.

1. INTRODUCTION

Functional data analysis (FDA) is concerned with data which are viewed as functions defined over some set T. Examples of functional data can be found in several application domains such as meteorology, medicine, economics and many others (for an overview, see Ramsay and Silverman, 2002). Comprehensive surveys about functional data analysis can be found in Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ramsay *et al.* (2009), Ramsay and Silverman (2002, 2005), Zhang (2013) and in the review papers Cuevas (2014) and Valderrama (2007). Many papers available in the literature are devoted to estimation and classification of functional data, e.g., cluster analysis (Jacques and Preda, 2014; Tokushige *et al.*, 2007; Yamamoto and Terada, 2014), confidence intervals (Lian, 2012), discriminant analysis (Górecki *et al.*, 2014; James and Hastie, 2001; Preda *et al.*, 2007), estimation (Attouch and Belabed, 2014; Chesneau *et al.*, 2013; Cuevas *et al.*, 2006, 2007; Prakasa Rao, 2010), principal component analysis (Berrendero *et al.*, 2011; Boente *et al.*, 2014; Boente and Fraiman, 2000; Jacques and Preda, 2014), variable selection (Gregorutti *et al.*, 2015). Hypothesis testing problems for functional data are also commonly considered, e.g., heteroscedastic ANOVA problem (Cuesta-Albertos and Febrero-Bande, 2010; Zhang, 2013), paired two-sample problem (Martínez-Camblor and Corral, 2011), the one-way ANOVA and MANOVA problem (Abramovich *et al.*, 2004; Cuevas *et al.*, 2004; Horváth and Rice, 2015; Górecki and Smaga, 2015, 2017), testing equality of covariance functions (Zhang, 2013), two-sample Behrens–Fisher problem (Zhang *et al.*, 2010b).

In this paper, the two-sample problem for functional data which are from the same subject (probably submitted to different conditions) is considered. We follow the notation of Martínez-Camblor and Corral (2011). Suppose we have a functional sample consisting of independent trajectories $X_1(t), ..., X_n(t)$ from a stochastic process which may be expressed in the following form

$$X_i(t) = m(t) + \varepsilon_i(t),\ t \in [0, 2], \tag{1.1}$$

where $\varepsilon_i(t)$ are random functions with $E(\varepsilon_i(t)) = 0$ and covariance function $\mathbb{C}(s, t)$. Hence, the null hypothesis is of the form

$$H_0 \colon m(t) = m(t + 1),\ \forall t \in [0, 1]. \tag{1.2}$$

Concerning $t \in [0, 2]$, we ignore (possible) period in which the subject is not monitored.

To illustrate the testing problem described above, we consider the orthosis data. Seven volunteers ($n = 7$) were participated in the experiment. First, they were stepping-in-place without orthosis. Second, they did the same with a spring-loaded orthosis on the right knee. Under each condition, the moment of

force at the knee was computed at 256 time points, equally spaced and scaled to the interval $[0, 1]$. So the orthosis data can be represented as curves. We are interested in testing if the mean curves of all volunteers are different under these two conditions (see Figure 1). As the curves obtained without and with orthosis are from the same subjects (volunteers), we have a paired two-sample problem for functional data. The detailed description of the experiment and its analysis are presented in Section 6.

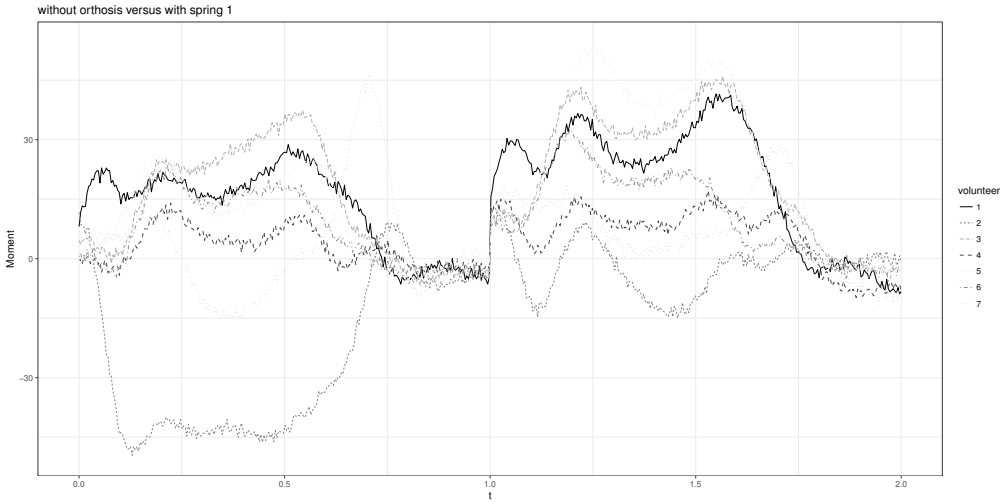

Figure 1: The mean curves of all volunteers of the 10 raw orthosis curves under without orthosis ($t \in [0, 1]$) and with spring 1 ($t \in [1, 2]$) conditions.

For testing (1.2), Martínez-Camblor and Corral (2011) proposed to use the test statistic

$$\mathcal{C}_n = n \int_0^1 (\bar{X}(t) - \bar{X}(t+1))^2 \, dt, \tag{1.3}$$

where $\bar{X}(t) = n^{-1} \sum_{i=1}^n X_i(t), \ t \in [0, 2]$. This test statistic is based on a simple idea that the null hypothesis should be rejected whenever the "between group variability" measured by the difference between sample means is large enough at a prescribed significance level. As in the standard ANOVA test statistic, appropriate "within group variability" measure may be also contained as denominator in $\mathcal{C}_n$. However, then it seems to be impossible to find the exact sampling distribution of such statistic, even under Gaussianity assumption. Moreover, since Martínez-Camblor and Corral (2011) used an asymptotic test (large sample sizes may be required) and such a denominator tends to some parameter connected with covariance function as $n \to \infty$, the denominator could be replaced by that parameter. Then it could be incorporated to the numerator so that it is only necessary to calculate the asymptotic distribution of the test statistic and replaced

by an estimator in that distribution. This reasoning can be used in homoscedastic as well as heteroscedastic case. Bearing in mind this motivation, Martínez-Camblor and Corral (2011) used only the numerator (i.e., $\mathcal{C}_n$), and avoided the homoscedasticity assumption in such a way (see also Cuevas *et al.*, 2004, for similar argumentation).

Martínez-Camblor and Corral (2011) derived a random expression of their test statistic (1.3), and approximated the null distribution by a parametric bootstrap method via re-sampling some Gaussian process involved in the limit random expression of $\mathcal{C}_n$ under the null hypothesis. Moreover, Martínez-Camblor and Corral (2011) considered nonparametric approach and proposed bootstrap and permutation tests. Although these methods work reasonably well in finite samples, they may be time-consuming. In this paper, we present the Box-type approximation (Box, 1954; Brunner *et al.*, 1997; also called two-cumulant approximation, see Zhang, 2013) for the asymptotic distribution of $\mathcal{C}_n$ under the null, and we propose the new test based on this approximation. It is shown to be root-n consistent. The new testing procedure is also much less computationally intensive than the re-sampling and permutation tests of Martínez-Camblor and Corral (2011). Moreover, it is comparable with these tests in terms of size control and power.

This paper is organized as follows. Section 2 presents the Box-type approximation for the asymptotic null distribution of test statistic $\mathcal{C}_n$ and the new test based on this approximation. Its root-n consistency is proved in Section 3. In Section 4, an intensive simulation study providing an idea of the size control and power of the new testing procedure and the tests proposed by Martínez-Camblor and Corral (2011) is given. The comparison of computational time required to perform the considered tests is presented in Section 5. Section 6 contains a real-data example of the use of those tests to the orthosis data. Some concluding remarks are given in Section 7. In the Appendix, proofs of theoretical results, numerical implementation of the new test, R code which performs it and additional simulations are presented.

2. THE TESTING PROCEDURE

In this section, we describe and discuss the new testing procedure for (1.2) which is based on the Box-type approximation for the asymptotic distribution of the test statistic $\mathcal{C}_n$ given by (1.3) under the null.

Let $X_1(t), ..., X_n(t)$ be independent trajectories from a stochastic process (with expectation function $m(t)$ and covariance function $\mathbb{C}(s,t)$, $s,t \in [0,2]$) expressed as in (1.1). For theoretical study, we list the following regularity assumptions.

Assumptions:

A1. The mean function $m(t) \in L^2[0,2]$ and $\mathrm{tr}(\mathbb{C}) \stackrel{\mathrm{def}}{=} \int_0^2 \mathbb{C}(t,t)\ dt < \infty$, where $L^2([0,2])$ denotes the set of all square-integrable functions over $[0,2]$.

A2. The subject-effect function $v_1(t) \stackrel{\mathrm{def}}{=} X_1(t) - m(t)$ satisfies

$$E\|v_1\|^4 = E\left(\int_0^2 v_1^2(t)\ dt\right)^2 < \infty.$$

A3. For any $t \in [0,2]$, $\mathbb{C}(t,t) > 0$, and $\max_{t\in[0,2]} \mathbb{C}(t,t) < \infty$.

A4. For any $(s,t) \in [0,2]^2$, $E\left(v_1^2(s)v_1^2(t)\right) < C < \infty$, where C is certain constant independent of any $(s,t) \in [0,2]^2$.

The given assumptions are quite common in functional data analysis literature (see, for instance, Zhang, 2013; Zhang and Liang, 2014). Assumption A1 is regular. It guarantees that as $n \to \infty$, the sample mean function will converge to Gaussian process weakly. Assumptions A2–A4 are additionally imposed to obtain the consistency of estimator of the covariance function. The uniformly boundedness of $E\left(v_1^2(s)v_1^2(t)\right)$ in assumption A4 is satisfied when the subject-effect function $v_1(t)$ is uniformly bounded in probability over $[0,2]$.

Under assumption A1, by (4.7) in Zhang (2013), we have $E\|X_1\|^2 = \|m\|^2 + \mathrm{tr}(\mathbb{C}) < \infty$. Hence, using the central limit theorem for random elements taking values in a Hilbert space (see, for example, Zhang, 2013, p. 91) and the continuous mapping theorem as in the proof of Theorem 1 in Martínez-Camblor and Corral (2011), under the null hypothesis, we obtain $\mathcal{C}_n \xrightarrow{d} \|\xi\|^2$, as $n \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution, and $\xi(t), t \in [0,1]$ is a Gaussian process with mean zero and covariance function

$$\mathbb{K}(s,t) = \mathbb{C}(s,t) - \mathbb{C}(s,t+1) - \mathbb{C}(s+1,t) + \mathbb{C}(s+1,t+1),\ s,t \in [0,1] \tag{2.1}$$

(see the proof of Theorem 1 in Martínez-Camblor and Corral, 2011, for more details). Under assumptions A1 and A3, we have $\mathrm{tr}(\mathbb{K})$ is finite, where we use the fact $\mathbb{C}(s,t) \le (\mathbb{C}(s,s)\mathbb{C}(t,t))^{1/2} \le \max_{t\in[0,2]} \mathbb{C}(t,t) < \infty$. Thus, Theorem 4.2 in Zhang (2013) implies $\|\xi\|^2$ has the same distribution as $\sum_{k\in\mathbb{N}} \lambda_k A_k$, where A_k, $k = 1, 2, ...$, is a sequence of independent random variables following a central chi-squared distribution with one degree of freedom, and λ_k, $k = 1, 2, ...$, is the non-negative sequence, satisfying $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_r \ge ... \ge 0$ and $\sum_{k\in\mathbb{N}} \lambda_k^2 < \infty$, of the eigenvalues of $\mathbb{K}(s,t)$ given by (2.1). Since $\mathcal{C}_n \xrightarrow{d} \|\xi\|^2$, as $n \to \infty$, we conclude that

$$\mathcal{C}_n \xrightarrow{d} \mathcal{C}_0^* = \sum_{k\in\mathbb{N}} \lambda_k A_k \tag{2.2}$$

under the null and assumptions A1 and A3. Hence, the test statistic $\mathcal{C}_n$ converges in distribution to a central χ^2-type mixture (see Zhang, 2005), under the null and

assumptions A1 and A3. On the basis of (2.2), the asymptotic null distribution of $\mathcal{C}_n$ is known except the unknown eigenvalues λ_k, $k = 1, 2, ...$, of $\mathbb{K}(s,t)$. These unknown eigenvalues can be estimated by the eigenvalues $\hat{\lambda}_k$, $k = 1, 2, ...$, of the following estimator of $\mathbb{K}(s,t)$:

$$\hat{\mathbb{K}}(s,t) = \hat{\mathbb{C}}(s,t) - \hat{\mathbb{C}}(s,t+1) - \hat{\mathbb{C}}(s+1,t) + \hat{\mathbb{C}}(s+1,t+1), \ \ s,t \in [0,1], \tag{2.3}$$

where $\hat{\mathbb{C}}(s,t) = (n-1)^{-1}\sum_{i=1}^{n}(X_i(s) - \bar{X}(s))(X_i(t) - \bar{X}(t)), \ \ s,t \in [0,2]$ is the unbiased estimator of $\mathbb{C}(s,t)$ (see Zhang, 2013, p. 108). Moreover, it is often sufficient to use only the positive eigenvalues of $\hat{\mathbb{K}}(s,t)$. With the sample size n growing to infinity, the estimator $\hat{\mathbb{K}}(s,t)$ is consistent in the sense of the following lemma. Let $\xrightarrow{P}$ denote convergence in probability.

Lemma 2.1. *Under the model* (1.1) *and assumptions A1–A4, we have* $\hat{\mathbb{K}}(s,t) \xrightarrow{P} \mathbb{K}(s,t)$ *uniformly over* $[0,1]^2$, *as* $n \to \infty$.

We now apply Box-type approximation (Box, 1954; Brunner *et al.*, 1997) for approximating the asymptotic null distribution of $\mathcal{C}_n$. This method is also known as two-cumulant approximation (see Zhang, 2013). It is an example of the approximation methods using cumulants, which are often considered in functional data analysis (see, for example, Górecki and Smaga, 2015; Zhang, 2013; Zhang and Liang, 2014; Zhang *et al.*, 2010b), so we also may name it "two-cumulant approximation". The key idea of this method is to approximate the distribution of $\mathcal{C}_0^*$ by that of a random variable of the form $\beta\chi_d^2$, where the parameters β and d are determined by matching the first two cumulants or moments of $\mathcal{C}_0^*$ and $\beta\chi_d^2$. By the results of Zhang (2013, Sections 4.3 and 4.5), we have

$$\beta = \frac{\mathrm{tr}(\mathbb{K}^{\otimes 2})}{\mathrm{tr}(\mathbb{K})}, \qquad d = \frac{\mathrm{tr}^2(\mathbb{K})}{\mathrm{tr}(\mathbb{K}^{\otimes 2})}, \tag{2.4}$$

where $\mathrm{tr}(\mathbb{K}) = \int_0^1 \mathbb{K}(t,t)\,dt$ and $\mathbb{K}^{\otimes 2} \overset{\mathrm{def}}{=} \int_0^1 \mathbb{K}(s,u)\mathbb{K}(u,t)\,du$. The approximation of the distribution of $\mathcal{C}_0^*$ by that of $\beta\chi_d^2$ seems to be sensible, since $\mathcal{C}_0^*$ is a χ^2-type mixture which is nonnegative and generally skewed, and so $\beta\chi_d^2$ is. Thus, $\mathcal{C}_0^*$ and $\beta\chi_d^2$ with β and d as in (2.4) have the same range, mean and variance and similar shapes. However, the distributions of these random variables are usually not the same. Moreover, the conditional distributions of the parametric and nonparametric bootstrap and permutation statistics of Martínez-Camblor and Corral (2011) can be different of the distribution of $\beta\chi_d^2$. Fortunately, these distributions are very similar to each other, and the distribution of $\beta\chi_d^2$ can have flexible shapes and be adaptive to different shapes of the underlying null distribution of $\mathcal{C}_n$, which is confirmed by simulation studies of Section 4. From those simulation studies, we can observe that both the previous and new approximations give very similar and satisfactory results for small and moderate sample sizes. The same holds for large samples. For instance, when $n = 2000$, the empirical sizes of the parametric and nonparametric bootstrap, permutation and new testing procedures were

equal to 5.2%, 4.8%, 5.2%, 4.8%, respectively, and the empirical power of all tests was equal to 100%. These results suggest that the type I error rate (resp. power) of each test tends to the nominal significance level or to value close to it (resp. one) as $n \to \infty$.

The natural estimators of β and d are obtained by replacing the covariance function $\mathbb{K}(s,t)$ in (2.4) by its estimator $\hat{\mathbb{K}}(s,t)$ given by (2.3), i.e.,

$$\hat{\beta} = \frac{\mathrm{tr}(\hat{\mathbb{K}}^{\otimes 2})}{\mathrm{tr}(\hat{\mathbb{K}})}, \qquad \hat{d} = \frac{\mathrm{tr}^2(\hat{\mathbb{K}})}{\mathrm{tr}(\hat{\mathbb{K}}^{\otimes 2})}. \tag{2.5}$$

Therefore, under the null, $\mathcal{C}_n \sim \hat{\beta}\chi^2_{\hat{d}}$ approximately, and hence the new test (the BT test) for (1.2) is conducted by computing the p-value of the form

$$P(\chi^2_{\hat{d}} > \mathcal{C}_n/\hat{\beta}), \tag{2.6}$$

or for given significance level α, the estimated critical value of $\mathcal{C}_n$ given by

$$\hat{\mathcal{C}}_{n,\alpha} = \hat{\beta}\chi^2_{\hat{d},\alpha}, \tag{2.7}$$

where $\chi^2_{r,\alpha}$ denotes the upper 100α percentile of χ^2_r. The critical region of the new testing procedure is of the form $\{\mathcal{C}_n > \hat{\beta}\chi^2_{\hat{d},\alpha}\}$. In the following theorem, we show that the estimated critical value $\hat{\mathcal{C}}_{n,\alpha}$ tends to theoretical critical value $\mathcal{C}_{0,\alpha} = \beta\chi^2_{d,\alpha}$, as $n \to \infty$. The consistency of the estimators $\hat{\beta}$ and $\hat{d}$ is also proved there.

Theorem 2.1. *Under the assumptions of Lemma 2.1, as $n \to \infty$, we have $\hat{\beta} \xrightarrow{P} \beta$ and $\hat{d} \xrightarrow{P} d$. Moreover, we have $\hat{\mathcal{C}}_{n,\alpha} \xrightarrow{P} \mathcal{C}_{0,\alpha} = \beta\chi^2_{d,\alpha}$, as $n \to \infty$.*

Numerical implementation of the BT test is described in the Appendix. This testing procedure is very easy to implement in the R language (R Core Team, 2015). In the Appendix, we also present and describe the R code which performs the new test.

3. ASYMPTOTIC POWER UNDER LOCAL ALTERNATIVES

In this section, we investigate the asymptotic power of the BT test under two kinds of local alternatives. Power of tests under similar types of alternatives was studied in the literature concerning the functional data analysis (see, for example, Zhang *et al.*, 2010a, Zhang and Liang, 2014). The formulas for the asymptotic powers of the BT test are given in the proofs of Theorems 3.1 and 3.2.

First, we consider the local alternatives of the form $H_{1n}^{(1)} : m(t) - m(t+1) = n^{-\tau/2}d(t), t \in [0,1]$, where $\tau \in [0,1)$ is fixed and $d(t)$ is any fixed real function such that $\|d\| \in (0,\infty)$. So, we study the power behavior when the alternatives tend to the null hypothesis (1.2) with a rate slightly slower than $n^{-1/2}$. In the following result, we establish the asymptotic power of the BT test tends to one, as $n \to \infty$, under $H_{1n}^{(1)}$ and under gaussianity assumption of processes $X_i(t), i = 1, ..., n$ in model (1.1).

Theorem 3.1. *Under model* (1.1)*, where* $X_i(t), i = 1, ..., n$ *are Gaussian processes, assumptions A1–A4 and the local alternatives* $H_{1n}^{(1)}$*,* $\tau \in [0,1)$*, the asymptotic power of the BT test tends to* 1 *as* $n \to \infty$*.*

We now consider the local alternatives, which tend to the null hypothesis (1.2) with the root-n rate, i.e., $H_{1n}^{(2)} : m(t) - m(t+1) = n^{-1/2}d(t), t \in [0,1]$, where $d(t)$ is any fixed real function such that $\|d\| \in (0,\infty)$. Here, we do not assume gaussianity of the observations, but the asymptotic power of the BT test tending to 1 is obtained when the information provided by $d(t)$ diverges to infinity. This is presented in the following theorem.

Theorem 3.2. *Under model* (1.1)*, assumptions A1–A4 and the local alternatives* $H_{1n}^{(2)}$*, as* $n \to \infty$*, the asymptotic power of the BT test tends to* 1 *as* $\|d\| \to \infty$*.*

Theorems 3.1 and 3.2 indicate that the BT test can detect the local alternatives $H_{1n}^{(1)}$ and $H_{1n}^{(2)}$ with probability tending to one under the assumptions given above. By the definition of Zhang and Liang (2014), we obtain that the BT test is root-n consistent.

4. SIMULATIONS

Simulations are conducted to compare the empirical sizes (type I error rates) and powers of the BT test with those of Martínez-Camblor and Corral (2011). As we mentioned, Martínez-Camblor and Corral (2011) proposed three approximation methods for the null distribution of $\mathcal{C}_n$ based on the asymptotic distribution (the A test), on bootstrap (the B test), and on permutation (the P test). Additional simulations considering different dependency structure than that in this section are given in the Appendix. All simulations were conducted with the help of the R computing environment (R Core Team, 2015).

4.1. Description of the simulation experiments

To be consistent with the results of Martínez-Camblor and Corral (2011) for the A, B and P tests, we present similar simulation experiments to those in that paper. We generated $X_i(t) = m_1(t) + \varepsilon_{i1}(t)$ and $X_i(t+1) = m_2(t) + \varepsilon_{i2}(t)$ for $t \in [0,1]$, $i = 1, ..., n$, where $m_j(t)$ and $\varepsilon_{ij}(t)$ are described below. Sample sizes $n = 25, 35, 50$ are considered. Let

$$m_{0,1}(t) = \sqrt{6t/\pi}\exp(-6t)I_{[0,1]}(t), \quad m_{1,1}(t) = \sqrt{13t/(2\pi)}\exp(-13t/2)I_{[0,1]}(t),$$
$$m_{2,1}(t) = \sqrt{11t/(2\pi)}\exp(-11t/2)I_{[0,1]}(t), \quad m_{3,1}(t) = \sqrt{5}t^{2/3}\exp(-7t)I_{[0,1]}(t),$$
$$m_{0,2}(t) = (\sin(2\pi t^2))^5 I_{[0,1]}(t), \quad m_{1,2}(t) = (\sin(2\pi t^2))^3 I_{[0,1]}(t),$$
$$m_{2,2}(t) = (\sin(2\pi t^2))^7 I_{[0,1]}(t), \quad m_{3,2}(t) = (\sin(2\pi t^{(9/5)}))^3 I_{[0,1]}(t).$$

Figure 2 depicts the shapes of $m_{i,j}(t)$. Because of the choice of $m_i(t), i = 1, 2$, we considered eight models. In models M0–M3, $m_1 = m_{0,1}$ and $m_2 = m_{j,1}$, $j = 0, ..., 3$ respectively, and in models M4–M7, $m_1 = m_{0,2}$ and $m_2 = m_{j,2}$, $j = 0, ..., 3$ respectively.

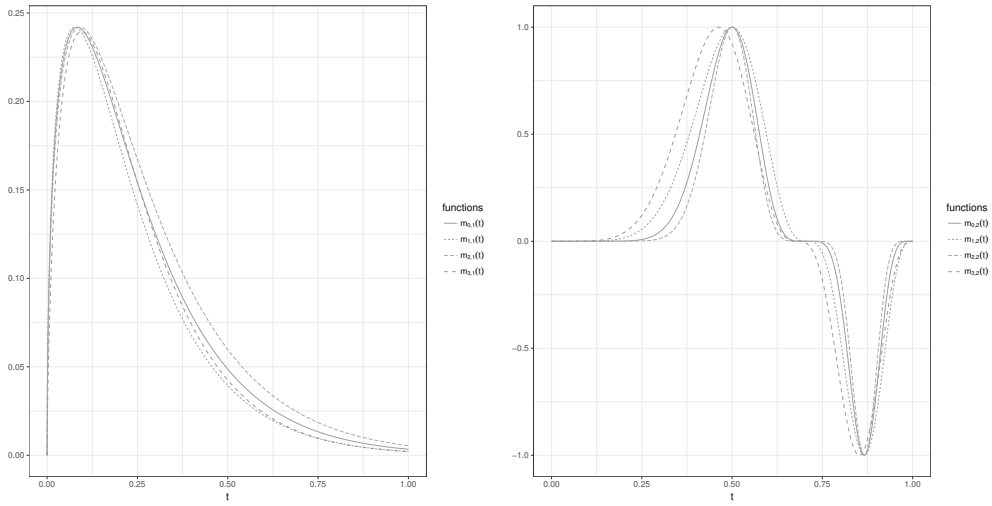

Figure 2: The shapes of functions $m_{i,j}(t), t \in [0,1]$ used in simulations of Section 4.

Three different types of errors were considered. In the *normal* case, $\varepsilon_{i1}(t) = \xi B_{i1}(t)$ and $\varepsilon_{i2}(t) = \rho\varepsilon_{i1}(t) + \xi\sqrt{1-\rho^2}B_{i2}(t)$, where $\rho = 0, 0.25, 0.5$, B_{i1} and B_{i2} are two independent standard Brownian Bridges, and $\xi = 0.05$ for models M0–M3 and $\xi = 0.5$ for the remaining. In the *lognormal* (resp. *mixed*) case, the error functions are $\exp(\varepsilon_{ij}(t))$, $j = 1, 2$ (resp. $\varepsilon_{i1}(t)$ and $\exp(\varepsilon_{i2}(t))$), where $\varepsilon_{ij}(t)$ are as above. The errors functions $\exp(\varepsilon_{ij}(t))$ are adequately centered.

In practice, the functional data are not usually continuously observed. The points, at which the functional data are observed, are called the design time points. So, the processes $X_i(t), X_i(t+1), t \in [0,1]$ were generated in discretized versions $X_i(t_r)$, $X_i(t_r+1)$, for $r = 1, ..., I$ and for $I = 26, 101, 251$, where the values t_r were chosen equispaced in the interval $[0,1]$.

Under various parameter configurations, the empirical sizes and powers (as percentages) of the tests were calculated at the nominal significance level $\alpha = 5\%$ and based on 1000 replications. In Tables 1–7, the results for models M0–M6 are displayed. For model M7, the empirical powers were always 100%. The empirical power in the omitted rows in these tables is always 100%. Similarly as in Martínez-Camblor and Corral (2011), the p-values of the A, B and P tests were estimated from 1000 replications.

Table 1: Empirical sizes (as percentages) of all tests obtained in model M0. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed).

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	**6.9**	**6.8**	6.0	**6.5**	6.1	6.2	4.9	6.1	5.3	5.4	5.0	5.3
		0.25	**6.8**	**7.1**	6.4	**6.8**	**6.8**	**6.9**	5.9	**7.1**	5.2	5.6	5.1	5.3
		0.50	**7.2**	**7.6**	6.4	**7.1**	**7.2**	**7.0**	6.2	**6.7**	5.1	5.7	4.6	5.2
	35	0.00	5.5	6.1	5.7	6.1	4.8	4.5	4.0	4.7	4.3	4.5	4.1	4.3
		0.25	5.4	6.1	5.4	5.5	5.4	5.6	5.1	5.1	4.2	4.4	3.6	4.2
		0.50	5.9	6.0	5.1	5.9	5.4	5.6	5.2	5.7	4.4	4.3	3.7	4.2
	50	0.00	5.4	5.2	4.8	5.2	6.0	6.1	5.7	5.8	**6.5**	5.8	5.7	6.1
		0.25	5.2	5.7	5.1	5.0	6.3	**6.6**	**6.5**	6.4	6.3	5.7	5.9	5.7
		0.50	5.3	5.4	5.0	5.0	**7.2**	**7.1**	6.3	**7.0**	5.4	5.1	4.9	5.2
L	25	0.00	**6.8**	**6.7**	**6.7**	6.4	5.4	5.7	5.7	5.5	5.1	4.8	5.5	5.0
		0.25	**7.2**	**7.4**	**7.2**	**7.1**	5.1	5.3	5.1	4.9	5.7	5.8	5.7	5.6
		0.50	**7.2**	**8.0**	**7.9**	**7.6**	5.2	5.3	5.3	5.4	5.6	5.9	6.0	5.9
	35	0.00	5.4	5.3	4.8	5.3	5.5	5.0	5.2	5.2	4.8	5.0	4.8	4.8
		0.25	4.7	4.3	4.6	4.5	5.2	5.2	5.0	5.2	5.0	4.8	4.8	4.7
		0.50	4.7	4.9	4.7	4.6	5.9	5.9	5.7	5.7	4.7	4.6	4.5	4.6
	50	0.00	5.0	5.3	4.9	5.1	5.3	5.3	4.9	4.8	4.9	4.9	5.2	5.3
		0.25	5.8	5.7	5.7	5.6	5.0	5.1	4.5	4.7	4.7	4.9	4.6	4.8
		0.50	5.5	5.5	5.7	5.8	5.0	5.3	5.1	5.1	4.4	4.6	4.4	4.6
M	25	0.00	5.1	5.6	4.8	5.2	5.6	6.0	5.6	5.6	5.8	5.6	5.8	5.6
		0.25	5.7	5.3	5.2	5.0	5.2	5.6	5.4	5.6	6.2	6.0	5.8	5.7
		0.50	5.8	6.1	5.6	5.9	5.9	5.9	5.6	6.1	6.1	6.1	5.3	6.0
	35	0.00	5.3	5.5	5.2	5.3	4.9	4.7	4.5	4.7	4.5	4.8	4.7	4.8
		0.25	4.8	5.0	4.7	4.9	4.9	5.3	4.8	5.0	4.8	5.0	4.9	4.6
		0.50	4.9	5.3	4.7	4.8	5.1	5.2	4.5	5.3	4.6	5.7	4.4	4.6
	50	0.00	5.1	5.1	5.0	5.1	5.5	5.3	5.4	5.2	5.8	5.7	5.6	5.8
		0.25	4.9	4.9	4.9	4.7	5.3	5.5	5.0	5.0	6.2	6.1	5.5	6.1
		0.50	4.5	4.8	4.5	4.6	5.7	5.8	5.5	5.5	5.9	6.4	5.6	6.0

Table 2: Empirical powers (as percentages) of all tests obtained in model M1. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	40.5	40.4	38.2	39.5	39.8	39.9	36.6	39.5	38.5	38.6	36.7	38.8
		0.25	49.2	49.6	46.8	48.9	51.4	51.5	48.3	51.0	50.2	49.8	47.4	49.5
		0.50	68.5	68.0	66.0	68.0	70.2	71.0	68.7	69.8	69.3	69.6	66.6	69.1
	35	0.00	53.1	53.8	51.6	52.7	52.1	52.0	50.2	52.3	54.4	54.5	53.1	54.5
		0.25	65.7	66.5	65.2	65.9	64.7	65.2	63.2	65.3	66.9	66.8	65.7	66.3
		0.50	85.4	85.7	84.6	84.9	81.6	81.3	80.1	81.1	85.0	85.0	84.2	84.8
	50	0.00	69.2	68.9	68.2	68.8	68.7	69.3	68.3	68.6	69.8	68.6	68.4	69.1
		0.25	81.8	81.1	81.2	81.1	81.0	81.4	80.9	81.1	82.2	82.9	81.7	82.4
		0.50	93.9	93.9	93.7	93.9	94.5	94.5	94.1	94.5	94.3	94.0	94.0	94.0
L	25	0.00	98.8	98.1	99.0	98.6	99.0	99.2	99.3	98.9	99.2	99.3	99.6	99.4
M	25	0.00	62.6	62.8	61.2	61.8	61.6	61.9	61.3	61.7	61.3	62.0	60.6	61.6
		0.25	67.2	67.6	65.7	67.1	67.3	68.6	66.4	67.5	67.6	67.6	66.8	67.3
		0.50	73.2	74.0	72.1	73.6	74.1	74.6	73.2	74.7	75.0	75.5	73.3	74.3
	35	0.00	76.8	77.5	76.5	76.9	77.2	76.5	76.9	76.7	78.6	79.8	78.4	78.8
		0.25	82.7	82.7	82.0	82.8	82.6	83.0	81.6	82.1	84.7	84.5	83.5	84.8
		0.50	87.5	88.1	87.4	87.1	87.3	87.2	86.5	87.0	88.5	88.1	88.2	88.3
	50	0.00	89.9	90.7	90.0	90.0	90.9	91.0	91.2	91.4	91.0	91.1	90.8	91.3
		0.25	93.7	93.8	92.9	93.6	94.3	94.0	93.9	93.9	94.4	94.3	94.5	94.5
		0.50	96.1	96.0	95.8	96.1	97.0	97.1	96.9	97.1	97.3	97.5	97.2	97.4

Table 3: Empirical powers (as percentages) of all tests obtained in model M2. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	49.3	49.0	46.9	48.5	50.0	50.1	48.0	49.6	49.1	49.8	46.5	48.7
		0.25	60.7	62.1	58.8	60.6	61.9	62.5	59.5	61.7	63.0	62.8	59.2	62.2
		0.50	78.6	79.0	77.0	78.6	79.5	79.4	78.5	79.3	79.4	79.4	77.7	79.1
	35	0.00	62.4	61.6	60.3	61.2	65.4	65.2	64.1	64.8	63.1	62.2	60.7	62.1
		0.25	76.8	77.3	75.9	77.2	78.1	78.1	77.3	78.0	75.4	75.4	73.5	74.5
		0.50	90.9	91.4	90.3	91.0	91.5	91.6	90.7	91.5	89.8	89.9	89.6	90.1
	50	0.00	79.9	80.2	79.2	79.9	79.9	80.1	78.7	79.8	79.7	80.5	79.2	79.6
		0.25	90.7	90.4	89.6	90.3	90.1	90.0	89.5	90.0	90.2	89.9	89.5	89.7
		0.50	97.9	97.8	97.6	97.8	98.5	98.2	98.3	98.3	98.1	98.3	97.8	98.2
L	25	0.00	100	100	100	99.9	99.9	99.9	100	100	100	99.9	100	100
M	25	0.00	74.1	74.2	72.8	73.5	75.5	75.4	74.4	74.8	75.4	76.0	74.3	75.3
		0.25	79.9	79.8	79.2	79.8	81.7	81.5	80.5	81.1	80.9	81.6	81.2	81.1
		0.50	85.7	85.6	84.0	85.5	86.8	87.5	85.9	87.2	85.7	85.8	85.1	85.7
	35	0.00	87.5	87.8	87.5	87.4	89.3	89.7	89.2	89.3	87.9	88.0	87.2	88.0
		0.25	92.7	92.5	92.2	92.4	93.0	92.4	92.3	92.9	92.0	91.8	91.8	91.7
		0.50	95.9	96.0	95.7	95.9	95.9	95.8	95.8	95.7	95.2	94.8	94.6	95.0
	50	0.00	97.8	97.6	97.8	97.7	97.0	96.9	97.0	96.7	97.6	97.7	97.7	97.5
		0.25	98.9	98.8	98.8	98.9	98.5	98.5	98.6	98.5	98.7	98.6	98.7	98.6
		0.50	99.4	99.2	99.2	99.5	99.3	99.4	99.4	99.3	99.4	99.4	99.4	99.5

Table 4: Empirical powers (as percentages) of all tests obtained in model M3. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	19.3	19.6	18.0	18.4	31.4	31.6	30.9	30.9	31.4	32.7	31.7	31.7
		0.25	25.1	25.2	23.6	24.3	46.9	46.2	47.3	45.3	46.3	46.3	47.0	44.5
		0.50	39.6	39.6	38.9	39.2	76.2	75.9	78.2	75.4	80.0	81.7	81.9	80.5
	35	0.00	26.0	25.7	26.0	26.3	46.6	45.9	47.3	46.1	50.5	50.7	50.1	49.6
		0.25	37.1	37.7	36.8	36.9	67.2	68.3	69.7	67.8	73.2	73.7	74.6	73.3
		0.50	60.3	60.4	60.4	59.7	96.2	95.5	96.2	96.0	97.7	97.5	98.2	97.7
	50	0.00	40.3	41.2	40.1	39.8	77.5	77.8	78.5	77.2	82.9	82.9	83.7	82.5
		0.25	54.1	53.9	54.5	53.3	95.0	94.9	96.1	95.1	97.6	97.7	97.4	97.7
		0.50	84.6	84.2	84.6	83.4	100	100	100	100	100	100	100	100
L	25	0.00	70.4	71.7	75.6	71.5	97.9	97.9	98.8	98.1	99.2	99.1	99.4	99.3
		0.25	91.5	91.9	94.4	91.1	99.8	99.8	100	99.7	100	100	100	100
		0.50	99.9	99.9	99.9	99.8	100	100	100	100	100	100	100	100
	35	0.00	93.0	92.7	94.7	92.6	100	100	100	100	100	100	100	100
		0.25	99.9	99.9	100	99.7	100	100	100	100	100	100	100	100
	50	0.00	100	100	99.9	100	100	100	100	100	100	100	100	100
M	25	0.00	33.5	34.1	34.6	32.9	62.8	62.6	64.8	62.2	66.6	67.3	69.9	67.0
		0.25	39.6	40.4	40.6	39.4	71.1	71.5	75.1	70.6	77.1	77.6	79.4	76.6
		0.50	46.8	46.9	47.4	45.6	79.1	80.0	82.0	79.3	84.7	84.4	85.5	84.2
	35	0.00	48.6	48.7	50.0	48.0	87.2	87.0	89.0	87.6	90.7	90.6	92.8	90.6
		0.25	57.2	56.8	58.3	56.9	92.8	93.0	93.5	92.6	95.3	95.6	96.2	95.3
		0.50	65.3	65.5	66.1	65.3	96.0	96.1	96.6	96.1	97.9	97.8	98.1	98.0
	50	0.00	75.0	75.1	75.8	74.8	99.4	99.6	99.6	99.6	99.8	99.5	99.8	99.9
		0.25	83.0	82.8	83.1	81.8	99.8	99.8	100	99.8	100	100	99.9	100
		0.50	89.4	89.1	89.3	88.8	99.9	99.9	100	100	100	100	100	100

4.2. Results

In this subsection, we describe the simulation results for the new method and the tests of Martínez-Camblor and Corral (2011).

Tables 1 and 5 display the empirical sizes of the tests obtained in models M0 and M4. Based on the binomial proportion confidence interval, for the nominal level $\alpha = 5\%$, the empirical size over the 1000 independent replications should belong to the interval $[3.6\%, 6.4\%]$ (resp. $[3.2\%, 6.8\%]$) with probability 95% (resp. 99%). Therefore in Tables 1 and 5, when the rejection proportions are outside the 95% significance limits, they are displayed in bold, and when they are outside the 99% significance limits they are underlined. The results for the BT test and the tests proposed in Martínez-Camblor and Corral (2011) are generally quite satisfactory, and the nominal level is well respected in most cases by the tests. Their empirical sizes are rarely larger than the upper endpoint of the 95% confidence interval, and they are not less than lower endpoint of that interval.

Under normal and mixed cases, the B test is the most liberal of all the tests, and it is slightly more liberal than the A and BT tests, which are more liberal than the P test. Nevertheless, the P test is not conservative. Under lognormal case, the empirical sizes do not express such a tendency in general. In model M0 and normal case, the empirical sizes of all tests decrease when I increases for $n = 25, 35$, and they increase when $n = 50$. In the other cases of model M0 and in model M4, this observation is not true generally, and the behavior of the empirical sizes is more complicated when I increases. Summarising, the new test respects the nominal level a bit better than the A and B tests and may be more liberal than the P test.

Table 5: Empirical sizes (as percentages) of all tests obtained in model M4. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed).

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	5.3	5.2	4.7	5.3	5.7	6.1	5.4	5.5	5.1	5.0	4.2	4.6
		0.25	5.4	5.4	5.0	5.3	5.2	5.4	5.1	5.4	5.4	5.4	4.3	5.4
		0.50	5.2	5.1	4.6	5.2	5.0	5.1	4.5	4.9	5.7	5.8	4.7	5.6
	35	0.00	4.9	5.0	4.3	4.7	**6.8**	**6.7**	6.4	**6.7**	4.7	5.0	4.5	4.8
		0.25	5.5	5.7	4.9	5.7	**6.8**	**7.0**	**6.5**	**6.8**	4.9	4.8	4.4	4.9
		0.50	5.9	6.3	5.6	6.3	**6.7**	**6.6**	6.3	6.4	4.9	5.0	5.1	5.2
	50	0.00	6.3	6.1	5.7	6.0	5.8	5.7	5.5	5.9	4.9	5.1	4.9	4.9
		0.25	6.3	6.4	5.9	6.4	5.3	5.5	4.8	5.2	5.3	5.5	5.0	5.4
		0.50	5.9	6.4	5.4	6.2	5.3	5.4	5.0	5.2	5.5	5.8	5.2	5.6
L	25	0.00	4.0	4.4	4.4	4.4	5.6	5.6	5.5	5.6	5.4	5.3	5.2	5.2
		0.25	3.9	4.0	4.0	4.0	6.0	6.0	6.2	5.9	5.6	5.5	5.4	5.4
		0.50	3.7	4.1	4.1	3.8	5.9	6.3	6.4	6.1	4.8	5.0	5.1	4.8
	35	0.00	4.4	4.9	4.9	4.9	4.6	4.7	4.9	4.8	5.4	5.7	5.2	5.3
		0.25	4.2	4.1	4.5	4.2	4.2	4.4	4.8	4.5	5.3	5.3	5.5	5.1
		0.50	4.4	4.6	4.8	4.5	4.6	4.9	5.6	5.2	5.4	5.2	5.1	5.2
	50	0.00	5.0	5.2	5.2	5.0	4.9	5.4	5.1	5.2	4.8	5.5	5.1	5.2
		0.25	5.4	5.2	5.7	5.6	5.3	5.2	5.0	5.0	4.9	5.9	5.5	5.4
		0.50	5.8	5.7	5.7	5.8	6.2	5.8	5.6	5.7	5.3	5.7	5.6	5.4
M	25	0.00	5.6	5.1	5.3	5.4	5.7	6.0	5.9	5.6	6.3	6.2	6.1	6.2
		0.25	5.5	5.3	5.1	5.4	5.7	6.4	5.4	5.9	6.2	6.2	6.1	6.3
		0.50	5.6	5.5	5.3	5.4	**6.6**	**6.8**	6.2	**6.6**	**6.6**	6.2	6.1	5.9
	35	0.00	5.4	5.8	5.5	5.5	5.8	5.9	5.7	5.8	4.7	4.8	4.6	4.6
		0.25	5.2	5.4	5.1	5.1	5.9	5.8	5.9	5.9	5.1	5.4	4.6	5.1
		0.50	5.8	5.6	4.9	5.1	5.9	6.1	5.6	5.8	5.4	5.3	5.2	5.1
	50	0.00	**7.8**	**8.0**	**7.5**	**7.5**	6.4	**7.0**	**6.6**	**6.7**	5.7	6.1	**6.5**	5.9
		0.25	**8.6**	**8.5**	**7.9**	**8.6**	6.4	**6.6**	6.4	6.4	6.0	**6.5**	**6.5**	6.2
		0.50	**8.5**	**8.5**	**7.9**	**8.5**	**6.6**	**6.7**	6.1	**6.6**	6.2	**6.7**	**6.7**	**6.5**

Table 6: Empirical powers (as percentages) of all tests obtained in model M5. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	71.9	71.9	72.0	71.5	74.3	74.2	73.7	73.3	72.0	73.2	73.9	72.2
		0.25	89.4	89.2	89.7	89.0	90.5	89.5	90.5	89.6	87.3	88.4	89.0	87.5
		0.50	99.3	99.2	99.6	99.5	99.2	99.0	99.4	99.2	99.3	99.2	99.4	99.1
	35	0.00	91.4	91.8	91.5	91.4	92.2	93.6	93.5	92.8	92.4	92.4	92.5	91.8
		0.25	99.1	98.9	99.2	99.1	99.1	99.1	99.4	99.1	99.1	99.1	99.5	99.3
	50	0.00	99.4	99.6	99.8	99.7	99.7	99.6	99.7	99.7	99.6	99.8	99.8	99.8
L	25	0.00	99.4	99.2	99.3	98.9	99.6	99.6	99.7	99.5	99.3	99.3	99.6	99.4
		0.25	99.9	99.9	99.9	99.9	100	100	100	100	100	100	100	100
	35	0.00	99.9	99.9	99.9	99.9	100	100	100	100	100	100	100	100
M	25	0.00	95.8	96.6	97.3	96.4	97.4	97.7	98.1	97.9	96.5	96.2	96.6	96.3
		0.25	98.8	98.6	98.7	98.7	99.1	99.2	99.3	99.2	98.5	98.8	99.0	98.7
		0.50	99.8	99.8	99.8	99.9	99.8	99.9	99.8	99.8	99.8	99.8	99.7	99.8
	35	0.00	99.9	99.9	99.9	99.9	99.9	99.8	100	100	99.9	99.9	99.9	99.9

Table 7: Empirical powers (as percentages) of all tests obtained in model M6. The column "R" refers to different residual types (N – normal, L – lognormal, M – mixed). The empirical power in the omitted rows is always 100%.

R	n	I	26				101				251			
		ρ	A	B	P	BT	A	B	P	BT	A	B	P	BT
N	25	0.00	23.1	23.4	22.6	22.5	23.1	23.0	22.8	23.2	23.3	22.8	22.2	23.0
		0.25	31.7	32.1	30.8	30.8	32.9	32.3	32.0	31.4	30.6	30.6	30.5	30.3
		0.50	51.9	51.9	51.4	50.3	53.8	54.0	55.1	52.8	51.3	52.5	52.1	50.7
	35	0.00	32.3	33.0	32.2	32.1	35.5	35.3	35.0	34.5	33.3	33.0	32.0	33.2
		0.25	47.8	47.8	48.6	47.5	49.2	49.8	49.9	49.6	47.4	48.9	48.9	48.1
		0.50	76.3	76.6	76.7	75.4	77.3	77.3	78.9	76.8	75.2	75.2	77.3	75.6
	50	0.00	52.8	50.9	52.9	51.4	54.1	54.7	55.1	53.8	56.9	57.4	56.2	56.7
		0.25	74.5	73.3	75.6	74.2	75.2	75.4	76.2	75.5	76.3	76.4	76.4	75.7
		0.50	95.3	95.4	96.3	95.6	95.4	95.8	96.6	95.9	96.6	96.7	96.9	96.3
L	25	0.00	61.4	61.1	63.8	62.2	63.7	64.3	65.7	63.4	61.6	63.2	64.8	61.8
		0.25	76.8	77.3	79.1	77.6	79.2	79.7	82.1	79.7	78.6	80.2	81.7	79.0
		0.50	93.8	94.5	95.7	94.6	95.3	95.7	96.2	95.4	95.8	96.0	96.9	96.0
	35	0.00	82.1	82.8	83.9	82.7	82.8	82.7	84.6	83.4	82.5	82.5	84.4	82.7
		0.25	93.3	93.4	94.0	93.1	95.3	95.1	95.8	95.2	94.3	94.5	95.4	94.5
		0.50	99.5	99.9	99.6	99.6	99.7	99.7	99.8	99.7	99.5	99.5	99.6	99.5
	50	0.00	95.9	95.7	96.5	95.9	96.6	96.9	97.2	97.1	96.8	96.8	97.6	97.3
		0.25	98.9	99.1	99.0	98.8	99.5	99.5	99.6	99.6	99.6	99.6	99.8	99.6
M	25	0.00	33.9	34.5	35.4	33.6	34.4	35.7	36.0	34.1	34.7	35.9	36.4	34.3
		0.25	41.0	40.7	41.4	39.8	41.9	42.2	43.7	42.0	42.1	42.4	43.7	41.8
		0.50	47.5	48.0	49.3	47.3	49.7	50.9	51.3	49.1	50.4	49.9	51.5	49.6
	35	0.00	50.8	50.7	52.3	50.4	51.0	51.6	53.6	51.7	51.4	52.0	53.9	51.8
		0.25	60.5	61.7	62.8	60.4	63.4	61.9	64.6	62.3	60.7	60.9	62.2	60.2
		0.50	71.8	71.9	73.4	71.6	75.8	76.1	76.9	75.7	72.9	74.0	74.4	73.1
	50	0.00	74.8	74.3	75.8	74.2	73.6	73.3	75.2	73.4	75.4	74.2	76.4	74.6
		0.25	84.6	85.3	85.9	84.6	84.6	84.0	85.3	83.9	86.9	86.1	87.1	86.2
		0.50	92.7	93.6	93.0	92.5	92.8	93.2	93.5	92.8	94.4	94.6	94.4	94.1

The empirical powers of the testing procedures obtained in models M1–M3 and M5–M6 are given in Tables 2–4 and 6–7. Similarly to the empirical sizes, the empirical powers are also quite satisfactory. The observed differences among the empirical powers of all tests are very small. In models M1–M2, the B test is usually a bit better than the other tests, while in models M3 and M5–M6, the P test has such property. In models M1–M2 and M5–M6, the empirical powers of each test are similar among different I's, while in model M3, they increase when I increases. They also increase with n or ρ. Since in models M3 and M6 the functions m_1 and m_2 are very close to each other, the observed empirical powers are usually moderate. In the other models, they are generally quite high even for small n and ρ in all considered situations. Thus, the empirical powers of the BT test are comparable with those of the tests proposed by Martínez-Camblor and Corral (2011), and their behavior is quite satisfactory.

5. SPEED COMPARISON

In this section, we study how the computational time required to perform the A, B, P and BT tests depends on the number of observations n and the number of design time points.

In the experiments, the functional data were generated under models considered in Section 4. We changed $n = 100, 200, ..., 1000$ and $I = 500, 1000$. As an example, Figure 3 shows the execution times against n for obtaining the final p-values of the A, B, P and BT tests when the data were generated as in model M4 under normal case and $\rho = 0.5$, $I = 500$ or $I = 1000$. The results obtained in the other models are similar.

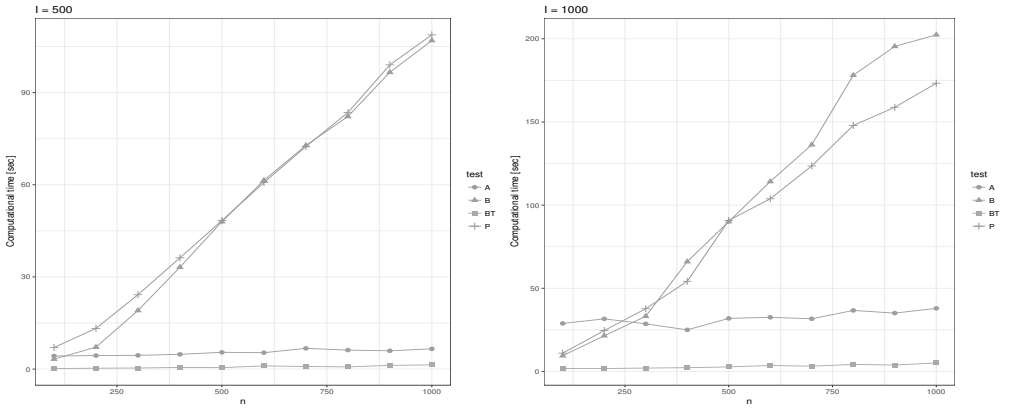

Figure 3: The execution times versus n for obtaining the final p-values of the A, B, P and BT tests when the number I of design time points in $[0, 1]$ as well as in $[1, 2]$ is equal to 500 or 1000. The data were generated as in model M4 under normal case and $\rho = 0.5$.

First of all, the BT test is the fastest among all considered ones, as was expected. It may be extremely faster than the testing procedures of Martínez-Camblor and Corral (2011), and works at most a few seconds. The execution time for the A test almost does not depend on the number of observations. This follows from that in the implementation of the A test, the data are used only to calculate the value of test statistic and the estimator of covariance function (This is done only once.). However, the execution time for this testing procedure increases significantly with an increase of the number of design time points, since the generation of artificial trajectories of the Gaussian process ξ described in Section 2 strongly depends on it. In most cases, the nonparametric bootstrap and permutation methods are the slowest ones. Their execution times are quite similar and increase much with an increase of n or I.

Summarizing, the BT test works very fast even for big data sets, in contrast to the other testing procedures under consideration.

6. APPLICATIONS TO THE ORTHOSIS DATA

In this section, we apply the new test and the testing procedures proposed by Martínez-Camblor and Corral (2011) to real-data example, using orthosis data, which are available on the website of Professor Jin-Ting Zhang (`http://www.stat.nus.edu.sg/ zhangjt/books/Chapman/FANOVA.htm`). These data were used for illustrative purposes in many problems for functional data (see, for instance, Abramovich *et al.*, 2004; Górecki and Smaga, 2015; Zhang and Liang, 2014).

Abramovich *et al.* (2004) reported the orthosis data were acquired and computed in an experiment by Dr. Amarantini David and Dr. Martin Luc (Laboratoire Sport et Performance Motrice, EA 597, UFRAPS, Grenoble University, France). The aim of their research was to investigate how muscle copes with an external perturbation. Seven young male volunteers participated in the experiment. They wore a spring-loaded orthosis of adjustable stiffness under the following four experimental conditions: a control condition (without orthosis); an orthosis condition (with orthosis); and two spring conditions (with spring 1 or with spring 2) in which stepping-in-place was perturbed by fitting a spring-loaded orthosis onto the right knee joint. All volunteers tried all four conditions 10 times for 20 seconds each. In order to avoid possible perturbations in the initial and final parts of the experiment, only the central 10 seconds were used in the study. The resultant moment of force at the knee was derived by means of body segment kinematics recorded with a sampling frequency of 200 Hz. For each stepping-in-place replication, the resultant moment was computed at 256 time points, equally spaced and scaled to the interval $[0, 1]$ so that a time interval corresponded to an individual gait cycle.

For illustrative purposes, we use the orthosis data under the first (without orthosis) and third (with spring 1) experimental conditions. For each volunteer, we calculate the mean curve of the 10 raw orthosis curves under these conditions. Figure 1 depicts the resulting curves. Of interest is to test if the mean curves of all volunteers are different under these two conditions ($t \in [0,1]$ — without orthosis; $t \in [1,2]$ — with spring 1). This is a paired two-sample problem for functional data. We applied the A, B, P and BT tests to this problem and the p-values of these tests are equal to 0.001, 0, 0, 0.0008123766 respectively. Hence all testing procedures suggest that the mean curves of all volunteers under without orthosis and with spring 1 conditions are unlikely the same. From Figure 1, however, we observe that the mean curves may be the same at the last stage of the experiment, i.e., for $t \in [0.8,1] \cup [1.8,2]$. In this case, the p-values of the A, B, P and BT tests are equal to 0.201, 0.204, 0.241, 0.2368321 respectively, and hence we fail to reject the equality of mean curves of all volunteers under without orthosis and with spring 1 conditions over $[0.8,1] \cup [1.8,2]$. Zhang and Liang (2014) also observed similar behavior of orthosis curves at the last stage of the experiment and confirmed its evidence by using appropriate tests. However, they considered the orthosis data under all four experimental conditions in the context of the functional analysis of variance.

7. CONCLUSIONS

In this paper, we studied the paired two-sample problem for functional data. We proposed the test for this problem based on the test statistic considered by Martínez-Camblor and Corral (2011) and the Box-type approximation for its asymptotic null distribution. This testing procedure is root-n consistent, easy to implement and much less computationally intensive than the re-sampling and permutation tests of Martínez-Camblor and Corral (2011). Moreover, it is comparable with those tests in terms of size control and power, and its finite sample behavior is very satisfactory. The illustrative real-data example indicates that the decisions suggested by the new test and the testing procedures of Martínez-Camblor and Corral (2011) seem to be similar in practice.

APPENDIX

A. Proofs

In the proofs, we use similar techniques as in Zhang *et al.* (2010a) and Zhang and Liang (2014).

Proof of Lemma 2.1: Under assumptions A1–A4, from the proof of Theorem 4.17 in Zhang (2013), it follows that $\hat{\mathbb{C}}(s,t) \xrightarrow{P} \mathbb{C}(s,t)$ uniformly over $[0,2]^2$, as $n \to \infty$. Hence, by (2.3) and the continuous mapping theorem, we obtain $\hat{\mathbb{K}}(s,t) \xrightarrow{P} \mathbb{K}(s,t)$. □

Proof of Theorem 2.1: By Lemma 2.1, we obtain $\hat{\mathbb{K}}(s,t) \xrightarrow{P} \mathbb{K}(s,t)$ uniformly over $[0,1]^2$. Hence

$$\lim_{n\to\infty} \operatorname{tr}(\hat{\mathbb{K}}) = \int_0^1 \lim_{n\to\infty} \hat{\mathbb{K}}(t,t)\, dt = \int_0^1 \mathbb{K}(t,t)\, dt = \operatorname{tr}(\mathbb{K}),$$
$$\lim_{n\to\infty} \operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2}) = \int_0^1 \int_0^1 \lim_{n\to\infty} \hat{\mathbb{K}}^2(s,t)\, dsdt = \int_0^1 \int_0^1 \mathbb{K}^2(s,t)\, dsdt = \operatorname{tr}(\mathbb{K}^{\otimes 2}).$$

Therefore, by (2.5) and (2.7) and the continuous mapping theorem, we conclude that

$$\hat{\beta} = \frac{\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2})}{\operatorname{tr}(\hat{\mathbb{K}})} \xrightarrow{P} \frac{\operatorname{tr}(\mathbb{K}^{\otimes 2})}{\operatorname{tr}(\mathbb{K})} = \beta, \quad \hat{d} = \frac{\operatorname{tr}^2(\hat{\mathbb{K}})}{\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2})} \xrightarrow{P} \frac{\operatorname{tr}^2(\mathbb{K})}{\operatorname{tr}(\mathbb{K}^{\otimes 2})} = d$$

and $\hat{\mathcal{C}}_{n,\alpha} = \hat{\beta}\chi^2_{\hat{d},\alpha} \xrightarrow{P} \beta\chi^2_{d,\alpha}$, as $n \to \infty$. The theorem is proved. □

Proof of Theorem 3.1: Under the local alternatives $H_{1n}^{(1)}$, we have

(1.1)
$$\begin{aligned}\mathcal{C}_n &= n\int_0^1 \left((\bar{X}(t) - m(t)) - (\bar{X}(t+1) - m(t+1)) + (m(t) - m(t+1))\right)^2 dt \\ &= \int_0^1 \left(n^{1/2}(\bar{X}(t) - m(t)) - n^{1/2}(\bar{X}(t+1) - m(t+1)) + n^{(1-\tau)/2} d(t)\right)^2 dt.\end{aligned}$$

Under gaussianity assumption, Theorem 4.14 of Zhang (2013, p. 109) implies $n^{1/2}(\bar{X}(t) - m(t)), t \in [0,2]$ is a Gaussian process with mean zero and covariance function $\mathbb{C}(s,t)$. Hence, the processes $n^{1/2}(\bar{X}(t) - m(t))$ and $n^{1/2}(\bar{X}(t+1) - m(t+1))$ for $t \in [0,1]$ are also Gaussian processes with such parameters.

Thus, $n^{1/2}(\bar{X}(t)-m(t))-n^{1/2}(\bar{X}(t+1)-m(t+1))+n^{(1-\tau)/2}d(t)$ is a Gaussian process with mean $n^{(1-\tau)/2}d(t)$ and covariance function $\mathbb{K}(s,t)$ given by (2.1). By the assumption of $d\in L^2([0,1])$ and since $\mathrm{tr}(\mathbb{K})$ is finite as noted in Section 2, from Theorem 4.2 in Zhang (2013, p. 86), it follows that $\mathcal{C}_n$ has the same distribution as $\sum_{r=1}^{l}\lambda_r A_r+n^{1-\tau}\sum_{r=l+1}^{\infty}\Delta_r^2$, where $A_r\sim\chi_1^2(n^{1-\tau}\lambda_r^{-1}\Delta_r^2)$ are independent, λ_r are the decreasing-ordered eigenvalues of $\mathbb{K}(s,t)$, $\Delta_r\overset{\mathrm{def}}{=}\int_0^1 d(t)\phi_r(t)dt$, $\phi_r(t)$ are the associated eigenfunctions of $\mathbb{K}(s,t)$, $r=1,2,...$, and l is the number of all positive eigenvalues. The possibility of $l=\infty$ is permitted. Using above observation and since $\sum_{r=1}^{\infty}\Delta_r^2=\|d\|^2$, we calculate the expected value and variance of the test statistic as follows

$$\begin{aligned}
E(\mathcal{C}_n)&=\sum_{r=1}^{l}\lambda_r E(A_r)+n^{1-\tau}\sum_{r=l+1}^{\infty}\Delta_r^2\\
&=\sum_{r=1}^{l}\lambda_r(1+n^{1-\tau}\lambda_r^{-1}\Delta_r^2)+n^{1-\tau}\sum_{r=l+1}^{\infty}\Delta_r^2\\
&=\sum_{r=1}^{l}\lambda_r+n^{1-\tau}\|d\|^2=\mathrm{tr}(\mathbb{K})+n^{1-\tau}\|d\|^2,\\
Var(\mathcal{C}_n)&=\sum_{r=1}^{l}\lambda_r^2 Var(A_r)=2\sum_{r=1}^{l}\lambda_r^2(1+2n^{1-\tau}\lambda_r^{-1}\Delta_r^2)=2\sum_{r=1}^{l}\lambda_r^2+4n^{1-\tau}\Delta_\lambda^2\\
&=2\mathrm{tr}(\mathbb{K}^{\otimes 2})+4n^{1-\tau}\Delta_\lambda^2,
\end{aligned}$$

where $\Delta_\lambda^2\overset{\mathrm{def}}{=}\sum_{r=1}^{l}\lambda_r\Delta_r^2$. The rest of the proof is divided into two cases.

Case 1. Let $\Delta_r=0$ for all $r=1,...,l$. Then, $\mathcal{C}_n$ has the same distribution as

$$\sum_{r=1}^{l}\lambda_r A_r+n^{1-\tau}\sum_{r=1}^{\infty}\Delta_r^2=\sum_{r=1}^{l}\lambda_r A_r+n^{1-\tau}\|d\|^2,$$

where $A_r\sim\chi_1^2$. Hence, the distributions of $\mathcal{C}_n$ and $\mathcal{C}_0^*+n^{1-\tau}\|d\|^2$ are the same, where $\mathcal{C}_0^*$ is given in (2.2). Theorem 2.1 implies the asymptotic power of the BT test is of the form $P(\mathcal{C}_n>\hat{\mathcal{C}}_{n,\alpha})=P(\mathcal{C}_0^*>\mathcal{C}_{0,\alpha}-n^{1-\tau}\|d\|^2)+o(1)$, and it is easy to see that this power tends to 1, as $n\to\infty$.

Case 2. Let $\Delta_r\neq 0$ for some $r\in\{1,...,l\}$. Since $A_r\sim\chi_1^2(n^{1-\tau}\lambda_r^{-1}\Delta_r^2)$, it has the same distribution as $(Y_r+n^{(1-\tau)/2}\lambda_r^{-1/2}\Delta_r)^2$, where $Y_r\sim N(0,1)$. Thus, the distribution of $\mathcal{C}_n$ is as that of $\sum_{r=1}^{l}\lambda_r Y_r^2+2n^{(1-\tau)/2}\Delta_\lambda Y+n^{1-\tau}\sum_{r=1}^{\infty}\Delta_r^2$, where $Y\overset{\mathrm{def}}{=}\sum_{r=1}^{l}\lambda_r^{1/2}\Delta_r Y_r/\Delta_\lambda\sim N(0,1)$. Therefore, $(\mathcal{C}_n-E(\mathcal{C}_n))/Var(\mathcal{C}_n)$ has the same distribution as

$$\frac{\sum_{r=1}^{l}\lambda_r(Y_r^2-1)}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2})+4n^{1-\tau}\Delta_\lambda^2}}+\frac{2n^{(1-\tau)/2}\Delta_\lambda Y}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2})+4n^{1-\tau}\Delta_\lambda^2}}.$$

Since $\tau \in [0,1)$, $\mathrm{tr}(\mathbb{K}^{\otimes 2})$ is finite (by the Cauchy–Schwarz inequality) and $0 < \Delta_\lambda^2 < \lambda_1 \sum_{r=1}^{l} \Delta_r^2 \leq \lambda_1 \|d\|^2 < \infty$, we have

$$\frac{\sum_{r=1}^{l} \lambda_r (Y_r^2 - 1)}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2}) + 4n^{1-\tau}\Delta_\lambda^2}} \xrightarrow{p} 0,$$

and

$$\frac{2n^{(1-\tau)/2}\Delta_\lambda Y}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2}) + 4n^{1-\tau}\Delta_\lambda^2}} = \frac{2\Delta_\lambda Y}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2})/n^{1-\tau} + 4\Delta_\lambda^2}} \xrightarrow{d} Y \sim N(0,1),$$

as $n \to \infty$. By Theorem 2.1, we obtain

$$P(\mathcal{C}_n > \hat{\mathcal{C}}_{n,\alpha}) = 1 - \Phi\left(\frac{\mathcal{C}_{0,\alpha} - \mathrm{tr}(\mathbb{K}) - n^{1-\tau}\|d\|^2}{\sqrt{2\mathrm{tr}(\mathbb{K}^{\otimes 2}) + 4n^{1-\tau}\Delta_\lambda^2}}\right) + o(1),$$

where Φ is the cumulative distribution function $N(0,1)$. Hence, $P(\mathcal{C}_n > \hat{\mathcal{C}}_{n,\alpha}) \to 1$, as $n \to \infty$, because $\tau \in [0,1)$ and $\mathcal{C}_{0,\alpha}$, $\mathrm{tr}(\mathbb{K})$, $\mathrm{tr}(\mathbb{K}^{\otimes 2})$ and $\Delta_\lambda^2 > 0$ are finite. □

Proof of Theorem 3.2: Under the local alternatives $H_{1n}^{(2)}$, by (1.1), we have

$$\mathcal{C}_n = \int_0^1 \left(n^{1/2}(\bar{X}(t) - m(t)) - n^{1/2}(\bar{X}(t+1) - m(t+1)) + d(t)\right)^2 dt.$$

Similarly as in the proof of Theorem 1 in Martínez-Camblor and Corral (2011), we obtain $n^{1/2}(\bar{X}(t) - m(t)) - n^{1/2}(\bar{X}(t+1) - m(t+1)) + d(t) \xrightarrow{d} \xi_d(t)$, as $n \to \infty$, where $\xi_d(t), t \in [0,1]$ is a Gaussian process with mean $d(t)$ and covariance function $\mathbb{K}(s,t)$ given by (2.1). Hence, by the continuous mapping theorem, we have $\mathcal{C}_n \xrightarrow{d} \|\xi_d\|^2$, as $n \to \infty$. Since $d \in L^2([0,1])$ and $\mathrm{tr}(\mathbb{K}) < \infty$ (see Section 2), Theorem 4.2 in Zhang (2013, p. 86) shows that $\|\xi_d\|^2$ has the same distribution as $\sum_{r=1}^{l} \lambda_r A_r + \sum_{r=l+1}^{\infty} \delta_r^2$, where $A_r \sim \chi_1^2(\lambda_r^{-1}\delta_r^2)$ are independent, λ_r are the decreasing-ordered eigenvalues of $\mathbb{K}(s,t)$, $\delta_r \stackrel{\mathrm{def}}{=} \int_0^1 d(t)\phi_r(t)dt$, $\phi_r(t)$ are the associated eigenfunctions of $\mathbb{K}(s,t)$, $r = 1, 2, ...$, and l is the number of all positive eigenvalues ($l = \infty$ is possible). Since A_r has the same distribution as $(Y_r + \lambda_r^{-1/2}\delta_r)^2$, $Y_r \sim N(0,1)$, the distribution of $\|\xi_d\|^2$ is the same as that of $\sum_{r=1}^{l} \lambda_r Y_r^2 + 2\delta_\lambda Y + \sum_{r=1}^{\infty} \delta_r^2$, where $\delta_\lambda^2 \stackrel{\mathrm{def}}{=} \sum_{r=1}^{l} \lambda_r \delta_r^2$ and $Y \stackrel{\mathrm{def}}{=} \sum_{r=1}^{l} \lambda_r^{1/2}\delta_r Y_r/\delta_\lambda \sim N(0,1)$. Observing that $\sum_{r=1}^{\infty} \delta_r^2 = \|d\|^2$ and by Theorem 2.1, the asymptotic power of the BT test, as $n \to \infty$, is given by $P(\mathcal{C}_n > \hat{\mathcal{C}}_{n,\alpha}) = P(\mathcal{C}_0^* + 2\delta_\lambda Y + \|d\|^2 > \mathcal{C}_{0,\alpha}) + o(1)$, where $\mathcal{C}_0^*$ and $\mathcal{C}_{0,\alpha}$ are given in (2.2) and Theorem 2.1. The rest of the proof runs as in the proof of Proposition 4 in Zhang and Liang (2014) taking $\delta^2 = \|d\|^2$. □

B. Numerical implementation

As we mentioned in Subsection 4.1, in practice, the n functional observations are not continuously observed. Each function is usually observed on a grid of design time points. In this paper, all individual functions $X_i(t)$ for $t \in [0,1]$ and $t \in [1,2]$ in the simulations and the example (also in the function `BT.test` given in the next section) are assumed to be observe on a common grid of design time points that are equally spaced in $[0,1]$ and in $[1,2]$. To implement the new test when the design time points are different for different individual functions, one first has to reconstruct the functional sample from the observed discrete functional sample using some smoothing technique, then discretize each individual function of the reconstructed functional sample on a common grid of time points, and finally apply the test accordingly (see Zhang, 2013, or Zhang and Liang, 2014, for more details).

Assume that $0 = t_1 \leq t_2 \leq \cdots \leq t_p = 1$ and $1 = t_1 + 1 \leq t_2 + 1 \leq \cdots \leq t_p + 1 = 2$ denote a grid of design time points that are equally spaced in $[0,1]$ and in $[1,2]$, at which the data are observed. Then, we have

$$\begin{aligned}
\mathcal{C}_n &= n \int_0^1 (\bar{X}(t) - \bar{X}(t+1))^2 \, dt \approx \frac{n}{p} \sum_{i=1}^{p} (\bar{X}(t_i) - \bar{X}(t_i+1))^2 = \frac{1}{p}\mathcal{C}_n^0, \\
\operatorname{tr}(\hat{\mathbb{K}}) &= \int_0^1 \hat{\mathbb{K}}(t,t) \, dt \approx \frac{1}{p} \sum_{i=1}^{p} \hat{\mathbb{K}}(t_i,t_i) = \frac{1}{p}\operatorname{trace}(\hat{\mathbf{K}}), \\
\operatorname{tr}(\hat{\mathbb{K}}^{\otimes 2}) &= \int_0^1 \int_0^1 \hat{\mathbb{K}}^2(s,t) \, dsdt \approx \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\mathbb{K}}^2(t_i,t_j) = \frac{1}{p^2}\operatorname{trace}(\hat{\mathbf{K}}^2),
\end{aligned}$$

where $\hat{\mathbf{K}} = (\hat{\mathbb{K}}(t_i,t_j))_{i,j=1}^{p}$. For example, similar approximations have previously been used by Zhang (2013, p. 117), and Zhang and Liang (2014). If the number p is very small, then we can first reconstruct the data as described in the last paragraph and then discretize the reconstructed functions on a greater number of design time points. The estimated parameters $\hat{\beta}$ and $\hat{d}$ in (2.5) are approximately expressed as

$$\hat{\beta} \approx \frac{\operatorname{trace}(\hat{\mathbf{K}}^2)}{p \cdot \operatorname{trace}(\hat{\mathbf{K}})} = \frac{1}{p}\hat{\beta}^0, \quad \hat{d} \approx \frac{\operatorname{trace}^2(\hat{\mathbf{K}})}{\operatorname{trace}(\hat{\mathbf{K}}^2)},$$

and hence the approximation of the p-value given in (2.6) is of the form

$$P(\chi^2_{\hat{d}} > \mathcal{C}_n/\hat{\beta}) \approx P(\chi^2_{\hat{d}} > \mathcal{C}_n^0/\hat{\beta}^0).$$

C. R code

The new test is performed by the R function `BT.test` given below. The notations in the program are consistent with or similar to those used in the paper. The argument `x` is a data frame or matrix of data, whose each row is a discretized version of a function $X_i(t)$, $t \in [0,2]$, $i = 1, ..., n$. It means that the columns of `x` represent the values of the sample functions at the design time points. The number of columns is even, and the first half of them is connected with the design time points in $[0,1]$, and the second half with those in $[1,2]$. As outputs, we obtain value of test statistic and p-value of the test.

```
BT.test = function(x){
 n = nrow(x); p = ncol(x); CC = var(x)
 Cn = n*sum((colMeans(x[, 1:(p/2)]) - colMeans(x[, (p/2+1):p]))^2)
 KK = CC[1:(p/2), 1:(p/2)] - CC[1:(p/2), (p/2+1):p] -
      CC[(p/2+1):p, 1:(p/2)] + CC[(p/2+1):p, (p/2+1):p]
 A = sum(diag(KK)); B = sum(diag(KK%*%KK)); beta = B/A; d = (A^2)/B
 p.value = 1 - pchisq(Cn/beta, d)
 return(c(Cn/(p/2), p.value))
}
```

D. Additional simulations

In this section, we present some additional simulations suggested by one of the reviewers. The simulation models are similar to those in Subsection 4.1, but we consider the functional autoregressive process of order one ($\mathrm{FAR}_f(1)$) instead of compound symmetric dependency structure. The $\mathrm{FAR}_f(1)$ process was considered, for example, by Didericksen *et al.* (2012) or Horváth and Rice (2015). The error functions are generated in the following way:

$$\varepsilon_{ij}(t) = \eta \int_0^1 f(t,u)\varepsilon_{i,j-1}(u)\, du + \xi B_{ij}(t),\ t \in [0,1], i = 1, ..., n, j = 1, 2, \tag{4.1}$$

where f is a kernel, $\eta = 0.005, 0.125$ and $\xi = 0.05, 0.5$ for models M0–M3 and M4–M7, respectively, and B_{ij} are independent standard Brownian Bridges. If $\|f\| < 1$, then (4.1) has a unique stationary and ergodic solution (see Bosq, 2000, $\eta = \xi = 1$). We consider $f(t,u) = c\exp((t^2+u^2)/2)$, where $c = 0.3416$ so that $\|f\| \approx 0.5$ (see Horváth and Rice, 2015). To obtain ε_{i0}, we use $\varepsilon_{i,-2} = \xi B_i$, where B_i is a standard Brownian Bridge, and then $\varepsilon_{i,-1}$ and ε_{i0} are generated according to (4.1). The errors functions ε_{ij} are adequately centered. The results are given in Table 8. They are very satisfactory. The conclusions are similar to those obtained in Subsection 4.2.

Table 8: Empirical sizes and powers (as percentages) of the A, B, P and BT tests obtained in $\mathrm{FAR}_f(1)$ case. The column "M" refers to different models.

M	I	26				101				251			
	n	A	B	P	BT	A	B	P	BT	A	B	P	BT
M0	25	4.7	4.6	4.6	4.6	5.5	5.3	5.3	5.4	4.4	4.4	4.5	4.4
	35	4.4	4.9	4.5	4.6	5.4	5.6	5.4	5.1	5.3	5.3	5.3	4.9
	50	4.9	4.5	5.3	4.8	4.8	4.8	4.3	4.9	6.0	5.2	5.5	5.6
M1	25	99.3	99.2	99.7	99.3	99.5	99.5	99.7	99.6	99.6	99.3	99.7	99.1
	35	100	100	100	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100
M2	25	100	100	100	100	100	100	100	100	100	100	100	100
	35	100	100	100	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100
M3	25	69.2	68.5	74.5	69.7	97.9	98.2	99.0	98.0	99.2	99.5	99.4	99.4
	35	94.9	94.7	96.5	95.3	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100
M4	25	5.2	5.5	4.9	5.1	4.3	4.4	4.1	4.5	6.1	6.4	5.7	5.9
	35	**6.6**	**6.7**	6.2	**6.6**	5.2	5.6	5.3	5.3	4.8	4.8	4.2	4.8
	50	5.5	5.4	4.9	5.5	5.9	5.5	5.9	5.7	5.4	5.3	4.8	5.1
M5	25	79.4	79.8	80.3	79.0	84.6	84.1	84.8	84.0	83.9	84.4	84.2	83.5
	35	96.1	95.9	96.6	95.8	97.0	97.0	97.5	97.3	96.9	96.8	97.5	97.0
	50	100	100	100	100	100	99.9	100	100	100	99.9	100	100
M6	25	30.1	29.7	29.4	29.2	29.7	30.1	29.9	29.4	31.8	32.0	31.6	30.9
	35	41.5	41.2	42.2	41.3	43.0	43.6	42.9	42.2	44.7	43.7	44.6	43.4
	50	58.7	57.7	58.6	57.4	64.8	65.2	65.9	63.9	65.9	65.8	66.7	65.5
M7	25	100	100	100	100	100	100	100	100	100	100	100	100
	35	100	100	100	100	100	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100

ACKNOWLEDGMENTS

The author is thankful to the referees for their constructive suggestions which improved the presentation of this article.

REFERENCES

[1] Abramovich, F.; Antoniadis, A.; Sapatinas, T. and Vidakovic, B. (2004). Optimal testing in a fixed-effects functional analysis of variance model, *International Journal of Wavelets, Multiresolution, Information Processing*, **2**, 323–349.

[2] Attouch, M.K. and Belabed, F.Z. (2014). The k nearest neighbors estimation of the conditional hazard function for functional data, *REVSTAT — Statistical Journal*, **12**, 273–297.

[3] Berrendero, J.R.; Justel, A. and Svarc, M. (2011). Principal components for multivariate functional data, *Computational Statistics & Data Analysis*, **55**, 2619–2634.

[4] Boente, G.; Barrera, M.S. and Tyler, D.E. (2014). A characterization of elliptical distributions, some optimality properties of principal components for functional data, *Journal of Multivariate Analysis*, **131**, 254–264.

[5] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components, *The Annals of Statistics*, **20**, 655–674.

[6] Bosq, D. (2000). *Linear Processes in Function Spaces*, Springer, New York.

[7] Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, *The Annals of Mathematical Statistics*, **25**, 290–302.

[8] Brunner, E.; Dette, H. and Munk, A. (1997). Box-type approximations in nonparametric factorial designs, *Journal of the American Statistical Association*, **92**, 1494–1502.

[9] Chesneau, C.; Kachour, M. and Maillot, B. (2013). Nonparametric estimation for functional data by wavelet thresholding, *REVSTAT — Statistical Journal*, **11**, 211–230.

[10] Cuesta-Albertos, J.A. and Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data, *Test*, **19**, 537–557.

[11] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data, *Journal of Statistical Planning and Inference*, **147**, 1–23.

[12] Cuevas, A.; Febrero, M. and Fraiman, R. (2004). An ANOVA test for functional data, *Computational Statistics & Data Analysis*, **47**, 111–122.

[13] Cuevas, A.; Febrero, M. and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data, *Computational Statistics & Data Analysis*, **51**, 1063–1074.

[14] Cuevas, A.; Febrero, M. and Fraiman, R. (2007). Robust estimation, classification for functional data via projection-based depth notions, *Computational Statistics*, **22**, 481–496.

[15] Didericksen, D.; Kokoszka, P. and Zhang, X. (2012). Empirical properties of forecasts with the functional autoregressive model, *Computational Statistics*, **27**, 285–298.

[16] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory, Practice*, Springer, New York.

[17] Górecki, T.; Krzyśko, M. and Waszak, Ł. (2014). Functional discriminant coordinates, *Communications in Statistics — Theory and Methods*, **43**, 1013–1025.

[18] Górecki, T. and Smaga, Ł. (2015). A comparison of tests for the one-way ANOVA problem for functional data, *Computational Statistics*, **30**, 987–1010.

[19] Górecki, T. and Smaga, Ł. (2017). Multivariate analysis of variance for functional data, *Journal of Applied Statistics*, **44**, 2172–2189.

[20] Gregorutti, B.; Michel, B. and Saint-Pierre, P. (2015). Grouped variable importance with random forests, application to multiple functional data analysis, *Computational Statistics & Data Analysis*, **90**, 15–35.

[21] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, Springer, New York.

[22] Horváth, L. and Rice, G. (2015). Testing equality of means when the observations are from functional time series, *Journal of Time Series Analysis*, **36**, 84–108.

[23] Jacques, J. and Preda, C. (2014). Functional data clustering: a survey, *Advances in Data Analysis and Classification*, **8**, 231–255.

[24] James, G.M. and Hastie, T.J. (2001). Functional linear discriminant analysis for irregularly sampled curves, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **63**, 533–550.

[25] Lian, H. (2012). Empirical likelihood confidence intervals for nonparametric functional data analysis, *Journal of Statistical Planning and Inference*, **142**, 1669–1677.

[26] Martínez-Camblor, P. and Corral, N. (2011). Repeated measures analysis for functional data, *Computational Statistics & Data Analysis*, **55**, 3244–3256.

[27] Prakasa Rao, B.L.S. (2010). Nonparametric density estimation for functional data by delta sequences, *Brazilian Journal of Probability and Statistics*, **24**, 468–478.

[28] Preda, C.; Saporta, G. and Lévéder, C. (2007). PLS classification of functional data, *Computational Statistics*, **22**, 223–235.

[29] Ramsay, J.O.; Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*, Springer.

[30] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis. Methods and Case Studies*, Springer, New York.

[31] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, 2nd edition*, Springer, New York.

[32] R Core Team (2015). R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`

[33] Tokushige, S.; Yadohisa, H. and Inada, K. (2007). Crisp and fuzzy k-means clustering algorithms for multivariate functional data, *Computational Statistics*, **22**, 1–16.

[34] Valderrama, M.J. (2007). An overview to modelling functional data, *Computational Statistics*, **22**, 331–334.

[35] Yamamoto, M. and Terada, Y. (2014). Functional factorial K-means analysis, *Computational Statistics & Data Analysis*, **79**, 133–148.

[36] Zhang, J.T. (2005). Approximate and asymptotic distributions of chi-squared-type mixtures with applications, *Journal of the American Statistical Association*, **100**, 273–285.

[37] Zhang, J.T. (2013). *Analysis of Variance for Functional Data*, Chapman & Hall, London.

[38] Zhang, J.T. and Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test, *Scandinavian Journal of Statistics*, **41**, 51–71.

[39] Zhang, J.T.; Liang, X. and Xiao, S. (2010a). On the two-sample Behrens–Fisher problem for functional data, *Journal of Statistical Theory and Practice*, **4**, 571–587.

[40] Zhang, C.Q.; Peng, H. and Zhang, J.T. (2010b). Two-sample tests for functional data, *Communications in Statistics — Theory and Methods*, **39**, 559–578.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 551–562

PREDICTIVE ESTIMATION OF POPULATION MEAN IN RANKED SET SAMPLING

Authors: Shakeel Ahmed
– Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
sahmed@stat.qau.edu.pk

Javid Shabbir
– Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
javidshabbir@gmail.com

Sat Gupta
– Department of Mathematics and Statistics, University of North Carolina at Greensboro, USA
sngupta@uncg.edu

Received: July 2016 Revised: May 2017 Accepted: August 2017

Abstract:

- The article presents predictive estimation of population mean of the study variable in Ranked Set Sampling (RSS). It is shown that the predictive estimators in RSS using mean per unit estimator, ratio estimator and regression estimator as predictor for non-sampled values are equivalent to the corresponding classical estimators in RSS. On the other hand, when product estimator is used as predictor, the resulting estimator differs from the classical product estimator under RSS. Expressions for the Bias and the Mean Squared Error (MSE) of the proposed estimators are obtained up to first order of approximation. A simulation study is conducted to observe the performance of estimators under predictive approach.

Key-Words:

- *efficiency; product estimator; ratio estimator; regression estimator; RSS.*

1. INTRODUCTION

It is very common to construct estimators for population parameters of a study variable using the information contained only in a sample of the study variable. However, in many situations, statisticians are interested in using some auxiliary information from the population itself which helps in finding more efficient estimators. In literature, a lot of work has been done on how to use the auxiliary information (see for example, Agrwal and Roy (1999), Upadhyaya and Singh (1999), Singh (2003), Singh and Tailor (2003), Kadilar and Cingi (2004, 2006), Yan and Tian (2010), and Singh *et al.* (2014)). In many situations, we may be interested in estimating the average value of the variable being measured for non-sampled units on the basis of sample data at hand. This approach is called predictive method of estimation. This approach is based on superpopulation models, and hence it is also called model-based approach. The approach assumes that the population under consideration is a realization of random variables following a superpopulation model. Under this model the prior information about the population parameters such as the mean, the variance, and other parameters is utilized to predict the non-sampled values of the study variable.

Basu (1971) constructed predictive estimators for population mean using mean per unit estimator, regression estimator, and ratio estimator as predictors for the mean of unobserved units in the population. Srivastava (1983) compared the estimator obtained by using the product estimator as a predictor for mean of unobserved units in the population with the customary product estimator. Recently, Yadav and Mishra (2015) have established predictive estimators using product estimator as predictor for the mean of unobserved units of the population.

Basic statistical principles play a vital role in making inference about the population of interest. If these principles are violated, even optimal statistical procedures will not allow us to make legitimate statistical inferences about the parameters of interest. Ranked Set Sampling (RSS) technique is a good alternative for Simple Random Sampling (SRS) for obtaining experimental data that are truly representative of the population under investigation. This is true across all of the sciences including agricultural, biological, environmental, engineering, physical, medical, and social sciences. This is because in RSS measurements are likely more regularly spaced than measurements in SRS. The RSS procedure creates stratification of the entire population at the sampling stage, i.e. we are randomly selecting samples from the subpopulations of small, medium and large units without constructing the subpopulation strata in advance. Ranked set sampling method, proposed originally by McIntyre (1952) to estimate mean pasture yields, has recently been modified by many authors to estimate the population parameters. Dell and Clutter (1972) showed that the mean estimator is an unbiased estimator of the population mean under RSS for both perfect as

well as imperfect ranking. Muttlak (1997) suggested median ranked set sampling (MRSS) for estimation of finite population mean. Al-Saleh and Al-Omari (2002) used multistage ranked set sampling (MSRSS) to increase the efficiency of the estimator of the population mean for certain value of the sample size. Jemain and Al-Omari (2006) suggested double quartile ranked set sampling (DQRSS) for estimating the population mean. Many other authors have worked on estimation of parameters in RSS (see Al-Omari and Jaber (2008), Bouza (2002), Al-Nasser (2007), Ohyama *et al.* (1999), and Samawi and Muttlak (1996) among others).

In this study, we propose a predictive estimator, using ratio, product and regression estimators as predictors for non-sampled observations under ranked set sampling scheme. In Section 2, we review the predictive estimators introduced by Basu (1971). Section 3 consists of the proposed estimators and their properties. An efficiency comparison is carried out through simulations in Section 4. Some concluding remarks are given in Section 5.

2. PREDICTIVE ESTIMATORS IN SIMPLE RANDOM SAMPLING

Let $U = \{U_1, U_2, ..., U_N\}$ be a population of size N. Let (y_i, x_i) be the values of the study variable y and the auxiliary variable x on the i-th $(0 \leq i \leq N)$ unit of U.

Let S be the set of all possible samples from U using simple random sampling with replacement (SRSWR). For any given $s \in S$, let $\vartheta(s)$ be the number of distinct units in s and let $\bar{s}$ denote the set of all those units of U which are not in s. Basu (1971) presented population mean as follows:

$$\bar{Y} = \frac{\vartheta(s)}{N}\bar{Y}_s + \frac{N - \vartheta(s)}{N}\bar{Y}_{\bar{s}}, \tag{2.1}$$

where $\bar{Y}_s = \frac{1}{\vartheta(s)}\sum_{i\in s} y_i$ and $\bar{Y}_{\bar{s}} = \frac{1}{N-\vartheta(s)}\sum_{i\in \bar{s}} y_i$. Under simple random sampling with size $\vartheta(s) = n$, the predictor for overall population mean is given by

$$\bar{Y} = \frac{n}{N}\bar{Y}_s + \frac{N - n}{N}\bar{Y}_{\bar{s}}, \tag{2.2}$$

where $\bar{Y}_s = \frac{1}{n}\sum_{i\in s} y_i$ and $\bar{Y}_{\bar{s}} = \frac{1}{N-n}\sum_{i\in \bar{s}} y_i$. An appropriate estimator of the population mean is then given by

$$t = \frac{n}{N}\bar{y}_s + \frac{N - n}{N}T, \tag{2.3}$$

where T is the predictor of $\bar{Y}_{\bar{s}}$. Basu (1971) used the mean per unit estimator $\bar{y} = \frac{1}{n}\sum_{i\in s} y_i$, ratio estimator $\bar{y}_r = \frac{\bar{y}_s}{\bar{x}_s}\bar{X}_{\bar{s}}$, product estimator $\bar{y}_p = \frac{\bar{y}_s}{\bar{X}_{\bar{s}}}\bar{x}_s$ and regression estimator $\bar{y}_{lr} = \bar{y}_s + \beta(\bar{X}_{\bar{s}} - \bar{x}_s)$ as predictors. Here, $\bar{X}_{\bar{s}} = \frac{1}{N-n}\sum_{i\in \bar{s}} x_i =$

$\frac{N\bar{X}-n\bar{x}_s}{N-n}$ and $\beta = {S_{yx}}/{S_x^2}$, where β is regression coefficient of Y on X, and $\bar{X}$ is the population mean of the auxiliary variable based on N units both are assumed to be known in advance. Also, let $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ and $S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})$.

It has been shown by Basu (1971) that while using simple mean per unit estimator, ratio estimator and regression estimator as T, the predictive estimator t becomes the corresponding classical simple mean estimator $\bar{y}$, ratio estimator $\bar{y}_r$ and regression estimator $\bar{y}_{lr}$respectively. However, when product estimator is used, then t becomes

$$t_p = \bar{y}_s \frac{n\bar{X} + (N-2n)\bar{x}_s}{N\bar{X} - n\bar{x}_s}. \tag{2.4}$$

It can be easily noticed that t_p is quite different from the usual product estimator.

The Bias and Mean Squared Error (MSE) of t with ratio and product estimators as predictor are given below up to 1^{st} order of approximation:

$$Bias(t_r) \cong \bar{Y}\frac{1}{n}\left(C_x^2 - \rho C_y C_x\right), \tag{2.5}$$

$$Bias(t_p) \cong \bar{Y}\frac{1}{n}\left(\theta C_x^2 + \rho C_y C_x\right) \tag{2.6}$$

and

$$MSE(t_r) \cong \bar{Y}^2\frac{1}{n}\left(C_y^2 + C_x^2 - 2\rho C_y C_x\right), \tag{2.7}$$

$$MSE(t_p) \cong \bar{Y}^2\frac{1}{n}\left(C_y^2 + C_x^2 + 2\rho C_y C_x\right), \tag{2.8}$$

where $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$, $\rho = \frac{S_{yx}}{S_y S_y}$, $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ and $\theta = \frac{n}{N-n}$. Also the bias and *MSE* of t_p are given by

$$Bias(\bar{y}_p) \cong \bar{Y}\frac{1}{n}\left(\rho C_y C_x\right) \tag{2.9}$$

and

$$MSE(\bar{y}_p) \cong \bar{Y}^2\frac{1}{n}\left(C_y^2 + C_x^2 + 2\rho C_y C_x\right). \tag{2.10}$$

From Equations (2.8) and (2.10), it is clear that $\bar{y}_p$ and t_p have same *MSE* when first order of approximation is used although they are different estimators. The variance of t_{lr} is given by

$$Var(t_{lr}) = \frac{1}{n}S_y^2\left(1 - \rho^2\right). \tag{2.11}$$

3. PREDICTIVE ESTIMATOR IN RANKED SET SAMPLING

To obtain a Ranked Set Sample from a superpopulation consisting of N units, an initial sample of m units is selected and ranked according to the attribute of interest. A variety of mechanisms are used for ranking purpose, i.e. visual inspection of units, expert opinion, or through the use of some concomitant variables. If ranking is performed on the auxiliary variable X, the unit that is judged to be the smallest ranked unit from the selected sample is called the first judgment order statistic and is denoted by $Y_{[1]}$. On the other hand, when ranking is performed on the study variable Y itself, the smallest ranked unit (called smallest order statistic) is selected from the sample and denoted by $Y_{(1)}$. Then a second sample of size m (independent of the first sample) is selected from the population and is ranked in the same manner as the first. From the second sample, we select the unit ranked as the second smallest in the sample (i.e. the second judgment order statistic) and is denoted by $Y_{[2]}$ or $Y(2)$ according to the above mentioned definitions. This process continues till inclusion of the largest ranked unit from the m-th sample selected for judgment. This entire process results into m observations and is called a cycle. We complete r cycles to obtain a ranked set sample of size $n = rm$ units.

Let Ω be the all possible samples of size $n = rm$ can be taken from a superpopulation U using a ranked set sampling scheme. Suppose that ω be a single set Ω having size $n = rm$. Let $\bar{\omega}$ denote the set of all those units of U which are not in ω. Let $y_{i[i]}$ and $x_{i(i)}$ be the values of the study variable Y and the auxiliary variable X for i-th unit taken from the i-th judgment ranked sample for actual quantification, where $i = 1, 2, ..., m$. It is assumed that ranking is performed with respect to the auxiliary variable X.

For a ranked set sample of size $n = rm$ (for simplicity, we use $r = 1$), we obtain the following estimators

$$t_{rss[j]} = \frac{m}{N}\bar{y}_{rss} + \frac{N-m}{N}T_{[j]}, (j = 1, 2, 3, 4), \tag{3.1}$$

where $\bar{y}_{rss} = \frac{1}{m}\sum_{i\in\omega} y_{i[i]}$ and $T_{[j]}$ is the predictor for mean of non-sampled observations $(\bar{Y}_{\bar{\omega}})$ which is defined by $T_{[1]} = \bar{y}_{rss}$, $T_{[2]} = \bar{y}_{rss(r)}$, $T_{[3]} = \bar{y}_{rss(lr)}$ and $T_{[4]} = \bar{y}_{rss(p)}$, where $\bar{y}_{rss[r]} = \bar{y}_{rss}\frac{\bar{X}_{\varpi}}{\bar{x}_{rss}}$, $\bar{y}_{rss[lr]} = \bar{y}_{rss} + \beta\left(\bar{X}_{\varpi} - \bar{x}_{rss}\right)$ and $\bar{y}_{rss[p]} = \bar{y}_{rss}\frac{\bar{x}_{rss}}{\bar{X}_{\varpi}}$. Here, $\bar{X}_{\varpi} = \frac{1}{N-m}\sum_{i\in\varpi} x_{i(i)} = \frac{N\bar{X}-m\bar{x}_{rss}}{N-m}$, and $\bar{x}_{rss} = \frac{1}{m}\sum_{i\in\omega} x_{i(i)}$. Inserting $T_{[j]}$ for $(j = 1, 2, 3, 4)$ in Equation (3.1), we have

$$t_{rss[1]} = \bar{y}_{rss}, \tag{3.2}$$

$$t_{rss[2]} = \bar{y}_{rss}\frac{\bar{X}}{\bar{x}_{rss}}, \tag{3.3}$$

$$t_{rss[3]} = \bar{y}_{rss} + \beta(\bar{x}_{rss} - \bar{X}), \tag{3.4}$$

and

$$t_{rss[4]} = \bar{y}_{rss}\frac{m\bar{X} - (N-2m)\bar{x}_{rss}}{N\bar{X} - m\bar{x}_{rss}}. \tag{3.5}$$

Equations (3.2), (3.3) and (3.4) show that $t_{rss[1]}$, $t_{rss[2]}$ and $t_{rss[3]}$ are equivalent to $\bar{y}_{rss}$, $\bar{y}_{rss[r]}$ and $\bar{y}_{rss[lr]}$ respectively. On the other hand $t_{rss[4]}$ differs from $\bar{y}_{rss[p]}$ (usual product estimator under RSS).

To obtain the Bias and the MSE of proposed predictive estimators, we consider the following error terms

$$\in_0 = \frac{\bar{y}_{rss}}{\bar{Y}} - 1 \quad \text{and} \quad \in_1 = \frac{\bar{x}_{rss}}{\bar{X}} - 1$$

such that $E(\in_0) = E(\in_1) = 0$ and

$$E(\in_0^2) = \bar{Y}^{-2}\left(\frac{S_y^2}{m} - \frac{1}{m^2}\sum_{i=1}^{m}\delta_{y[i]}^2\right),$$
$$E(\in_1^2) = \bar{X}^{-2}\left(\frac{S_x^2}{m} - \frac{1}{m^2}\sum_{i=1}^{m}\delta_{x(i)}^2\right)$$

and

$$E(\in_0\in_1) = \bar{Y}^{-1}\bar{X}^{-1}\left(\frac{S_{yx}}{m} - \frac{1}{m^2}\sum_{i=1}^{m}\delta_{y[i]}\delta_{x(i)}\right),$$

where $\delta_{y[i]} = \bar{Y}_{[i]} - \bar{Y}$ and $\delta_{x(i)} = \bar{X}_{(i)} - \bar{X}$ for $i = 1, 2, ..., m$. Here, $\bar{Y}_{[i]}$ and $\bar{X}_{(i)}$ are population means of the study variable and the auxiliary variable respectively for i-th order statistic. It is easy to show that $t_{rss[1]}$ is an unbiased estimator of the population mean $\bar{Y}$ with

$$Var(t_{rss[1]}) = \frac{S_y^2}{m} - \frac{1}{m^2}\sum_{i=1}^{m}\delta_{y[i]}^2. \tag{3.6}$$

It is clear that $Var(t_{rss[1]}) \leq \frac{S_y^2}{m}$. This indicates that $t_{rss[1]}$ is more efficient than $\bar{y}_s$(sample mean under SRSWR). Similarly, the bias and the *MSE* of $t_{rss[2]}$, up to first order of approximation, are given by

$$Bias(\bar{y}_{rss[2]}) \cong \frac{\bar{Y}}{m}\left[\left(C_x^2 - \rho C_y C_x\right) - \frac{1}{m}\left(\sum_{i=1}^{m}W_{x(i)}^2 - \sum_{i=1}^{m}W_{y[i]}W_{x(i)}\right)\right] \tag{3.7}$$

and

$$MSE(t_{rss[2]}) \cong \frac{1}{m}\left(S_y^2 + R^2S_x^2 - 2R\rho S_yS_x\right) - \frac{1}{m^2}\sum_{i=1}^{m}\kappa_{[i]}^2, \tag{3.8}$$

where $\kappa_{[i]} = W_{y[i]} - RW_{x(i)}$, $W_{y[i]} = \frac{\delta_{y[i]}}{\bar{Y}}$, $W_{x(i)} = \frac{\delta_{x(i)}}{\bar{X}}$ and $R = \frac{\bar{Y}}{\bar{X}}$. From Equations (2.7) and (3.8), it is obvious that $MSE(t_{rss[2]}) \leq MSE(t_r)$, i.e. $t_{rss[2]}$ is more efficient than the predictive type ratio estimator under SRSWR. Further, we can show that $t_{rss[3]}$ is an unbiased estimator of $\bar{Y}$ with variance

$$Var(t_{rss[3]}) = \frac{S_y^2}{m}\left(1-\rho^2\right) - \frac{1}{m^2}\sum_{i=1}^{m} A_{[i]}^2, \tag{3.9}$$

where $A_{[i]} = W_{y[i]} - \beta W_{x(i)}$, $\forall\, i = 1, 2..., m$. Equation (3.9) shows the superiority of the predictive type regression estimator as compared to its counterpart in SRSWR.

Finally, to compute the Bias and the MSE of $t_{rss[4]}$, note that

$$\begin{aligned} t_{rss[4]} &= \bar{Y}(1+\in_0)\frac{m\bar{X} + (N-2m)\bar{X}(1+\in_1)}{N\bar{X} - m\bar{X}(1+\in_1)}, \\ &= \bar{Y}(1+\in_0)\left(1 + \frac{(N-2m)\in_1}{N-m}\right)\left(1 + \frac{m\in_1}{N-m}\right)^{-1}. \end{aligned}$$

Assuming $\left|\frac{m}{N-m}\right| < 1$, and expanding up to first order of approximation using binomial expansion, we have

$$t_{rss[4]} - \bar{Y} \cong \bar{Y}\left(\in_0 + \in_1 + \in_0\in_1 + \phi\in_1^2\right) \tag{3.10}$$

where $\phi = \frac{m}{N-m}$. Taking expectation of Equation (3.10), we get

$$Bias(t_{rss[i]}) \cong \frac{\bar{Y}}{m}\left[C_{yx} + \phi C_x^2 - \frac{1}{m}\sum_{i=1}^{m}\left(\delta_{yx[i]} + \phi\delta_{x(i)}^2\right)\right]. \tag{3.11}$$

MSE of $t_{rss[4]}$ can be obtained by squaring and taking expectation in Equation (3.10). This gives

$$MSE(t_{rss[4]}) \cong \frac{1}{m}\left(S_y^2 + R^2S_x^2 + 2R\rho S_yS_x\right) - \frac{1}{m^2}\sum_{i=1}^{m} B_{[i]}^2, \tag{3.12}$$

where $B_{[i]} = W_{y[i]} + RW_{x(i)}$ for $i = 1, 2, ..., m$.

From Equations (2.6), (2.8), (3.11) and (3.12) it can be noticed that the expression for bias of $t_{rss[4]}$ is different from that of usual product estimator although they have the same MSE for first order of approximation.

4. SIMULATION STUDY

To compare the efficiencies of the proposed estimators, we conduct a simulation study as follows:

1. Generate a hypothetical population on two variables X and Y, where X is generated using three different distributions with some specific values of parameters as described in first row of Table 1.
2. Then Y is generated as $Y = \rho \times X + e$, where e is generated using a standard normal distribution and ρ is the correlation coefficient between X and Y which is fixed at 0.5, 0.7 and 0.9.
3. Take an RSS and a SRSWR, each having size $n = rm$, and compute the proposed estimators and corresponding estimators in SRSWR, where $r = 5, 10$ and $m = 2, 4, 6$.
4. Repeat Step 2, 10,000 times. Then compute the mean squared error of each estimator to obtain relative efficiency of the proposed estimators.

Table 1 provides relative efficiency of proposed predictive estimators in RSS with respect to simple mean estimator in SRS, i.e.

$$RE[j] = \frac{Var(\bar{y}_s)}{MSE(t_{rss[j]})} \quad \text{for} \quad j = 1, 2, 3, 4.$$

Table 1 shows that the relative efficiencies of the RSS increases with the increase of the correlation between the auxiliary variable and the study variable. *RE* also increases with the increase of the set size m. Predictive estimator using ratio estimator and regression estimator as predictors are almost equally efficient for all the case that considered in this study. However, the product estimator gives worse performance as the correlation between the study variable and the auxiliary variable increases. But this because product estimator is not preferable for prediction in ranked set sampling, when ranking is performed based on an auxiliary variable that has positive correlation with the variable of interest. Efficiencies of the proposed estimators are significantly higher when uniform distribution is used to generate data in the interval [0, 10]. Efficiency is at its peak for uniform distribution with high positive correlation between the study variable and the auxiliary variable.

Table 1: Efficiency Comparison.

	r	m	Normal(5,1)				Exponential(1)				Uniform(0,10)			
			$RE(1)$	$RE(2)$	$RE(3)$	$RE(4)$	$RE(1)$	$RE(2)$	$RE(3)$	$RE(4)$	$RE(1)$	$RE(2)$	$RE(3)$	$RE(4)$
$\rho = 0.5$	5	2	1.1035	1.315	1.3171	0.7292	1.0758	1.0675	1.297	0.6952	1.3359	3.366	3.5067	0.4542
		4	1.1208	1.2557	1.2581	0.8375	1.1616	1.1833	1.3128	0.8434	1.7646	3.5969	3.6251	0.6835
		6	1.223	1.3371	1.3383	0.97	1.1961	1.1827	1.3148	0.9336	2.0576	3.5757	3.5902	0.8916
	10	2	1.0587	1.2869	1.2897	0.691	1.0842	1.0733	1.3092	0.6919	1.3304	3.485	3.6089	0.4504
		4	1.1147	1.2583	1.2598	0.8299	1.1925	1.1805	1.3412	0.8673	1.7856	3.5659	3.6008	0.7013
		6	1.1966	1.3087	1.3097	0.9537	1.1762	1.2036	1.289	0.918	2.048	3.5403	3.5553	0.8895
$\rho = 0.7$	5	2	1.2067	1.9023	1.9053	0.5654	1.1752	1.5186	1.8926	0.5362	1.4461	8.1124	8.4514	0.4061
		4	1.3263	1.8048	1.8084	0.7308	1.3749	1.6819	1.9129	0.7272	2.1256	8.6801	8.7483	0.6425
		6	1.5132	1.9278	1.9295	0.9165	1.4852	1.6825	1.9196	0.8758	2.699	8.5921	8.6269	0.8745
	10	2	1.1361	1.8449	1.849	0.5286	1.1812	1.5155	1.9115	0.5315	1.4348	8.3593	8.6565	0.4023
		4	1.3311	1.8299	1.8321	0.7327	1.4121	1.6694	1.9515	0.7478	2.1744	8.6288	8.7132	0.6629
		6	1.4946	1.8999	1.9013	0.914	1.4611	1.7073	1.8795	0.8616	2.6987	8.5361	8.5724	0.8752
$\rho = 0.9$	5	2	1.3744	4.9954	5.0033	0.4283	1.3408	4.0352*	5.0519*	0.4086	1.5073	33.3303**	34.723**	0.3841
		4	1.7944	4.7412	4.7506	0.6223	1.8154	4.4429*	5.0647*	0.6081	2.3887	35.7477**	36.0284**	0.6244
		6	2.2503	5.0197*	5.0242*	0.8455	2.1969	4.4629*	5.0952*	0.804	3.2381	35.1926**	35.335**	0.8656
	10	2	1.2823	4.8347*	4.8455*	0.4006	1.3338	3.9467*	5.0819*	0.4022	1.4964	34.1889**	35.4045**	0.3803
		4	1.8255	4.8757*	4.8816*	0.6354	1.8519	4.3654*	5.1294*	0.6201	2.4573	35.5361**	35.8836**	0.6447
		6	2.2714	5.0086*	5.0122*	0.8629	2.1789	4.5009*	5.0205*	0.7967	3.2513	35.1079**	35.257**	0.8692

* Stands for higher relative efficiency;
** Stands for highest relative efficiency.

5. CONCLUSION

Assuming a superpopulation model, we developed some predictive type estimators in ranked set sampling as RSS is more efficient method of sample selection for actual measurements. Properties (bias and efficiency) are examined up to first order of approximation. It is observed that the predictive estimators are equivalent to the corresponding classical estimators in RSS when simple mean estimator, ratio estimator and regression estimator are used as predictors for non-sampled values. On the other hand, predictive estimator has different form as compared to the corresponding classical product estimator when product estimator is used as predictor.

This study can be extended by using exponential type estimators and some other efficient estimators as predictor

ACKNOWLEDGMENTS

The authors would like to express their most sincere gratitude to the editor and the referee for their careful reading of the paper and for making many useful suggestions, all of which helped substantially improve the presentation of the paper.

REFERENCES

[1] Agrwal, M. and Roy, D. (1980). Efficient estimators of population variance with regression-type and ratio-type predictor-inputs, *Metron*, **57**(3), 4–13.

[2] Rothwell, D.P. (2007). L-ranked set sampling: a generalization procedure for robust visual sampling, *Communication in Statistics — Simulation and Computation*, **36**(1), 33–44.

[3] Al-Omari, A. and Jaber, K. (2008). Percentile double ranked set sampling, *Journal of Mathematics and Statistics*, **44**, 903–915.

[4] Al-Saleh, M. and Al-Omari, A. (2002). Multistage ranked set sampling, *Journal of Statistical Planning and Inference*, **102**(1), 273–286.

[5] Basu, D. (1971). *An essay on the logical foundations of survey sampling, Part I (with discussion).* In "Foundation of Statistical Inference" (V.P. Godambe and D.A. Sprott, Eds.), Toronto, Holt Rinehart and Winston, 203–243.

[6] Bouza, C. (2002). Ranked set subsampling the non-response strata for estimating the difference of means, *Biometrical Journal*, **1**, 203–243.

[7] Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics background, Biometrika, **28**, 545–555.

[8] Jemain, A. and Al-Omari, A. (2006). Double quartile ranked set samples, Pakistan Journal of Statistics, **22**(3), 203–243.

[9] Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling, Applied Mathematics and Computation, **151**, 893–902.

[10] Kadilar, C. and Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient, Hacettepe Journal of Mathematics and Statistics, **35**(1), 103–109.

[11] McIntyre, G. (1952). A method for unbiased selective sampling using ranked sets, Crop and Pasture Science, **3**, 385–390.

[12] Muttlak, H. (1997). Median ranked set sampling, Journal of Applied Statistical Sciences, **6**(4), 577–586.

[13] Ohyama, T.; Doi, J. and Yanagawa, T. (1999). Estimating population characteristics by incorporating prior values in stratified random sampling/ranked set sampling, Journal of Statistical Planning and Inference, **138**, 4021–4032.

[14] Samawi, H.M. and Muttlak, H.A. (1996). Estimation of ratio using rank set sampling, *Biometrical Journal*, **38**, 753–764.

[15] Singh, G. (2003). On the improvement of product method of estimation in sample surveys, *JISAS*, **56**, 267–275.

[16] Singh, H. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population means, *Statistics in Transition*, **6**(4), 555–560.

[17] Singh, H.; Tailor, R.; Tailor, R. and Kakran, M. (2014). An improved estimator of population mean using power transformation, *Journal of the Indian Society of Agricultural Statistics*, **58**(2), 223–230.

[18] Srivastava, S. (1983). Predictive estimation of finite population mean using product estimator, *Metrika*, **30**, 93–99.

[19] Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*, **41**, 627–636.

[20] Yadav, S.K. and Mishra, S.S. (2015). Developing improved predictive estimator for finite population mean using auxiliary information, *Statistica*, **95**(1), 76–85.

[21] Yan, Z. and Tian, B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable, *ICICA, PART II, CCIS*, **106**, 103–110.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 563–597

A REVIEW OF THE BEHRENS–FISHER PROBLEM AND SOME OF ITS ANALOGS: DOES THE SAME SIZE FIT ALL?

Authors: Sudhir Paul
– Department of Mathematics and Statistics, University of Windsor, Ontario, Canada
smjp@uwindsor.ca

You-Gan Wang
– School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia
you-gan.wang@qut.edu.au

Insha Ullah
– School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia
insha.ullah@qut.edu.au

Received: April 2017 Revised: August 2017 Accepted: September 2017

Abstract:

- The traditional Behrens–Fisher (B-F) problem is to test the equality of the means μ_1 and μ_2 of two normal populations using two independent samples, when the quotient of the population variances is unknown. Welch [43] developed a frequentist approximate solution using a fractional number of degrees of freedom t-distribution. We make a a comprehensive review of the existing procedures, propose new procedures, evaluate these for size and power, and make recommendation for the B-F and its analogous problems for non-normal populations. On the other hand, we investigate and answer a question: does the same size fit all all, i.e. is the t-test with Welch's degree of freedom correction robust enough for the B-F problem analogs, and what sample size is appropriate to use a normal approximation to the Welch statistic.

Key-Words:

- *the Behrens–Fisher problem; the beta-binomial model; the negative binomial model; the Weibull model.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

The traditional Behrens–Fisher (B-F) [5, 20] problem is to test the equality of the means μ_1 and μ_2 of two independent normal populations where the variances σ_1^2 and σ_2^2 are unknown and unspecified. The problem arises when the ratio of the population variances is unknown as well. In the case of known Importance of this problem is well understood and its application is widespread [1, 12, 14, 15, 16].

Ever since the solution of this problem by [43], many papers have been written. See, for example, [7], [19], and [30]. These and similar other papers [9, 39] have attempted improvement, in terms of level and power, over the Welch procedure. More recently, non-parametric [14, 16, 21] and Bayesian [24, 46] procedures have also been developed.

However, independent samples from two two-parameter populations (other than the normal) arise in many situations. The problem then is to test the equality of two location (or some analogous) parameters when the dispersion (or some analogous) parameters are unknown and possibly different. These problems are analogous to the traditional Behrens–Fisher problem. Prior to 2014 not much have been written on the solution of the Behrens–Fisher analogous problems. Some (to our knowledge) problems analogous to the B-F problem that have been dealt with recently are

(i) testing equality of two negative binomial means in presence of unequal dispersion parameters [31];

(ii) testing equality of scale parameters of two Weibull distributions in the presence of unequal shape parameters [2], and

(iii) testing equality of two beta binomial proportions in the presence of unequal dispersion parameters [3].

When the sample sizes are small the two sample t-test (T_1) with Welch's [43] degree of freedom and for large sample sizes ($N = n_1 + n_2 > 30$) the standard normal statistic (T_N) (see, Section 2) are recommended by standard text books [23]. Many evidences have been shown in favour of the preference of the Welch T_1 over other procedures. See, for example, [7, 12, 30] for the standard BF problem. More recently [39] developed a jackknife based procedure and [9] developed a computationally intensive procedure for the BF problem. However, no systematic study has been conducted so far to determine the overall sample size required under which the normal approximation of the statistic T_N works.

The primary purpose of this paper is to make a comprehensive review of the existing procedures, evaluate these for size and power, and make recommendations for the standard BF and its analogous problems in some sense.

For the standard BF and some of its analogous problems we also investigate performance of a new Monte-Carlo approach, the bootstrap and the rank counterparts. A recent study [31] suggests that the Welch T_1 does well in some non-normal situations, such as for samples from two negative binomial populations. Along with some other procedures performances of the Welch T_1 and the new Monte-Carlo approach are investigated for samples from normal, two discrete models (count data and data in the form of proportions) and a survival model for a wide range of parameter spaces to reflect comparison of the means for variances which are same to very different.

The secondary purpose is to investigate and answer a question: does the same size fit all or in other words is the t-test with Welch's [43] degree of freedom correction robust enough for the BF problem analogs and what sample sizes are appropriate for the normal approximation of the statistic T_N.

Review, possible new procedures, simulations, and recommendations for the standard BF problem are given in Section 2. The BF analogues corresponding to the negative binomial, the beta binomial, and the Weibull are dealt with in Sections 3, 4 and 5 respectively. The concluding section (Section 6) provides some guide lines as to which procedure(s) to be used in each case. Some recommendations for possible future study are also provided in this section.

2. THE BEHRENS–FISHER PROBLEM: TWO NORMAL POPULATIONS

2.1. Welch's t-Statistic

The well-known Behrens–Fisher (B-H) problem is to test the equality of the means μ_1 and μ_2 of two independent normal populations where the variances σ_1^2 and σ_2^2 are unknown and possibly unequal.

Let Y_{i1}, ..., Y_{in_i} be a random sample from a population, $i = 1, 2$. Now, let y_{i1}, ..., y_{in_i} be a corresponding sample realization with mean $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$ and variance $s_i^2 = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2/(n_i - 1)$. If the samples come from normal populations with means μ_1 and μ_2 and unknown and possibly unequal variances σ_1^2 and σ_2^2, then

$$T_N = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

is asymptotically normally distributed with mean 0 and variance 1 when both n_1 and n_2 are sufficiently large. This is stated in many undergraduate text books in Mathematical Statistics [23].

However, when the sample sizes n_1 and n_2 are smaller the distribution of T_N, henceforth denoted by T_1, is approximately distributed as Student's t with degrees of freedom

$$f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}\right)}$$

[43]. It is shown by [19] and [43] using simulations that the statistic

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{(n_1-1)s_1^2}{(n_1^2-3n_1)} + \frac{(n_2-1)s_2^2}{(n_2^2-3n_2)}}}$$

might be preferable to the statistic T_1 because the former would maintain nominal level better than the later. However, [19] does not provide a degree of freedom for the above Z to be used as an approximation to the t-distribution. To this end [7] derive degrees of freedom and compare performance of T_1 with a few other statistics, such as the Wald, likelihood ratio and score statistics and the statistic Z, in terms of level and power and find that T_1 is still the best. However, there is an error in the degrees of freedom formula which later was corrected by [30]. After carrying out further simulations [30] finds that in addition to all the reasons given by [7] to prefer T_1 over Z, the former shows better power performance than the latter. See, [30] for further details.

To the best of our knowledge, to-date, the statistic T_1 is the best and is referred as the statistic to use in recent text books [23]. In this paper we attempt to do a comprehensive review of all available methods and develop a new Monte Carlo procedure.

2.2. The Likelihood, Score and Wald Tests [7]

The likelihood ratio statistic (LR), score statistic and Wald statistic, denoted by L, S and W, derived by Best and Rayner (1987) are

$$L = n_1 \log[(n_1-1)s_{10}^2/((n_1-1)s_1^2)] + n_2 \log[(n_2-1)s_{20}^2/((n_2-1)s_2^2)],$$

$$S = (\bar{y}_1 - \bar{y}_2)^2/((n_1-1)s_{10}^2/n_1^2 + (n_2-1)s_{20}^2/n_2^2),$$

and

$$W = (\bar{y}_1 - \bar{y}_2)^2/((n_1-1)s_1^2/n_1^2 + (n_2-1)s_2^2/n_2^2),$$

where $s_{i0}^2 = \sum_{j=1}^{n_i}(y_{ij} - \mu_0)^2/(n_i - 1)$ and μ_0 is the solution to the cubic equation

$$\begin{aligned}
&-(n_1+n_2)\mu_0^3 + [(n_1+2n_2)\bar{y_1} + (n_2+2n_1)\bar{y_2}]\mu_0^2 \\
&-[n_1(n_2-1)s_2^2/n_2 + n_2(n_1-1)s_1^2/n_1 + 2(n_1+n_2)\bar{y_1}\bar{y_2} + n_2\bar{y_1}^2 + n_1\bar{y_2}^2]\mu_0 \\
&+[n_1\bar{y_1}\{(n_2-1)s_2^2/n_2 + \bar{y_2}^2\} + n_2\bar{y_2}\{(n_1-1)s_1^2/n_1 + \bar{y_1}^2\}] = 0.
\end{aligned}$$

[32] give a brief description on the construction mechanism as well as the advantages of the $C(\alpha)$ or score tests over the LR and the Wald tests (see, [30] for details).

2.3. A Monte Carlo Procedure developed Using T_1

By examining the T_1-statistic, it is clear that the denominator is a convex combination of $\chi^2_{(n_1-1)}/(n_1-1)$ and $\chi^2_{(n_2-1)}/(n_2-1)$, and the combination proportion depends on the ratio of the two underlying population variances and the sample sizes. The t-distribution approximation becomes exact when $\tau = \sigma_2^2 n_1/\sigma_1^2 n_2 = 1$, and we expect the Monte Carlo method works better when τ is very different from 1. Theoretically, the p-value cannot be calculated under the null unless τ is specified. Under the null, the T_1 statistic follows an exact t distribution with degree of freedom being n_1-1, n_2-1 and (n_1+n_2-2) when τ takes 0, ∞ and 1. The new statistic, henceforth denoted by T, is

$$T = \frac{\frac{\bar{y}_1-\bar{y}_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}}{\sqrt{\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}}} = \frac{\mathcal{N}}{\sqrt{\mathcal{K}}}.$$

Here $\mathcal{N} \sim N(0,1)$. We now study the distribution of $\mathcal{K}$.

$$\begin{aligned}
\mathcal{K} &= \frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \\
&\sim \frac{\frac{\chi^2_{n_1-1}}{n_1-1}\frac{\sigma_1^2}{n_1} + \frac{\chi^2_{(n_2-1)}}{n_2-1}\frac{\sigma_2^2}{n_2}}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \\
&\sim \lambda\kappa_1 + (1-\lambda)\kappa_2,
\end{aligned}$$

where λ is a proportion parameter, $(\sigma_1^2/n_1)/(\sigma_1^2/n_1+\sigma_2^2/n_2)$, $\kappa_1 \sim \chi^2_{n_1-1}/(n_1-1)$, and $\kappa_2 \sim \chi^2_{n_2-1}/(n_2-1)$.

In order to simulate the Monte Carlo numbers from $\mathcal{K}$, we will need to provide a value for λ. Clearly, we can estimate λ by

$$\hat{\lambda} = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}.$$

We therefore obtained an approximate distribution for $\mathcal{K}$,

$$\tilde{\mathcal{K}} \sim \hat{\lambda}\kappa_1 + (1-\hat{\lambda})\kappa_2,$$

whose distribution can be easily obtained. The final distribution, using Monte Carlo procedure, can be approximated by $Z/\sqrt{\tilde{\mathcal{K}}}$ which is obtained by a random

number from $N(0,1)$ and two independent random numbers from $\chi^2_{(n_1-1)}$ and $\chi^2_{(n_2-1)}$. Because κ_1 and κ_2 are independently simulated from $\hat{\lambda}$, we have $\mathrm{E}(\tilde{\mathcal{K}}) = 1$ and $\mathrm{var}(\tilde{\mathcal{K}}) = 2\hat{\lambda}^2/(n_1-1) + 2(1-\hat{\lambda})^2/(n_2-1)$.

If the variance ratio σ_2^2/σ_1^2 is known, the distribution of $\mathcal{K}$ above is known as a mixture of two χ^2 distributions and T (§2.3) becomes pivotal but it is generally not an exact t distribution. However, if the variance ratio is given, one can use the pooled variance estimator and form a t-statistic with $n_1 + n_2 - 2$ degrees of freedom.

If t-distribution is used to approximate T, i.e., $\tilde{\mathcal{K}}$ is approximated by a chi-square distribution, the "best" degree of freedom by matching the variance $(\tilde{\mathcal{K}})$ to $\chi^2_{(d)}/d$ is

$$d = 2/\mathrm{var}(\tilde{\mathcal{K}}) = \frac{(n_1-1)(n_2-1)}{(n_2-1)\hat{\lambda}^2 + (n_1-1)(1-\hat{\lambda})^2},$$

which is exactly the same as Welch's formula!

After developing this procedure we found that [18] also developed the same statistic. Similar idea has also been explored by [4] and [43]. However, they used an exact distribution which is complex to use and showed that the Welch approximation is remarkably accurate, even for small n_1 and n_2, provided that n_1 and n_2 are equal or nearly equal. Singh, Saxena, and Srivastava [39] developed a procedure similar to the one given above and [9] developed another Monte Carlo based procedure "Computational Approach Test" (CAT). Using a simulation study [9] find that the procedure developed by [39] is not as good as it has been claimed [9]. On the other hand the CAT procedure is quite computationally involved. For small sample sizes the CAT is quite conservative. In contrast our method, which is also Monte Carlo, is very easy to use and its performance is much better that that of CAT. This issue will be dealt with in a separate paper.

2.4. A Bootstrap Procedure [13]

A bootstrap test for the Behrens–Fisher problem is developed by [13]. Among the re-sampling methods, the two sample bootstrap test is the one that neither assumes equal variances nor does it require any distributional assumptions and offer a possible solution to the Behrens–Fisher problem [13]. All we need is a suitable test statistic and a null distribution under the hypothesis of equal population means. Manly (1997) recommends to use T_N as a test statistic, where,

$$T_N = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

is asymptotically normally distributed with mean 0 and variance 1 when both n_1 and n_2 are sufficiently large. The null distribution is approximated by the distribution of B values of T_N evaluated at each of the B bootstrap samples. The detailed algorithm proceeds as follows:

1. Calculate T_N using the observed two sample data.
2. Obtain a bootstrap samples of size n_i; say y^*_{ij}, from the adjusted y_{ij}, that is, from $y^{adj}_{ij} = y_{ij} - \bar{y}_i + \bar{y}$, where $\bar{y}$ is the overall mean.
3. Calculate
$$T^*_N = \frac{\bar{y}^*_1 - \bar{y}^*_2}{\sqrt{s^{2*}_1/n_1 + s^{2*}_2/n_2}}.$$
4. Repeat step 2 and 3 B times ($B = 999$); thereby obtaining 999 bootstrap values of T_N*.
5. For a two sided test, a difference between the means is significant if the observed value of $|T_N| > 100(1-\alpha/2)th$ values of T^*_N.

2.5. A Non-Parametric Procedure [21]

To address the Behrens–Fisher problem, the Mann–Whitney–Wilcoxon test [28, 44] is modified in [21]. Define P_{1i}, the number of y_2 observations less than y_{1i}, for $i = 1, ..., n_1$. Similarly, define P_{2j}, the number of y_1 observations less than y_{2j}, for $j = 1, ..., n_2$. The P_{1i} and P_{2j} are called the placements of y_1 and y_2, respectively [21]. Let $\bar{P}_1$ denotes the mean of y_1 placements and $\bar{P}_2$ the mean of y_2 placements. Also compute the quantities $V_1 = \sum_{i=1}^{n_1}(P_{1i} - \bar{P}_1)^2$ and $V_2 = \sum_{j=1}^{n_2}(P_{2j} - \bar{P}_2)^2$, then the Fligner–Policello statistic (modified Mann–Whitney–Wilcoxon statistic) is given by

$$\hat{U} = \frac{\sum_{i=1}^{n_1} P_{2i} - \sum_{j=1}^{n_2} P_{1j}}{2(V_1 + V_2 + \bar{P}_1\bar{P}_2)^{1/2}}.$$

For a two-sided test the null hypothesis of equal medians is rejected if $|\hat{U}| \geq u_{\alpha/2}$. The critical value $u_{\alpha/2}$ can be calculated exactly or estimated using Monte Carlo simulation for large n_1 and n_2. The procedure is also available in contributed R package *NSM3*.

2.6. Simulations

We have conducted a simulation study to compare the performance, in terms of level and power, of 10 statistics, namely, the statistic T_N, the Welch Statistic T_1, the new procedure T, the likelihood ratio statistic L, the Wald

Test W, the score statistic S, the Fenstad statistic Z, the bootstrap procedure BT, the Wilcoxon two sample non parametric procedure WC and the recent non-parametric procedure FP by [21]. To perform WC we used R function *wilcox.test()*.

To compare the statistics in terms of size, we considered $\mu_1 = \mu_2 = 1$, a range of values of $VR = \sigma_1^2/\sigma_2^2 = 1/25, 2/24, 3/23, ..., 24/2, 25/1$, and a nominal level $\alpha = .05$. Note that this choice of variance ratios ensures comparison of the means for variances which are same to very different.

For sample sizes we considered equal and unequal n_1 and n_2. So, for example, n_1 was fixed at 5, 10, 15, 20, 25, 30. Then, for each fixed n_1, empirical levels were obtained for $n_2 = 5, 10, 15, 20, 25, 30$. These results are all given as graphs in Figures 1–6 in Appendix A1 in supplementary material. The graphs are in terms of size against $\rho = \log(\sigma_1^2/\sigma_2^2)$. All simulation results are based on 10,000 samples.

We now discuss the size results of the 10 statistics:

i. The statistics T_N and T_1: The statistic T_N is liberal, highly liberal for smaller n_1 and n_2. Even for $n_1 = n_2 = 30$, for which basic text books recommend its use, it is liberal, empirical level ranging, on average, from 0.0504 (when $VR \approx 1$) to 0.0618 (as VR is further and further away from 1). We then wanted to see what happens for larger n_1 and n_2. For this we extended the simulation study for (n_1, n_2)= (35,35), (40,40), (50,50), (60,60), (70,70), (80,80). Results are presented as graphs in Figure 7 in Appendix A1 in supplementary material. For $n_1 = n_2 = 35$, it holds level when $-1 < \rho < 1$. Otherwise, empirical level improves as the sample size increases. However, even at $n_1 = n_2 = 80$, this statistic is somewhat liberal, specially near $\rho = \pm 3$.
For a close comparison between T_N and T_1 empirical level results for $n_1 = n_2 = 35, 40, 50, 60, 70, 80$ are given as graphs in Figure 1. It shows that even at $n_1 = n_2 = 80$ empirical levels of T_N are slightly larger than those of T_1; T_N is still slightly liberal.

ii. The statistics T_1 and T: For all situations studied, even for $n_1 = n_2 = 5$, these two statistics hold level very closely having almost identical empirical levels. For a more close comparison between these two statistics some graphs containing empirical levels are given in Figure 2. From these graphs we conclude that T performs better than T_1 only

(**a**) for $n_1 = n_2$ and the variance ratio is moderate ($-.05 < \rho < .05$) and

(**b**) for $n_1 \neq n_2$ and sample size of the sample with larger variance is larger.

In all other situations, T_1, in general, performs better than or same as T.

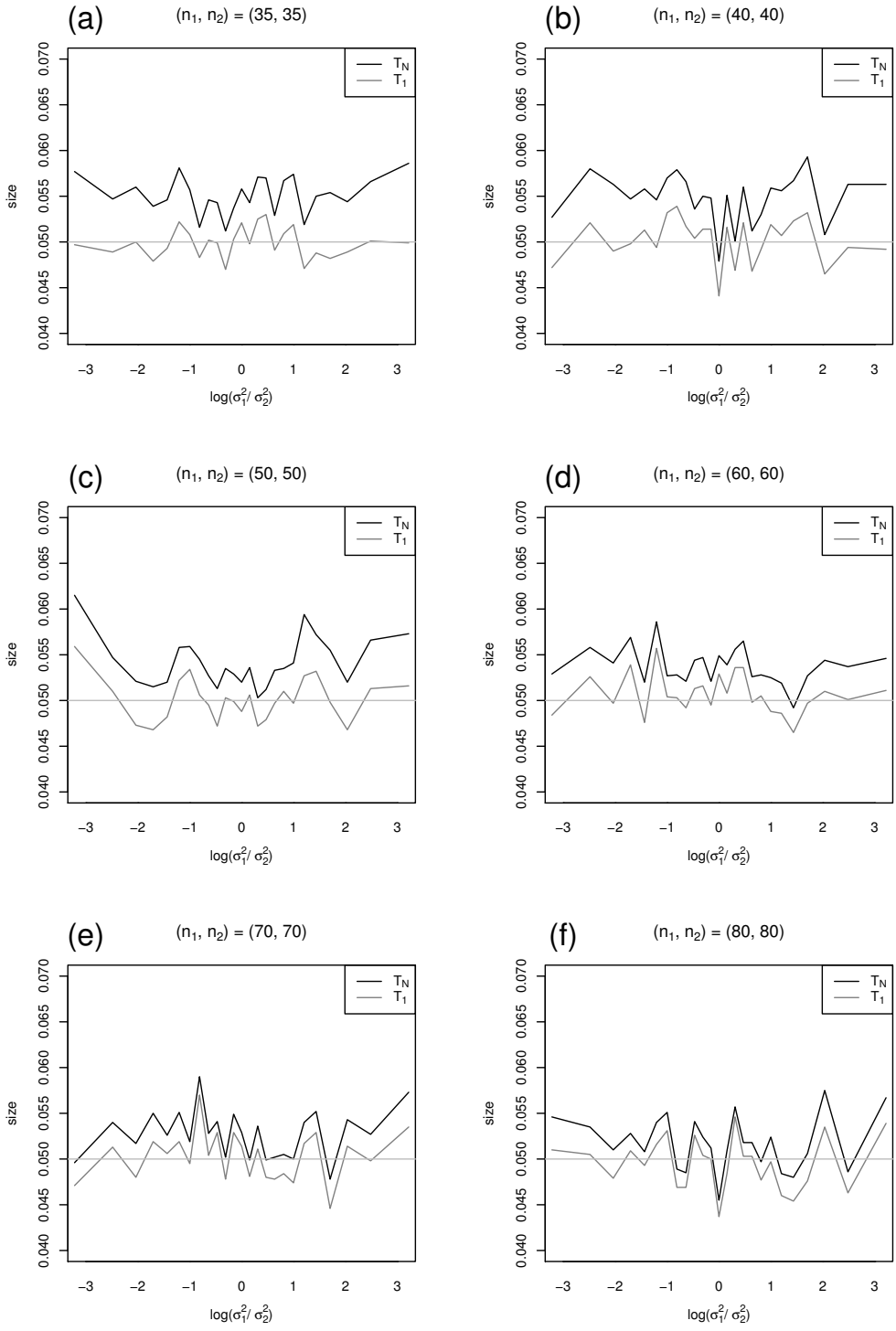

Figure 1: Plots of graphs showing empirical levels of the statistics T_N and T_1 for large sample sizes.

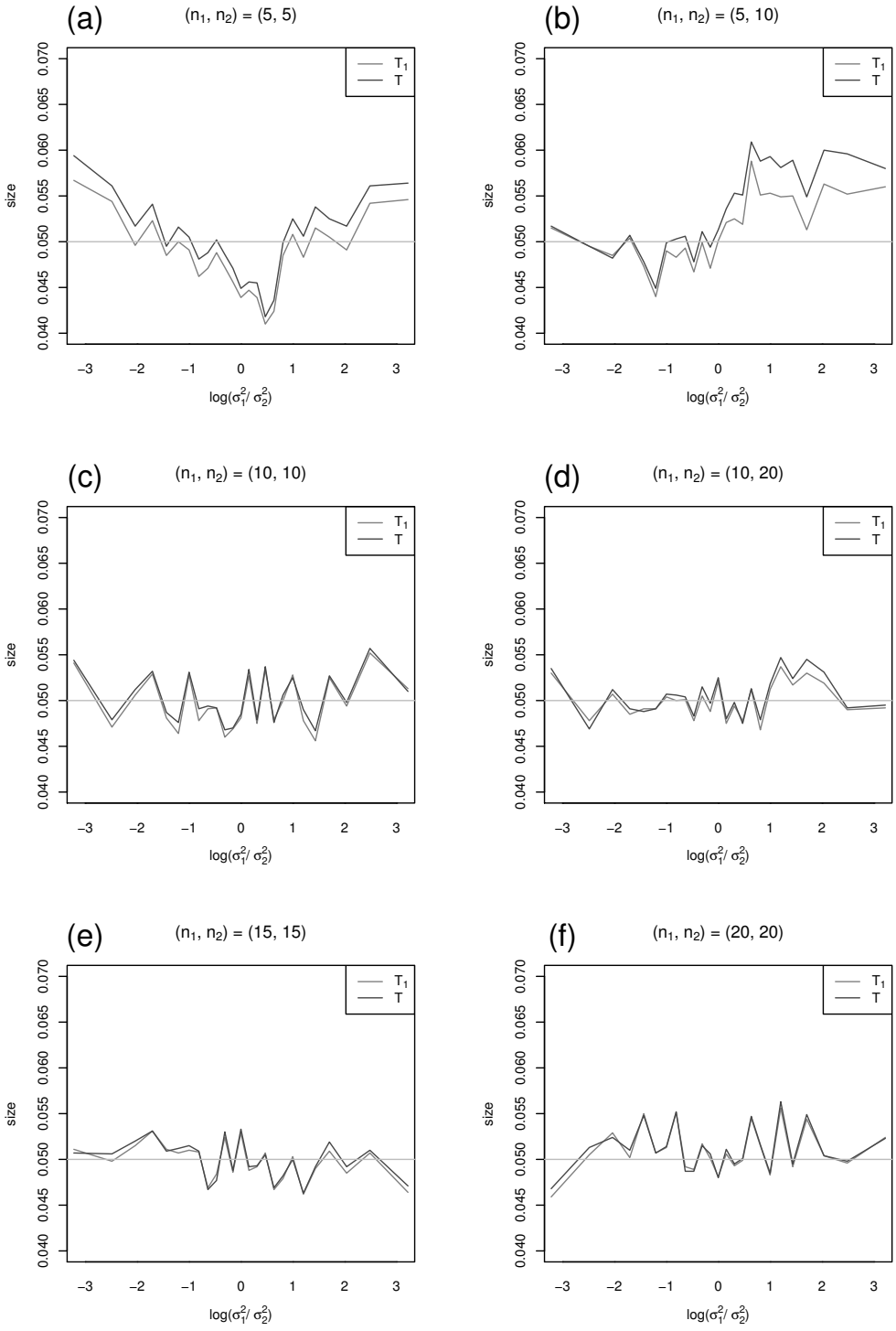

Figure 2: Plots of graphs showing empirical levels of the statistics T_1 and T under certain conditions explained in the text.

iii. The non-parametric procedures WC and FP: The Wilcoxon test WC, in general, shows extreme behaviour. It is either conservative or liberal depending on the value of ρ or whether $n_1 < n_2$ or $n_1 > n_2$. The improved non-parametric procedure that is most recently introduced and is available in the R package, is substantially better than WC. The extreme behaviour moderates a lot compared to WC. However, in general, it also does not hold level. Only for $n_1 = n_2$ empirical level performance of this procedure is very close to that of T_1 and T (slightly better than that of T_1 and T when $n_1 = n_2 = 5$ and ρ is not too far from zero).

iv. The bootstrap procedure BT: Only in some instances, for example, for $n_1 = 5$, $n_2 = 25$ and $n_1 = 5$, $n_2 = 30$ and $\rho \leq 0$, level performance of this statistics is similar to those of T_1 and T. However, this is a computer intensive procedure.

v. The Fenstad Statistic Z: This statistic is conservative for smaller sample sizes and liberal for larger sample sizes. Its best performance is for $n_1 = n_2 = 20$, even then it is conservative.

vi. The Statistics S, LR and W: The statistics LR and W are in general liberal and the statistic S is conservative. In a lot of situations, for example, for larger sample sizes the statistic S holds nominal level reasonably well (empirical size being very close to those of T_1 and T). Otherwise it is conservative.

For power comparison we considered all combinations of the sample sizes $n_1 = 5, 10, 15, 20, 25, 30$ and $n_2 = 5, 10, 15, 20, 25, 30$. The variance ratios considered were $VR = 1/16, 1/4, 1, 4, 16$. As in the study of performance in terms of size, the power study was done for the nominal level $\alpha = 0.05$. We use $\mu_1 = 1$ and $\mu_2 = \mu_1 + \tau$. The shift parameter τ is calculated as $\tau = \delta\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ (see, [7]), where $\delta = 1, 2, 3$. Departure from equality of means for fixed but unequal variance is measured by τ. The power results are given in Tables 1 to 36 in Appendix A2 in supplementary material.

We now discuss the power results.

i. The statistic T_N: It shows highest power which is not surprising as it is also highly liberal. It is interesting to note that, even though T_N is more liberal than T_1 and T for $n_1 = n_2 = 30$, it is only slightly more powerful. For large and equal sample sizes ($n_1 = n_2 = 80$) in which its empirical level is close to the nominal level power of this statistic is similar to that of T_1. A Power graph of T_N, T_1 and T for $n_1 = n_2 = 80$ and $\delta = 2$ against $VR = 1/16, 1/4, 1, 4, 16$ is given in Figure 3(a). The statistics T_1 and T show almost indistinguishable power, where as T_N shows slightly larger power. This is in line with the finding that T_N is slightly liberal.

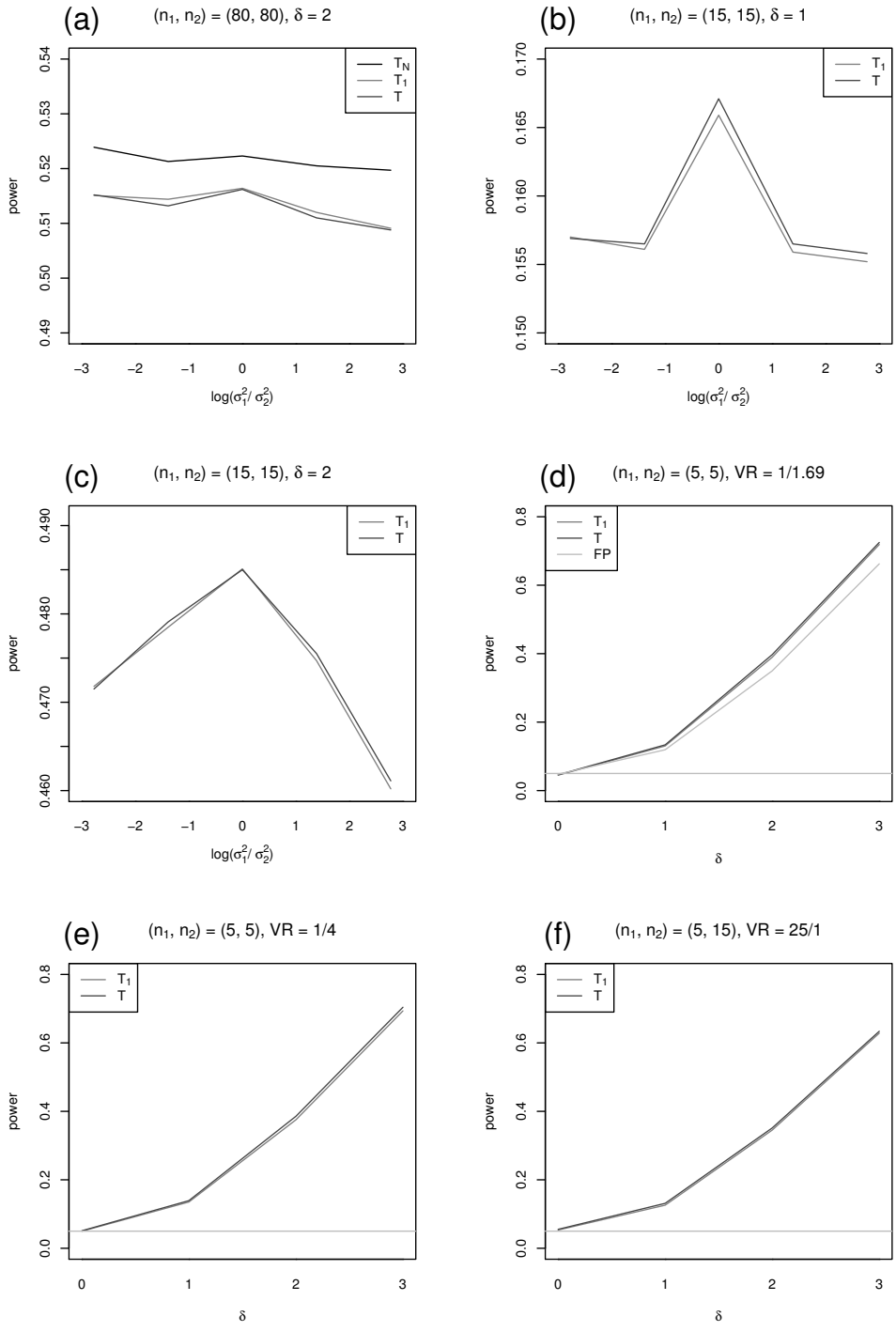

Figure 3: Plots of graphs showing empirical power:
(a) of the statistics T_N, T_1, and T for $(n_1, n_2) = (80, 80)$ and $\delta = 2$;
(b) of the statistics T_1 and T for $(n_1, n_2) = (15, 15)$ and $\delta = 1$;
(c) of the statistics T_1 and T for $(n_1, n_2) = (15, 15)$ and $\delta = 2$;
(d) of the statistics T_1, T, and FP for $(n_1, n_2) = (5, 5)$ and $VR = 1/1.69$;
(e) of the statistics T_1 and T for $(n_1, n_2) = (5, 5)$ and $VR = 1/4$;
(f) of the statistics T_1 and T for $(n_1, n_2) = (5, 15)$ and $VR = 25/1$.

ii. The statistics T_1 and T: Both these statistics show similar power. Power increases as δ increases. See, for instance, power graphs of both these statistics for $n_1 = n_2 = 15$, $\delta = 1$ and $\delta = 2$ against $VR = 1/16, 1/4, 1, 4, 16$ in Figures 3(b, c).

iii. As expected, power of all the other statistics L, W and Z or the procedures BT, WC and FP is more or less than that of T_1 and T depending on whether they are liberal or conservative.
We now examine a situation $n_1 = n_2 = 5$, $\rho = 1/1.69$ from $n_1 = n_2$ in which empirical level performance of the procedure FP is very close to that of T_1 and T. The power graph is given in Figure 3(d) (power against $\delta = 0, 1, 2, 3$). It shows that power of all three procedures increase as δ increases (as expected). However, as δ increases, power of FP does not increase as fast as the power of T_1 and T. In general, for smaller and equal sample sizes, level performances of the statistics T_1, T, BT, and FP are similar and hold level reasonably close to the nominal. However, in these situations power of the procedure FP is similar or somewhat smaller in comparison to that of the other three statistics or procedures.

iv. The Statistic S: In all those situations in which (for larger sample sizes and for $\rho < 0$) this statistic holds nominal level reasonably well (empirical size being very close to those of T_1 and T) the power of this statistic is also close to those of T_1 and T. Otherwise it is less powerful as expected.

2.7. An Example

This is a set of data from [26, p. 83]. The data which refer to driving times from a person's home to work, measured for two different routes, are 6.5, 6.8, 7.1, 7.3, 10.2 ($n_1 = 5$, $\bar{x}_1 = 7.58$, $s_1^2 = 2.237$) and 5.5, 5.8, 5.9, 6.0, 6.0, 6.0, 6.3, 6.3, 6.4, 6.5, 6.5 ($n_2 = 11$, $\bar{x}_2 = 6.136$, $s_2^2 = 0.073$). The means are different with very different variances. By examining the overall findings of the simulation results above, we see that the only statistic that is appropriate here is the statistic T_1 as $n_1 = 5$, $n_2 = 11$, $s_1^2 = 2.237$, $s_2^2 = 0.073$ are contrary to the situation in which the statistic T or the procedure FP is appropriate.

For these data the p-values of the statistics T_N, T_1, T, L, W, S, Z, BT, WC, FP are 0.0321, 0.0968, 0.0961, 0.0500, 0.0167, 0.1009, 0.0327, 0.3395, 0.0030, 0.0000 respectively.

Now, the value of $T_1 = 2.1426$ with p-value = .0968 indicates that means of the two groups are not different at 10% level of significance.

However, note that (from Figure 1(b) of the supplementary material) both T and T_1 hold level for $n_1 = 5$, $n_2 = 10$ and $\rho > 3$ and their p-values (.0968 and .0961) are also very similar. The same is more or less true for S whose empirical level is below 0.05 but not too much (again from Figure 1(b) of the supplementary material). The p-value of 0.10 for S is also not too different from those of T and T_1. The overall conclusion using the p-values coincide with the findings in Figure 1(b) of the supplementary material. But, since $n_1 = 5$, $n_2 = 11$ and $\hat{\rho} > 3$ for these data the conclusion is that the the hypothesis of equality of the means can be accepted at 10% level of significance. However, at 5% level of significance there is evidence that the two means are different.

3. TWO NEGATIVE BINOMIAL POPULATIONS

3.1. The Negative Binomial Formulation

The most convenient form of the negative binomial distribution, henceforth denoted by $NB(\mu, c)$ is

$$f(y|\mu, c) = Pr(Y = y|\mu, c) = \frac{\Gamma(y + c^{-1})}{y!\Gamma(c^{-1})} \left(\frac{c\mu}{1 + c\mu}\right)^y \left(\frac{1}{1 + c\mu}\right)^{c^{-1}}, \tag{3.1}$$

for $y = 0, 1, ...$, $\mu > 0$ [33, 34]. See, [31] for further details.

Now, let y_{i1}, ..., y_{in_i} be a sample realization from $NB(\mu_i, c_i)$, $i = 1, 2$. Our problem is to test $H_0 : \mu_1 = \mu_2$, where c_1 and c_2 are unspecified. To test this hypothesis [31] develop a likelihood ratio test L, a likelihood ratio test based on the bias corrected maximum likelihood estimates of the nuisance parameters $L(bc)$, a score test T^2_{NB} (henceforth denoted by S), a score test based on the bias corrected maximum likelihood estimates of the nuisance parameters $S(bc)$, a $C(\alpha)$ test based on the method of moments estimates of the nuisance parameters. [31] show that this later statistic, if Welch's [43] degree of freedom correction is applied, becomes identical to Welch's t-statistic T_1.

[31] investigated by simulations, for level and power, the statistics L, $L(bc)$, S, $S(bc)$, T_1, and the statistic T_N (pretending that negative binomial data can be treated as normal $N(\mu, \sigma^2)$ data). Their simulation study showed no advantage of the bias corrected statistics $L(bc)$ and $S(bc)$ over their uncorrected counterparts. So, here and in subsequent sections any statistic based on bias corrected estimates of the nuisance parameters will not be discussed. The remaining four statistics and the new statistic T developed in Section 2 for normal data are given below.

3.2. The likelihood Ratio Test

The likelihood ratio test is fully described and all necessary results are developed in [31]. So, to save space we omit this from presentation in this paper and refer the reader to that paper.

3.3. The Score Test

The score test statistic (for derivation see, [31]) is

$$S = \sum_{i=1}^{2} \frac{n_i(\bar{y}_i - \tilde{\mu}_0)^2}{\tilde{\mu}_0(1 + \tilde{\mu}_0\tilde{c}_{i0})},$$

which has an asymptotic $\chi^2(1)$ as $n \to \infty$, where $n = n_1 + n_2$.

3.4. The Other Three Statistics T_N, T_1 and T

These three statistics are given in Section 2.1 for data that come from normal distribution. Here the same statistics are used for negative binomial data as if these are normally distributed data.

Apart from the statistic T, which is newly introduced in Section 2.3, [31] show by simulations that for moderate to large sample sizes, in general, the statistic T_1 shows best overall performance in terms of size and power and it is easy to calculate. For large sample sizes, for example, for $n_1 = n_2 = 50$, all four statistics, L, S, T_1, T_N do well in terms of level and their power performances are also similar.

3.5. Simulations

We have conducted a simulation study to compare the 5 statistics T_N, T_1, T, L, and S, the bootstrap procedure BT and the two non-parametric procedures WC and FP. The three statistics T_N, T_1, and T and the three procedures BT, WC, and the FP are applied here exactly the same way as in the case of normally distributed data in Sections 2.4 and 2.5 respectively.

To compare the statistics in terms of size, we considered all combinations of the sample sizes $n_1 = 5, 10, 15, 20, 25, 30$ and $n_2 = 5, 10, 15, 20, 25, 30$, $\mu_1 = \mu_2 = 2$,

$c_1 = .10, .25, .40, .55, .70, .85, 1$, $c_2 = .10, .25, .40, .55, .70, .85, 1$, and a nominal level $\alpha = .05$. These results are all given as graphs in Figures 1–6 in Appendix B1 in Supplementary Material. The graphs are in terms of size against $\rho = \log(c_1/c_2)$. All simulation results are based on 10,000 samples. A discussion of the size results is given in what follows.

i. For $n_1 = n_2 = 5, 10$, the L statistic holds level most effectively (though somewhat conservative for $n_1 = n_2 = 5$ and somewhat liberal for $n_1 = n_2 = 10$), This finding is in line with Paul and Alam (2014). In these situations another statistic that is competing with L having very similar level is T_N.

ii. For the smaller of n_1 and n_2 equal to 5 and the other equal to 10 to 30, the L statistic performs best, although consistently somewhat conservative. In these situations, for all other statistics no consistent pattern emerges. For example, T_N is mainly very highly liberal, only in a very few situations its empirical level is close to the nominal level. For the smaller of n_1 and n_2 equal to 10 and the other equal to 10 to 30, the L statistic performs best, although consistently somewhat liberal. In these situations the other statistics are either liberal or conservative. For unequal sample sizes, smaller of n_1 and n_2 less than 20 and the other up to 30 the L statistic seems to perform best.

iii. For the smaller of n_1 and n_2 equal to or greater than 20 and the other also equal to or greater than 20, overall, the best performing procedures are through the use of the statistic T_1 or T or the score test statistic S. At $n_1 = n_2 = 30$ empirical level of all these 3 procedures are very close to the nominal level.

For power comparison we consider all combinations of the sample sizes $n_1 = 5, 10, 15, 20, 25, 30$ and $n_2 = 5, 10, 15, 20, 25, 30$. We use $\mu_1 = 1$, $c_1 = .1$, $c_2 = .10, .25, .40, .55, .70, .85, 1$, and $\mu_2 = \mu_1 + \delta$, for $\delta = 1.0, 1.5, 2.0$. As in the study of performance in terms of size, the power study was done for the nominal level $\alpha = 0.05$. All simulation results are based on 10,000 samples. A discussion of the power results is given in what follows.

We first concentrate on the L statistic which seems to be doing better in terms of size for the smaller of n_1 and n_2 less than 20 and the other up to 30. The power results are given in Tables 1 to 27 in Appendix B2 in supplementary material. In general, the L statistic shows highest power. Only in some situations the statistic T_N or T_1 or T show higher power, but in these situations these later statistics are also liberal.

Now we discuss power performance of the statistics T_1, T and S which perform best in terms of size starting at $n_1 = 20$ and $n_2 = 20$. Here we compare these only with the L statistic as it is, in general, liberal or conservative but not

too much. The power results are given in Tables 28 to 36 in Appendix B2 in supplementary material. The L statistic, in general, is somewhat more powerful than the other three statistics, but it is also slightly liberal in comparison to the other three statistics. The other 3 statistic show similar power. For example, for $n_1 = n_2 = 20$ and $c_2 = .7$ empirical level of L is close to 0.06 and those of the other three are close 0.05 (see, graph for $n_1 = n_2 = 20$ in Figure 4). The powers for L, T_1, T and S, $\delta = 2$, are 0.694, 0.554, 0.555 and 0.572 respectively (see, Table 22).

In general, power decreases as the value of c_2 goes further away from $c_1 = .10$ and increases as the sample size increases.

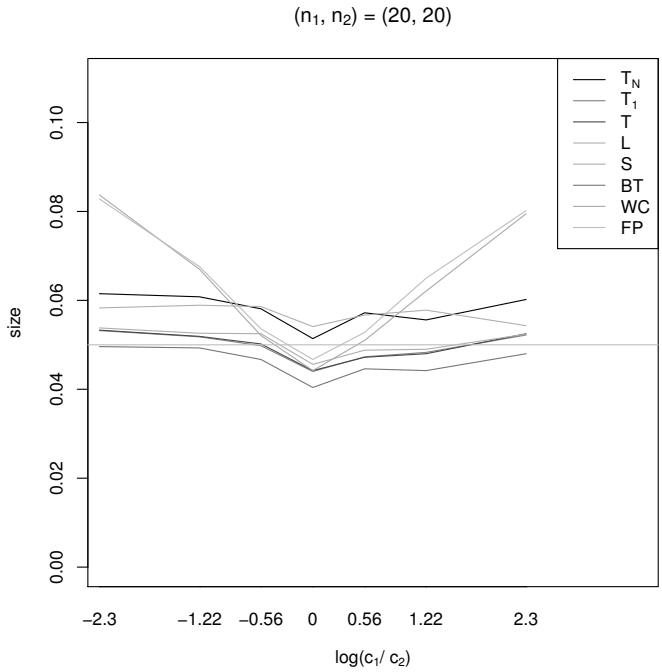

Figure 4: Plots of graphs showing empirical levels of all the statistics for $(n_1, n_2) = (20, 20)$.

3.6. An Example

[37] presents a set of data, originally given by [6], to see the effectiveness of a treatment (Cholestyramin), in comparison to a placebo, in reducing the number of vascular lesions. The data are given in Table 1, which refer to the observed number of vascular lesions on each patient's angiogram in the treatment group as well as in the control group (placebo).

Table 1: Frequency of patients by number of lesions on each patient's angiogram [6].

Number of lesions (y_{ij})	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	$\hat{\mu}$	$\hat{c}$
Cholestyramine	5	4	6	5	7	7	6	6	7	2	2	1	0	0	1	4.932	0.250
Placebo	2	4	6	4	6	9	7	5	2	4	4	2	0	2	0	5.509	0.185

The maximum likelihood estimates of μ and c based on a negative binomial model for the two groups are given in this table as well. The $\hat{\mu}$'s and the $\hat{c}$'s both differ. We now apply the statistics T_1, T and T^2_{NB} to test the equality of the two means. The values of T_1, T and S with p-value in the parenthesis are $-0.379(.705)$, $-0.379(.704)$, and $0.146(.702)$ respectively. Based of the p-values which are very close the difference is not significant.

We now show how to apply the bootstrap critical value method using the likelihood ratio statistic L for small sample sizes. For this we take a sample of size $n_1 = 15$ with replacement from the treatment group and a sample of size $n_1 = 10$ with replacement from the control group which are given below

Treatment group: 8 8 10 5 2 0 0 7 3 1 1 3 8 6 0;
Placebo group: 1 1 2 9 13 4 6 9 10 6.

Suppose these are the observed data for the two groups. For these data the value of L is 1.26 and the bootstrap 95% critical value is 5.16 which indicates that the difference between the two means is not significant.

The bootstrap critical value is obtained as: from the sampled data of $n_1 = 15$ and $n_2 = 10$ above we take 10000 pairs of samples (one sample of size 15 from the treatment group and one sample of size 10 from the control group) with replacement. For each pair of samples we obtain the value of L. Then the bootstrap critical value is the 9500^{th} value of the ordered (from smallest) L values.

4. TWO BETA-BINOMIAL POPULATIONS

4.1. The Beta-Binomial Formulation

For modelling data in the form of proportions with extra-dispersion the most popular model is the extended beta-binomial distribution of [35]. Let $y|p \sim$ binomial(m, p), where p is a beta random variable with mean π and variance

$\pi(1-\pi)\phi$, where ϕ is an extra dispersion parameter. Then the unconditional distribution of y is the extended beta-binomial distribution of [35] for which the pmf is given in what follows.

$$Pr(y|\pi,\phi) = \binom{m}{y} \frac{\prod_{r=0}^{y-1}[\pi(1-\phi)+r\phi] \prod_{r=0}^{m-y-1}[(1-\pi)(1-\phi)+r\phi]}{\prod_{r=0}^{m-1}[(1-\phi)+r\phi]} \tag{4.1}$$

with mean $m\pi$ and variance $m\pi(1-\pi)(1+(m-1)\phi)$, where $0 \leq \pi \leq 1$, and $\phi \geq \max[-\pi/(m-1), -(1-\pi)/(m-1)]$.

Denote this probability mass function by $\text{BB}(m,\pi,\phi)$. Now, let y_{i1}/m_{i1}, ..., y_{in_i}/m_{in_i} be a sample realization from $\text{BB}(m_{ij},\pi_i,\phi_i)$, $i=1,2$, $j=1,...,m_{in_i}$. Our purpose is to test $H_0 : \pi_1 = \pi_2$ with ϕ_1 and ϕ_2 being unspecified. [3] develop eight tests, namely, a likelihood ratio test, a $C(\alpha)$ (score) test based on the maximum likelihood estimates of nuisance parameters, a $C(\alpha)$ test based on the [25] method of moments estimates of the nuisance parameters, a $C\,(\alpha)$ test based on the quasi-likelihood and the method of moments estimates of the nuisance parameters by [8], a $C(\alpha)$ test based on the quasi-likelihood and the method of moments estimates of the nuisance parameters by [40], a $C(\alpha)$ test based on extended quasi-likelihood estimates of the nuisance parameters, and two non-parametric tests by [36]. See, [3] for further details.

By doing an extensive simulation study [3] show that none of the statistics, except the $C(\alpha)$ statistic C_{BB}, does well in terms of level and power. The statistic C_{BB} holds nominal level most effectively (close to the nominal level) and it is at least as powerful as any other statistic which is not liberal. It has the simplest formula, is based on estimates of the nuisance parameters only under the null hypothesis and is easiest to calculate. Also, it is robust in the sense that no distributional assumption is required to develop this statistic.

In this paper we compare the performance C_{BB} with the statistics T_N, T_1 and T, the bootstrap procedure BT and the two non-parametric procedures WC and FP. These are described below for the application to data in the form of proportions.

4.2. The Statistic C_{BB}

The statistic C_{BB} is (detailed derivation is given in [3] $C_{BB} = C^2/(A - A^2/B)$, which is distributed as chi-squared, asymptotically, as $n \to \infty$ ($n = n_1 + n_2$), with

1 degree of freedom, where

$$C = \sum_{j=1}^{n_1} \left[\frac{1}{1+(m_{1j}-1)\phi_1} \left\{ \frac{y_{1j}}{\pi} - \frac{m_{1j}-y_{1j}}{1-\pi} \right\} \right],$$

$$A = \sum_{j=1}^{n_1} \left[\frac{1}{1+(m_{1j}-1)\phi_1} \left\{ \frac{m_{1j}}{\pi(1-\pi)} \right\} \right]$$

and

$$B = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[\frac{1}{1+(m_{ij}-1)\phi_i} \left\{ \frac{m_{ij}}{\pi(1-\pi)} \right\} \right].$$

The parameters π, ϕ_1 and ϕ_2 in C, A and B are replaced by the maximum extended quasi-likelihood estimates $\hat{\pi}$, $\hat{\phi}_1$ and $\hat{\phi}_2$ obtained by solving

$$\sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[\frac{1}{1+(m_{ij}-1)\phi_i} \left\{ \frac{y_{ij}}{\pi} - \frac{m_{ij}-y_{ij}}{1-\pi} \right\} \right] = 0,$$

$$\sum_{j=1}^{n_1} \left[\frac{m_{1j}-1}{\{1+(m_{1j}-1)\phi_1\}^2} \left\{ y_{1j} \log\left(\frac{z_{1j}}{\pi}\right) + (m_{1j}-y_{1j}) \log\left(\frac{1-z_{1j}}{1-\pi}\right) - \frac{1+(m_{1j}-1)\phi_1}{2} \right\} \right] = 0$$

and

$$\sum_{j=1}^{n_2} \left[\frac{m_{2j}-1}{\{1+(m_{2j}-1)\phi_2\}^2} \left\{ y_{2j} \log\left(\frac{z_{2j}}{\pi}\right) + (m_{2j}-y_{2j}) \log\left(\frac{1-z_{2j}}{1-\pi}\right) - \frac{1+(m_{2j}-1)\phi_2}{2} \right\} \right] = 0$$

simultaneously.

4.3. The Bootstrap Procedure

The bootstrap procedure is developed here for data in the form of proportions (e.g. x/n) as follows:

1. Calculate the continuous data in the form of proportions for the two samples as $p_{ij} = y_{ij}/m_{ij}$, $i = 1, 2$, $j = 1, ..., m_{in_i}$. Let $\bar{p}_i = \sum_{j=1}^{n_i} p_{ij}/n_i$ and $s_{iP}^2 = \sum_{j=1}^{n_i} (p_{ij} - \bar{p}_i)^2/(n_i - 1)$. Then, define a statistic T_P, analogous to T_N, as

$$T_P = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{s_{1P}^2}{n_1} + \frac{s_{2P}^2}{n_2}}}.$$

2. Obtain a bootstrap sample of size n_i; say p^*_{ij}, from the adjusted p_{ij}, that is, from $p^{adj}_{ij} = p_{ij} - \bar{p}_i + \bar{p}$, where $\bar{p}$ is the overall mean of p_{ij} .

3. Calculate

$$T^*_P = \frac{\bar{p}^*_1 - \bar{p}^*_2}{\sqrt{s^{2*}_{1P}/n_1 + s^{2*}_{2P}/n_2}}.$$

4. Repeat step 2 and 3 B times ($B = 999$); thereby obtaining 999 bootstrap values of $T_{P}*$.

5. For a two sided test, a difference between the means is significant if the observed value of $|T_P| >$ ($100(1-\alpha/2)th$ values of T^*_P.

4.4. The Other Three Statistics T_N, T_1, and T and the Three Procedures BT, WC, and FP

Calculation of the three statistics T_N, T_1, and T and the three procedures BT, WC, and FP proceed by considering the p_{ij}, as y_{ij} in Section 2.

4.5. Simulations

We have conducted a simulation study to compare, in terms of level and power, the statistics C_{BB}, T_N, T_1 and T, the bootstrap procedure BT and the two non-parametric procedures WC and FP.

To generate data y_{ij} from $BB(m_{ij}, \pi_i, \phi_i)$, we take random samples with replacement of $n_1 = 5, 10, 15, 20, 25, 30$ litters with the litter sizes m_{1j}, $j = 1, ..., 27$ of the control group (Group 1) and $n_2 = 5, 10, 15, 20, 25, 30$ litters with the litter sizes m_{2j}, $j = 1, ..., 21$ of the medium group (Group 2) of Paul (1982). The m_{1j}, $j = 1, ..., 27$ of group 1 were 12, 7, 6, 6, 7, 8, 10, 7, 8, 6, 11, 7, 8, 9, 2, 7, 9, 7, 11, 10, 4, 8, 10, 12, 8, 7, 8 and m_{2j} of group 2 were 4, 4, 9, 8, 9, 7, 8, 9, 6, 4, 6, 7, 3, 13, 6, 8, 11, 7, 6, 10, 6. Note that our simulation study is much more extensive in comparison to [3]. Where as [3] consider fixed sample sizes ($n_1 = 27$ and $n_1 = 21$), we consider random samples of different sizes given above. The different combinations of parameter values are also much more extensive in our study.

For empirical levels we considered $\pi_1 = \pi_2 = \pi = 0.05, 0.10, 0.20, 0.40, 0.50$ and $(\phi_1, \phi_2) = (0.05, 0.50)$, $(0.10, 0.40)$, $(0.15, 0.30)$, $(0.20, 0.20)$, $(0.30, 0.15)$, $(0.40, 0.10)$, $(0.50, 0.05)$.

For power comparison the values of π_1 and π_2 considered were according to the formula $\pi_2 = \pi_1 + \delta$ with $\pi_1 = 0.05, 0.10, 0.20, 0.40$ and $\delta = 0.05, 0.10, 0.20$. That is, for each value of π_1 power has been simulated for three increments $0.05, 0.10, 0.20$. The same combination of values (ϕ_1, ϕ_2) were chosen as in the study of level performance.

All simulation results are based on 10,000 good samples. The definition of good samples here is "those samples for which the estimating equations converged within the permitted range $\cap_j \left(-1/(n_{ij} - 1)\right) < \phi_i < 1$, $i = 1, 2$. For more details see [3].

The empirical level results are summarized in Figures 1–36 in Appendix C1 and empirical power results are summarized in Tables 1–36 in Appendix C2 in Supplementary Material. The Level results are graphed against $\log(\phi_1/\phi_2)$ and power tables are in terms of VR $= (\phi_1/\phi_2)$.

We now discuss the size results of the 7 statistics:

(**i**) The statistics T_N: In general, the statistic T_N does not show any consistent behaviour, although shows mostly highly liberal behaviour.

(**ii**) The statistics T_1 and T: In general, level performance of these two statistics are similar. These two statistics hold level reasonably well when n_1, n_2 and π are all large, for example, for $n_1 \geq 20$ and $n_2 \geq 20$ and π ($\geq .2$). See Figures 22, 23, 24, 28, 29, 30, 34, 35 and 36 in Appendix C1 of the supplementary material. For some other situations, for example for $n_1 = n_2 = 10, 15$ and $\pi \geq .20$, performance of these two statistics are also the best and hold nominal level reasonably well. See Figures 8 and 15 in Appendix C1 of the supplementary material.

(**iii**) The statistic C_{BB}, recommended by Alam and Paul (2017): In some small sample size situations this statistic holds level reasonably well. See, for example, the situations in which one of the sample size is large and π is not too large ($\pi = .20$) (graphs (c) in figures 8–13, 15–18, 21–24, 28–36 in Appendix C of the supplementary material). For small π (in case of some of π=.05, .10, and .20) C_{BB} performs best (see Figures 8(a,b,c), 9(c), 10(c), 11(c), 14(c), 15(c), 16(c), 17(c), 21(c) in Appendix C1 of the supplementary material).
For large sample sizes ($n_1 \geq 20$ and $n_2 \geq 20$) level performance of C_{BB} is close to those of T_1 and T for $\pi = .2$. However, as π increases from .2 it shows conservative behaviour (see Figures 22, 23, 24, 28, 29, 30, 34, 35 and 36 in Appendix C1 of the supplementary material).

(**iv**) Performance of all other statistics are erratic at the best.

Next we discuss power performance.

(**i**) Since level performance of the statistic T_N, the bootstrap procedure BT and the two non-parametric procedures WC and FP are, in general, not satisfactory, we do not discuss their power performances, although power results are given in the supplementary material.

(**ii**) Power of T_1 and T are similar in all situations studied. Note from the level results that for large sample sizes ($n_1 \geq 20$ and $n_2 \geq 20$) level performance of C_{BB} is close to those of T_1 and T for $\pi = .2$ and as π increases from .2 it shows conservative behaviour. In all these situations power of C_{BB} is the best. That means, C_{BB} shows higher power even in situations where it is conservative but T_1 and T hold level. So, in these situations, unless C_{BB} can be adjusted to hold level we can not recommend its use. Power of C_{BB}, in most small sample sizes and small π ($< .2$) situations in which it holds level, in general, is larger or similar to those of T_1 and T.

4.6. Two Examples

Example 1. Here, for illustrative purposes, we use data from an experiment, given in [41], to identify in utero damage in laboratory rodents after exposer to boric acid. The study design involved four doses of boric acid. The compound was administered to pregnant female mice during the first 17 days of gestation, after which the dams were sacrificed and their litters examined. Table 2 lists the number of dead embryos and total number of implants at each of the four exposure doses: $d1 = 0$ (control), $d2 = 0.1$, $d3 = 0.2$, and $d4 = 0.4$ (as percent boric acid in feed).

The maximum likelihood estimates of the parameters (π, ϕ) for the four dose groups are also given in Table 2. It shows that the estimates of the $\hat{\pi}$'s are different and also the estimates of the $\hat{\phi}$'s are different. Now, suppose we want to compare π of the control group ($d_1 = 0$) with that of the dose group 4 ($d_4 = .4$). That is, we want to test $H_0 : \pi_1 = \pi_4$.

Now, the maximum likelihood estimate of π_1 is 0.069. If we assume that 0.069 is the true value of π_1 and $H_0 : \pi_1 = \pi_4$ is true, then, under the null hypothesis, the value of the common π is 0.069. Further, the sample sizes in the two groups are 27 and 26 which are between (25,25) and (30,30). Now, looking at Figures 29, 30, 35 and 36 in Appendix C1 of the supplementary material we see that none of the statistics hold nominal level for $\pi = 0.069$ and sample sizes $n_1 = 27$ and $n_2 = 26$. So, we apply a Monte Carlo Procedure (MCP) similar to

the parametric bootstrap. For this we consider

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Note that if we apply a t-test with Welch's degree of freedom, it becomes the procedure T_1. We now do the test by obtaining approximate critical values, for a two sided test, of the exact distribution of t which are calculated as what is given below.

Table 2: Per-litter data from Teratological study of boric acid (Stalon, *et al.* (2000). (i) Number of dead embryos. (ii) Total number of implants. Doses $d_1 = 0$ (control), $d_2 = 0.1$, $d_3 = 0.2$, $d_4 = 0.4$.

Dose Group															$\hat{\pi}$	$\hat{\phi}$
$d_1 = 0$	(i)	0	0	1	1	1	2	0	0	1	2	0	0	3	.0692	.0219
	1	0	0	2	3	0	2	0	0	2	1	1	0	0		
	(ii)	15	3	9	12	13	13	16	11	11	8	14	13	14		
	13	8	13	14	14	11	12	15	15	14	11	16	12	14		
$d_2 = 0.1$	(i)	0	1	1	0	2	0	0	3	0	2	3	1	1	.0968	.0058
	0	0	0	1	0	2	2	2	3	1	0	1	1	1		
	(ii)	6	14	12	10	14	12	14	14	10	12	13	11	11		
	11	13	10	12	11	10	12	15	12	12	12	12	13	15		
$d_3 = 0.2$	(i)	1	0	0	0	0	0	4	0	0	1	2	0	1	.0521	.0245
	1	0	0	1	0	1	0	0	1	2	1	0	0	1		
	(ii)	12	12	11	13	12	14	15	14	12	6	13	10	14		
	12	10	9	12	13	14	13	14	13	12	14	13	12	7		
$d_4 = 0.4$	(i)	12	1	0	2	2	4	0	1	0	1	3	0	1	.2234	.2497
	0	3	2	3	3	1	1	8	0	2	8	4	2			
	(ii)	12	12	13	8	12	13	13	13	12	9	9	11	14		
	10	12	21	10	11	11	11	14	15	13	11	12	12			

Keep m_{ij} fixed as given in the two groups, $j = 1, ..., 27$ for $i = 1$ and $j = 1, ..., 26$ for $i = 2$. Now, generate random numbers from $BB(m_{1j}, 0.069, 0.0218)$ for $j = 1, ..., 27$ and random numbers from $BB(m_{2j}, 0.069, 0.2496)$ for $j = 1, ..., 26$. This gives one sample for which calculate the value of t. Repeat this procedure and generate 100,000 samples and thereby 100,000 values of t. Order these 100,000 values from the smallest to the largest. The 2500^{th} and the 97500^{th} values are the 2.5% and the 97.5% critical values.

Now, the value of t from the data in the dose groups $d_1 = 0$ and $d_4 = .4$ is -2.8182. If -2.8182 does not fall between the 2.5% and the 97.5% critical values reject the null hypothesis of equality of the two proportions at 5% level of significance.

Following the procedure described above, the 2.5% and the 97.5% critical values obtained are -1.673003 and 2.637581 respectively. Since $T_1 = -2.8182$ falls in the rejection region the null hypothesis $H_0 : \pi_1 = \pi_4$ is rejected.

To check whether this procedure works we did some further simulations. For empirical level we again obtained 100,000 values of t as above with $\pi = 0.069$. We then calculated the proportion of t values that fall outside $(-1.673003, 2.637581)$. When this proportion is multiplied by 100 we obtain the empirical level. For power we do exactly the same as above but now take $\pi = 0.069 + \delta$, where $\delta = 0.02, .04, ..., .14$. The power results are given in Table 3.

Table 3: Power Table of T_1 and MCP, $\pi = 0.069 + \delta$, $\delta = 0, 02, .04, ..., .14$.

δ	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14
T_1	0.068	0.054	0.119	0.248	0.424	0.604	0.756	0.866
MCP	0.052	0.100	0.227	0.406	0.599	0.757	0.868	0.938

To compare the performance of the above Monte Carlo method with that of T_1 we extended the simulation study by obtaining the proportion of the 100,000 samples for which $|t| >$ the critical value of T_1 with Welch's degree of freedom. Results are also given in Table 3, which show that the new Monte Carlo procedure holds level almost exactly, the Welch T_1-test is somewhat liberal and yet the new procedure shows higher power compared to T_1.

Example 2. A data set from [45] of an in vivo cytogenetic assay is given Table 4. In this example, the sample sizes $n_1 = n_2 = 10$ are small in which the extended quasi-likelihood based score test C_{BB} does well (see Figure 8(a,b,c) in Appendix C1 of the supplementary material). For illustrative purpose, we test the equality of proportions in the first two groups. For this the value of $C_{BB} = 0.0171$ with p-value $= 0.8660$ showing strong support for the null hypothesis of the two proportions.

Table 4: Data from an in vivo cytogenetic assay [45].

Dose Group	No. of aberrant cells in 50 cells per animal										$\hat{\pi}$	$\hat{\phi}$
Negative control	0	4	0	0	4	0	1	1	0	0	0.0199	0.0447
Low Dose	1	0	3	0	1	0	3	0	0	1	0.0180	0.0125
Medium Dose	6	5	0	3	7	1	1	0	0	0	0.0454	0.0690
High Dose	3	2	1	6	4	0	0	0	0	5	0.0417	0.0476

5. TWO WEIBULL POPULATIONS

5.1. The Weibull Formulation

Data in the form of survival times arise in many fields of studies such as engineering, manufacturing, aeronautics and bio-medical sciences. See [29] for a recent review. The two parameter Weibull random variable Y with shape parameter β and scale parameter α has the probability density function

$$f(y;\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{(\beta-1)} \exp\left[-\left(\frac{y}{\alpha}\right)^{\beta}\right];\ y \geq 0;\ \beta,\alpha > 0. \tag{5.1}$$

The mean and variance of Y are $\mu = \alpha\Gamma(1+1/\beta)$ and $\sigma^2 = \alpha^2[\Gamma(1+2/\beta) - \{\Gamma(1+1/\beta)\}^2]$ respectively.

In some practical data analytic problems lifetimes or survival times data arise in the form of two samples following two independent Weibull populations with different shape and scale parameters. Let $y_{11}, ..., y_{1n_1}$ and $y_{21}, ..., y_{2n_2}$ be samples from two independent Weibull populations with parameters (α_1, β_1) and (α_2, β_2) respectively. In such a situation it may be of interest to test the equality of the scale parameters with the shape parameters being unspecified. That is to test the null hypothesis $H_0 : \alpha_1 = \alpha_2$, where β_1 and β_2 are unspecified.

For this problem [2] develop four test statistics, namely, a likelihood ratio statistic, a score statistic, and two $C(\alpha)$ statistics; one of which is based on the method of moments estimates of the nuisance parameters by [11] and the other is based on the method of moments estimates of the nuisance parameters by [42]. However, through a simulation study they show that the two statistics based on the method of moments estimates of the nuisance parameters perform best.

However, the actual analog of the Behrens–Fisher problem is to test $H_0 : \mu_1 = \mu_2$ with σ_1^2 and σ_2^2 being unspecified. To deal with this problem we develop a score test in Section 5.2. In Section 5.3 we conduct a simulation study to compare this statistic for level and power with the statistics T_N, T_1 and T, and the procedures BT, WC and the FP.

5.2. The Score Test

A score test statistic (derivation is given in Appendix E of the supplementary material) for testing $H_0 : \mu_1 = \mu_2$, where σ_1^2 and σ_2^2 are unknown and

unspecified is given by $S_w = S^2/I$, where

$$S = \frac{1}{\Gamma(1+\beta_1^{-1})}\left[-\frac{n_1\beta_1}{\mu} + \frac{\beta_1\{\Gamma(1+\beta_1^{-1})\}^{\beta_1}}{\mu^{\beta_1+1}}\sum_{j=1}^{n_1} y_{1j}^{\beta_1}\right] + \frac{1}{\Gamma(1+\beta_2^{-1})}\left[\frac{n_2\beta_2}{\mu} - \frac{\beta_2\{\Gamma(1+\beta_2^{-1})\}^{\beta_2}}{\mu^{\beta_2+1}}\sum_{j=1}^{n_2} y_{2j}^{\beta_2}\right]$$

and

$$I = \frac{1}{\Gamma(1+\beta_1^{-1})}\left\{\frac{n_1\beta_1}{\mu^2} - \frac{\beta_1(\beta_1+1)\{\Gamma(1+\beta_1^{-1})\}^{\beta_1}}{\mu^{\beta_1+2}}\sum_{j=1}^{n_1}\mathrm{E}(y_{1j}^{\beta_1})\right\} - \frac{1}{\Gamma(1+\beta_2^{-1})}\left\{\frac{n_2\beta_2}{\mu^2} - \frac{\beta_2(\beta_2+1)\{\Gamma(1+\beta_2^{-1})\}^{\beta_2}}{\mu^{\beta_2+2}}\sum_{j=1}^{n_2}\mathrm{E}(y_{2j}^{\beta_2})\right\}.$$

In S and I the quantity, such as $\mathrm{E}(y^{\beta_i})$ is calculated as $\mathrm{E}(y^{\beta_i}) = \int_0^\infty y_i^{\beta} f(y,\mu,\beta_i)dy$. Of course, the parameters μ, β_1 and β_2 in S and I are to be replaced by their maximum likelihood estimates $\hat{\mu}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ which are obtained by maximizing the log-likelihood function

$$l = \sum_{i=1}^{2}\left[\frac{1}{\Gamma(1+\beta_i^{-1})}\left\{n_i\log\left(\frac{\beta_i\Gamma(1+\beta_i^{-1})}{\mu}\right) + (\beta_i-1)\left\{\sum_{j=1}^{n_i}\log(y_{ij}) - n_i\log\left(\frac{\mu}{\Gamma(1+\beta_i^{-1})}\right)\right\}\right\}\right] - \sum_{i=1}^{2}\frac{\{\Gamma(1+\beta_i^{-1})\}^{\beta_i-1}}{\mu^{\beta_i}}\sum_{j=1}^{n_i} y_{ij}^{\beta_i}$$

with respect to the parameters μ, β_1 and β_2. The distribution of S_w is asymptotically distributed as chi-square with one degrees of freedom.

5.3. Simulations

We have conducted a simulation study to compare the statistic S_w, in terms of level and power, with the three statistics T_N, T_1 and T, and the three procedures BT, WC, and FP. These statistics are applied here exactly the same way as in the case of normally distributed data studied in Sections 2.4 and 2.5. As in the two previous sections we use the Weibull data as if the data come from two normal populations.

To compare the statistics in terms of size and power, we considered the sample sizes $n_1 = 5, 10, 15, 20, 25, 30$ and $n_2 = 5, 10, 15, 20, 25, 30$. We generate data from the Weibull (α_1, β_1) and Weibull (α_2, β_2) populations. For size comparison, in order to comply with equal means condition, we fix the values of α_1, β_1, and β_2;

and evaluate the expression $\{\alpha_1\Gamma(1+1/\beta_1)\}/\{\alpha_2\Gamma(1+1/\beta_2)\} = 1$ to obtain the value of α_2. For power comparison, we again fix the values of α_1, β_1, and β_2; but evaluate the expression $\{\alpha_1\Gamma(1+1/\beta_1)\}/\{\alpha_2\Gamma(1+1/\beta_2)\} = 1/(1+\delta)$ with $\delta = .1, .2, .3$, to obtain the value of α_2. Both the size and power are calculated for all combinations of $\beta_1 = 1, 2, 3, 4, 5$ and $\beta_2 = 2, 3, 4$ while fixing $\alpha_1 = 1$ and determining α_2 from the expressions given above.

The size results are all given as graphs in Figures 1–36 and the power results are all given in Tables 1–36 in Appendix D2 in supplementary material. The graphs are in terms of size against $\rho = \log(\sigma_1^2/\sigma_2^2)$. All simulation results are based on 10,000 samples.

We now discuss the size results, of the 7 statistics, given in Figures 1–36 in Appendix D1 in the supplementary material:

(**i**) The statistic T_N: The statistic T_N is liberal, highly liberal for smaller n_1 and n_2. Even for $n_1 = n_2 = 30$ it is liberal, empirical level ranging, on average, from 0.0525 (when $VR \approx 1$) to 0.0781 (as VR is further and further away from 1).

(**ii**) The statistics T_1 and T: Overall, these two statistics perform best, even for smaller sample sizes, holding empirical levels closer to the nominal. Only exceptions are when the sample size differences are large as well as when the differences between the variances are large, also when $n_2 > n_1$ as well as $\sigma_2^2 > \sigma_1^2$. In these situations both of these statistics can be quite liberal, although T_1 is slightly better than T. See, for example, Figures 4, 5, 11, 12, 25 of Appendix D1 of the the supplementary material.

(**iii**) Behaviour of the remaining four statistics or procedures are inconsistent, sometimes very liberal and sometimes very conservative. The exceptions are for

  (**a**) FP for $n_1 = n_2$ which does as well as T_1 and T in some cases (see, for example, Figure 1),

  (**b**) BT, irrespective of sample sizes, which does as well or better than T_1 and T (see, for example, Figure 5).

Next we discuss power performance using the power results given Tables 1–36 in Appendix D2 in the supplementary material.

Since the procedures T_N, WC, and S_w have highly inconsistent behaviour in terms of level, we omit these from power discussion. Power of T_1 and T are similar. However, T shows some edge over T_1. In general, these show higher power than FP and BT. Even in the situations in which FP and BT have slight advantage in terms level, T_1 and T maintain higher power.

5.4. An Example

[17] give data on survival times (in weeks) for two groups of patients who died of acute myelogenous leukemia. Patients were classified into the two groups according to the presence or absence of a morphologic characteristic of white cells. Patients termed AG positive were identified by the presence of Auer rods and/or significant granulature of the leukemic cells in the bone marrow at diagnosis. For the AG negative patients these factors were absent. The survival times for 17 patients in the AG positive group were: 65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65 and those for 16 patients in the AG negative group were: 56, 65, 17, 17, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43.

We now test the equality of the mean survival times in the two groups. As the performance of the five statistics or procedures T_N, WC, BT, FP and C_W are far less than satisfactory we do not consider them any further. The values of T_1 and T with corresponding p-values in the parenthesis are 3.1124 (0.0054) and 3.1124 (0.0047) respectively leading the conclusion that the two means are not the same.

6. DISCUSSION

We do a comprehensive review of the standard Behrens–Fisher (BF) problem and some of its analogs. Among the B-F analogous problems we deal with the two parameters negative binomial, the Beta-binomial, and the two parameter Weibull. In each case a number of procedures are either reviewed or developed and extensive simulation studies are conducted to study the properties of the procedures in terms of size and power. Some new results and findings are shown and examples of application are given in all cases.

If the variance ratio is known, the mixing parameter λ in $\mathcal{K}$ is then known, so the distribution of T (§2.3) becomes pivotal, which is not an exact t-distribution. In fact, if the variance ratio is given, one should use the pooled variance estimator which can lead to a t-statistic. For other distributions other than the normal cases, it is the same story but in an asymptotical sense. The tests based on t-distributions or chi-square distributions or any other derived from "normal" distributions all become asymptotical approximations. Therefore, if there is some reason to specify the variance ratio σ_2/σ_1, the traditional two independent samples Student t-test or Welch test are usable but both are approximations.

A review paper can possibly be never complete given that a vast literature is available. Here also we do not make such a claim. For example, we do not consider the Bayesian methods [24, 46] to the solve Behrens–Fisher problem.

For the standard Behrens–Fisher problem we studied 10 procedures T_N, T_1, T, L, W, S, Z, BT, WC and FP including a new procedure T. Based on the finding through extensive simulation study we recommend that the statistic T_N be used only when the two means are visibly different or if the sample sizes are large, such as, $min(n_1, n_2) \geq 80$ (only at this sample size level the Central Limit Theorem reasonably takes hold); otherwise use T_1 except for

(**i**) (**a**) $n_1 = n_2$ and the variance ratio is not extreme (close to 1/25 or 25/1 limits),
(**b**) for $n_1 \neq n_2$ and sample size of the sample with larger variance is larger, in which case use T;

(**ii**) for smaller and equal sample sizes use the procedure FP.

For the negative binomial BF Problem we studied five statistics T_N, T_1, T, LR_{NB} and T^2_{NB} and the bootstrap procedure BT and two non-parametric procedures WC and FP. Note that six of these T_N, T_1, T, BT, WC and FP are the same as those used for the standard BF problem. We recommend that for the smaller of n_1 and n_2 less than 20 and the other up to 30 the LR statistic, although somewhat liberal or conservative, be used. In these situations, in general, it is most powerful. However, for some extra effort, it would be advisable to use the bootstrap p-value based on this statistic. For the sample sizes stating at $n_1 = 20$ and $n_2 = 20$ (n_1 equal to or not equal to n_2) the statistics T_1, T and S all hold level reasonably well and at $n_1 = n_2 = 30$ empirical level of all these 3 procedures are very close to the nominal level. In these situations these statistics are also, in general, most powerful and therefore recommended. The practitioner can use any one of them.

For the beta-binomial BF problem we have studied seven statistics or procedures C_{BB}, T_N, T_1, T, BT, WC, and FP. For larger sample sizes (n_1 or n_2 ≥ 20) and for large π ($\geq .2$) the statistics T_1 and T are the best and therefore recommended. For small sample sizes and small π ($< .2$) we recommend to use the statistic C_{BB}. In all other situations we recommend a procedure similar to the parametric bootstrap given in example 1 in Section 4.5.

The results of the statistics T_1 and T are interesting. Even though here we are not dealing with normal data, the level properties, for large sample sizes and large π (n_1, $n_2 \geq 20$, and $\geq .2$), show to be similar to those for normally distributed data. The reason, in our opinion, is that the transformation of the discrete (binomial) data y_{ij} to continuous (proportions) data $p_{ij} = n_{ij}/y_{ij}$ does the trick in this situation.

For the Weibull BF problem also we have studied seven statistics S_w, T_N, T_1, T, BT, WC and FP. Based on extensive simulation studies we recommend that the statistic T_1 or T be used for larger sample sizes (n_1 and n_2 both larger than 25), otherwise use the bootstrap p-value or the approximate critical value of the exact distribution of the statistic based on T_1 or T.

The interesting overall finding is that the statistic T_1 or T can be used for all the cases studied here for sample sizes larger than 25 except for the beta-binomial samples in which the additional requirement is that π be large ($\geq .2$). For smaller sample sizes, specific recommendations given above, on a case by case basis, should be followed. The statistic T_N should never be used in the BF or BF analogous problems unless the two sample sizes are very large.

It will be interesting to find through further studies whether these recommendations are applicable in other BF analogous problems, such as, testing equality of means of two gamma, extreme value and log-normal or other similar survival populations having possibly different variances. In some large sample size situations or in sparse (beta-binomial with $\pi \leq .1$) situations for data in the form of proportions we recommended using a parametric bootstrap type procedure. Further research in this area should focus on improvements in performance, specially in terms of levels, of some of the statistics, such as the statistic C_{BB}.

For testing the equality of the scale parameters with the shape parameters being unspecified of two Weibull populations [2] develop four test statistics of which they recommend the statistics based on two different method of moments estimates of the nuisance parameters. It will be of interest to develop these later two statistics for testing $H_0 : \mu_1 = \mu_2$ with σ_1^2 and σ_2^2 being unspecified and compare with the statistics recommended in this paper.

SUPPLEMENTARY MATERIAL

Supplementary material for "Empirical level and Power" that includes graphs of empirical levels and tables of empirical power referred to in Sections 2, 3, 4 and 5 and derivation of the score test referred to in Section 5.2 are available as Appendix A, Appendix B, Appendix C, Appendix D, and Appendix E in `https://dataverse.scholarsportal.info/dataverse/sudhirpaul`. Empirical level graphs and empirical power tables for the normal BF problem are in Appendix A1 and Appendix A2 respectively. The (level graphs, power tables) for the negative binomial, beta-binomial, and the Weibull BF analogous problems are in Appendices (B1, B2), (C1, C2) and, (D1, D2) respectively.

ACKNOWLEDGMENTS

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Australian Research Council Discovery Project (ARC DP130100766 and DP160104292). The paper was completed while Sudhir Paul was visiting School of Mathematical Sciences, Queensland University of Technology during January–March, 2016 and 2017.

REFERENCES

[1] Algina, J.; Oshima, T.C. and Lin, W.-Y. (1994). Type I error rates for Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups, *Journal of Educational Statistics*, **19**(3), 275–291.

[2] Alam, K. and Paul, S. (2015). Testing equality of scale parameters of two Weibull distributions in the presence of unequal shape parameters, *Acta Comment. Univ. Tartu. Math.*, **19**(1), 11–26.

[3] Alam, K. and Paul, S. (2017). Testing Equality of Two Beta Binomial Proportions in the Presence of Unequal Extra-dispersion Parameters, *Comm. Statist. Simulation Comput.*, **45**(4), 1–25.

[4] Barnard, G.A. (1984). Comparing the means of two independent samples, *Applied Statistics*, **33**, 266–271.

[5] Behrens W.H.V. (1929). Ein Beitrang Zur Fehlerberchnung bei wenigen beobachtungen, *Landwirtsch Jahrbucher*, **68**, 807–837.

[6] Brensike, J.F.; Kelsey, S.F.; Passamani, E.R.; Fisher, M.R.; Richardson, J.M.; Loh, I.K.; Stone, N.J.; Aldrich, R.F.; Battaglini, J.W.; Moriarty, D.J.; Myrianthopoulos, M.B.; Detre, K.M.; Epstein, S.E. and Levy, R.I. (1982). National Heart, Lung, and Blood Institute Type II coronary intervention study: design, methods, and baseline characteristics. *Control Clin. Trials*, **3**(2), 91–111.

[7] Best, D.J. and Rayner J.C.W. (1987). Welch's Approximate Solution for the Behrens–Fisher Problem, *Technometrics*, **29**(2), 205–210.

[8] Breslow, N.E. (1990). Tests of hypotheses in over-dispersed Poisson regression and other quasi-likelihood models, *J. Amer. Statist. Assoc.*, **85**, 565–571.

[9] Chang, C.-H. and Pal, N. (2008). A revisit to the Behrens–Fisher problem: comparison of five test methods, *Communications in Statistics — Simulation and Computation*, **37**(6), 1064–1085.

[10] Cox, D.R. and Snell, E.J. (1968). A general definition of residuals, *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, **30**, 248–275.

[11] Cran, G.W. (1988). Moment estimators for the 3-parameter Weibull distribution, *IEEE Trans. Rel.*, **37**(4), 360–363.

[12] Derrick, B.; Toher, D. and White, P. (2016). Why Welch's test is Type I error robust, *The Quantitative Methods in Psychology*, **12**(1), 30–38.

[13] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London.

[14] Fagerland, M.W. and Sandvik, L. (2009). The Wilcoxon–Mann–Whitney test under scrutiny, *Statistics in Medicine*, **28**(10), 1487–1497.

[15] Fagerland, M.W. and Sandvik, L. (2009). Performance of five two-sample location tests for skewed distributions with unequal variances, *Contemporary Clinical Trials*, **30**(5), 490–496.

[16] Fay, M.P. and Proschan, M.A. (2010). Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules, *Statistics Surveys*, **4**, 1–39.

[17] Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information, *Biometrics*, **21**, 826–838.

[18] Feiveson, A.H. and Delaney, F.C. (1968). The Distribution and Properties of a Weighted Sum of Chi Squares, *Manned Spacecraft Center, Houston, Texas*, **1**, NASA TN D–4575.

[19] Fenstad, G.V. (1983). A Comparison between the U and V tests in the Behrens–Fisher Problem, *Biometrika*, **76**(1), 300–302.

[20] Fisher, R.A. (1936). The Statistical Utilization of Multiple Measurements, *A. Eug.*, **11**, 377–386.

[21] Fligner, M.H. and Policello, C.C. (1981). Robust rank procedures for the Behrens–Fisher problem, *J. Amer. Statist. Assoc.*, **76**(4), 162–168.

[22] Gail, M.H.; Santner, T.J. and Brown, C.C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumour, *Biometrics*, **36**, 225–231.

[23] Hogg, R.V. and Tanis, E.A. (2010). *Probability and Statistical Inference*, Pearson, New York.

[24] Kim, S-H. and Cohen, A.H. (1998). On the Behrens–Fisher Problem: A Review, *J. Educ. Behav. Stat.*, **23**(4), 356–377.

[25] Kleinman, J.C. (1973). Proportions with extraneous variance: single and independent samples, *J. Amer. Statist. Assoc.*, **68**(341), 46–54.

[26] Lehman, E.L. and D'Abrera, H.J.M. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, Halden-Day: U.S.A.

[27] Manly, B.F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd ed., London, Chapman and Hall.

[28] Mann, H.B. and Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics*, **18**(1), 50–60.

[29] Murthy, D.N.P.; Xie, M. and Jiang, R. (2004). *Weibull Models*, John Wiley & Sons, Inc., New Jersey.

[30] Paul, S.R. (1992). Comment on Best and Rayner (1987), *Technometrics*, **34**(2), 249–250.

[31] Paul, S. and Alam, K. (2014). Testing equality of two negative binomial means in presence of unequal over-dispersion parameters: a Behrens–Fisher problem analog, *J. Stat. Comput. Simul.*, **45**(15), 3140–3153.

[32] Paul, S.R. and Islam, A. (1995). Analysis of proportions in the presence of over-/under-dispersion, *Biometrics*, **51**, 1400–1410.

[33] Paul, S.R. and Plackett, R.L. (1978). Inference sensitivity for Poisson mixtures, *Biometrika*, **65**(3), 591–602.

[34] Piegorsch, W.W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter, *Biometrics*, **46**, 863–867.

[35] Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors, *J. Amer. Statist. Assoc.*, **81**(394), 321–327.

[36] Rao, J.N.K. and Scott, A.J. (1992). A simple method for the analysis of clustered binary data, *Biometrics*, **48**, 577–585.

[37] Saha, K. (2013). Interval estimation of the mean difference in the analysis of over-dispersed count data, *Biom. J.*, **55**(1), 114–133.

[38] Saha, K. and Paul, S.R. (2005). Bias corrected maximum likelihood estimator of the negative binomial dispersion parameter, *Biometrics*, **61**(1), 179–185.

[39] Singh, P.; Saxena, K.K. and Srivastava, O.P. (2002). Power comparisons of solutions to the Behrens–Fisher problem, *American Journal of Mathematical and Management Sciences*, **22**(3–4), 233–250.

[40] Srivastava, M.S. and Wu, Y. (1993). Local efficiency of moment estimators in beta-binomial model, *Comm. Statist. — Theory and Methods*, **22**(9), 257–261.

[41] Slaton, T.L.; Piegorsch, W.W. and Durham, S.D. (2000). Estimation and Testing with Overdispersed Proportions Using the Beta-Logistic Regression Model of Heckman and Willis, *Biometrics*, **56**(1), 125–133.

[42] Teimouri, M. and Gupta, A.K. (2013). On the three-parameter Weibull distribution shape parameter estimation, *J. Data Sci.*, **11**(3), 403–414.

[43] Welch, B.L. (1938). The significance of the difference between two means when the population variances are unequal, *Biometrika*, **29**(3/4), 350–362.

[44] Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin*, **1**(6), 80–83.

[45] Williams, D.A. (1988). Extra-binomial variation in toxicology, *Proceedings of the XIVth International Workshop on Statistical Modelling*, 165–174.

[46] Yin, Y. and Li, B. (2014). Analysis of the Behrens–Fisher Problem based on Bayesian Evidence, *J. Appl. Math.*, **2014**.

REVSTAT – Statistical Journal
Volume 17, Number 4, October 2019, 599–616

FORECASTING DAILY EXCHANGE RATES: A COMPARISON BETWEEN SSA AND MSSA

Authors: Rahim Mahmoudvand
– Department of Statistics, Bu-Ali Sina University, Hamedan, Iran
R.mahmoudvand@basu.ac.ir

Paulo Canas Rodrigues
– CAST, Faculty of Natural Sciences, University of Tampere, Tampere, Finland
AND
Department of Statistics, Federal University of Bahia, Salvador, Brazil
paulocanas@gmail.com

Masoud Yarmohammadi
– Department of Statistics, Payame Noor University, Tehran, Iran

Received: August 2016 Revised: September 2017 Accepted: September 2017

Abstract:

- In this paper, daily exchange rates in four of the BRICS emerging economies: Brazil, India, China and South Africa, over the period 2001 to 2015 are considered. In order to predict the future of exchange rate in these countries, it is possible to use both univariate and multivariate time series techniques.

  Among different time series analysis methods, we choose singular spectrum analysis (SSA), as it is a relatively powerful non-parametric technique and requires the fewest assumptions to be hold in practice. Both multivariate and univariate versions of SSA are considered to predict the daily currency exchange rates. The results show the superiority of MSSA, when compared with univariate SSA, in terms of mean squared error.

Key-Words:

- *multivariate singular spectrum analysis; univariate singular spectrum analysis; forecasting; exchange rates.*

AMS Subject Classification:

- 37M10, 15A18, 62M15.

1. INTRODUCTION

Exchange rates are among the most important economic indices in the international monetary markets, as they powerfully affect cross-border economic transactions and have the greatest attention in monetary policy debates. Therefore, central banks should pay special attention to exchange rates and the value of their domestic currency (Dilmaghani and Tehranchian, 2015). Significant impact of economic growth, trade development, interest rates and inflation rates on exchange rates make it extremely difficult to predict them (Yu *et al.*, 2007). Therefore, exchange rates forecasting has become a very important and challenge research issue for both academic and industrial communities. By now, there is a vast literature considering the problem of exchange rate forecasting. We categorise them into three types:

(**i**) Explanation based methods: In these methods, the economic theory describes the evolution path of exchange rates based on the variability of economic variables. Depending on the type of economic variables, macroeconomic or microeconomic, have been introduced two different methods:

  (**a**) Monetary exchange rate models that use macroeconomic variables. Investigation on these methods imply that, over long horizons, the fluctuations in fundamentals can be used successfully for exchange rate forecasting. More informations about these methods and a literature review can be found for example in Engle and West (2005), Della Corte and Tsiakas (2011) and Plakandaras (2015).

  (**b**) Microstructural based models that use microeconomic variables. In these methods, exchange rate fluctuations are related to short run changes in microeconomic variables. More details can be found for example in Papaioannou *et al.* (2013) and Janetzko (2014).

(**ii**) Extrapolation based methods: These methods use only historical data on the exchange rates and can be categorized in two groups:

  (**a**) Parametric methods: Autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH) and vector autoregressive (VAR) models are the most widely used methods in this category. A good review on related works is provided by Plakandaras (2015).

  (**b**) Non-parametric methods: Machine learning methodologies and more specifically Artificial Neural Network (ANN) and Support Vector Machines (SVM) gained significant merit in exchange rate forecasting (see for example Yu *et al.*, 2007).

Overall, according to the existing literature, the methods that incorporate denoised series in the analysis produce better results than other methods (see, for example, Fu (2010) and Lin *et al.* (2012)).

In the light of the above discussion, in this study, we apply Singular Spectrum Analysis (SSA), which is a powerful non-parametric technique for time series analysis. SSA incorporates the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing (Golyandina *et al.*, 2001). SSA is designed to look for nonlinear, non?stationary, and intermittent or transient behaviour in an observed time series, and has gained successful application in the various sciences such as meteorological, bio-mechanical, hydrological, physical sciences, economics and finance, engineering and so on. By now, many studies used SSA and its applications (see, for example, Hassani *et al.* (2009a, 2013, 2015), Mahmoudvand *et al.* (2015, 2017), and Mahmoudvand and Rodrigues (2016, 2017)). In particular, Ghodsi and Yarmohammadi (2014) and Beneki and Yarmohammadi (2014) evaluated the forecasting performance of neural networks (NN), and univariate singular SSA, for forecasting exchange rates in some countries. They concluded that SSA is able to outperform NN. In addition, Hassani *et al.* (2009b) used three time series of daily exchange rates: UK Pound/US Dollar, Euro/US Dollar and Japanese yen/US Dollar, and found that the multivariate singular spectrum analysis (MSSA) predictions compare favourably to the random walk (RW) predictions, both for predicting the value and the direction of changes in the exchange rate.

In this paper we compare the performances of SSA and MSSA in forecasting exchange rates. The differences between this study and SSA-based related works are as follows:

- The studies by Ghodsi and Yarmohammadi (2014) and Beneki and Yarmohammadi (2014) used only the univariate SSA, whereas we consider both univariate and multivariate SSA.
- The study by Hassani *et al.* (2009) used both univariate and multivariate SSA, but they considered only one multivariate SSA forecasting algorithm, whereas we apply four multivariate SSA algorithms to produce forecasts.

The rest of this paper is organised as follows: Section 2 gives a brief description of MSSA and its forecasting algorithms. Section 3 presents a comparison between SSA and MSSA with a real data set based on daily currency exchange rates in four of the BRICS emerging economies: Brazil, India, China and South Africa. We finish the paper by a summary conclusion in Section 4.

2. MULTIVARIATE SINGULAR SPECTRUM ANALYSIS

In this section we provide a brief description of MSSA. A more detailed theoretical description can be found, for example, in Hassani and Mahmoudvand (2013).

Let $Y_t = \left[y_t^{(1)}, ..., y_t^{(M)}\right]$, $t = 1, ..., T$, denote a sample of a M-variate time series with length T. We assume that the M-variate time series with T observations $\mathbf{Y}_T$, whose rows are $Y_1, ..., Y_T$, can be written in terms of a signal plus noise model as $\mathbf{Y}_T = \mathbf{S}_T + \mathbf{N}_T$, where $\mathbf{S}_T$ and $\mathbf{N}_T$ are the corresponding matrices containing the signal and noise, respectively. Then, basic version of MSSA can be divided in six steps, as briefly described below.

Step 1: Embedding. The results of this step is a block Hankel trajectory matrix $\mathbf{X}$. Denote by $\mathbf{X}^{(m)}, m = 1, ..., M$, the Hankel matrix associated to the m^{th} time series, $y_1^{(m)}, ..., y_T^{(m)}$. Using window length L, where $2 \leq L \leq T$, and considering $k = T - L + 1$, we have:

$$\mathbf{X}^{(m)} = \begin{bmatrix} y_1^{(m)} & y_2^{(m)} & y_3^{(m)} & ... & y_k^{(m)} \\ y_2^{(m)} & y_3^{(m)} & y_4^{(m)} & ... & y_{k+1}^{(m)} \\ y_3^{(m)} & y_4^{(m)} & y_5^{(m)} & ... & y_{k+2}^{(m)} \\ \vdots & \vdots & \vdots & ... & \vdots \\ y_L^{(m)} & y_{L+1}^{(m)} & y_{L+2}^{(m)} & ... & y_T^{(m)} \end{bmatrix}. \tag{2.1}$$

The trajectory matrix $\mathbf{X}$ in MSSA can be defined by stacking the trajectory matrices horizontally or vertically, i.e.

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(M)} \end{bmatrix} \qquad \text{or} \qquad \mathbf{X} = \left[\mathbf{X}^{(1)} \ ... \ \mathbf{X}^{(M)}\right]. \tag{2.2}$$

A similar procedure can be done to transform matrices $\mathbf{S}_T$ and $\mathbf{N}_T$ in the block Hankel matrices $\mathbf{S}$ and $\mathbf{N}$, respectively. Let

$$\mathbf{S}^{(m)} = \begin{bmatrix} s_1^{(m)} & s_2^{(m)} & s_3^{(m)} & ... & s_k^{(m)} \\ s_2^{(m)} & s_3^{(m)} & s_4^{(m)} & ... & s_{k+1}^{(m)} \\ s_3^{(m)} & s_4^{(m)} & s_5^{(m)} & ... & s_{k+2}^{(m)} \\ \vdots & \vdots & \vdots & ... & \vdots \\ s_L^{(m)} & s_{L+1}^{(m)} & s_{L+2}^{(m)} & ... & s_T^{(m)} \end{bmatrix} \tag{2.3}$$

and

$$\mathbf{N}^{(m)} = \begin{bmatrix} n_1^{(m)} & n_2^{(m)} & n_3^{(m)} & ... & n_k^{(m)} \\ n_2^{(m)} & n_3^{(m)} & n_4^{(m)} & ... & n_{k+1}^{(m)} \\ n_3^{(m)} & n_4^{(m)} & n_5^{(m)} & ... & n_{k+2}^{(m)} \\ \vdots & \vdots & \vdots & ... & \vdots \\ n_L^{(m)} & n_{L+1}^{(m)} & n_{L+2}^{(m)} & ... & n_T^{(m)} \end{bmatrix}. \tag{2.4}$$

The block Hankel matrix $\mathbf{S}$ can then be defined by stacking the trajectory matrices horizontally or vertically, i.e.

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}^{(1)} \\ \vdots \\ \mathbf{S}^{(M)} \end{bmatrix} \qquad \text{or} \qquad \mathbf{S} = \left[\mathbf{S}^{(1)} \ ... \ \mathbf{S}^{(M)}\right], \tag{2.5}$$

and the block Hankel matrix $\mathbf{N}$ can then be defined by stacking the trajectory matrices horizontally or vertically, i.e.

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}^{(1)} \\ \vdots \\ \mathbf{N}^{(M)} \end{bmatrix} \qquad \text{or} \qquad \mathbf{N} = \left[\mathbf{N}^{(1)} \ ... \ \mathbf{N}^{(M)}\right]. \tag{2.6}$$

The MSSA algorithms that use these forms as their trajectory matrix, are called HMSSA and VMSSA, respectively.

Step 2: SVD. In this step, $\mathbf{X}$ will be decomposed by singular value decomposition, as follows:

$$\mathbf{X} = \mathbf{X}_1 + ... + \mathbf{X}_d, \tag{2.7}$$

where $\mathbf{X}_i$'s are unitary matrices and d represents the rank of $\mathbf{X}$. Denoting by $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d \geq 0$ the eigenvalues of $\mathbf{X}\mathbf{X}'$ and $U_1, U_2, ..., U_d$, the corresponding eigenvectors, we have:

$$\mathbf{X}_j = U_j U_j' \mathbf{X} \quad , \quad j = 1, 2..., d.$$

Step 3: Grouping. Considering $\mathbf{X}_i$ to be associated to the i-th largest singular value of $\mathbf{X}$, this step intends to separate the signal and noise components as follows:

$$\mathbf{X} = \underbrace{\mathbf{X}_1 + ... + \mathbf{X}_r}_{\widehat{\mathbf{S}}=Signal} + \underbrace{\mathbf{X}_{r+1} + ... + \mathbf{X}_d}_{\widehat{\mathbf{N}}=Noise}, \tag{2.8}$$

where $r < d$.

Step 4: In this step, using anti-diagonal averaging on each block of $\widehat{\mathbf{S}}$ (see Equation (2.8)), the denoised time series will be reconstructed. We use notation $\widetilde{\mathbf{S}}$ to show the results of this step.

Step 5: The forecast engine of MSSA, which is a linear function of the last L observations of the denoised time series, will be constructed in this step. Details of these engines are given in the next subsection.

Step 6: In this step, h-steps ahead forecasts will be obtained by using the forecast engine.

In general we have four different MSSA forecasting algorithms for MSSA, as shown in Table 1. Computational formulas for these methods are provided in the next subsection.

Table 1: Possible forecasting algorithms for multivariate SSA.

Trajectory form	**Forecasting method**	**Abbreviation**
Horizontal	Recurrent Vector	HMSSA-R HMSSA-V
Vertical	Recurrent Vector	VMSSA-R VMSSA-V

Note that VMSSA and HMSSA can be used for an univariate time series and, in this case, are equivalent and equivalent to the univariate SSA. In fact, there are two different univariate SSA algorithms to obtain forecasts: the recurrent SSA (RSSA) and the vector SSA (VSSA).

2.1. Details about the forecasting engine in MSSA

For simplicity in notation, denote by $\mathbf{Z}[,j]$ and $\mathbf{Z}[i,]$, the j-th column, and the i-th row of the matrix $\mathbf{Z}$, respectively. Denote also $W^h[\ell,]$ the l-th row of $\mathbf{W}^h$. It should be mentioned that the forecasting algorithms presented by Hassani and Mahmoudvand (2013) are based on the recurrent formulas. Here, we obtained a new representation of the algorithm by matrix power. This new representation help us to compute and evaluate the algorithms easier than the forms based on recurrent formulas.

The main idea to construct the forecast engine for MSSA is based on the partitioning of the eigenvector matrix into two parts: the first partition as regressor and the second as response. Then, regressing the second part on the first by the least square method, it produces the forecast model.

Horizontal form

Let $U_j = [u_{1,j}, ..., u_{L,j}]'$, $j = 1, ..., d$, be the j-th eigenvector of $\mathbf{XX}'$. Denote by $\mathbf{U}_r$ the matrix of its first r eigenvectors, corresponding to the r largest singular values of $\mathbf{X}$. We can then do the partition as follows:

$$\mathbf{U}_r = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,r} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,r} \\ \vdots & \vdots & \cdots & \vdots \\ u_{L-1,1} & u_{L-1,2} & \cdots & u_{L-1,r} \\ u_{L,1} & u_{L,2} & \cdots & u_{L,r} \end{bmatrix} \tag{2.9}$$

The gray colour row corresponds to the response and the remaining rows are considered to be the regressors. In the next two subsections we give more details about the HMSSA-R and HMSSA-V.

HMSSA-R

Assume that $\mathbf{U}_r^{\nabla}$ and $\mathbf{U}_{\nabla r}$ are the first $L-1$ rows of $\mathbf{U}_r$ and last row of $\mathbf{U}_r$, respectively (see equation (2.9)). In addition, let us define:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ 0 & \widehat{\mathcal{A}} \end{bmatrix} \quad , \quad \widehat{\mathcal{A}} = \left(1 - \mathbf{U}_{\nabla r}\mathbf{U}_{\nabla r}'\right)^{-1} \mathbf{U}_{\nabla r}\mathbf{U}_r^{\nabla\prime}, \tag{2.10}$$

where $\mathbf{I}$ is the $(L-1) \times (L-1)$ identity matrix and $\mathbf{0}$ is a column vector with $L-1$ zeros. Then, the h-steps ahead forecasts can be obtained by:

$$\widehat{y}_{T+h}^{(m)} = \mathbf{W}^h[L,]\widetilde{\mathbf{S}}[,mK], \quad m = 1, ..., M, \quad h = 1, 2, ... \tag{2.11}$$

The coefficients $\mathbf{W}^h[L,]$ are generated by the whole system of time series, i.e., they consider the correlation among time series. In addition, $\widetilde{\mathbf{S}}[,mK]$ is smoothed again based on the information of all time series. It should be noticed, however, that the forecasts for all individual time series are made by using the same coefficients.

HMSSA-V

Considering the same notation as in HMSSA-R, we can define:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{\Pi} \\ 0 & \widehat{\mathcal{A}} \end{bmatrix} \quad , \quad \mathbf{\Pi} = \mathbf{U}_r^{\nabla}\mathbf{U}_r^{\nabla\prime} + \widehat{\mathcal{A}}'(1 - U_{\nabla r}U_{\nabla r}')\widehat{\mathcal{A}}, \tag{2.12}$$

where $\mathbf{0}$ is column vector with $L-1$ zeros. Then, the h-steps ahead forecasts can be obtained by:

$$\widehat{y}_{T+h}^{(m)} = \frac{1}{L}\sum_{\ell=h}^{h+L-1} W^{\ell}[L-\ell+h,]\widehat{\mathbf{S}}[,mK], \quad m=1,...,M, \quad h=1,2,... \tag{2.13}$$

To better understand how HMSSA-R and HMSSA-V differ, we need to compare Equations (2.11) and (2.13). Note that $\widetilde{\mathbf{S}}$ in Equation (2.11) is obtained by diagonal averaging (see Step 4), and then multiplied by the coefficient $\mathbf{W}^h[L,]$ to produce the forecasts. However, $\widehat{\mathbf{S}}$ in Equation (2.13) is the result of grouping (see Step 3), which is then multiplied by the coefficients $W^{\ell}[L-\ell+h,]$ and the forecasts are produced by averaging.

Both methods, HMSSA-R and HMSSA-V, employ a fixed coefficients for all time series to produce the forecasts. In the approach that considers vertical based methods, we consider different coefficients to produce forecasts for different time series in the multivariate framework. In what follows, we describe how the vertical based methods produce forecasts.

Vertical form

Denote by $\mathbf{U}_r$ the matrix of the first r eigenvectors of $\mathbf{XX}'$ corresponding to the r largest singular values of $\mathbf{X}$. This matrix has dimension $LM \times r$ and we can partitioning as follows:

$$\mathbf{U}_r = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,r} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,r} \\ \vdots & \vdots & \cdots & \vdots \\ u_{L-1,1} & u_{L-1,2} & \cdots & u_{L-1,r} \\ u_{L,1} & u_{L,2} & \cdots & u_{L,r} \\ u_{L+1,1} & u_{L+1,2} & \cdots & u_{L+1,r} \\ u_{L+2,1} & u_{L+2,2} & \cdots & u_{L+2,r} \\ \vdots & \vdots & \cdots & \vdots \\ u_{2L-1,1} & u_{2L-1,2} & \cdots & u_{2L-1,r} \\ u_{2L,1} & u_{2L,2} & \cdots & u_{2L,r} \\ \vdots & \vdots & \cdots & \vdots \\ u_{(M-1)L+1,1} & u_{(M-1)L+1,2} & \cdots & u_{(M-1)L+1,r} \\ u_{(M-1)L+2,1} & u_{(M-1)L+2,2} & \cdots & u_{(M-1)L+2,r} \\ \vdots & \vdots & \cdots & \vdots \\ u_{(ML-1,1} & u_{ML-1,2} & \cdots & u_{ML-1,r} \\ u_{ML,1} & u_{ML,2} & \cdots & u_{ML,r} \end{bmatrix} \tag{2.14}$$

The gray colour rows correspond to the response and the remaining rows are considered to be the regressors. In the next two subsections we give more details about the VMSSA-R and VMSSA-V.

VMSSA-R

Assume that $\mathbf{U}_r^{\triangledown}$ is constructed by removing the rows L, $2L$, ..., ML, from $\mathbf{U}_r$, and $\mathbf{U}_{\triangledown r}$ is the matrix that is constructed by stacking the rows L, $2L$, ..., ML, of $\mathbf{U}_r$ (see equation (2.14)). In addition, let us define:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ 0 & \widehat{\mathcal{A}}_0[1,] \\ \mathbf{0} & \mathbf{I} \\ 0 & \widehat{\mathcal{A}}_0[2,] \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{I} \\ 0 & \widehat{\mathcal{A}}_0[M,] \end{bmatrix} \quad , \quad \widehat{\mathcal{A}} = \left(I_{M\times M} - \mathbf{U}_{\triangledown r}\mathbf{U}_{\triangledown r}'\right)^{-1}\mathbf{U}_{\triangledown r}\mathbf{U}_r^{\triangledown\prime}, \tag{2.15}$$

where $\mathbf{I}$ is the $(L-1)\times(L-1)$ identity matrix, $\mathbf{0}$ is a column vector with $L-1$ zeros and $[0, \widehat{\mathcal{A}}_0[i,]]$ is a vector of size LM where, before each $L-1$ elements of $\widehat{\mathcal{A}}[i,]$, $i = 1, ..., M$, a zero is added. Then, the h-steps ahead forecasts can be obtained by:

$$\widehat{y}_{T+h}^{(m)} = \mathbf{W}^h[mL,]\widetilde{\mathbf{S}}[,K], \quad m = 1, ..., M, \quad h = 1, 2, ... \tag{2.16}$$

VMSSA-V

Considering the notation as in VMSSA-R, we can define:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{\Pi}_1 \\ 0 & \widehat{\mathcal{A}}_0[1,] \\ \mathbf{0} & \mathbf{\Pi}_2 \\ 0 & \widehat{\mathcal{A}}_0[2,] \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{\Pi}_M \\ 0 & \widehat{\mathcal{A}}_0[M,] \end{bmatrix} \quad , \quad \mathbf{\Pi} = \mathbf{U}_r^{\triangledown}\mathbf{U}_r^{\triangledown\prime} + \widehat{\mathcal{A}}'(I_{M\times M} - U_{\triangledown r}U_{\triangledown r}')\widehat{\mathcal{A}}, \tag{2.17}$$

where $\mathbf{0}$ is a column vector with $L-1$ zeros and $\mathbf{\Pi}_j$ represents the rows number $(j-1)(L-1)+1$, ..., $j(L-1)$ of $\mathbf{\Pi}$, $j = 1, ..., M$. Then, the h-steps ahead forecasts can be obtained by:

$$\widehat{y}_{T+h}^{(m)} = \frac{1}{L}\sum_{\ell=h}^{h+L-1} W^{\ell}[mL-\ell+h,]\widehat{S}[,K], \quad m = 1, ..., M, \quad h = 1, 2, ... \tag{2.18}$$

The comparison between VMSSA-R and VMSSA-V is similar to the comparison between HMSSA-R and HMSSA-V, i.e., the part of the time series that is used to produce forecasts in VMSSA-R comes from an diagonal averaging process, whereas the the part of the time series that is used to produce forecasts in VMSSA-V comes from the grouping step which then is subjected to a weighted average.

2.2. MSSA choices

There are two main decisions the user has to make while fitting a MSSA model: the window length, L, and the number of singular values used to reconstruct the series and to construct the forecast engine, r. Despite of the importance of these choices, there have been just a few studies about these choices in the multivariate case. Regarding to the window length, Hassani and Mahmoudvand (2013) showed that a value close the $MT/(M+1)$ and $T/(M+1)$ is optimal for HMSSA and VMSSA, respectively. There are also several studies in the univariate case that can be used similarly to find a suitable value for the multivariate case (see for example Golyandina *et al.* (2001) and Golyandina and Zhigljavsky (2013)). A weighted correlation and screen plots of the singular values are among the simplest ways to find a proper value for r.

2.3. Prediction intervals for MSSA forecasts

Prediction intervals can be very useful in assessing the quality of the forecasts. There are two different types of prediction interval for SSA forecasts, but here we will focus on the bootstrap based method. More details can be found in Golyandina *et al.* (2001) and Golyandina and Zhigljavsky (2013). To obtain the bootstrap prediction interval for the h-steps-ahead forecast, the first step is to obtain the MSSA decomposition $\mathbf{Y}_T = \widetilde{\mathbf{S}}_T + \widetilde{\mathbf{N}}_T$. Then, we simulate p independent copies $\widetilde{\mathbf{N}}_{T,i}$, $i = 1, ..., p$, of the residual series $\mathbf{N}_T$. Adding each of these residual series to the signal series $\widetilde{\mathbf{S}}_T$, we get p time series $\mathbf{Y}_{T,i} = \widetilde{\mathbf{S}}_T + \widetilde{\mathbf{N}}_{T,i}$. Applying the MSSA forecasting algorithm, keeping unchanged the window length L and the number r of eigenvalues/eigenvectors used for reconstruction, to the series $\mathbf{Y}_{T,i}$, $i = 1, ..., p$, we can obtain p forecasting results h-steps-ahead $\widehat{y}^{(m)}_{T+h,i}$, $m = 1, ..., M$. The empirical $\alpha/2$ and $1-\alpha/2$ quantiles of the p h-steps-ahead forecasts $\widehat{y}^{(m)}_{T+h,1}, ...\widehat{y}^{(m)}_{T+h,p}$, correspond to the bounds of the bootstrap prediction interval with confidence level $1-\alpha$.

3. NUMERICAL RESULTS

3.1. Description of the data

In this section, we consider daily currency exchange rate data for the BRICS countries (Brazil–BRL, Russia–RUB, India–IND, China–CHN and South Africa–RAND). However the complete data from Russia could not be found which made us discard this country from our study, which does not interfere with the results as the recent behaviour is very similar to India. Fourteen years of data, between September 2001 and September 2015, were considered. The data was collected from the Board of Governors of the Federal Reserve System (US) — `https://reserach.stlouisfed.org`. Figure 1 shows the behaviour of the daily exchange rates for the four considered countries, between September 2001 and September 2015, when compared with USD.

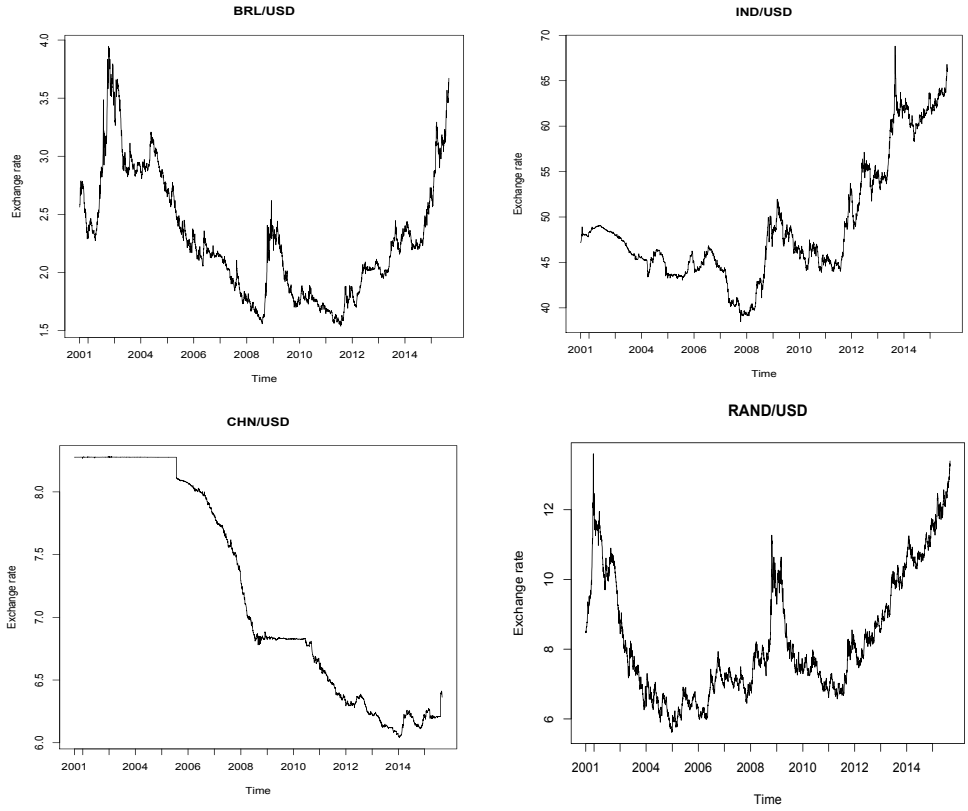

Figure 1: Daily exchange rates for Brazil, India, China and South Africa between September 2001 and September 2015, when compared with USD.

3.2. Preliminary analysis

In this section, we will assess the evidence provided by data in favour of using methods such as MSSA. In particularly, we check stationarity and causality.

Stationarity testing

We use the Augmented Dickey–Fuller method to test for the presence of unit root in the exchange rate time series. Results given in Table 2 indicate that the exchange rates are non-stationary processes. In this way, the non-stationary time series should be differentiated before using a standard time series approach, or we might apply directly methods that do not depend on the stationarity assumption such as SSA and MSSA.

Table 2: Augmented Dickey–Fuller test for the four exchange rates.

	BRL	IND	CHN	RAND
Test statistics	0.256	−1.265	−0.239	−0.962
P-value	0.991	0.889	0.992	0.945

Testing causality

A question that frequently arises in time series analysis is whether one economic variable can help to forecast another economic variable. Here the question is whether one exchange rate time series can help us in forecasting other exchange rate time series and vice versa. One way to address this question was proposed by Granger.

Table 3: Pairwise Granger causality tests.

Series		Null hypothesis:			
		Series 2 does not Granger-Cause Series 1		Series 1 does not Granger-Cause Series 2	
Series 1	Series 2	F-Statistics	P-value	F-Statistics	P-value
BRL	IND	11.95	0.00061	0.50	0.47771
BRL	CHN	1.49	0.22171	1.93	0.16461
BRL	RAND	6.50	0.01081	0.99	0.32012
IND	CHN	9.82	0.00174	0.89	0.34661
IND	RAND	0.04	0.85452	17.89	0.00002
CHN	RAND	0.19	0.66181	12.31	0.00051

The results of this test, for the differentiated time series, are reported in Table 3 for all six pairs of exchange rates series. P-values in Table 3 suggest us to reject all null hypotheses with a significance level of 10%, except one case which has a high P-value. So in general, the exchange rates can help to forecast each other which, again, motivates us to use MSSA.

3.3. Accuracy of forecasts

As it is usual in forecasting literature (see for example Hyndman, 2010), the mean square error (MSE) of forecasts is used to compare the accuracy of the methods under analysis. In order to find reliable values for MSE, we divide the observations into two parts: training and testing sets. Since the length of our data set is large, we decide to produce the results with several different segmentation: 17, 35, 70 and 140 observations for testing sets and remaining for training sets. Note that when considering 35 observations in the testing set, we consider about 99% of the observations (3481 observations) for modelling and the remained 1% are considered for testing.

Let us now explain how we obtain the one-step-ahead forecasts in this case. We considered 3481 observations and find forecasts for the 3482-th observation by all methods. Then we considered 3482 observations and forecast the 3483-th observation by all methods, and repeat until the end of the series (i.e. until observation 3515). In this way, we find 35 predictions for each method that can be compared with the observed values using the MSE. Note that in this way, we begin with 3477 [3472] observations for 5 [10] steps ahead and we only consider the 5-th [10-th] forecasts, in each stage, to compute MSE. The results for 1, 5 and 10 steps ahead forecasts and different sizes of the testing sets are presented in Tables 4, 5, 6 and 7. The results in these tables, indicate a better performance of the MSSA related algorithms when compared with the univariate SSA related algorithms. This improvement of MSSA related algorithms is visible in all time series under consideration, except the 10 steps ahead prediction of the USD/RAND currency.

Table 4: MSE based on 17 forecasts for each combination of forecasting method, time series and number of steps ahead.

Method	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
VMSSA-V	0.0154	0.0228	0.0333	0.2084	**0.7419**	2.0543	0.0137	0.0155	0.0171	0.0115	**0.0234**	0.0699
VMSSA-R	0.0234	0.0338	0.0489	0.2604	0.7553	1.9602	0.0138	0.0152	**0.0162**	0.0119	0.0262	0.0874
HMSSA-V	0.0022	0.0076	0.0110	**0.2048**	0.8737	**1.7994**	**0.0016**	0.0100	0.0186	**0.0113**	0.0423	0.0714
HMSSA-R	**0.0021**	**0.0058**	**0.0109**	0.2065	0.8752	1.8802	**0.0016**	**0.0097**	0.0185	0.0114	0.0372	0.0818
VSSA	0.0025	0.0074	0.0160	0.2946	0.8913	2.0659	0.0017	0.0112	0.0206	0.0138	0.0372	0.0699
RSSA	0.0026	0.0074	0.0119	0.3164	0.8661	2.0407	0.0017	0.0110	0.0206	0.0128	0.0271	**0.0639**

Table 5: MSE based on 35 forecasts for each combination of forecasting method, time series and number of steps ahead.

Method	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
VMSSA-V	0.0159	0.0212	0.0292	**0.1182**	**0.3904**	1.1842	0.0071	0.0081	0.0090	**0.0104**	**0.0254**	0.0816
VMSSA-R	0.0201	0.0267	0.0373	0.1451	0.3934	1.0789	0.0078	0.0084	**0.0089**	0.0166	0.035	0.0891
HMSSA-V	**0.0021**	0.0114	0.0241	0.1289	0.5369	1.0701	**8e-04**	0.0049	0.0092	0.0124	0.0483	0.0854
HMSSA-R	**0.0021**	**0.0104**	**0.0235**	0.1275	0.4897	**1.0522**	**8e-04**	**0.0047**	0.0090	0.0119	0.0400	0.0803
VSSA	0.0032	0.0145	0.0296	0.1692	0.5578	1.3482	**8e-04**	0.0055	0.0101	0.0148	0.0631	0.0892
RSSA	0.0034	0.0121	0.0248	0.1769	0.4725	1.1398	**8e-04**	0.0054	0.0100	0.0133	0.0335	**0.0636**

Table 6: MSE based on 70 forecasts for each combination of forecasting method, time series and number of steps ahead.

Method	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
VMSSA-V	0.0107	0.0136	0.0185	**0.0906**	0.3247	0.8630	0.0043	0.0049	0.0055	0.0325	0.0755	0.1727
VMSSA-R	0.0139	0.0175	0.0231	0.1044	**0.2683**	**0.6190**	0.0059	0.0062	0.0065	0.0443	0.0839	0.1497
HMSSA-V	**0.0019**	0.0091	0.0169	0.1057	0.3628	0.6917	**4e-04**	0.0025	0.0047	0.0131	0.0588	0.1220
HMSSA-R	**0.0019**	**0.0081**	**0.0164**	0.1009	0.3196	0.6519	**4e-04**	**0.0024**	**0.0046**	**0.0124**	0.0507	0.1124
VSSA	0.0023	0.0137	0.0286	0.1189	0.3828	0.9635	**4e-04**	0.0029	0.0053	0.0154	0.0615	0.0882
RSSA	0.0025	0.0097	0.0194	0.1212	0.3087	0.6991	**4e-04**	0.0027	0.0051	0.0146	**0.0425**	**0.0815**

Table 7: MSE based on 140 forecasts for each combination of forecasting method, time series and number of steps ahead.

Method	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
VMSSA-V	0.0248	0.0296	0.0362	**0.0784**	0.2743	0.6804	0.0037	0.0042	0.0048	0.0374	0.0807	0.1589
VMSSA-R	0.0309	0.0361	0.0423	0.0882	**0.2241**	**0.512**	0.0049	0.0052	0.0056	0.0482	0.0869	0.1448
HMSSA-V	**0.0020**	0.0102	0.0202	0.0861	0.3171	0.5706	**2e-04**	0.0015	**0.0026**	0.0145	0.0789	0.1368
HMSSA-R	**0.0020**	**0.0092**	**0.0189**	0.0831	0.2724	0.5334	**2e-04**	**0.0013**	**0.0026**	**0.0141**	0.0708	0.1276
VSSA	0.0022	0.0146	0.0343	0.1117	0.4013	0.8755	**2e-04**	0.0017	0.0033	0.0179	0.0779	**0.0991**
RSSA	0.0023	0.0109	0.0244	0.1128	0.3083	0.604	3e-04	0.0016	0.0029	0.0177	**0.0611**	0.1000

In order to show the gains in MSE, one may compare the ratio of the minimum of MSE by MSSA related algorithms over the minimum of MSE by univariate related algorithms. The results are reported in Table 8. As it can be seen in this table, improvement by MSSA when its MSE compare with univariate SSA, varied between 0.66 to 1.38 and in most of cases MSSA produces an improvement over SSA.

The results for length and coverage ratio for the 95% prediction intervals can be found in Tables 9 and 10, respectively. The performance of both multivariate methods in the horizontal form, HMSSA-R and HMSSA-V, is overall better in terms of coverage ratios, despite having also overall larger length in the prediction intervals. Although the univariate methods give smaller length for the prediction intervals, their coverage ratio is, generally, much worse than the multivariate methods.

Table 8: Ratio of best MSE by MSSA over the best MSE by univariate SSA based on 17, 35, 70 and 140 forecasts and 1, 5 and 10 steps ahead.

Testing size	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
17	0.84	0.78	0.92	0.70	0.86	0.88	0.94	0.88	0.79	0.88	0.86	1.09
35	0.66	0.86	0.95	0.70	0.83	0.92	1.00	0.87	0.89	0.78	0.76	1.26
70	0.83	0.84	0.85	0.76	0.87	0.89	1.00	0.89	0.90	0.85	1.20	1.36
140	0.91	0.84	0.77	0.70	0.73	0.85	1.00	0.81	0.90	0.80	1.16	1.29

Table 9: Length of 95% prediction interval based on 35 forecasts for each combination of forecasting method, time series and number of steps ahead.

Method	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
VMSSA-V	0.292	0.312	0.324	0.363	0.672	0.660	0.271	0.283	0.296	0.187	0.288	0.359
VMSSA-R	0.261	0.362	0.377	0.335	0.544	0.568	0.332	0.355	0.363	0.179	0.249	0.273
HMSSA-V	0.385	0.608	0.625	0.389	0.605	0.644	0.381	0.598	0.647	0.362	0.595	0.650
HMSSA-R	0.365	0.539	0.592	0.370	0.545	0.607	0.346	0.552	0.573	0.356	0.552	0.618
VSSA	0.063	0.111	0.094	0.621	1.188	1.120	0.020	0.035	0.031	0.247	0.339	0.368
RSSA	0.059	0.101	0.088	0.589	1.008	0.911	0.032	0.029	0.001	0.235	0.315	0.344

Table 10: Coverage ratio for 95% prediction interval based on 35 forecasts for each combination of forecasting method, time series and number of steps ahead.

Method	Currency											
	BRL			IND			CHN			RAND		
	1	5	10	1	5	10	1	5	10	1	5	10
VMSSA-V	0.89	0.88	0.79	0.43	0.51	0.33	0.97	0.95	0.96	0.80	0.76	0.67
VMSSA-R	0.92	0.91	0.91	0.67	0.64	0.60	0.79	0.70	0.69	0.77	0.70	0.64
HMSSA-V	0.99	0.97	0.86	0.49	0.40	0.23	0.99	0.99	0.99	0.91	0.77	0.66
HMSSA-R	0.99	0.99	0.86	0.43	0.40	0.23	0.86	0.99	0.99	0.99	0.83	0.66
VSSA	0.26	0.40	0.26	0.63	0.66	0.37	0.66	0.66	0.60	0.69	0.57	0.54
RSSA	0.29	0.31	0.20	0.66	0.63	0.31	0.66	0.63	0.60	0.69	0.66	0.43

4. CONCLUSION

In this paper, we used univariate and multivariate SSA to forecasts the daily exchange rates of Brazil, India, China and South Africa. As a preliminary analysis, we conducted the traditional time series analysis of unit root test and found that all series are non-stationary. We also used Granger test to see whether series support each other. With the exception of the forecasts for 5 and 10 steps ahead for RAND, MSSA outperformed SSA in terms of forecasting accuracy. Accordingly, we can conclude that MSSA can be of great help to forecast exchange rates.

ACKNOWLEDGMENTS

The authors of this paper acknowledge financial support by CAPES — Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Coordination for the Improvement of Higher Education Personnel), Brazil, grant number 88881.062137/2014-01.

REFERENCES

[1] Beneki, C. and Yarmohammadi, M. (2014). Forecasting exchange rates: An optimal approach, *Journal of Systems Science and Complexity*, **27**(1), 21–28.

[2] Della Corte, P.; Sarno, L. and Tsiakas, L. (2011). Spot and forward volatility in foreign exchange, *Journal of Financial Economics*, **100**, 496–513.

[3] Engle, R. and Wesy, K. (2005). Exchange Rates and Fundamentals, *Journal of Political Economy*, **113**(3), 485–517.

[4] Fu, T. (2010). Herding in China equity market, *International Journal of Economics and Finance*, **2**(2), 148.

[5] Ghodsi, M. and Yarmohammadi, M. (2014). Exchange rate forecasting with optimum singular spectrum analysis, *Journal of Systems Science and Complexity*, **27**(1), 47–55.

[6] Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series*, Springer, New York, London.

[7] Golyandina, N.; Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York, London.

[8] Hassani, H. and Mahmoudvand, R. (2013). Multivariate Singular Spectrum Analysis: A General View and New Vector Forecasting Algorithm, *International Journal of Energy and Statistics*, **1**(1), 55–83.

[9] Hassani, H.; Heravi, S. and Zhigljavsky, A. (2009a). Forecasting European industrial production with singular spectrum analysis, *International Journal of Forecasting*, **25**, 103–118.

[10] Hassani, H.; Soofi, A. and Zhigljavsky, A. (2009b). Predicting Daily Exchange Rate with Singular Spectrum Analysis, *Nonlinear Analysis: Real World Applications*, **11**, 2023–2034.

[11] Hassani, H.; Soofi, A.S. and Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis, *Journal of the Royal Statistical Society: Series A*, **176**(3), 743–760.

[12] Hassani, H.; Webster, A.; Silva, E.S. and Heravi, S. (2015). Forecasting US tourist arrivals using optimal singular spectrum analysis, *Tourism Management*, **46**, 322–335.

[13] Hyndman, R.J. (2010). Why every statistician should know about cross-validation, `http://robjhyndman.com/researchtips/crossvalidation/`

[14] Janetzko, D. (2014). Using Twitter to Model the ERO/USD Exchange rate, arXiv preprint arXiv:1402.1624.

[15] Kapl, M. and Müller W.G. (2010). Prediction of steel prices: A comparison between a conventional regression model and MSSA, *Statistics and Its Interface*, **3**(3), 369–375.

[16] Dilmaghani, A.K. and Tehranchian, A.M. (2015). The Impact of Monetary Policies on the Exchange Rate: A GMM Approach, *Iranian Economic Review*, **19**(2), 177–191.

[17] Lin, C.S.; Chiu, S.H. and Lin, T.Y. (2012). Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting, *Economic Modelling*, **29**, 2583–2590.

[18] Mahmoudvand, R.; Konstantinides, D. and Rodrigues, P.C. (2017). Forecasting Mortality Rate by Multivariate Singular Spectrum Analysis, *Applied Stochastic Models in Business and Industry*, DOI: 10.1002/asmb.2274.

[19] Mahmoudvand, R. and Rodrigues, P.C. (2017). A New Parsimonious Recurrent Forecasting Model in Singular Spectrum Analysis, *Journal of Forecasting*, DOI: 10.1002/for.2484.

[20] Mahmoudvand, R. and Rodrigues, P.C. (2016). Missing value imputation in time series using Singular Spectrum Analysis, *International Journal of Energy and Statistics*, **4**, DOI: 10.1142/S2335680416500058.

[21] Mahmoudvand, R.; Alehosseini, F. and Rodrigues, P.C. (2015). Mortality Forecasting with Singular Spectrum Analysis, *REVSTAT — Statistical Journal*, **13**, 193–206.

[22] Papaioannou, P.; Russo, L.; Papaioannou, G. and Siettos, C.I. (2013). Can social microblogging be used to forecast intraday exchange rates? *Netnomics*, **14**, 47–68.

[23] Patterson, K.; Hassani, H.; Heravi, S. and Zhigljavsky, A. (2011). Forecasting the final vintage of the industrial production series, *Journal of Applied Statistics*, **38**(10), 2183–2211.

[24] Plakandaras, V. (2015). *Forecasting financial time series with machine learning techniques*, Ph.D. Thesis, Department Of Economics, Democritus University of Thrace, Greece.

[25] Stepanov, D. and Golyandina, N. (2005). SSA-based approaches to analysis and forecast of multidimensional time series. In: *Proceedings of the 5th St. Petersburg Workshop on Simulation, June 26–July 2, 2005*, St. Petersburg State University, St. Petersburg, 293–298.

[26] Yu, L.; Wang, S. and Lai, K.K. (2007). *Foreign-Exchange Rate Forecasting with Artificial Neural Networks*, Springer, New York.

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 January, one in April, one in July and the other in October.

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.

REVSTAT is published, in English, by Statistics Portugal.

REVSTAT - Statistical Journal is an open access peer reviewed journal.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews® (MathSciNet®)
- Science Citation Index Expanded
- Zentralblatt für Mathematic
- Scimago Journal & Country Rank
- Scopus

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and submitted in PDF format to the Editor-in chief (revstat@fc.ul.pt) and to the executive editor (revstat@ine.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, published papers and forthcoming papers are available at REVSTAT website.

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.

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal website (http://www.ine.pt).

After assigning copyright, authors may use their own material in other publications provided that REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Editorial Board