ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Special issue on
«Distribution Theory, Estimation and Inference»
in honor of Professors Bimal Sinha and Bikas Sinha 70th Birthday

Guest Editors:
Carlos A. Coelho
Filipe J. Marques

Volume 15, No.2
April 2017

REVSTAT

Statistical Journal

Catalogação Recomendada

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

CREDITS

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

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, i-ii

Foreword

This issue of REVSTAT is dedicated to Professors Bimal K. Sinha and Bikas K. Sinha on the occasion of their 70th birthday.

Professors Bimal K. Sinha and Bikas K. Sinha [Statistical Twins – as mentioned in the IMS Bulletin, 1990] were born on the 16th of March, 1946 in undivided India in the village of Atgharia, District Pabna, now in Bangladesh. The Sinha family migrated to India in the 1950's. Along the years Sinha Brothers emerged as two very distinct Statisticians, taking distinct paths in their professional lives.

Bimal K. Sinha got his Ph.D. in 1973 from the University of Calcutta, India. A former faculty of the Indian Statistical Institute and the University of Pittsburgh, Bimal Sinha joined UMBC (University of Maryland Baltimore County) in 1985 as a Professor of Statistics and Founder of Statistics Graduate Program. Professor Sinha received the prestigious recognition of UMBC's Presidental Research Professorship and also the University System of Maryland Board of Regents Research Professorship. A Fellow of both the Institute of Mathematical Statistics (IMS) and the American Statistical Association (ASA), Professor Sinha is the author of four books, more than 150 research articles, and supervised about 30 PhD students. For over twenty years, Professor Sinha worked on many aspects of environmental statistics, and his recent research focuses on 'Data analysis under confidentiality protection' at the US Census Bureau.

Bikas K. Sinha got his Ph.D. in 1972 also from the University of Calcutta, India. A former faculty of the Calcutta University Department of Statistics (1972–1975) and a Visiting Faculty at the Institute of Mathematics, Federal University of Bahia, Brazil (1975–1979), Bikas Sinha joined the world famous Indian Statistical Institute at its Calcutta Centre as an Associate Professor of Statistics (1979–1984) and was promoted to Professor of Statistics (1985–2011). Bikas Sinha served extensively several universities in the USA as Visiting Scientist/Visiting Professor with teaching and research assignments. He has collaborated with statisticians within India and abroad as also spanning over several continents with an impressive record of more than 90 research collaborators. An Expert on Missions of the United Nations, Professor Bikas Sinha is the author/co-author of Springer-Verlag Monographs and Graduate-Level Text Books (Wiley & Sage) and has over 140 published research articles. His expertise centers around Optimal Experimental Designs and Finite Population Inference. He had one doctoral student each at Calcutta University and ISI, Calcutta. Besides, numerous students within India and abroad sought his academic mentorship for their research studies.

Professor Bimal K. Sinha served on Editorial Boards in a number of statistical Journals. He was co-editor of Sankhya, Series A (1981–1997) and Series B (1998–1999), Editorial Board Member of the Journal of Multivariate Analysis (1981–1982) and of the Handbook of Statistics, Vol. 2, Editor of the Calcutta Statistical Association Bulletin (2003–2007), Associate Editor of the Journal of Applied Statistical Science (1992 on), of the Journal of Environmental Statistics (1992 on), of the Journal of Environmental Modeling and Assessment (1994 on), of Statistics and Decisions (1993–2001) and of the Journal of Statistical Planning and Inference (1994–2002).

In addition to the awards and distinctions already mentioned above, Professor Bimal Sinha is also an Elected Member of the International Statistical Institute, and received in 2002 the American Statistical Association, Section on Statistics and the Environment, Distinguished Achievement Medal Award and in 2012 the University System of Maryland Board of Regents Award for Excellence in Research.

Professor Bikas K. Sinha was also a member of the Editorial Board of a large number of statistical Journals. He served as Editorial Board Member of Sankhya (1985–1995), as Associate Editor of the Journal of Statistical Planning and Inference (1992–1997) and of the Calcutta Statistical Association Bulletin (1988–1992) where he also served as Editor, from 1993 to 2000. He also had editorship duties in the Pakistan Journal of Statistics, International Journal of Statistical Science, Statistical Methodology, Statistics, Statistics & Decisions and the Journal of Combinatorics & Information System Sciences.

Professor Bikas Sinha also received a number of awards, among which are the P. C. Mahalanobis Gold Medal in 1980. He is an Elected Member of the International Statistical Institute since 1985, and was President of the Indian Science Congress, Statistics Section (2000–2001) and Member of the National Statistical Commission of the Government of India (2006–2009). In the recent past, Professor Bikas Sinha has been a regular visitor at the US Bureau of Census under the 'Summer at Census' visitor plan.

Carlos A. Coelho
Filipe J. Marques

INDEX

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

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 155-186

INFERENCE FOR MULTIVARIATE REGRESSION MODEL BASED ON SYNTHETIC DATA GENERATED UNDER FIXED-POSTERIOR PREDICTIVE SAMPLING: COMPARISON WITH PLUG-IN SAMPLING

Authors: Ricardo Moura
– CMA, Faculty of Sciences and Technology, Nova University of Lisbon Portugal (rp.moura@campus.fct.unl.pt)

Martin Klein*
– Center for Statistical Research and Methodology, U.S. Census Bureau, U.S.A. (martin.klein@census.gov)

Carlos A. Coelho
– CMA and Mathematics Department, Faculty of Sciences and Technology, Nova University of Lisbon, Portugal (cmac@fct.unl.pt)

Bimal Sinha*
– Department of Mathematics and Statistics,
University of Maryland, Baltimore County
and Center for Disclosure Avoidance Research, U.S. Census Bureau
U.S.A. (sinha@umbc.edu)

Received: March 2016 Revised: May 2016 Accepted: July 2016

Abstract:

- The authors derive likelihood-based exact inference methods for the multivariate regression model, for singly imputed synthetic data generated via Posterior Predictive Sampling (PPS) and for multiply imputed synthetic data generated via a newly proposed sampling method, which the authors call Fixed-Posterior Predictive Sampling (FPPS). In the single imputation case, our proposed FPPS method concurs with the usual Posterior Predictive Sampling (PPS) method, thus filling the gap in the existing literature where inferential methods are only available for multiple imputation. Simulation studies compare the results obtained with those for the exact test procedures under the Plug-in Sampling method, obtained by the same authors. Measures of privacy are discussed and compared with the measures derived for the Plug-in Sampling method. An application using U.S. 2000 Current Population Survey data is discussed.

Key-Words:

- *finite sample inference; maximum likelihood estimation; pivotal quantity; plug-in sampling; statistical disclosure control; unbiased estimators.*

AMS Subject Classification:

- 62H10, 62H15, 62H12, 62J05, 62F10, 62E15, 62E10, 62E17, 62D99.

*__Disclaimer:__ This article is released to inform interested parties of ongoing research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1. INTRODUCTION

When releasing microdata to the public, methods of statistical disclosure control (SDC) are used to protect confidential data, that is "data which allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information" (Regulation No. 223/2009 of the European Parliament [7]), while enabling valid statistical inference to be drawn on the relevant population. SDC methods include data swapping, additive and multiplicative noise, top and bottom coding, and also the creation of synthetic data. In this paper, the authors provide inferential tools for the statistical analysis of a singly imputed synthetic dataset when the real dataset cannot be released. The multiple imputation case is also addressed, using a new adapted method of generating synthetic data, which the authors call Fixed-Posterior Predictive Sampling (FPPS).

The use of synthetic data for SDC started with Little [4] and Rubin [10] using multiple imputation (Rubin [9]). Reiter [8] was the first to present methods for drawing inference based on partially synthetic data. Moura et al. [5] complemented this work with the development of likelihood-based exact inference methods for both single and multiple imputation, that is, inferential procedures developed based on exact distributions, and not on asymptotic results, in the case where synthetic datasets were generated via Plug-in Sampling. The procedures of Reiter [8] are general in that they can be applied to a variety of estimators and statistical models, but these procedures are only applicable in the multiple imputation case, and are based on large sample approximations.

There are two major objectives in the present research. First, to make available likelihood-based exact inference for singly imputed synthetic data via Posterior Predictive Sampling (PPS) where the usual available procedures are not applicable, therefore extending the work of Klein and Sinha [2], under the multivariate linear regression (MLR) model. Second, to propose a different approach for release of multiple synthetic datasets, FPPS, which can use a similar way of gathering information from the synthetic datasets to that used in [5], when these synthetic datasets are generated via the Plug-in Sampling method. This second objective arises from the fact that when using the classical PPS it is too hard to construct an exact joint probability density function (pdf) for the estimators, under the MLR model, since one would face the problem of deriving the distribution of a sum of variables that follow Wishart distributions with different parameter matrices. It is with this problem in mind, that we propose an adapted method that we will call the FPPS method. We show that this method offers a higher level of confidentiality than the Plug-in Sampling method, and it still allows one to draw inference for the unknown parameters using a joint pdf of the proposed estimators.

A brief description of the PPS and FPPS methods follows. Suppose that $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$ are the original data which are jointly distributed according to the pdf $f_{\boldsymbol{\theta}}(\mathbf{Y})$, where $\boldsymbol{\theta}$ is the unknown (scalar, vector or matrix) parameter. A prior $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ is assumed and then the posterior distribution of $\boldsymbol{\theta}$ is obtained as $\pi(\boldsymbol{\theta}|Y) \propto \pi(\boldsymbol{\theta}) f_{\boldsymbol{\theta}(x)}$, and used to draw a replication $\boldsymbol{\theta}_f^{\bullet}$ of $\boldsymbol{\theta}$, when applying the FPPS, or draw $M \geq 1$ independent replications $\boldsymbol{\theta}_1^{\bullet}, ..., \boldsymbol{\theta}_M^{\bullet}$ of $\boldsymbol{\theta}$, when applying the PPS. In the case of FPPS, we generate M replicates of $\mathbf{Y}$, namely, $\mathbf{W}_j = (\mathbf{w}_{j1}, ..., \mathbf{w}_{jn})$, $j = 1, ..., M$ drawn all independently from the same $f_{\boldsymbol{\theta}_f^{\bullet}}$, where $f_{\boldsymbol{\theta}_f^{\bullet}}$ is the joint pdf of the original $\mathbf{Y}$ with $\boldsymbol{\theta}_f^{\bullet}$ replacing the unknown $\boldsymbol{\theta}$. In the case of the usual PPS method for each j-th generated synthetic dataset we would use the corresponding j-th posterior draw $\boldsymbol{\theta}_j^{\bullet}$ and corresponding j-th joint pdf's $f_{\boldsymbol{\theta}_j^{\bullet}}$, for $j = 1, ..., M$. In either case, these synthetic datasets $\mathbf{W}_1, \ldots, \mathbf{W}_M$ will be the datasets available to the general public. One may observe that, for $M = 1$, the Posterior Predictive Sampling and Fixed-Posterior Predictive Sampling methods concur.

Regarding the MLR model, in our context, we consider the *sensitive* response variables y_j $(j = 1, ..., m)$ forming the vector of response variables $\mathbf{y} = (y_1, ..., y_m)'$, and a set of p non-*sensitive* explanatory variables $\mathbf{x} = (x_1, ..., x_p)'$. It is assumed that $\mathbf{y}|\mathbf{x} \sim N_m(\mathbf{B}'\mathbf{x}, \boldsymbol{\Sigma})$, with $\mathbf{B}$ and $\boldsymbol{\Sigma}$ unknown, and the original data consist of $\mathcal{Y} = \{(y_{1i}, ..., y_{mi}, x_{1i}, ..., x_{pi}), i = 1, ..., n\}$, where n will be the sample size. Let us consider $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$ with $\mathbf{y}_i = (y_{1i}, ..., y_{mi})'$ and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)$ with $\mathbf{x}_i = (x_{1i}, ..., x_{pi})'$. We assume $rank(\mathbf{X} : p \times n) = p < n$ and $n \geq m + p$. Therefore the following regression model is considered

$$\mathbf{Y}_{m\times n} = \mathbf{B}'_{m\times p}\mathbf{X}_{p\times n} + \mathbb{E}_{m\times n}, \tag{1.1}$$

where $\mathbb{E}_{m\times n}$ is distributed as $N_{mn}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$. Based on the original data,

$$\hat{\mathbf{B}} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{Y}' \tag{1.2}$$

is the Maximum Likelihood Estimator (MLE) and the Uniformly Minimum-Variance Unbiased Estimator (UMVUE) of $\mathbf{B}$, distributed as $N_{pm}(\mathbf{B}, \boldsymbol{\Sigma}\otimes(\mathbf{X}\mathbf{X}')^{-1})$, independent of $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}(\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X})(\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X})'$ which is the MLE of $\boldsymbol{\Sigma}$, with $n\hat{\boldsymbol{\Sigma}} \sim W_m(\boldsymbol{\Sigma}, n - p)$. Therefore

$$\mathbf{S} = \frac{n\hat{\boldsymbol{\Sigma}}}{n - p} \tag{1.3}$$

will be the UMVUE of $\boldsymbol{\Sigma}$.

The organization of the paper is as follows. In Section 2, based on singly and multiply imputed synthetic datasets generated via Fixed-Posterior Predictive Sampling, two procedures are proposed to draw inference for the matrix of regression coefficients. Under the single imputation case, we recall that the FPPS and the PPS methods coincide. The test statistics proposed will be pivot statistics, different from the classical test statistics for $\mathbf{B}$ under the MLR model (see [1, Secs 8.3 and 8.6]) since it is shown that these classical test statistics are not pivotal in the present context. Section 3 presents some simulations in order to check the accuracy of theoretically derived results. Also in this section, the authors use a measure for the *radius* (distance between the center and the edge) of the confidence sets for the regression coefficients adapted from [5], computed for the original data and also for the synthetic data generated via FPPS. These *radius* measures are compared with the ones obtained when synthetic datasets are generated via Plug-in Sampling. Section 4 presents data analyses under the proposed methods in the context of public use data from the U.S. Current Population Survey comparing with the same data analysis given by [5] under the Plug-in Sampling method. In Section 5, we compare the level of privacy protection obtained via our FPPS method and via Plug-in Sampling method. Some concluding remarks are added in Section 6. Proofs of the theorems, and other technical derivations are presented in Appendices A and B.

2. ANALYSIS FOR SINGLE AND MULTIPLE IMPUTATION

In this section, we present two new exact likelihood-based procedures for the analysis of synthetic data generated using Fixed-Posterior Predictive Sampling method, under the MLR model in (1.1). For the single imputation case, the two new procedures developed also offer the possibility of drawing inference for a single synthetic dataset generated via Posterior Predictive Sampling.

2.1. A First New Procedure

In this subsection, the synthetic data will consist of M synthetic versions of $\mathbf{Y}$ generated based on the FPPS method.

Consider the joint prior distribution $\pi(\mathbf{B},\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\alpha/2}$, leading to the posterior distributions for $\boldsymbol{\Sigma}$ and $\mathbf{B}$

$$\boldsymbol{\Sigma}|_{\mathcal{Y},\mathbf{S}} \sim W_m^{-1}((n-p)\mathbf{S}, n+\alpha-p) \tag{2.1}$$

and

$$\mathbf{B}|_{\mathcal{Y},\boldsymbol{\Sigma}} \sim N_{pm}(\hat{\mathbf{B}}, \boldsymbol{\Sigma}\otimes(\mathbf{X}\mathbf{X}')^{-\mathbf{1}}), \tag{2.2}$$

where we assume that $n+\alpha > p+m+1$ (see proof in Appendix B.1). Consequently, we draw $\tilde{\boldsymbol{\Sigma}}$ from (2.1) and $\tilde{\mathbf{B}}$ from (2.2), upon replacing $\boldsymbol{\Sigma}$ by $\tilde{\boldsymbol{\Sigma}}$ in this latter expression. We then generate the M synthetic datasets, denoted as $\mathbf{W}_j = (\mathbf{w}_{j1}, ..., \mathbf{w}_{jn})$, for $j = 1, ..., M$, where $\mathbf{w}_{ji} = (w_{1ji}, ..., w_{mji})'$, are independently distributed as

$$\mathbf{w}_{ji}|_{\tilde{\mathbf{B}},\tilde{\boldsymbol{\Sigma}}} \sim N_m(\tilde{\mathbf{B}}'\mathbf{x_i}, \tilde{\boldsymbol{\Sigma}}), \quad i = 1, ..., n, j = 1, ..., M. \tag{2.3}$$

For $i = 1, ..., n$ and $j = 1, ..., M$, let $\mathbf{B}^\bullet_j = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{W}'_j$ and $\mathbf{S}^\bullet_j = \frac{1}{n-p}(\mathbf{W}_j - \mathbf{B}^{\bullet\prime}_j\mathbf{X})(\mathbf{W}_j - \mathbf{B}^{\bullet\prime}_j\mathbf{X})'$ be the estimators of $\mathbf{B}$ and $\boldsymbol{\Sigma}$, based on the synthetic data $(w_{1ji}, ..., w_{mji}, x_{1i}, ..., x_{pi})$, which by Lemma 1.1 in [5] are jointly sufficient.

Conditional on $(\tilde{\mathbf{B}}, \tilde{\mathbf{\Sigma}})$, for every $j = 1, ..., M$, $\mathbf{B}_j^\bullet$ is independent of $\mathbf{S}_j^\bullet$ and $\{(\mathbf{B}_1^\bullet, \mathbf{S}_1^\bullet), ..., (\mathbf{B}_M^\bullet, \mathbf{S}_M^\bullet)\}$ are jointly sufficient estimators for $\mathbf{B}$ and $\mathbf{\Sigma}$. Define then

$$\overline{\mathbf{B}}_M^\bullet = \frac{1}{M}\sum_{j=1}^{M} \mathbf{B}_j^\bullet \quad \text{and} \quad \overline{\mathbf{S}}_M^\bullet = \frac{1}{M}\sum_{j=1}^{M} \mathbf{S}_j^\bullet, \tag{2.4}$$

which are also mutually independent, given $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{\Sigma}}$. For $p \geq m$ and $n + \alpha > p + 2m + 2$, we derive the following main results.

1. The MLE of $\mathbf{B}$ is $\overline{\mathbf{B}}_M^\bullet$, which is unbiased for $\mathbf{B}$, with $Var(\overline{\mathbf{B}}_M^\bullet) = N_{M,n,m,p,\alpha}\mathbf{\Sigma} \otimes (\mathbf{XX}')^{-1}$, where $N_{M,n,m,p,\alpha} = \frac{2M(n+\frac{\alpha}{2}-p-m-1)+n-p}{M(n+\alpha-p-2m-2)}$ (see Theorem 2.1 and Appendix B.3).

2. An unbiased estimator (UE) of $\mathbf{\Sigma}$ will be $\hat{\mathbf{S}}_M = \frac{n+\alpha-p-2m-2}{n-p}\overline{\mathbf{S}}_M^\bullet$ (see Theorem 2.1 and Appendix B.3); for $\alpha = 2m+2$, $\overline{\mathbf{S}}_M^\bullet$ will also be an UE for $\mathbf{\Sigma}$,

3. In Theorem 2.2 (see below), we prove that

$$T_M^\bullet = \frac{|(\overline{\mathbf{B}}_M^\bullet - \mathbf{B})'(\mathbf{XX}')(\overline{\mathbf{B}}_M^\bullet - \mathbf{B})|}{|M(n-p)\overline{\mathbf{S}}_M^\bullet|}, \tag{2.5}$$

   a statistic somewhat related with the Hotelling T^2, this one built to make inference on a matrix parameter, is a pivotal quantity, and that for $\mathbf{A}_1 \sim W_m(\mathbf{I}_m, n+\alpha-p-m-1)$, $\mathbf{A}_2 \sim W_m(\mathbf{I}_m, n-p)$ and $F_i \sim F_{p-i+1,M(n-p)-i+1}$ $(i = 1, ..., m)$, all independent random variables,

$$T_M^\bullet|\mathbf{\Omega} \overset{st}{\sim} \left\{\prod_{i=1}^{m} \frac{p-i+1}{M(n-p)-i+1}F_i\right\}\left|\frac{M+1}{M}\mathbf{I}_m + \mathbf{\Omega}\right|,$$

   where $\mathbf{\Omega}$ has the same distribution as $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{-1}\mathbf{A}_1^{\frac{1}{2}}$ and where $\overset{st}{\sim}$ means 'stochastic equivalent to'.

4. If one wants to test a linear combination of the parameters in $\mathbf{B}$, namely, $\mathbf{C} = \mathbf{AB}$ where $\mathbf{A}$ is a $k \times p$ matrix with $rank(\mathbf{A}) = k \leq p$ and $k \geq m$, one defines

$$T_{M,\mathbf{C}}^\bullet = \frac{|(\mathbf{A}\overline{\mathbf{B}}_M^\bullet - \mathbf{C})'(\mathbf{A}(\mathbf{XX}')^{-1}\mathbf{A}')^{-1}(\mathbf{A}\overline{\mathbf{B}}_M^\bullet - \mathbf{C})|}{|M(n-p)\overline{\mathbf{S}}_M^\bullet|}$$

   and proceeds by noting that

$$T_{M,\mathbf{C}}^\bullet|\mathbf{W} \overset{st}{\sim} \left\{\prod_{i=1}^{m} \frac{k-i+1}{M(n-p)-i+1}F_{k,i}\right\}\left|\frac{M+1}{M}\mathbf{I}_m + \mathbf{\Omega}\right|, \tag{2.6}$$

   with $F_{k,i} \sim F_{k-i+1,M(n-p)-i+1}$ being independent random variables and $\mathbf{\Omega}$ defined as in the previous item.

(i) *Test for the significance of* $\mathbf{C}$*:* in order to test $H_0 : \mathbf{C} = \mathbf{C}_0$ versus $H_1 : \mathbf{C} \neq \mathbf{C}_0$, we reject H_0 whenever $T^{\bullet}_{M,\mathbf{C}_0}$ exceeds $\delta_{M,k,m,p,n;\gamma}$ where $\delta_{M,k,m,p,n;\gamma}$ satisfies $(1-\gamma) = Pr(T^{\bullet}_{M,\mathbf{C}_0} \leq \delta_{M,k,m,p,n;\gamma})$ when H_0 is true. To perform a test for $\mathbf{B} = \mathbf{B}_0$ one has to take $\mathbf{A} = \mathbf{I}_p$.

(ii) *Confidence set for* $\mathbf{C}$*:* a $(1-\gamma)$ level confidence set for $\mathbf{C}$ is given by

$$\Delta_M(\mathbf{C}) = \{\mathbf{C} : T^{\bullet}_{M,\mathbf{C}} \leq \delta_{M,k,m,n,p;\gamma}\}, \tag{2.7}$$

where the value of $\delta_{M,k,m,n,p;\gamma}$ can be obtained by simulating the distribution in (2.6).

Results in 1-4 are derived based on Theorems 2.1 and 2.2 below.

Theorem 2.1. The joint pdf of $\overline{\mathbf{B}}^{\bullet}_M, \overline{\mathbf{S}}^{\bullet}_M$ and $\tilde{\mathbf{\Sigma}}^{-\mathbf{1}}$, for $\overline{\mathbf{B}}^{\bullet}_M$ and $\overline{\mathbf{S}}^{\bullet}_M$ defined in (2.4), is proportional to

$$e^{-\frac{1}{2}tr\{(\frac{\mathbf{M+1}}{\mathbf{M}}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{-1}(\overline{\mathbf{B}}^{\bullet}_M-\mathbf{B})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}^{\bullet}_M-\mathbf{B})+M(n-p)\tilde{\mathbf{\Sigma}}^{-\mathbf{1}}\overline{\mathbf{S}}^{\bullet}_M\}}$$
$$\times \frac{|\overline{\mathbf{S}}^{\bullet}_M|^{\frac{M(n-p)-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{M(n-p)+n+\alpha}{2}-m-1}} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\tfrac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-\mathbf{1}} + \mathbf{\Sigma}^{-\mathbf{1}}|^{-p/2} |\tilde{\mathbf{\Sigma}}^{-\mathbf{1}} + \mathbf{\Sigma}^{-\mathbf{1}}|^{-\frac{2n+\alpha-2p-m-1}{2}},$$

so that $\overline{\mathbf{B}}^{\bullet}_M$ and $\overline{\mathbf{S}}^{\bullet}_M$, given $\tilde{\mathbf{\Sigma}}$, are independent, with

$$\overline{\mathbf{B}}^{\bullet}_M|_{\tilde{\mathbf{\Sigma}}} \sim N_{pm}\left(\mathbf{B}, \left(\frac{M+1}{M}\tilde{\mathbf{\Sigma}} + \mathbf{\Sigma}\right) \otimes (\mathbf{X}\mathbf{X}')^{-1}\right)$$

and

$$\overline{\mathbf{S}}^{\bullet}_M|_{\tilde{\mathbf{\Sigma}}} \sim W_m\left(\frac{1}{M(n-p)}\tilde{\mathbf{\Sigma}}, M(n-p)\right).$$

Proof: See Appendix A. □

Theorem 2.2. The distribution of the statistic $T^{\bullet}_M$ defined in (2.5) can be obtained from the decomposition

$$T^{\bullet}_M|_{\mathbf{\Omega}} \overset{st}{\sim} \left\{\prod_{i=1}^{m} \frac{p-i+1}{M(n-p)-i+1} F_i\right\} \left|\frac{M+1}{M}\mathbf{I}_m + \mathbf{\Omega}\right|$$

where $F_i \sim F_{p-i+1, M(n-p)-i+1}$ are independent random variables, themselves independent of $\mathbf{\Omega}$, which has the same distribution as $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{-1}\mathbf{A}_1^{\frac{1}{2}}$ with $\mathbf{A}_1 \sim W_m(\mathbf{I}_m, n+\alpha-p-m-1)$ and $\mathbf{A}_2 \sim W_m(\mathbf{I}_m, n-p)$, two independent random variables.

Proof: See Appendix A. □

Remark 2.1. When $m = 1$ and $M = 1$, the statistic in (2.5) reduces to the statistic T^2 used in [2] whose pdf is obtained by noting that

$$T^2|_{\Omega=\omega} \sim \frac{p}{n-p}(2+\omega)F_{p,n-p} \quad \text{where} \quad f_\Omega(\omega) \propto \frac{\omega^{\frac{n+\alpha-p-4}{2}}}{(1+\omega)^{\frac{2n+\alpha-2p-2}{2}}}.$$

Remark 2.2. We remark that the statistic $T^\bullet_M$ in (2.5) degenerates towards zero when $n \to \infty$ or $M \to \infty$, but

$$(M(n-p))^m\, T^\bullet_M|_{\mathbf{\Omega}} \xrightarrow[n\to\infty]{d} \left\{\prod_{i=1}^m \chi^2_{p-i+1}\right\} \left|\frac{M+1}{M}\mathbf{I}_m + \mathbf{\Omega}\right|$$

and

$$(M(n-p))^m\, T^\bullet_M|_{\mathbf{\Omega}} \xrightarrow[M\to\infty]{d} \left\{\prod_{i=1}^m \chi^2_{p-i+1}\right\} |\mathbf{I}_m + \mathbf{\Omega}|\,,$$

where $\xrightarrow{d}$ represents convergence in distribution. Consequently, if instead of using $T^\bullet_M$ one uses $T^\bullet_{M2} = (M(n-p))^m\, T^\bullet_M = \frac{|(\overline{\mathbf{B}}^\bullet_M - \mathbf{B})'(XX')(\overline{\mathbf{B}}^\bullet_M - \mathbf{B})|}{|\mathbf{S}^\bullet_M|}$ one would have

$$T^\bullet_{M2}|_{\mathbf{\Omega}} \xrightarrow[n\to\infty]{d} \left\{\prod_{i=1}^m \chi^2_{p-i+1}\right\} \left|\frac{M+1}{M}\mathbf{I}_m + \mathbf{\Omega}\right|$$

and

$$T^\bullet_{M2}|_{\mathbf{\Omega}} \xrightarrow[M\to\infty]{d} \left\{\prod_{i=1}^m \chi^2_{p-i+1}\right\} |\mathbf{I}_m + \mathbf{\Omega}|\,,$$

which corresponds to the use of a simple scale change.

In Table 1, we list the simulated 0.05 cut-off points for $T^\bullet_M$, for $M = 1$ for some values of p, m and n.

Table 1: Cut-off points of the 95% confidence set for the regression coefficient $\mathbf{B}$

n	$p=3$, $m=1$, $\alpha=2$	$p=3$, $m=1$, $\alpha=4$	$p=3$, $m=3$, $\alpha=4$	$p=3$, $m=3$, $\alpha=6$
10	6.568	7.433	20.11	29.08
50	5.502E-01	5.581E-01	9.277E-03	9.691E-03
100	2.518E-01	2.542E-01	9.212E-04	9.443E-04
200	1.207E-01	1.208E-01	1.049E-04	1.064E-04
n	$p=4$, $m=1$, $\alpha=2$	$p=4$, $m=1$, $\alpha=4$	$p=4$, $m=3$, $\alpha=4$	$p=4$, $m=3$, $\alpha=6$
10	11.08	12.69	239.2	372.7
50	6.884E-01	6.984E-01	3.550E-02	3.697E-02
100	3.108E-01	3.128E-01	3.487E-03	3.564E-03
200	1.487E-01	1.490E-01	3.674E-04	3.723E-04

Similar to what was done in [5], one could suggest the following adaptations of the classical test criterion for the multivariate regression model (see [1, Secs 8.3 and 8.6] for the classical criteria):

(a) $T^{\bullet}_{1,M} = |\overline{\mathbf{S}}^{\bullet}_M||\overline{\mathbf{S}}^{\bullet}_M + (\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(XX')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})|^{-1}$ (Wilks' Lambda Criterion),

(b) $T^{\bullet}_{2,M} = tr\left[(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(\mathbf{XX}')(\overline{\mathbf{B}}^{\bullet} - \mathbf{B})(\overline{\mathbf{S}}^{\bullet}_M)^{-1}\right]$ (Pillai's Trace Criterion),

(c) $T^{\bullet}_{3,M} = tr\left[(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(\mathbf{XX}')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})[(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(\mathbf{XX}')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B}) + \overline{\mathbf{S}}^{\bullet}_M]^{-1}\right]$ (Hotelling-Lawley Trace Criterion),

(d) $T^{\bullet}_{4,M} = \lambda_1$ where λ_1 denotes the largest eigenvalue of $(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(\mathbf{XX}')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})(\overline{\mathbf{S}}^{\bullet}_M)^{-1}$ (Roy's Largest Root Criterion).

However, these statistics are non-pivotal, since their distributions are function of $\mathbf{\Sigma}$ (see Appendix B.3).

2.2. A Second New Procedure

We propose yet another likelihood-based approach for exact inference about $\mathbf{B}$ where one may gather more information from the released synthetic data, following a somewhat similar procedure to the one used in [5]. Let us start by recalling that $\mathbf{W_j}$ $(j = 1, ..., M)$ are $m \times n$ matrices formed by the vectors $(\mathbf{w}_{j1}, ..., \mathbf{w}_{jn})$ as columns, generated from $\mathbf{w}_{ji}|_{\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}} \sim N_m(\tilde{\mathbf{B}}'\mathbf{x_i}, \tilde{\mathbf{\Sigma}})$ $(i = 1, ..., n)$. Note that, conditionally on $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{\Sigma}}$, $(\mathbf{w}_{1i}, ..., \mathbf{w}_{Mi})$ is a random sample from $N_m(\tilde{\mathbf{B}}'\mathbf{x}_i, \tilde{\mathbf{\Sigma}})$, for $i = 1, ..., n$. Consider $\overline{\mathbf{w}}_i = \frac{1}{M}\sum_{j=1}^{M} \mathbf{w}_{ji}$ and $\mathbf{S}_{\mathbf{w}i} = \sum_{j=1}^{M}(\mathbf{w}_{ji} - \overline{\mathbf{w}}_i)(\mathbf{w}_{ji} - \overline{\mathbf{w}}_i)'$ which are sufficient statistics for $\mathbf{\Sigma}$, based on the i-th covariate vector. Defining $\mathbf{S_w} = \sum_{i=1}^{n} \mathbf{S}_{\mathbf{w}i}$, we have $(\overline{\mathbf{w}}_1, ..., \overline{\mathbf{w}}_n, \mathbf{S_w})$ as the joint sufficient statistics for $(\mathbf{B}, \mathbf{\Sigma})$. Conditionally on $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{\Sigma}}$, we have $\overline{\mathbf{w}}_i \sim N_m(\tilde{\mathbf{B}}'\mathbf{x}_i, \frac{1}{M}\tilde{\mathbf{\Sigma}})$ and $\mathbf{S}_{\mathbf{w}i} \sim W_m(\tilde{\mathbf{\Sigma}}, M-1)$.

From the M released synthetic data matrices $\mathbf{W}_j$ $(j = 1, ..., M)$, we may define $\overline{\mathbf{W}}_M = \frac{1}{M}\sum_{j=1}^{M} \mathbf{W}_j$ and define for $\mathbf{B}$ its estimator

$$\overline{\mathbf{B}}^{\bullet}_M = (\mathbf{XX}')^{-\mathbf{1}}\mathbf{X}\overline{\mathbf{W}}'_M, \tag{2.8}$$

and for $\mathbf{\Sigma}$ its estimator

$$\mathbf{S}^{\bullet}_{comb} = \frac{\mathbf{S_w} + M \times \mathbf{S}^{\bullet}_{mean}}{Mn - p}, \tag{2.9}$$

where we define $\mathbf{S}^{\bullet}_{mean} = (\overline{\mathbf{W}}_M - \overline{\mathbf{B}}^{\bullet\prime}_M\mathbf{X})(\overline{\mathbf{W}}_M - \overline{\mathbf{B}}^{\bullet\prime}_M\mathbf{X})'$.

In fact, if the M synthetic datasets are treated as a single synthetic dataset of size nM, the estimators obtained for $\mathbf{B}$ and $\mathbf{\Sigma}$ will be exactly the same as the ones obtained in (2.8) and (2.9). The proof of this fact may be analyzed in Appendix C.

Analogous to what was done in the previous subsection, one can derive the following inferential results, for $p \geq m$ and $n + \alpha > p + 2m + 2$.

1. An UE of $\boldsymbol{\Sigma}$ will be $\hat{\mathbf{S}}_M = \frac{n+\alpha-p-2m-2}{n-p}\mathbf{S}^{\bullet}_{comb}$ (see Corollary 2.3 Appendix B.4), and for $\alpha = 2m+2$, $\mathbf{S}^{\bullet}_{comb}$ will also be an UE for $\boldsymbol{\Sigma}$.

2. In Corollary 2.3 (see below), we prove that

$$T^{\bullet}_{comb} = \frac{|(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(XX')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})|}{|(Mn-p)\mathbf{S}^{\bullet}_{comb}|} \tag{2.10}$$

is a pivotal quantity, and that for $\mathbf{A}_1 \sim W_m(\mathbf{I}_m, n+\alpha-p-m-1)$, $\mathbf{A}_2 \sim W_m(\mathbf{I}_m, n-p)$ and $F_i \sim F_{p-i+1,Mn-p-i+1}$ $(i = 1, ..., m)$, all independent random variables,

$$T^{\bullet}_{comb}|\boldsymbol{\Omega} \overset{st}{\sim} \left\{\prod_{i=1}^{m} \frac{p-i+1}{Mn-p-i+1}F_i\right\} \left|\frac{M+1}{M}\mathbf{I}_m + \boldsymbol{\Omega}\right|,$$

where $\boldsymbol{\Omega}$ has the same distribution as $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{-1}\mathbf{A}_1^{\frac{1}{2}}$.

3. If one wants to test a linear combination of the parameters in $\mathbf{B}$, namely, $\mathbf{C} = \mathbf{AB}$ where $\mathbf{A}$ is a $k \times p$ matrix with $rank(\mathbf{A}) = k \leq p$ and $k \geq m$, one may define

$$T^{\bullet}_{comb,\mathbf{C}} = \frac{|(\mathbf{A}\overline{\mathbf{B}}^{\bullet}_M - \mathbf{C})'(\mathbf{A}(\mathbf{XX}')^{-1}\mathbf{A}')^{-1}(\mathbf{A}\overline{\mathbf{B}}^{\bullet}_M - \mathbf{C})|}{|(Mn-p)\overline{\mathbf{S}}^{\bullet}_{comb}|},$$

and proceed by noting that

$$T^{\bullet}_{comb,\mathbf{C}}|\mathbf{W} \overset{st}{\sim} \left\{\prod_{i=1}^{m} \frac{k-i+1}{Mn-p-i+1}F_{k,i}\right\} \left|\frac{M+1}{M}\mathbf{I}_m + \boldsymbol{\Omega}\right|, \tag{2.11}$$

with $F_{k,i} \sim F_{k-i+1,Mn-p-i+1}$ being independent random variables and $\boldsymbol{\Omega}$ defined as in the previous item.

(i) *Test for the significance of* $\mathbf{C}$*:* in order to test $H_0 : \mathbf{C} = \mathbf{C}_0$ versus $H_1 : \mathbf{C} \neq \mathbf{C}_0$, we reject H_0 whenever $T^{\bullet}_{comb,\mathbf{C}_0}$ exceeds $\delta_{comb,k,m,p,n;\gamma}$ where $\delta_{comb,k,m,p,n;\gamma}$ satisfies $(1-\gamma) = Pr(T^{\bullet}_{comb,\mathbf{C}_0} \leq \delta_{comb,k,m,p,n;\gamma})$ when H_0 is true. To perform a test for $\mathbf{B} = \mathbf{B}_0$ one has to take $\mathbf{A} = \mathbf{I}_p$.

(ii) *Confidence set for* $\mathbf{C}$*:* a $(1-\gamma)$ level confidence set for $\mathbf{C}$ is given by

$$\Delta_{comb}(\mathbf{C}) = \{\mathbf{C} : T^{\bullet}_{comb,\mathbf{C}} \leq \delta_{comb,k,m,n,p;\gamma}\}, \tag{2.12}$$

where the value of $\delta_{comb,k,m,n,p;\gamma}$ can be obtained by simulating the distribution in (2.11).

Results in 1-3 are derived based on the following Corollaries 2.3 and 2.4, of Theorems 2.1 and 2.2, respectively.

Corollary 2.3. The joint pdf of $\overline{\mathbf{B}}^{\bullet}_M$, $\mathbf{S}^{\bullet}_{comb}$ and $\tilde{\mathbf{\Sigma}}^{\mathbf{-1}}$, for $\overline{\mathbf{B}}^{\bullet}_M$ and $\mathbf{S}^{\bullet}_{comb}$ defined in (2.8) and (2.9), is proportional to

$$e^{-\frac{1}{2}tr\{(\frac{\mathbf{M+1}}{\mathbf{M}}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{-1}(\overline{\mathbf{B}}^{\bullet}_M-\mathbf{B})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}^{\bullet}_M-\mathbf{B})+(Mn-p)\tilde{\mathbf{\Sigma}}^{\mathbf{-1}}\mathbf{S}^{\bullet}_{comb}\}}$$
$$\times \frac{|\mathbf{S}^{\bullet}_{comb}|^{\frac{Mn-p-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{Mn-p+n+\alpha}{2}-m-1}}|\mathbf{\Sigma}|^{-\frac{n}{2}}|\tfrac{M}{M+1}\tilde{\mathbf{\Sigma}}^{\mathbf{-1}}+\mathbf{\Sigma}^{\mathbf{-1}}|^{-p/2}|\tilde{\mathbf{\Sigma}}^{\mathbf{-1}}+\mathbf{\Sigma}^{\mathbf{-1}}|^{-\frac{2n+\alpha-2p-m-1}{2}},$$

so that $\overline{\mathbf{B}}^{\bullet}_M$ and $\mathbf{S}^{\bullet}_{comb}$, given $\tilde{\mathbf{\Sigma}}$, are independent, with

$$\overline{\mathbf{B}}^{\bullet}_M|_{\tilde{\mathbf{\Sigma}}} \sim N_{pm}\left(\mathbf{B}, \left(\frac{M+1}{M}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma}\right)\otimes(\mathbf{X}\mathbf{X}')^{-1}\right)$$

and

$$\mathbf{S}^{\bullet}_{comb}|_{\tilde{\mathbf{\Sigma}}} \sim W_m\left(\frac{1}{Mn-p}\tilde{\mathbf{\Sigma}}, M(n-p)\right).$$

Proof: See Appendix A. □

Corollary 2.4. The distribution of the statistic $T^{\bullet}_{comb}$ defined in (2.10) can be obtained from the decomposition

$$T^{\bullet}_{comb}|_{\mathbf{\Omega}} \overset{st}{\sim} \left\{\prod_{i=1}^{m}\frac{p-i+1}{Mn-p-i+1}F_i\right\}\left|\frac{M+1}{M}\mathbf{I}_m+\mathbf{\Omega}\right|$$

where $F_i \sim F_{p-i+1,Mn-p-i+1}$ are independent random variables, themselves independent of $\mathbf{\Omega}$, which has the same distribution as $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{-1}\mathbf{A}_1^{\frac{1}{2}}$ with $\mathbf{A}_1 \sim W_m(\mathbf{I}_m, n+\alpha-p-m-1)$ and $\mathbf{A}_2 \sim W_m(\mathbf{I}_m, n-p)$, two independent random variables.

Proof: See Appendix A. □

Remark 2.3. Similar to what happens with the statistic $T^{\bullet}_M$ in (2.5), the statistic $T^{\bullet}_{comb}$ in (2.10) also degenerates towards zero when $n\to\infty$ or $M\to\infty$, and similarly to what happens with $T^{\bullet}_M$,

$$(Mn-p)^m\,T^{\bullet}_{comb}|_{\mathbf{\Omega}} \xrightarrow[n\to\infty]{d} \left\{\prod_{i=1}^{m}\chi^2_{p-i+1}\right\}\left|\frac{M+1}{M}\mathbf{I}_m+\mathbf{\Omega}\right|$$

and

$$(Mn-p)^m\,T^{\bullet}_{comb}|_{\mathbf{\Omega}} \xrightarrow[M\to\infty]{d} \left\{\prod_{i=1}^{m}\chi^2_{p-i+1}\right\}|\mathbf{I}_m+\mathbf{\Omega}|\,.$$

Using the simple scale change $T^{\bullet}_{comb2} = (Mn-p)^m\, T^{\bullet}_{comb} = \frac{|(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(XX')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})|}{|\overline{\mathbf{S}}^{\bullet}_{comb}|}$ one would have

$$T^{\bullet}_{comb2}|\boldsymbol{\Omega} \xrightarrow[n\to\infty]{d} \left\{\prod_{i=1}^{m} \chi^2_{p-i+1}\right\} \left|\frac{M+1}{M}\mathbf{I}_m + \boldsymbol{\Omega}\right|$$

and

$$T^{\bullet}_{comb2}|\boldsymbol{\Omega} \xrightarrow[M\to\infty]{d} \left\{\prod_{i=1}^{m} \chi^2_{p-i+1}\right\} |\mathbf{I}_m + \boldsymbol{\Omega}|\,,$$

similar to what happens with $T^{\bullet}_M$.

3. SIMULATION STUDIES

In order to compare the PPS and the FPPS methods with the Plug-in Sampling method we present the results of some simulations analogous to the ones presented in [5]. The objectives of these simulations are: (i) to show that the inference methods developed in Section 2 perform as predicted, and (ii) to compare the measures (*radius*) obtained from our methods with the ones from the Plug-in method. All simulations were carried out using the software Mathematica®. To conduct the simulation, we take the population distribution as a multivariate normal distribution with expected value given by the right hand side of (1.1), for $m = 2$ and $p = 3$, with matrix of regressor coefficients

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix}$$

and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

We set $\alpha = 6$ in order to have both $\overline{\mathbf{S}}^{\bullet}_M$ and $\mathbf{S}^{\bullet}_{comb}$ as the unbiased estimators of $\boldsymbol{\Sigma}$. The regressor variables $x_{1i}, x_{2i}, x_{3i}, i = 1, ..., n$ are generated as i.i.d. $N(1,1)$ and held fixed for the entire simulation. Based on Monte Carlo simulation with 10^5 iterations, we compute an estimate of the coverage probability of the confidence regions for $\mathbf{B}$ and $\mathbf{C} = \mathbf{AB}$ given by (2.7) and (2.12), defined as percentage of observed values of the statistics smaller than the respective theoretical cut-off points, with $\mathbf{A} = \left(\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$, using the methodologies described in Section 2. For $M = 1$, $M = 2$ and $M = 5$, the estimated coverage probabilities of the confidence sets are shown in Table 2 under the columns $\mathbf{B}(1)$ and $\mathbf{AB}(1)$ for the first new procedure in Subsection 2.1, and under the columns $\mathbf{B}(2)$ and $\mathbf{AB}(2)$ for the second new procedure in Subsection 2.2. For $M = 1$, a single column is shown for each confidence region since the two new procedures are the same.

The results reported in Table 2 for samples of size $n = 10, 50, 100, 200$, show that, based on singly and multiply imputed synthetic data, the 0.95 confidence

Table 2: Average coverage for **B** and **AB**

n	$M=1$		$M=2$				$M=5$			
	B	**AB**	1st Approach		2nd Approach		1st Approach		2nd Approach	
			B(1)	**AB**(1)	**B**(2)	**AB**(2)	**B**(1)	**AB**(1)	**B**(2)	**AB**(2)
10	0.949	0.951	0.949	0.949	0.951	0.949	0.951	0.950	0.949	0.951
50	0.949	0.950	0.951	0.951	0.950	0.951	0.951	0.950	0.949	0.948
100	0.949	0.949	0.951	0.950	0.949	0.951	0.949	0.951	0.951	0.950
200	0.951	0.951	0.949	0.951	0.951	0.949	0.950	0.951	0.950	0.951

sets for **B** and **AB** have an estimated coverage probability approximately equal to 0.95, confirming that the confidence sets perform as predicted.

In order to measure the *radius* (distance between the center and the edge) of the confidence sets, we use the same measure proposed in [5], which is

$$\Upsilon_M = d^*_{M,m,n,p,\gamma} \times |\tilde{\mathbf{S}}^\bullet_M|,$$

where $d^*_{M,m,n,p,\gamma}$ is the cut-off point in (2.7) or (2.12). Here we take $M=0$ for the original data, with $\tilde{\mathbf{S}}^\bullet_0 = (n-p)\mathbf{S}$, $M=1$ for the singly imputed synthetic data and $M=2,5$ for the multiply imputed synthetic data, with $\tilde{\mathbf{S}}^\bullet_M = M(n-p)\overline{\mathbf{S}}^\bullet_M$ for the first new procedure, and $\tilde{\mathbf{S}}^\bullet_M = (Mn-p)\mathbf{S}^\bullet_{comb}$ for the second new procedure. The expected value of this measure will be

$$E(\Upsilon_M) = d^*_{M,m,n,p,\gamma} \times \frac{(n-p)!}{(n-p-m)!} \times K_{M,n,p,m}|\mathbf{\Sigma}|$$

where $K_{0,n,p,m} = 1$ for the original data,

$$K_{M,n,p,m} = \frac{(-2+\kappa_{n,p,\alpha,m}-m)!}{(-2+\kappa_{n,p,\alpha,m})!}\frac{(Mn-Mp)!}{(Mn-Mp-m)!}$$

for the procedure in Subsection 2.1 and

$$K_{M,n,p,m} = \frac{(-2+\kappa_{n,p,\alpha,m}-m)!}{(-2+\kappa_{n,p,\alpha,m})!}\frac{(Mn-p)!}{(Mn-p-m)!}$$

for the procedure in Subsection 2.2, where $\kappa_{n,\alpha,p,m} = n+\alpha-p-m-1$, assuming $n+\alpha > p+2m+2$. For more details about these expected values we refer to Appendix B.5.

We present in Table 3 the average of the simulated values of the *radius* Υ_M and its expected value $E(\Upsilon_M)$ for the confidence sets $\Delta_M(\mathbf{B})$ (first procedure) and $\Delta_{comb}(\mathbf{B})$ (second procedure), and in Table 4 the same values for the confidence sets $\Delta_M(\mathbf{C})$ (first procedure) and $\Delta_{comb}(\mathbf{C})$ (second procedure), for $M=0,1,2,5$ and $n=10,50,200$. These values may be compared with the values obtained in [5] for the Plug-in Sampling.

Observing Tables 3 and 4 and comparing the entries in these tables with the results in [5] for Plug-in Sampling, we may see that when synthetic data are generated under FPPS, larger *radius* are obtained. In the singly imputed case, one can observe that the PPS synthetic datasets will lead to a *radius* that is

Table 3: Average values of Υ_M and the values of $E(\Upsilon_M)$ for the confidence set for $\mathbf{B}$.

n	Orig	$M = 1$		$M = 2$			
				1st Procedure		2nd Procedure	
		avg	exp	avg	exp	avg	exp
10	36.97	507.25	512.19	251.55	252.55	237.64	238.68
50	19.11	176.36	176.53	121.23	121.52	121.23	121.48
200	17.52	154.93	156.06	105.81	106.61	105.90	106.72

n	$M = 5$			
	1st Procedure		2nd Procedure	
	avg	exp	avg	exp
10	175.34	176.18	163.82	168.92
50	92.25	92.80	92.28	92.84
200	81.89	82.39	81.91	82.40

Table 4: Average values of Υ_M and the values of $E(\Upsilon_M)$ for the confidence set for $\mathbf{C} = \mathbf{AB}$.

n	Orig	$M = 1$		$M = 2$			
				1st Procedure		2nd Procedure	
		avg	exp	avg	exp	avg	exp
10	13.43	172.64	172.32	92.23	92.44	86.24	86.61
50	7.33	68.93	68.99	47.75	47.86	47.45	47.55
200	7.10	60.65	61.09	41.74	42.05	41.74	42.05

n	$M = 5$			
	1st Procedure		2nd Procedure	
	avg	exp	avg	exp
10	63.07	63.38	61.34	61.74
50	35.32	35.52	35.08	35.27
200	32.47	32.51	32.54	32.53

approximately two and half times that of the *radius* under Plug-in Sampling. As the number M of released synthetic datasets increases, Υ_M slowly decreases, increasing however the difference of the *radius* between the FPPS and the Plug-in methods. Eventually, one may need very large values of M, in order to have values of Υ_M close to the value of Υ_0. As in [5] we also observe that the values of Υ_M $(M > 1)$, for both new FPPS procedures become identical for larger sample sizes.

4. AN APPLICATION USING CURRENT POPULATION SURVEY DATA

In this section, we provide an application based on the same real data used in [5] to compare the original data inference with the one obtained via PPS, for the single imputation case, and via FPPS, for the multiple imputation case. The data are from the U.S. 2000 Current Population Survey (CPS) March supplement, available online at http://www.census.gov.cps/. Further details on the data may be found in [5].

In this application, $\mathbf{x}$, the vector of regressor variables, is defined as

$$\mathbf{x} = \Big(1, N, L, A, I(E=34), ..., I(E=37), I(E=39), ..., I(E=46),$$
$$I(M=3), ..., I(M=7), I(R=2), I(R=4), I(S=2)\Big)',$$

where N, L, A, are respectively, the number of people in household, the number of people in the household who are less than 18 years old and the age for the head of household, E, M, R and S, are respectively, the education level for the head of the household (coded to take values 31, 34-37, 39-46), the marital status for the head of the household (coded to take values 1,3-7), the race of the head of the household (coded to take values 1,2,4) and the sex of the head of the household (coded to take values 1,2). I(E = 34) is the indicator variable for $E = 34$, I(E = 35) is the indicator variable for $E = 35$, and so on, and where the indicator variable for the first code present in the sample for each variable is taken out in order to make the model matrix full rank. The vector $\mathbf{y}$ of response variables will be formed by the same three numerical variables used in [5], namely, *total household income*, *household alimony payment* and *household property tax*. After deleting all entries where at least one of these variables are reported as 0, we were left with a sample size of 141, and as such the model matrix $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]$ has thus $p = 24$ rows, $n = 141$ columns, with rank equal to 24. Throughout this section we will assume $\alpha = 8$ in order to have $\mathbf{S}^\bullet_M$ and $\mathbf{S}^\bullet_{comb}$ as unbiased estimators of $\mathbf{\Sigma}$. Via PPS method we generate a single synthetic dataset and show in expression (4.1) the realizations of the unbiased estimator $\mathbf{S}^\bullet$ for $\mathbf{\Sigma}$ and of the estimator $\mathbf{S}$ for the original data, respectively denoted by $\widetilde{\mathbf{S}}^\bullet_1$ and $\widetilde{\mathbf{S}}$

(4.1)
$$\widetilde{\mathbf{S}}^\bullet_1 = \begin{pmatrix} 1.58572 & -0.20443 & 0.27981 \\ -0.20443 & 1.61395 & 0.16089 \\ 0.27981 & 0.16089 & 0.34648 \end{pmatrix}, \quad \widetilde{\mathbf{S}} = \begin{pmatrix} 1.1980 & -0.0375 & 0.2970 \\ -0.0375 & 1.0699 & 0.1175 \\ 0.2970 & 0.1175 & 0.4045 \end{pmatrix}.$$

In Table 5 we show the realizations of the unbiased estimator $\mathbf{B}^\bullet_1$ of $\mathbf{B}$ and of the estimator $\hat{\mathbf{B}}$ of the original data, respectively denoted by $\widetilde{\mathbf{B}}^\bullet_1$ and $\widetilde{\hat{\mathbf{B}}}$.

At a first glance the estimates originated via Plug-in Sampling (see [5]) seem to be more in agreement with the original data estimates than the ones drawn from PPS. Nevertheless, this is only one draw and it could be a question of chance to originate 'better' or 'worse' data. Therefore, one must conduct inferences on the regression coefficients based on multiple draws.

Inferences on regression coefficients are obtained by applying the methodologies in Subsections 2.1 and 2.2, to analyze the singly imputed synthetic dataset and multiply imputed synthetic datasets, considering $M = 1$, $M = 2$ and $M = 5$, using the statistics $T^\bullet_M$ and $T^\bullet_{comb}$ and their empirical distributions based on simulations with 10^4 iterations, to test the fit of the model and the significance of some regressors for $\gamma = 0.05$. Regarding the test of fit of the model one will find, for all values of M, results equivalent to the ones obtained for the case when synthetic data are generated via Plug-in Sampling, i.e., concluding that the explanatory variables in $\mathbf{x}$ have a significant role in determining the values of the

Table 5: Estimates of the regressor coefficients from the FPPS synthetic data ($\widetilde{\mathbf{B}}^\bullet$), Plug-in synthetic data ($\widetilde{\mathbf{B}}^*$) and from the original data.

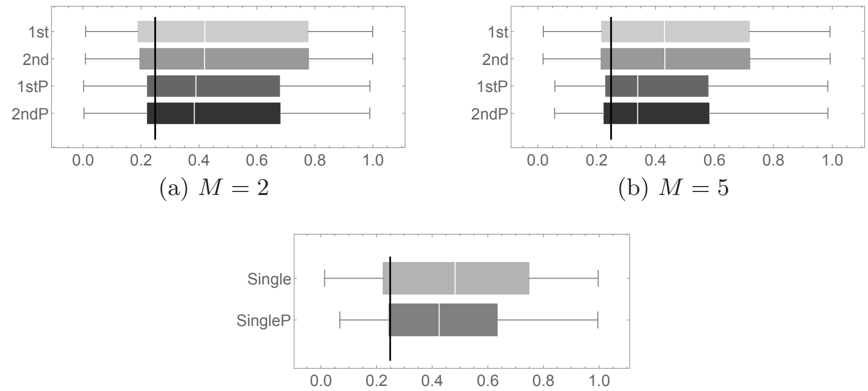
regressor	$FPPS$ $SyntheticData$ ($\widetilde{\mathbf{B}}^\bullet$)			$Plug-in$ $SyntheticData$ ($\widetilde{\mathbf{B}}^*$)			$OriginalData$ ($\widetilde{\mathbf{B}}$)		
	I	AP	PT	I	AP	PT	I	AP	PT
Intercept	11.4996	3.3381	8.1713	10.1829	3.7094	10.9787	9.8339	4.6663	10.1095
N	0.2801	−0.2562	0.6317	−0.0938	0.1435	0.6189	0.0457	0.0375	0.4585
L	−0.3996	0.4960	−0.6017	0.0812	0.0163	−0.5932	0.0186	0.1310	−0.3851
A	−0.0061	0.0223	0.0018	0.0075	0.0285	−0.0097	0.0118	0.0181	−0.0020
I(E=34)	−4.7732	0.3476	−0.4662	−6.6680	1.2055	−2.0664	−4.4348	0.5944	−1.2291
I(E=35)	−5.5990	2.8081	1.9914	−1.2231	−0.0154	−0.7091	−1.4060	0.9188	−0.1468
I(E=36)	−4.2467	2.2712	0.6907	−0.4478	2.1718	−0.9172	−2.3100	1.0416	−0.5002
I(E=37)	−3.5281	0.7339	1.4653	−1.1547	1.3009	−1.0659	−2.0490	0.7410	0.2335
I(E=39)	−3.3369	1.5590	1.0109	−2.5737	0.7234	−1.1346	−2.2208	0.4054	−0.4136
I(E=40)	−2.8766	1.7608	1.2350	−1.8032	1.0617	−0.6940	−1.8834	0.8519	0.0852
I(E=41)	−2.8266	2.7954	2.3165	−1.5615	1.6881	−0.0291	−1.9468	1.4222	0.1094
I(E=42)	−3.5901	2.3990	0.7908	−2.4543	2.0378	−1.1494	−2.3381	1.3840	−0.0808
I(E=43)	−1.9852	2.1149	1.9765	−1.7090	1.1722	−0.4341	−1.5057	1.0766	0.5309
I(E=44)	−3.2012	2.0495	1.7665	−2.2668	1.5629	−0.2140	−1.8082	1.1301	0.4936
I(E=45)	0.1813	1.1103	1.7535	−1.8984	2.1024	−0.4636	−0.9893	0.7958	0.3057
I(E=46)	0.5791	2.3091	3.5534	0.4558	1.4836	1.1497	−0.6198	1.0766	1.0624
I(M=3)	−2.3691	0.8545	−0.3594	−1.9077	−0.4988	−0.4836	−2.7258	0.0964	−0.2156
I(M=4)	−4.4234	2.2640	−1.2282	−0.0088	0.5609	−0.2349	−0.0134	0.5887	0.3864
I(M=5)	−1.0787	1.5611	0.1170	0.3767	0.6729	0.1184	0.1455	0.4770	0.1558
I(M=6)	−0.8300	−0.2358	−0.2713	0.3948	−0.3092	−0.1046	−0.7122	−0.4448	−0.4025
I(M=7)	−2.8242	2.9533	0.5456	1.0576	0.5476	0.5187	−0.1990	1.1750	0.6685
I(R=2)	0.3378	3.8443	1.4196	−1.0805	3.0078	−0.1619	−0.9205	1.3432	0.4696
I(R=4)	0.0340	1.9168	−0.4519	0.6883	−0.3211	0.3639	−0.7040	0.0975	−0.1618
I(S=2)	1.3582	−0.4793	−0.1588	0.0564	−0.2309	−0.2849	0.1236	−0.1355	−0.4025

response variables in $\mathbf{y}$ since the obtained p-values, computed as the fraction of values of the empirical distribution of the corresponding statistic that are larger than the computed value of the statistic, were all approximately zero. The cut-off points obtained from the empirical distributions of $T_M^\bullet$ and $T_{comb}^\bullet$ (respectively associated with the first and second procedures in Subsections 2.1 and 2.2) are approximately equal to 0.50357, for $M = 1$ (where first and second procedures coincide), to 0.03460 and 0.02569, for $M = 2$, and to 0.00149 and 0.00094, for $M = 5$.

In order to test the significance of some regressors, we propose to study two different cases, using in each case the same sets of regressors as in [5]. Therefore, we will test the significance of regressor variables R and S, for the first case, and regressor variables A and E, for the second case. As such, in the first case, we will consider a 3×24 matrix

$$\mathbf{A} = \left(\mathbf{0}_{3\times 21} \middle| \mathbf{I}_3 \right)$$

and we will be interested in testing the hypothesis $H_0 : \mathbf{AB} = \mathbf{C_0}$, where $\mathbf{C}_0$ is a 3×3 matrix consisting of only zeros. We now generate 100 draws of $M = 1$, $M = 2$ and $M = 5$ synthetic datasets and gather the different p-values obtained when using the statistics in (2.5) and (2.10). In Figure 1, one may analyze the box-plots of the p-values obtained for each procedure together with the ones obtained in [5] for the same sets of variables, where under Single, 1st and 2nd, one has the box-plots associated with the new procedures developed in this paper and under SingleP, 1stP and 2ndP, the box-plots associated to the Plug-in Sampling method. The existing line in the box-plots marks the original data p-value 0.249, obtained using the $T_{O,\mathbf{C}}$ statistic in (3) of [5]. It is important to note that in the case of single imputation ($M = 1$) the FPPS method reduces to the usual PPS method.



(a) $M = 2$ (b) $M = 5$

(c) $M = 1$

Figure 1: Box-plots of p-values obtained, when testing the joint significance of I(R=2), I(R=4) and I(S=2), from 100 draws of synthetic datasets using procedures in Section 2 and using Plug-in Sampling method from [5], for $M = 1$, $M = 2$ and $M = 5$.

In general, from Figure 1, we may note in both new procedures a larger spread of the p-values when compared with the p-values gathered from Plug-in Sampling, presenting a distribution of p-values with larger values than the original, nonetheless with the majority of these p-values leading to similar conclusions as those obtained from the original data for $\gamma = 0.05$, that is, to not reject the null hypothesis that variables R and S do not have significant influence on the response variables.

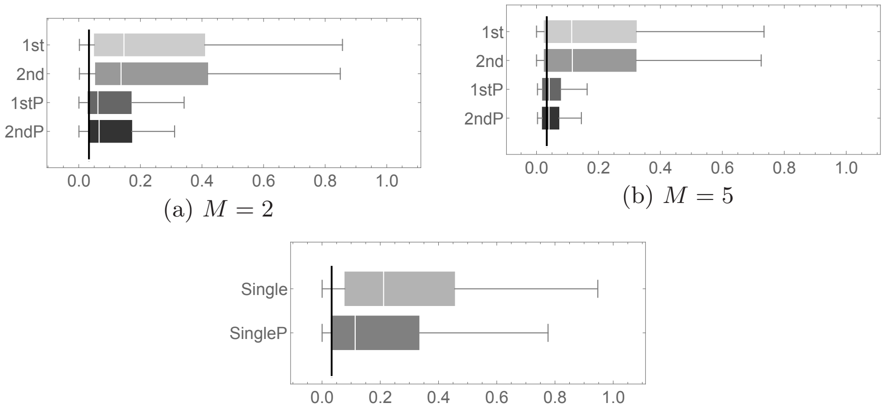
We may note that in general, in cases where the p-value obtained from the original data is rather low, we expect to obtain larger p-values for the synthetic data, given the inherent variability of these synthetic data and the "need" of the inferential exact methods to preserve the $1-\gamma$ coverage level, and impossibility of compressing the synthetic data p-values towards zero.

For the second case, we are interested in testing the hypothesis H_0 : $\mathbf{AB} = \mathbf{C_0}$, where $\mathbf{C_0}$ is a 13×13 matrix consisting of only zeros, with

$$\mathbf{A} = \left(\mathbf{0}_{13\times 3} \middle| \mathbf{I}_{13} \middle| \mathbf{0}_{13\times 8} \right),$$

corresponding to the test of joint significance of variables A and E. The p-value obtained for the original data, based on (3) in [5], was 0.033, thus rejecting their non-significance for $\gamma = 0.05$. In Figure 2, we can compare the box-plots obtained for the FPPS and Plug-in Sampling methods obtained by generating 100 draws of synthetic datasets, for $M = 1$, $M = 2$ and $M = 5$. The vertical line represents again the original data's p-value.

From Figure 2, we note that the spread of p-values is again larger for our new procedures based on FPPS than the ones from the Plug-in method, majorly



(a) $M = 2$ (b) $M = 5$

(c) $M = 1$

Figure 2: Box-plots of p-values obtained, when testing the joint significance of A and E, from 100 draws of synthetic datasets using procedures in Section 2 and using Plug-in Sampling method from [5], for $M = 1$, $M = 2$ and $M = 5$.

leading to a different conclusion from the inference obtained from the original data.

For the single imputation case, even if the spread of the p-values gathered from the PPS is larger than the ones from the Plug-in Sampling, the distributions of p-values are not that different for the two methods.

For the two cases studied, the two new FPPS multiple imputation procedures presented have very similar p-values. As M increases the spread of the p-values from FPPS becomes smaller and closer to the original data's p-value but at a smaller rate than the p-values from the Plug-in Sampling.

Nevertheless, this larger spread of the p-values from FPPS will be compensated by an increase of the level of confidentiality, as it can be seen in the next section.

Next, we present the power for the tests

$$\begin{array}{l} H_0 : \mathbf{B} = \mathbf{B_0}(\neq \mathbf{0}) \ vs \ \ H_1 : \mathbf{B} = \mathbf{B_1} \quad \text{and} \\ H_0 : \mathbf{AB} = \mathbf{C_0}(\neq \mathbf{0}) \ vs \ \ H_1 : \mathbf{AB} = \mathbf{C_1} \end{array} \tag{4.2}$$

for $\mathbf{B_0}$ equal to $\tilde{\hat{\mathbf{B}}}$, rounded to two decimal places,

$$\mathbf{A} = \left(\mathbf{0}_{12\times 4} \middle| \mathbf{I}_{12} \middle| \mathbf{0}_{12\times 8} \right),$$

a 12×12 matrix defined appropriately in order to isolate the indicator variables associated with the variable E, and $\mathbf{C}_1 = \mathbf{AB}_1$ where $\mathbf{B}_1$ takes different values, found in Table 6, with $\mathbf{D}$ a $p \times m$ matrix of 1's.

The power for the synthetic data obtained via FPPS was then simulated as well as the power for the case when these synthetic datasets are treated as if they were the original data. We also simulated the power from the original data and refer to [5] for the power values for the synthetic data generated via Plug-in Sampling.

Table 6: Power for the tests to the hypothesis (4.2), with $\mathbf{B}(1)$, $\mathbf{C}(1)$ and $\mathbf{B}(2)$ and $\mathbf{C}(2)$ denoting the first and second procedures proposed by the authors in Subsections 2.1 and 2.2 for FPPS and in [5] for Plug-in method.

Power for $\mathbf{B}_1 =$	orig data $\mathbf{B}$	Methods	M=1 $\mathbf{B}$	M=2 $\mathbf{B}(1)$	M=2 $\mathbf{B}(2)$	M=5 $\mathbf{B}(1)$	M=5 $\mathbf{B}(2)$	synt as orig $\mathbf{B}$
$\mathbf{B}_0 + 0.005\mathbf{D}$	0.537	FPPS	0.215	0.252	0.253	0.275	0.279	1.000
		Plug-in	0.279	0.382	0.385	0.471	0.472	1.000
$\mathbf{B}_0 * 0.95$	0.945	FPPS	0.535	0.634	0.637	0.700	0.700	1.000
		Plug-in	0.679	0.840	0.841	0.906	0.909	1.000
Power for $\mathbf{C}_1 =$	orig data $\mathbf{C}$	Methods	M=1 $\mathbf{C}$	M=2 $\mathbf{C}(1)$	M=2 $\mathbf{C}(2)$	M=5 $\mathbf{C}(1)$	M=5 $\mathbf{C}(2)$	synt as orig $\mathbf{C}$
$\mathbf{A}(\mathbf{B}_0 + 3\mathbf{D})$	0.465	FPPS	0.185	0.202	0.207	0.245	0.246	0.996
		Plug-in	0.284	0.334	0.343	0.416	0.418	0.975
$\mathbf{A}(\mathbf{B}_0 * 0.5)$	0.393	FPPS	0.136	0.160	0.161	0.179	0.181	0.996
		Plug-in	0.197	0.271	0.279	0.326	0.327	0.959

From the power values in Table 6 we may see that tests based on the synthetic data via FPPS show lower values for its power than the ones based in Plug-in generation, as expected, since we are using a method which is supposed to give more confidentiality by generating more perturbed datasets. We may see that these values increase along with the value of M, but with a smaller rate than that for Plug-in Sampling, leading to the conclusion that one will need larger values of M to obtain a closer power value to the one registered when testing using the original data. If synthetic data is treated as original, we obtain a larger power than the one obtained for the original data, which is obviously misleading, since the estimated coverage probability will be in fact much smaller than the desired 0.95.

5. PRIVACY PROTECTION OF SINGLY VERSUS MULTIPLY IMPUTED SYNTHETIC DATA

In order to evaluate the level of protection and at the same time compare it with the level obtained from synthetic data generated via Plug-in Sampling, we perform, in this section, a similar evaluation as in [5] using CPS data. Let us consider $\mathbf{W}_l = (\mathbf{w}_{1l}, ..., \mathbf{w}_{nl})$, $l = 1, ..., M$, M synthetic datasets generated via FPPS, where $\mathbf{w}_{il} = (w_{1il}, ..., w_{mil})'$, $i = 1, ..., n$. The estimate of the original values $\mathbf{y}_i = (y_{1i}, ..., y_{mi})'$ will be $\hat{\mathbf{y}}_i = \frac{1}{M}\sum_{l=1}^{M} \mathbf{w}_{il}$. Let us recall the three criteria used in [5] as measures of the level of privacy protection:

$$\Gamma_{1,\epsilon} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} Pr\left[\left|\frac{\hat{y_{ji}} - y_{ji}}{y_{ji}}\right| < \epsilon \,\middle|\, \mathbf{Y}\right]; \tag{5.1}$$

$$(5.2)\qquad \Gamma_{2,\epsilon} = \frac{1}{n}\sum_{i=1}^{n} Pr\left[\sqrt{\frac{1}{m}\sum_{j=1}^{m}\frac{(\hat{y_{ji}} - y_{ji})^2}{y_{ji}^2}} < \epsilon \,\middle|\, \mathbf{Y}\right];$$

$$(5.3)\qquad \Gamma_{3,\epsilon} = Pr\left[\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\left|\frac{\hat{y_{ji}} - y_{ji}}{y_{ji}}\right| < \epsilon \,\middle|\, \mathbf{Y}\right].$$

Let us also consider, from $\Gamma_{1,\epsilon}$, the following quantity, for $i = 1, ...n$ and $j = 1, .., m$,

$$D_{1,\epsilon,ji} = Pr\left[\left|\frac{\hat{y_{ji}} - y_{ji}}{y_{ji}}\right| < \epsilon \,\middle|\, \mathbf{Y}\right]$$

and, from $\Gamma_{3,\epsilon}$,

$$D_3 = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\left|\frac{\hat{y_{ji}} - y_{ji}}{y_{ji}}\right|.$$

We use a Monte Carlo simulation with 10^4 iterations to estimate all three measures in (5.1)–(5.3) based on the $n = 141$ households in the CPS data. In Table 7, we show the values of $\Gamma_{1,0.01}$, $\Gamma_{2,0.01}$ and the minimum, 1st quartile (Q_1), median, 3rd quartile (Q_3) and maximum of $D_{1,\epsilon}$, displaying also the values gathered when using Plug-in Sampling. In Table 8, we show the values of $\Gamma_{3,0.1}$ and the minimum, Q_1, median, Q_3 and maximum of D_3 also displaying the values gathered when using Plug-in Sampling.

Table 7: Values of $\Gamma_{1,0.01}$, $\Gamma_{2,0.01}$ and a summary of the distribution of $D_{1,0.01}$.

M	Method	$\Gamma_{1,0.01}$	$\Gamma_{2,0.01}$	Min	Q_1	Median	Q_3	Max
$M = 1$	FPPS	0.0602	0.0005	0	0.0385	0.0507	0.0784	0.1455
	Plug-in	0.0631	0.0006	0	0.0398	0.0552	0.0854	0.1491
$M = 2$	FPPS	0.0702	0.0009	0	0.0357	0.0624	0.0910	0.1945
	Plug-in	0.0754	0.0010	0	0.0331	0.0697	0.0954	0.2134
$M = 5$	FPPS	0.0797	0.0012	0	0.0214	0.0711	0.1136	0.2785
	Plug-in	0.0879	0.0018	0	0.0110	0.0792	0.1284	0.3279

Table 8: Values of $\Gamma_{3,0.1}$ and a summary of the distribution of D_3.

M	Method	$\Gamma_{3,0.1}$	Min	Q_1	Median	Q_3	Max
$M = 1$	FPPS	0.0000	0.1091	0.1248	0.1287	0.1325	0.1544
	Plug-in	0.0000	0.1050	0.1202	0.1233	0.1264	0.1379
$M = 2$	FPPS	0.0021	0.0960	0.1088	0.1116	0.1145	0.1324
	Plug-in	0.0694	0.0948	0.1026	0.1051	0.1072	0.1159
$M = 5$	FPPS	0.5008	0.0896	0.0980	0.1000	0.1020	0.1131
	Plug-in	1.0000	0.0846	0.0905	0.0920	0.0936	0.0992

Looking at Tables 7 and 8, we observe that the values of the privacy measures in (5.1)–(5.3) increase for increasing values of M for both procedures developed in Subsections 2.1 and 2.2, showing that the disclosure risk increases with the increase in the number of released synthetic datasets. Compared with the measures obtained under Plug-in Sampling, we may observe a smaller disclosure risk in all cases, leading to the conclusion that the proposed FPPS procedures have an overall higher level of confidentiality. Regarding measures $\Gamma_{2,\epsilon}$ and $\Gamma_{3,\epsilon}$ this increase reaches in some cases an increase of 50% or more in confidentiality. In the single imputation case, under the PPS we also register an increase of confidentiality when comparing the same measure under Plug-in Sampling, nevertheless this increase is relatively small.

6. CONCLUDING REMARKS

In this paper the authors derive likelihood-based exact inference for single and multiple imputation cases where synthetic datasets are generated via Fixed-Posterior Predictive Sampling (FPPS). If only one synthetic dataset is released, then FPPS is equivalent to the usual Posterior Predictive Sampling (PPS) method. Thus the proposed methodology can be used to analyze a singly imputed synthetic data set generated via PPS under the multivariate linear regression (MLR) model. Therefore this work fills a gap in the literature because the state of the art methods apply only to multiply imputed synthetic data. Under the MLR model, the authors derived two different exact inference procedures for the matrix of regression coefficients, when multiply imputed synthetic datasets are released. It is shown that the methodologies proposed lead to confidence sets matching the expected level of confidence, for all sample sizes. Furthermore, while the second proposed procedure displays a better precision for smaller samples and/or smaller values of M by yielding smaller confidence sets, the two procedures concur for larger sample sizes and larger values of M, as it is corroborated in theory by remarks 2.2 and 2.3. When compared with inference procedures for Plug-in Sampling, the procedures proposed based on FPPS lead to synthetic datasets that give respondents a higher level of confidentiality, that is, a reduced disclosure risk, nevertheless at the expense of accuracy, since the confidence sets are larger, as illustrated in the application with the CPS data. Once likelihood-based exact inferential methods are now made available both for FPPS/PPS and Plug-in Sampling, it is therefore the responsibility of those in charge of releasing the data to decide which method to use in order to better respect the demands and objectives of their institution.

ACKNOWLEDGMENTS

Ricardo Moura's research is supported by a Fulbright Research Grant, and he sincerely thanks the faculty of Mathematics and Statistics at UMBC for their support and encouragement. Ricardo Moura and Carlos A. Coelho also thank FCT (Portuguese Foundation for Science and Technology) project UID/MAT/00297/2013 awarded through CMA/UNL. Martin Klein and Bimal Sinha thank Laura McKenna, Eric Slud, William Winkler, and Tommy Wright at the U.S. Census Bureau for their support. The authors would also like to thank the referees for the helpful comments and suggestions leading to the improvement of the paper.

A. Proof of Theorems 2.1 and 2.2 and Corollaries 2.3 and 2.4

Proof of Theorem 2.1: Given $(\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}})$, from (2.3) we have that, for every $j=1,...,M$,

$$\mathbf{W}_j'|_{\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}} \sim N_{nm}(\mathbf{X}'\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}\otimes\mathbf{I}_n) \implies \mathbf{B}_j^{\bullet}|_{\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}} \sim N_{pm}(\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}\otimes(\mathbf{X}\mathbf{X}')^{-1})$$

and

$$(n-p)\mathbf{S}_j^{\bullet}|_{\tilde{\mathbf{\Sigma}}} \sim W_m(\tilde{\mathbf{\Sigma}},n-p).$$

Therefore, we have for $\overline{\mathbf{B}}_M^{\bullet}$ and $\overline{\mathbf{S}}_M^{\bullet}$ in (2.4),

$$\overline{\mathbf{B}}_M^{\bullet}|_{\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{B}_j^{\bullet}|_{\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}} \sim N_{pm}\left(\tilde{\mathbf{B}},\frac{1}{M}\tilde{\mathbf{\Sigma}}\otimes(\mathbf{X}\mathbf{X}')^{-1}\right)$$

and

$$M(n-p)\overline{\mathbf{S}}_M^{\bullet}|_{\tilde{\mathbf{\Sigma}}} = (n-p)\sum_{j=1}^{M}\overline{\mathbf{S}}_j^{\bullet}|_{\tilde{\mathbf{\Sigma}}} \sim W_m(\tilde{\mathbf{\Sigma}},M(n-p)).$$

Since $\overline{\mathbf{B}}_M^{\bullet}$ and $\overline{\mathbf{S}}_M^{\bullet}$ are independent, the conditional joint pdf of $(\overline{\mathbf{B}}_M^{\bullet},\overline{\mathbf{S}}_M^{\bullet})$, given $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{\Sigma}}$, is

$$\text{(A.1)}\quad \begin{aligned} &f(\overline{\mathbf{B}}_M^{\bullet},\overline{\mathbf{S}}_M^{\bullet}|\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}) \propto \\ &\qquad e^{-\frac{1}{2}tr\{M\tilde{\mathbf{\Sigma}}^{-1}\left[(\overline{\mathbf{B}}_M^{\bullet}-\tilde{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}_M^{\bullet}-\tilde{\mathbf{B}})+M(n-p)\overline{\mathbf{S}}_M^{\bullet}\right]\}} \times \frac{|\overline{\mathbf{S}}_M^{\bullet}|^{\frac{M(n-p)-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{M(n-p)+p}{2}}}, \end{aligned}$$

while, due to the independence of $\tilde{\mathbf{\Sigma}}^{-\mathbf{1}}$ and $\tilde{\mathbf{B}}$, generated from (2.1) and (2.2), respectively, the joint pdf of $(\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}^{-\mathbf{1}})$, given $\mathbf{S}$, is

$$\text{(A.2)}\quad f(\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}^{-\mathbf{1}}|\mathbf{S}) \propto |\tilde{\mathbf{\Sigma}}|^{-p/2}e^{-\frac{1}{2}tr\{\tilde{\mathbf{\Sigma}}^{-1}\left[(\tilde{\mathbf{B}}-\hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\tilde{\mathbf{B}}-\hat{\mathbf{B}})+(n-p)\mathbf{S}\right]\}}\frac{|\mathbf{S}|^{\frac{n+\alpha-p-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{n+\alpha-p}{2}-m-1}}.$$

On the other hand, given the independence of $\hat{\mathbf{B}}$ and $\mathbf{S}$, defined in (1.2) and (1.3), the joint pdf of $(\hat{\mathbf{B}},\mathbf{S})$ is given by

$$\text{(A.3)}\qquad f(\hat{\mathbf{B}},\mathbf{S}) \propto e^{-\frac{1}{2}tr\{\mathbf{\Sigma}^{-1}\left[(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{X}\mathbf{X}'(\hat{\mathbf{B}}-\mathbf{B})+(n-p)\mathbf{S}\right]\}}\frac{|\mathbf{S}|^{\frac{n-p-m-1}{2}}}{|\mathbf{\Sigma}|^{\frac{n}{2}}}.$$

Thus, by multiplying the three pdf's in (A.1), (A.2) and (A.3), we obtain the joint pdf of $(\overline{\mathbf{B}}_M^{\bullet},\overline{\mathbf{S}}_M^{\bullet},\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}}^{-\mathbf{1}},\hat{\mathbf{B}},\mathbf{S})$.

Since

$$tr\{M(\overline{\mathbf{B}}_M^{\bullet}-\tilde{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}_M^{\bullet}-\tilde{\mathbf{B}})\} = tr\{M(\tilde{\mathbf{B}}-\overline{\mathbf{B}}_M^{\bullet})'\mathbf{X}\mathbf{X}'(\tilde{\mathbf{B}}-\overline{\mathbf{B}}_M^{\bullet})\},$$

and since from Appendix B.2 we may write

$$M(\tilde{\mathbf{B}}-\overline{\mathbf{B}}^{\bullet}_{M})'\mathbf{X}\mathbf{X}'(\tilde{\mathbf{B}}-\overline{\mathbf{B}}^{\bullet}_{M})+(\tilde{\mathbf{B}}-\hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\tilde{\mathbf{B}}-\hat{\mathbf{B}})=$$
$$=(M+1)\left[\tilde{\mathbf{B}}-\frac{1}{M+1}(\mathbf{B}^{\bullet}+\hat{\mathbf{B}})\right]'\mathbf{X}\mathbf{X}'\left[\tilde{\mathbf{B}}-\frac{1}{M+1}(\mathbf{B}^{\bullet}+\hat{\mathbf{B}})\right]$$
$$+\frac{M}{M+1}(\mathbf{B}^{\bullet}-\hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\mathbf{B}^{\bullet}-\hat{\mathbf{B}}),$$

by integrating out $\tilde{\mathbf{B}}$, we obtain the joint pdf of $(\overline{\mathbf{B}}^{\bullet}_{M},\overline{\mathbf{S}}^{\bullet}_{M},\tilde{\mathbf{\Sigma}}^{-1},\hat{\mathbf{B}},\mathbf{S})$ proportional to

$$e^{-\frac{1}{2}tr\{\tilde{\mathbf{\Sigma}}^{-1}\left[\frac{M}{M+1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})+(n-p)(M\overline{\mathbf{S}}^{\bullet}_{M}+\mathbf{S})\right]+\mathbf{\Sigma}^{-1}\left[(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{X}\mathbf{X}'(\hat{\mathbf{B}}-\mathbf{B})+(n-p)\mathbf{S}\right]\}}$$
$$\times\frac{|\overline{\mathbf{S}}^{\bullet}_{M}|^{\frac{M(n-p)-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{M(n-p)+n-\alpha}{2}-m-1}}\frac{|\mathbf{S}|^{n+\frac{\alpha}{2}-p-m-1}}{|\mathbf{\Sigma}|^{\frac{n}{2}}}.$$

Since

$$tr\left\{\frac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})'(\mathbf{X}\mathbf{X}')(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})+\mathbf{\Sigma}^{-1}(\hat{\mathbf{B}}-\mathbf{B})'(\mathbf{X}\mathbf{X}')(\hat{\mathbf{B}}-\mathbf{B})\right\}=$$
$$tr\left\{\mathbf{X}\mathbf{X}'\left[\frac{M}{M+1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})\tilde{\mathbf{\Sigma}}^{-1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})'+(\hat{\mathbf{B}}-\mathbf{B})\mathbf{\Sigma}^{-1}(\hat{\mathbf{B}}-\mathbf{B})'\right]\right\}$$

and since from the identities in 1.-3. in Appendix B1 in [5] we may write

$$\frac{M}{M+1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})\tilde{\mathbf{\Sigma}}^{-1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})'+(\hat{\mathbf{B}}-\mathbf{B})\mathbf{\Sigma}^{-1}(\hat{\mathbf{B}}-\mathbf{B})'=$$
$$=\left[\hat{\mathbf{B}}-\left(\frac{M}{M+1}\overline{\mathbf{B}}^{\bullet}_{M}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{B}\mathbf{\Sigma}^{-1}\right)\left(\frac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}^{-1}\right)^{-1}\right]$$
$$\left(\frac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}^{-1}\right)\left[\hat{\mathbf{B}}-\left(\frac{M}{M+1}\overline{\mathbf{B}}^{\bullet}_{M}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{B}\mathbf{\Sigma}^{-1}\right)\left(\frac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}^{-1}\right)^{-1}\right]'$$
$$+(\overline{\mathbf{B}}^{\bullet}_{M}-\mathbf{B})\left(\frac{M+1}{M}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma}\right)^{-1}(\overline{\mathbf{B}}^{\bullet}_{M}-\mathbf{B})',$$

integrating out $\hat{\mathbf{B}}$ we will have the joint pdf of $(\overline{\mathbf{B}}^{\bullet}_{M},\overline{\mathbf{S}}^{\bullet}_{M},\tilde{\mathbf{\Sigma}}^{-1},\mathbf{S})$ proportional to

$$e^{-\frac{1}{2}tr\{(\frac{M+1}{M}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{-1}(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}^{\bullet}_{M}-\hat{\mathbf{B}})+(n-p)\tilde{\mathbf{\Sigma}}^{-1}(M\overline{\mathbf{S}}^{\bullet}_{M}+\mathbf{S})+(n-p)\mathbf{\Sigma}^{-1}\mathbf{S}\}}$$
$$\times\frac{|\overline{\mathbf{S}}^{\bullet}_{M}|^{\frac{M(n-p)-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{M(n-p)+n-\alpha}{2}-m-1}}\frac{|\mathbf{S}|^{n+\frac{\alpha}{2}-p-m-1}}{|\mathbf{\Sigma}|^{\frac{n}{2}}}\left|\frac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}^{-1}\right|^{-p/2}.$$

Consequently, if we integrate out $\mathbf{S}$ we will end up with the joint pdf of $(\overline{\mathbf{B}}^{\bullet}_{M},\overline{\mathbf{S}}^{\bullet}_{M},\tilde{\mathbf{\Sigma}}^{-1})$ proportional to

(A.4)

$$e^{-\frac{1}{2}tr\{(\frac{M+1}{M}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{-1}(\overline{\mathbf{B}}^{\bullet}_{M}-\mathbf{B})'\mathbf{X}\mathbf{X}'(\overline{\mathbf{B}}^{\bullet}_{M}-\mathbf{B})+M(n-p)\tilde{\mathbf{\Sigma}}^{-1}\overline{\mathbf{S}}^{\bullet}_{M}\}}$$
$$\times\frac{|\overline{\mathbf{B}}^{\bullet}_{M}|^{\frac{M(n-p)-m-1}{2}}}{|\tilde{\mathbf{\Sigma}}|^{\frac{M(n-p)+n-\alpha}{2}-m-1}}|\mathbf{\Sigma}|^{-\frac{n}{2}}\left|\frac{M}{M+1}\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}^{-1}\right|^{-p/2}|\tilde{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}^{-1}|^{-\frac{2n+\alpha-2p-m-1}{2}}$$

as we wanted to prove. It is easy to see that in (A.4), $\overline{\mathbf{S}}^{\bullet}_M$ and $\overline{\mathbf{B}}^{\bullet}_M$, given $\tilde{\boldsymbol{\Sigma}}^{-1}$, are separable, with the distributions in the body of the Theorem. $\square$

Proof of Theorem 2.2: From the distributions of $\overline{\mathbf{S}}^{\bullet}_M$ and $\overline{\mathbf{B}}^{\bullet}_M$ in Theorem 2.1, and by Theorem 2.4.1 in [3] we have that, for $p \geq m$,

$$(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(XX')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})|_{\tilde{\boldsymbol{\Sigma}}^{-1}} \sim W_m\left(\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}, p\right).$$

From Theorem 2.4.2 in [3] and Subsection 7.3.3 in [1] we have

(A.5)
$$\mathbf{H} = \left(\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}\right)^{-\frac{1}{2}}(\overline{\mathbf{B}}^{\bullet}_M-\mathbf{B})'(\mathbf{XX}')(\overline{\mathbf{B}}^{\bullet}_M-\mathbf{B})\left(\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}\right)^{\prime-\frac{1}{2}} \sim W_m(\mathbf{I}, p)$$

and

$$\mathbf{G} = M(n-p)\tilde{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\overline{\mathbf{S}}^{\bullet}_M\tilde{\boldsymbol{\Sigma}}^{\prime-\frac{1}{2}} \sim W_m(\mathbf{I}, M(n-p)). \tag{A.6}$$

We may thus write $T^{\bullet}_M$ in (2.5) as

$$T^{\bullet}_M = \frac{|(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})'(XX')(\overline{\mathbf{B}}^{\bullet}_M - \mathbf{B})|}{|M(n-p)\overline{\mathbf{S}}^{\bullet}_M|} = \frac{\left|\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}\right|}{|\tilde{\boldsymbol{\Sigma}}|} \times \frac{|\mathbf{H}|}{|\mathbf{G}|},$$

where, $|\mathbf{G}| \sim \prod_{i=1}^m \chi^2_{n-p-i+1}$ and $|\mathbf{H}| \sim \prod_{i=1}^m \chi^2_{p-i+1}$, with independent chi-square random variables in each product, we end up with a product of independent F-distributions, due to the independence of $\mathbf{H}$ and $\mathbf{G}$, inherited from the independence of $\overline{\mathbf{B}}^{\bullet}_M$ and $\overline{\mathbf{S}}^{\bullet}_M$. So, conditionally on $\tilde{\boldsymbol{\Sigma}}^{-1}$, we have

$$T^{\bullet}_M|_{\tilde{\boldsymbol{\Sigma}}^{-1}} \sim \left\{\prod_{i=1}^m \frac{p-i+1}{M(n-p)-i+1}F_{p-i+1,n-p-i+1}\right\} \times \left|\tilde{\boldsymbol{\Sigma}}^{-1}\left(\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}\right)\right|,$$

where

$$\begin{aligned}\left|\tilde{\boldsymbol{\Sigma}}^{-1}\left(\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}\right)\right| &= \left|\frac{M+1}{M}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\right| = \left|\frac{M+1}{M}\boldsymbol{\Sigma}^{-1} + \tilde{\boldsymbol{\Sigma}}^{-1}\right||\boldsymbol{\Sigma}| \\ &= \left|\boldsymbol{\Sigma}^{1/2}\right|\left|\frac{M+1}{M}\boldsymbol{\Sigma}^{-1} + \tilde{\boldsymbol{\Sigma}}^{-1}\right|\left|\boldsymbol{\Sigma}^{1/2}\right| = \left|\frac{M+1}{M}\mathbf{I} + \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^{1/2}\right|.\end{aligned}$$

As such, from (A.4), integrating out $\overline{\mathbf{B}}^{\bullet}_M$ and $\overline{\mathbf{S}}^{\bullet}_M$, we end up with the pdf of $\tilde{\boldsymbol{\Sigma}}^{-1}$ proportional to

$$\begin{aligned}&|\tilde{\boldsymbol{\Sigma}}|^{\frac{M(n-p)}{2}}\left|\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}\right|^{\frac{p}{2}}\frac{1}{|\tilde{\boldsymbol{\Sigma}}|^{\frac{M(n-p)+n-\alpha}{2}-m-1}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \\ &\qquad\times\left|\frac{M}{M+1}\tilde{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}\right|^{-p/2}|\tilde{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}|^{-\frac{2n+\alpha-2p-m-1}{2}} \\ &= |\tilde{\boldsymbol{\Sigma}}^{-1}|^{\frac{n+\alpha-2m-2}{2}}\left|\frac{M+1}{M}\tilde{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}\right|^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \\ &\qquad\times\left|\frac{M}{M+1}\tilde{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}\right|^{-p/2}|\tilde{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}|^{-\frac{2n+\alpha-2p-m-1}{2}}.\end{aligned}$$

Making the transformation $\mathbf{\Omega}{=}\mathbf{\Sigma}^{\frac{1}{2}}\tilde{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}^{\frac{1}{2}}$, which implies $\tilde{\mathbf{\Sigma}}^{-1}{=}\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Omega}\mathbf{\Sigma}^{-\frac{1}{2}}$, with the Jacobian of the transformation from $\tilde{\mathbf{\Sigma}}^{-1}$ to $\mathbf{\Omega}$ being $|\mathbf{\Sigma}|^{-\frac{m+1}{2}}$, we have the pdf of $\mathbf{\Omega}$ proportional to

$$|\mathbf{\Omega}|^{\frac{n+\alpha-2m-2}{2}}\left|\frac{M+1}{M}\mathbf{\Omega^{-1}}+\mathbf{I_m}\right|^{\frac{p}{2}}\left|\frac{M}{M+1}\mathbf{\Omega}+\mathbf{I_m}\right|^{-p/2}|\mathbf{\Omega}+\mathbf{I_m}|^{-\frac{2n+\alpha-2p-m-1}{2}}.$$

Since $|\frac{M+1}{M}\mathbf{\Omega^{-1}}+\mathbf{I_m}|^{\frac{p}{2}}=\left(\frac{M+1}{M}\right)^{p/2}|\frac{M}{M+1}\mathbf{\Omega}+\mathbf{I_m}|^{\frac{p}{2}}|\mathbf{\Omega}|^{-\frac{p}{2}}$ we end up with

$$f(\mathbf{\Omega})\propto|\mathbf{\Omega}|^{\frac{n+\alpha-p-2m-2}{2}}\times|\mathbf{\Omega}+\mathbf{I_m}|^{-\frac{2n+\alpha-2p-m-1}{2}}$$

independent of $\mathbf{\Sigma}$. Therefore, we may conclude that

$$T_M^{\bullet}|_{\mathbf{\Omega}}\sim\left\{\prod_{i=1}^{m}\frac{p-i+1}{n-p-i+1}F_{p-i+1,M(n-p)-i+1}\right\}\left|\frac{M+1}{M}\mathbf{I_m}+\mathbf{\Omega}\right|$$

where from [6, Theorem 8.2.8.] $\mathbf{\Omega}$ has the same distribution as $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{-1}\mathbf{A}_1^{\frac{1}{2}}$ with $\mathbf{A}_1\sim W_m(\mathbf{I}_m,n+\alpha-p-m-1)$ and $\mathbf{A}_2\sim W_m(\mathbf{I}_m,n-p)$, two independent random variables. □

Proof of Corollary 2.3: The proof is identical to the proof of Theorem 2.1 replacing the joint pdf of $(\overline{\mathbf{B}}_M^{\bullet},\overline{\mathbf{S}}_M^{\bullet})$ by the joint pdf of $(\overline{\mathbf{B}}_M^{\bullet},\mathbf{S}_{comb}^{\bullet})$, noting that we have

$$(Mn-p)\mathbf{S}_{comb}^{\bullet}|_{\tilde{\mathbf{\Sigma}}}\sim W_m(\tilde{\mathbf{\Sigma}},Mn-p).\qquad\square$$

Proof of Corollary 2.4: The proof is identical to that of Theorem 2.2 replacing $\overline{\mathbf{S}}_M^{\bullet}$ by $\mathbf{S}_{comb}^{\bullet}$, noting that from Corollary 2.3, conditional on $\tilde{\mathbf{\Sigma}}$, $\overline{\mathbf{B}}_M^{\bullet}$ is $N_{pm}(\mathbf{B},(\mathbf{\Sigma}+\frac{\mathbf{1}}{\mathbf{M}}\tilde{\mathbf{\Sigma}})\otimes(\mathbf{XX'})^{\mathbf{-1}})$ and $(Mn-p)\mathbf{S}_{comb}^{\bullet}$ is $W_m(\tilde{\mathbf{\Sigma}},Mn-p)$, independent of $\overline{\mathbf{B}}_M^{\bullet}$. □

B. Details on several results

B.1. The posterior distributions for Σ and B

Let us start by observing that $\mathbf{Y}|_{\mathbf{B},\mathbf{\Sigma}}\sim N_{mn}(\mathbf{B'X},\mathbf{I_n}\otimes\mathbf{\Sigma})$ and that the likelihood function for $\mathbf{Y}$ will be

$$l(\mathbf{B},\mathbf{\Sigma}|y)\propto|\mathbf{\Sigma}|^{-n/2}e^{-\frac{1}{2}tr\{\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{B'X})(\mathbf{Y}-\mathbf{B'X})'\}}.$$

We may then get the joint posterior distribution of $(\mathbf{B},\mathbf{\Sigma})$ from the product of the prior and likelihood functions as

$$\pi(\mathbf{B},\mathbf{\Sigma}|y)\propto|\mathbf{\Sigma}|^{-\frac{n+\alpha}{2}}e^{-\frac{1}{2}tr\{\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{B'X})(\mathbf{Y}-\mathbf{B'X})'\}}.\qquad\text{(B.1)}$$

The exponent in (B.1) may be written as

$$\begin{aligned}
tr\{\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{B}'\mathbf{X})(\mathbf{Y}-\mathbf{B}'\mathbf{X})'\} &= tr\big\{\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X}+\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X}) \\
&\qquad\qquad \times(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X}+\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})'\big\} \\
= tr\left\{\boldsymbol{\Sigma}^{-1}\left[(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})'\right]\right\} \\
+ tr\Big\{\boldsymbol{\Sigma}^{-1}\big[(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})(\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})' &+ (\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})' \\
&+ (\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})(\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})'\big]\Big\} \\
= tr\left\{\boldsymbol{\Sigma}^{-1}\left[(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})'\right] + (\mathbf{B}-\hat{\mathbf{B}})'(\mathbf{X}\mathbf{X}')(\mathbf{B}-\hat{\mathbf{B}})\right\} \\
&+ 2tr\left\{\boldsymbol{\Sigma}^{-1}\left[(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})(\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})'\right]\right\},
\end{aligned}$$

where, using $\hat{\mathbf{B}}' = \left[(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{Y}'\right]' = \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-\mathbf{1}}$,

$$\begin{aligned}
(\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X})(\hat{\mathbf{B}}'\mathbf{X}-\mathbf{B}'\mathbf{X})' &= \mathbf{Y}\mathbf{X}'\hat{\mathbf{B}} - \mathbf{Y}\mathbf{X}'\mathbf{B} + \hat{\mathbf{B}}\mathbf{X}\mathbf{X}'\hat{\mathbf{B}} + \hat{\mathbf{B}}\mathbf{X}\mathbf{X}'\mathbf{B} \\
&= \mathbf{Y}\mathbf{X}'\hat{\mathbf{B}} - \mathbf{Y}\mathbf{X}'\mathbf{B} + \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-\mathbf{1}}\mathbf{X}\mathbf{X}'\hat{\mathbf{B}} \\
&\qquad + \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{X}'\mathbf{B} \\
&= \mathbf{Y}\mathbf{X}'\hat{\mathbf{B}} - \mathbf{Y}\mathbf{X}'\mathbf{B} - \mathbf{Y}\mathbf{X}'\hat{\mathbf{B}} + \mathbf{Y}\mathbf{X}'\mathbf{B} = 0.
\end{aligned}$$

Therefore, the joint posterior distribution of $(\mathbf{B},\boldsymbol{\Sigma})$ is proportional to

$$|\boldsymbol{\Sigma}|^{-\frac{n+\alpha-p}{2}} e^{-\frac{n-p}{2}tr\{\boldsymbol{\Sigma}^{-1}\mathbf{S}\}} \times |\boldsymbol{\Sigma}|^{-\frac{p}{2}} e^{-\frac{1}{2}tr\{\boldsymbol{\Sigma}^{-1}(\mathbf{B}-\hat{\mathbf{B}})'(\mathbf{X}\mathbf{X}')(\mathbf{B}-\hat{\mathbf{B}})\}}$$

In conclusion, by Corollary 2.4.6.2. in [3], the posterior distribution for $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma}|_{\mathbf{S}} \sim W_m^{-1}\left((n-p)\mathbf{S}, n+\alpha-p\right) \implies \boldsymbol{\Sigma}^{-1}|_{\mathbf{S}} \sim W_m\left(\frac{1}{n-p}\mathbf{S}^{-1}, n+\alpha-p-m-1\right)$$

and the posterior distribution for $\mathbf{B}$ is

$$\mathbf{B}|_{\hat{\mathbf{B}},\boldsymbol{\Sigma}} \sim N_{pm}(\hat{\mathbf{B}}, \boldsymbol{\Sigma}\otimes(\mathbf{X}\mathbf{X}')^{-\mathbf{1}}),$$

assuming $n+\alpha > p+m+1$.

B.2. Matrix calculations required in the proof of Theorem 2.1

For $\tilde{\mathbf{B}}$, $\mathbf{B}$ and $\mathbf{X}$ defined as in Section 2 we have

$$\begin{aligned}
M(\tilde{\mathbf{B}}-\overline{\mathbf{B}}^{\bullet}_M)'\mathbf{X}\mathbf{X}'(\tilde{\mathbf{B}}-\overline{\mathbf{B}}^{\bullet}_M) + (\tilde{\mathbf{B}}-\hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'(\tilde{\mathbf{B}}-\hat{\mathbf{B}}) = \\
= (M+1)\tilde{\mathbf{B}}'\mathbf{X}\mathbf{X}'\tilde{\mathbf{B}} - M\overline{\mathbf{B}}^{\bullet\prime}_M\mathbf{X}\mathbf{X}'\tilde{\mathbf{B}} - M\tilde{\mathbf{B}}'\mathbf{X}\mathbf{X}'\overline{\mathbf{B}}^{\bullet}_M + M\overline{\mathbf{B}}^{\bullet\prime}_M\mathbf{X}\mathbf{X}'\overline{\mathbf{B}}^{\bullet}_M \\
- \hat{\mathbf{B}}'\mathbf{X}\mathbf{X}'\tilde{\mathbf{B}} - \tilde{\mathbf{B}}'\mathbf{X}\mathbf{X}'\hat{\mathbf{B}} + \hat{\mathbf{B}}'\mathbf{X}\mathbf{X}'\hat{\mathbf{B}} \\
= (M+1)\tilde{\mathbf{B}}'\mathbf{X}\mathbf{X}'\tilde{\mathbf{B}} - \tilde{\mathbf{B}}'\mathbf{X}\mathbf{X}'(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}) - (M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})'\mathbf{X}\mathbf{X}'\tilde{\mathbf{B}} \\
+ M\overline{\mathbf{B}}^{\bullet\prime}_M\mathbf{X}\mathbf{X}'\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}'\mathbf{X}\mathbf{X}'\hat{\mathbf{B}}
\end{aligned}$$

$$= (M+1)\left[\tilde{\mathbf{B}} - \frac{1}{M+1}(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})\right]' \mathbf{XX}' \left[\tilde{\mathbf{B}} - \frac{1}{M+1}(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})\right]$$
$$+ M\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}'\mathbf{XX}'\hat{\mathbf{B}} - \frac{1}{M+1}(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})'\mathbf{XX}'(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}).$$

Since,

$$\begin{aligned}
&M\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}'\mathbf{XX}'\hat{\mathbf{B}} - \frac{1}{M+1}(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})'\mathbf{XX}'(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}) \\
&= M\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}}'\mathbf{XX}'\hat{\mathbf{B}} \\
&\qquad - \frac{M^2}{M+1}\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M - \frac{1}{M+1}\hat{\mathbf{B}}'\mathbf{XX}'\hat{\mathbf{B}} \\
&\qquad\qquad - \frac{M}{M+1}\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\hat{\mathbf{B}} - \frac{M}{M+1}\hat{\mathbf{B}}'\mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M \\
&= \frac{M}{M+1}\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M + \frac{M}{M+1}\hat{\mathbf{B}}'\mathbf{XX}'\hat{\mathbf{B}} - \frac{M}{M+1}\overline{\mathbf{B}}^{\bullet\prime}_M \mathbf{XX}'\hat{\mathbf{B}} \\
&\qquad\qquad - \frac{M}{M+1}\hat{\mathbf{B}}'\mathbf{XX}'\overline{\mathbf{B}}^{\bullet}_M \\
&= \frac{M}{M+1}(\overline{\mathbf{B}}^{\bullet}_M - \hat{\mathbf{B}})'\mathbf{XX}'(\overline{\mathbf{B}}^{\bullet}_M - \hat{\mathbf{B}})
\end{aligned}$$

we may write

$$\begin{aligned}
&M(\tilde{\mathbf{B}} - \overline{\mathbf{B}}^{\bullet}_M)'\mathbf{XX}'(\tilde{\mathbf{B}} - \overline{\mathbf{B}}^{\bullet}_M) + (\tilde{\mathbf{B}} - \hat{\mathbf{B}})'\mathbf{XX}'(\tilde{\mathbf{B}} - \hat{\mathbf{B}}) = \\
&= (M+1)\left[\tilde{\mathbf{B}} - \frac{1}{M+1}(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})\right]' \mathbf{XX}' \left[\tilde{\mathbf{B}} - \frac{1}{M+1}(M\overline{\mathbf{B}}^{\bullet}_M + \hat{\mathbf{B}})\right] \\
&\qquad\qquad + \frac{M}{M+1}(\overline{\mathbf{B}}^{\bullet}_M - \hat{\mathbf{B}})'\mathbf{XX}'(\overline{\mathbf{B}}^{\bullet}_M - \hat{\mathbf{B}})\,.
\end{aligned}$$

B.3. Details about the derivations of results 1, 2 and 5 in Section 2.1

Details on Result 1

From (A.4) we may immediately conclude that the MLE of $\mathbf{B}$ based on the synthetic data will be $\overline{\mathbf{B}}^{\bullet}_M$ with

$$E(\overline{\mathbf{B}}^{\bullet}_M) = (\mathbf{XX}')^{-1}\mathbf{X}\frac{1}{M}\sum_{j=1}^{M} E(\mathbf{W}'_j) = (\mathbf{XX}')^{-1}\mathbf{XX}'E(\tilde{\mathbf{B}}) = E(\hat{\mathbf{B}}) = \mathbf{B}$$

and

$$Var(\overline{\mathbf{B}}^{\bullet}_M) = Var[E(\overline{\mathbf{B}}^{\bullet}_M|\tilde{\mathbf{B}}, \tilde{\boldsymbol{\Sigma}})] + E[Var(\overline{\mathbf{B}}^{\bullet}_M|\tilde{\mathbf{B}}, \tilde{\boldsymbol{\Sigma}})]. \tag{B.2}$$

For the first term in (B.2), we have

$$Var[E(\overline{\mathbf{B}}^{\bullet}_{M}|\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}})] = Var[\tilde{\mathbf{B}}] = Var[E(\tilde{\mathbf{B}}|\hat{\mathbf{B}},\tilde{\mathbf{\Sigma}})] + E[Var(\tilde{\mathbf{B}}|\hat{\mathbf{B}},\tilde{\mathbf{\Sigma}})] =$$

$$= Var(\hat{\mathbf{B}}) + E[\tilde{\mathbf{\Sigma}}\otimes(\mathbf{XX}')^{-\mathbf{1}}] = \mathbf{\Sigma}\otimes(\mathbf{XX}')^{-\mathbf{1}} + \frac{n-p}{n+\alpha-p-2m-2}\mathbf{\Sigma}\otimes(\mathbf{XX}')^{-\mathbf{1}}$$

and for the second term, we have

$$E[Var(\overline{\mathbf{B}}^{\bullet}_{M}|\tilde{\mathbf{B}},\tilde{\mathbf{\Sigma}})] = E\left[\frac{1}{M}\tilde{\mathbf{\Sigma}}\otimes(\mathbf{XX}')^{-\mathbf{1}}\right] = \frac{1}{M}\frac{n-p}{n+\alpha-p-2m-2}\mathbf{\Sigma}\otimes(\mathbf{XX}')^{-\mathbf{1}},$$

so that

$$Var(\overline{\mathbf{B}}^{\bullet}_{M}) = \frac{2M(n-p-m-1)+n-p+M\alpha}{M(n+\alpha-p-2m-2)}\mathbf{\Sigma}\otimes(\mathbf{XX}')^{-\mathbf{1}}$$

under the condition that $n+\alpha > p+2m+2$.

Details on Result 2

$$E(\overline{\mathbf{S}}^{\bullet}_{M}) = E(\tilde{\mathbf{\Sigma}}) = E\left(\frac{n-p}{n+\alpha-p-2m-2}\mathbf{S}\right) = \frac{n-p}{n+\alpha-p-2m-2}\mathbf{\Sigma}.$$

Details on Result 5

Let us consider $\mathbf{H}$ and $\mathbf{G}$ given by (A.5) and (A.6). We will begin by rewriting all four classical statistics $T^{\bullet}_{1,M}$, $T^{\bullet}_{2,M}$, $T^{\bullet}_{3,M}$ and $T^{\bullet}_{4,M}$ in Subsection 2.1, in order to make them assume the same kind of form and then we will prove why all of them are non-pivotal, without loss of generality considering $M=1$. The first statistic, $T^{\bullet}_{1,M}$ may be rewritten as

$$T^{\bullet}_{1,1} = \frac{|\mathbf{G}|}{|\mathbf{G}+(n-p)\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\mathbf{H}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}|}.$$

while $T^{\bullet}_{2,M}$ and $T^{\bullet}_{3,M}$ may be rewritten as

$$T^{\bullet}_{2,1} = (n-p)tr\left[\mathbf{H}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}\mathbf{G}^{-1}\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\right],$$

$$T^{\bullet}_{3,1} = tr\{\mathbf{H}\times[\mathbf{H}+(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{-\mathbf{1/2}}\tilde{\mathbf{\Sigma}}^{\mathbf{1/2}}\times(n-p)\mathbf{G}\times\tilde{\mathbf{\Sigma}}^{\mathbf{1/2}}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{-\mathbf{1/2}}]^{-\mathbf{1}}\}.$$

Concerning $T^{\bullet}_{4,1}$, we have $T^{\bullet}_{4,1}=\lambda_1$ where λ_1 denotes the largest eigenvalue of

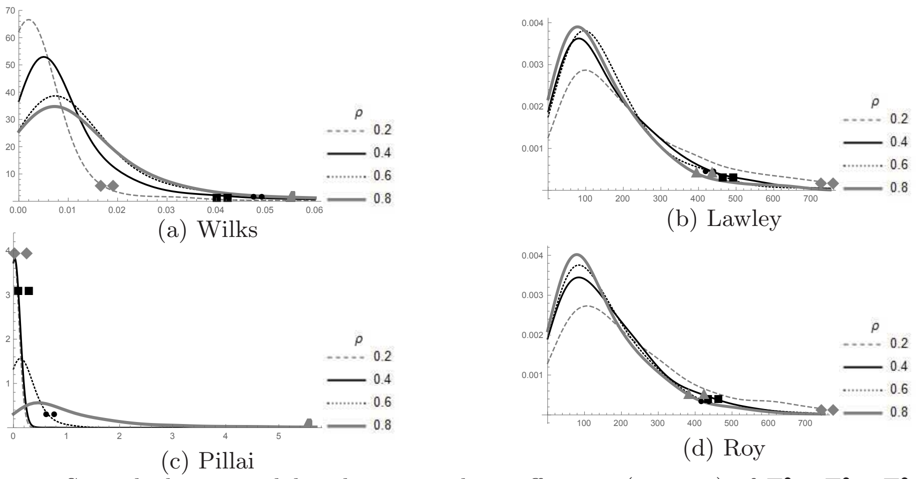
$$(n-p)\mathbf{H}\times(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}\times\mathbf{G}^{-\mathbf{1}}\times\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}.$$

We can observe that a term in the denominator of the expression $T^{\bullet}_{1,1}$ is

$$\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\mathbf{H}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{\mathbf{1/2}}\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}|_{\tilde{\mathbf{\Sigma}}^{-\mathbf{1}}} \sim W_m((\mathbf{2I}+\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}\mathbf{\Sigma}\tilde{\mathbf{\Sigma}}^{-\mathbf{1/2}}),p),$$

while in the expressions for the other statistics there are similar terms. These terms involve a product similar to $\tilde{\mathbf{\Sigma}}^{-1/2}(\mathbf{2}\tilde{\mathbf{\Sigma}}+\mathbf{\Sigma})^{1/2}$ that cannot be simplified to an expression which is not a function of $\mathbf{\Sigma}$, therefore making these statistics non-pivotal.

Thus, in order to illustrate how these statistics are dependent on $\mathbf{\Sigma}$, we can analyze in Figure 3 the empirical distributions of $T^{\bullet}_{1,1}$, $T^{\bullet}_{2,1}$, $T^{\bullet}_{3,1}$ and $T^{\bullet}_{4,1}$ when we consider a simple case where $m = 2$, $p = 3$, $\alpha = 4$, $n = 100$ and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho = \{0.2, 0.4, 0.6, 0.8\}$ for a simulation size of 1000.



(a) Wilks (b) Lawley (c) Pillai (d) Roy

Figure 3: Smoothed empirical distributions and cut-off points ($\gamma = 0.05$) of $T^{\bullet}_{1,1}$, $T^{\bullet}_{2,1}$, $T^{\bullet}_{3,1}$ and $T^{\bullet}_{4,1}$ for $\rho = \{0.2, 0.4, 0.6, 0.8\}$.

B.4. Details about the derivation of result 1 in Subsection 2.2

Recalling that $(Mn - p)\mathbf{S}^{\bullet}_{comb}|_{\tilde{\mathbf{\Sigma}}} \sim W_m(\tilde{\mathbf{\Sigma}}, Mn - p)$ and that $\tilde{\mathbf{\Sigma}}^{-1}|_{\mathbf{S}} \sim W_m(\frac{1}{n-p}\mathbf{S}^{-1}, n + \alpha - p - m - 1)$ we immediately obtain

$$E(\mathbf{S}^{\bullet}_{comb}) = E(\tilde{\mathbf{\Sigma}}) = E\left(\frac{n-p}{n+\alpha-p-2m-2}\mathbf{S}\right) = \frac{n-p}{n+\alpha-p-2m-2}\mathbf{\Sigma}.$$

B.5. Details about the derivations of the results in Section 3

Details on the Expected Values in Section 3

Recall that $(n-p)\mathbf{S} \sim W_m(\mathbf{\Sigma}, n-p)$, thus implying that

$$E(|(n-p)\mathbf{S}|) = |\mathbf{\Sigma}|E(\prod_{i=1}^{m}\chi^2_{n-p-i+1}) = \frac{(n-p)!}{(n-p-m)!}|\mathbf{\Sigma}|,$$

and recall that

$$\tilde{\mathbf{\Sigma}}|_{\mathbf{S}} \sim W_m^{-1}((n-p)\mathbf{S}, n+\alpha-p) \implies \tilde{\mathbf{\Sigma}}^{-1}|_{\mathbf{S}} \sim W_m\left(\frac{1}{n-p}\mathbf{S}^{-1}, n+\alpha-p-m-1\right)$$

thus implying that, making $\kappa_{n,\alpha,p,m} = n+\alpha-p-m-1$, given $\mathbf{S}$,

$$\begin{aligned} E(|\tilde{\mathbf{\Sigma}}|) = E(|\tilde{\mathbf{\Sigma}}^{-1}|^{-1}) &= |(n-p)\mathbf{S}|E\left(\frac{1}{\prod_{i=1}^m \chi^2_{\kappa_{n,\alpha,p,m}-i+1}}\right) \\ &= |(n-p)\mathbf{S}|\frac{(-2+\kappa_{n,\alpha,p,m}-m)!}{(-2+\kappa_{n,\alpha,p,m})!}, \end{aligned}$$

since $\prod_{i=1}^m \chi^2_{\kappa_{n,\alpha,p,m}-i+1}$ is a product of independent χ^2 variables. Also recalling that, given $\tilde{\mathbf{\Sigma}}$, we have $M(n-p)\overline{\mathbf{S}}^{\bullet}{}_M \sim W_m(\tilde{\mathbf{\Sigma}}, M(n-p))$ and $(Mn-p)\mathbf{S}^{\bullet}_{comb} \sim W_m(\tilde{\mathbf{\Sigma}}, Mn-p)$, we may conclude that, given $\tilde{\mathbf{\Sigma}}$,

$$E(|M(n-p)\mathbf{S}^{\bullet}_M|) = \frac{(Mn-Mp)!}{(Mn-Mp-m)!} \times |\tilde{\mathbf{\Sigma}}|$$

and

$$E(|(Mn-p)\mathbf{S}^{\bullet}_{comb}|) = \frac{(Mn-p)!}{(Mn-p-m)!} \times |\tilde{\mathbf{\Sigma}}|.$$

Combining the results for $E(|(n-p)\mathbf{S}|)$ and $E(|\tilde{\mathbf{\Sigma}}|)|_{\mathbf{S}}$ with each of the expected values for $|M(n-p)\mathbf{S}^{\bullet}_M|$ and $|(Mn-p)\mathbf{S}^{\bullet}_{comb}|$, we end up with the expression for $E(\Upsilon_M)$ found in Section 3.

C. Joining multiple datasets into a single dataset

Let us consider the M synthetic datasets as one only dataset of size nM

$$\begin{pmatrix} \mathbf{W}_a \\ \mathbf{X}_a \end{pmatrix} = \left(\begin{array}{c|c|c|c} \mathbf{W}_1 & \mathbf{W}_2 & \ldots & \mathbf{W}_M \\ \hline \mathbf{X} & \mathbf{X} & \ldots & \mathbf{X} \end{array} \right),$$

where $\mathbf{W}_a = (\mathbf{W}_1|...|\mathbf{W}_M)$ is the $m \times nM$ matrix of the synthesized data under FPPS and $\mathbf{X}_a = (\mathbf{X}|...|\mathbf{X})$ the $p \times nM$ matrix of the M repeated 'fixed' sets of covariates, from the original data.

Let

$$\mathbf{B}_a = (\mathbf{X}_a\mathbf{X}_a')^{-1}\mathbf{X}_a\mathbf{W}_a'$$

be the estimator for $\mathbf{B}$, based on the dataset of size nM, obtained by joining the M synthetic datasets in one only dataset. Consequently one has that

$$\begin{aligned} \mathbf{B}_a &= (\mathbf{X}_a\mathbf{X}_a')^{-1}\mathbf{X}_a\mathbf{W}_a' = (M(\mathbf{X}\mathbf{X}'))^{-1}\mathbf{X}_a\mathbf{W}_a' = \frac{1}{M}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}_a\mathbf{W}_a' \\ &= \frac{1}{M}(\mathbf{X}\mathbf{X}')^{-1}\Big(\underbrace{\mathbf{X}| \ \ldots \ |\mathbf{X}}_{M\ times}\Big)\mathbf{W}_a' = \frac{1}{M}\left((\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{W}_1 + \ldots + (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{W}_M\right) \\ &= \frac{1}{M}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\left(\mathbf{W}_1 + \ldots + \mathbf{W}_M\right) = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\overline{\mathbf{W}}_M = \overline{\mathbf{B}}^{\bullet}_M, \end{aligned}$$

which is same estimator for $\mathbf{B}$ as in (2.8).

Now let

$$\mathbf{S}_a = \frac{1}{nM-p}(\mathbf{W}_a - \mathbf{B}_a'\mathbf{X}_a)(\mathbf{W}_a - \mathbf{B}_a'\mathbf{X}_a)'$$

be the estimator for $\boldsymbol{\Sigma}$, based on the dataset of size nM, obtained by joining the M synthetic datasets in one only dataset.

Observe that $\overline{\mathbf{W}}_M = \frac{1}{M}\sum_{j=1}^{M}\mathbf{W}_j$, defined before expression (2.8), can be written as

$$\overline{\mathbf{W}}_M = \frac{1}{M}\mathbf{W}_a\mathbf{R}$$

with $\mathbf{R} = \left(\overrightarrow{\mathbf{1}}_M \otimes \mathbf{I}_n\right)$ where $\overrightarrow{\mathbf{1}}_M$ is a vector of 1's of size M.

Now let us consider the estimator $\mathbf{S}_\mathbf{w}$ of $\boldsymbol{\Sigma}$, defined in the text, before expression (2.8). This estimator may be written as

$$\mathbf{S}_\mathbf{w} = \sum_{i=1}^{n}\sum_{j=1}^{M}(\mathbf{w}_{ji} - \overline{\mathbf{w}}_i)(\mathbf{w}_{ji} - \overline{\mathbf{w}}_i)',$$

where $\mathbf{w}_{ji}$ is the i-th column of $\mathbf{W}_j$ $(i = 1, \ldots, n; j = 1, \ldots, M)$. We may thus write

$$\begin{aligned}
\mathbf{S}_\mathbf{w} &= \left(\mathbf{W}_a - \overrightarrow{\mathbf{1}}'_M \otimes \overline{\mathbf{W}}_M\right)\left(\mathbf{W}_a - \overrightarrow{\mathbf{1}}'_M \otimes \overline{\mathbf{W}}_M\right)' \\
&= \left(\mathbf{W}_a - \frac{1}{M}\overrightarrow{\mathbf{1}}'_M \otimes (\mathbf{W}_a\mathbf{R})\right)\left(\mathbf{W}_a - \frac{1}{M}\overrightarrow{\mathbf{1}}'_M \otimes (\mathbf{W}_a\mathbf{R})\right)' \\
&= \left(\mathbf{W}_a - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\right)\left(\mathbf{W}_a - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\right)'
\end{aligned}$$

and the estimator $\mathbf{S}_{mean}$ of $\boldsymbol{\Sigma}$, defined right after expression (2.9) as

$$\mathbf{S}_{mean} = \left(\frac{1}{M}\mathbf{W}_a\mathbf{R} - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\right)\left(\frac{1}{M}\mathbf{W}_a\mathbf{R} - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\right)'.$$

We may therefore write the combination estimator $\mathbf{S}_{comb}$ defined in (2.9) as

$$\begin{aligned}
\mathbf{S}_{comb} = \frac{1}{nM-p}&\left[\left(\mathbf{W}_a - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\right)\left(\mathbf{W}_a - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\right)'\right] \\
&+\frac{1}{nM-p}\left[M \times \left(\frac{1}{M}\mathbf{W}_a\mathbf{R} - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\right)\left(\frac{1}{M}\mathbf{W}_a\mathbf{R} - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\right)'\right]
\end{aligned}$$

To prove that $\mathbf{S}_{comb}$ is equal to $\mathbf{S}_a$ it will only be necessary to focus on

$$\begin{aligned}
&\left(\mathbf{W}_a - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\right)\left(\mathbf{W}_a - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\right)' \\
&\qquad +M \times \left(\frac{1}{M}\mathbf{W}_a\mathbf{R} - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\right)\left(\frac{1}{M}\mathbf{W}_a\mathbf{R} - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\right)'
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{W}_a\mathbf{W}_a' - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' + \frac{1}{M^2}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{R}\mathbf{R}'\mathbf{W}_a' \\
&\qquad + \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' - \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' \\
&\qquad - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{X}_a'\mathbf{B}_a + \frac{1}{M}\mathbf{B}_a'\mathbf{X}_a\mathbf{R}\mathbf{R}'\mathbf{X}_a'\mathbf{B}_a\,,
\end{aligned}
$$

which, using the fact that $\frac{1}{M}\mathbf{X}_a\mathbf{R}\mathbf{R}' = \mathbf{X}_a$ and $\frac{1}{M}\mathbf{R}\mathbf{R}'\mathbf{R}\mathbf{R}' = \mathbf{R}\mathbf{R}'$, may be written as

$$
\begin{aligned}
&\mathbf{W}_a\mathbf{W}_a' - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' - \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' + \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' \\
&\qquad + \frac{1}{M}\mathbf{W}_a\mathbf{R}\mathbf{R}'\mathbf{W}_a' - \mathbf{B}_a'\mathbf{X}_a\mathbf{W}_a' - \mathbf{W}_a\mathbf{X}_a'\mathbf{B}_a + \mathbf{B}_a'\mathbf{X}_a\mathbf{X}_a'\mathbf{B}_a \\
&\quad = \mathbf{W}_a\mathbf{W}_a' - \mathbf{B}_a'\mathbf{X}_a\mathbf{W}_a' - \mathbf{W}_a\mathbf{X}_a'\mathbf{B}_a + \mathbf{B}_a'\mathbf{X}_a\mathbf{X}_a'\mathbf{B}_a \\
&\quad = (\mathbf{W}_a - \mathbf{B}_a'\mathbf{X}_a)(\mathbf{W}_a - \mathbf{B}_a'\mathbf{X}_a)' = (nM-p)\mathbf{S}_a\,.
\end{aligned}
$$

REFERENCES

[1] Anderson, T.W. (2003). *An Introduction To Multivariate Statistical Analysis*, 3rd ed., Wiley, New Jersey.

[2] Klein, M. and Sinha, B. (2015). Inference for singly imputed synthetic data based on posterior predictive sampling under multivariate normal and multiple linear regression models. *Sankhya* B, **77**, 2, 293–311.

[3] Kollo, T. and Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, Springer, New York

[4] Little, R. (1993). Statistical analysis of masked data. *Journal of Official Statistics*, **9**, 407–426.

[5] Moura, R., Klein, M., Coelho, C. A. and Sinha, B. (2016). Inference for multivariate regression model based on synthetic data generated using plug-in sampling. *Technical Report - Centro de Matemática e Aplicações (CMA)* 2/2016, (submited for publication).

[6] Muirhead, R.J. (1985). *Aspects of Multivariate Statistical Theory*, 2nd ed., John Wiley & Sons, Inc., New Jersey.

[7] Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009. *Official Journal of the European Union*, **L 87**, 164–173.

[8] Reiter, J. (2003). Inference for partially synthetic public use microdata sets. *Survey Methodology*, **29**, 181–188.

[9] Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley, New Jersey.

[10] Rubin, D. (1993). Discussion: Statistical disclosure limitation. *Journal of Official Statistics*, **9**, 461–468.

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 187-204

VALUE-AT-RISK ESTIMATION AND THE PORT MEAN-OF-ORDER-P METHODOLOGY

Authors: Fernanda Figueiredo
– Faculdade de Economia, Universidade do Porto, and CEAUL, Universidade de Lisboa, Portugal (otilia@fep.up.pt)

M. Ivette Gomes
– CEAUL and DEIO, FCUL, Universidade de Lisboa, Portugal (ivette.gomes@fc.ul.pt)

Lígia Henriques-Rodrigues
– Universidade de São Paulo, IME, São Paulo, Brasil, and CEAUL, Portugal (ligiahr@ime.usp.br)

Received: May 2016 Revised: October 2016 Accepted: November 2016

Abstract:

- In finance, insurance and statistical quality control, among many other areas of application, a typical requirement is to estimate the *value-at-risk* (VaR) at a small level q, i.e. a high quantile of probability $1-q$, a value, high enough, so that the chance of an exceedance of that value is equal to q, small. The semi-parametric estimation of high quantiles depends strongly on the estimation of the *extreme value index* (EVI), the primary parameter of extreme events. And most semi-parametric VaR-estimators do not enjoy the adequate behaviour, in the sense that they do not suffer the appropriate linear shift in the presence of linear transformations of the data. Recently, and for heavy tails, i.e. for a positive EVI, new VaR-estimators were introduced with such a behaviour, the so-called PORT VaR-estimators, with PORT standing for *peaks over a random threshold*. Regarding EVI-estimation, new classes of PORT-EVI estimators, based on a powerful generalization of the Hill EVI-estimator related to adequate *mean-of-order-p* (MO_p) EVI-estimators, were even more recently introduced. In this article, also for heavy tails, we introduce a new class of PORT-MO_p VaR-estimators with the above mentioned behaviour, using the PORT-MO_p class of EVI-estimators. Under convenient but soft restrictions on the underlying model, these estimators are consistent and asymptotically normal. The behaviour of the PORT-MO_p VaR-estimators is studied for finite samples through Monte-Carlo simulation experiments.

Key-Words:

- *asymptotic behaviour; heavy tails; high quantiles; mean-of-order-p estimation; Monte-Carlo simulation;* PORT *methodology; semi-parametric methods; statistics of extremes; value-at-risk.*

AMS Subject Classification:

- 62G32, 62E20; 65C05.

1. INTRODUCTION AND SCOPE OF THE ARTICLE

In the field of *extreme value theory* (EVT) it is usually said that a *cumulative distribution function* (CDF) F has a heavy right-tail whenever the right tail function, given by $\overline{F} := 1-F$, is a regularly varying function with a negative index of regular variation $\alpha = -1/\xi$, i.e. for every $x > 0$, $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = x^{-1/\xi}$, $\xi > 0$. Then we are in the domain of attraction for maxima of an *extreme value* (EV) CDF,

$$\mathrm{EV}_\xi(x) = \exp(-(1+\xi x)^{-1/\xi}), \quad x > -1/\xi, \quad \xi > 0, \tag{1.1}$$

and we write $F \in \mathcal{D}_\mathcal{M}(\mathrm{EV}_{\xi>0})$. More generally, we can have $\xi \in \mathbb{R}$, i.e. the CDF $\mathrm{EV}_\xi(x) = \exp\big(-(1+\xi x)^{-1/\xi}\big)$, $1+\xi x > 0$, if $\xi \neq 0$, and by continuity the so-called Gumbel CDF, $\mathrm{EV}_0(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$, for $\xi = 0$. The parameter ξ is the *extreme value index* (EVI), one of the primary parameters in probabilistic and statistical EVT.

In a context of heavy tails, and with the notation $U(t) := F^{\leftarrow}(1-1/t)$, $t \geq 1$, $F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$ the generalized inverse function of the underlying model F, the positive EVI appears, for every $x > 0$, as the limiting value, as $t \to \infty$, of the quotient $(\ln U(tx) - \ln U(t))/\ln x$ (de Haan, 1970). Indeed, with the usual notation $\mathcal{R}_\mathrm{a}$ for the class of regularly varying functions with an index of regular variation a, we can further say that

$$F \in \mathcal{D}_\mathcal{M}^+ := \mathcal{D}_\mathcal{M}(\mathrm{EV}_{\xi>0}) \iff \overline{F} = 1-F \in \mathcal{R}_{-1/\xi} \iff U \in \mathcal{R}_\xi, \tag{1.2}$$

with the first necessary and sufficient condition given in Gnedenko (1943) and the second one in de Haan (1984). Heavy-tailed distributions have recently been accepted as realistic models for various phenomena in the most diverse areas of application, among which we mention bibliometrics, biometry, economics, ecology, finance, insurance, and statistical quality control.

For small values of a level q, and as usual in the area of statistical EVT, we want to extrapolate beyond the sample, estimating the *value-at-risk* (VaR) at a level q, denoted by VaR_q, or equivalently, a high quantile χ_{1-q}, i.e. a value such that $F(\chi_{1-q}) = 1-q$, i.e.

$$\mathrm{VaR}_q \equiv \chi_{1-q} := U(1/q), \quad q = q_n \to 0, \quad \text{as } n \to \infty. \tag{1.3}$$

We further often assume that $nq_n \to K$ as $n \to \infty$, $K \in [0,\ 1]$, and base inference on the $k+1$ upper *order statistics* (OSs). As usual in semi-parametric estimation of parameters of extreme events, we shall assume that k is an *intermediate* sequence of integers in $[1, n[$, i.e.

$$k = k_n \to \infty \ \text{ and } \ k/n \to 0, \quad \text{as} \quad n \to \infty. \tag{1.4}$$

To derive the asymptotic non-degenerate behaviour of estimators of parameters of extreme events under a semi-parametric framework, it is further convenient to assume a bit more than the first-order condition, $U \in \mathcal{R}_\xi$, provided in (1.2). A common condition for heavy tails, also assumed now, is the second-order condition that guarantees that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \xi \ln x}{A(t)} = \begin{cases} \frac{x^\rho - 1}{\rho}, & \text{if } \rho < 0, \\ \ln x, & \text{if } \rho = 0, \end{cases} \tag{1.5}$$

being ρ (≤ 0). Note that the limit function in (1.5) is necessarily of the given form and $|A| \in \mathcal{R}_\rho$ (Geluk and de Haan, 1987). Sometimes, for sake of simplicity and for technical reasons, we assume to be working in a sub-class of Hall-Welsh class of models (Hall and Welsh, 1985), where there exist $\xi > 0$, $\rho < 0$, $C > 0$ and $\beta \neq 0$, such that, as $t \to \infty$,

$$U(t) = C\, t^\xi \Big(1 + A(t)\big(1 + o(1)\big)/\rho\Big), \quad \text{with} \quad A(t) = \xi\beta t^\rho. \tag{1.6}$$

The parameters β and ρ are the so-called generalized scale and shape second-order parameters, respectively. Typical heavy-tailed models, like the $\mathrm{EV}_{\xi>0}$ in (1.1) $(\rho = -\xi)$, the Fréchet CDF, $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $x \geq 0$, $\alpha > 0$ $(\xi = 1/\alpha$, $\rho = -1)$, the Generalized Pareto, $\mathrm{GP}_{\xi>0}(x) = 1 + \ln \mathrm{EV}_\xi(x)$, $x \geq 0$ $(\rho = -\xi)$, and the well-known Student-t_ν $(\xi = 1/\nu,\ \rho = -2/\nu)$ belong to such a class. Then, the second-order condition in equation (1.5) holds, with $A(t) = \xi\beta t^\rho$, $\beta \neq 0$, $\rho < 0$, as given in (1.6). Further details on these semi-parametric frameworks can be seen in Beirlant *et al.* (2004), de Haan and Ferreira (2006) and Fraga Alves *et al.* (2007), among others. Semi-parametric statistical choice tests of $F \in \mathcal{D}^+_\mathcal{M}$ can be seen in Fraga Alves and Gomes (1996) and Dietrich *et al.* (2002), also among others.

Under the validity of condition (1.6), and using the notation $a(t) \sim b(t)$ if and only if $\lim_{t\to\infty} a(t)/b(t) = 1$, we can guarantee that $U(t) \sim Ct^\xi$, as $t \to \infty$, and from (1.3), we have

$$\mathrm{VaR}_q = U(1/q) \ \sim\ Cq^{-\xi}, \quad \text{as} \quad q \to 0.$$

An obvious estimator of VaR_q is thus $\widehat{C}q^{-\hat{\xi}}$, with $\widehat{C}$ and $\hat{\xi}$ any consistent estimators of C and ξ, respectively. Given a sample $\underline{\mathbf{X}}_n := (X_1, \ldots, X_n)$, let us denote $(X_{1:n} \leq \cdots \leq X_{n:n})$ the set of associated ascending OSs. A common estimator of C, proposed in Hall (1982), is

$$\widehat{C} \equiv C_{k,n,\hat{\xi}} := X_{n-k:n}(k/n)^{\hat{\xi}}$$

and

$$Q_{k,q,\hat{\xi}} = \widehat{C}\ q^{-\hat{\xi}} = X_{n-k:n}\big(k/(nq)\big)^{\hat{\xi}} \tag{1.7}$$

is the straightforward VaR-estimator at the level q (Weissman, 1978). In classical approaches, we often consider for $\hat{\xi}$ the Hill (H) estimator (Hill, 1975), the average of the log-excesses, i.e.

$$\mathrm{H}_k \equiv \mathrm{H}_k(\underline{\mathbf{X}}_n) := \frac{1}{k}\sum_{i=1}^{k} \left(\ln X_{n-i+1:n} - \ln X_{n-k:n}\right). \tag{1.8}$$

But the Hill EVI-estimator is the logarithm of the *geometric mean* (or *mean of order* 0) of

$$(1.9) \qquad U_{ik} := X_{n-i+1:n}/X_{n-k:n}, \quad 1 \le i \le k < n.$$

It is thus sensible to consider the *mean-of-order-p* (MO_p) of U_{ik}, $1 \le i \le k$, as done in Brilhante *et al.* (2013), for $p \ge 0$, and in Gomes and Caeiro (2014) for any $p \in \mathbb{R}$. See also, Paulauskas and Vaičiulis (2013, 2015), Beran *et al.* (2014), Gomes *et al.* (2015a, 2016a) and Caeiro *et al.* (2016a). We then more generally get the class of MO_p EVI-estimators,

$$(1.10) \qquad \mathrm{H}_k(p) = \mathrm{H}_k(p; \underline{\mathbf{X}}_n) := \begin{cases} \frac{1}{p}\left(1 - k/\sum\limits_{i=1}^{k} U_{ik}^p\right), & \text{if } p < 1/\xi,\ p \ne 0, \\ \mathrm{H}_k, & \text{if } p = 0, \end{cases}$$

with $\mathrm{H}_k(0) \equiv \mathrm{H}_k$, given in (1.8), and U_{ik} given in (1.9), $1 \le i \le k < n$. Associated PORT MO_p VaR-estimators are thus a sensible generalization of the Weissman-Hill VaR-estimators.

The MO_p EVI-estimators, in (1.10), depend now on this *tuning* parameter $p \in \mathbb{R}$, are highly flexible, but, as often desirable, they are not location-invariant, depending strongly on possible shifts in the underlying data model. Also, most of the semi-parametric VaR-estimators in the literature, like the ones in Beirlant *et al.* (2008), Caeiro and Gomes (2008), the MO_p VaR-estimators in Gomes *et al.* (2015b), as well as in other papers on semi-parametric quantile estimation prior to 2008 (see also, the functional equation in (1.7), Beirlant *et al.*, 2004, and de Haan and Ferreira, 2006), do not enjoy the adequate behaviour in the presence of linear transformations of the data, a behaviour related to the fact that for any high-quantile, VaR_q, we have

$$(1.11) \qquad \mathrm{VaR}_q(\lambda + \delta X) = \lambda + \delta \mathrm{VaR}_q(X)$$

for any model X, real λ and positive δ. Recently, and for $\xi > 0$, Araújo Santos *et al.* (2006) provided VaR-estimators with the linear property in (1.11), based on a *sample of excesses* over a random threshold $X_{n_s:n}$, $n_s := \lfloor ns \rfloor + 1$, $0 \le s < 1$, where $\lfloor x \rfloor$ denotes the integer part of x, being s possibly null only when the underlying parent has a finite left endpoint (see Gomes *et al.*, 2008b, for further details on this subject). Those VaR-estimators are based on the sample of size $n^{(s)} = n - n_s$, defined by

$$(1.12) \qquad \underline{\mathbf{X}}_n^{(s)} := \big(X_{n:n} - X_{n_s:n}, \ldots, X_{n_s+1:n} - X_{n_s:n}\big).$$

Such estimators were named PORT-VaR estimators, with PORT standing for *peaks over a random threshold*, and were based on the PORT-Hill, $\mathrm{H}_k(\underline{\mathbf{X}}_n^{(s)})$, $k < n - n_s$, with $\mathrm{H}_k(\underline{\mathbf{X}}_n)$ provided in (1.8). Now, we further suggest for an adequate VaR-estimation, the use of the PORT-MO_p EVI-estimators,

$$(1.13) \qquad \mathrm{H}_k(p,s) := \mathrm{H}_k\big(p; \underline{\mathbf{X}}_n^{(s)}\big), \quad k < n - n_s,$$

introduced and studied both theoretically and for finite samples in Gomes *et al.* (2016c), with $\mathrm{H}_k(p; \underline{\mathbf{X}}_n)$ and $\underline{\mathbf{X}}_n^{(s)}$ respectively provided in (1.10) and (1.12). Such PORT-MO_p VaR-estimators are given by

$$\widehat{\mathrm{VaR}}_q(k;p,s) := (X_{n-k:n} - X_{n_s:n}) \left(\frac{k}{nq}\right)^{\mathrm{H}_k(p,s)} + X_{n_s:n}. \tag{1.14}$$

Under convenient restrictions on the underlying model, this class of VaR-estimators is consistent and asymptotically normal for adequate k, with $k+1$ the number of upper OSs used in the semi-parametric estimation of VaR_q.

In Section 2 of this paper, and following closely Henriques-Rodrigues and Gomes (2009) and Gomes *et al.* (2016c), we present a few introductory technical details and asymptotic results associated with the PORT methodology. A few comments on the asymptotic behaviour of the PORT-classes of VaR-estimators under study will be provided in Section 3. In Section 4, through the use of Monte-Carlo simulation techniques, we shall exhibit the performance of the PORT-MO_p VaR-estimators in (1.14), comparatively to the classical Weissman-Hill, MO_p and a PORT version of the most simple *reduced-bias* (RB) VaR-estimators in Gomes and Pestana (2007). In Section 5, we refer possible methods for the adaptive choice of the tuning parameters (k,p,s), either based on the bootstrap or on heuristic methodologies, and provide some concluding remarks.

2. A FEW TECHNICAL DETAILS ASSOCIATED WITH THE PORT METHODOLOGY

First note that if there is a shift $\lambda \in \mathbb{R}$ in the model, i.e. if the CDF $F(x) = F_\lambda(x) = F_0(x-\lambda)$, the EVI does not change with λ. Indeed, if a shift λ is induced in data associated with a *random variable* (RV) X, i.e. if we consider $Y = X + \lambda$, $U_\lambda(t) \equiv U_Y(t) = U_X(t) + \lambda$. Consequently, and due to the fact that $F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$ if and only if there exists a function $a(\cdot)$ such that

$$\frac{U(tx)-U(t)}{a(t)} \underset{t\to\infty}{\longrightarrow} (x^\xi - 1)/\xi \quad \text{(de Haan, 1984)},$$

the EVI, ξ, does not depend on any shift λ, i.e., $U_\lambda \in \mathcal{R}_\xi$. However, the same does not happen to the second-order parameters. Indeed, condition (1.5) can be rewritten as

$$\lim_{t\to\infty} \frac{\ln U_\lambda(tx) - \ln U_\lambda(t) - \xi \ln x}{A_\lambda(t)} = \frac{x^{\rho_\lambda} - 1}{\rho_\lambda}, \tag{2.1}$$

for all $x > 0$, with $|A_\lambda| \in \mathcal{R}_{\rho_\lambda}$, and for $\lambda \neq 0$,

$$\rho_\lambda = \begin{cases} \rho_0 & \text{if } \rho_0 > -\xi, \\ -\xi & \text{if } \rho_0 \le -\xi. \end{cases}$$

Furthermore, and again for $\lambda \neq 0$, the function $A_\lambda(t)$ in (2.1) can be chosen as

$$A_\lambda(t) := \begin{cases} -\frac{\xi\ \lambda}{U_0(t)} & \text{if } \rho_0 < -\xi, \\ A_0(t) - \frac{\xi\ \lambda}{U_0(t)} & \text{if } \rho_0 = -\xi, \\ A_0(t) & \text{if } \rho_0 > -\xi. \end{cases} \tag{2.2}$$

In Hall-Welsh class of models, in (1.6), we can thus consider the parameterization $A_\lambda(t) = \xi\beta_\lambda t^{\rho_\lambda}$. Further details on the influence of such a shift in $\big(\beta_0, \rho_0, A_0(\cdot)\big)$ and on the estimation of generalized shape and scale second-order parameters can be found in Henriques-Rodrigues *et al.* (2014, 2015).

2.1. Asymptotic behaviour of the PORT EVI-estimators

In this section we present, under the validity of the second-order condition in (1.5), the asymptotic distributional representations of the PORT-MO$_p$ EVI-estimators, $\mathrm{H}_k(p,s)$, in (1.13). Generalizing the results of Theorem 2.1 in Araújo Santos *et al.* (2006), and on the basis of the asymptotic behaviour of the MO$_p$ EVI-estimators derived in Brilhante *et al.* (2013), Gomes *et al.* (2016c), proved the following theorem:

Theorem 2.1 (Gomes *et al.*, 2016c)**.** *If the second order condition* (1.5) *holds, $k = k_n$ is an intermediate sequence of positive integers, i.e.* (1.4) *holds, for any real s, $0 \le s < 1$, with $\chi_s := F^{\leftarrow}(s)$, finite, we have for* $\mathrm{H}_k(p,s)$, *in* (1.13), *an asymptotic distributional representation of the type,*

$$\mathrm{H}_k(p,s) \stackrel{d}{=} \xi + \frac{\sigma_{\mathrm{H}(p)}\, P_k^{\mathrm{H}(p)}}{\sqrt{k}} + \left(b_{\mathrm{H}(p)} A_0(n/k) + \frac{c_{\mathrm{H}(p)}\chi_s}{U_0(n/k)}\right)(1 + o_p(1)), \tag{2.3}$$

where $P_k^{\mathrm{H}(p)}$ is a sequence of asymptotically standard normal RVs,

$$\sigma_{\mathrm{H}(p)} := \frac{\xi(1-p\xi)}{\sqrt{1-2p\xi}}, \quad b_{\mathrm{H}(p)} := \frac{1-p\xi}{1-p\xi-\rho}, \quad c_{\mathrm{H}(p)} := \frac{\xi(1-p\xi)}{1-(p-1)\xi}. \tag{2.4}$$

3. ASYMPTOTIC BEHAVIOUR OF PORT VAR–ESTIMATORS

Assuming that we are working with data from $F_\lambda(x) = F_0(x-\lambda)$, i.e. an underlying model with location parameter $\lambda \in \mathbb{R}$, we first present the following result on the asymptotic behaviour of intermediate OSs, proved in Ferreira *et al.* (2003).

Proposition 3.1 (Ferreira *et al.*, 2003)**.** *Under the second-order framework in* (2.1) *and for intermediate sequences of positive integers k, i.e. if* (1.4) *holds,*

$$X_{n-k:n} \stackrel{d}{=} U_\lambda(n/k)\Big(1 + \frac{\xi\ B_k}{\sqrt{k}} + o_p\Big(\frac{1}{\sqrt{k}}\Big) + o_p(A_\lambda(n/k))\Big)$$

with $U_\lambda(t) = \lambda + U_0(t)$, $A_\lambda(t)$ given in (2.2), *and where B_k is asymptotically standard normal. Moreover, for $i<j$, $\mathbb{C}ov(B_i, B_j) = \sqrt{i\ j}\ (1-j/n)/(j-1)$.*

Straightforward generalizations of Theorem 3.1 in Araújo Santos *et al.* (2006) and Theorem 4.1 in Henriques-Rodrigues and Gomes (2009), enable us to state the following theorem.

Theorem 3.1. *Let us assume that the second-order condition in* (2.1) *holds, with* $A_\lambda(t) = \xi\beta_\lambda t^{\rho_\lambda}$, *that* k *is an intermediate sequence of integers, i.e.* (1.4) *holds, and that* $\ln(nq)/\sqrt{k} \to 0$, *as* $n \to \infty$, *with* $q = q_n$ *given in* (1.3). *Let us further use the notation* $r_n := k/(nq)$, *assuming that* $nq = o(\sqrt{k})$ *so that* $r_n \to \infty$. *Then, for any real* s, $0 \le s < 1$, $\chi_s = F^{\leftarrow}(s)$, *finite, and the* PORT-*quantile estimator in* (1.14),

$$
(3.1)\quad \frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s)}{\mathrm{VaR}_q} - 1\right) \stackrel{d}{=} \sigma_{\mathrm{H}(p)} P_k^{\mathrm{H}(p)} + \sqrt{k}\big(b_{\mathrm{H}(p)} A_0(n/k) + c_{\mathrm{H}(p)}\ \chi_s/U_0(n/k)\big)(1+o_p(1)),
$$

with $\big(\sigma_{\mathrm{H}(p)}, b_{\mathrm{H}(p)}, c_{\mathrm{H}(p)}\big)$ *given in* (2.4), *and where* $P_k^{\mathrm{H}(p)}$ *is asymptotically standard normal.*

Proof: The PORT-quantile estimator in (1.14) can be written as

$$
\widehat{\mathrm{VaR}}_q(k;p,s) := X_{n-k:n}\left\{\left(1 - \frac{X_{n_s:n}}{X_{n-k:n}}\right) r_n^{\mathrm{H}_k(p,s)} + \frac{X_{n_s:n}}{X_{n-k:n}}\right\},
$$

with $r_n = k/(nq)$. Therefore,

$$
\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q = X_{n-k:n}\left\{\left(1 - \frac{X_{n_s:n}}{X_{n-k:n}}\right) r_n^{\mathrm{H}_k(p,s)} + \frac{X_{n_s:n}}{X_{n-k:n}} - \frac{\mathrm{VaR}_q}{X_{n-k:n}}\right\}.
$$

The use of the delta method enables us to write

$$
r_n^{\mathrm{H}_k(p,s)} \stackrel{d}{=} r_n^{\xi}\Big(1 + \ln r_n\big(\mathrm{H}_k(p,s) - \xi\big)\big(1 + o_p(1)\big)\Big).
$$

Since $\mathrm{VaR}_q = U_\lambda(1/q) = U_\lambda\,(n r_n/k)$, we can write

$$
\frac{\mathrm{VaR}_q}{X_{n-k:n}} = \frac{U_\lambda\,(n r_n/k)}{U_\lambda\,(n/k)} \times \frac{U_\lambda\,(n/k)}{X_{n-k:n}} =: A \times B.
$$

Using the results in Proposition 3.1 and the first-order Taylor series approximation for $(1+x)^{-1}$, as $x \to 0$, we get

$$
B = \frac{U_\lambda\,(n/k)}{X_{n-k:n}} \stackrel{d}{=} \left(1 + \frac{\xi B_k}{\sqrt{k}} + o\left(A_\lambda\,(n/k)\right)\right)^{-1} \stackrel{d}{=} 1 - \frac{\xi B_k}{\sqrt{k}} + o\left(A_\lambda\,(n/k)\right).
$$

The second order condition in (1.5), and the first-order Taylor series approximation for $\exp(x)$, again as $x \to 0$, enable us to get

$$
A = \frac{U_\lambda\,(n r_n/k)}{U_\lambda\,(n/k)} \stackrel{d}{=} r_n^{\xi} \exp\left(A_\lambda\,(n/k)\,\frac{r_n^{\rho_\lambda} - 1}{\rho_\lambda}\right) \stackrel{d}{=} r_n^{\xi}\left(1 - \frac{A_\lambda\,(n/k)}{\rho_\lambda} + o_p(A_\lambda\,(n/k))\right).
$$

Consequently,

$$A \times B \overset{d}{=} r_n^{\xi}\left(1 - \frac{\xi B_k}{\sqrt{k}} - \frac{A_\lambda\,(n/k)}{\rho_\lambda}(1 + o_p\,(1))\right).$$

Therefore, as $X_{n_s:n}/X_{n-k:n} = o_p(1)$ and using again the result in Proposition 3.1, we can write

$$\begin{aligned}\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q &= \mathrm{VaR}_q\left(\frac{\widehat{\mathrm{VaR}}_q}{\mathrm{VaR}_q} - 1\right)\\ &\overset{d}{=} \mathrm{VaR}_q\left(\ln r_n\big(\mathrm{H}_k(p,s) - \xi\big) + \frac{\xi B_k}{\sqrt{k}} + \frac{A_\lambda(n/k)}{\rho_\lambda}\right)(1 + o_p(1))\\ &\overset{d}{=} \ln r_n \mathrm{VaR}_q\big(H_k(p,s) - \xi\big)(1 + o_p(1)),\end{aligned}$$

and from (2.3), the result in (3.1) follows. □

Corollary 3.1. *Under the conditions of* Theorem 3.1, *with* $\mathcal{N}(\mu,\sigma^2)$ *denoting a normal* RV *with mean value* μ *and variance* σ^2, $\big(\sigma_{\mathrm{H}(p)}, b_{\mathrm{H}(p)}, c_{\mathrm{H}(p)}\big)$ *given in* (2.4), *and* $P_k^{\mathrm{H}(p)}$ *an asymptotically standard normal* RV, *the following results hold:*

- *For values of* $\xi + \rho_0 < 0$ *and* $\chi_s \neq 0$,

$$\frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q}{\mathrm{VaR}_q}\right) \overset{d}{=} \sigma_{\mathrm{H}(p)} P_k^{\mathrm{H}(p)} + \sqrt{k}\Big(c_{\mathrm{H}(p)}\frac{\chi_s}{U_0(n/k)}\Big)(1+o_p(1)).$$

If $\sqrt{k}/U_0(n/k) \to \lambda_U$ *finite, then*

$$\frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q}{\mathrm{VaR}_q}\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}(\lambda_U c_{\mathrm{H}(p)}\chi_s, \sigma^2_{\mathrm{H}(p)}).$$

- *For values of* $\xi + \rho_0 > 0$ *or* $\xi + \rho_0 \le 0$ *and* $\chi_s = 0$,

$$\frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q}{\mathrm{VaR}_q}\right) \overset{d}{=} \sigma_{\mathrm{H}(p)} P_k^{\mathrm{H}(p)} + \sqrt{k}\left(b_{\mathrm{H}(p)} A_0(n/k)\right)(1+o_p(1)).$$

If $\sqrt{k}A_0(n/k) \to \lambda_A$ *finite, then*

$$\frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q}{\mathrm{VaR}_q}\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}(\lambda_A b_{\mathrm{H}(p)}, \sigma^2_{\mathrm{H}(p)}).$$

- *For values of* $\xi + \rho_0 = 0$ *and* $\chi_s \neq 0$,

$$\begin{aligned}&\frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s) - \mathrm{VaR}_q}{\mathrm{VaR}_q}\right)\\ &\quad\overset{d}{=} \sigma_{\mathrm{H}(p)} P_k^{\mathrm{H}(p)} + \sqrt{k}\Big(b_{\mathrm{H}(p)} A_0(n/k) + c_{\mathrm{H}(p)}\frac{\chi_s}{U_0(n/k)}\Big)(1 + o_p(1)).\end{aligned}$$

If $\sqrt{k}/U_0(n/k) \to \lambda_{_U}$ *and* $\sqrt{k}A_0(n/k) \to \lambda_{_A}$*, with* $\lambda_{_U}$ *and* $\lambda_{_A}$ *both finite, then*

$$\frac{\sqrt{k}}{\ln r_n}\left(\frac{\widehat{\mathrm{VaR}}_q(k;p,s)-\mathrm{VaR}_q}{\mathrm{VaR}_q}\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}(\lambda_{_U} c_{_{\mathrm{H}(p)}}\chi_s + \lambda_{_A} b_{_{\mathrm{H}(p)}}, \sigma^2_{_{\mathrm{H}(p)}}).$$

4. A MONTE-CARLO SIMULATION STUDY

Monte-Carlo multi-sample simulation experiments, of size 5000×20, have been implemented for the classes of MO_p and PORT-MO_p VaR-estimators associated with $p = p_\ell = 2\ell/(5\xi)$, $\ell = 0, 1, 2$. Apart from the MO_p and PORT-MO_p VaR-estimators, we have further considered in the VaR-estimator in (1.7), the replacement of the estimator $\hat{\xi}(k)$ by one of the most simple classes of *corrected-Hill* (CH) EVI-estimators, the one in Caeiro *et al.* (2005). Such a class is defined as

$$\mathrm{CH}(k) \equiv \mathrm{CH}(k;\hat{\beta},\hat{\rho}) := \mathrm{H}(k)\Big(1-\hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})\Big). \tag{4.1}$$

The estimators in (4.1) can be second-order *minimum-variance reduced-bias* (MVRB) EVI-estimators, for adequate levels k and an adequate external estimation of the vector of second-order parameters, (β, ρ), introduced in (1.6), i.e. the use of $\mathrm{CH}(k)$ can enable us to eliminate the dominant component of bias of the Hill estimator, $\mathrm{H}(k)$, keeping its asymptotic variance. Indeed, from the results in Caeiro *et al.* (2005), we know that it is possible to adequately estimate the second-order parameters β and ρ, so that we get

$$\sqrt{k}\,(\mathrm{CH}(k)-\xi) \overset{d}{=} \mathcal{N}\left(0,\xi^2\right) + o_p\big(\sqrt{k}(n/k)^\rho\big),$$

i.e. $\mathrm{CH}(k)$ overpasses $\mathrm{H}(k)$ for all k. Overviews on reduced-bias estimation can be found in Chapter 6 of Reiss and Thomas, 2007, Gomes *et al.* (2008a), Beirlant *et al.* (2012) and Gomes and Guillou (2015). For the estimation of the vector of second-order parameters (β, ρ), and just as in the aforementioned review articles, we propose an algorithm of the type of the ones presented in Gomes and Pestana (2007), where the authors used the β-estimator in Gomes and Martins (2002) and the simplest ρ-estimator in Fraga Alves *et al.* (2003), both computed at a level $k_1 = \lfloor n^{0.999} \rfloor$. More recent estimators of β can be found in Caeiro and Gomes (2006), Gomes *et al.* (2010) and Henriques-Rodrigues *et al.* (2015). For alternative estimation of ρ, not later than 2014, see Gomes and Guillou (2015). See also, Caeiro and Gomes (2014, 2015b) and Henriques-Rodrigues *et al.* (2014).

It is well-known that the PORT methodology works efficiently only when the left endpoint of the underlying parent is negative, and $q = 0$ does not work when the left endpoint is infinite, like happens with the Student model (see Araújo Santos *et al.*, 2006, Gomes *et al.*, 2008b, 2011, 2016c, Caeiro *et al.*, 2016b, for further details related to the topic of PORT estimation). Consequently, only models with this characteristic have been considered, the EV_ξ, in (1.1) and the

Student-t_ν, with a probability density function

$$f(x;\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\big(1+x^2/\nu\big)^{-(\nu+1)/2}, \quad x \in \mathbb{R}.$$

The values $s = 0$ (for the EV_ξ parents), the value of s associated with the best performance of the PORT methodology for these models, and $s = 0.1$ (for the Student parents) were the ones used for illustration of the results. Sample sizes from $n = 100(100)500$ and $n = 1000(1000)5000$ were simulated from the aforementioned underlying models, for different values of ξ.

4.1. Mean values and mean square error patterns as k-functionals

For each value of n and for each of the aforementioned models, we have first simulated, on the basis of the initial 5000 runs, the mean value (E) and the *root mean square error* (RMSE) of the scale normalized VaR-estimators, i.e. the Var-estimators over VaR_q, as functions of k. For the EVI-estimation, apart from H_p, in (1.10), $p = 0$ $\big(\mathrm{H}_0 \equiv \mathrm{H}\big)$ and $p = p_\ell = 2\ell/(5\xi)$, $\ell = 1$ (for which asymptotic normality holds), and $\ell = 2$ (where only consistency was proved), and the MVRB (CH) EVI-estimators, in (4.1), we have also included their PORT versions, for the above mentioned values of s, using the notation $\bullet|s$, where $\bullet$ refers to the acronymous of the EVI-estimator.

The results are illustrated in Figure 1, for samples of size $n = 1000$ from an EV_ξ underlying parent, with $\xi = 0.1$ and $s = 0$. In this case, and for all k, there is a clear reduction in RMSE, as well as in bias, with the obtention of estimates closer to the target value ξ, particularly when we consider the PORT-version associated with H_{p_1}. Further note that, at optimal levels, in the sense of minimal RMSE, even the H_{p_2} beat the PORT-MVRB VaR-estimators.

Similar patterns were obtained for all other simulated models.

4.2. Mean values at optimal levels

Table 1 is also related to the EV_ξ model, with $\xi = 0.1$. We there present, for different sample sizes n, the simulated mean values at optimal levels (levels where RMSEs are minima as functions of k) of some of the normalized VaR-estimators, under consideration in this study. Information on 95% confidence intervals are also given. Among the estimators considered, and distinguishing 3 regions, a first one with (H, CH, H_{p_1}), a second one with the associated PORT versions, (H|0, CH|0, $\mathrm{H}_{p_1}|0$), and a third one with (H_{p_2}, $\mathrm{H}_{p_2}|0$), the one providing the smallest squared bias is written in **bold** whenever there is an out-performance of the behaviour achieved in the previous regions.

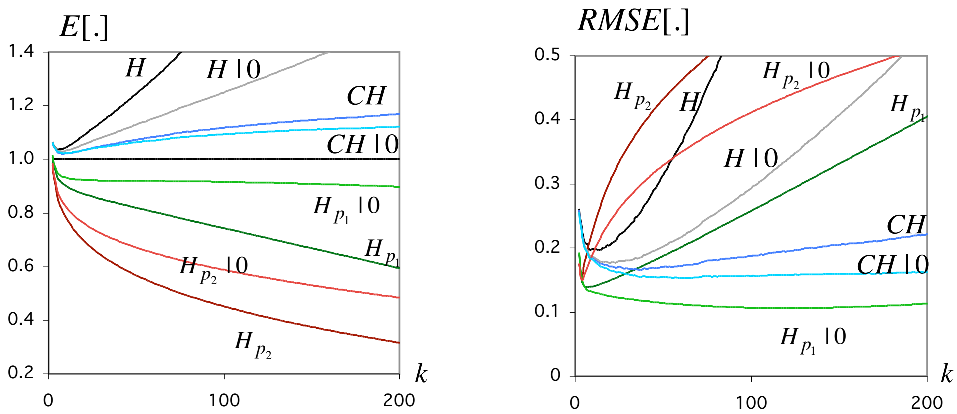


Figure 1: Mean values (*left*) and RMSEs (*right*) of the normalized H, CH, and H_p, $p = p_\ell = 2\ell/(5\xi)$, $\ell = 1, 2$ VaR-estimators for $q = 1/n$, together with their PORT versions, associated with $s = 0$ and generally denoted $\bullet|0$, for $EV_{0.1}$ underlying parents and sample size $n = 1000$

Table 1: Simulated mean values of semi-parametric normalized VaR-estimators at their optimal levels for underlying $EV_{0.1}$ parents

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
H	**1.089** $\pm$ 0.0048	1.073 $\pm$ 0.0042	1.061 $\pm$ 0.0031	1.058 $\pm$ 0.0030	1.053 $\pm$ 0.0018
CH	0.905 $\pm$ 0.0081	**0.930** $\pm$ 0.0049	**0.983** $\pm$ 0.0073	**1.056** $\pm$ 0.0035	**1.052** $\pm$ 0.0025
H_{p_1}	0.885 $\pm$ 0.0014	0.901 $\pm$ 0.0056	0.910 $\pm$ 0.0029	0.915 $\pm$ 0.0022	0.918 $\pm$ 0.0006
H|0	**1.078** $\pm$ 0.0037	1.069 $\pm$ 0.0033	1.063 $\pm$ 0.0037	1.060 $\pm$ 0.0032	1.057 $\pm$ 0.0027
CH|0	0.922 $\pm$ 0.0036	**0.945** $\pm$ 0.0038	1.025 $\pm$ 0.0006	1.116 $\pm$ 0.0005	1.060 $\pm$ 0.0021
H_{p_1}|0	0.887 $\pm$ 0.0037	0.898 $\pm$ 0.0031	0.893 $\pm$ 0.0009	0.915 $\pm$ 0.0005	**0.998** $\pm$ 0.0002
H_{p_2}	0.865 $\pm$ 0.0014	0.889 $\pm$ 0.0012	0.912 $\pm$ 0.0006	0.924 $\pm$ 0.0008	0.926 $\pm$ 0.0065
H_{p_2}|0	0.889 $\pm$ 0.0014	0.909 $\pm$ 0.0012	0.920 $\pm$ 0.0070	0.926 $\pm$ 0.0050	0.928 $\pm$ 0.0006

Tables 2, 3 and 4 are similar to Table 1, but respectively associated with $EV_{0.25}$, Student-t_4 and t_2 underlying parents.

Table 2: Simulated mean values of semi-parametric normalized VaR-estimators at their optimal levels for underlying $EV_{0.25}$ parents

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
H	1.143 $\pm$ 0.0068	1.125 $\pm$ 0.0070	1.108 $\pm$ 0.0048	1.106 $\pm$ 0.0052	1.094 $\pm$ 0.0034
CH	0.848 $\pm$ 0.0092	0.874 $\pm$ 0.0041	0.925 $\pm$ 0.0027	**1.036** $\pm$ 0.0041	1.094 $\pm$ 0.0036
H_{p_1}	**0.862** $\pm$ 0.0023	**0.912** $\pm$ 0.0014	**0.993** $\pm$ 0.0013	1.083 $\pm$ 0.0038	**1.049** $\pm$ 0.0014
H|0	1.133 $\pm$ 0.0059	1.109 $\pm$ 0.0052	1.104 $\pm$ 0.0047	1.101 $\pm$ 0.0048	1.088 $\pm$ 0.0012
CH|0	0.878 $\pm$ 0.0004	0.906 $\pm$ 0.0031	0.941 $\pm$ 0.0020	**0.965** $\pm$ 0.0018	1.063 $\pm$ 0.0004
H_{p_1}|0	**0.983** $\pm$ 0.0017	**1.060** $\pm$ 0.0022	1.048 $\pm$ 0.0021	1.055 $\pm$ 0.0022	1.064 $\pm$ 0.0017
H_{p_2}	0.854 $\pm$ 0.0046	0.848 $\pm$ 0.0014	0.868 $\pm$ 0.0043	0.869 $\pm$ 0.0035	0.881 $\pm$ 0.0024
H_{p_2}|0	0.848 $\pm$ 0.0050	0.859 $\pm$ 0.0034	0.867 $\pm$ 0.0025	0.872 $\pm$ 0.0023	0.851 $\pm$ 0.0009

Note that contrarily to what happens with the non-PORT and PORT EVI-estimation, where the values associated with p_2 are better than the ones associated with p_1, things work the other way round for the VaR-estimation.

Table 3: Simulated mean values of semi-parametric normalized VaR-estimators at their optimal levels for underlying Student t_4 parents ($\xi = 0.25$)

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
H	1.114 ± 0.0056	1.099 ± 0.0043	1.089 ± 0.0037	1.085 ± 0.0037	1.077 ± 0.0037
CH	0.903 ± 0.0292	0.903 ± 0.0053	0.922 ± 0.0030	$\mathbf{0.978} \pm 0.0028$	1.056 ± 0.0015
H_{p_1}	$\mathbf{0.932} \pm 0.0014$	$\mathbf{1.009} \pm 0.0019$	$\mathbf{1.035} \pm 0.0023$	1.032 ± 0.0019	$\mathbf{1.054} \pm 0.0017$
H $\|0.1$	1.095 ± 0.0063	1.081 ± 0.0027	1.070 ± 0.0027	1.061 ± 0.0020	1.035 ± 0.0015
CH $\|0.1$	0.890 ± 0.0030	0.950 ± 0.0031	$\mathbf{0.980} \pm 0.0020$	$\mathbf{0.990} \pm 0.0012$	$\mathbf{0.998} \pm 0.0006$
$\mathrm{H}_{p_1}\|0.1$	$\mathbf{1.056} \pm 0.0027$	1.055 ± 0.0023	1.057 ± 0.0019	1.056 ± 0.0023	1.041 ± 0.0012
$\mathrm{H}_{p_2}\|0.1$	0.876 ± 0.0011	0.904 ± 0.0008	0.953 ± 0.0005	0.982 ± 0.0005	0.998 ± 0.0002
H_{p_2}	0.875 ± 0.0062	0.882 ± 0.0029	0.886 ± 0.0022	0.889 ± 0.0020	0.877 ± 0.0005

Table 4: Simulated mean values of semi-parametric normalized VaR-estimators at their optimal levels for underlying Student t_2 parents ($\xi = 0.5$)

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
H	1.236 ± 0.0090	1.198 ± 0.0107	1.168 ± 0.0043	1.145 ± 0.0038	1.106 ± 0.0038
CH	1.115 ± 0.1919	0.809 ± 0.0072	0.825 ± 0.0053	0.848 ± 0.0030	0.848 ± 0.0043
H_{p_1}	$\mathbf{1.094} \pm 0.0073$	$\mathbf{1.082} \pm 0.0048$	$\mathbf{1.084} \pm 0.0031$	$\mathbf{1.080} \pm 0.0040$	$\mathbf{1.062} \pm 0.0021$
H $\|0.1$	1.163 ± 0.0056	1.121 ± 0.0048	1.077 ± 0.0030	$\mathbf{1.049} \pm 0.0027$	$\mathbf{1.007} \pm 0.0021$
CH $\|0.1$	0.793 ± 0.0053	0.813 ± 0.0048	0.828 ± 0.0036	0.840 ± 0.0038	0.864 ± 0.0028
$\mathrm{H}_{p_1}\|0.1$	1.098 ± 0.0058	1.087 ± 0.0034	$\mathbf{1.072} \pm 0.0033$	1.051 ± 0.0023	1.010 ± 0.0017
$\mathrm{H}_{p_2}\|0.1$	0.836 ± 0.0014	0.868 ± 0.0013	0.915 ± 0.0008	0.949 ± 0.0007	1.065 ± 0.0004
H_{p_2}	0.803 ± 0.0036	0.796 ± 0.0026	0.795 ± 0.0012	0.813 ± 0.0008	0.873 ± 0.0005

4.3. RMSEs and relative efficiency indicators at optimal levels

We have further computed the Weissman-Hill VaR-estimator, i.e. the VaR-estimator $Q_{k,q,\hat{\xi}}$, in (1.7), with $\hat{\xi}$ replaced by the H EVI-estimator, in (1.8), at the simulated optimal k in the sense of minimum RMSE. Such an estimator is denoted by Q_{00}. For any of the VaR-estimators under study, generally denoted $\mathrm{Q}(k)$, we have also computed Q_0, the estimator $\mathrm{Q}(k)$ computed at the simulated value of $k_{0|Q} := \arg\min_k \mathrm{RMSE}\big(\mathrm{Q}(k)\big)$. The simulated indicators are

$$\mathrm{REFF}_{\mathrm{Q}|0} := \frac{\mathrm{RMSE}\,(\mathrm{Q}_{00})}{\mathrm{RMSE}\,(\mathrm{Q}_0)}. \tag{4.2}$$

Remark 4.1. Note that, as usual, an indicator higher than one means that the estimator has a better performance than the Weissman-Hill VaR-estimator. Consequently, the higher the indicators in (4.2) are, the better the associated VaR-estimators perform, comparatively to Q_{00}.

Again as an illustration of the obtained results, we present Tables 5–8. In the first row, we provide RMSE_0, the RMSE of Q_{00}, so that we can easily recover the RMSE of all other estimators. The following rows provide the REFF-indicators for the different VaR-estimators under study. A similar mark (**bold**) is

Table 5: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying $\mathrm{EV}_{0.1}$ parents

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
RMSE_0	0.329 ± 0.1224	0.273 ± 0.1209	0.225 ± 0.1059	0.200 ± 0.0754	0.157 ± 0.0324
CH	1.287 ± 0.0154	1.323 ± 0.0147	1.252 ± 0.0123	1.202 ± 0.0083	1.073 ± 0.0041
H_{p_1}	$\mathbf{1.566} \pm 0.0174$	$\mathbf{1.505} \pm 0.0129$	$\mathbf{1.460} \pm 0.0103$	$\mathbf{1.440} \pm 0.0093$	$\mathbf{1.545} \pm 0.0113$
H|0	1.132 ± 0.0093	1.121 ± 0.0060	1.118 ± 0.0049	1.122 ± 0.0049	1.136 ± 0.0057
CH|0	1.659 ± 0.0196	$\mathbf{1.833} \pm 0.0179$	1.548 ± 0.0202	1.373 ± 0.0110	1.202 ± 0.0077
$\mathrm{H}_{p_1}\|0$	$\mathbf{1.695} \pm 0.0190$	1.626 ± 0.0149	$\mathbf{1.614} \pm 0.0128$	$\mathbf{1.874} \pm 0.0160$	$\mathbf{4.988} \pm 0.0340$
H_{p_2}	1.450 ± 0.0177	1.379 ± 0.0117	1.316 ± 0.0084	1.279 ± 0.0086	1.189 ± 0.0063
$\mathrm{H}_{p_2}\|0$	1.529 ± 0.0184	1.440 ± 0.0113	1.359 ± 0.0082	1.323 ± 0.0097	1.240 ± 0.0066

Table 6: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying $\mathrm{EV}_{0.25}$ parents

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
RMSE_0	0.469 ± 0.1207	0.394 ± 0.1350	0.329 ± 0.1453	0.294 ± 0.1498	0.231 ± 0.1538
CH	1.393 ± 0.0144	1.431 ± 0.0155	1.681 ± 0.0215	1.908 ± 0.0257	1.197 ± 0.0045
H_{p_1}	$\mathbf{2.132} \pm 0.0218$	$\mathbf{2.522} \pm 0.0233$	$\mathbf{3.802} \pm 0.0333$	$\mathbf{3.866} \pm 0.0248$	$\mathbf{3.108} \pm 0.0229$
H|0	1.178 ± 0.0081	1.174 ± 0.0101	1.185 ± 0.0053	1.206 ± 0.0060	1.245 ± 0.0043
CH|0	1.837 ± 0.0164	1.907 ± 0.0206	2.215 ± 0.0222	2.678 ± 0.0251	2.681 ± 0.0180
$\mathrm{H}_{p_1}\|0$	$\mathbf{3.527} \pm 0.0300$	$\mathbf{2.754} \pm 0.0221$	1.703 ± 0.0135	1.584 ± 0.0128	1.443 ± 0.0102
H_{p_2}	1.771 ± 0.0148	1.658 ± 0.0154	1.540 ± 0.0118	1.464 ± 0.0118	1.283 ± 0.095
$\mathrm{H}_{p_2}\|0$	1.896 ± 0.0176	1.757 ± 0.0162	1.614 ± 0.0144	1.526 ± 0.0131	1.338 ± 0.0110

Table 7: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying Student t_4 parents ($\xi = 0.25$)

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
RMSE_0	0.378 ± 0.1445	0.320 ± 0.1507	0.270 ± 0.1556	0.240 ± 0.1572	0.185 ± 0.1554
CH	1.217 ± 0.1176	1.310 ± 0.0129	1.480 ± 0.0134	$\mathbf{1.881} \pm 0.0114$	$\mathbf{1.531} \pm 0.0095$
H_{p_1}	$\mathbf{2.143} \pm 0.0187$	$\mathbf{2.483} \pm 0.0209$	$\mathbf{1.821} \pm 0.0148$	1.422 ± 0.0100	1.151 ± 0.0088
H |0.1	1.243 ± 0.0105	1.273 ± 0.0081	1.359 ± 0.0066	1.457 ± 0.0064	1.808 ± 0.0069
CH |0.1	1.773 ± 0.0160	2.038 ± 0.0181	$\mathbf{2.599} \pm 0.0252$	$\mathbf{3.082} \pm 0.0198$	$\mathbf{4.431} \pm 0.0269$
$\mathrm{H}_{p_1}\|0.1$	1.640 ± 0.0119	1.516 ± 0.0138	1.463 ± 0.0161	1.477 ± 0.0206	1.664 ± 0.0240
H_{p_2}	1.713 ± 0.0167	1.631 ± 0.0152	1.518 ± 0.0122	1.427 ± 0.0066	1.270 ± 0.0069
$\mathrm{H}_{p_2}\|0.1$	2.080 ± 0.0205	2.288 ± 0.0238	$\mathbf{3.045} \pm 0.0248$	$\mathbf{4.026} \pm 0.0291$	$\mathbf{6.345} \pm 0.0502$

Table 8: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying Student t_2 parents ($\xi = 0.5$)

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
RMSE_0	0.675 ± 0.1735	0.559 ± 0.1793	0.449 ± 0.1804	0.379 ± 0.1789	0.255 ± 0.1684
CH	0.684 ± 0.3593	1.359 ± 0.0145	$\mathbf{1.388} \pm 0.0099$	$\mathbf{1.371} \pm 0.0080$	$\mathbf{1.449} \pm 0.0343$
H_{p_1}	$\mathbf{1.728} \pm 0.0148$	$\mathbf{1.468} \pm 0.0116$	1.308 ± 0.0085	1.240 ± 0.0047	1.209 ± 0.0510
H |0.1	1.318 ± 0.0097	1.382 ± 0.0099	1.532 ± 0.0117	$\mathbf{1.667} \pm 0.0079$	$\mathbf{2.110} \pm 0.0132$
CH |0.1	$\mathbf{1.969} \pm 0.0160$	$\mathbf{1.786} \pm 0.0150$	$\mathbf{1.573} \pm 0.0103$	1.419 ± 0.0091	1.123 ± 0.0065
$\mathrm{H}_{p_1}\|0.1$	1.609 ± 0.0149	1.516 ± 0.0116	1.560 ± 0.0109	1.647 ± 0.0072	2.037 ± 0.0108
H_{p_2}	2.271 ± 0.0214	1.992 ± 0.0179	1.766 ± 0.0126	1.665 ± 0.0151	1.691 ± 0.0124
$\mathrm{H}_{p_2}\|0.1$	$\mathbf{2.810} \pm 0.0276$	$\mathbf{2.821} \pm 0.0287$	$\mathbf{3.116} \pm 0.0227$	$\mathbf{3.547} \pm 0.0280$	$\mathbf{2.848} \pm 0.0169$

used for the highest REFF indicator, again considering the aforementioned three regions and $q = 1/n$.

For a better visualization of the results presented in some of the tables above, we further present Figure 2, associated with an $EV_{0.1}$ underlying parent.

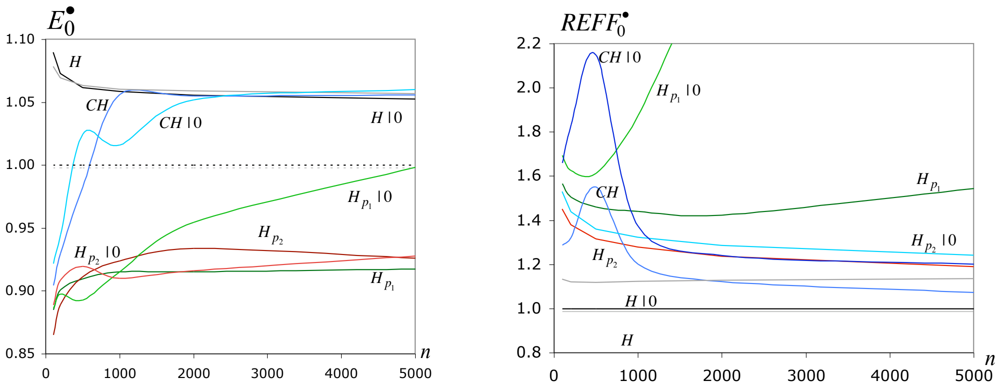


Figure 2: Mean values (*left*) and REFF-indicators (*right*) at optimal levels of the different normalized VaR-estimators under study, for $q = 1/n$, an underlying $EV_{0.1}$ parent and sample sizes $n = 100(100)500$ and $500(500)5000$

5. CONCLUSIONS

The new PORT-MO_p VaR-estimators, defined in (1.14), generalize the Weissman-Hill PORT-quantile estimator studied in Araújo Santos *et al.* (2006). Consequently, both asymptotically and for finite sample sizes, we were expecting a much better behaviour of the new PORT-MO_p VaR-estimator. The gain in efficiency of the PORT-MO_p VaR-estimators is, in most cases, greater than the one obtained with the MVRB and PORT-MVRB VaR-estimators. The simulated mean values of the normalized PORT-MO_p VaR-estimators are always better, for moderate to large values of n, in the Student-t_ν parents. For the EV_ξ-parents, we have different behaviours accordingly to the size of the sample but there is a general out-performance of the PORT-MO_p VaR-estimators. And indeed, for an adequate choice of k, p and s, the PORT-MO_p VaR–estimators are able to outperform the MVRB and even the PORT-MVRB VaR-estimators, in most cases. The choice of (k, p, s) can be done through heuristic sample-path stability algorithms, like the ones in Gomes *et al.* (2013) or through a bootstrap algorithm of the type of the ones presented in Caeiro and Gomes (2015a) and in Gomes *et al.* (2016b), where R-scripts are provided.

ACKNOWLEDGMENTS

Research partially supported by COST Action IC1408, and National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, project UID/MAT/00006/2013 (CEA/UL), and by COST Action IC1408—CroNos. We also acknowledge the valuable suggestions from a referee.

REFERENCES

[1] Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation. *Revstat* **4**:3, 227–247.

[2] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, England.

[3] Beirlant, J., Figueiredo, F., Gomes, M.I. and Vandewalle, B. (2008). Improved reduced-bias tail index and quantile estimators. *J. Statistical Planning and Inference* **138**:6, 1851–1870.

[4] Beirlant, J., Caeiro, F. and Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* **10**:1, 1–31.

[5] Beran, J., Schell, D. and Stehlík, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Statist. Math.* **66**, 193–220.

[6] Brilhante, M.F., Gomes, M.I. and Pestana, D. (2013). A simple generalization of the Hill estimator. *Computational Statistics and Data Analysis* **57**:1, 518–535.

[7] Caeiro, F. and Gomes, M.I. (2006). A new class of estimators of a scale second order parameter. *Extremes* **9**, 193–211.

[8] Caeiro, F. and Gomes, M.I. (2008). Semi-parametric second-order reduced-bias high quantile estimation. *Test* **18**:2, 392–413.

[9] Caeiro, F. and Gomes, M.I. (2014). A semi-parametric estimator of a shape second order parameter. In A. Pacheco *et al.* (Eds.), *New Advances in Statistical Modeling and Application*, Springer-Verlag, Berlin, Heidelberg, 163–175.

[10] Caeiro, F. and Gomes, M.I. (2015a). Threshold selection in extreme value analysis. In Dipak Dey and Jun Yan, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman-Hall/CRC, Chapter 4, 69–87.

[11] Caeiro, F. and Gomes, M.I. (2015b). Bias reduction in the estimation of a shape second-order parameter of a heavy tailed model. *J. Statistical Computation and Simulation* **85**:17, 3405–3419.

[12] Caeiro, F., Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3**:2, 113-136.

[13] Caeiro, F., Gomes, M.I., Beirlant, J. and de Wet, T. (2016a). Mean-of-order p reduced-bias extreme value index estimation under a third-order framework. *Extremes* **19**:4, 561–589.

[14] Caeiro, F., Gomes, M.I. and Henriques-Rodrigues, L. (2016b). A location invariant probability weighted moment EVI estimator. *International J. Computer Mathematics* **93**:4, 676–695.

[15] Dietrich, D., de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions. *Extremes* **5**, 71–85.

[16] Ferreira, A., Haan, L. de and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics* **37**:5, 401–434.

[17] Fraga Alves, M.I. and Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction—a comparative analysis. *Comm. Stat. Theory Methods* **25**:4, 789–811.

[18] Fraga Alves, M.I., Gomes, M.I. and Haan, L. de (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica* **60**:2, 193–213.

[19] Fraga Alves, M.I., Gomes, M.I., de Haan, L. and Neves, C. (2007). A note on second order conditions in extreme value theory: linking general and heavy tails conditions. *Revstat* **5**:3, 285–305.

[20] Geluk, J. and Haan, L. de (1980). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

[21] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**:6, 423–453.

[22] Gomes, M.I. and Caeiro, F. (2014). Efficiency of partially reduced-bias mean-of-order-p versus minimum-variance reduced-bias extreme value index estimation. In M. Gilli *et al.* (eds.). *Proceedings of COMPSTAT* 2014, The International Statistical Institute/International Association for Statistical Computing, 289–298.

[23] Gomes, M.I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A Review. *International Statistical Review* **83**:2, 263–292.

[24] Gomes, M.I. and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the extreme value index based on external estimation of the second order parameter. *Extremes* **5**:1, 5–31.

[25] Gomes, M.I. and Pestana, D. (2007). A sturdy reduced bias extreme quantile (VaR) estimator. *J. American Statistical Association* **102**:477, 280–292.

[26] Gomes, M.I., Canto e Castro, L., Fraga Alves, M.I. and Pestana, D. (2008a). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. *Extremes* **11**:1, 3–34.

[27] Gomes, M.I., Fraga Alves, M.I. and Araújo Santos, P. (2008b). PORT Hill and moment estimators for heavy-tailed models. *Communications in Statistics—Simulation and Computation* **37**, 1281–1306.

[28] Gomes, M.I., Henriques-Rodrigues, L., Pereira, H. and Pestana, D. (2010). Tail index and second order parameters' semi-parametric estimation based on the log-excesses. *J. Statistical Computation and Simulation* **80**:6, 653–666.

[29] Gomes, M.I., Henriques-Rodrigues, L. and Miranda, C. (2011). Reduced-bias location-invariant extreme value index estimation: a simulation study. *Communications in Statistics—Simulation and Computation* **40**:3, 424–447.

[30] Gomes, M.I., Henriques-Rodrigues, L., Fraga Alves, M.I. and Manjunath, B.G. (2013). Adaptive PORT-MVRB estimation: an empirical comparison of two heuristic algorithms. *J. Statististical Computation and Simulation* **83**:6, 1129–1144.

[31] Gomes, M.I., Brilhante, M.F., Caeiro, F. and Pestana, D. (2015a). A new partially reduced-bias mean-of-order-p class of extreme value index estimators. *Computational Statistics and Data Analysis* **82**, 223–237.

[32] Gomes, M.I., Brilhante, F. and Pestana, D. (2015b). A mean-of-order-p class of value-at-risk estimators. In Kitsos, C., Oliveira, T., Rigas, A. and Gulati, S. (eds.), *Theory and Practice of Risk Assessment*, Springer Proceedings in Mathematics and Statistics 136, Springer International Publishing, Switzerland, 305–320.

[33] Gomes, M.I., Brilhante, M.F. and Pestana, D. (2016a). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics—Simulation and Computation* **45**, 1–30.

[34] Gomes, M.I., Caeiro, F., Henriques-Rodrigues, L. and Manjunath, B.G. (2016b). Bootstrap methods in statistics of extremes. In Longin, F. (ed.), *Handbook of Extreme Value Theory and Its Applications to Finance and Insurance*. Handbook Series in Financial Engineering and Econometrics (Ruey Tsay Adv.Ed.). John Wiley and Sons, Chapter 6, 117–138.

[35] Gomes, M.I., Henriques-Rodrigues, L. and Manjunath, B.G. (2016c). Mean-of-order-p location-invariant extreme value index estimation. *Revstat* **14**: 3, 273–296.

[36] Haan, L. de (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract 32, Amsterdam.

[37] Haan, L. de (1984). Slow variation and characterization of domains of attraction. In Tiago de Oliveira, ed., *Statistical Extremes and Applications*, 31–48, D. Reidel, Dordrecht, Holland.

[38] Haan, L. de and Ferreira, A. (2006). *Extreme Value Theory: an Introduction.* Springer.

[39] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statistical Society B* **44**:1, 37–42.

[40] Hall, P. and Welsh, A.H. (1985). Adaptative estimates of parameters of regular variation. *Annals of Statistics* **13**, 331–341.

[41] Henriques-Rodrigues, L. and Gomes, M.I. (2009). High quantile estimation and the PORT methodology. *Revstat* **7**:3, 245–264.

[42] Henriques-Rodrigues, L., Gomes, M.I., Fraga Alves, M.I. and Neves, C. (2014). PORT estimation of a shape second-order parameter. *Revstat* **13**:3, 299–328.

[43] Henriques-Rodrigues, L., Gomes, M.I. and Manjunath, B.G. (2015). Estimation of a scale second-order parameter related to the PORT methodology. *J. Statistical Theory and Practice* **9**:3, 571–599.

[44] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**:5, 1163–1174.

[45] Paulauskas, V. and Vaičiulis, M. (2013). On the improvement of Hill and some others estimators. *Lithuanian Mathematical J.* **53**, 336–355.

[46] Paulauskas, V. and Vaičiulis, M. (2015). A class of new tail index estimators. *Annals Institute of Statistical Mathematics*, DOI: 10.1007/s10463-015-0548-3.

[47] Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*, 3rd edition, Birkhäuser Verlag.

[48] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. American Statistical Association* **73**, 812–815.

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 205-222

WEIGHTED-TYPE WISHART DISTRIBUTIONS WITH APPLICATION

Authors: Mohammad Arashi
– Department of Statistics, School of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran. Department of Statistics, Faculty of Natural and Agricultural Sciences,University of Pretoria, Pretoria, South Africa. (m_arashi_stat@yahoo.com)

Andriëtte Bekker
– Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa. (andriette.bekker@up.ac.za)

Janet van Niekerk
– Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa. (janet.vanniekerk@up.ac.za)

Received: March 2016 Revised: October 2016 Accepted: November 2016

Abstract:

- In this paper, we consider a general framework for constructing new valid densities regarding a random matrix variate. However, we focus specifically on the Wishart distribution. The methodology involves coupling the density function of the Wishart distribution with a Borel measurable function as a weight. We propose three different weights by considering trace and determinant operators on matrices. The characteristics for the proposed weighted-type Wishart distributions are studied and the enrichment of this approach is illustrated. A special case of this weighted-type distribution is applied in the Bayesian analysis of the normal model in the univariate and multivariate cases. It is shown that the performance of this new prior model is competitive using various measures.

Key-Words:

- *Bayesian analysis; eigenvalues; Kummer gamma; Kummer Wishart; matrix variate; weight function; Wishart distribution.*

AMS Subject Classification:

- 62E15, 60E05, 62H12, 62F15

1. INTRODUCTION

The modeling of real world phenomena is constantly increasing in complexity and standard statistical distributions cannot model these adequately. The question arises whether we can introduce new models to compete with and enhance the standard approaches available in the literature. Various generalizations and extensions have been proposed for standard statistical models, since more complex models are needed to solve the modeling complications of real data. To mention a few: Sutradhar et al. (1989) generalized the Wishart distribution for the multivariate elliptical models, however Teng et al. (1989) considered matrix variate elliptical models in their study. Wong and Wang (1995) defined the Laplace-Wishart distribution, while Letac and Massam (2001) defined the normal quasi-Wishart distribution. In the context of graphical models, Roverato (2002) defined the hyper-inverse Wishart and Wang and West (2009) extended the inverse Wishart distribution for using hyper-Markov properties (see Dawid and Lauritzen (1993)), while Bryc (2008) proposed the compound Wishart and q-Wishart in graphical models. Abul-Magd et al. (2009) proposed a generalization to Wishart-Laguerre ensembles. Adhikari (2010) generalized the Wishart distribution for probabilistic structural dynamics, and Díaz-García et al. (2011) extended the Wishart distribution for real normed division algebras. Munilla and Cantet (2012) also formulated a special structure for the Wishart distribution to apply in modeling the maternal animal. These generalizations justify the speculative research to propose new models based on the concept of weighted distributions Rao (1965). Assuming special cases of these new models as priors for an underlying normal model in a Bayesian analysis exhibit interesting behaviour.

In this paper we propose a weighted-type Wishart distribution, making use of the mathematical mechanism frequently used in proposing weighted-type distributions, from length-biased viewpoint, and consider its applications in Bayesian analysis. The building block of our contribution is an extension of the mathematical formulation of univariate weighted-type distributions to multivariate weighted-type distributions. Specifically, if $f(x;\sigma^2)$ is the main/natural probability density function (pdf) which is imposed by a scalar weight function $h(x;\phi)$ (not necessarily positive), then the weighted-type distribution is given by

$$g(x;\boldsymbol{\theta}) = Ch(x;\phi)f(x;\sigma^2), \quad \boldsymbol{\theta} = (\sigma^2,\phi), \tag{1.1}$$

where $C^{-1} = E_{\sigma^2}[h(X;\phi)]$ and the expectation $E_{\sigma^2}[.]$ is taken over the same probability measure as $f(.)$. The parameter ϕ can be seen as an *enriching parameter.*

For the multivariate case, one can simply use the pdf of a multivariate random variable for $f(.)$ in (1.1). Further, the parameter space can be multidimensional. However, the weight function $h(.)$ should remain of scalar form. Thus the question that arises is: Why not replace $f(.)$ in (1.1) with the pdf of a matrix variate random variable? To address this issue and using (1.1) as departure, we define matrix variate weighted-type distributions, from where new matrix variate distributions originate.

Initially let $\mathcal{S}_m$ be the space of all positive definite matrices of dimension m. To set the platform for what we are proposing, consider a random matrix variate $\boldsymbol{X} \in \mathcal{S}_m$ having a pdf $f(.;\boldsymbol{\Psi})$ with parameter $\boldsymbol{\Psi}$. We construct matrix variate distributions, with pdf $g(.;\boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = (\boldsymbol{\Psi}, \boldsymbol{\Phi})$ and enrichment parameter $\boldsymbol{\Phi} \in \mathcal{S}_m$, by utilizing one of the following mechanisms:

1. (Loading with a weight of trace form)

$$g(\boldsymbol{X};\boldsymbol{\Theta}) = C_1 h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])f(\boldsymbol{X};\boldsymbol{\Psi}), \boldsymbol{\Theta} = (\boldsymbol{\Psi}, \boldsymbol{\Phi}). \tag{1.2}$$

2. (Loading with a weight of determinant form)

$$g(\boldsymbol{X};\boldsymbol{\Theta}) = C_2 h_2(|\boldsymbol{X}\boldsymbol{\Phi}|)f(\boldsymbol{X};\boldsymbol{\Psi}), \boldsymbol{\Theta} = (\boldsymbol{\Psi}, \boldsymbol{\Phi}). \tag{1.3}$$

3. (Loading with a mixture of weights of trace and determinant forms)

$$g(\boldsymbol{X};\boldsymbol{\Theta}) = C_3 h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|)f(\boldsymbol{X};\boldsymbol{\Psi}), \boldsymbol{\Theta} = (\boldsymbol{\Psi}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2), \tag{1.4}$$

where $h_i(.), i = 1, 2$ is a Borel measurable function (weight function) which admits Taylor's series expansion, C_j is a normalizing constant and $f(.)$ can be referred to as a generator.
In this paper, we consider the $f(.)$ in (1.2)-(1.4) to be the pdf of the Wishart distribution with parameters n and $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\Psi} = (n, \boldsymbol{\Sigma})$, given by

$$\frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right), \tag{1.5}$$

with $\boldsymbol{X}, \boldsymbol{\Sigma} \in \mathcal{S}_m$, denoted by $W_m(n, \boldsymbol{\Sigma})$, and incorporate a weight function, $h_i(.)$, as given by (1.2)-(1.4). Note that $\Gamma_m(.)$ is the multivariate gamma function and $etr(.) = exp(tr(.))$.

We organize our paper as follows: In Section 2, we discuss the weighted-type Wishart distribution that originated from (1.2) and propose some of its important properties. The enrichment of this approach is illustrated by the graphical display of the joint density function of the eigenvalues of the random matrix for certain cases. In Section 3, the weighted-type Wishart distributions emanating from (1.3) and (1.4) are proposed. The significance of this approach of extending the well-known Wishart distribution, will be demonstrated in Section 4, by assuming special cases as a priors for the underlying univariate and multivariate normal model. Comparison results of these cases with well-known priors path the way for integrating these models in Bayesian analysis. Finally, some thoughts of other possible applications are given in Section 5.

2. WEIGHTED-TYPE I WISHART DISTRIBUTION

In this section we consider the construction methodology of a weighted-type I Wishart distribution according to (1.2).

Definition 2.1. The random matrix $\mathbf{X} \in \mathcal{S}_m$ is said to have a weighted-type I Wishart distribution (W1WD) with parameters $\boldsymbol{\Psi}$, $\boldsymbol{\Phi} \in \mathcal{S}_m$ and the weight function $h_1(.)$, if it has the following pdf

$$\begin{aligned} g(\mathbf{X};\boldsymbol{\Theta}) &= \frac{h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])f(\boldsymbol{X};\boldsymbol{\Psi})}{E\left[h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])\right]} \\ &= c_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}]), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi},\boldsymbol{\Phi}), \end{aligned} \tag{2.1}$$

with

$$\{c_{n,m}(\boldsymbol{\Theta})\}^{-1} = 2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\sum_{k=0}^{\infty}\frac{2^k h_1^{(k)}(0)}{k!}\sum_{\kappa}\left(\frac{n}{2}\right)_{\kappa} C_{\kappa}\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}\right), \tag{2.2}$$

written as $\boldsymbol{X} \sim \mathbb{W}_m^I(n,\boldsymbol{\Sigma},\boldsymbol{\Phi})$. In (2.1) $f(\boldsymbol{X};\boldsymbol{\Psi})$ is the pdf of the Wishart distribution ($W_m(n,\boldsymbol{\Sigma})$) (see 1.5) i.e. $\boldsymbol{\Psi} = (\boldsymbol{\Sigma}, n)$, $n > m-1$, $\boldsymbol{\Sigma} \in \mathcal{S}_m$ and $h_1(.)$ is a Borel measurable function that admits Taylor's series expansion, $(a)_\kappa = \frac{\Gamma_m(a,\kappa)}{\Gamma_m(a)}$ and $\Gamma_m(a,\kappa)$ is the generalized gamma function. The parameters are restricted to take those values for which the pdf is non-negative.

Remark 2.1. Note that using Taylor's series expansion for $h_1(.)$ in (2.1) it follows that

$$h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}]) = \sum_{k=0}^{\infty}\frac{h_1^{(k)}(0)}{k!}\operatorname{tr}(\boldsymbol{X}\boldsymbol{\Phi})^k = \sum_{k=0}^{\infty}\frac{h_1^{(k)}(0)}{k!}C_\kappa(\boldsymbol{X}\boldsymbol{\Phi}), \tag{2.3}$$

from Definition 7.21, p.228 of Muirhead (2005) where $h_1^{(k)}(0)$ is the k-th derivative of $h_1(.)$ at the point zero. Therefore using Theorem 7.2.7, p.248 of Muirhead (2005) follows from Definition 2.1 that

$$\begin{aligned} E\left[h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])\right] &= \int_{\mathcal{S}_m} h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}])f(\boldsymbol{X};\boldsymbol{\Psi})d\boldsymbol{X} \\ &= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)}\sum_{k=0}^{\infty}\frac{h_1^{(k)}(0)}{k!}\sum_{\kappa}\int_{\mathcal{S}_m}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}}\operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) \\ &\qquad C_\kappa(\boldsymbol{X}\boldsymbol{\Phi})d\boldsymbol{X} \\ &= \frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}}{2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)}\sum_{k=0}^{\infty}\frac{2^{\frac{nm}{2}+k}\Gamma_m\left(\frac{n}{2}\right)|\boldsymbol{\Sigma}|^{\frac{n}{2}}h_1^{(k)}(0)}{k!}\sum_{\kappa}\left(\frac{n}{2}\right)_\kappa C_\kappa\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}\right) \\ &= \sum_{k=0}^{\infty}\frac{2^k h_1^{(k)}(0)}{k!}\sum_{\kappa}\left(\frac{n}{2}\right)_\kappa C_\kappa\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}\right), \end{aligned}$$

and (2.2) follows ($C_\kappa\left(a\boldsymbol{X}\right) = a^k C_\kappa\left(\boldsymbol{X}\right)$ and $C_\kappa(.)$ is the zonal polynomial corresponding to κ (Muirhead (2005)).

Remark 2.2. Here we consider some thoughts related to Definition 2.1 and (2.1).

(1) As formerly noticed, the weight function should be a scalar function. In Definition 2.1, we used the trace operator, however any relevant operator can be used. The determinant operator will be discussed in Section 3. Another interesting operator can be the eigenvalue. In this respect one may use the result of Arashi (2013) to get closed expression for the expected value of the weight function.

(2) For $h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}]) = etr(\boldsymbol{X}\boldsymbol{\Phi})$ in (2.1) we obtain an enriched Wishart distribution with scale matrix $\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Phi}$.

(3) For $h_1(x\phi) = exp(x\phi)$ and $m = 1$ in (2.1) the pdf simplifies to

$$g(x;\theta) = c_n(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}-1} \exp\left(-\left(\frac{1}{2\sigma^2} - \phi\right) x\right), \tag{2.4}$$

which is the pdf of a gamma random variable with parameters $\frac{n}{2}$ and $\frac{1}{2\sigma^2} - \phi$, with $c_n(\theta) = \frac{\left(\frac{1}{2\sigma^2}-\phi\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}$, $\theta = (\sigma^2, \phi)$, written as $G(\alpha = \frac{n}{2}, \beta = \frac{1}{2\sigma^2} - \phi)$.

(4) For $h_1(x) = x$ and $m = 1$ in (2.1) the pdf simplifies to

$$g(x;\theta) = c_n(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}-1} \exp\left(-\frac{1}{2\sigma^2}x\right) x = c_n(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}x\right),$$

with $c_n(\theta) = \frac{\left(\frac{1}{2\sigma^2}-\phi\right)^{\frac{n+2}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}$, $\theta = (\sigma^2, \phi)$, hence $X \sim G(\alpha = \frac{n}{2}+1, \beta = \frac{1}{2\sigma^2} - \phi)$. This is also called the length-biased or size-biased gamma distribution (see Patil and Ord (1976)) with parameters $\frac{n}{2}$ and $\frac{1}{2\sigma^2} - \phi$.

(5) For $h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}]) = (1 + tr[\boldsymbol{X}\boldsymbol{\Phi}])$ in (2.1) the pdf simplifies to

$$g(\boldsymbol{X};\boldsymbol{\Theta}) = c_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)(1 + tr(\boldsymbol{X}\boldsymbol{\Phi})), \tag{2.5}$$

with $c_{n,m}(\boldsymbol{\Theta})$ as in (2.2), $\boldsymbol{\Theta} = (n, \boldsymbol{\Sigma}, \boldsymbol{\Phi})$, which is defined as the Kummer Wishart distribution and denoted as $KW_m(n, \boldsymbol{\Sigma}, \boldsymbol{\Phi})$.

(6) For $h_1(x\phi) = (1 + x\phi)^\gamma$, where γ is a known fixed constant, and $m = 1$ in (2.1) the pdf simplifies to

$$g(x;\theta) = c_n(\theta)(\sigma^2)^{-\frac{n}{2}} x^{\frac{n}{2}-1} \exp\left(-\frac{1}{2\sigma^2}x\right)(1 + \phi x)^\gamma \tag{2.6}$$

with $c_n(\theta) = 2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\sum_{k=0}^{\infty}\frac{(2\phi^2\sigma^2)^k\gamma!}{(\gamma-k)!k!}\sum_\kappa\left(\frac{n}{2}\right)_\kappa$, $\theta = (\sigma^2, \phi)$. If $\phi = 1$ then this is also known as the Kummer gamma or generalized gamma distribution, written as $KG(\alpha = \frac{n}{2}, \beta = \frac{1}{2\sigma^2}, \gamma)$, by expanding the term $(1 + \phi x)^\gamma$ (see Pauw et al. (2010)).

(7) Various functional forms of $h_1(.)$ are explored and the joint density of eigenvalues of the matrix variates are graphically illustrated to show the flexibility built in by this construction, see Table 1.

2.1. Characteristics

In this section some statistical properties of the W1WD (Definition 2.1) are derived. Most of the computations here deal with the relevant use of (2.3) and Theorem 7.2.7, p.248 of Muirhead (2005), though we do not mention every time.

Theorem 2.1. *Let* $\mathbf{X} \sim \mathbb{W}_m^I(n, \mathbf{\Sigma}, \mathbf{\Phi})$, *then the* r^{th} *moment of* $|\mathbf{X}|$ *is given by*

$$E\left(|\mathbf{X}|^r\right) = \frac{c_{n,m}(\mathbf{\Theta})}{c_{2(\frac{n}{2}+r),m}(\mathbf{\Theta})}|\mathbf{\Sigma}|^r,$$

where $c_{n,m}(\mathbf{\Theta})$ *and* $c_{2(\frac{n}{2}+r),m}(\mathbf{\Theta})$ *as in* (2.2).

Proof: Similarly as in Remark 2.1, by using (2.1),

$$\begin{aligned} E\left(|\mathbf{X}|^r\right) &= c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \int_{\mathcal{S}_m} |\mathbf{X}|^{r+\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Sigma^{-1}X}\right) h_1(tr[\boldsymbol{X\Phi}])d\mathbf{X} \\ &= c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \sum_{k=0}^{\infty} \frac{h_1^{(k)}(0)}{k!} \sum_{\kappa} \int_{\mathcal{S}_m} |\boldsymbol{X}|^{r+\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\boldsymbol{X}\right) \\ &\quad \times C_\kappa(\boldsymbol{X\Phi})d\boldsymbol{X}, \end{aligned}$$

the result follows. □

In the following, we give the exact expression for the moment generating function (MGF) of the W1WD, provided its existence.

Theorem 2.2. *Let* $\mathbf{X} \sim \mathbb{W}_m^I(n, \mathbf{\Sigma}, \mathbf{\Phi})$, *then the moment generating function of* $\mathbf{X}$ *is given by*

$$M_{\mathbf{X}}(\mathbf{T}) = c_{n,m}(\mathbf{\Theta}) d_{n,m} |\boldsymbol{I}_m - 2\mathbf{\Sigma}\boldsymbol{T}|^{-\frac{n}{2}},$$

with $c_{n,m}(\mathbf{\Theta})$ *as in (2.2) and* $d_{n,m} = 2^{\frac{nm}{2}}\Gamma_m\left(\frac{n}{2}\right)\sum_{k=0}^{\infty}\sum_{\kappa}\frac{2^k h_1^{(k)}(0)}{k!}\left(\frac{n}{2}\right)_\kappa$ $\times\, C_\kappa\left(\mathbf{\Phi}(\mathbf{\Sigma}^{-1} - 2\boldsymbol{T})^{-1}\right)$.

Proof: Using equation (2.1) we have

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{T}) &= E(\operatorname{etr}(\boldsymbol{TX})) \\ &= c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \int_{\mathcal{S}_m} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\boldsymbol{X} + \boldsymbol{TX}\right) h_1(tr[\boldsymbol{X\Phi}])d\boldsymbol{X} \\ &= c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \sum_{k=0}^{\infty}\sum_{\kappa} \frac{2^{\frac{nm}{2}+k} h_1^{(k)}(0)(\frac{n}{2})_\kappa \Gamma_m(\frac{n}{2})\Gamma(\frac{nm}{2}+k)}{k!\Gamma(\frac{nm}{2}+k)} \\ &\quad \times |\mathbf{\Sigma}^{-1} - 2\boldsymbol{T}|^{-\frac{n}{2}} C_\kappa(\mathbf{\Phi}(\mathbf{\Sigma}^{-1} - 2\boldsymbol{T})^{-1}) \end{aligned}$$

and the proof is complete. □

Another important statistical characteristic is the joint pdf of the eigenvalues of $\mathbf{X}$, which is given in the next theorem.

Theorem 2.3. *Let* $\mathbf{X} \sim \mathbb{W}_m^I(n, \mathbf{\Sigma}, \mathbf{\Phi})$, *then the joint pdf of the eigenvalues* $\lambda_1 > \lambda_2 > ... > \lambda_m > 0$ *of* $\mathbf{X}$ *is*

$$\frac{c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}}\pi^{\frac{1}{2}m^2}}{\Gamma_m\left(\frac{m}{2}\right)} \prod_{i<j}^{m} (\lambda_i - \lambda_j)\, |\mathbf{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}}$$
$$\times \sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\rho}\sum_{\kappa}\sum_{\phi\in\rho,\kappa} \frac{h_1^{(k)}(0) C_\phi^{\rho,\kappa}\left(\mathbf{I}_m, \mathbf{I}_m\right) C_\phi^{\rho,\kappa}\left(-\frac{1}{2}\mathbf{\Sigma^{-1}}, \mathbf{\Phi}\right)}{r!k!\left[C_\phi\left(\mathbf{I}_m\right)\right]^2} C_\phi\left(\mathbf{\Lambda}\right).$$

Proof: From Theorem 3.2.17, p.104 of Muirhead (2005) the pdf of $\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_m)$ is

$$\frac{\pi^{\frac{1}{2}m^2}}{\Gamma_m\left(\frac{m}{2}\right)} \prod_{i<j}^{m} (\lambda_i - \lambda_j) \int_{\mathcal{O}(m)} g(\mathbf{H}\mathbf{\Lambda}\mathbf{H}'; \mathbf{\Theta}) d\mathbf{H},$$

where $\mathcal{O}(m)$ is the space of all orthogonal matrices $\boldsymbol{H}$ of order m.

Note that

$$\begin{aligned}
\mathbb{I} &= \int_{O(m)} g(\mathbf{H}\mathbf{\Lambda}\mathbf{H}'; \mathbf{\Theta}) d\mathbf{H} \\
&= c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} \int_{\mathcal{O}(m)} |\mathbf{H}\mathbf{\Lambda}\mathbf{H}'|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Sigma^{-1}}\mathbf{H}\mathbf{\Lambda}\mathbf{H}'\right) h_1(tr[\mathbf{H}\mathbf{\Lambda}\mathbf{H}'\mathbf{\Phi}]) d\mathbf{H}
\end{aligned}$$

By using (2.3), we get

$$\begin{aligned}
\mathbb{I} &= c_{n,m}(\mathbf{\Theta})|\mathbf{\Sigma}|^{-\frac{n}{2}} |\mathbf{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}} \sum_{r=0}^{\infty}\frac{1}{r!}\sum_{k=0}^{\infty}\frac{h_1^{(k)}(0)}{k!} \\
&\times \sum_{\rho}\sum_{\kappa}\int_{O(m)} C_\rho\left(-\frac{1}{2}\mathbf{\Sigma^{-1}}\mathbf{H}\mathbf{\Lambda}\mathbf{H}'\right) C_\kappa\left(\mathbf{\Phi}\mathbf{H}\mathbf{\Lambda}\mathbf{H}'\right) d\mathbf{H}.
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_{O(m)} C_\rho\left(-\frac{1}{2}\mathbf{\Sigma}\mathbf{H}\mathbf{\Lambda}\mathbf{H}'\right) C_\kappa\left(\mathbf{\Phi}\mathbf{H}\mathbf{\Lambda}\mathbf{H}'\right) d\mathbf{H} \\
&= \sum_{\phi\in\rho,\kappa} \frac{C_\phi\left(\mathbf{\Lambda}\right) C_\phi^{\rho,\kappa}\left(\mathbf{I}_m, \mathbf{I}_m\right) C_\phi^{\rho,\kappa}\left(-\frac{1}{2}\mathbf{\Sigma}, \mathbf{\Phi}\right)}{\left[C_\phi\left(\mathbf{I}_m\right)\right]^2}
\end{aligned}$$

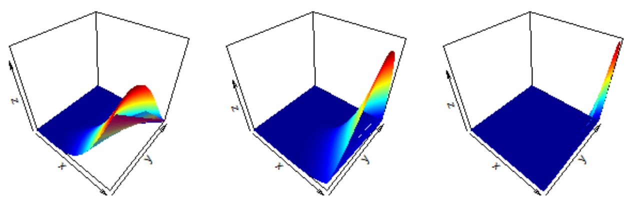
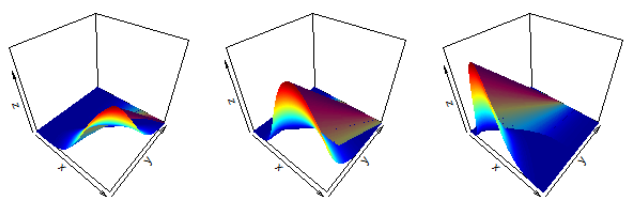
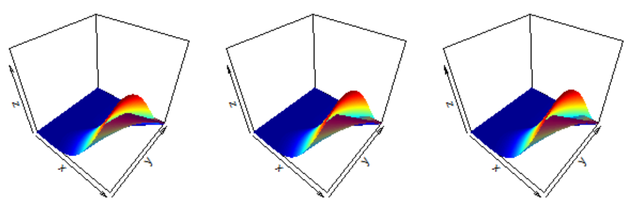
from (1.2), p.468 of Davis (1979) and the result follows. □

Remark 2.3. For $\boldsymbol{\Sigma} = c_1\boldsymbol{I}$ and $\boldsymbol{\Phi} = c_2\boldsymbol{I}$ the result can be obtained from Theorem 3.2.17, p.104 of Muirhead (2005) as follows:

$$\mathbb{I} = c_{n,m}(\boldsymbol{\Theta})c_1^{-\frac{mn}{2}}|\boldsymbol{\Lambda}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{c_1}{2}\boldsymbol{\Lambda}\right) h_1\left(c_2\boldsymbol{\Lambda}\right). \tag{2.7}$$

Based on Remark 2.3, Table 1 illustrates the joint pdf of the eigenvalues of $\boldsymbol{X}_{2\times 2}$ for specific c_1, c_2 and n and different weight functions using (2.7). It is evident that the functional form of the weight function provides increased flexibility for the user. Negative and positive correlations amongst the eigenvalues can be obtained using different weight functions, $h_1(.)$.

Table 1: Joint pdf of the eigenvalues for $n = 9, c_2 = 1$ and $c_1 = 0.1$ (Left), $c_1 = 0.5$ (Middle) and $c_1 = 1.5$ (Right)

$h_1(c_2x) = exp(c_2x)$	
$h_1(c_2x) = exp(\frac{1}{c_2x})$	
$h_1(c_2x) = 1 + c_2x$	

3. FURTHER DEVELOPMENTS

3.1. Weighted-type II Wishart distribution

In this section we focus on the construction of a weighted-type Wishart distribution for which the weight function is of determinant form (see (1.3)). Before exploring the form of the weighted-type Wishart distribution based on a weight of determinant form, let $\boldsymbol{X} \sim W_m(n, \boldsymbol{\Sigma})$ and μ_k denote the k-th moment of $|\boldsymbol{X}|$. Then from (15), p.101 of Muirhead (2005)

$$\mu_k = E\left[|\boldsymbol{X}|^k\right] = \frac{2^k \Gamma_m\left(\frac{n}{2}+k\right)}{\Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{\Sigma}|^k.$$

Thus for any Borel measurable function $h_2(.)$, making use of Taylor's series expansion, we have

$$h_2(|\boldsymbol{X\Phi}|) = \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} |\boldsymbol{X\Phi}|^k. \tag{3.1}$$

Hence

$$E\left[h_2(|\boldsymbol{X\Phi}|)\right] = \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0)}{k!} |\boldsymbol{\Phi}|^k \mu_k = \sum_{k=0}^{\infty} \frac{2^k \Gamma_m\left(\frac{n}{2}+k\right) h_2^{(k)}(0)}{k! \Gamma_m\left(\frac{n}{2}\right)} |\boldsymbol{\Phi\Sigma}|^k.$$

Accordingly, we have the following definition for a weighted-type Wishart distribution with weight of determinant form (see (1.3)).

Definition 3.1. The random matrix $\mathbf{X} \in S_m$ is said to have a weighted-type II Wishart distribution (W2WD) with parameters $\boldsymbol{\Psi}$, $\boldsymbol{\Phi} \in \mathcal{S}_m$ and the weight function $h_2(.)$, if it has the following pdf

$$\begin{aligned} g(\mathbf{X}; \boldsymbol{\Theta}) &= \frac{h_2(|\boldsymbol{X\Phi}|) f(\boldsymbol{X}; \boldsymbol{\Psi})}{E\left[h_2(|\boldsymbol{X\Phi}|)\right]} \\ &= c^*_{n,m}(\boldsymbol{\Theta}) |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_2(|\boldsymbol{X\Phi}|), \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi}, \boldsymbol{\Phi}) \end{aligned}$$

with

$$\begin{aligned} \left\{c^*_{n,m}(\boldsymbol{\Theta})\right\}^{-1} &= \int_{\mathcal{S}_m} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_2(|\boldsymbol{X\Phi}|) d\boldsymbol{X} \\ &= \sum_{k=0}^{\infty} \frac{h_2^{(k)}(0) 2^{\frac{(n+2k)m}{2}} \Gamma_m\left(\frac{n+2k}{2}\right)}{k!} |\boldsymbol{\Phi\Sigma}|^k. \end{aligned}$$

and $f(\boldsymbol{X}; \boldsymbol{\Psi})$ is the pdf of the Wishart distribution $(W_m(n, \boldsymbol{\Sigma}))$ i.e. $\boldsymbol{\Psi} = (\boldsymbol{\Sigma}, n)$, $n > m-1$, $\boldsymbol{\Sigma} \in \mathcal{S}_m$ and $h_2(.)$ is a Borel measurable function that admits Taylor's series expansion. The parameters are restricted to take those values for which the pdf is non-negative. We write this as $\boldsymbol{X} \sim \mathbb{W}^{II}_m(n, \boldsymbol{\Sigma}, \boldsymbol{\Phi})$.

3.2. Weighted-type III Wishart distribution

As before, in this section we give the definition of the weighted-type III Wishart distribution (W3WD). Utilizing a more extended version of (1.4) (allowing more parameters) we have the following definition:

Definition 3.2. The random matrix $\mathbf{X} \in S_m$ is said to have a weighted-type III Wishart distribution (W3WD) with parameters $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2 \in \mathcal{S}_m$ and the weight functions $h_1(.)$ and $h_2(.)$, if it has the following pdf

$$\begin{aligned} g(\mathbf{X};\boldsymbol{\Theta}) &= \frac{h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|)f(\boldsymbol{X};\boldsymbol{\Psi})}{E\left[h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|)\right]}, \quad \boldsymbol{\Theta} = (\boldsymbol{\Psi}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) \\ &= c^{**}_{n,m}(\boldsymbol{\Theta})|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{X}|^{\frac{n}{2}-\frac{m+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) h_1(tr[\boldsymbol{X}\boldsymbol{\Phi}_1])h_2(|\boldsymbol{X}\boldsymbol{\Phi}_2|) \end{aligned}$$

with

$$\begin{aligned} \{c^{**}_{n,m}(\boldsymbol{\Theta})\}^{-1} = \frac{1}{2^{\frac{nm}{2}}} \sum_{k=0}^{\infty}\sum_{t=0}^{\infty} \frac{h_1^{(k)}(0)h_2^{(t)}(0)}{k!t!} 2^{mt+k}\Gamma_m\left(\frac{n}{2}+t\right)|\boldsymbol{\Phi}_1\boldsymbol{\Sigma}|^t \sum_{\kappa}\left(\frac{n}{2}+t\right)_{\kappa} \\ \times C_{\kappa}(\boldsymbol{\Phi}_1\boldsymbol{\Sigma}). \end{aligned}$$

where $f(\boldsymbol{X};\boldsymbol{\Psi})$ is the pdf of the Wishart distribution ($W_m(n,\boldsymbol{\Sigma})$) i.e. $\boldsymbol{\Psi} = (\boldsymbol{\Sigma}, n)$, $n > m-1$, $\boldsymbol{\Sigma} \in \mathcal{S}_m$ and $h_1(.)$ and $h_2(.)$ are Borel measurable functions that admit Taylor's series expansion. We denote this as $\boldsymbol{X} \sim \mathbb{W}_m^{III}(n,\boldsymbol{\Sigma},\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2)$. The parameters are restricted to take those values for which the pdf is non-negative.

4. APPLICATION

In this section special cases of Definition 1 are applied as priors for the normal model under the squared error loss function. First the Kummer gamma distribution ((2.6) with $\phi = 1$) as a prior for the variance of the univariate normal distribution and secondly the Kummer Wishart (2.5) as a prior for the covariance matrix of the matrix variate normal distribution. Bekker and Roux (1995) considered the Wishart prior as a competitor for the conjugate inverse-Wishart prior for the covariance matrix of the matrix variate normal distribution. Van Niekerk et al. (2016a) confirmed the value added of the latter by a numerical study. This is the stimulus to consider other possible priors.

4.1. Univariate Bayesian illustration

In this section a special univariate case of the weighted-type I Wishart distribution is applied as a prior for the variance of the normal model. Consider a random sample of size n_1 from a univariate normal distribution with unknown mean and variance, i.e. $X_i \sim N(\mu, \sigma^2), \boldsymbol{X} = (X_1, ..., X_{n_1})$. Let $h(x) = (1+x)^\gamma$, $m = 1$ and $\boldsymbol{\Sigma}_{1\times 1} = \sigma^2$ in Definition 2.1 (see (2.6)), and consider this distribution as a prior for σ^2, and an objective prior for μ. This prior model is compared with the well-known inverse gamma and gamma priors in terms of coverage and median credible interval width. The three priors under consideration are

- Inverse gamma prior ($IG(\alpha_1, \beta_1)$) with pdf
$$g(x;\alpha_1,\beta_1) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} x^{-\alpha_1-1} \exp\left(-\frac{\beta_1}{x}\right), \quad x > 0$$
- Gamma prior (2.4) ($G(\alpha_2, \beta_2)$)
- Kummer gamma prior (2.6) ($KG(\alpha_3, \beta_3, \gamma = 1)$).

The marginal posterior pdf and Bayes estimator of σ^2 under the Kummer gamma prior are calculated using Remark 5 of Van Niekerk et al. (2016b) as

$$q(\sigma^2|\boldsymbol{X}) = \frac{(\sigma^2)^{\alpha_3 - \frac{n_1}{2} - \frac{1}{2}} \exp\left(-\beta_3\sigma^2\right)\left(1+\phi\sigma^2\right) \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)}{\Gamma\left(\alpha_3+\frac{1}{2}\right)\beta_3^{\alpha_3+\frac{1}{2}} E_{\sigma_1^2}\left[(\sigma_1^2)^{-\frac{n_1}{2}}(1+\phi\sigma_1^2)\exp\left(-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)\right]},$$

and

$$\widehat{\sigma^2} = \frac{\beta_3\Gamma\left(\alpha_3+\frac{3}{2}\right) E_{\sigma_2^2}\left[(\sigma_2^2)^{-\frac{n_1}{2}}(1+\phi\sigma_2^2)\exp\left(-\frac{1}{2\sigma_2^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)\right]}{\Gamma\left(\alpha_3+\frac{1}{2}\right) E_{\sigma_1^2}\left[(\sigma_1^2)^{-\frac{n_1}{2}}(1+\phi\sigma_1^2)\exp\left(-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{n_1} X_i^2 - \overline{X}^2\right]\right)\right]},$$

where $\sigma_1^2 \sim G\left(\alpha_3 + \frac{1}{2}, \beta_3\right)$ and $\sigma_2^2 \sim G\left(\alpha_3 + \frac{3}{2}, \beta_3\right)$.

A normal sample of size 18 is simulated with mean $\mu = 0$ and variance $\sigma^2 = 1$. The hyperparameters are chosen such that $E(\sigma^2) = \sigma_0^2 = 0.9$.

Four combinations of hyperparameter values are investigated and summarized in Table 2. Note that in combination 4, the prior belief for the Kummer gamma is not 0.9 but 5.23, which is quite far from the target value of 1 and the prior information is clearly misspecified. It is clear from Table 2 that the Kummer gamma prior, with parameter combination 4, is very vague when compared to the other two priors. To evaluate the performance of the new prior structure, 1000 independent samples are simulated and for each one the posterior densities and estimates are calculated. This enables the calculation of the coverage probabilities and median credible interval width as given in Table 3.

The coverage probability obtained under the Kummer gamma prior is higher than for the inverse-gamma and gamma priors, while the median width of the credible interval (indicated in brackets) is competitive. It is interesting to note that even under total misspecification (see combination 4 in Table 2), the Kummer gamma prior is still performing well. The performance superiority of the Kummer gamma prior is clear from Table 3.

Table 2: Influence of hyperparameters on the prior pdf's (– Kummer gamma prior, - - Inverse gamma prior, ... Gamma prior)

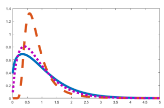
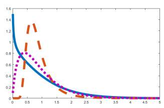
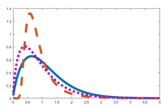
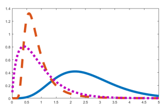
Combination	1	2	3	4
Inverse gamma prior	$\alpha_1 = 3.22$, $\beta_1 = 2$	$\alpha_1 = 4.33$, $\beta_1 = 3$	$\alpha_1 = 3.22$, $\beta_1 = 2$	$\alpha_1 = 3.22$, $\beta_1 = 2$
Gamma prior	$\alpha_2 = 1.8$, $\beta_2 = 2$	$\alpha_2 = 1.8$, $\beta_2 = 2$	$\alpha_2 = 1.8$, $\beta_2 = 2$	$\alpha_2 = 1.8$, $\beta_2 = 2$
Kummer gamma prior	$\alpha_3 = 1.2$, $\beta_3 = 2.1$	$\alpha_3 = 0.8$, $\beta_3 = 1.8$	$\alpha_3 = 1.8$, $\beta_3 = 2.5$	$\alpha_3 = 5.0$, $\beta_3 = 2.5$
				

Table 3: Coverage probabilities (median credible interval width) calculated from the posterior density functions

Combination	Inverse gamma prior	Gamma prior	Kummer gamma prior
1	74.5%(0.8)	77.4%(2.15)	90.8%(0.85)
2	72%(0.75)	78.1%(2.125)	89.2%(0.9)
3	76.8%(0.85)	77.7%(2.05)	92.1%(0.9)
4	73.2%(0.8)	77.6%(1.8)	87.8%(1.4)

4.2. Multivariate Bayesian illustration

In this section the Kummer Wishart (2.5) prior is considered for the covariance matrix of the matrix variate normal model. Consider a random sample of size n_1 from a matrix variate normal distribution with unknown mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{X}_i \sim N_{m,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_p)$ with likelihood function

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \overline{\boldsymbol{X}}, \boldsymbol{V}) \propto |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right].$$

The three priors for $\boldsymbol{\Sigma}$ under consideration are

- Inverse Wishart prior ($IW_m(p_1, \boldsymbol{\Phi})$) with pdf

$$g(\boldsymbol{X}; p_1, \boldsymbol{\Phi}) = \left[2^{\frac{m(p_1-m-1)}{2}}\Gamma_m\left(\frac{p_1-m-1}{2}\right)\right]^{-1} |\boldsymbol{X}|^{-\frac{p_1}{2}} |\boldsymbol{\Phi}|^{\frac{p_1-m-1}{2}} \times etr\left[-\frac{1}{2}\boldsymbol{X}^{-1}\boldsymbol{\Phi}\right], \qquad \boldsymbol{X} \in S_m$$

- Wishart prior (1.5) ($W_m(p_2, \boldsymbol{\Phi})$)
- Kummer Wishart prior (2.5) ($KW_m(p_3, \boldsymbol{I}, \boldsymbol{\Phi})$).

The conditional posterior pdf's of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with a Kummer Wishart prior and objective prior for $\boldsymbol{\mu}$, necessary for the simulation of the posterior samples, are

$$q\left(\boldsymbol{\mu}|\boldsymbol{\Sigma}, \overline{\boldsymbol{X}}, \boldsymbol{V}\right) \propto etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right],$$

and

$$\begin{aligned} q\left(\boldsymbol{\Sigma}|\boldsymbol{\mu}, \overline{\boldsymbol{X}}, \boldsymbol{V}\right) \;\propto\; & |\boldsymbol{\Sigma}|^{-\frac{n_1 p}{2} + \frac{n}{2} - \frac{m+1}{2}} etr\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{V} + n_1(\overline{\boldsymbol{X}} - \boldsymbol{\mu})(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\right]\right] \\ & \times \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}\right)(1 + tr(\boldsymbol{\Sigma}\boldsymbol{\Theta})). \end{aligned}$$

with $\boldsymbol{V} = \sum_{i=1}^{n_1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$. A sample of size 10 is simulated from a multivariate normal distribution ($p = 1$) with $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_m$. The hyperparameters are chosen as $\boldsymbol{\Phi} = \boldsymbol{I}_m, m = 3, p_1 = 9.5, p_2 = p_3 = 3$, according to the methodology of Van Niekerk et al. (2016a). Posterior samples of size 5000, are simulated using a Gibbs sampling scheme with an additional Metropolis-Hastings algorithm, similarly to Van Niekerk et al. (2016a).

The estimates calculated for $\boldsymbol{\Sigma}$ under the three different priors as well as the MLE are

$$\widehat{\boldsymbol{\Sigma}}_{MLE} = \begin{bmatrix} 1.8719 & 0.2168 & 0.9523 \\ 0.2168 & 2.9553 & -0.2471 \\ 0.9523 & -0.2471 & 1.0715 \end{bmatrix}, \widehat{\boldsymbol{\Sigma}}_{IW} = \begin{bmatrix} 0.6600 & 0.0772 & 0.3355 \\ 0.0772 & 1.0256 & -0.0873 \\ 0.3355 & -0.0873 & 0.3762 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_{W} = \begin{bmatrix} 0.5547 & 0.0627 & -0.2247 \\ 0.0627 & 0.7968 & 0.0255 \\ -0.2247 & 0.0255 & 1.2348 \end{bmatrix}, \widehat{\boldsymbol{\Sigma}}_{KW} = \begin{bmatrix} 1.1389 & 0.0115 & -0.0098 \\ 0.0115 & 1.0401 & -0.0132 \\ -0.0098 & -0.0132 & 1.0763 \end{bmatrix}$$

The above estimates are obtained for one posterior sample. The Frobenius norm (see Golub and Van Loan (1996)) of the errors, defined as $||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_F = \sqrt{tr\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)'\left(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right)}$, are calculated for each estimate and given in Table 4.

The Kummer Wishart prior results in the smallest Frobenius norm of the error. For further investigation, this sampling scheme is repeated 100 times to obtain 100 estimates under each prior as well as the MLE for each of the 100 simulated samples. The Frobenius norm of the error for each estimate and every repetition is calculated and the empirical cumulative distribution function (ecdf) of each set of Frobenius norms calculated for each estimator is obtained and given in Figure 1. The ecdf which is most left in the figure is regarded as the best since for a specific value of the error norm, a higher proportion of estimates

Table 4: Frobenius norm of the error of the estimates calculated from the simulated sample

Frobenius norm	Value
$\|\|\widehat{\mathbf{\Sigma}}_{MLE} - \mathbf{\Sigma}\|\|_F$	1.2336
$\|\|\widehat{\mathbf{\Sigma}}_{IW} - \mathbf{\Sigma}\|\|_F$	0.9928
$\|\|\widehat{\mathbf{\Sigma}}_{W} - \mathbf{\Sigma}\|\|_F$	1.5766
$\|\|\widehat{\mathbf{\Sigma}}_{KW} - \mathbf{\Sigma}\|\|_F$	0.1468

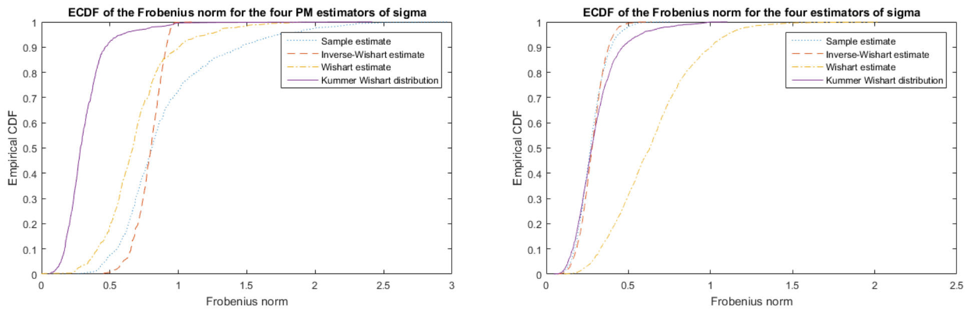


Figure 1: The empirical cumulative distribution function (ecdf) of the Frobenius norm of the estimation errors for $n = 10$(Left) and $n = 100$(Right)

from that particular prior results in less error. It is evident from Figure 1 that the performance of the sample estimate improves as the sample size increases, which is to be expected, and the performance of the Kummer Wishart prior is still competitive. From Figure 1 we conclude that the Kummer Wishart prior results in an estimate of $\mathbf{\Sigma}$, for small and larger sample sizes, with less error and preference should be given to this prior. To validate the graphical interpretation, a two-sample Kolmogorov-Smirnov test is performed for $n = 10$, pairwise, on the three different ecdf's and the p-value for some pairs are given in Table 5.

Table 5: p-values of the Kolmogorv-Smirnov two-sample test based on samples ($n = 10$) of the Frobenius norms

Pairwise comparison	p-value
MLE and IW	< 0.001
IW and KW	< 0.001
W and KW	< 0.001
MLE and KW	< 0.001

From Table 5 it is clear that the ecdf of the errors under the Kummer Wishart prior is significantly different from the other priors. Therefore, the assertion can be made that the Kummer Wishart prior structure produces an estimate that results in statistically significant less error.

5. DISCUSSION

In this paper, we proposed a construction methodology for new matrix variate distributions with specific focus on the Wishart distribution, followed by weighting the Wishart distribution with different weight functions. It was shown from Bayesian viewpoint, by simulation studies, that the Kummer gamma and Kummer Wishart priors, as special cases of the weighted-type I Wishart distribution, outperformed the well-known priors. The weighted-type III Wishart distribution gives rise to a Wishart distribution with larger degrees of freedom and scaled covariance matrix that might have application in missing value analysis.

In the following we list some thoughts that might be considered as plausible applications of the proposed distributions.

(i) Let N_1 and N_2 observations be independently and identically derived from $Z_1 \sim N_m(\mathbf{0}, \mathbf{\Sigma}_1)$ and $Z_2 \sim N_m(\mathbf{0}, \mathbf{\Sigma}_2)$, respectively. Then the statistic $\boldsymbol{T} = \sum_{j=1}^{N}(Z_1 + Z_2)(Z_1 + Z_2)^T$, has the Wishart distribution $W_m(N_1 + N_2, \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)$. Suppose the focus of the paper is on the covariance structure $\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2$ (similar to standby systems), then, to reduce the cost of sampling, one may only consider N_1 observations from the W1WD and take $h_1(.)$ to be of exponential form in (2.1).

(ii) A weight of the form $h_2(|\boldsymbol{X}\mathbf{\Phi}|) = |\boldsymbol{X}|^{\frac{q}{2}}$, where q is a known fixed constant with $\mathbf{\Phi} = \boldsymbol{I}_m$ in Definition 3.1 has applications in missing value analysis. To see this, let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n+q}$ be a random sample from $N_m(\mathbf{0}, \mathbf{\Sigma})$. Define $\boldsymbol{T} = \sum_{i=1}^{n} \boldsymbol{Y}_i \boldsymbol{Y}_i^T$. Then $\boldsymbol{T} \sim W_m(n, \mathbf{\Sigma})$. Now, using the weight $h_2(\boldsymbol{X}) = |\boldsymbol{X}|^{\frac{q}{2}}$ one obtains the $W_m(n + q, \mathbf{\Sigma})$ distribution and without having the observations $n+1, \ldots, n+q$ we can find the distribution of the full sample and the relative analysis.

(iii) Finally, one may ask what is the sampling distribution regarding Definition 3.2? To answer this question, we recall that if $\boldsymbol{Y} \sim N(\mathbf{0}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$, then $\boldsymbol{X} = \boldsymbol{Y}^T\boldsymbol{Y} \sim W_m(n, \mathbf{\Sigma})$. Now, assume matrices $\boldsymbol{A} \in \mathcal{S}_m$ and $\boldsymbol{B} \in \mathcal{S}_m$ exist such that $[\mathbf{\Sigma} + \mathbf{\Phi}]^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$. Then if we sample $\boldsymbol{Y}^* \sim N\left(\mathbf{0}, \boldsymbol{I}_{n+\alpha} \otimes [\mathbf{\Sigma} + \mathbf{\Phi}]\right)$, the quadratic form $\boldsymbol{X}^* = \boldsymbol{Y}^T\boldsymbol{Y}$ will have the distribution as in Definition 3.2, where $\boldsymbol{A} = \mathbf{\Sigma}$, $\boldsymbol{B} = \mathbf{\Phi}_1$ and $\mathbf{\Phi}_2 = \boldsymbol{I}$. In other words, if we enlarge both the covariance and number of samples in a normal population and consider the distribution of the quadratic form, we are indeed weighting a Wishart distribution with a Wishart.

ACKNOWLEDGMENTS

The authors would like to hereby acknowledge the support of the StatDisT group. This work is based upon research supported by the UP Vice-chancellor's post-doctoral fellowship programme, the National Research foundation grant (Re:CPRR3090132066 No 91497) and the vulnerable discipline-academic statistics (STAT) fund. The authors thank the reviewers and associate editor for their helpful recommendations.

REFERENCES

[1] Abul-Magd,A.Y., Akemann, G. and Vivo, P. (2009). Superstatistical generalizations of Wishart-Laguerre ensembles of random matrices, *Journal of Physics A: Mathematical and Theoretical*, **42**(17), 175–207.

[2] Adhikari, S. (2010). Generalized Wishart distribution for probabilistic structural dynamics, *Computational Mechanics*, **45**(5), 495–511.

[3] Arashi, M. (2013). On Selberg-type square matrices integrals, *Journal of Algebraic Systems*, **1**,1, 53–65.

[4] Bekker, A. and Roux, J.J.J. (1995). Bayesian multivariate normal analysis with a Wishart prior, *Communications in Statistics - Theory and Methods*, **24**(10), 2485–2497.

[5] Bryc, W. (2008). Compound real Wishart and q-Wishart matrices, *International Mathematics Research Notices*, rnn079.

[6] Davis, A.W. (1979). Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory, *Annals of the Institute of Statistical Mathematics*, **31**(1), 465–485.

[7] Dawid, A.P. and Lauritzen, S.L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models, *The Annals of Statistics*, 1272–1317.

[8] Díaz-García, J.A. and Gutiérrez-Jáimez, R. and others (2011). On Wishart distribution: some extensions, *Linear Algebra and its Applications*, **435**(6), 1296–1310.

[9] Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations, 3rd ed*, Baltimore, MD: Johns Hopkins.

[10] Letac, G. and Massam, H. (2001). The normal quasi-Wishart distribution, *Contemporary Mathematics*, **287**,231–240.

[11] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York.

[12] Munilla, S. and Cantet, R.J.C. (2012). Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic parameters–an application to the maternal animal model, *Journal of Animal Breeding and Genetics*, **129**(3), 173–187.

[13] Patil, G. P. and Ord, J. K. (1976). On size-biased sampling and related form-invariant weighted distributions, *Sankhyā*, **38B**, 48-61.

[14] Pauw, J., Bekker, A. and Roux, J.J.J. (2010). Densities of composite weibullized generalized gamma variables, *South African Statistical Journal*, **44**, 17–42.

[15] Rao, C. R. (1965). *On discrete distributions arising out of methods of ascertainment*, In Classical and Contagious Pergamon Press and Statistical Publishing Society, 320–332.

[16] Rènyi, A. (1961). On measures of information and entropy, *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, 1960*, 547–561.

[17] Roverato, A. (2002). Hyper Inverse Wishart Distribution for Non-decomposable Graphs and its Application to Bayesian Inference for Gaussian Graphical Models, *Scandinavian Journal of Statistics*, **29**(3),391–411.

[18] Shannon, C.E. (1948). A Mathematical Theory of Communication, *Bell System Technical Journal*, **27**(3), 379–423.

[19] Sutradhar, B.C. and Ali, M.M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model, *Journal of Multivariate Analysis*, **29**(1),155–162.

[20] Teng, C., Fang H. and Deng, W. (1989). The generalized noncentral Wishart distribution, *Journal of Mathematical Research and Exposition*, **9**(4), 479–488.

[21] Van Niekerk, J., Bekker, A., Arashi, M. and De Waal, D.J. (2016a). Estimation under the matrix variate elliptical model, *South African Statistical Journal*, **50**(1), 149–171.

[22] Van Niekerk, J., Bekker, A., and Arashi, M. (2016b). A gamma-mixture class of distributions with Bayesian application, *Communications in Statistics-Simulation and Computation*, Accepted.

[23] Wang, H. and West, M. (2009). Bayesian analysis of matrix normal graphical models, *Biometrika*, **96**(4),821–834.

[24] Wong, C.S. and Wang, T. (1995). Laplace-Wishart distributions and Cochran theorems, *Sankhyā: The Indian Journal of Statistics, Series A*, 342–359.

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 223-250

BIVARIATE BETA AND KUMARASWAMY MODELS DEVELOPED USING THE ARNOLD-NG BIVARIATE BETA DISTRIBUTION

Authors: Barry C. Arnold
– Department of Statistics, University of California, Riverside, USA (barry.arnold@ucr.edu)

Indranil Ghosh
– Department of Mathematics & Statistics, University of North Carolina, Wilmington, USA (ghoshi@uncw.edu)

Received: March 2016 Revised: October 2016 Accepted: January 2017

Abstract:

- In this paper we explore some mechanisms for constructing bivariate and multivariate beta and Kumaraswamy distributions. Specifically, we focus our attention on the Arnold-Ng (2011) eight parameter bivariate beta model. Several models in the literature are identified as special cases of this distribution including the Jones-Olkin-Liu-Libby-Novick bivariate beta model, and certain Kotz and Nadarajah bivariate beta models among others. The utility of such models in constructing bivariate Kumaraswamy models is investigated. Structural properties of such derived models are studied. Parameter estimation for the models is also discussed. For illustrative purposes, a real life data set is considered to exhibit the applicability of these models in comparison with rival bivariate beta and Kumaraswamy models.

Key-Words:

- *bivariate beta, second kind beta distribution, gamma components, copulas, bivariate Kumaraswamy models.*

AMS Subject Classification:

- 62E, 60F.

1. INTRODUCTION

Kumaraswamy (1980) introduced a two parameter absolutely continuous distribution which compares extremely favorably, in terms of simplicity, with the beta distribution. The Kumaraswamy distribution (hereafter the K distribution) on the interval $(0, 1)$, has its probability density function (pdf) and its cdf with two shape parameters $\delta > 0$ and $\beta > 0$ defined by

$$f(x) = \delta\beta x^{\delta-1}(1-x^{\delta})^{\beta-1}I(0 < x < 1) \quad \text{and } F(x) = 1-(1-x^{\delta})^{\beta}. \tag{1.1}$$

If a random variable X has (1.1) as its density then we will write $X \sim K(\delta, \beta)$.

The density function in (1.1) has similar properties to those of the beta distribution. The Kumaraswamy pdf is unimodal, uniantimodal, increasing, decreasing or constant depending (similar to the beta distribution) on the values of the parameters. The construction of bivariate Kumaraswamy distributions has received limited attention.

Barreto-Souza and Lemonte (2013) introduced a bivariate Kumaraswamy distribution related to a Marshall-Olkin survival copula. They discussed some structural properties of their bivariate Kumaraswamy distribution, including a detailed discussion of estimation of the model parameters. Recently Arnold and Ghosh (2016) discussed some different strategies for constructing legitimate bivariate Kumaraswamy models via conditional specification, conditional survival function specification and via a copula based approach. In this paper, we consider several specialized approaches to the problem of constructing bivariate K distributions based on sub-models of the Arnold-Ng 8-parameter bivariate beta distribution. Included is discussion of the Jones-Olkin-Liu-Libby-Novick bivariate beta distribution and two Kotz and Nadarajah (2007) bivariate beta models.

To carry out this program, we make use of the observation that a Kumaraswamy distribution is a special case of the generalized beta distribution which is that of a positive power of a beta random variable.

In this paper we will make repeated use of the fact that a Kumaraswamy variable can be viewed as a power of a beta variable. Thus,

$$\text{if } \; Y \sim Beta(1, \beta), \quad \text{then for } \; \delta > 0, \quad X = Y^{1/\delta} \sim K(\delta, \beta).$$

Our proposed flexible families of bivariate Kumaraswamy distributions will be obtained by applying such marginal power transformations to suitable bivariate beta models. It will be convenient to begin with a careful discussion of the 8-parameter bivariate beta distribution introduced by Arnold and Ng (2011), together with its related sub-models and possible higher dimensional extensions. Note that in Arnold and Ghosh (2016), use was made of the simpler 5- parameter Arnold-Ng model in an analogous program for developing bivariate Kumaraswamy models. The present paper thus represents a natural extension of some of the results in that earlier paper.

We will begin with a detailed discussion of the 8-parameter Arnold-Ng model with marginals of the second kind beta type. The corresponding models with classical (first kind) beta marginals are then obtained via simple marginal transformations, using the observation that

$$\text{if } U \sim Beta^{(2)}(\alpha,\beta) \text{ then } (1+U^{-1})^{-1} \sim Beta(\alpha,\beta).$$

With suitable parametric restrictions, corresponding bivariate Kumaraswamy models are then readily derived.

The remainder of this article is organized as follows: In section 2, as mentioned above, we review the Arnold-Ng (2011) eight parameter bivariate second kind beta model. We consider the many sub-models that are obtained via parametric restrictions, and we discuss higher dimensional versions of this model. In section 3, we discuss the parallel models with marginals that are of the first or classical beta kind. In Section 4, we briefly consider the construction of bivariate generalized beta distributions. In Section 5, the useful concepts of reciprocation closure and closure under reflection about the point 1/2 are reviewed. Section 6 deals with a catalog of bivariate Kumaraswamy distributions obtained via marginal power transformations applied to certain bivariate beta variables. In section 7, we revisit the concept of reflection about 1/2. Section 8 includes some discussion of possible parameter estimation strategies for the models. Section 9 includes an illustrative application in which one of the bivariate Kumaraswamy models is compared with some competing models when fitted to a particular data set. Some concluding remarks are contained in Section 10.

2. BIVARIATE SECOND KIND BETA DISTRIBUTIONS

A random variable X is said to have a second kind beta distribution with positive parameters α_1 and α_2, if its density is of the form

$$f_X(x) = \frac{1}{B(\alpha_1,\alpha_2)} \frac{x^{\alpha_1-1}}{(1+x)^{\alpha_1+\alpha_2}} I(x>0)$$

and, in such a case, we write $X \sim B^{(2)}(\alpha_1,\alpha_2)$.

In our subsequent discussion we make considerable use of the observation that if V_1, V_2 are independent gamma distributed random variables with $V_i \sim \Gamma(\alpha_i, 1)$, $i = 1, 2$, then $X = V_1/V_2 \sim B^{(2)}(\alpha_1,\alpha_2)$.

The construction of the Arnold-Ng (2011) 8-parameter bivariate second kind beta distribution begins with 8 independent gamma distributed random variables, $U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8$ with $U_i \sim \Gamma(\alpha_i, 1)$, $i = 1, 2, ..., 8$. The random vector (V_1, V_2) is then defined by

$$V_1 = \frac{U_1+U_5+U_7}{U_3+U_6+U_8} \tag{2.1}$$

and

$$V_2 = \frac{U_2 + U_5 + U_8}{U_4 + U_6 + U_7}. \tag{2.2}$$

Utilizing the fact that sums of independent gamma variables with the same scale parameter again are gamma distributed, we see that

$$V_1 \sim B^{(2)}(\alpha_1 + \alpha_5 + \alpha_7, \alpha_3 + \alpha_6 + \alpha_8)$$

and

$$V_2 \sim B^{(2)}(\alpha_2 + \alpha_5 + \alpha_8, \alpha_4 + \alpha_6 + \alpha_7).$$

This 8-parameter model is the most general bivariate second kind beta model that can be constructed via ratios of sums of independent gamma variables. Some indication of how it was developed will be useful for envisioning how to construct higher dimensional versions. There are four places in which a gamma distributed variable can appear in (2.1)-(2.2). They are: in the numerator of (2.1), in the denominator of (2.1), in the numerator of (2.2) and in the denominator of (2.2). A random variable U might appear only once in the two ratios. This is the case for U_1, U_2, U_3 and U_4, each of which appears in a different one of the four available places. A random variable U might appear in two of the available four places, but to retain the independence of the numerator from its corresponding denominator in a given ratio, the same U cannot appear in both. There are four different ways in which a variable U can appear in two places as illustrated by U_5, U_6, U_7, and U_8. Thus U_5 appears in both numerators, U_6 appears in both denominators, etc.. A random variable U cannot appear in more two of the four places without violating the required independence of numerators from their corresponding denominators. This thus results in the appearance of the 8 U_i's in the general model (2.1)-(2.2). The addition of any more gamma distributed U's to the model in any one or two places will not yield a more general model since they could be combined with already present gamma variables to keep the dimension of the parameter vector at the value 8.

The three dimensional version of this construction will involve 26 U_i's. This number can be verified by noting that a trivariate model (V_1, V_2, V_3) expressed as ratios of independent linear combinations of independent gamma variables (with unit scale parameter), will involve 6 places where a particular U can appear, three numerators and three denominators. But a particular U cannot appear in both the numerator and denominator of any of the three V_i's. There will be 6 U's which appear in one of the 6 possible places. These will be denoted by $U_1, U_2, ..., U_6$. There will be 12 U's that appear in exactly two of the 6 possible positions, denoted by $U_7, U_8, ..., U_{18}$. Finally there are 8 U's that appear in 3 places, namely $U_{19}, U_{20}, ..., U_{26}$. No U can appear in more than 3 places without violating the requirement that numerators must be independent of their corresponding denominators.

Thus, there are a total of 26 parameters in the model where U_i, $i = 1, 2, ..., 26$ are independent variables with $U_i \sim \Gamma(\alpha_i, 1)$ for each i. The model can then be expressed in the following, somewhat daunting, form.

$$V_1 = \frac{U_1 + U_7 + U_8 + U_9 + U_{10} + U_{19} + U_{20} + U_{21} + U_{22}}{U_4 + U_{11} + U_{12} + U_{13} + U_{14} + U_{23} + U_{24} + U_{25} + U_{26}},$$

$$V_2 = \frac{U_2 + U_7 + U_{11} + U_{15} + U_{16} + U_{19} + U_{20} + U_{23} + U_{24}}{U_5 + U_9 + U_{13} + U_{17} + U_{18} + U_{21} + U_{22} + U_{25} + U_{26}},$$

and

$$V_3 = \frac{U_3 + U_8 + U_{12} + U_{15} + U_{17} + U_{19} + U_{21} + U_{23} + U_{25}}{U_6 + U_{10} + U_{14} + U_{16} + U_{18} + U_{20} + U_{22} + U_{24} + U_{26}}.$$

The pattern for the dimensions of parameter spaces of the multivariate models becomes clear. The univariate model involves 2 U's, i.e., $3^1 - 1$. The bivariate model involves 8 U's, i.e., $3^2 - 1$. The trivariate case involves 26 U's, i.e., $3^3 - 1$., and so on. The general four dimensional model has 80 parameters! The enormous number of parameters involved in the completely general 3 and 4 dimensional models (i.e., 26 and 80) will compel us to consider simplified sub-models, of somewhat restricted flexibility, obtained by setting some of the α's equal to 0. This may well be desirable, even in the bivariate case. The full array of sub-models of the 8 parameter model (2.1)-(2.2) can be enumerated as follows.

There is, to begin with, the full 8-parameter model in which all of the α_i's are positive. We can label the various sub-models by listing the subscripts of the α_i's which remain in the sub-model, i.e., which have not been set equal to 0. Thus $B^{(2)}(1, 2, 3, 4, 5.6, 7, 8)$ denotes the full model, while for example $B^{(2)}(1, 5, 6)$ denotes the model in which only α_1, α_5 and α_6 have not been set equal to 0. Note that the list of subscripts of the α_i's that are set equal to zero cannot include any of the four triples $(1, 5, 7), (3, 6, 8), (2, 5, 8)$ or $(4, 6, 7)$ in order to retain the second kind beta form for the marginal distributions. Thus, the list of permissible sub-models includes:

- $\binom{8}{1} = 8$ models in which just one of the α_i's has been set equal to 0,
- $\binom{8}{2} = 28$ models in which exactly two of the α_i's have been set equal to 0,
- $\binom{8}{3} - 4 = 52$ permissible models in which exactly three of the α_i's have been set equal to 0,
- $\binom{8}{4} - 20 = 50$ permissible models in which exactly four of the α_i's have been set equal to 0,
- $\binom{8}{5} - 36 = 20$ permissible models in which exactly five of the α_i's have been set equal to 0,
- $\binom{8}{6} - 26 = 2$ permissible models in which exactly six of the α_i's have been set equal to 0. Of these models, $BB^{(2)}(5, 6)$ has $V_1 = V_2$, while $BB^{(2)}(7, 8)$ has $V_1 = 1/V_2$, so that they are of little interest.

In all there are 161 models which might be considered, of which 159 are non trivial. As we shall see in the next section, several of the corresponding bivariate beta of the first kind models (but not many) have received detailed coverage in the literature. It should be noted that very few of these models have available analytic expressions for the corresponding joint density. Typically those models with more than 3 parameters will not have tractable joint densities.

Returning to the general 8-parameter model (2.1)-(2.2), we may readily write down the moments of the V_i's since they have second kind beta distributions. Thus, for any integer j less than $\alpha_3 + \alpha_6 + \alpha_8$, we have

$$\begin{aligned} E(V_1^j) &= E[(U_1 + U_5 + U_7)^j] E[(U_3 + U_6 + U_8)^{-j}] \\ &= \frac{\Gamma(\alpha_1 + \alpha_5 + \alpha_7 + j)}{\Gamma(\alpha_1 + \alpha_5 + \alpha_7)} \frac{\Gamma(\alpha_3 + \alpha_6 + \alpha_8 - j)}{\Gamma(\alpha_3 + \alpha_6 + \alpha_8)}, \end{aligned}$$

and similarly, for any integer $k < \alpha_4 + \alpha_6 + \alpha_7$,

$$E(V_2^k) = \frac{\Gamma(\alpha_2 + \alpha_5 + \alpha_8 + k)}{\Gamma(\alpha_2 + \alpha_5 + \alpha_8)} \frac{\Gamma(\alpha_4 + \alpha_6 + \alpha_7 - k)}{\Gamma(\alpha_4 + \alpha_6 + \alpha_7)}.$$

Expressions for the variances are then readily written down. However mixed moments are more difficult to deal with. For example, we have

$$E(V_1 V_2) = E\left[\left(\frac{U_1 + U_5 + U_7}{U_3 + U_6 + U_8}\right)\left(\frac{U_2 + U_5 + U_8}{U_4 + U_6 + U_7}\right)\right]$$

which appears to be difficult to evaluate analytically, unless most of the α_i's are equal to 0. Thus, analytic expressions for the covariance between V_1 and V_2 will be usually unavailable. Nevertheless, the covariance and any mixed moments of the form $E(V_1^\ell V_2^m)$ can be readily approximated by repeated simulation of the U_i's, thanks to the strong law of large numbers.

3. BIVARIATE BETA DISTRIBUTIONS (OF THE FIRST, OR CLASSICAL, KIND)

If $U \sim B^{(2)}(\alpha_1, \alpha_2)$, i.e., if $U =^d W_1/W_2$ where the W_i's are independent with $W_i \sim \Gamma(\alpha_i, 1),\ i = 1, 2$, then the random variable $V = (1 - U^{-1})^{-1}$ has a (classical) beta distribution or beta distribution of the first kind, and we denote this by $V \sim B(\alpha_1, \alpha_2)$. Here, V can be represented in the form

$$V = \frac{W_1}{W_1 + W_2}.$$

Application of such a transformation to the marginals of the model (2.1)-(2.2) yields a parallel 8-parameter bivariate (classical) beta distribution with the following structure

$$W_1 = \frac{U_1 + U_5 + U_7}{(U_1 + U_5 + U_7) + (U_3 + U_6 + U_8)} \tag{3.1}$$

and

$$W_2 = \frac{U_2 + U_5 + U_8}{(U_2 + U_5 + U_8) + (U_4 + U_6 + U_7)}, \tag{3.2}$$

where the U_i's are independent gamma distributed random variables with $U_i \sim \Gamma(\alpha_i, 1),\ i = 1, 2, ..., 8$. In this case we write

$$(W_1, W_2) \sim BB(1, 2, 3, 4, 5, 6, 7, 8),$$

indicating that all 8 of the U_i's are involved in the distribution. This is the 8-parameter bivariate beta distribution introduced in Section 6.1 of Arnold and Ng (2011). As was the case for the bivariate beta of the second kind distribution discussed in Section 2, it will often be of interest to consider sub-models in which some of the α_i's are set equal to zero, so that the corresponding U_i's do not appear in the expressions (2.1) and (2.2). Thus for example the model $BB(1, 2, 6, 7, 8)$ may be recognized as the 5-parameter bivariate beta model discussed extensively in Arnold and Ng (2011), while the simpler 3-parameter models $BB(1, 2, 6)$, $BB(3, 5, 6)$, $BB(4, 5, 6)$ and $BB(6, 7, 8)$ have also appeared in the literature, as has the 4-parameter model $BB(5, 6, 7, 8)$.

The $BB(6, 7, 8)$ model is recognizable as a Dirichlet distribution, the $BB(1, 2, 6)$ model is identifiable as the Libby-Novak (1982)-Jones (2002)-Olkin-Liu (2003) model, the $BB(3, 5, 6)$ and $BB(4, 5, 6)$ models are the same as the first two models discussed in Nadarajah and Kotz (2005), and the $BB(5, 6, 7, 8)$ has been discussed by Olkin and Trikalinos (2015). Finally we mention that the BB(1,2,3,4,5,6) model was introduced by Magnussen (2004). Of course, not all bivariate beta models can be viewed as sub-models of (3.1)-(3.2). For example the third model in Nadarajah and Kotz (2005) (which is defined in terms of three independent beta variables) is not of this form, nor are the various copula based models obtained by marginally transforming quite arbitrary bivariate distributions to obtain beta marginals. Moreover some bivariate beta models, such as for example the one in Nadarajah (2007) only have beta marginals in special sub-cases.

In this setting also, there are 161 models which might be considered, of which 159 are non trivial. It, once more, should be noted that very few of these models have available analytic expressions for the corresponding joint density. Typically those models with more than 3 parameters will not have tractable joint densities.

Returning to the general 8-parameter model (3.1)-(3.2), we may readily write down the moments of the W_i's since they have (classical) beta distributions. Thus, for example, the means and variances are given by

$$E(W_1) = \frac{\alpha_1 + \alpha_5 + \alpha_7}{\alpha_1 + \alpha_5 + \alpha_7 + \alpha_3 + \alpha_6 + \alpha_8},$$

$$E(W_2) = \frac{\alpha_2+\alpha_5+\alpha_8}{\alpha_2+\alpha_5+\alpha_8+\alpha_4+\alpha_6+\alpha_7},$$

$$var(W_1) = \frac{(\alpha_1+\alpha_5+\alpha_7)(\alpha_3+\alpha_6+\alpha_8)}{(\alpha_1+\alpha_5+\alpha_7+\alpha_3+\alpha_6+\alpha_8)^2(\alpha_1+\alpha_5+\alpha_7+\alpha_3+\alpha_6+\alpha_8+1)},$$

and

$$var(W_2) = \frac{(\alpha_2+\alpha_5+\alpha_8)(\alpha_4+\alpha_6+\alpha_7)}{(\alpha_2+\alpha_5+\alpha_8+\alpha_4+\alpha_6+\alpha_7)^2(\alpha_2+\alpha_5+\alpha_8+\alpha_4+\alpha_6+\alpha_7+1)}.$$

Although expressions for the variances are readily written down, mixed moments are more difficult to deal with. For example, we have

$$E(W_1W_2) = E\left[\left(\frac{U_1+U_5+U_7}{U_1+U_5+U_7+U_3+U_6+U_8}\right)\left(\frac{U_2+U_5+U_8}{U_2+U_5+U_8+U_4+U_6+U_7}\right)\right]$$

which will be difficult to evaluate analytically, unless most of the α_i's are equal to 0. Thus, analytic expressions for the covariance between W_1 and W_2 will be usually unavailable. As was the case for the second kind beta models, this covariance and any mixed moments of the form $E(W_1^\ell W_2^m)$ can be readily approximated by repeated simulation of the U_i's, using the strong law of large numbers.

4. RECIPROCATION AND REFLECTION ABOUT 1/2

If $X \sim B(\alpha_1,\alpha_2)$ then it follows readily that $1-X \sim B(\alpha_2,\alpha_1)$. Similarly, if $X \sim B^{(2)}(\alpha_1,\alpha_2)$ then $1/X \sim B^{(2)}(\alpha_2,\alpha_1)$. In words, the family of beta distributions is closed under reflection about the point 1/2, and the family of second kind beta distributions is closed under reciprocation. If one of these transformations is applied to one of the coordinates of a bivariate beta random variable, a new bivariate beta random variable will be obtained, but with a modified dependence structure. Thus if (W_1,W_2) has a bivariate beta distribution with positive correlation, then $(W_1, 1-W_2)$ will again have a bivariate beta distribution, but now it will have negative correlation (since $cov(W_1, 1-W_2) = cov(W_1, -W_2) = -cov(W_1,W_2)$). Similarly, if (W_1,W_2) has a bivariate second kind beta distribution, then $(W_1, 1/W_2)$ will again have a bivariate second kind beta distribution, but typically with correlation opposite in sign to that of (W_1,W_2).

The $BB^{(2)}(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7,\alpha_8)$ family of distributions is closed under marginal reciprocation and, likewise, the $BB(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7,\alpha_8)$ family of distributions is closed under marginal reflection about 0. Specifically we have:

If $(W_1,W_2) \sim BB^{(2)}(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7,\alpha_8)$, then

- $(W_1, 1/W_2) \sim BB^{(2)}(\alpha_1,\alpha_4,\alpha_3,\alpha_2,\alpha_7,\alpha_8,\alpha_5,\alpha_6)$,

- $(1/W_1, W_2) \sim BB^{(2)}(\alpha_3, \alpha_2, \alpha_1, \alpha_4, \alpha_8, \alpha_7, \alpha_6, \alpha_5)$,

and

- $(1/W_1, 1/W_2) \sim BB^{(2)}(\alpha_3, \alpha_4, \alpha_1, \alpha_2, \alpha_6, \alpha_5, \alpha_8, \alpha_7)$.

In a parallel fashion, if $(W_1, W_2) \sim BB(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8)$, then

- $(W_1, 1 - W_2) \sim BB(\alpha_1, \alpha_4, \alpha_3, \alpha_2, \alpha_7, \alpha_8, \alpha_5, \alpha_6)$,
- $(1 - W_1, W_2) \sim BB(\alpha_3, \alpha_2, \alpha_1, \alpha_4, \alpha_8, \alpha_7, \alpha_6, \alpha_5)$,

and

- $(1 - W_1, 1 - W_2) \sim BB(\alpha_3, \alpha_4, \alpha_1, \alpha_2, \alpha_6, \alpha_5, \alpha_8, \alpha_7)$.

See Singapurwalla et al. (2016) for further discussion of bivariate beta models related by marginal reflection about 1/2.

5. BIVARIATE GENERALIZED BETA MODELS

If $X \sim B(\alpha_1, \alpha_2)$ then for $\gamma > 0$, $W = X^{1/\gamma}$ is said to have a generalized beta distribution, written

$$W \sim GB(\alpha_1, \alpha_2, \gamma).$$

Similarly, if $X \sim B^{(2)}(\alpha_1, \alpha_2)$ then for $\gamma > 0$, $W = X^{1/\gamma}$ is said to have a generalized second kind beta distribution, written

$$W \sim GB^{(2)}(\alpha_1, \alpha_2, \gamma).$$

Analogous generalizations of our bivariate beta models are defined as follows.

If $(V_1, V_2) \sim BB(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8)$ and if $W_1 = V_1^{1/\gamma_1}$ and $W_2 = V_1^{1/\gamma_2}$ then (W_1, W_2) has a bivariate generalized beta distribution and we write

$$(W_1, W_2) \sim GBB(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8; \gamma_1, \gamma_2).$$

Analogously, if $(V_1, V_2) \sim BB^{(2)}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8)$ and if $W_1 = V_1^{1/\gamma_1}$ and $W_2 = V_1^{1/\gamma_2}$ then (W_1, W_2) has a bivariate generalized second kind beta distribution and we write

$$(W_1, W_2) \sim GBB^{(2)}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8; \gamma_1, \gamma_2).$$

Additional flexibility for the bivariate generalized second kind beta distribution can be achieved by introducing location, scale and rotation parameters. Thus for $\underline{\mu} \in (-\infty, \infty)^2$ and a 2×2 matrix A, we will define (using column vectors)

$$\underline{Z} = \underline{\mu} + A\underline{W}$$

where $\underline{W} \sim GBB^{(2)}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8; \gamma_1, \gamma_2)$.

6. BIVARIATE KUMARASWAMY MODELS

If $X \sim B(1, \beta)$ and $Y = X^{1/\gamma}$, then Y is said to have a Kumaraswamy (1980) distribution, and we write $Y \sim K(\gamma, \beta)$. This distribution is a special case of the generalized beta distribution, but it has one attractive feature. Unlike other generalized beta distributions, the Kumaraswamy distribution has a simple analytic expression available for its distribution function. Thus, if $Y \sim K(\gamma, \beta)$ then

$$F_Y(y) = 1 - (1 - y^\gamma)^\beta I(0 < y < 1).$$

As a consequence, the Kumaraswamy distribution has emerged as a serious competitor to the beta distribution for modeling data taking values in the unit interval. In Arnold and Ghosh (2016), several bivariate Kumaraswamy distributions were discussed in some detail. In this Section we will focus on bivariate Kumaraswamy distributions that can be constructed by marginal power transformations applied to the 8-parameter Arnold-Ng bivariate beta model (3.1)-(3.2), incorporating the needed parametric restrictions to ensure that the marginal distributions of the bivariate beta model have their first parameters equal to 1.

Thus we begin with (V_1, V_2) having the distribution of the form (3.1)-(3.2), but with the following constraints on the α parameters.

$$\alpha_1 + \alpha_5 + \alpha_7 = 1 \tag{6.1}$$

and

$$\alpha_2 + \alpha_5 + \alpha_8 = 1, \tag{6.2}$$

to ensure that $V_1 \sim B(1, \alpha_3 + \alpha_6 + \alpha_8)$ and $V_2 \sim B(1, \alpha_4 + \alpha_6 + \alpha_7)$.

We then define

$$(W_1, W_2) = (V_1^{1/\delta_1}, V_2^{1/\delta_2}),$$

for positive parameters δ_1 and δ_2, to obtain a bivariate Kumaraswamy model, and we write

$$(W_1, W_2) \sim BK(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8; \delta_1, \delta_2).$$

This appears to be a 10-parameter model but, because of the two parametric restrictions (6.1)-(6.2), the parameter space is actually of dimension 8. The parameters of the model, $\alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8; \delta_1$ and δ_2, are constrained as follows:

- $\delta_1, \delta_2 > 0$,
- $\alpha_3, \alpha_4, \alpha_6 \in [0, \infty)$,
- $\alpha_7, \alpha_8 \in [0, 1]$,
- $0 \leq \alpha_5 \leq \min\{1-\alpha_7, 1-\alpha_8\}$,

while $\alpha_1 = 1 - \alpha_5 - \alpha_7$ and $\alpha_2 = 1 - \alpha_5 - \alpha_8$.

As was the case for the bivariate beta and the bivariate second kind beta models discussed in Sections 2 and 3, simplified and more manageable sub-models can be identified by setting some of the α parameters equal to 0. Below we consider in some detail some of these simplified models.

6.1. The Dirichlet bivariate Kumaraswamy model

For this model, we set $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ and, in order to satisfy (6.1)-(6.2), we set $\alpha_7 = \alpha_8 = 1$, while $\alpha_6 \in (0, \infty)$. This results in a three parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_7}{U_6 + U_7 + U_8}\right)^{1/\delta_1},$$

$$W_2 = \left(\frac{U_8}{U_6 + U_7 + U_8}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and U_7, U_8 are i.i.d. $\Gamma(1,1)$ variables, while $U_6 \sim \Gamma(\alpha_6, 1)$ is independent of U_7 and U_8.

Since there is only one α parameter remaining in the model, we may drop the subscript "6" and write

$$\underline{W} \sim \textit{Dirichlet-BK}(\alpha, \delta_1, \delta_2).$$

The corresponding joint density is of the form

$$f_{\underline{W}}(\underline{w}) = \alpha(\alpha+1)\delta_1\delta_2 w_1^{\delta_1-1} w_2^{\delta_2-1}(1 - w_1^{\delta_1} - w_2^{\delta_2})^{\alpha-1}\ I(w_1, w_2 > 0, w_1^{\delta_1} + w_2^{\delta_2} < 1)$$

The marginal densities are, by construction, of the Kumaraswamy type. Thus

$$f_{W_1}(w_1) = (\alpha+1)\delta_1 w_1^{\delta_1-1}\left(1 - w_1^{\delta_1}\right)^{\alpha} \quad I(0 < w_1 < 1).$$

and

$$f_{W_2}(w_2) = (\alpha+1)\delta_2 w_2^{\delta_2-1}\left(1 - w_2^{\delta_2}\right)^{\alpha} \quad I(0 < w_2 < 1).$$

The corresponding conditional densities correspond to scaled Kumaraswamy distributions. Thus, the conditional density of W_2 given $W_1 = w_1$ will be

$$\begin{aligned}
f_{W_2|W_1}(w_2|w_1) &= \alpha\delta_2 \frac{\left(1 - w_1^{\delta_1} - w_2^{\delta_2}\right)^{\alpha-1}}{\left(1 - w_1^{\delta_1}\right)^{\alpha}} \\
&= \frac{\alpha\delta_2}{\left(1 - w_1^{\delta_1}\right)} \left(1 - \frac{w_2^{\delta_2}}{\left(1 - w_1^{\delta_1}\right)}\right)^{\alpha-1} \qquad I(0 < w_2 < (1 - w_1^{\delta_1})^{1/\delta_2}).
\end{aligned}$$

An analogous expression is available for the conditional density of W_1 given $W_2 = w_2$.

Using known results for the Beta and the Dirichlet distribution, we may verify that

$$\begin{aligned}
E(W_1^{\gamma_1}) &= \frac{\Gamma(1 + \gamma_1\delta_1^{-1})\Gamma(2 + \alpha)}{\Gamma(2 + \alpha + \gamma_1\delta_1^{-1})}, \\
E(W_2^{\gamma_2}) &= \frac{\Gamma(1 + \gamma_2\delta_2^{-1})\Gamma(2 + \alpha)}{\Gamma(2 + \alpha + \gamma_2\delta_2^{-1})},
\end{aligned}$$

and

$$E(W_1^{\gamma_1} W_2^{\gamma_2}) = \frac{\Gamma(2 + \alpha)\Gamma(1 + \gamma_1\delta_1^{-1})\Gamma(1 + \gamma_2\delta_2^{-1})}{\Gamma(\alpha + 2 + \gamma_1\delta_1^{-1} + \gamma_2\delta_2^{-1})},$$

from which one can obtain the covariance and correlation (which, for this model, are necessarily non-positive).

By differentiating $\log f_{\underline{W}}(\underline{w})$ it is possible to locate the mode of this joint density. It will be located at the point (w_1^*, w_2^*) where

$$w_1^* = \left\{ \frac{\delta_2(\alpha - 1)}{(\alpha\delta_2 - 1)(\alpha\delta_1 - 1) + (1 - \delta_2)} \right\}^{1/\delta_1}.$$

and

$$w_2^* = \left\{ \frac{\delta_1(\alpha - 1)}{(\alpha\delta_2 - 1)(\alpha\delta_1 - 1) + (1 - \delta_1)} \right\}^{1/\delta_2},$$

provided that this point is an interior point of the support set, i.e., provided that

$$w_1^*, w_2^* > 0, \text{ and } w_1^{*\delta_1} + w_2^{*\delta_2} < 1.$$

In other cases, the mode will occur on the boundary of the support set.

It must be remarked that the restrictive nature of the support of this bivariate Kumaraswamy model will limit its potential for applications.

6.2. The Libby-Novick-Jones-Olkin-Liu bivariate Kumaraswamy model

For this model, we set $\alpha_3 = \alpha_4 = \alpha_5 = \alpha_7 = \alpha_8 = 0$ and, in order to satisfy (6.1)-(6.2), we set $\alpha_1 = \alpha_2 = 1$, while $\alpha_6 \in (0, \infty)$. This results in a three parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_1}{U_1 + U_6}\right)^{1/\delta_1}, \tag{6.3}$$

$$W_2 = \left(\frac{U_2}{U_2 + U_6}\right)^{1/\delta_2}, \tag{6.4}$$

where $\delta_1, \delta_2 > 0$ and U_1, U_2 are i.i.d. $\Gamma(1,1)$ variables, while $U_6 \sim \Gamma(\alpha_6, 1)$ is independent of U_1 and U_2.

Since there is only one α parameter remaining in the model, here too we may drop the subscript "6" and write

$$\underline{W} \sim LNJOL\text{-}BK(\alpha, \delta_1, \delta_2).$$

The corresponding joint density is of the form

$$f_{\underline{W}}(\underline{w}) = \alpha(\alpha+1)\delta_1\delta_2 w_1^{\delta_1 - 1} w_2^{\delta_2 - 1} \frac{(1 - w_1^{\delta_1})^{\alpha}(1 - w_2^{\delta_2})^{\alpha}}{(1 - w_1^{\delta_1} w_2^{\delta_2})^{\alpha+2}} I(0 < w_1, w_2 < 1). \tag{6.5}$$

Since the W_i's can be represented as powers of Beta random variables we can easily get the following expressions for their moments.

$$E(W_i^{\gamma_i}) = \frac{\alpha\Gamma(\frac{\gamma_i}{\delta_i} + 1)}{\Gamma(\frac{\gamma_i}{\delta_i} + \alpha + 1)}, \quad i = 1, 2.$$

A simple expression for $E(W_1 W_2)$ is not available, although it is possible to provide a series expansion for it, and hence for the covariance. As Olkin and Liu (2003) noted in the bivariate beta case (with the δ_i's equal to one) it is possible to verify a strong version of positive dependence for this model. For two points $(w_1, w_2), (w_1', w_2')$ (with $w_1 < w_1', w_2 < w_2'$) it is readily verified that

$$\frac{f_{W_1,W_2}(w_1, w_2) f_{W_1,W_2}(w_1', w_2')}{f_{W_1,W_2}(w_1, w_2') f_{W_1,W_2}(w_1', w_2)} \geq 1,$$

so the joint density is positive likelihood ratio dependent. Consequently the correlation is always positive in this model.

6.3. The Nadarajah-Kotz bivariate Kumaraswamy model of the first kind

For this model, we set $\alpha_1 = \alpha_2 = \alpha_4 = \alpha_7 = \alpha_8 = 0$ and, in order to satisfy (6.1)-(6.2), we set $\alpha_5 = 1$, while $\alpha_3, \alpha_6 \in (0, \infty)$. This results in a four parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_5}{U_3+U_5+U_6}\right)^{1/\delta_1},$$

$$W_2 = \left(\frac{U_5}{U_5+U_6}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_5 \sim \Gamma(1,1)$, $U_3 \sim \Gamma(\alpha_3, 1)$ and $U_6 \sim \Gamma(\alpha_6, 1)$. In this case we write

$$\underline{W} \sim NK(1)\text{-}BK(\alpha_3, \alpha_6, \delta_1, \delta_2).$$

The corresponding joint density is of the form

$$f_{\underline{W}}(\underline{w}) = \alpha_6\delta_1\delta_2 \frac{(w_2^{\delta_2} - w_1^{\delta_1})^{\alpha_3-1}(1-w_2^{\delta_2})^{\alpha_6-1}}{w_1^{1-\delta_1} w_2^{\delta_2(\alpha_3+\alpha_6-1)+1} B(\alpha_6+1, \alpha_3)} \; I(0 < w_1^{\delta_1} < w_2^{\delta_2} < 1)\,.$$

Because of the structure of the $NK(1)$ bivariate beta model, it is possible to obtain expressions for arbitrary mixed moments as follows. For arbitrary $\tau_1, \tau_2 > 0$, we have

$$\begin{aligned} E(W_1^{\tau_1} W_2^{\tau_2}) &= E\left(\left(\frac{U_5}{U_3+U_5+U_6}\right)^{\tau_1/\delta_1}\left(\frac{U_5}{U_5+U_6}\right)^{\tau_2/\delta_2}\right) \\ &= E\left(\left(\frac{U_5}{U_5+U_6}\frac{U_5+U_6}{U_3+U_5+U_6}\right)^{\tau_1/\delta_1}\left(\frac{U_5}{U_5+U_6}\right)^{\tau_2/\delta_2}\right), \end{aligned}$$

where $U_5/(U_5+U_6)$ and $(U_5+U_6)/(U_3+U_5+U_6)$ are independent beta distributed random variables. Thus

$$\begin{aligned} E(W_1^{\gamma_1} W_2^{\gamma_2}) &= E\left(\left(\frac{U_5}{U_5+U_6}\right)^{(\gamma_1/\delta_1)+(\gamma_2/\delta_2)}\right) E\left(\left(\frac{U_5+U_6}{U_3+U_5+U_6}\right)^{\gamma_1/\delta_1}\right) \\ &= \frac{B(1+(\gamma_1/\delta_1)+(\gamma_2/\delta_2), \alpha_6)}{B(1,\alpha_6)} \frac{B(1+\alpha_6+(\gamma_1/\delta_1), \alpha_3)}{B(1+\alpha_6, \alpha_3)}. \end{aligned}$$

From this we may obtain the following expression for the covariance in this model

$$\begin{aligned} Cov(W_1, W_2) &= E(W_1W_2) - E(W_1)E(W_2) \\ &= \left(\frac{B(1+1/\delta_1+1/\delta_2, \alpha_6)}{B(1,\alpha_6)}\right)\left(\frac{B(1+1/\delta_1+\alpha_6, \alpha_3)}{B(1+\alpha_6, \alpha_3)}\right) \\ &\quad - \left(\frac{B(1+1/\delta_1, \alpha_3+\alpha_6)}{B(1, \alpha_3+\alpha_6)}\right)\left(\frac{B(1+1/\delta_2, \alpha_6)}{B(1,\alpha_6)}\right). \end{aligned}$$

In the special case in which $\delta_1 = \delta_2 = 1$, it is possible to verify that this covariance is always non-negative. For other values of the δ's, negative covariance is possible. Sufficient conditions for negative covariance (and hence, correlation) are that

$$\frac{1}{\delta_2} > \max(\alpha_6, \delta_1), \quad \alpha_6 > \alpha_3 \quad \text{and} \quad \alpha_3 + \alpha_6 > \frac{1}{\delta_1} > (\alpha_6 - 1).$$

By differentiating $\log f_{\underline{W}}(\underline{w})$ it is possible to locate the mode of this joint density. It will be located at the point (w_1^*, w_2^*) where

$$w_1^{*\delta_1} = \frac{w_2^{*\delta_2}(\delta_1 - 1)}{(\alpha_3\delta_1 - 1)}$$

and

$$w_2^* = \left\{ \frac{(-1-\alpha_6\delta_2) + \frac{(-1+\delta_1)}{(\alpha_3\delta_1-1)}(1+(-1+\alpha_3+\alpha_6)\delta_2)}{(1+\delta_2) - \frac{(1+\alpha_3\delta_2)(-1+\delta_1)}{(\alpha_3\delta_1-1)}} \right\}^{1/\delta_2},$$

provided that this point is an interior point of the support set, i.e., provided that

$$0 < w_1^{*\delta_1} < w_2^{*\delta_2} < 1.$$

In other cases, the mode will occur on the boundary of the support set.

In this case too, unless $\delta_1 = \delta_2$, the restrictive nature of the support of this bivariate Kumaraswamy model will limit its potential for applications.

6.4. The Nadarajah-Kotz bivariate Kumaraswamy model of the second kind

For this model, we set $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_7 = \alpha_8 = 0$ and, in order to satisfy (6.1)-(6.2), we set $\alpha_5 = 1$, while $\alpha_3, \alpha_6 \in (0, \infty)$. This results in a four parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_5}{U_5 + U_6}\right)^{1/\delta_1},$$

$$W_2 = \left(\frac{U_5}{U_4 + U_5 + U_6}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_5 \sim \Gamma(1,1)$, $U_4 \sim \Gamma(\alpha_4, 1)$ and $U_6 \sim \Gamma(\alpha_6, 1)$. However, this can be recognized as a re-parameterized version of the $NK(1)BK$ distribution, with the subscripts of the W_i's interchanged. It is thus not necessary to list expressions for the joint density, moments, etc., since that material can easily be gleaned from Section 6.3.

6.5. The Olkin-Trikalinos bivariate Kumaraswamy model

For this model, we set $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ and, in order to satisfy (6.1)-(6.2), we set $\alpha_5 \in (0,1)$, while $\alpha_7 = \alpha_8 = 1 - \alpha_5$ and $\alpha_6 \in (0, \infty)$. This results in a four parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_5 + U_7}{U_5 + U_6 + U_7 + U_8}\right)^{1/\delta_1},$$

$$W_2 = \left(\frac{U_5 + U_8}{U_5 + U_6 + U_7 + U_8} \right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_i \sim \Gamma(\alpha_i, 1), \quad i = 5, 6, 7, 8$. In this case we write

$$\underline{W} \sim OT\text{-}BK(\alpha_5, \alpha_6, \delta_1, \delta_2).$$

In this case also, an analytic expression for the joint density is not available, but we can make the following observations about this joint distribution.

Marginal moments are of course Kumaraswamy moments and thus are readily written down. Mixed moments are more troublesome, except in the case when $\delta_1 = \delta_2 = 1$ in which case we reduce to an Olkin-Trikalinos model and the W_i's can be represented as sums of coordinates of a three dimensional Dirichlet variable. For example, in this case as observed by Olkin and Trikalinos, a simple expression for the covariance can be obtained in the following form

$$(6.6) \qquad cov(W_1, W_2) = \frac{(\alpha_5\alpha_6 - \alpha_7\alpha_8)}{(\alpha_5 + \alpha_6 + \alpha_7 + \alpha_8)(\alpha_5 + \alpha_6 + \alpha_7 + \alpha_8 + 1)},$$

when $\delta_1 = \delta_2 = 1$,

Recall that our model is:

$$(W_1, W_2) = \left(\left(\frac{U_5 + U_7}{U_5 + U_6 + U_7 + U_8} \right)^{1/\delta_1}, \left(\frac{U_5 + U_8}{U_5 + U_6 + U_7 + U_8} \right)^{1/\delta_2} \right)$$

where the U_j's are independent with

$$U_5 \sim \Gamma(\alpha_5, 1)\ , \alpha_5 \in (0, 1),\ U_6 \sim \Gamma(\alpha_6, 1)\ , \alpha_6 \in (0, \infty)$$

and

$$U_7 \sim \Gamma(1 - \alpha_5, 1),\ U_8 \sim \Gamma(1 - \alpha_5, 1).$$

To study moments of this distribution, consider the following three dimensional Dirichlet model, which has four positive parameters:

$$(Y_1, Y_2, Y_3) = \left(\frac{U_5}{U_5 + U_6 + U_7 + U_8}, \frac{U_7}{U_5 + U_6 + U_7 + U_8}, \frac{U_8}{U_5 + U_6 + U_7 + U_8} \right)$$

with a $Dirichlet(\alpha_5, 1 - \alpha_5, 1 - \alpha_5, \alpha_6)$ distribution. So we have available expressions for

$$E(Y_1), E(Y_2), E(Y_3), E(Y_1^2), E(Y_2^2), E(Y_3^2), E(Y_1Y_2), E(Y_1Y_3), E(Y_2Y_3)$$

and indeed for

$$E(Y_1^{\tau_1}), E(Y_2^{\tau_2}), E(Y_3^{\tau_3}), E(Y_1^{\tau_1}Y_2^{\tau_2}), E(Y_1^{\tau_1}Y_3^{\tau_3}), E(Y_2^{\tau_2}Y_3^{\tau_3})$$

and for

$$E(Y_1^{\tau_1} Y_2^{\tau_2} Y_1 Y_3^{\tau_3}).$$

But note that

$$(W_1, W_2) = \left((Y_1 + Y_2)^{1/\delta_1}, (Y_1 + Y_3)^{1/\delta_2}\right).$$

In general only a series expansion for $E(W_1^{\nu_1} W_2^{\nu_2})$ will be available. However, in the unlikely case in which $\nu_1/\delta_1 = k_1$, a positive integer and $\nu_2/\delta_2 = k_2$ is also a positive integer then we can write:

$$\begin{aligned} E(W_1^{\nu_1} W_2^{\nu_2}) &= E\left((Y_1 + Y_2)^{\nu_1/\delta_1} (Y_1 + Y_3)^{\nu_2/\delta_2}\right) \\ &= E\left((Y_1 + Y_2)^{k_1} (Y_1 + Y_3)^{k_2}\right) \\ &= \sum_{\ell_1=0}^{k_1} \sum_{\ell_2=0}^{k_2} \binom{k_1}{\ell_1} \binom{k_2}{\ell_2} E(Y_1^{\ell_1+\ell_2} Y_2^{k_1-\ell_1} Y_3^{k_2-\ell_2}), \end{aligned}$$

which is then computable. In particular, if $\delta_1 = \delta_2 = 1$, we get

$$E(W_1 W_2) = E[(Y_1 + Y_2)(Y_1 + Y_3)] = E(Y_1^2) + E(Y_1 Y_2) + E(Y_1 Y_3) + E(Y_2 Y_3)$$

which is easy to evaluate and then subtracting $E(W_1)E(W_2)$ we eventually re-confirm the result in (6.6).

$$\begin{aligned} cov(W_1, W_2) &= \frac{(\alpha_5\alpha_6 - \alpha_7\alpha_8)}{(\alpha_5 + \alpha_6 + \alpha_7 + \alpha_8)(\alpha_5 + \alpha_6 + \alpha_7 + \alpha_8 + 1)} \\ &= \frac{[\alpha_5\alpha_6 - (1 - \alpha_5)^2]}{(\alpha_6 - \alpha_5 + 2)(\alpha_6 - \alpha_5 + 3)} \end{aligned}$$

where we have imposed the constraints $\alpha_7 = \alpha_8 = 1 - \alpha_5$.

When $\delta_1 = \delta_2 = 1$, the model encompasses a full range of values for its covariance and correlation. In particular we have

- The OT-BK model with $\delta_1 = \delta_2 = 1$, will exhibit positive correlation if $\alpha_6 \geq \alpha_5 + 2$, and $\alpha_5 > 1/4$.
- The OT-BK model with $\delta_1 = \delta_2 = 1$, will exhibit negative correlation if $\alpha_6 \leq \alpha_5 - 3$, and $\alpha_5 < 1/4$.

More specifically, with $\delta_1 = \delta_2 = 1$,

- When $\alpha_5 = 0$, $Cov(W_1, W_2) = -\frac{1}{(\alpha_6+2)(\alpha_6+3)} < 0$, for any choice of $\alpha_6 \in (0, \infty)$.
- When $\alpha_5 = 1$, $Cov(W_1, W_2) = \frac{\alpha_6}{(\alpha_6+1)(\alpha_6+2)} > 0$, for any choice of $\alpha_6 \in (0, \infty)$.

In cases in which the δ's are not both equal to 1, the covariances and correlations will have to be evaluated numerically in order to determine when they are positive and when negative.

6.6. The Ghosh bivariate Kumaraswamy model

For this model, suggested by I. Ghosh, we set $\alpha_5 = \alpha_7 = \alpha_8 = 0$ and, in order to satisfy (6.1)-(6.2), we set $\alpha_1 = \alpha_2 = 1$, while $\alpha_3, \alpha_4, \alpha_6 \in (0, \infty)$. This results in a five parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_1}{U_1 + U_3 + U_6}\right)^{1/\delta_1},$$

$$W_2 = \left(\frac{U_2}{U_2 + U_4 + U_6}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_1, U_2 \sim \Gamma(1,1)$ and $U_i \sim \Gamma(\alpha_i, 1), \quad i = 3, 4, 6$. In this case we write

$$\underline{W} \sim G\text{-}BK(\alpha_3, \alpha_4, \alpha_6, \delta_1, \delta_2).$$

In this case also, an analytic expression for the joint density is not available.

6.7. The Magnussen Kumaraswamy model

Magnussen (2004) described a bivariate beta distribution which can be identified as a special case of the Arnold-Ng(8) bivariate beta model, obtained by setting $\alpha_7 = \alpha_8 = 0$. It is thus of the form:

$$\left(\frac{U_1 + U_5}{U_1 + U_3 + U_5 + U_6}, \frac{U_2 + U_5}{U_2 + U_4 + U_5 + U_6}\right).$$

In order to satisfy (6.1)-(6.2), we must have $\alpha_1 + \alpha_5 = 1$ and $\alpha_2 + \alpha_5 = 1$, while $\alpha_3, \alpha_4, \alpha_6 \in (0, \infty)$. This results in a six parameter bivariate Kumaraswamy distribution of the form

$$W_1 = \left(\frac{U_1 + U_5}{U_1 + U_3 + U_5 + U_6}\right)^{1/\delta_1},$$

$$W_2 = \left(\frac{U_2 + U_5}{U_2 + U_4 + U_5 + U_6}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_5 \sim \Gamma(\alpha_5, 1)$ where $\alpha_5 \in [0, 1]$, $U_i \sim \Gamma(1 - \alpha_5, 1), \quad i = 1, 2$ and $U_i \sim \Gamma(\alpha_i, 1), \alpha_i \in (0, \infty), \quad i = 3, 4, 6$. In this case we write

$$\underline{W} \sim M\text{-}BK(\alpha_3, \alpha_4, \alpha_5, \alpha_6, \delta_1, \delta_2).$$

7. VARIATIONS, USING REFLECTION ABOUT 1/2

It is possible to construct other bivariate Kumaraswamy models by applying one or two marginal reflections about the point 1/2 to the bivariate beta model, before imposing the necessary parameter constraints and the marginal power transformations. For example the model (6.3)-(6.4), was derived by first considering a bivariate beta model of the form

$$(V_1, V_2) = \left(\frac{U_1}{U_1 + U_6}, \frac{U_2}{U_2 + U_6}\right).$$

Instead, we can consider starting with the doubly reflected model, $(1-V_1, 1-V_2)$, i.e.,

$$\left(\frac{U_6}{U_1 + U_6}, \frac{U_6}{U_2 + U_6}\right).$$

However, note that, according to our notation of Section 4, U_1 is playing the role of a gamma variable usually denoted by U_3, U_2 is playing the role of a variable usually denoted by U_4, and U_6 would be better labeled U_5. Thus we eventually arrive at the following four parameter bivariate Kumaraswamy model

$$W_1 = \left(\frac{U_5}{U_3 + U_5}\right)^{1/\delta_1}, \quad W_2 = \left(\frac{U_5}{U_4 + U_5}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_5 \sim \Gamma(1,1)$ and $U_i \sim \Gamma(\alpha_i, 1),\ \ i = 3, 4$. If, instead we only reflect V_2 about 1/2, we eventually arrive at the model,

$$W_1 = \left(\frac{U_1}{U_1 + U_8}\right)^{1/\delta_1}, \quad W_2 = \left(\frac{U_8}{U_4 + U_8}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_i \sim \Gamma(1,1)\ \ i = 1, 8$ and $U_4 \sim \Gamma(\alpha_4, 1)$.

Finally, if we only reflect V_1 about 1/2, we eventually arrive at the model,

$$W_1 = \left(\frac{U_7}{U_3 + U_7}\right)^{1/\delta_1}, \quad W_2 = \left(\frac{U_2}{U_2 + U_7}\right)^{1/\delta_2},$$

where $\delta_1, \delta_2 > 0$ and the U_i's are independent gamma variables with $U_i \sim \Gamma(1,1),\ \ i = 2, 7$ and $U_3 \sim \Gamma(\alpha_3, 1)$.

This approach can be applied to each of the bivariate models discussed in this section to obtain three related but distinct models in each case. Recall that such modifications of the models may be useful since reflection of one of the coordinates in the model about 1/2 will typically change the sign of the correlations in the original model.

8. PARAMETER ESTIMATION

The reader will have noticed that many of the models discussed in this paper do not have available analytic expressions for their joint densities. In addition, in many cases, it is difficult to identify functions of (W_1, W_2), say $g(W_1, W_2)$ for which $E(g(W_1, W_2))$ can be evaluated as a tractable function of the parameters of the model. We do have well behaved marginal distributions with available densities and moments, since the coordinate variables have Beta, second kind Beta, generalized Beta or Kumaraswamy distributions. Exceptions to this rule are the Libby-Novick-Jones-Olkin-Liu models for which, at least, the joint density is available, though mixed moments are only available in series form. Having observed this, we recognize that the old standby's maximum likelihood and the method of moments will require some modification if they are to be used to provide estimates of the model parameters. The same can be said for Bayesian estimation since it, also, typically utilizes a likelihood function. Arnold and Ng (2011) described a hybrid estimation strategy for parameter estimation in a 5-parameter sub-model of the BB(1,2,3,4,5,6,7,8) model, namely the BB(1,2,6,7,8) model. Unfortunately, their approach will not work for the associated bivariate Kumaraswamy model. In addition, an approximate Bayesian analysis of the BB(1,2,6,7,8) model was presented in Crackel (2015).

However, all is not lost because, without exception, all of the models discussed in this paper are easy to simulate. This means that, for given values of the parameters, highly accurate approximate values of moments, mixed moments, values of the joint distribution function and values of the joint moment generating function can be obtained. Admittedly, this will result in computer intensive estimation strategies, but it will allow selection among the sub-models for the one best adapted to a given data set. More details on these approximate estimation strategies will be the subject of a subsequent report.

9. A DATA SET

To illustrate the applicability of the bivariate beta and Kumaraswamy models developed in this paper, we consider the following data from the official website of the United Nations Development Program which can be found at (datalink: http://hdr.undp.org/en/composite/trends.) It consists of data on the Human Development Index (HDI) and is provided by the United Nations Development Program (UNDP). Specifically, we look at the 49 countries which are labeled as having very high HDI values for two specific years, the years 2010 and 2014. The reason of choosing these two particular time periods is that 2010 is right after the global financial turmoil (which started during the year 2008) which affected the entire economic sphere and related development and 2014 is the period where most of the countries in Europe were getting out of a recession.

Thus, it is quite interesting to see the change in the HDI values among countries over this period of 4 years.

We consider the following: Let X denote the HDI value for these 49 countries for the year 2010 and Y be the same for the year 2014. Noticeably, all the data points are within the range $(0,1)$, thereby a reasonable approach will be to fit bivariate distributions on the unit square, $[0,1]^2$. At this point we argue that (X,Y) can be modeled well by the bivariate Kumaraswamy and beta distributions developed and discussed in this paper.

1. Model I: The Libby- Novick-Jones-Olkin-Liu bivariate Kumaraswamy distribution. This absolutely continuous distribution has the following density (repeating (6.5)

$$f_{\underline{W}}(\underline{w}) = \alpha(\alpha+1)\delta_1\delta_2 w_1^{\delta_1-1} w_2^{\delta_2-1} \frac{(1-w_1^{\delta_1})^{\alpha}(1-w_2^{\delta_2})^{\alpha}}{(1-w_1^{\delta_1}w_2^{\delta_2})^{\alpha+2}}\; I(0<w_1,w_2<1).$$

2. Model II: The bivariate generalized beta distribution of the first kind [Equation (20) of Sarabia et al. (2014)], with density

$$f(x,y) = \frac{a_1a_2}{B(p_1,p_2,q)} \frac{x^{a_1p_1-1}y^{a_2p_2-1}\,(1-x^{a_1})^{p_2+q-1}\,(1-y^{a_2})^{p_1+q-1}}{(1-x^{a_1}y^{a_2})^{-(p_1+p_2+q)}},$$

for $0<(x,y)<1$, where $B(p_1,p_2,q)$ is the normalizing constant.

3. Model III: The Nadarajah (2007) bivariate generalized beta distribution given by

$$f(x,y) = \frac{Cx^{\alpha-1}y^{\beta-1}\,(1-x)^{\gamma-\alpha-1}\,(1-y)^{\gamma-\beta-1}}{(1-xy\delta)^{\gamma}},$$

for $0<x<1$, $0<y<1$, $\gamma>\alpha>0$, $\gamma>\beta>0$ and $0\le\delta<1$ where C is the normalizing constant given by

$$\frac{1}{C} = \frac{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}{\Gamma^2(\gamma)}\,{}_2F_1\left(\alpha,\beta;\gamma;\delta\right).$$

4. Model IV: Olkin & Liu (2003) bivariate beta distribution given by

$$f(x,y) = \frac{1}{B(\alpha_0,\alpha_1,\alpha_2)} \frac{x^{\alpha_1-1}y^{\alpha_2-1}\,(1-x)^{\alpha_0+\alpha_2-1}\,(1-y)^{\alpha_0+\alpha_1-1}}{(1-xy)^{\alpha_0+\alpha_1+\alpha_2}},$$

for $0<(x,y)<1$, where $B(\alpha_0,\alpha_1,\alpha_2) = \frac{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_0+\alpha_1+\alpha_2)}$.

The other bivariate beta Kumaraswamy models, are not considered in this application because either they do not have a closed form expression for the density, or if they have one, their support set does not match the range of points in the data set.

- The NK- bivariate Kumaraswamy model is not appropriate since it has support $0 < w_1 < w_2 < 1$ (if we were willing to accept the constraint $\delta_1 = \delta_2$). If the δ' s are unequal then the support set is unusual and it is difficult to envision a data set for which such a model will be appropriate.
- For the Dirichlet- bivariate Kumaraswamy model, the situation is similar.

To check the goodness of fit of all four statistical models, a χ^2 goodness-of-fit statistic is used and is computed using the computational package Mathematica. The MLEs are computed using the Nmaximize technique.

Table 1. Parameter estimates for HDI data set.

Model	Model I	Model II	Model III	Model IV
Parameter Estimates	$\hat{\alpha} = 3.5287(0.3335)$ $\hat{\delta_1} = 1.1845(0.9723)$ $\hat{\delta_2} = 3.2424(0.1065)$	$\hat{\alpha_1} = 0.692(0.0894)$ $\hat{\alpha_2} = 1.362(2.246)$ $\hat{p_1} = 3.016(0.9852)$ $\hat{p_2} = 0.782(5.681)$ $\hat{q} = 1.2218(0.3678)$	$\hat{\alpha} = 1.798(0.1142)$ $\hat{\beta} = 1.834(0.2794)$ $\hat{\gamma} = 4.038(0.3677)$ $\hat{\delta} = 0.587(1.2468)$	$\hat{\alpha_0} = 4.016(0.6436)$ $\hat{\alpha_1} = 3.7649(2.1873)$ $\hat{\alpha_2} = 6.172(0.5837)$
Log likelihood	-168.45	-205.38	-217.63	-196.39
χ^2 goodness p-value	0.6132	0.4821	0.4593	0.5041

For this particular data set, it appears that the best model, of the four that were considered, is the Libby- Novick-Jones-Olkin-Liu bivariate Kumaraswamy model.

10. AN ALTERNATIVE APPROACH USING COPULAS

The bivariate Kumaraswamy models discussed in this paper are constructed by focusing on bivariate beta random variables with the first parameter of each marginal beta distribution equal to one. An alternative approach, still using the Arnold-Ng bivariate model, is available.

Many researchers make use of what are called copula based bivariate models. For such models, one begins with a copula, a bivariate distribution with $Uniform(0,1)$ marginals, and makes marginal transformations to obtain a bivariate model with desired marginal distributions. The dependence structure of the resulting model is thus "inherited" from that of the particular copula used in the construction. Typically, one parameter families of copulas are used to build models in this way. More flexible models can be expected if multiparameter families of copulas are used.

A copula based bivariate Kumaraswamy model will be of the form

$$(X_1, X_2) = (\left[1 - (1 - Y_1)^{1/\delta_1}\right]^{1/\gamma_1}, \left[1 - (1 - Y_2)^{1/\delta_2}\right]^{1/\gamma_2}), \tag{10.1}$$

where (Y_1, Y_2) has the desired copula as its distribution (with $Uniform(01)$ marginals).

In (10.1) each X_i has been obtained from the corresponding Y_i by transforming using a Kumaraswamy quantile function, to obtain Kumaraswamy marginals.

Looking back at the Arnold-Ng bivariate beta model (3.1)-(3.2), it is evident that it contains many distributions with $Uniform(0,1)$ marginals since a $Uniform(0,1)$ can be identified as a $Beta(1,1)$ distribution. In fact the Arnold-Ng model contains a four parameter family of such distributions, i.e., of copulas. The subfamily of of the Arnold-Ng distributions that correspond to copulas is obtained by setting

$$\alpha_1 + \alpha_5 + \alpha_7 = 1, \tag{10.2}$$

$$\alpha_2 + \alpha_5 + \alpha_8 = 1, \tag{10.3}$$

$$\alpha_3 + \alpha_6 + \alpha_8 = 1, \tag{10.4}$$

$$\alpha_4 + \alpha_6 + \alpha_7 = 1. \tag{10.5}$$

In addition, recall that all α_i's are non-negative. The resulting four dimensional parameter space may be described as follows:

$$\alpha_5 \in [0,1], \alpha_6 \in [0,1], 0 \leq \alpha_7 \leq 1 - \max\{\alpha_5, \alpha_6\}, 0 \leq \alpha_8 \leq 1 - \max\{\alpha_5, \alpha_6\}.$$

The remaining α_i's, $i = 1, 2, 3, 4$, are then determined by equations (10.2)-(10.5).

Such models will be referred to as Arnold-Ng (henceforth AN) copulas.

In a separate report, Arnold and Ghosh (2016) investigate the use of this multiparameter family of copulas in the construction of eight parameter bivariate Kumaraswamy models. The enhanced flexibility of a four parameter copula model, when compared with typical one parameter families, makes such an approach an attractive alternative. See Arnold and Ghosh (2016) for detailed discussion of all submodels (with one, two or three parameters) of the AN four parameter copula family. These can be used to construct (using (10.1)) five, six and seven parameter bivariate Kumaraswamy distributions.

11. CONCLUDING REMARKS

In this paper we consider several different strategies for constructing bivariate beta (and also bivariate generalized beta) distributions as well as several types of bivariate Kumaraswamy distributions using the gamma based methodology for construction of bivariate beta models as suggested by Arnold and Ng (2011). It has been observed that for most of the constructed bivariate beta models, a corresponding closed form expression for the joint density is unavailable. Our proposed bivariate beta models are significantly different than those discussed and studied in detail in Sarabia et al. (2014).

However, one can readily simulate data from those models using an appropriate algorithm. We have also constructed various bivariate Kumaraswamy models starting from a 8 parameter bivariate Kumaraswamy models by setting the first parameter for the associated beta random variables to 1 and then making a positive power transformation. During the discussion, we have considered some structural properties of the resultant models, such as moments, dependence structure, etc.

However, in many applications it might be desirable to first test the hypothesis $H : \delta_1 = \delta_2 = 1$, using perhaps a generalized likelihood ratio test, before settling on the use of a bivariate Kumaraswamy model as opposed to a bivariate beta or generalized beta model. A preliminary visual inspection of the sample marginals might be a useful first step. Bivariate beta and bivariate Kumaraswamy) distributions could play a useful role in modeling dependent risks (in a typical financial setting) where the individual risks are transformed to be bounded on the interval $[0, 1]$.

Estimation of the model parameters (especially, for those models without a closed form of the density) using an approximate Bayesian approach as well as using an appropriate method of moments strategy (using marginal, joint and/or conditional moments) is currently under investigation and, as noted in Section 8, will be reported elsewhere.

Bivariate Kumaraswamy distributions might be considered as models in certain bivariate reliability contexts. Howver, the absence of corresponding density functions will typically not allow one to identify bivariate failure rate functions and other distributional features of interest in reliability. Numerical evaluations or simulation based approximations will be needed in almost all cases. One case in which a density exists is the Libby- Novick-Jones-Olkin-Liu bivariate Kumaraswamy distribution, displayed in (6.5). In this case, for example, it is possible to obtain a rather complicated series expansion for the reliability quantity $P(W_1 < W_2)$. See Appendix B. Expressions for other reliability features can be expected to be equally or more complicated and, even as in this simple case, will be of doubtful utility.

APPENDIX A

In Figure 1 we provide contour plots for some specific choices of the parameters α and δ_j for $j = 1, 2$ for some representative 3 parameter BK models. The following choices are made for each of these selected representative 3 parameter BK models:

- Choice 1 (c1): $\alpha = 1.2, \delta_1 = 0.5, \delta_2 = 0.5$.
- Choice 2 (c2): $\alpha = 1.8, \delta_1 = 1.3, \delta_2 = 0.9$.

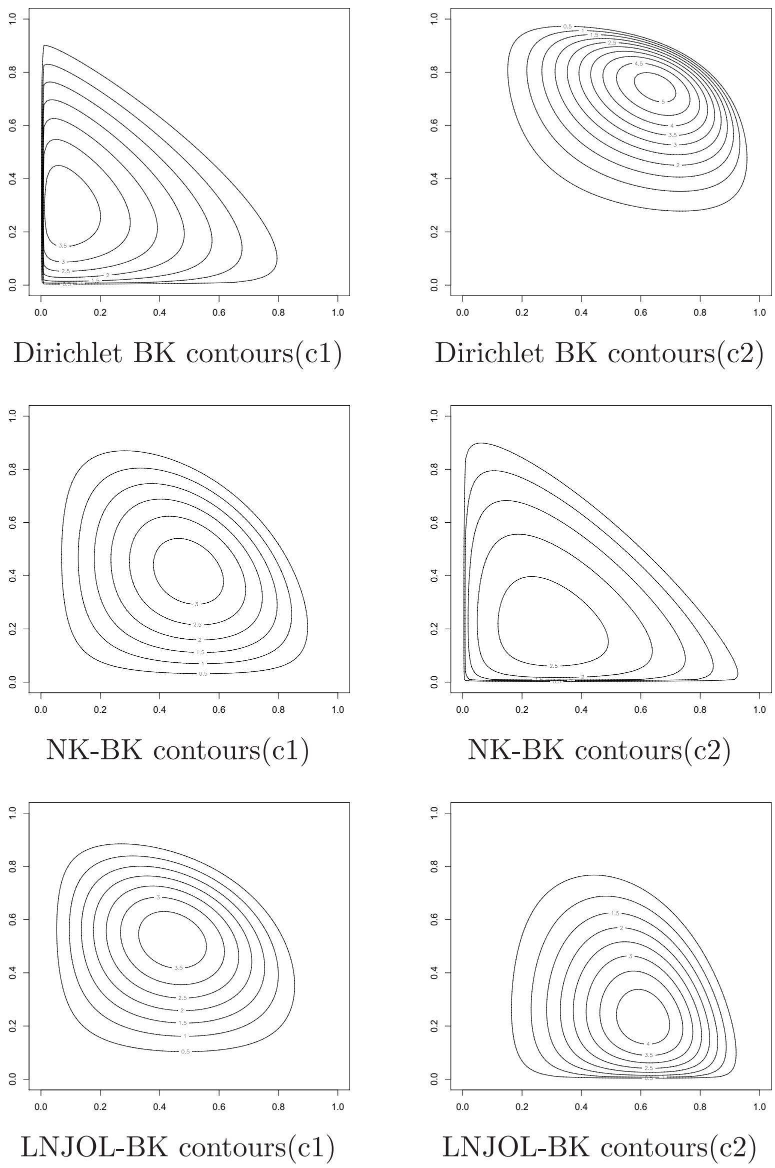


Figure 1: Contour plots for representative BK models.

APPENDIX B

The joint density of the LNJOLBK distribution is of the form

$$f_{\underline{W}}(\underline{w}) = \alpha(\alpha+1)\delta_1\delta_2 w_1^{\delta_1-1} w_2^{\delta_2-1} \frac{(1-w_1^{\delta_1})^{\alpha}(1-w_2^{\delta_2})^{\alpha}}{(1-w_1^{\delta_1}w_2^{\delta_2})^{\alpha+2}} \; I(0 < w_1, w_2 < 1).$$

In this case,

$$R = P(W_1 < W_2) = \int_0^1 \int_{w_1}^1 f(w_1, w_2) dw_2 dw_1.$$

First, let us consider the integral

$$\begin{aligned} I_1 &= \int_{w_1}^{1} \delta_2 w_2^{\delta_2-1} \frac{(1-w_2^{\delta_2})^{\alpha}}{(1-w_1^{\delta_1}w_2^{\delta_2})^{\alpha+2}} dw_2 \\ &= \int_{w_1^{\delta_2}}^{1} (1-t)^{\alpha} \left(1-tw_1^{\delta_1}\right)^{-(\alpha+2)} dt, \quad \text{on substitution} t = w_1^{\delta_1} \\ &= \sum_{k=0}^{\infty} w_1^{k\delta_1} \binom{\alpha+2+k-1}{k} \int_{w_1^{\delta_2}}^{1} t^k (1-t)^{\alpha} dt \end{aligned}$$

using the expansion

$$(1-z)^{-m} = \sum_{k=0}^{\infty} \binom{m+k-1}{k} z^k.$$

Next, consider the integral on I_1

$$\int_{w_1^{\delta_2}}^{1} t^k (1-t)^{\alpha} dt = B(k+1,\alpha-1) - \left(w_1^{\delta_1}\right)^{k+1} \sum_{n=0}^{\infty} \frac{(1-\alpha)_{(n)} w_1^{n\delta_1}}{n!(k+n)},$$

using the series expansion of the incomplete Beta function

$$B(z,a,b) = \int_0^z u^{a-1}(1-u)^{b-1} du = z^a \sum_{n=0}^{\infty} \frac{(1-b)_{(n)} z^n}{n!(a+n)},$$

where $T_{(n)}$ is the descending factorial.

Hence, the expression I_1 reduces to

$$I_1 = \sum_{k=0}^{\infty} \binom{\alpha+2+k-1}{k} B(k+1,\alpha-1) w_1^{k\delta_1+\delta_2} - \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \frac{w_1^{(2k+n+1)\delta_1}(1-\alpha)_{(n)}}{n!(k+n)}.$$

Therefore, the expression of R, the reliability parameter for this bivariate KW model can be expressed in the form

$$\begin{aligned} R &= \int_0^1 \alpha(\alpha+1)\delta_1 w_1^{\delta_1-1} \left(1-w_1^{\delta_1}\right)^{\alpha} I_1 dw_1 \\ &= \alpha(\alpha+1) \left[\sum_{k=0}^{\infty} \binom{\alpha+2+k-1}{k} B(k+1,\alpha-1)\delta_1 \int_0^1 w_1^{\delta_1(1+k)+\delta_2-1} \right. \\ &\qquad\qquad \times \left(1-w_1^{\delta_1}\right)^{\alpha} dw_1 \\ &\quad \left. - \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \frac{(1-\alpha)_{(n)}}{n!(k+n)} \delta_1 \int_0^1 w_1^{\delta_1(2+2k+n)-1} \left(1-w_1^{\delta_1}\right)^{\alpha} dw_1 \right] \\ &= \alpha(\alpha+1) \left[\sum_{k=0}^{\infty} \binom{\alpha+2+k-1}{k} B(k+1,\alpha-1) B(k\delta_1+\delta_2+1,\alpha+1) \right. \\ &\quad \left. - \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \frac{a}{n!(k+n)} B(2k+n+2,\alpha+1) \right], \end{aligned}$$

provided $\alpha > 1$.

REFERENCES

[1] Arnold, B.C.; Castillo, E., and Sarabia, J.M. (1999). *Conditional Specification of Statistical Models*, Springer-Verlag, New York.

[2] Arnold, B.C. and Ghosh, I. (2016). Some alternative bivariate Kumaraswamy models. *To appear in Communications in Statistics-Theory & Methods.*

[3] Arnold, B.C. and Ghosh, I. (2016). Bivariate Kumaraswamy models involving use of Arnold-Ng copulas. *To appear in Journal of Applied Statistical Sciences.*

[4] Barreto-Souza, W. and Lemonte, A.J. (2013). Bivariate Kumaraswamy distribution: properties and a new method to generate bivariate classes. *Statistics*, **47**, 1–22.

[5] Crackel, R.C. (2015). *Likelihood Free Inference for a Flexible Class of Bivariate Beta Distributions.* Thesis (Ph.D.) University of California, Riverside. ISBN: 978-1321-73477-5.

[6] Holland, P.W. and Wang, Y.J. (1987). Dependence function for continuous bivariate densities, *Communications in Statistics-Theory & Methods,* **16**, 863–876.

[7] Jones, M.C. (2002). Multivariate t and beta distributions associated with the multivariate F distribution, *Metrika,* **54**, 215–231.

[8] Kumaraswamy, P. (1980). Generalized probability density-function for double-bounded random-processes, *Journal of Hydrology,* **462**, 79–88.

[9] Libby, D.L., Novick, M.R. (1982). Multivariate generalized beta distributions with applications to utility assessment, *J. Educ. Stat.,* **7**, 271–294.

[10] Magnussen, S. (2004). An algorithm for generating positively correlated Beta-distributed random variables with known marginal distributions and a specified correlation, *Computational Statistics & Data Analysis*, **46**, 397–406.

[11] Nadarajah, S. (2007). A new bivariate beta distribution with application to drought data, *Metron,* **LXV**, 153–174.

[12] Nadarajah, S.; Cordeiro, G.M. and Ortega, E.M.M. (2012). General results for the Kumaraswamy-G distribution. *Journal of Statistical Computation and Simulation*, **12**, 951–979.

[13] Nadarajah, S. and Kotz, S. (2005). Some bivariate beta distributions, *Statistics,* **39**, 457–466.

[14] Olkin, I., Liu, R. (2003). A bivariate beta distribution. *Statistics and Probability Letters*, **62**, 407–412.

[15] Olkin, I., Trikalinos, A.T. (2015). A bivariate beta distribution. *Statistics and Probability Letters*, **96**, 54–60.

[16] Sarabia, J.M., Prieto, F. and Jorda, V. (2014). Bivariate beta-generated distributions with applications to well-being data. *Journal of Statistical Distributions and Applications*, **DOI: 10.1186/2195-5832-1-15.**

[17] Singapurwalla, N.D., Arnold, B.C., Gastwirth, J.L., Gordon, A.S. and Ng, H.K.T. (2016). Adversarial and amiable inference in medical diagnosis, reliability and survival. *International Statistical Review*, **84**, 390–412.

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 251-276

IMPROVED PENALTY STRATEGIES in LINEAR REGRESSION MODELS

Authors: Bahadir Yüzbaşi
– Department of Econometrics,
Inonu University, Turkey
b.yzb@hotmail.com

S. Ejaz Ahmed
– Department of Mathematics and Statistics,
Brock University, Canada
sahmed5@brocku.ca

Mehmet Güngör
– Department of Econometrics,
Inonu University, Turkey
mgungor44@gmail.com

Received: March 2016 Revised: January 2017 Accepted: January 2017

Abstract:

- We suggest pretest and shrinkage ridge estimation strategies for linear regression models. We investigate the asymptotic properties of suggested estimators. Further, a Monte Carlo simulation study is conducted to assess the relative performance of the listed estimators. Also, we numerically compare their performance with Lasso, adaptive Lasso and SCAD strategies. Finally, a real data example is presented to illustrate the usefulness of the suggested methods.

Key-Words:

- *Sub-model; Full Model; Pretest and Shrinkage Estimation; Multicollinearity; Asymptotic and Simulation.*

AMS Subject Classification:

- 62J05, 62J07.

1. INTRODUCTION

Consider a linear regression model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where y_i's are random responses, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ are known vectors, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is a vector denoting unknown coefficients, ε_i's are unobservable random errors and the superscript $\left(^\top\right)$ denotes the transpose of a vector or matrix. Further, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$ has a cumulative distribution function $\mathcal{F}(\boldsymbol{\varepsilon})$; $\mathcal{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathcal{V}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$, where σ^2 is finite and $\boldsymbol{I}_n$ is an identity matrix of dimension $n \times n$. In this paper, we consider that the design matrix has rank p $(p \leq n)$.

It is usually assumed that the explanatory variables are independent of each other in a multiple linear regression model. However, this assumption may not be valid in real life, that is, the independent variables in model may be correlated which cause to multicollinearity problem. In literature, some biased estimations, such as shrinkage estimation, principal components estimation (PCE), ridge estimation, partial least squares (PLS) estimation and Liu-type estimators were proposed to combat this problem. The ridge estimation is proposed by Hoerl and Kennard (1970), and is one of the most effective methods is the most popular one. This estimator has less mean squared error (MSE) than the least squares estimation (LSE) estimation.

The multiple linear regression model is used by data analysts in nearly every field of science and technology as well as economics, econometrics, finance. This model is also used to obtain information about unknown parameters based on sample information and, if available, other relevant information. The other information may be considered as non-sample information (NSI), see Ahmed (2001). This is also known as uncertain prior information (UPI). Such information, which is usually available from previous studies, expert knowledge or researcher's experience, is unrelated to the sample data. The NSI may or may not positively contribute to the estimation procedure. However, it may be advantageous to use the NSI in the estimation process when sample information may be rather limited and may not be completely reliable.

In this study, we consider a linear regression model (1.1) in a more realistic situation when the model is assumed to be sparse. Under this assumption, the vector of coefficients $\boldsymbol{\beta}$ can be partitioned as $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_1$ is the coefficient vector for main effects, and $\boldsymbol{\beta}_2$ is the vector for nuisance effects or insignificant coefficients. We are essentially interested in the estimation of $\boldsymbol{\beta}_1$ when it is reasonable that $\boldsymbol{\beta}_2$ is close to zero. The full model estimation may be subject to high variability and may not be easily interpretable. On the other hand, a sub-model strategy may result with an under-fitted model with large bias. For this reason, we consider pretest and shrinkage strategy to control the magnitude of the bias. Ahmed (2001) gave a detailed definition of shrinkage estimation, and discussed large sample estimation techniques in a regression model. For more recent work on the subject, we refer to Ahmed et al. (2012), Ahmed and Fallahpour (2012), Ahmed (2014a), Ahmed (2014b), Hossain et al. (2016), Gao et al. (2016). Further, for some related work on shrinkage estimation we refer to Prakash and Singh (2009) and Shanubhogue and Al-Mosawi (2010), and others.

In this study, we also consider L_1 type estimators, and compare them with pretest and shrinkage estimators. Yüzbaşı and Ahmed (2015) provided some numerical comparisons of these estimators. The novel aspects of this manuscript, we investigate the asymptotic properties of pretest and shrinkage estimators when the number of observations is larger than the number of covariates.

The paper is organized as following. The full and sub-model estimators based on ridge regression are given in Section 2. The pretest, shrinkage estimators and penalized estimations are also presented in this section. The asymptotic properties of the pretest and shrinkage estimators are given in Section 3. The results of a Monte Carlo simulation study that include a comparison with some penalty estimators are given in Section 4. A real data example is given in Section 5. The concluding remarks are presented in Section 6.

2. ESTIMATION STRATEGIES

The ridge estimator can be obtained from the following model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \;\text{ subject to } \boldsymbol{\beta}^\top \boldsymbol{\beta} \leq \phi,$$

where ϕ is inversely proportional to k, $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$, which is equal to

$$\arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right)^2 + k \sum\nolimits_{j=1}^{p} \beta_j^2 \right\}.$$

It yields

$$\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}} = \left(\mathbf{X}^\top \mathbf{X} + k \boldsymbol{I}_p \right)^{-1} \mathbf{X}^\top \mathbf{y}, \tag{2.1}$$

where $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}}$ is called a ridge full model estimator and $k \in [0, \infty]$ is tuning ridge parameter. If $k = 0$, then $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}}$ is the LSE estimator, and $k = \infty$, then $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}} = \mathbf{0}$. In this study, we select optimal the value of k which minimizes the mean square error of the equation (2.1) via 10-fold cross validation.

We let $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$, where $\mathbf{X_1}$ is an $n \times p_1$ sub-matrix containing the regressors of interest and $\mathbf{X_2}$ is an $n \times p_2$ sub-matrix that may or may not be relevant in the analysis of the main regressors. Similarly, $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top \right)^\top$ be the vector of parameters, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions p_1 and p_2, respectively, with $p_1 + p_2 = p$, $p_i \geq 0$ for $i = 1, 2$.

A sub-model or restricted model is defined as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{subject to } \boldsymbol{\beta}^{\top}\boldsymbol{\beta} \ \leq \phi \text{ and } \boldsymbol{\beta_2} = \mathbf{0},$$

then we have the following restricted linear regression model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \text{ subject to } \boldsymbol{\beta}_1^{\top}\boldsymbol{\beta}_1 \leq \phi. \tag{2.2}$$

We denote $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ as the full model or unrestricted ridge estimator of $\boldsymbol{\beta}_1$ is given by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} = \left(\mathbf{X}_1^{\top}\boldsymbol{M}_2^{R}\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^{\top}\boldsymbol{M}_2^{\mathrm{R}}\mathbf{y},$$

where $\boldsymbol{M}_2^{\mathrm{R}} = \boldsymbol{I}_n - \mathbf{X}_2\left(\mathbf{X}_2^{\top}\mathbf{X}_2 + k\boldsymbol{I}_{p_2}\right)^{-1}\mathbf{X}_2^{\top}$. For model (2.2), the sub-model or restricted estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ of $\boldsymbol{\beta}_1$ has the form

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} = \left(\mathbf{X}_1^{\top}\mathbf{X}_1 + k_1\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^{\top}\mathbf{y},$$

where k_1 is ridge parameter for sub-model estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$.

Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ performs better than $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ when $\boldsymbol{\beta}_2$ is close to zero. However, for $\boldsymbol{\beta}_2$ away from the zero, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ can be inefficient. But, the estimate $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is consistent for departure of $\boldsymbol{\beta}_2$ from zero.

The idea of penalized estimation was introduced by Frank and Friedman (1993). They suggested the notion of bridge regression as follows. For a given penalty function $\pi(\cdot)$ and tuning parameter that controls the amount of shrinkage λ, bridge estimators are estimated by minimizing the following penalized least square criterion

$$\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda\pi(\beta),$$

where $\pi(\beta)$ is $\sum_{j=1}^{p}|\beta_j|^{\gamma}$, $\gamma > 0$. This penalty function bounds the L_γ norm of the parameters.

2.1. Pretest and Shrinkage Ridge Estimation

The pretest is a combination of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ and $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ through an indicator function $I\left(\mathscr{L}_n \leq c_{n,\alpha}\right)$, where $\mathscr{L}_n$ is appropriate test statistic to test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}.$ Moreover, $c_{n,\alpha}$ is an $\alpha-$level critical value using the distribution of $\mathscr{L}_n$. We define test statistics as follows:

$$\mathscr{L}_n = \frac{n}{\widehat{\sigma}^2}\left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{LSE}}\right)^{\top}\mathbf{X}_2^{\top}\boldsymbol{M}_1\mathbf{X}_2\left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{LSE}}\right),$$

where $\widehat{\sigma}^2 = \frac{1}{n-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}})^{\top}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}})$ is consistent estimator of σ^2, $\boldsymbol{M}_1 = \boldsymbol{I}_n - \mathbf{X}_1\left(\mathbf{X}_1^{\top}\mathbf{X}_1\right)^{-1}\mathbf{X}_1^{\top}$ and $\widehat{\boldsymbol{\beta}}_2^{\mathrm{LSE}} = \left(\mathbf{X}_2^{\top}\boldsymbol{M}_1\mathbf{X}_2\right)^{-1}\mathbf{X}_2^{\top}\boldsymbol{M}_1\mathbf{y}$. Under H_0, the test statistic $\mathscr{L}_n$ follows chi-square distribution with p_2 degrees of freedom for large n values. The pretest test ridge regression estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) I\left(\mathscr{L}_n \leq c_{n,\alpha}\right),$$

where $c_{n,\alpha}$ is an $\alpha-$ level critical value.

The shrinkage or Stein-type ridge regression estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ of $\boldsymbol{\beta}_1$ is defined by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} + \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1 - (p_2 - 2)\mathscr{L}_n^{-1}\right),\ p_2 \geq 3.$$

The estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ is general form of the Stein-rule family of estimators, where shrinkage of the base estimator is towards the restricted estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$. The Shrinkage estimator is pulled towards the restricted estimator when the variance of the unrestricted estimator is large. Also, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ is the smooth version of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$.

The positive part of the shrinkage ridge regression estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}$ of $\boldsymbol{\beta}_1$ defined by

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}} = \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} + \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1 - (p_2 - 2)\mathscr{L}_n^{-1}\right)^{+},$$

where $z^{+} = max(0, z)$.

2.1.1. Lasso strategy

For $\gamma = 1$, we obtain the L_1 penalized least squares estimator, which is commonly known as Lasso (least absolute shrinkage and selection operator).

$$\widehat{\boldsymbol{\beta}}^{\mathrm{Lasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}.$$

The parameter $\lambda \geq 0$ controls the amount of shrinkage.

2.1.2. Adaptive Lasso strategy

The adaptive Lasso estimator is defined as

$$\widehat{\boldsymbol{\beta}}^{\mathrm{aLasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \widehat{\xi}_j\, |\beta_j| \right\},$$

where the weight function is

$$\widehat{\xi}_j = \frac{1}{|\beta_j{}^*|^\gamma};\ \gamma > 0.$$

The $\beta_j{}^*$ a root–n consistent estimator of β. For computational details we refer to Zou (2006).

2.1.3. SCAD strategy

The smoothly clipped absolute deviation (SCAD) is proposed by Fan and Li (2001). Given $a > 2$ and $\lambda > 0$, the SCAD penalty at β is

$$J_\lambda(\beta; a) = \begin{cases} \lambda\,|\beta|\,, & |\beta| \le \lambda \\ -\left(\beta^2 - 2a\lambda\,|\lambda| + \lambda^2\right) / \left[2(a-1)\right], & \lambda < |\beta| \le a\lambda \\ (a+1)\lambda^2/2 & |\beta| > a\lambda. \end{cases}$$

Hence, the SCAD estimation is given by

$$\widehat{\boldsymbol{\beta}}^{\text{SCAD}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{p} J_\lambda(\beta_j; a) \right\}.$$

For estimation strategies based on $\gamma = 2$, we establish some useful asymptotic results in the following section.

3. ASYMPTOTIC ANALYSIS

Consider a sequence of local alternatives $\{K_n\}$ given by

$$K_n : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2(n)} = \frac{\boldsymbol{\kappa}}{\sqrt{n}},$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_{p_2})^\top$ is a fixed vector. The asymptotic bias of an estimator $\boldsymbol{\beta}_1^*$ is defined as $\mathcal{B}(\boldsymbol{\beta}_1^*) = \mathcal{E} \lim_{n\to\infty} \{\sqrt{n}\,(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1)\}$, the asymptotic covariance of an estimator $\boldsymbol{\beta}_1^*$ is $\boldsymbol{\Gamma}(\boldsymbol{\beta}_1^*) = \mathcal{E} \lim_{n\to\infty} \left\{ n\,(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1)\,(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1)^\top \right\}$, and by using asymptotic covariance matrix $\boldsymbol{\Gamma}$, the asymptotic risk of an estimator $\boldsymbol{\beta}_1^*$ is given by $\mathcal{R}(\boldsymbol{\beta}_1^*) = \operatorname{tr}(\boldsymbol{W\Gamma})$, where $\boldsymbol{\kappa}$ is a positive definite matrix of weights with dimensions of $p \times p$, and $\boldsymbol{\beta}_1^*$ is one of the suggested estimators.

We consider two regularity conditions as the following to establish the asymptotic properties of the estimators.

(i) $\frac{1}{n} \max\limits_{1\le i\le n} \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i \to 0$ as $n \to \infty$, where $\mathbf{x}_i^\top$ is the ith row of $\mathbf{X}$

(ii) $\lim_{n\to\infty} n^{-1}(\mathbf{X}^\top\mathbf{X}) = \boldsymbol{C}$, for finite $\boldsymbol{C}$.

Theorem 3.1. *When $k \neq \infty$, if $k/\sqrt{n} \to \lambda_0 \geq 0$ and $\boldsymbol{C}$ is non-singular, then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}} - \boldsymbol{\beta}\right) \xrightarrow{d} \left(-\lambda_0\boldsymbol{C}^{-1}\boldsymbol{\beta},\, \sigma^2\boldsymbol{C}^{-1}\right).$$

For proof, see Knight and Fu (2000).

Proposition 3.1. *Assuming above regularity conditions (i) and (ii) hold, then, together with Theorem 3.1, under $\{K_n\}$ as $n \to \infty$ we have*

$$\begin{pmatrix} \vartheta_1 \\ \vartheta_3 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} -\boldsymbol{\mu}_{11.2} \\ \boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \sigma^2\boldsymbol{C}_{11.2}^{-1} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} \end{pmatrix}\right],$$

$$\begin{pmatrix} \vartheta_3 \\ \vartheta_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{\delta} \\ -\boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \sigma^2\boldsymbol{C}_{11}^{-1} \end{pmatrix}\right],$$

where $\vartheta_1 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)$, $\vartheta_2 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} - \boldsymbol{\beta}_1\right)$, $\vartheta_3 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)$, $\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix}$, $\boldsymbol{\gamma} = \boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}$ *and* $\boldsymbol{\delta} = \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega}$, $\boldsymbol{\Phi} = \sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}$, $\boldsymbol{\mu} = -\lambda_0\boldsymbol{C}^{-1}\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ *and* $\boldsymbol{\mu}_{11.2} = \boldsymbol{\mu}_1 - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\left((\boldsymbol{\beta}_2 - \boldsymbol{\kappa}) - \boldsymbol{\mu}_2\right)$.

The expressions for bias for listed estimators are:

Theorem 3.2.

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= -\boldsymbol{\mu}_{11.2}\,, \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= -\boldsymbol{\gamma}\,, \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta} H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right), \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= -\boldsymbol{\mu}_{11.2} - (p_2-2)\boldsymbol{\delta}\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right), \\
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta} H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right), \\
&\quad -(p_2-2)\boldsymbol{\delta}\mathcal{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\, I\left(\chi^2_{p_2+2}(\Delta) > p_2-2\right)\right\},
\end{aligned}$$

where $\Delta = \left(\boldsymbol{\kappa}^\top\boldsymbol{C}_{22.1}^{-1}\boldsymbol{\kappa}\right)\sigma^{-2}$, $\boldsymbol{C}_{22.1} = \boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}$, *and* $H_v(x,\Delta)$ *is the cumulative distribution function of the non-central chi-squared distribution with non-centrality parameter* Δ *and* v *degree of freedom, and*

$$\mathcal{E}\left(\chi_v^{-2j}(\Delta)\right) = \int_0^\infty x^{-2j}\, dH_v(x,\Delta)\,.$$

Proof: See Appendix. ∎

Now, we define the following asymptotic quadratic bias ($\mathcal{QB}$) of an estimator $\boldsymbol{\beta}_1^*$ by converting them into the quadratic form since the bias expression of all the estimators are not in the scalar form.

$$\mathcal{QB}\left(\boldsymbol{\beta}_1^*\right) = \left(\mathcal{B}\left(\boldsymbol{\beta}_1^*\right)\right)^\top \boldsymbol{C}_{11.2}\mathcal{B}\left(\boldsymbol{\beta}_1^*\right),$$

where $\boldsymbol{C}_{11.2} = \boldsymbol{C}_{11} - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}$.

$$\begin{aligned}
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2},\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \boldsymbol{\gamma}^\top \boldsymbol{C}_{11.2}\boldsymbol{\gamma},\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta} H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad + \boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad + \boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta} H^2_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right),\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + (p_2-2)\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\mathcal{E}\left(\chi^{-2}_{p_2+2}\left(\Delta\right)\right)\\
&\quad + (p_2-2)\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}\mathcal{E}\left(\chi^{-2}_{p_2+2}\left(\Delta\right)\right)\\
&\quad + (p_2-2)^2\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\left(\mathcal{E}\left(\chi^{-2}_{p_2+2}\left(\Delta\right)\right)\right)^2,\\
\mathcal{QB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + \left(\boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2} + \boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\right)\\
&\quad \cdot \left[H_{p_2+2}\left(p_2-2;\Delta\right)\right.\\
&\quad \left. + (p_2-2)\mathcal{E}\left\{\chi^{-2}_{p_2+2}\left(\Delta\right) I\left(\chi^{-2}_{p_2+2}\left(\Delta\right) > p_2-2\right)\right\}\right]\\
&\quad + \boldsymbol{\delta}^\top \boldsymbol{C}_{11.2}\boldsymbol{\delta}\left[H_{p_2+2}\left(p_2-2;\Delta\right)\right.\\
&\quad \left. + (p_2-2)\mathcal{E}\left\{\chi^{-2}_{p_2+2}\left(\Delta\right) I\left(\chi^{-2}_{p_2+2}\left(\Delta\right) > p_2-2\right)\right\}\right]^2.
\end{aligned}$$

The $\mathcal{QB}$ of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is $\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}$ and the $\mathcal{QB}$ of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ is an unbounded function of $\boldsymbol{\gamma}^\top \boldsymbol{C}_{11.2}\boldsymbol{\gamma}$. The $\mathcal{QB}$ of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ starts from $\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}$ at $\Delta = 0$, and when Δ increases it increases to the maximum point and then decreases to zero. For the $\mathcal{QB}$s of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ and $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}$, they similarly start from $\boldsymbol{\mu}_{11.2}^\top \boldsymbol{C}_{11.2}\boldsymbol{\mu}_{11.2}$, and increase to a point, and then decrease towards zero.

Theorem 3.3. *Under local alternatives and assumed regularity conditions the risks of the estimators are:*

$$\begin{aligned}
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= \sigma^2 tr\left(\boldsymbol{W}\boldsymbol{C}_{11.2}^{-1}\right) + \boldsymbol{\mu}_{11.2}^\top \boldsymbol{W}\boldsymbol{\mu}_{11.2},\\
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \sigma^2 \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\right) + \boldsymbol{\gamma}^\top \boldsymbol{W}\boldsymbol{\gamma},
\end{aligned}$$

$$
\begin{aligned}
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) - 2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad -\sigma^2\mathrm{tr}(\boldsymbol{W}\boldsymbol{C}_{11.2}^{-1} - \boldsymbol{W}\boldsymbol{C}_{11}^{-1})H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) \\
&\quad +\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left\{2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) - H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right\}, \\
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) + 2(p_2-2)\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -(p_2-2)\sigma^2\mathrm{tr}\left(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\right)\{2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -(p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+2}(\Delta)\right)\} \\
&\quad +(p_2-2)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\{2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right) \\
&\quad -2\mathcal{E}\left(\chi^{-2}_{p_2+4}(\Delta)\right) - (p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+4}(\Delta)\right)\}, \\
\mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \mathcal{R}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) \\
&\quad -2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\delta}\mathcal{E}\left(\left\{1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta)\right\}I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right) \\
&\quad +(p_2-2)\sigma^2\mathrm{tr}\left(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\right) \\
&\quad \cdot\Big[2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right) \\
&\quad -(p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\Big] \\
&\quad -\sigma^2\mathrm{tr}\left(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\right)H_{p_2+2}(p_2-2;\Delta) \\
&\quad +\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\left[2H_{p_2+2}(p_2-2;\Delta) - H_{p_2+4}(p_2-2;\Delta)\right] \\
&\quad -(p_2-2)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\Big[2\mathcal{E}\left(\chi^{-2}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right) \\
&\quad -2\mathcal{E}\left(\chi^{-2}_{p_2+4}(\Delta)I\left(\chi^2_{p_2+4}(\Delta)\le p_2-2\right)\right) \\
&\quad +(p_2-2)\mathcal{E}\left(\chi^{-4}_{p_2+2}(\Delta)I\left(\chi^2_{p_2+2}(\Delta)\le p_2-2\right)\right)\Big].
\end{aligned}
$$

Proof: See Appendix. ∎

Noting that if $\boldsymbol{C}_{12} = \boldsymbol{0}$, then all the risks reduce to common value $\sigma^2\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{C}_{11}^{-1}\right) + \boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{W}\boldsymbol{\mu}_{11.2}$ for all $\boldsymbol{\omega}$. For $\boldsymbol{C}_{12} \neq \boldsymbol{0}$, the risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ remains constant while the risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ is an unbounded function of Δ since $\Delta \in [0,\infty)$. The risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ increases as Δ moves away from zero, achieves it maximum and then decreases towards the risk of the full model estimator. Thus, it is a bounded function of Δ. The risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is smaller than the risk of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ for some small values of Δ and opposite conclusions holds for rest of the parameter space. It can be seen that $\mathcal{R}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}) \le \mathcal{R}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}) \le \mathcal{R}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}})$, strictly inequality holds for small values of Δ. Thus positive shrinkage is superior to the shrinkage estimator. However, both shrinkage estimators outperform the full model estimator in the entire parameter space induced by Δ. On the other hand, the pretest estimator performs better than the shrinkage estimators when Δ takes small values and outside this interval the opposite conclusion holds.

4. SIMULATION STUDIES

In this section, we conduct a Monte Carlo simulation study. The design matrix is generated to be correlated with different magnitudes. We simulate the response from the following model:

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + ... + x_{pi}\beta_p + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$ with $\sigma^2 = 1$. We generate the design matrix $\mathbf{X}$ from a multivariate normal distribution with mean vector $\boldsymbol{\mu} = \mathbf{0}_{p_1}$ and covariance matrix $\boldsymbol{\Sigma}_x$. Further, we consider the off-diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_x$ are equal to be ρ, which is the coefficient of correlation between any two predictors, with $\rho = 0.25, 0.5, 0.75$. The ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\mathbf{X}^\top\mathbf{X}$ is calculated as the condition number test (CNT) which is helpful in detecting the existence of multicollinearity in the design matrix. If the CNT is larger than 30, then the model may have significant multicollinearity, for which we refer to Belsley (1991).

For $H_0 : \beta_j = 0$, $j = p_1 + 1, p_1 + 2, \ldots, p$, with $p = p_1 + p_2$, the regression coefficients are set $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top = \left(\boldsymbol{\beta}_1^\top, \mathbf{0}_{p_2}^\top\right)^\top$ with $\boldsymbol{\beta}_1 = (1, 1, 1, 1)^\top$. In order to investigate the behaviour of the estimators, we define $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_1^\top, \mathbf{0}_{p_2}^\top\right)^\top$ and $\|\cdot\|$ is the Euclidean norm. We considered Δ values between 0 and 4. If $\Delta^* = 0$, then it means that we will have $\boldsymbol{\beta} = (1, 1, 1, 1, \underbrace{0, 0, \ldots, 0}_{p_2})^\top$ to generated the response under null hypothesis. On the other hand, when $\Delta^* > 0$, say $\Delta^* = 2$, we will have $\boldsymbol{\beta} = (1, 1, 1, 1, 2, \underbrace{0, 0, \ldots, 0}_{p_2-1})^\top$ to generated the response under the local alternative hypotheses. When we increase the number of Δ, it indicates the degree of violation of null hypothesis. In our simulation study, we consider the sample size of $n = 60, 100$. Also, the number of predictor variables: $(p_1, p_2) \in \{(4, 4), (4, 8), (4, 16), (4, 32)\}$. Finally, each realization was repeated 1000 times to calculate the MSE of suggested estimators and $\alpha = 0.05$. All computations were conducted using the statistical package R (R Development Core Team, 2010). The performance of one of the suggested estimator was evaluated by using MSE criterion. Also, the relative mean square efficiency (RMSE) of the $\boldsymbol{\beta}_1^\blacktriangle$ to the $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is indicated by

$$\mathrm{RMSE}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} : \boldsymbol{\beta}_1^\blacktriangle\right) = \frac{\mathrm{MSE}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right)}{\mathrm{MSE}\left(\boldsymbol{\beta}_1^\blacktriangle\right)},$$

where $\boldsymbol{\beta}_1^\blacktriangle$ is one of the listed estimators.

For brevity, we report the results for the values of $n = 60, 100$, $p_1 = 4$, $p_2 = 32$ and $\rho = 0.75$ in Table 1, and we plot the simulation results in Figures 1 and 2.

Table 1: RMSE of estimators for $p_1 = 4$, $p_2 = 32$ and $\rho = 0.75$.

Δ	$n = 60$					$n = 100$				
	CNT	**RSM**	**RPT**	**RS**	**RPS**	**CNT**	**RSM**	**RPT**	**RS**	**RPS**
0.000		2.240	1.964	2.015	2.158		1.871	1.737	1.749	1.802
0.200		2.152	2.016	1.910	2.074		1.695	1.546	1.623	1.667
0.400		2.099	1.689	1.918	2.082		1.541	1.283	1.508	1.543
0.600		1.621	1.323	1.615	1.708		1.298	1.058	1.387	1.400
0.800		1.396	1.027	1.554	1.589		1.156	0.955	1.392	1.396
1.000	1798.267	1.193	0.908	1.465	1.500	599.313	0.815	0.925	1.209	1.209
1.250		1.037	0.885	1.410	1.410		0.700	0.962	1.217	1.217
1.500		0.798	0.982	1.352	1.352		0.540	0.993	1.111	1.111
1.750		0.628	0.985	1.238	1.238		0.411	1.000	1.078	1.078
2.000		0.586	0.995	1.227	1.227		0.319	1.000	1.060	1.060
4.000		0.198	1.000	1.058	1.058		0.098	1.000	1.018	1.018

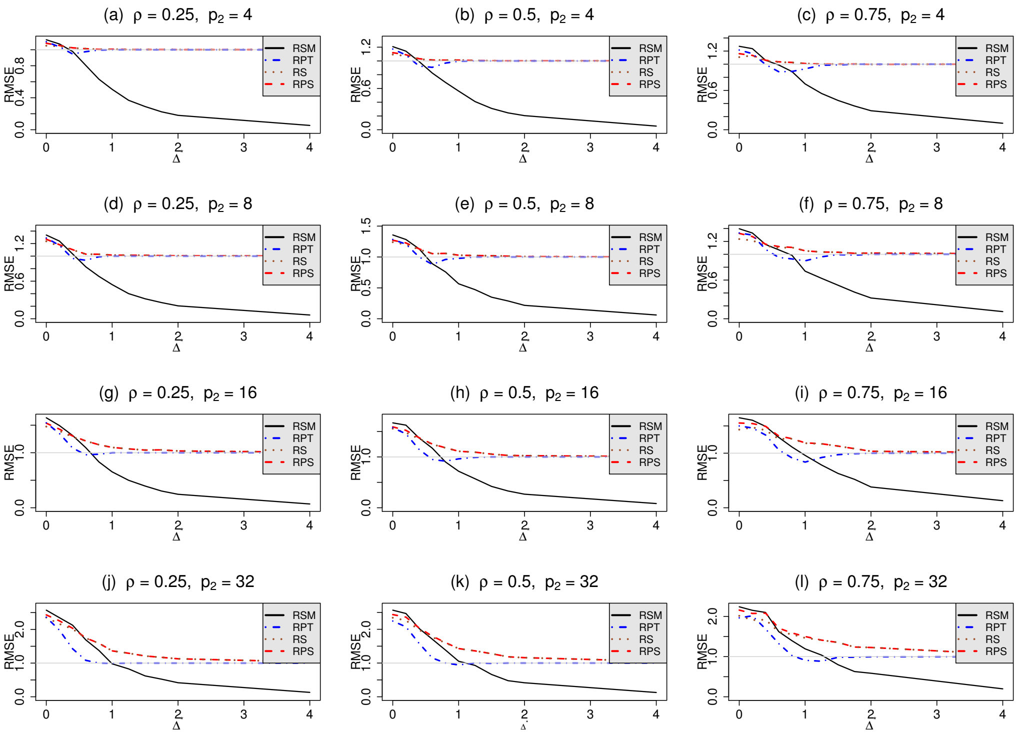


Figure 1: RMSE of the estimators as a function of the non-centrality parameter Δ^* when $n = 60$ and $p_1 = 4$.

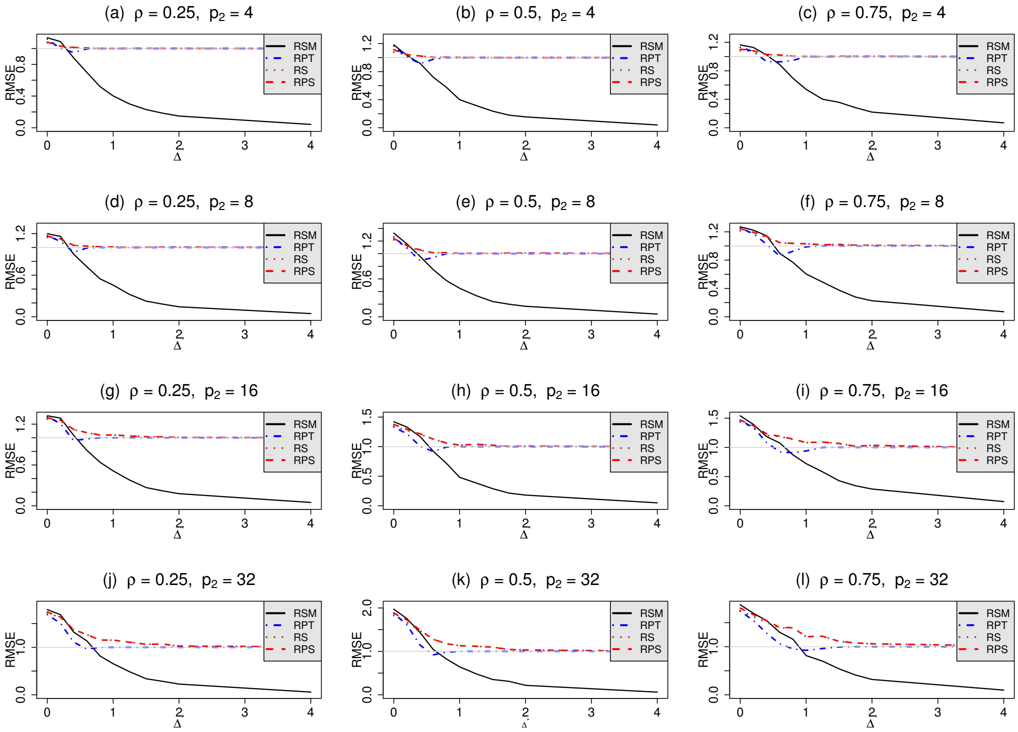


Figure 2: RMSE of the estimators as a function of the non-centrality parameter Δ^* when $n = 100$ and $p_1 = 4$.

The findings can be summarized as follows:

a) When $\Delta^* = 0$, RSM outperforms all listed estimators. In contrast to this, after the small interval near Δ^*, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ decreases and goes to zero.

b) The RPT outperforms RS and RPS in case of $\Delta^* = 0$. However, for large p_2 values while keeping p_1 and n fixed, RPT is less efficient than RPS. When Δ^* is larger than zero, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ decreases, and it remains below 1 for intermediate values of Δ^*, after that the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ increases and approaches one for larger values of Δ^*.

c) Clearly, RPS performs better than RS in the entire parameter space induced by Δ^*. Both shrinkage estimators outshine the full model estimator regardless the correctness of the selected sub-model at hand. This is consistent with the asymptotic theory we presented earlier. Recalling that Δ^* measures the degree of deviation from the assumption on the parameter space, it is clear that one cannot go wrong with the use of shrinkage estimators even if the selected sub-model is not correctly specified. As evident from the table and graphs, if the selected sub-model is correct, that is, $\Delta^* = 0$ then shrinkage estimators are relatively highly efficient than the full model estimator. In other words risk reduction is substantial. On the other hand, the gain slowly diminishes if the sub-model is grossly misspecified. Nevertheless, the shrinkage estimators are at least as good as the full model estimator in terms of risk. Hence, the use of shrinkage estimators make sense in real-life applications when a sub-model cannot be correctly specified, which is the case in most applications.

d) Generally speaking, ridge-type estimators perform better than classical estimator in the presence of multicollinearity among predictors. Our simulation results strongly corroborates to this effect; the RMSE of the ridge-type estimators are increasing function of the amount of multicollinearity.

4.1. Comparison Lasso, aLasso and SCAD

For comparison purposes, we considered $n = 50, 75$, $p_2 = 5, 9, 15, 20$ and $p_1 = 5$ at $\Delta^* = 0$. Here, we used *cv.glmnet* function in *glmnet* package in R for Lasso and aLasso, and *cv.ncvreg* function in *ncvreg* package for SCAD method. The weights for aLasso are obtained from the 10-fold CV Lasso. Results are presented in Table 2.

Table 2: CNT and RMSE of estimators for $p_1 = 5$.

n	ρ	p_2	**CNT**	**LSE**	**RSM**	**RPT**	**RS**	**RPS**	**Lasso**	**aLasso**	**SCAD**
50	0.25	5	10.88	0.90	2.63	2.54	1.61	1.83	1.30	1.46	1.39
		9	20.36	0.85	3.84	3.50	2.34	2.78	1.54	1.92	1.89
		15	46.75	0.76	5.51	4.12	2.85	4.17	1.92	2.56	2.77
		20	90.66	0.65	7.32	5.36	4.24	5.44	2.25	3.18	3.14
	0.5	5	29.39	0.85	2.75	2.39	1.04	1.93	1.22	1.29	1.16
		9	53.39	0.77	4.10	3.32	2.22	2.94	1.46	1.67	1.38
		15	126.44	0.67	5.93	4.69	3.06	4.38	1.76	2.13	1.65
		20	245.60	0.54	8.07	5.77	4.13	5.70	1.99	2.38	1.94
	0.75	5	79.58	0.71	3.93	2.80	1.66	2.12	1.01	0.93	0.73
		9	156.47	0.64	4.75	3.14	2.07	3.01	1.18	1.03	0.73
		15	385.42	0.48	6.50	4.18	2.72	4.41	1.35	1.26	0.83
		20	718.83	0.39	8.94	4.64	3.40	5.91	1.50	1.25	0.83
75	0.25	5	8.90	0.94	2.20	1.97	1.53	1.64	1.25	1.53	1.48
		9	15.12	0.91	3.44	2.96	2.25	2.68	1.60	2.12	2.09
		15	28.11	0.85	5.54	3.26	3.63	3.88	2.05	3.13	2.99
		20	43.77	0.78	7.15	4.11	4.94	5.47	2.63	4.13	3.78
	0.5	5	22.77	0.88	2.59	2.11	1.45	1.79	1.25	1.45	1.19
		9	38.33	0.86	4.03	2.95	2.32	2.73	1.56	1.96	1.59
		15	77.16	0.78	5.79	4.34	3.42	4.35	1.97	2.71	2.45
		20	122.80	0.72	7.30	5.52	4.13	5.50	2.28	3.15	2.75
	0.75	5	65.35	0.80	3.21	2.63	1.33	2.02	1.13	1.11	0.89
		9	113.78	0.76	5.27	3.67	2.30	3.32	1.42	1.52	1.17
		15	225.06	0.66	6.81	4.24	3.80	4.68	1.61	1.71	1.20
		20	359.89	0.57	7.59	5.52	4.11	5.58	1.82	1.91	1.41

Not surprisingly, the performance of the sub-model estimator is the best. The pretest estimator also performs better than other estimators. However, the performance of RPS is better than RPT for larger values of p_2. The performance LSE estimator is worse than listed estimators since the designed matrix is ill-conditioned. The performance of the Lasso, aLasso and SCAD are comparable when ρ is small. On the other hand, pretest and shrinkage estimators remain stable for a given value of ρ. Also, for large values of p_2, the shrinkage and pretest estimators indicate their superiority over L_1 penalty estimators. Thus, we recommend using shrinkage estimators in the presence of multicollinearity.

5. APPLICATION

We use the air pollution and mortality rate data by McDonald and Schwing (1973). This data includes $p = 15$ measurements on mortality rate and explanatory variables, which are air-pollution, socio-economic and meteorological, for $n = 60$ US cities in 1960. The data are freely available from Carnegie Mellon University's StatLib (`http://lib.stat.cmu.edu/datasets/`). In Table 3, we listed variables. Also, the CNT value is calculated as 882.081.574 which implies the existence of multicollinearity in the data set.

Table 3: Lists and Descriptions of Variables

Variables	Descriptions
Dependent Variable	
mort	Total age-adjusted mortality rate per 100.000
Covariates	
Air-Pollution	
prec	Average annual precipitation in inches
jant	Average January temperature in degrees F
jult	Average July temperature in degrees F
humid	Annual average % relative humidity at 1pm
Socio-Economic	
ovr65	% of 1960 SMSA population aged 65 or older
popn	Average household size
educ	Median school years completed by those over 22
hous	% of housing units which are sound & with all facilities
dens	Population per sq. mile in urbanized areas, 1960
nonw	% non-white population in urbanized areas, 1960
wwdrk	% employed in white collar occupations
poor	% of families with income $<$ 3000
Meteorological	
hc	Relative hydrocarbon pollution potential of hydrocarbons
nox	Relative hydrocarbon pollution potential of nitric oxides
so2	Relative hydrocarbon pollution potential of sulphur dioxides

In order to apply the proposed methods, we use two step approach since the prior information is not available here. In the first step, one might do usual variable selection to select the best sub-model. We use the Best Subset Selection (BSS). It showed that prec, jant, jult, educ, dens, nonw, hc and nox are the most important covariates for prediction of the response variable mort and the other variables may be ignored since they are not significantly important. In the second step, we have two model which are the full model with all the covariates and the sub-model with covariates via the BSS. Finally, we construct the shrinkage techniques from the full-model and the sub-model. We fit the full and sub-model which are given in Table 4.

Table 4: Fittings of full and sub-models

Models	**Formulas**
Full model	log(mort) $= \beta_0 + \beta_1$prec $+\beta_2$jant$+\beta_3$jult$+\beta_4$ovr65$+\beta_5$popn$+\beta_6$educ$+\beta_7$hous $+\beta_8$dens$+\beta_9$nonw$+\beta_{10}$wwdrk$+\beta_{11}$poor$+\beta_{12}$hc$+\beta_{13}$nox$+\beta_{14}$so2$+\beta_{15}$humid
Sub-Model	log(mort) $= \beta_0 + \beta_1$prec $+\beta_2$jant$+\beta_3$jult$+\beta_6$educ$+\beta_8$dens$+\beta_9$nonw$+\beta_{12}$hc$+\beta_{13}$nox

To evaluate the performance of the suggested estimators, we calculate the predictive error (PE) of an estimator. Furthermore, we define the relative predictive error (RPE) of $\widehat{\boldsymbol{\beta}}^*$ in terms of the full model ridge regression estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}}$ to easy comparison, is evaluated by as follows

$$\mathrm{RPE}\left(\widehat{\boldsymbol{\beta}}^*\right) = \frac{\mathrm{PE}(\widehat{\boldsymbol{\beta}}^{\mathrm{RFM}})}{\mathrm{PE}(\widehat{\boldsymbol{\beta}}^*)},$$

where $\widehat{\boldsymbol{\beta}}^*$ can be any of the listed estimators. If the RPE is larger than one, it indicates the superior to RFM.

Our results are based on 2500 case resampled bootstrap samples. Since there is no noticeable variation for larger number of replications, we did not consider further values. The average prediction errors were calculated via 10-fold CV for each bootstrap replicate. The predictors were first standardized to have zero mean and unit standard deviation before fitting the model. Figure 3 shows that prediction errors of estimators. As expected, the RSM has the smallest prediction error since the suggested sub-model is correct. Also, the Lasso, aLasso and SCAD have higher prediction error than the suggested techniques.

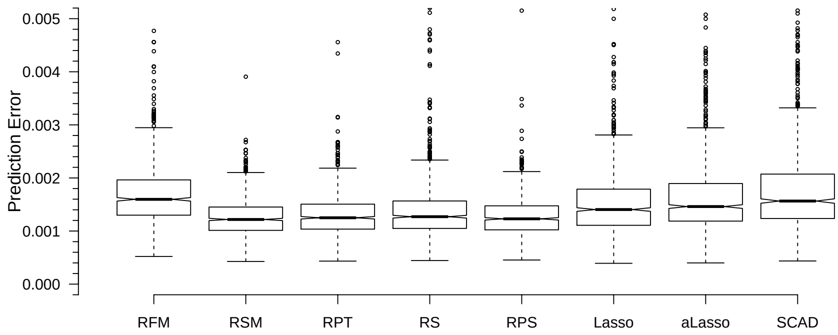


Figure 3: Prediction errors of listed estimators based on bootstrap simulation

Table 5: Estimate (first row) and standard error (second row) for significant coefficients for the air pollution and mortality rate data. The RPE column gives the relative efficiency based on bootstrap simulation with respect to the RFM.

	(Const.)	prec	jant	jult	educ	dens	nonw	hc	nox	RPE
RFM	6.846	0.013	-0.010	-0.002	-0.008	0.010	0.019	-0.007	0.009	1.000
	0.005	0.005	0.006	0.005	0.006	0.004	0.007	0.015	0.017	
RSM	6.845	0.016	-0.018	-0.012	-0.013	0.015	0.042	-0.080	0.081	1.336
	0.005	0.007	0.007	0.006	0.006	0.004	0.007	0.026	0.022	
RPT	6.845	0.016	-0.018	-0.011	-0.013	0.015	0.040	-0.072	0.073	1.288
	0.005	0.007	0.007	0.006	0.007	0.004	0.009	0.029	0.026	
RS	6.845	0.017	-0.017	-0.010	-0.012	0.015	0.038	-0.063	0.065	1.160
	0.005	0.007	0.007	0.007	0.008	0.005	0.012	0.039	0.037	
RPS	6.845	0.016	-0.016	-0.009	-0.012	0.014	0.035	-0.055	0.057	1.316
	0.005	0.006	0.006	0.006	0.007	0.004	0.008	0.025	0.023	
Lasso	6.845	0.019	-0.019	-0.008	-0.012	0.014	0.035	-0.029	0.032	1.060
	0.005	0.009	0.009	0.007	0.011	0.006	0.011	0.046	0.049	
aLasso	6.845	0.022	-0.022	-0.013	-0.012	0.015	0.039	-0.037	0.040	0.965
	0.006	0.010	0.009	0.008	0.012	0.006	0.012	0.050	0.053	
SCAD	6.845	0.019	-0.022	-0.010	-0.014	0.014	0.039	-0.035	0.038	0.897
	0.006	0.012	0.011	0.009	0.013	0.007	0.014	0.052	0.055	

Table 5 reveals that the RPE of the sub-model estimator, pretest, shrinkage and positive part of shrinkage estimators outperform the full model estimator. On the other hand, the sub-model estimator has the highest RPE since it is computed based on the assumption that the selected sub-model is the true model. As expected due to the presence of multicollinearity, the performance of both ridge-type shrinkage and pretest estimators is good and better than estimators based on L_1 criteria. Thus, the data analysis corroborates with our simulation and theoretical findings.

6. CONCLUSIONS

In this study we assessed the performance of least squares, pretest ridge, shrinkage ridge and L_1 estimators when predictors are correlated. We established the asymptotic properties of the pretest ridge and shrinkage ridge estimators. We demonstrated that shrinkage ridge estimators outclass the full model estimator and relatively perform better than sub-model estimator in a wide range of the parameter space. We conducted a Monte Carlo simulation to investigate the behavior of proposed estimators when a selected sub-model may or may not be a true model. Not surprisingly, the sub-model ridge regression estimator outshines all other estimators when the selected sub-model is the true one. However, when this assumption is violated, the performance of the sub-model estimator is profoundly poor. Further, the shrinkage estimators outperform pretest ridge estimators when p_2 is large. Our asymptotic theory is well supported by numerical analysis.

We also analyze the relative performance Lasso, adaptive Lasso and SCAD with other listed estimators. We observe that the performance of pretest and shrinkage ridge regression estimators are superior to L_1 estimators when predictors are highly correlated. The result of a data analysis is very consistent with theoretical and simulated analysis. In conclusion, we suggest to use ridge-type shrinkage estimators when the design matrix is ill-conditioned. The result of this paper are general in nature and consistent with the available results in the reviewed literature. Further, the result of this paper maybe extended to host of models and applications.

APPENDIX

By using $\widetilde{\boldsymbol{y}} = \mathbf{y} - \mathbf{X}_2\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}$

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} &= \underset{\boldsymbol{\beta_1}}{\arg\min}\left\{\|\widetilde{\boldsymbol{y}} - \mathbf{X}_1\boldsymbol{\beta}_1\| + k\,\|\boldsymbol{\beta}_1\|^2\right\} \\
&= \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\widetilde{\boldsymbol{y}} \\
&= \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{y} - \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}} \\
&= \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} - \left(\mathbf{X}_1^\top\mathbf{X}_1 + k\boldsymbol{I}_{p_1}\right)^{-1}\mathbf{X}_1^\top\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}. \qquad (6.1)
\end{aligned}$$

Using the equation (6.1), under local alternative $\{K_n\}$, $\boldsymbol{\Phi}$ is derived as follows:

$$\begin{aligned}
\boldsymbol{\Phi} &= Cov\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) \\
&= \mathcal{E}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)^{\top}\right] \\
&= \mathcal{E}\left[\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)^{\top}\right] \\
&= \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\mathcal{E}\left[\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)^{\top}\right]\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1} \\
&= \sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1} = \sigma^2(\boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1}).
\end{aligned}$$

Lemma 6.1. *Let* $\mathbf{X}$ *be* $q-$*dimensional normal vector distributed as* $\mathcal{N}\left(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_q\right)$, *then, for a measurable function of* φ, *we have*

$$\begin{aligned}
\mathcal{E}\left[\mathbf{X}\varphi\left(\mathbf{X}^{\top}\mathbf{X}\right)\right] &= \boldsymbol{\mu}_x\mathcal{E}\left[\varphi\chi_{q+2}^2\left(\Delta\right)\right] \\
\mathcal{E}\left[\boldsymbol{X}\boldsymbol{X}^{\top}\varphi\left(\mathbf{X}^{\top}\mathbf{X}\right)\right] &= \boldsymbol{\Sigma}_q\mathcal{E}\left[\varphi\chi_{q+2}^2\left(\Delta\right)\right] + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\top}\mathcal{E}\left[\varphi\chi_{q+4}^2\left(\Delta\right)\right],
\end{aligned}$$

where $\chi_v^2\left(\Delta\right)$ *is a non-central chi-square distribution with* v *degrees of freedom and non-centrality parameter* Δ.

Proof: It can be found in Judge and Bock (1978) ∎

Proof of Theorem 3.2: $\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) = -\boldsymbol{\mu}_{11.2}$ is provided by Proposition 3.1, and

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}} - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)\right\} - \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\widehat{\boldsymbol{\beta}}_2^{\mathrm{RFM}}\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2} - \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\omega} = -\left(\boldsymbol{\mu}_{11.2} + \boldsymbol{\delta}\right) = -\boldsymbol{\gamma}.
\end{aligned}$$

Hence, by using Lemma 6.1, it can be written as follows:

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}} - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n \le c_{n,\alpha}\right) - \boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \boldsymbol{\beta}_1\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n \le c_{n,\alpha}\right)\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2} - \boldsymbol{\delta}H_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right).
\end{aligned}$$

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}-\boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}-(p_2-2)\,\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right).
\end{aligned}$$

$$\begin{aligned}
\mathcal{B}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}-\boldsymbol{\beta}_1\right)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1-(p_2-2)\,\mathscr{L}_n^{-1}\right)\right. \\
&\quad \left. I\left(\mathscr{L}_n>p_2-2\right)-\boldsymbol{\beta}_1)\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(1-I\left(\mathscr{L}_n\le p_2-2\right)\right)\right.\right. \\
&\quad \left.\left. -\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}I\left(\mathscr{L}_n>p_2-2\right)-\boldsymbol{\beta}_1\right]\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n\le p_2-2\right)\right\} \\
&\quad -\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\,\mathscr{L}_n^{-1}I\left(\mathscr{L}_n>p_2-2\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}-\boldsymbol{\delta}H_{p_2+2}\left(p_2-2;(\Delta)\right) \\
&\quad -\boldsymbol{\delta}\,(p_2-2)\,\mathcal{E}\left\{\chi_{p_2+2}^{-2}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)>p_2-2\right)\right\}.
\end{aligned}$$

∎

Proof of Theorem 3.3: Firstly, the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}$ is given by

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)^{\top}\right\} \\
&= \mathcal{E}\left(\vartheta_1\vartheta_1^{\top}\right)=Cov\left(\vartheta_1\vartheta_1^{\top}\right)+\mathcal{E}\left(\vartheta_1\right)\mathcal{E}\left(\vartheta_1^{\top}\right)=\sigma^2\boldsymbol{C}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}.
\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}$ is given by

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}-\boldsymbol{\beta}_1\right)^{\top}\right\} \\
&= \mathcal{E}\left(\vartheta_2\vartheta_2^{\top}\right)=Cov\left(\vartheta_2\vartheta_2^{\top}\right)+\mathcal{E}\left(\vartheta_2\right)\mathcal{E}\left(\vartheta_2^{\top}\right)=\sigma^2\boldsymbol{C}_{11}^{-1}+\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top},
\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}$ is given by

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}-\boldsymbol{\beta}_1\right)^{\top}\right\}\\
&= \mathcal{E}\left\{\lim_{n\to\infty} n\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]\right.\\
&\qquad \left.\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]^{\top}\right\}\\
&= \mathcal{E}\left\{\left[\vartheta_1-\vartheta_3 I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]\left[\vartheta_1-\vartheta_3 I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right]^{\top}\right\}\\
&= \mathcal{E}\left\{\vartheta_1\vartheta_1^{\top}-2\vartheta_3\vartheta_1^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)+\vartheta_3\vartheta_3^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}.
\end{aligned}$$

Considering,

$$\begin{aligned}
&\mathcal{E}\left\{\vartheta_3\vartheta_1^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}\\
&= \mathcal{E}\left\{\mathcal{E}\left(\vartheta_3\vartheta_1^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)|\vartheta_3\right)\right\}=\mathcal{E}\left\{\vartheta_3\mathcal{E}\left(\vartheta_1^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)|\vartheta_3\right)\right\}\\
&= \mathcal{E}\left\{\vartheta_3\left[-\mu_{11.2}+\left(\vartheta_3-\boldsymbol{\delta}\right)\right]^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}\\
&= -\mathcal{E}\left\{\vartheta_3\boldsymbol{\mu}_{11.2}^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}+\mathcal{E}\left\{\vartheta_3\left(\vartheta_3-\boldsymbol{\delta}\right)^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}\\
&= -\boldsymbol{\mu}_{11.2}^{\top}\mathcal{E}\left\{\vartheta_3 I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}+\mathcal{E}\left\{\vartheta_3\vartheta_3^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}\\
&\quad -\mathcal{E}\left\{\vartheta_3\boldsymbol{\delta}^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}
\end{aligned}$$

and based on Lemma 6.1, we have

$$\begin{aligned}
&\mathcal{E}\left\{\vartheta_3\vartheta_1^{\top}I\left(\mathscr{L}_n\le c_{n,\alpha}\right)\right\}\\
&= -\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\left\{Cov(\vartheta_3\vartheta_3^{\top})H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\right.\\
&\quad \left.+\mathcal{E}\left(\vartheta_3\right)\mathcal{E}\left(\vartheta_3^{\top}\right)H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)-\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\right\}\\
&= -\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\boldsymbol{\Phi}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)-\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right).
\end{aligned}$$

Then,

$$\begin{aligned}
&\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPT}}\right)\\
&= \boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\sigma^2\boldsymbol{C}_{11.2}^{-1}-\boldsymbol{\Phi}H_{p_2+2}\left(\chi^2_{p_2,\alpha};(\Delta)\right)\\
&\quad -\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)+2\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&= \sigma^2\boldsymbol{C}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\sigma^2(\boldsymbol{C}_{11.2}^{-1}-\boldsymbol{C}_{11}^{-1})H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left[2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)-H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right].
\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}$ is given by

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)^{\top}\right\} \\
&= \mathcal{E}\left\{\lim_{n\to\infty} n\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\mathscr{L}_n^{-1}\right]\right. \\
&\qquad \left.\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)(p_2-2)\mathscr{L}_n^{-1}\right]^{\top}\right\} \\
&= \mathcal{E}\left\{\vartheta_1\vartheta_1^{\top}-2(p_2-2)\vartheta_3\vartheta_1^{\top}\mathscr{L}_n^{-1}+(p_2-2)^2\vartheta_3\vartheta_3^{\top}\mathscr{L}_n^{-2}\right\}.
\end{aligned}$$

Considering,

$$\begin{aligned}
\mathcal{E}\left\{\vartheta_3\vartheta_1^{\top}\mathscr{L}_n^{-1}\right\} &= \mathcal{E}\left\{\mathcal{E}\left(\vartheta_3\vartheta_1^{\top}\mathscr{L}_n^{-1}|\vartheta_3\right)\right\} = \mathcal{E}\left\{\vartheta_3\mathcal{E}\left(\vartheta_1^{\top}\mathscr{L}_n^{-1}|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\left[-\boldsymbol{\mu}_{11.2}+(\vartheta_3-\boldsymbol{\delta})\right]^{\top}\mathscr{L}_n^{-1}\right\} \\
&= -\mathcal{E}\left\{\vartheta_3\boldsymbol{\mu}_{11.2}^{\top}\mathscr{L}_n^{-1}\right\}+\mathcal{E}\left\{\vartheta_3(\vartheta_3-\boldsymbol{\delta})^{\top}\mathscr{L}_n^{-1}\right\} \\
&= -\boldsymbol{\mu}_{11.2}^{\top}\mathcal{E}\left\{\vartheta_3\mathscr{L}_n^{-1}\right\}+\mathcal{E}\left\{\vartheta_3\vartheta_3^{\top}\mathscr{L}_n^{-1}\right\}-\mathcal{E}\left\{\vartheta_3\boldsymbol{\delta}^{\top}\mathscr{L}_n^{-1}\right\}
\end{aligned}$$

by using Lemma 6.1, we have

$$\begin{aligned}
\mathcal{E}\left\{\vartheta_3\vartheta_1^{\top}\mathscr{L}_n^{-1}\right\} &= -\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)+\left\{Cov(\vartheta_3\vartheta_3^{\top})\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\right. \\
&\qquad \left.+\mathcal{E}(\vartheta_3)\,\mathcal{E}\left(\vartheta_3^{\top}\right)\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)-\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+2}\left(\chi_{p_2,\alpha}^{2};\Delta\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)+\boldsymbol{\Phi}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) \\
&\qquad +\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)-\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right).
\end{aligned}$$

Then,

$$\begin{aligned}
&\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) \\
&= \sigma^2\boldsymbol{C}_{11.2}^{-1}+\boldsymbol{\mu}_{11.2}\boldsymbol{\mu}_{11.2}^{\top}+2(p_2-2)\boldsymbol{\mu}_{11.2}^{\top}\boldsymbol{\delta}\mathcal{E}\left(\chi_{p_2+2,\alpha}^{-2}(\Delta)\right) \\
&\quad -(p_2-2)\boldsymbol{\Phi}\left\{2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)-(p_2-2)\mathcal{E}\left(\chi_{p_2+2}^{-4}(\Delta)\right)\right\} \\
&\quad +(p_2-2)\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left\{-2\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)+2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)+(p_2-2)\mathcal{E}\left(\chi_{p_2+4}^{-4}(\Delta)\right)\right\}.
\end{aligned}$$

Finally,

$$\begin{aligned}
\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right) &= \mathcal{E}\left\{\lim_{n\to\infty} n\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}-\boldsymbol{\beta}_1\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}-\boldsymbol{\beta}_1\right)^\top\right\} \\
&= \boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) - 2\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}-\boldsymbol{\beta}_1\right)^\top\right.\right. \\
&\quad \left.\left.\times\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right]\right\} \\
&\quad +\mathcal{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RFM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{RSM}}\right)^\top\right.\right. \\
&\quad \left.\left.\times\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\}^2 I\left(\mathscr{L}_n \le p_2-2\right)\right]\right\} \\
&= \boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) - 2\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +2\mathcal{E}\left\{\vartheta_3\vartheta_3^\top(p_2-2)\mathscr{L}_n^{-1} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad -2\mathcal{E}\left\{\vartheta_3\vartheta_3^\top(p_2-2)^2\mathscr{L}_n^{-2} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +\mathcal{E}\left\{\vartheta_3\vartheta_3^\top I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad -2\mathcal{E}\left\{\vartheta_3\vartheta_3^\top(p_2-2)\mathscr{L}_n^{-1} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +\mathcal{E}\left\{\vartheta_3\vartheta_3^\top(p_2-2)^2\mathscr{L}_n^{-2} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&= \boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right) - 2\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad -\mathcal{E}\left\{\vartheta_3\vartheta_3^\top(p_2-2)^2\mathscr{L}_n^{-2} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&\quad +\mathcal{E}\left\{\vartheta_3\vartheta_3^\top I\left(\mathscr{L}_n \le p_2-2\right)\right\}.
\end{aligned}$$

Considering,

$$\begin{aligned}
&\mathcal{E}\left\{\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&= \mathcal{E}\left\{\mathcal{E}\left(\vartheta_3\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\mathcal{E}\left(\vartheta_1^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)|\vartheta_3\right)\right\} \\
&= \mathcal{E}\left\{\vartheta_3\left[-\boldsymbol{\mu}_{11.2}+(\vartheta_3-\boldsymbol{\delta})\right]^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right\} \\
&= -\boldsymbol{\mu}_{11.2}\mathcal{E}\left(\vartheta_3\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right) \\
&\quad +\mathcal{E}\left(\vartheta_3\vartheta_3^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right) \\
&\quad -\mathcal{E}\left(\vartheta_3\boldsymbol{\delta}^\top\left\{1-(p_2-2)\mathscr{L}_n^{-1}\right\} I\left(\mathscr{L}_n \le p_2-2\right)\right) \\
&= -\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^\top\boldsymbol{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^2(\Delta)\le p_2-2\right)\right) \\
&\quad \boldsymbol{\Phi}\mathcal{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^2(\Delta)\le p_2-2\right)\right) \\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathcal{E}\left(\left\{1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta)\right\} I\left(\chi_{p_2+4}^2(\Delta)\le p_2-2\right)\right) \\
&\quad -\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathcal{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\} I\left(\chi_{p_2+2}^2(\Delta)\le p_2-2\right)\right),
\end{aligned}$$

we have

$$
\begin{aligned}
&\boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RPS}}\right)\\
&= \boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right)+2\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -2\boldsymbol{\Phi}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{-2}(\Delta)\le p_2-2\right)\right)\\
&\quad -2\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+4}^{-2}(\Delta)\right\}I\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +2\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -(p_2-2)^2\,\boldsymbol{\Phi}\mathcal{E}\left(\chi_{p_2+2,\alpha}^{-4}(\Delta)\,I\left(\chi_{p_2+2,\alpha}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -(p_2-2)^2\,\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\mathcal{E}\left(\chi_{p_2+4}^{-4}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +\boldsymbol{\Phi}H_{p_2+2}(p_2-2;\Delta)+\boldsymbol{\delta}\boldsymbol{\delta}^{\top}H_{p_2+4}(p_2-2;\Delta)\\
&= \boldsymbol{\Gamma}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{RS}}\right)+2\boldsymbol{\delta}\boldsymbol{\mu}_{11.2}^{\top}\mathcal{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +(p_2-2)\,\sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\\
&\quad \times\Big[2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -(p_2-2)\,\mathcal{E}\left(\chi_{p_2+2}^{-4}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\Big]\\
&\quad -\sigma^2\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}H_{p_2+2}(p_2-2;\Delta)\\
&\quad +\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left[2H_{p_2+2}(p_2-2;\Delta)-H_{p_2+4}(p_2-2;\Delta)\right]\\
&\quad -(p_2-2)\,\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\Big[2\mathcal{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad -2\mathcal{E}\left(\chi_{p_2+4}^{-2}(\Delta)\,I\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad +(p_2-2)\,\mathcal{E}\left(\chi_{p_2+2}^{-4}(\Delta)\,I\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\Big].
\end{aligned}
$$

∎

Now, the proof of Theorem 3.3 can be easily obtained by following the definition of asymptotic risk.

ACKNOWLEDGMENTS

The authors thank the Associate Editor and two anonymous referees for their valuable comments that improved the paper. Bahadır Yüzbaşı was supported by The Scientific and Research Council of Turkey under grant Tubitak-Bideb-2214/A during this study at Brock University in Canada, and S. Ejaz Ahmed is supported by the Natural Sciences and the Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Ahmed, S.E. (2001). *Shrinkage estimation of regression coefficients from censored data with multiple observations. In Ahmed, S. E. and Reid, N., (Eds.),* Lecture Notes in Statistics, 148, 103–120. Springer-Verlag, New York.

[2] Ahmed, S.E. (2014a). *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*, Springer, New York.

[3] Ahmed, S.E. (Ed.) (2014b). *Perspectives on Big Data Analysis: Methodologies and Applications*, (Vol. 622). American Mathematical Society.

[4] Ahmed, S.E.; Hossain, S. and Doksum, K.A. (2012). Lasso and shrinkage estimation in Weibull censored regression models, *Journal of Statistical Inference and Planning*, **12**, 1273–1284.

[5] Ahmed, S.E. and Fallahpour, S. (2012). Shrinkage Estimation Strategy in Quasi-Likelihood Models, *Statistics & Probability Letters*, **82**, 2170–2179.

[6] Belsley, D.A. (1991). *Conditioning diagnostics*, Encyclopedia of Statistical Sciences. 2.

[7] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**,1348–1360.

[8] Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools, *Technometrics*, **35**,109–148.

[9] Gao, X.; Ahmed, S.E. and Feng, Y. (2016) Post selection shrinkage estimation for high-dimensional data analysis. *Applied Stochastic Models in Business and Industry*, doi: 10.1002/asmb.2193.

[10] Hoerl, A.E. and Kennard, R.W. (1970). Ridge Regression: Biased estimation for non-orthogonal problems, *Technometrics*, **12**, 69–82.

[11] Hossain, S.; Ahmed, S.E.; and Doksum, K.A. (2015). Shrinkage, pretest, and penalty estimators in generalized linear models. *Statistical Methodology*, **24**, 52–68.

[12] Judge, G.G. and Bock, M.E. (1978). *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics*, North Holland, Amsterdam.

[13] Knight, K. and Fu, W. (2000). Asymptotics for Lasso-Type Estimators, *The Annals of Statistics*, **28**(5), 1356–1378.

[14] McDonald, G.C.. and Schwing, R.C. (1973). Instabilities of regression estimates relating air pollution to mortality. *Technometrics*, **15**(3), 463–481.

[15] Prakash, G. and Singh, D.C. (2009). A Bayesian shrinkage approach in Weibull type-II censored data using prior point information, *REVSTAT*, **7**(2), 171–187.

[16] R Development Core Team (2010). *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0.

[17] Shanubhogue, A. and Al-Mosawi, R.R. (2010). On estimation following subset selection from truncated Poisson distributions under Stein loss function, *REVSTAT*, **8**(1), 1–20.

[18] Yüzbaşı, B. and Ahmed, S.E. (2015). *Shrinkage Ridge Regression Estimators in High-Dimensional Linear Models*, In Proceedings of the Ninth International Conference on Management Science and Engineering Management, Springer Berlin Heidelberg, 793 – 807

[19] Zou, H. (2006). The adaptive Lasso and its oracle properties, *Journal of the American Statistical Association*, **101**(456), 1418–1429.

REVSTAT – Statistical Journal
Volume 15, Number 2, April 2017, 277-296

OPEN MARKOV CHAIN SCHEME MODELS FED BY SECOND ORDER STATIONARY AND NON STATIONARY PROCESSES

Authors: Manuel L. Esquível
– Departamento de Matemática, FCT NOVA,
Portugal (mle@fct.unl.pt)

Gracinda R. Guerreiro
– Departamento de Matemática, FCT NOVA,
Portugal (grg@fct.unl.pt)

José M. Fernandes
– Departamento de Ciências e Tecnologia, Universidade de Cabo Verde,
Cape Verde (jose.fernandes@docente.unicv.edu.cv)

Received: April 2016 Revised: October 2016 Accepted: January 2017

Abstract:

- We introduce a schematic formalism for the time evolution of a random open population divided into classes. With a Markov chain model, allowing for population entrances, we consider the flow of incoming members modeled by a time series - either ARIMA for the number of new incomings or SARMA for the residuals of a deterministic sigmoid type trend - and we detail the time series structure of the elements in each class.

  A practical application to real data from a credit portfolio is presented.

Key-Words:

- *Markov chains, Open Markov chain models, Second order processes, ARIMA, SARMA, Credit Risk.*

AMS Subject Classification:

- 60J10, 91D99, 62M10, 91B8.

1. INTRODUCTION

An usual application of a Markov chain model considers a *closed* population with each member being assigned a certain class at each date; the *random* transition of each element among the classes is governed by the transition probabilities. In the homogeneous case - the transition probabilities do not depend on the date at which the transitions occur - and, in the case where there are both transient and recurrent states in the Markov chain, the main emphasis is on the asymptotic behavior. Under that perspective, the transient type events do not matter on long run distributions. In a more realistic model, the population under scrutiny may be changing by the persistent arrival of new members and the events related to the so called "transient" states acquire new significance, as they may persist in time, as the inflow of new elements in the population continues indefinitely.

The consideration of Markov models with a population inflow, the so called *open* Markov models, may be set to start , according to [2], with a work by Gani (see [9]) and was much developed in subsequent years, as perfectly shown in the references mentioned in Bartholomew's work [2, p. 80]. Previously, [24] have obtained a mathematical model to predict distributions of staff and analyse long term impacts on patterns of recruitment and promotions. The case of Poisson recruitment in discrete time open Markov chain model was first dealt in [15], where expressions for the first and second moments of the classes probability distributions were obtained.

There has been remarkable work on the extension of discrete to continuous time Markov and semi-Markov models, such as the developments obtained in [18], [11], [12] and [13]. An important set of contributions to this theme has been detailed in [22] and, in particular, we would like to highlight the works [23], [21], [20], [14] and [16].

The motivation for the present work lays mainly in extending the previous results in the characterization of stable populations lead by discrete time Markov chain transitions: for instance, already in [7] the asymptotic behavior of the classes subpopulation averages is obtained in the case of an exponential input process and a detailed study of stability in terms of relative proportions among classes is also presented; in [14] the asymptotic behavior is described when the Poisson input parameter satisfies general regularity conditions. Having in mind the study not only of the expected values but also of the laws of the subpopulations in the classes, in [4] we considered a sequence of Poisson inflows and studied the probability distributions of the subpopulation classes relying on the fact that, by the *randomized sampling principle* (see [8, p. 268]), the subpopulations are Poisson distributed and independent of each other.

In this work, we consider that the inflow of new population elements is modeled by a time series - to wit, a second order stationary process or stationary with deterministic trend - and we study possible descriptions of the subpopulations, in particular in the transient states, as time flows.

Section 2 introduces the model and some preliminary results and notations that we be the basis of our developments. Section 3 contains the main results obtained in this paper which allowed us to perform an application to credit consumption in Section 4.

2. OPEN MARKOV CHAIN SCHEME MODEL

2.1. The model

Consider a population model driven by a Markov chain defined by a sequence of initial distributions given, for $n \geq 1$, by $(\mathbf{q}^n)^t = (q_1^n, q_2^n, \dots, q_{r_\star}^n)$ and a transition matrix $\mathbf{P} = [p_{ij}], 1 \leq i, j \leq r_\star$. After the first transition, supposing that the initial distribution is performed according to $(\mathbf{q}^1)^t$ at date $n = 1$, the new value of the proportion of the population, for instance, in state 1, is the proportion of those which stay in state 1 plus the proportion of those who come to state 1 from states 2 to $r_\star$. That is:

$$p_{11}q_1^1 + p_{21}q_2^1 + \cdots + p_{r1}q_{r_\star}^1 = \sum_{i=1}^{r_\star} p_{i1}q_i^1$$

and so, the new values of the proportions in all states, after one transition, can be recovered from $\mathbf{P}^\intercal \mathbf{q} = (\mathbf{q}^\intercal \mathbf{P})^\intercal$ and, after n transitions, by $(\mathbf{P}^{(n)})^\intercal \mathbf{q} = (\mathbf{q}^\intercal \mathbf{P}^{(n)})^\intercal$, with $\mathbf{P}^{(0)} = \mathbf{I}$, $\mathbf{P}^{(1)} = \mathbf{P}$ and, by induction, $\mathbf{P}^{(n+1)} = \mathbf{P} \circ \mathbf{P}^{(n)}$. Let us stress, as a notation convention, that all vectors are column vectors.

Now suppose that we want to account for the evolution of the **expected** number of elements in each class supposing that, at each date $k \in \{1, \dots, n\}$, a random number X_k of new elements enters the population. Just after the second cohort enters the population, a first transition occurs in the first cohort driven by the Markov chain law and so on and so forth. Table 1 summarizes this accounting process. Remark that at each step k we distribute multinomially the new random arrivals X_k according to the probability vector $\mathbf{q}^k$ and the elements in each class are redistributed according to the Markov chain transition matrix $\mathbf{P}$.

Table 1: Accounting of n Markov cohorts each with an initial distribution

Date	1	2	…	$n-1$	n
1	$\mathbb{E}[X_1](\mathbf{q}^1)^\intercal$	$\mathbb{E}[X_1](\mathbf{q}^1)^\intercal\mathbf{P}$	…	$\mathbb{E}[X_1](\mathbf{q}^1)^\intercal\mathbf{P}^{(n-2)}$	$\mathbb{E}[X_1](\mathbf{q}^1)^\intercal\mathbf{P}^{(n-1)}$
2	–	$\mathbb{E}[X_2](\mathbf{q}^2)^\intercal$	…	$\mathbb{E}[X_2](\mathbf{q}^2)^\intercal\mathbf{P}^{(n-3)}$	$\mathbb{E}[X_2](\mathbf{q}^2)^\intercal\mathbf{P}^{(n-2)}$
…	…	…	…	…	…
n	–	–	–	–	$\mathbb{E}[X_n](\mathbf{q}^n)^\intercal$

At date n, if we suppose that each new set of customers, a cohort, evolves independently from any one of the already existing sets of customers but, accordingly to the same Markov chain model, we may recover the total **expected** number of elements in each class at date n by computing the sum:

$$\overline{\mathbf{Y}_n} = \sum_{k=1}^{n} \mathbb{E}[X_k](\mathbf{q}^k)^\intercal \, \mathbf{P}^{(n-k)} \ . \tag{2.1}$$

Each vector component corresponds precisely to the **expected** number of elements in each class. This formula - for a constant initial distribution, i.e., $\mathbf{q}^k \equiv \mathbf{q}$ - is well known; see, for a deduction using conditional expectations, [2, p. 52: (3.2)]. In this paper, in order to further study the properties of $(\overline{\mathbf{Y}_n})_{n\geq 1}$, given the properties of a stochastic process $\mathbb{X} = (X_k)_{k\geq 1}$, we will randomize formula (2.1) by considering, instead, for $n \geq 1$:

$$\mathbf{Y}_n = \sum_{k=1}^{n} X_k(\mathbf{q}^k)^\intercal \, \mathbf{P}^{(n-k)} \ . \tag{2.2}$$

Despite the fact that the expressions in (2.1) and (2.2) share the same expected value, i.e, $\overline{\mathbf{Y}_n} = \mathbb{E}[Y_n]$, there is no obvious way to study the probability distribution of the number of elements in each of the population classes, except in the case where the $(X_k)_{k\geq 1}$ new elements are Poisson distributed or independent (see [4]). However, this is not the case for a typical ARMA time series.

Ideally, the most fruitful approach comes from knowing the joint distribution of the entrances $(X_k)_{k\geq 1}$ and of the Markov chain. As this is not the case here, we will call the stochastic process $(\mathbf{Y}_n)_{n\geq 1}$ an **open Markov chain scheme model** for the time evolution of the number of elements in each class. Observe that $\sum_{j=1}^{r_\star} \sum_{i=1}^{r_\star} p_{ij} q_i^j = \sum_{i=1}^{r_\star} \left(\sum_{j=1}^{r_\star} p_{ij} \right) q_i = 1$, if, for instance, the initial distribution does not depend on j; the same being true for the powers of the transition matrix.

For the case of a non-homogeneous Markov chain, the denomination non-homogeneous Markov system was used, in the context of this work, for the first time in [20], according to [19].

We note that some preliminary results on this problem have already been developed in [6].

2.2. Preliminary results and notations

We will introduce now the notions and main results, allowing to give meaning to the Cramer spectral representation theorem (see [3], [17] or [1]).

In the following, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The torus $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ is identified with $[-\pi, +\pi[$ by the map $\lambda \mapsto e^{i\lambda}$.

Definition 2.1. A **centered uncorrelated random field (CURF)** Z on $\mathbb{T}$ is a map from $\mathcal{B}(\mathbb{T})$, the Borel subsets of $\mathbb{T}$, into $L^2((\Omega, \mathcal{A}, \mathbb{P}), (\mathbb{C}, \mathcal{B}(\mathbb{C}))$, the Lebesgue space of (classes) of square integrable random variables taking values in the complex numbers $\mathbb{C}$, such that:

1. Z is centered: $\forall A \in \mathcal{B}(\mathbb{T})\ ,\ \mathbb{E}\left[Z(A)\right] = 0\ .$
2. The images of disjoint Borel sets are uncorrelated, i.e.: $\forall A, B \in \mathcal{B}(\mathbb{T})$: $A \cap B = \emptyset \Rightarrow \left(\mathbb{E}\left[Z(A) \cdot \overline{Z(B)}\right] = 0 \text{ and } Z(A \cup B) = Z(A) + Z(B)\right)$
3. Z is mean-square upper continuous:

$$\forall (A_n)_{n \geq 1}\ :\ A_n \downarrow \emptyset \Rightarrow \lim_{n \to +\infty} Z(A_n) =_{L^2} 0\ .$$

The following result characterizes the structure of a CURF by means of bounded positive measure defined over the Borel sets of the torus.

Theorem 2.1. *A map from $\mathcal{B}(\mathbb{T})$ into the centered random variables of the Lebesgue space $L^2((\Omega, \mathcal{A}, \mathbb{P}), (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ is a centered uncorrelated random field (CURF) if and only if there exists a bounded positive measure μ, named the **basis** of Z such that:*

$$\forall A, B \in \mathcal{B}(\mathbb{T})\ ,\ \mathbb{E}\left[Z(A) \cdot \overline{Z(B)}\right] = \mu(A \cap B)\ .$$

The next result gives sense to the stochastic integral naturally associated to a CURF by means of an isometry between Hilbert spaces of square integrable functions.

Theorem 2.2 (CURF stochastic integral)**.** *Let Z be a CURF on $\mathbb{T}$ with basis μ. There exists an unique isometry $\widetilde{Z}$ from $L^2((\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu))$ into $L^2(((\Omega, \mathcal{A}, \mathbb{P}), (\mathbb{C}, \mathcal{B}(\mathbb{C})))$ such that for all $A \in \mathcal{B}(\mathbb{T})$, $\widetilde{Z}(\mathbb{1}_A) = Z(A)$. We have that:*

1. *$\widetilde{Z}$ is a centered isometry*

$$\forall f \in L^2((\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu))\ ,\ \mathbb{E}\left[\widetilde{Z}(f)\right] = 0\ .$$

2. *The image of $\widetilde{Z}$ is the closure of the vector space spanned by the random variables obtained from Z, that is:*

$$\widetilde{Z}\left(L^2((\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu))\right) = \overline{\mathcal{V}(\{Z(A) : A \in \mathcal{B}(\mathbb{T})\}}\ .$$

Remark 2.1. For each $f \in L^2((\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu))$ we denote the isometry $\widetilde{Z}$ as a **stochastic integral** as follows:

$$\widetilde{Z}(f) = \int_{[-\pi,+\pi[} f(\lambda) dZ(\lambda).$$

Remark 2.2. Moreover, we stress the important result that all the centered isometries between the L^2 spaces mentioned above are generated by a CURF.

Covariances of stochastic processes are nonnegative-definite functions and these, in turn, are represented by positive bounded measures on the torus.

Definition 2.2. A function γ from $\mathbb{Z}$ into $\mathbb{C}$ is **nonnegative-definite** if and only if $\gamma(n) = \overline{\gamma(-n)}$, for $n \in \mathbb{Z}$ and

$$\forall r \geq 1,\ \forall z_1, \ldots, z_r \in \mathbb{C},\ \forall n_1, \ldots, n_r \in \mathbb{Z},\ \sum_{i,j=1}^{r} z_i \overline{z}_j \gamma(n_i - n_j) \geq 0 \ .$$

Theorem 2.3 (Bochner-Herglotz)**.** *A necessary and sufficient condition for a function γ to be nonnegative-definite is that there exists a positive bounded measure on $\mathbb{T}$, which is unique, such that:*

$$\forall n \in \mathbb{Z},\ \gamma(n) = \int_{[-\pi,+\pi[} e^{i\lambda n} d\mu(\lambda) \ .$$

Definition 2.3. A stochastic process $\mathbb{X} = (X_n)_{n\in\mathbb{Z}}$ is **second order stationary** if and only if:

1. All random variables are square integrable, that is:

$$\forall n \in \mathbb{Z} \quad \mathbb{E}\left[|X_n|^2\right] < +\infty \ .$$

2. Both the mean and the covariance functions (sequences) of the process, given, for all $n, m \in \mathbb{Z}$, respectively, by $M(n) := \mathbb{E}[X_n]$ and $\Gamma(m, n) := \mathbb{E}\left[(X_m - \mathbb{E}[X_m])\,\overline{(X_n - \mathbb{E}[X_n])}\right]$, are invariant by time translations, and so:

$$\forall m, m \in \mathbb{Z}\,,\ M(n) = M \in \mathbb{R} \text{ and } \Gamma(n, m) = \gamma(m - n)\ ,$$

for some function γ defined on $\mathbb{Z}$.

Remark 2.3. We may verify that γ is a nonnegative-definite function as defined in Definition 2.2, thus justifying the application of the Bochner-Herglotz theorem to obtain a representation of a second order stationary process.

Example 2.1 (White noise). A process $\mathbb{W} = (W_n)_{n\in\mathbb{Z}}$ is a **white noise** if the random variables are centered, square integrable and, moreover, uncorrelated, that is, if:

$$\forall n, m \in \mathbb{Z} \ : \ n \neq m \Rightarrow \Gamma(n,m) = 0 \ .$$

An example of white noise is given by a sequence of independent centered random variables with common variance.

Example 2.2 (ARMA process). A process $\mathbb{X} = (X_n)_{n\in\mathbb{Z}}$ is an ARMA(p,q) process if there exists a white noise $\mathbb{W} = (W_n)_{n\in\mathbb{Z}}$ and two complex sequences $a_1, a_2, \dots a_p$ and $b_1, b_2, \dots b_q$ such that

$$\sum_{k=0}^{p} a_k X_{n-k} = \sum_{l=0}^{q} b_l W_{n-l}. \tag{2.3}$$

Formula (2.3) is called a **canonical ARMA relation** (see [1, p. 80]) if the polynomials $P(z) = \sum_{k=0}^{p} a_k z^k$ and $Q(z) = \sum_{l=0}^{q} b_l z^l$ have no common factor, P has all his roots with modulus *strictly greater* than 1, Q has all his roots with modulus *greater or equal* than 1 and $P(0) = Q(0) = 1$. It is a remarkable result (see [1, p. 81]), that will prove useful in the following, that, if a stochastic process $\mathbb{X}$ satisfies a canonical ARMA relation with a white noise $\mathbb{W}$ then, this white noise is unique and it is named the **innovation** of $\mathbb{X}$.

We now obtain the representation of a second order stationary stochastic process by the positive bounded measure associated to its covariance.

Definition 2.4 (Spectral measure). Let $\mathbb{X} = (X_n)_{n\in\mathbb{Z}}$ be a second order stationary process. The **spectral measure** of $\mathbb{X}$ is the unique positive bounded measure $\mu_{\mathbb{X}}$ on $\mathbb{T}$ representing the covariance of the process that is, such that,

$$\forall m, n \in \mathbb{Z},\ \Gamma(m,n) = \gamma(m-n) = \int_{[-\pi,+\pi[} e^{i\lambda(m-n)} d\mu_{\mathbb{X}}(\lambda) \ .$$

In vue of future application the particular case of real valued processes deserves special mention.

Remark 2.4. In the case that $\mathbb{X}$ is real valued then the spectral measure $\mu_{\mathbb{X}}$ on $\mathbb{T}$ is invariant by the symmetry ϕ defined on $\mathbb{T}$ by $\phi(z) = \overline{z}$ for all $z \in \mathbb{T}$.

Remark 2.5. If the spectral measure $\mu_{\mathbb{X}}$ is absolutely continuous with respect to the Lebesgue measure on the torus then, by Radon-Nikodym theorem, $\mu_{\mathbb{X}}$ admits a density $f_{\mathbb{X}}$ with respect to the Lebesgue measure and we call this density the **spectral density** of $\mathbb{X}$.

Example 2.3 (White noise). A white noise $\mathbb{W} = (W_n)_{n\in\mathbb{Z}}$ with the random variables having common variance σ^2 has a spectral density given by:

$$f_{\mathbb{W}}(\lambda) = \frac{\sigma^2}{2\pi} \ .$$

Example 2.4 (ARMA process). The spectral density $f_{\mathbb{X}}$ corresponding to the canonical ARMA relation in Example 2.2 is given, using the same notations, by:

$$f_{\mathbb{X}}(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})} \right|^2 .$$

We now state the theorem allowing to represent second order stationary stochastic process as a CURF.

Theorem 2.4 (Cramer theorem). *Let $\mathbb{X} = (X_n)_{n\in\mathbb{Z}}$ be a second order stationary process with spectral measure $\mu_{\mathbb{X}}$. Then, there exists an unique CURF $Z_{\mathbb{X}}$ on $\mathbb{T}$ with basis $\mu_{\mathbb{X}}$ such that:*

$$\forall n \in \mathbb{Z},\ X_n = \int_{[-\pi,+\pi[} e^{i\lambda n} dZ_{\mathbb{X}}(\lambda) .$$

We will need the following observation clarifying the structure of the Cramer representation of a time inverted process.

Remark 2.6 (Time inversion). Let μ^{ϕ} be the image of μ by the symmetry ϕ. As the map $f \mapsto f \circ \phi$ is an isometry from $L^2(\mu^{\phi})$ onto $L^2(\mu)$, then the map

$$f \mapsto \int_{[-\pi,+\pi[} f \circ \phi(\lambda) dZ_{\mathbb{X}}(\lambda)$$

is also an isometry from $L^2(\mu^{\phi})$ into $L^2(\Omega)$. Now, by Remark 2.2 above, there exists an unique CURF $Z_{\mathbb{X}}^{\phi}$ with basis μ^{ϕ}, the **symmetric** CURF of $Z_{\mathbb{X}}$, such that for all $f \in L^2(\mu^{\phi})$:

$$\int_{[-\pi,+\pi[} f(\lambda) dZ_{\mathbb{X}}^{\phi}(\lambda) = \int_{[-\pi,+\pi[} f \circ \phi(\lambda) dZ_{\mathbb{X}}(\lambda) . \tag{2.4}$$

As a consequence, $\mathbb{X}^{\leftarrow} = (X_{-n})_{n\in\mathbb{Z}}$, the time inversion of $\mathbb{X}$, has a spectral representation given by:

$$X_{-n} = \int_{[-\pi,+\pi[} e^{-i\lambda n} dZ_{\mathbb{X}}(\lambda) = \int_{[-\pi,+\pi[} e^{i\lambda n} dZ_{\mathbb{X}}^{\phi}(\lambda) .$$

We will now introduce a special class of processes that will prove useful in the following results.

Definition 2.5 (Evanescent process). A centered stochastic $\mathbb{X} = (X_n)_{n\in\mathbb{Z}}$ process is called an **evanescent process** at time $+\infty$ iff:

$$\lim_{n\to+\infty} \mathbb{E}\left[|X_n|^2\right] = 0 .$$

Remark 2.7. Any linear combination of centered evanescent processes at time $+\infty$ is a centered evanescent process at time $+\infty$.

3. SECOND ORDER FEDS OF A MARKOV CHAIN SCHEME

In this section we considere a Markov chain scheme fed by a stochastic process. Let $\mathbf{P}$ be the transition matrix of the Markov chain. We will suppose that the transition matrix may be written in the following form:

$$\mathbf{P} = \begin{bmatrix} \mathbf{T} & \mathbf{S}_1 \\ \mathbf{0} & \mathbf{R} \end{bmatrix} , \tag{3.1}$$

where $\mathbf{T}$ is the $t_\star \times t_\star$ transition matrix between transient states, $\mathbf{S}_1$ the $t_\star \times (r_\star - t_\star)$ matrix of one step transitions between the transient and the recurrent states and $\mathbf{R}$ the $(r_\star - t_\star) \times (r_\star - t_\star)$ transition matrix between the recurrent states. A straitghforward computation shows that:

$$\mathbf{P}^{(n)} = \begin{bmatrix} \mathbf{T}^{(n)} & \mathbf{S}_n \\ \mathbf{0} & \mathbf{R}^{(n)} \end{bmatrix} , \; n > 1$$

with $\mathbf{S}_n = \mathbf{S}_{n-1}\mathbf{R} + \mathbf{T}^{(n-1)}\mathbf{S}_1 = \sum_{i=0}^{n-1} \mathbf{T}^{(i)}\mathbf{S}_1\mathbf{R}^{(n-1-i)}$.

We now write the successive cohorts vectors of classifications for new arriving elements, at time period k, as

$$(\mathbf{q}^k)^\intercal = \left[(\mathbf{t}^k)^\intercal \,\middle|\, (\mathbf{r}^k)^\intercal \right] , \tag{3.2}$$

with $(\mathbf{t}^k)^\intercal$ the vector of the initial classification probabilities for the transient states and $(\mathbf{r}^k)^\intercal$ the vector of the initial classification probabilities for the recurrent states. Using (3.1) and (3.2), formula (2.2) may be written as

$$\mathbf{Y}_n = \left[\mathbf{Y}_n^1 \mid \mathbf{Y}_n^2\right] = \left[\sum_{k=1}^{n} X_k (\mathbf{t}^k)^\intercal \mathbf{T}^{(n-k)} \,\middle|\, \sum_{k=1}^{n} X_k \left((\mathbf{t}^k)^\intercal \mathbf{S}_{n-k} + (\mathbf{r}^k)^\intercal \mathbf{R}^{(n-k)} \right) \right] . \tag{3.3}$$

Formula (3.3) allow us to estimate the number of elements in each subpopulation (transient or recurrent). However, for the reasons pointed in the introduction and for technical reasons that will become apparent in the following, we will consider only the transient states part of the transition matrix.

At first, we will suppose that the feeding process is stationary. The main result will be the following:

Theorem 3.1. *Consider an open Markov chain scheme model, with a diagonizable matrix* $\mathbf{P}$*, written as in* (3.1)*, and a constant vector of initial classification probabilities* $(\mathbf{q}^k)^\intercal \equiv (\mathbf{q})^\intercal$*, defined as in* (3.2)*.*

If the open Markov chain scheme model is fed by a real valued ARMA process then the population in each of the transient states may be described as a sum of an ARMA processes with an evanescent process.

Proof: Suppose that $\mathbb{X} = (X_k)_{k\in\mathbb{Z}}$ is a second order stationary time series. Recall that by the Cramer representation theorem (see [1, p. 51]) also stated above, we have that, for all $k \in \mathbb{Z}$,

$$X_k = \int_{[-\pi,+\pi[} e^{i\lambda k} dZ_{\mathbb{X}}(\lambda) \ ,$$

with $Z_{\mathbb{X}}$ the spectral field of $\mathbb{X}$, the unique CURF associated to $\mathbb{X}$ (see [1, p. 38] for a definition). Reporting this representation in the Markov chain scheme given by formula (2.2) we get that

$$\mathbf{Y}_n = \sum_{k=1}^{n} \left(\int_{[-\pi,+\pi[} e^{i\lambda k} dZ_{\mathbb{X}}(\lambda) \right) (\mathbf{q}^k)^\intercal \, \mathbf{P}^{(n-k)} \ . \tag{3.4}$$

Considering that the transition matrix of the transient states $\mathbf{T}$ is diagonalizable, it may be written as:

$$\mathbf{T} = \sum_{j=1}^{t_\star} \eta_j \boldsymbol{\alpha}_j \boldsymbol{\beta}_j^\intercal \ ,$$

with $(\eta_j)_{j\in\{1,\dots,t_\star\}}$ the eigenvalues, $(\boldsymbol{\alpha}_j)_{j\in\{1,\dots,t_\star\}}$ the left eigenvectors and with $(\boldsymbol{\beta}_j)_{j\in\{1,\dots,t_\star\}}$ the right eigenvectors of $\mathbf{T}$ (see [8] or [10]).We observe that $j \in \{1,\dots,t_\star\}$ corresponds to a transient state if and only if $\mid \eta_j \mid< 1$. Considering also that, for $k \geq 1$, we have $\mathbf{t}^k \equiv \mathbf{t}$, we will have, for $n \geq 1$,

$$\begin{aligned} \mathbf{Y}_n^1 &= \sum_{j=1}^{t_\star} \left(\int_{[-\pi,+\pi[} \left(\sum_{k=1}^{n} e^{i\lambda k} \, \eta_j^{(n-k)} \right) dZ_{\mathbb{X}}(\lambda) \right) \mathbf{t}^\intercal \, \boldsymbol{\alpha}_j \boldsymbol{\beta}_j^\intercal = \\ &= \sum_{j=1}^{t_\star} \left(\int_{[-\pi,+\pi[} e^{-i\lambda n} \left[\frac{1 - (e^{i\lambda} \, \eta_j)^{n+1}}{1 - e^{i\lambda} \, \eta_j} \right] dZ_{\mathbb{X}}(\lambda) \right) \mathbf{t}^\intercal \, \boldsymbol{\alpha}_j \boldsymbol{\beta}_j^\intercal \ . \end{aligned} \tag{3.5}$$

We now define, for each $j \in \{1,\dots,s\}$ and $n \geq 1$, $W_n^j = W_n^{1,j} - W_n^{2,j}$ with

$$W_n^{1,j} := \int_{[-\pi,+\pi[} e^{-i\lambda n} \left[\frac{1}{1 - e^{i\lambda}\eta_j} \right] dZ_{\mathbb{X}}(\lambda) \ , W_n^{2,j} := \int_{[-\pi,+\pi[} \left[\frac{e^{i\lambda} \, \eta_j^{n+1}}{1 - e^{i\lambda}\eta_j} \right] dZ_{\mathbb{X}}(\lambda)$$

and we observe that $h^{1,j}(\lambda) := \frac{1}{1-e^{i\lambda}\eta_j}$ and $h^{2,j}(\lambda) := \frac{e^{i\lambda} \, \eta_j^{n+1}}{1-e^{i\lambda}\eta_j}$ are both $L^2([-\pi,+\pi])$ functions due to $|\eta_j| < 1$. We will deal separately with these two components.

Firstly, we show that $\mathbb{W}^{2,j} = (W_n^{2,j})_{n\geq 1}$ is an *evanescent process* at $+\infty$, according to Definition 2.5. In fact, with $\mu_{\mathbb{X}}$ the spectral measure of $\mathbb{X}$, we have that:

$$\mathbb{E}\left[\left| W_n^{2,j} \right|^2 \right] = \int_{[-\pi,+\pi[} \left| \frac{e^{i\lambda} \, \eta_j^{n+1}}{1 - e^{i\lambda}\eta_j} \right|^2 d\mu_{\mathbb{X}}(\lambda) \leq \frac{|\eta_j|^{2n+2}}{|1 - |\eta_j||^2} \, \mu_{\mathbb{X}}\left([-\pi,+\pi[\right) \ ,$$

and so, as $\mu_{\mathbb{X}}$ is bounded and $|\eta_j| < 1$, we have, with exponential rate given by $|\eta_j|^{2n+2}$,

$$\lim_{n\to+\infty} \mathbb{E}\left[\left|W_n^{2,j}\right|^2\right] = 0\,.$$

In fact, due to the exponential convergence to zero of the second order moments, the convergence of the process $\mathbb{W}^{2,j}$ to zero is in the almost sure sense. Let $0 < \epsilon < 1$, then, as,

$$\mathbb{P}\left[\left|W_n^{2,j}\right| > |\eta_j|^{\epsilon n}\right] \leq \frac{\mathbb{E}\left[\left|W_n^{2,j}\right|^2\right]}{|\eta_j|^{2\epsilon n}} \leq \frac{|\eta_j|^{2n+2}}{|\eta_j|^{2\epsilon n}\left|1-|\eta_j|\right|^2}\mu_{\mathbb{X}}\left([-\pi,\pi[\right) =$$
$$= |\eta_j|^{2n(1-\epsilon)}\frac{|\eta_j|^2}{\left|1-|\eta_j|\right|^2}\mu_{\mathbb{X}}\left([-\pi,\pi[\right)\ ,$$

we have that, for some constant c,

$$\sum_{n=1}^{+\infty}\mathbb{P}\left[\left|W_n^{2,j}\right| > |\eta_j|^{\epsilon n}\right] < \sum_{n=1}^{+\infty}|\eta_j|^{2n(1-\epsilon)} < +\infty\ ,$$

with $\lim_{n\to+\infty}|\eta_j|^{2n(1-\epsilon)} = 0$, thus showing (see [3, p. 370]) the almost sure convergence to zero at $+\infty$ of the process $\mathbb{W}^{2,j}$.

Secondly, we have that $\mathbb{W}^{1,j} = (W_n^{1,j})_{n\geq 1}$ defines a second-order stationary stochastic process obtained from $\mathbb{X}$ from time inversion (see Remark 2.6 above) and via the filter given by $\overline{h^{1,j}}$ see [1, p. 58]). In fact, by formula (2.4) we have

$$W_n^{1,j} = \int_{[-\pi,+\pi[} e^{-i\lambda n}h^{1,j}(\lambda)dZ_{\mathbb{X}}(\lambda) = \int_{[-\pi,+\pi[} e^{i\lambda n}h^{1,j}(-\lambda)dZ_{\mathbb{X}}^{\phi}(\lambda) =$$
$$= \int_{[-\pi,+\pi[} e^{i\lambda n}\overline{h^{1,j}}(\lambda)dZ_{\mathbb{X}}^{\phi}(\lambda)$$

and so, considering the filter defined by $\overline{h^{1,j}}$ and the second order stationary process defined by the CURF $Z_{\mathbb{X}}^{\phi}$, we prove the stated result.

We will now suppose that $\mathbb{X}$ is a real valued *ARMA* process. For real processes, as stated in Remark 2.4, the spectral measure is invariant under ϕ and then, the spectral density is also invariant under ϕ. As the CURF $Z_{\mathbb{X}}^{\phi}$ has basis $\mu_{\mathbb{X}}^{\phi} \equiv \mu_{\mathbb{X}}$ and $f_{\mathbb{X}}^{\phi} \equiv f_{\mathbb{X}}$ (If the spectral density of an ARMA process is a rational function with real functions the the AR and MA polynomials may be chosen with real coefficients, see [1, p. 77]). then, we see that each $\mathbb{W}^{1,j}$ the filtered process, is also an *ARMA* process, due to the fact that by multiplying the spectral density by $\left|h^{1,j}\right|^2$ we only introduce roots strictly larger than 1 in the denominator and so, by resorting to the form of the spectral density in the canonical ARMA relation as stated in Examples 2.2 and 2.4, we still have an ARMA process. In fact, let $f_{\mathbb{W}^{1,j}}$ be the spectral density of the process $\mathbb{W}^{1,j}$, obtained from $\mathbb{X}$ by filtering by the square integrable function $h^{1,j}$. We have that, with the notations being used,
(3.6)

$$f_{\mathbb{W}^{1,j}}(\lambda) = f_{\mathbb{X}}^{\phi}(\lambda)\left|h^{1,j}(\lambda)\right|^2 = f_{\mathbb{X}}(\lambda)\left|h^{1,j}(\lambda)\right|^2 = \frac{\sigma^2}{2\pi}\left|\frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})(1-e^{-i\lambda}\eta_j)}\right|^2\ .$$

The polynomial $R(z) := P(z)(1 - z\eta_j)$ still has all its roots with modulus strictly greater than one and still verifies $R(0) = 1$. If $1/\eta_j$ is not a root of Q then, the representation of $f_{\mathbb{W}^{1,j}}$ still is a canonical ARMA relation. If Q admits $(1 - z\eta_j)$ as a factor then, writing $Q(z) = S(z)(1 - z\eta_j)$ we have the representation

$$f_{\mathbb{W}^{1,j}}(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{S(e^{-i\lambda})}{P(e^{-i\lambda})} \right|^2 ,$$

with S having all its roots with modulus strictly greater than one, still verifying $S(0) = 1$ and with P and S still not having any common roots. So, this representation is still a canonical ARMA relation. We may so observe that $\mathbb{Y}$ is asymptotically - due to the evanescent process - a linear combination of $ARMA$ processes.

We are now going to show that any linear combination of the $\mathbb{W}^{i,j}$ still is an ARMA process. That results from the fact that the innovation noise of each $\mathbb{W}^{i,j}$ coincides with the innovation noise of $\mathbb{X}$ (see [3, p. 210], for the general idea).

Let the spectral density of the process $\mathbb{X}$ be written according to the canonical ARMA representation as,

$$f_{\mathbb{X}}(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})} \right|^2 .$$

As the process $\mathbb{W}^{1,j} = (W_n^{1,j})_{n \geq 1}$ admits the spectral representation given by

$$W_n^{1,j} = \int_{[-\pi,+\pi[} e^{i\lambda n} \overline{h^{1,j}(\lambda)} dZ_{\mathbb{X}}^{\phi}(\lambda) ,$$

this process has a density that may be written according the canonical ARMA representation as

$$f_{\mathbb{W}^{1,j}}(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{Q^{1,j}(e^{-i\lambda})}{P^{1,j}(e^{-i\lambda})} \right|^2$$

with $Q^{1,j}$ and $P^{1,j}$ such that, either

$$Q(e^{-i\lambda}) = Q^{1,j}(e^{-i\lambda}) \left(1 - e^{-i\lambda}\eta_j\right) \text{ and } P^{1,j}(e^{-i\lambda}) = P(e^{-i\lambda})$$

or

$$Q^{1,j}(e^{-i\lambda}) = Q(e^{-i\lambda}) \text{ and } P^{1,j}(e^{-i\lambda}) = P(e^{-i\lambda})(1 - e^{-i\lambda}) .$$

We observe that as $d\mu_{\mathbb{X}}(\lambda) = f_{\mathbb{X}}(\lambda) d\mathrm{Leb}(\lambda)$, then we have that

$$\mu_{\mathbb{X}}\left(\left\{Q^{1,j}(e^{-i\lambda}) = 0\right\}\right) = \mu_{\mathbb{X}}\left(\left\{Q(e^{-i\lambda}) = 0\right\}\right) = 0 .$$

Now, let ε^j be the innovation noise of $\mathbb{W}^{1,j}$. We then have:

$$\begin{aligned} d\mu_{\varepsilon^j}(\lambda) &= \frac{1}{f_{\mathbb{W}^{1,j}}(\lambda)} 1\!\!1_{\left\{Q^{1,j}(e^{-i\lambda}) \neq 0\right\}}(\lambda) d\mu_{\mathbb{W}^{1,j}}(\lambda) = \\ &= \frac{1}{f_{\mathbb{W}^{1,j}}(\lambda)} 1\!\!1_{\left\{Q^{1,j}(e^{-i\lambda}) \neq 0\right\}}(\lambda) \left|h^{1,j}(\lambda)\right|^2 d\mu_{\mathbb{X}}(\lambda) = \\ &= 1\!\!1_{\left\{Q(e^{-i\lambda}) \neq 0\right\}}(\lambda) d\mathrm{Leb}(\lambda) = \frac{1}{f_{\mathbb{X}}(\lambda)} 1\!\!1_{\left\{Q(e^{-i\lambda}) \neq 0\right\}}(\lambda) d\mu_{\mathbb{X}}(\lambda) . \end{aligned}$$

We are now going to show that ε^j is the innovation noise of $\mathbb{X}$, that is, for $m < n$, $\mathbb{E}\left[X_m \varepsilon_n^j\right] = 0$, and so the innovation noise of $\mathbb{W}^{1,j}$ does not depend on j. For that we use the spectral representation of both processes, and so, considering any function ψ such that

$$|\psi(\lambda)|^2 = \frac{1}{f_{\mathbb{X}}(\lambda)} \mathbb{I}_{\left\{Q(e^{-i\lambda}) \neq 0\right\}}(\lambda) \, ,$$

we have, using the isometry property of the stochastic integral and the Cauchy theorem,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{[-\pi,+\pi[} e^{i\lambda m} dZ_{\mathbb{X}}(\lambda)\right) \cdot \overline{\left(\int_{[-\pi,+\pi[} e^{i\lambda n} dZ_{\varepsilon^j}(\lambda)\right)}\right] \\
&= \mathbb{E}\left[\left(\int_{[-\pi,+\pi[} e^{i\lambda m} dZ_{\mathbb{X}}(\lambda)\right) \cdot \overline{\left(\int_{[-\pi,+\pi[} e^{i\lambda n} \psi(\lambda) dZ_{\mathbb{X}}(\lambda)\right)}\right] \\
&= \int_{[-\pi,+\pi[} e^{i\lambda(m-n)} \overline{\psi(\lambda)} d\mu_{\mathbb{X}}(\lambda) \\
&= \int_{[-\pi,+\pi[} e^{i\lambda(m-n)} \frac{\sqrt{2\pi}}{\sigma} \overline{\left(\frac{P(e^{-i\lambda})}{Q(e^{-i\lambda})}\right)} \frac{\sigma^2}{2\pi} \frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})} \overline{\left(\frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})}\right)} \mathbb{I}_{\left\{Q(e^{-i\lambda}) \neq 0\right\}} d\mathrm{Leb}(\lambda) \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{[-\pi,+\pi[} e^{-i\lambda(n-m)} \frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})} d\mathrm{Leb}(\lambda) \\
&= \frac{\sigma i}{\sqrt{2\pi}} \int_{\mathbb{T}} z^{n-(m+1)} \frac{Q(z)}{P(z)} dz = 0 \, .
\end{aligned}$$

Let $\varepsilon = (\varepsilon_k)_{k \in \mathbb{Z}}$ be the common innovation noise of all the processes $\mathbb{W}^{1,j}$. We then have (see [1, p. 81]) that for each j and some (square) integrable sequence $(c_k^j)_{k \geq 0}$ we may write,

$$\mathbb{W}_n^{1,j} = \sum_{k \geq 0} c_k^j \varepsilon_{n-k} \, ,$$

and so for $\alpha_j, \alpha_l \in \mathbb{C}$, as $\left(\alpha_j c_k^j + \alpha_l c_k^l\right)_{k \geq 0}$ is a (square) integrable sequence,

$$\alpha_j \mathbb{W}_n^{1,j} + \alpha_l \mathbb{W}_n^{1,l} = \sum_{k \geq 0} \left(\alpha_j c_k^j + \alpha_l c_k^l\right) \varepsilon_{n-k} \, ,$$

and $\alpha_j \mathbb{W}^{1,j} + \alpha_l \mathbb{W}^{1,l}$ is an ARMA process. □

We now deal with the case of some non stationary processes which are relevant for the applications.

Theorem 3.2. *Under the same conditions of Theorem 3.1, if the open Markov chain scheme model is fed by a real valued ARIMA or SARMA process then the population in each of the transient states may be described as a sum of a deterministic trend plus a linear combination of ARMA processes plus an evanescent process.*

Proof: Let $\mathbb{X} = (X_n)_{n\in\mathbb{Z}}$ be an ARMA process. Let $s, d \geq 1$ be integers and consider the following functions defined, for $i, j \in \{0, 1, \ldots, s-1\}$ and $\alpha, \beta \in \{0, 1, \ldots, d-1\}$, by:

$$U_{i,\alpha}(x) = x^\alpha \cos\left(\frac{2\pi i}{s} x\right) \quad \text{and} \quad V_{j,\beta}(x) = x^\beta \sin\left(\frac{2\pi j}{s} x\right) .$$

Consider now the function given by linear combinations with complex coefficients of the functions $U_{i,\alpha}$ and $V_{j,\beta}$ as

$$P_{(s,d)}(x) = \sum_{0\leq i\leq s-1, 0\leq\alpha\leq d-1} a_{i,\alpha} U_{i,\alpha}(x) + \sum_{0\leq j\leq s-1, 0\leq\beta\leq d-1} b_{j,\beta} V_{j,\beta}(x) .$$

Then the process $\mathbb{T} = (T_n)_{n\in\mathbb{Z}}$ represented as

$$T_n = P_{(s,d)}(n) + \sum_{j=0}^{n} \gamma_j X_{n-j} ,$$

is an identifiable ARIMA or SARMA process for an appropriate choice of s, d and the complex coefficients $a_{i,\alpha}$, $b_{j,\beta}$ and γ_j (see [1, p. 87, 89]). Moreover, every identifiable ARIMA or SARMA process can be represented in that form. Consider now an open Markov chain scheme fed by $\mathbb{T}$. We have the obvious decomposition:

$$\begin{aligned}
\mathbf{Y}_n &= \sum_{k=1}^{n} T_k (\mathbf{q}^k)^\intercal\, \mathbf{P}^{(n-k)} = \\
&= \sum_{k=1}^{n} P_{(s,d)}(k) (\mathbf{q}^k)^\intercal\, \mathbf{P}^{(n-k)} + \sum_{k=1}^{n} \left(\sum_{j=0}^{k} \gamma_j X_{k-j}\right) (\mathbf{q}^k)^\intercal\, \mathbf{P}^{(n-k)} = \\
&= \sum_{k=1}^{n} P_{(s,d)}(k) (\mathbf{q}^k)^\intercal\, \mathbf{P}^{(n-k)} + \sum_{j=0}^{n} \gamma_j \left(\sum_{k=j}^{n} X_{k-j} (\mathbf{q}^k)^\intercal\, \mathbf{P}^{(n-k)}\right)
\end{aligned}$$

and so, as the right-hand term of the last sum is a linear combination of ARMA processes the result follows. □

Remark 3.1. Using the results in [4], we note that, at least on average, the asymptotic behavior of the subpopulations can be described.

4. AN APPLICATION TO COMSUMPTION CREDIT

4.1. Real data

In this section we will present fittings of a second order processes to real data of consumption credit portfolio from a Cape Verdean bank.

In this portfolio, we defined five risk classes, according to the number of days in delay of the monthly reimbursements, as shown in Table 2, and an extra class for the clients leaving the portfolio.

Table 2: Portfolio risk classes

Risk Class	**Number of days in delay**
RC1	0 - 30
RC2	31-60
RC3	61-90
RC4	91-120
RC5	> 120
RC6	Leaving

In each month, each client is classified into the risk class that refers to his number of days in delay of reimbursments. Only fully paid contracts are allowed to move to risk class 6.

The transition matrix, estimated from portfolio data, is given by:

$$\mathbf{P} = \begin{bmatrix} 0.934735 & 0.026566 & 0 & 0 & 0 & 0.038698 \\ 0.518363 & 0.285733 & 0.195903 & 0 & 0 & 0 \\ 0.009076 & 0.372018 & 0.248963 & 0.369943 & 0 & 0 \\ 0 & 0.007835 & 0.335464 & 0.205361 & 0.450928 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.1}$$

Naturally, each new client is initially placed in risk class 1, and so,

$$(\mathbf{q}^k)^\intercal \equiv (\mathbf{q})^\intercal = [\ 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ]$$

In previous works (see [4] and [5]), using this data and the related client information in the consumption credit database, we provided models for the spread to be applied to each client and related this to the global spread of the portfolio, estimated using an open Markov chain model with the number of entrances modeled by a sequence of Poisson laws. We recall that one of the motivations for the present work is to develop a model with the entrances of new clients modeled by a time series.

The data on the number of new monthly clients arriving to the portfolio corresponds to a monthly sequence of 106 observations. The fitting was performed using Wolfram Mathematica and, for illustration purposes, we adopted two different approaches. In the first one, we fitted a time series directly to data. In the second, we firstly fitted a sigmoid type function to data, as in [4], and then, a time series to the residuals of the sigmoid fitting.

The results obtained for the first approach are illustrated in Table 3.

As shown in Table 3 the best model for the entrance data, under both the *AIC* and the *BIC* criterias, is an ARIMA$[0, 1, 1]$ model.

For the second approach, we show, in Figure 1, on the left side graphic, both the data and the fitted sigmoid type function and, in the right side graphic, the correspondent residuals.

Table 3: Fitting the entrance data directly

	Candidate	BIC
1	**ARIMAProcess[0, 1, 1]**	**909.593**
2	ARIMAProcess[1, 1, 0]	912.706
3	ARIMAProcess[0, 1, 2]	913.637
4	ARIMAProcess[1, 1, 1]	918.713
5	ARIMAProcess[1, 1, 2]	919.631
6	ARIMAProcess[0, 1, 0]	925.378
7	ARIMAProcess[1, 2, 1]	929.179
8	ARProcess[2]	932.164
9	ARProcess[3]	935.583
10	ARIMAProcess[2, 2, 2]	935.643

	Candidate	AIC
1	**ARIMAProcess[0, 1, 1]**	**897.294**
2	ARIMAProcess[0, 1, 2]	898.366
3	ARIMAProcess[1, 1, 0]	900.518
4	ARIMAProcess[1, 1, 2]	901.762
5	ARIMAProcess[1, 1, 1]	903.686
6	ARIMAProcess[1, 2, 1]	914.666
7	ARIMAProcess[2, 2, 2]	916.189
8	ARIMAProcess[0, 1, 0]	916.863
9	ARProcess[2]	917.8
10	ARProcess[3]	918.702

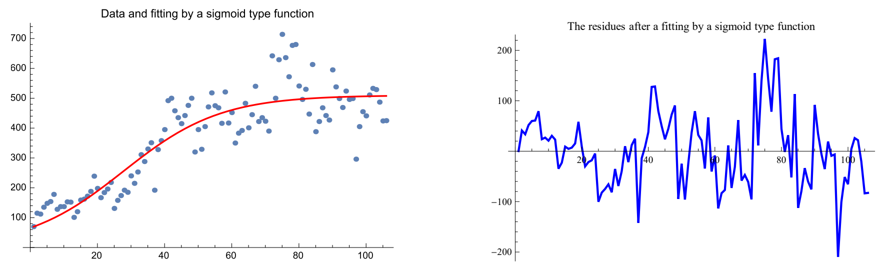


Figure 1: Fitting a sigmoid type function to data and to the residuals.

In Table 4 it is shown that the best model for the residuals of the fitting of the entrance data by a sigmoid type function, under both the *AIC* and the *BIC* criterias, is the SARMA$[(1,0),(1,0)_{34}]$ model.

Table 4: Fitting the residues of a fitting by a sigmoid function

	Candidate	AIC
1	**SARMAProcess[{1, 0}, {1, 0}$_{34}$]**	**876.091**
2	SARMAProcess[{1, 0}, {2, 0}$_{34}$]	876.237
3	SARMAProcess[{1, 0}, {1, 1}$_{34}$]	878.53
4	SARMAProcess[{2, 0}, {1, 0}$_{34}$]	879.641
5	SARMAProcess[{1, 0}, {2, 1}$_{34}$]	880.596
6	SARMAProcess[{1, 1}, {1, 0}$_{34}$]	883.077
7	ARProcess[2]	884.778
8	SARMAProcess[{2, 1}, {1, 0}$_{34}$]	885.059
9	SARMAProcess[{1, 0}, {0, 1}$_{34}$]	886.115
10	ARProcess[1]	886.819

	Candidate	BIC
1	**SARMAProcess[{1, 0}, {1, 0}$_{34}$]**	**882.283**
2	SARMAProcess[{1, 0}, {2, 0}$_{34}$]	883.098
3	SARMAProcess[{1, 0}, {1, 1}$_{34}$]	885.067
4	SARMAProcess[{2, 0}, {1, 0}$_{34}$]	886.016
5	SARMAProcess[{1, 0}, {2, 1}$_{34}$]	887.306
6	SARMAProcess[{1, 1}, {1, 0}$_{34}$]	888.927
7	ARProcess[2]	889.879
8	SARMAProcess[{2, 1}, {1, 0}$_{34}$]	890.959
9	SARMAProcess[{1, 0}, {0, 1}$_{34}$]	891.028
10	ARProcess[1]	891.08

In Table 5 we present the results on the parameter estimation for both the ARIMA and the SARMA processes.

4.2. A simulation study

In this section we will compare, by means of a simulation study, two ways of obtaining the distribution of the number of elements in each risk class. First,

Table 5: The ARIMA and SARMA parameter tables

ARIMA model

	Estimate	Standard Error	t-Statistic	P-Value
b_1	−0.398645	0.0890771	−4.47528	9.63596×10^{-6}

SARMA model

	Estimate	Standard Error	t-Statistic	P-Value
a_1	0.438344	0.0949723	4.6155	5.522×10^{-6}
α_1	0.478033	0.0931484	5.13196	6.53092×10^{-7}

we simulate 300 paths of the Markov chain model and compute the observed proportions of elements in each one of the six classes. We will also simulate the number of elements in each class according to the two models fitted in the previous section. The results are presented in Tables 6 and 7. The results in the first table show that the sub-population in class 5 is slightly larger in both ARIMA and SARMA models, when compared with the direct simulation of the Markov chain. As the class 5 population measures the most part of the risk of the portfolio, both the ARIMA and SARMA models are conservative, but not excessively.

Table 6: Proportions in each class by simulation of the Markov chain and of the Markov chain scheme models ARIMA and SARMA

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
MARKOV	0.0366667	0.00333333	0.00333333	0.00333333	0.0266667	0.926667
ARIMA	0.106427	0.00594949	0.00328688	0.00362677	0.0305018	0.850208
SARMA	0.130596	0.00705139	0.00362903	0.00378776	0.0307309	0.824204

We also computed the relative proportions of elements in the population in each one of the five transient classes. The results show a remarkable difference between the Markov chain simulation and the Markov chain scheme fed with the sum of a sigmoid trend plus a SARMA process, thus showing the advantage of an open model.

Table 7: Conditional proportions for the 5 transient classes - simulation

	Class 1	Class 2	Class 3	Class 4	Class 5
MARKOV	0.5	0.0454545	0.0454545	0.0454545	0.363637
ARIMA	0.710498	0.0397184	0.021943	0.0242121	0.203628
SARMA	0.742888	0.0401113	0.0206435	0.0215464	0.17481

We simulated 300 paths for each of the two models fitted in section 4, to wit, the ARIMA$[0,1,1]$ and the SARMA$[(1,0),(1,0)_{34}]$. We computed the mean and the standard deviation for each classe and the correspondent one standard deviation confidence interval. Despite the paths in the ARIMA$[0,1,1]$ possibly taking negative values we computed the corresponding number of elements in each class. The results in Table 8 clearly show that the SARMA$[(1,0),(1,0)_{34}]$ model, for the residuals of a sigmoid type function fitting, is much more adequate to describe the evolution of the entrance of new clients in the credit portfolio.

In Figure 2 we observe that the results given by the SARMA$[(1,0),(1,0)_{34}]$ model are more meaningful. In fact, in the empirical distribution of the simulated number of elements, negative numbers occur in both classes 1 and 5.

Table 8: Data at date 106 and confidence intervals from models

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Data	5307	284	144	149	1203	32256
ARIMA	[-1614,5904]	[-81,321]	[-36,168]	[-33,179]	[-248,1478]	[-6164,40436]
SARMA	[4669,5593]	[255,299]	[134,151]	[142,156]	[1158,1257]	[31112,33654]

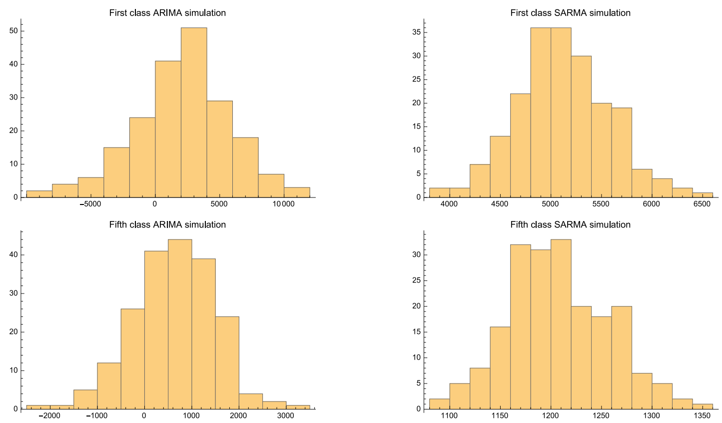


Figure 2: Simulated empirical distributions in classes 1 and 5.

ACKNOWLEDGMENTS

This work has been partially supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, under project UID/MAT/00297/2013 (CMA/UNL).

REFERENCES

[1] Azencott, R. and Dacunha-Castelle, D. (1986). *Series of Irregular Observations. Forecasting and Model Building*, Springer-Verlag.

[2] Bartholomew, D.J. (1982). *Stochastic Models for Social Processes*, John Wiley & Sons, Ltd. Chichester, Third Edition. Wiley Series in Probability and Mathematical Statistics.

[3] Bremaud, P. (2014). *Fourier Analysis and Stochastic Processes*, Cham: Springer, New York.

[4] Esquível, M.L.; Fernandes, J.M. and Guerreiro, G.R. (2014). On the evolution and asymptotic analysis of open Markov populations: application to consumption credit, *Stochastic Models*, **30**, 3, 365–389.

[5] Esquível, M.L.; Fernandes, J.M.; Guerreiro, G.R. and Silva, A.F. (2015). On a spread model for portfolio credit risk modeling, *AIP Conference Proceedings*, **1648**, 540004, 1–4.

[6] Esquível, M.L.; Fernandes, J.M. and Guerreiro, G.R. (2016). An open Markov chain scheme model for a credit comsumption portfolio fed by ARIMA and SARMA processes, *AIP Conference Proceedings*, **1738**, 190003.

[7] Feichtinger, G. (1976). On the generalization of stable age distributions to Gani-type person-flow models, *Advances in Applied Probability*, **3**, 3, 433–445.

[8] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, Third Edition, John Wiley & Sons Inc., New York.

[9] Gani, J. (1963). Formulae for projecting enrolments and degrees awarded in universities, *Journal of the Royal Statistical Society. Series A (General)*, **126**, 3, 400–409.

[10] Laha, R.G. and Rohatgi, V.K. (1979). *Probability Theory*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

[11] McClean, S.I. (1976). A continuous-time population model with poisson recruitment, *Journal of Applied Probability*, **13**, 2, 348–354.

[12] McClean, S.I. (1976). Continuous-time stochastic models of a multigrade population, *Journal of Applied Probability*, **15**, 1, 26–37.

[13] McClean, S.I. (1980). A semi-markov model for a multigrade population with poisson recruitment, *Journal of Applied Probability*, **17**, 3, 846–852.

[14] Mehlmann, A. (1977). A note on the limiting behaviour of discrete-time markovian manpower models with inhomogeneous independent poisson input, *Journal of Applied Probability*, **14**, 3, 611–613.

[15] Pollard, J.H. (1967). Hierarchical population models with poisson recruitment, *Journal of Applied Probability*, **4**, 1, 209–213.

[16] Papadopoulou, A.A. and Vassiliou, P.-C.G. (1994). Asymptotic behavior of non-homogeneous semi-markov systems, *Linear Algebra and its Applications*, **210**, 153–198.

[17] Shiryaev, A.N. (1996). *Probability*, volume 95 of Graduate Texts in Mathematics, Springer-Verlag, New York.

[18] Taylor, G.J.; McClean, S.I. and Millard, P.H. (1998). Using a continuous-time markov model with poisson arrivals to describe the movements of geriatric patients, *Applied Stochastic Models and Data Analysis*, **14**, 2, 165–174.

[19] Tsantas, N. (1982). Stochastic analysis of a non-homogeneous Markov system, *Europena Journal of Operational Research*, **85**, 670–685.

[20] Vassiliou, P.-C.G. (1982). Asymptotic behavior of markov systems, *Journal of Applied Probability*, **19**, 4, 851–857.

[21] Vassiliou, P.-C.G. (1982). On the Limiting Behaviour of a Non-Homogeneous Markovian Manpower Model with Independent Poisson Input, *Journal of Applied Probability*, **19**, 2, 433–438.

[22] Vassiliou, P.-C.G. (1997). The evolution of the theory of non-homogeneous Markov systems, *Applied Stochastic Models and Data Analysis*, **13**, 3-4, 159–176.

[23] Vassiliou, P.-C.G.; Georgiou, A.C. and Tsantas, N. (1982). Control of Asymptotic Variability in Non-Homogeneous Markov Systems, *Journal of Applied Probability*, **27**, 4, 756–766.

[24] Young, A. and Almond, G. (1961). Predicting Distribtions of Staff, *The Computer Journal*, **3**, 4, 246–250.

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.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Copyright