ISSN 1645-6726

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT

Statistical Journal

Volume 10, No.3
November 2012

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.: +351 218 426 100
    Fax: +351 218 426 364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 226 050 748
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *300 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT included]

- Single issue …….. **€ 11**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)……. **€ 26**
- Annual subscription (No. 2, No. 3) …….. **€ 18**

INDEX

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

REVSTAT – Statistical Journal
Volume 10, Number 3, November 2012, 285–308

SMALL AREA ESTIMATION USING A SPATIO-TEMPORAL LINEAR MIXED MODEL

Authors: Luis N. Pereira
– Escola Superior de Gestão, Hotelaria e Turismo,
Centro de Investigação sobre o Espaço e as Organizações,
Universidade do Algarve, Portugal
lmper@ualg.pt

Pedro S. Coelho
– ISEGI – Universidade Nova de Lisboa, Portugal
Faculty of Economics, Ljubljana University, Slovenia
psc@isegi.unl.pt

Received: February 2012 Accepted: April 2012

Abstract:

- In this paper it is proposed a spatio-temporal area level linear mixed model involving spatially correlated and temporally autocorrelated random effects. An empirical best linear unbiased predictor (EBLUP) for small area parameters has been obtained under the proposed model. Using previous research in this area, analytical and bootstrap estimators of the mean squared prediction error (MSPE) of the EBLUP have also been worked out. An extensive simulation study using time-series and cross-sectional data was undertaken to compare the efficiency of the proposed EBLUP estimator over other well-known EBLUP estimators and to study the properties of the proposed estimators of MSPE.

Key-Words:

- *empirical best linear unbiased prediction; linear mixed model; mean squared prediction error estimation; small area estimation; spatial correlation; temporal autocorrelation.*

AMS Subject Classification:

- 62J05, 62F40, 62D05.

1. INTRODUCTION

Large scale repeated sample surveys are usually designed to produce reliable estimates of several characteristics of interest for large subgroups of a population, from which samples are drawn. A subgroup may be a geographical region or a group obtained by cross-classification of demographic factors such as age or gender. However, for effective planning in a wide variety of fields, there is a growing demand to produce similar estimates for smaller subgroups for which adequate sample sizes are not available. In fact, sample sizes are frequently very small or even zero in many subgroups of interest (small areas), resulting in unreliable direct design-based small area estimates. This creates a need to employ indirect estimators that "borrow information" from related small areas and time periods through linking models using recent census or current administrative data, in order to increase the effective sample size and thus precision. Such indirect estimators are often based on explicit Linear Mixed Models (LMM) that provide a link to a related small area through the use of supplementary data. The empirical best linear unbiased prediction (EBLUP) approach is one of the most popular methods for the estimation of small area parameters of interest. This approach uses an appropriate LMM which captures several salient features of the areas and combines information from censuses or administrative records conjointly with the survey data. When time-series and cross-sectional data are available, longitudinal LMM might be useful to take simultaneously advantage of both the possible spatial similarities among small areas and the expected time-series relationships of the data in order to improve the efficiency of the small area estimators. Although there is some research on temporal (e.g. Rao & Yu, 1994; Datta *et al.*, 2002; Saei & Chambers, 2003; Pereira & Coelho, 2010) and on spatial small area estimation using LMM (e.g. Salvati, 2004; Petrucci *et al.*, 2005; Petrucci & Salvati, 2006; Chandra *et al.*, 2007; Pratesi & Salvati, 2008), there is a need of research in the field of small area estimation using LMM with spatio-temporal information. Such kind of estimation might account simultaneously for the spatial dependence and the chronological autocorrelation in order to strengthen the small area estimates. This can be achieved by incorporating in the model both area specific random effects and area-by-time specific random effects. The area specific random effects could then be linked by a spatial process, while the area-by-time random effects could be linked by a temporal process. This approach is definitely more complex than a simple regression method and its success depends on the ability to define a suitable spatial neighbourhood, to specify properly spatial and temporal processes and to estimate additional parameters.

Thus, the main goal of this paper is to propose a simple and intuitive spatio-temporal LMM involving spatially correlated and temporally autocorrelated random area effects, using both time-series and cross-sectional data. The proposed model is an extension of three well-known small area models in the literature.

Under the proposed model, two research questions are addressed. Firstly, we analyse the extent to which the spatial and the temporal relationships in the data justify the introduction of a spatial and a temporal autoregressive parameters in the model. This is carried out via a simulation study which compares the efficiency of the proposed EBLUP estimator against other well-known EBLUP estimators, by taking into account the joint effects of the following components: the sampling variances of the direct estimators of the small area parameters; the variance components of the random effects; and the spatial and temporal autocorrelation parameters. Secondly, we discuss how to measure the uncertainty of the proposed EBLUP. This is carried out via a simulation study which compares the accuracy of the analytical and the bootstrap estimators of the mean squared prediction error (MSPE) introduced in this paper.

Singh *et al.* (2005) proposed the only existing work on small area estimation using spatial-temporal approaches. They proposed a spatio-temporal state space model via Kalman filtering estimation which, like our model, borrows strength from past surveys, neighbour small areas and a set of covariates. However, our model, unlike the model due to Singh *et al.* (2005), makes use of a different estimation method and incorporates independent specific random effects. Note that the model due to Singh *et al.* (2005) considers an interaction between the spatial dependence and the temporal autocorrelation, since the spatial process is stated in the design matrix of random effects.

This paper is organized as follows. In Section 2 it is proposed a spatio-temporal area level LMM. The BLUP and EBLUP of the mixed effects are also provided in this section. Section 3 discusses the measure of uncertainty of the proposed EBLUP. In this section it is proposed both an analytical and a parametric bootstrap method to estimate the MSPE of the EBLUP. The design of the simulation study, as well as its empirical results on the efficiency of the proposed EBLUP and on the properties of the proposed estimators of MSPE, is reported in Section 4. Finally, the paper ends with a conclusion in Section 5, which summarizes the main advantages of the proposed methodology and identifies further issues of research.

2. THE SPATIO-TEMPORAL MODEL

2.1. Proposed model

Let $\boldsymbol{\theta} = \text{col}_{1\leq i\leq m}(\boldsymbol{\theta}_i)$ be a $mT\times 1$ vector of the parameters of inferential interest and assume that $\mathbf{y} = \text{col}_{1\leq i\leq m}(\mathbf{y}_i)$ is its design-unbiased direct survey estimator. Here $\boldsymbol{\theta}_i = \text{col}_{1\leq t\leq T}(\theta_{it})$, $\mathbf{y}_i = \text{col}_{1\leq t\leq T}(y_{it})$ and y_{it} is the direct survey

estimator of the parameter of interest for small area i at time t, $\theta_{it}(i = 1, ..., m; t = 1, ..., T)$. Thus the sampling error model is given by:

$$\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{2.1}$$

where $\boldsymbol{\epsilon} = \text{col}_{1\leq i\leq m;1\leq t\leq T}(\varepsilon_{it})$ is a $mT\times 1$ vector of sampling errors. We assume that $\boldsymbol{\epsilon} \overset{iid}{\sim} N(\mathbf{0}; \mathbf{R})$, where $\mathbf{R} = \text{diag}_{1\leq i\leq m;1\leq t\leq T}(\sigma^2_{it})$ is a $mT \times mT$ matrix with known sampling variances of the direct estimators. We propose the following linking model in which the parameters of inferential interest are related to area-by-time specific auxiliary data through a linear model with random effects:

$$\boldsymbol{\theta} \;=\; \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{v} + \mathbf{u}_2 \tag{2.2}$$

where $\mathbf{X} = \text{col}_{1\leq i\leq m;1\leq t\leq T}(\mathbf{x}'_{it})$ is a $mT\times p$ design matrix of area-by-time specific auxiliary variables with rows given by $\mathbf{x}'_{it} = (x_{it1}, ..., x_{itp})$ $(1\times p)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ $(p\times 1)$ is a column vector of regression parameters, $\mathbf{v} = \text{col}_{1\leq i\leq m}(\nu_i)$ is a $m\times 1$ vector of random area specific effects and $\mathbf{u}_2 = \text{col}_{1\leq i\leq m;1\leq t\leq T}(u_{2,it})$ is a $mT\times 1$ vector of random area-by-time specific effects. Further, $\mathbf{Z}_1 = \mathbf{I}_m \otimes \mathbf{1}_T$ $(mT\times m)$ where $\mathbf{I}_m$ is an identity matrix of order m and $\mathbf{1}_T$ $(T\times 1)$ is a column vector of ones. We assume that $\mathbf{X}$ has full column rank p and $\mathbf{v}$ is the vector of the second order variation.

In order to take into account for the spatial dependence among small areas we propose the use of a simple spatial model in the random area specific effects. In particular, we propose the use of the simultaneous autoregressive (SAR) process (Anselin, 1992), where the vector $\mathbf{v}$ satisfies:

$$\mathbf{v} = \phi\,\mathbf{W}\mathbf{v} + \mathbf{u}_1 \;\;\Rightarrow\;\; \mathbf{v} = (\mathbf{I}_m - \phi\,\mathbf{W})^{-1}\mathbf{u}_1\ , \tag{2.3}$$

where ϕ is a spatial autoregressive coefficient which defines the strength of the spatial relationship among the random effects associated with neighboring areas and $\mathbf{W} = \{w_{ij}\}$ $(m\times m)$ is a known spatial proximity matrix which indicates whether the small areas are neighbors or not $(i, j = 1, ..., m)$. A simple common way to specify $\mathbf{W}$ is to define $w_{ij} = 1$ if the area i is physically contiguous to area j and $w_{ij} = 0$ otherwise. In this case $\mathbf{W}$ is a contiguity matrix. The most common way to define $\mathbf{W}$ is in the row standardized form, that is, restricting rows to satisfy $\sum_{j=1}^{m} w_{ij} = 1$, for $i = 1, ..., m$. It is yet possible to create more elaborate weights as functions of the length of common boundary between the small areas (Wall, 2004). Further, $\mathbf{u}_1 = \text{col}_{1\leq i\leq m}(u_{1i})$ is a $m\times 1$ vector of independent error terms satisfying $\mathbf{u}_1 \overset{iid}{\sim} N(\mathbf{0}; \sigma^2_u\mathbf{I}_m)$.

In order to borrow strength across time we propose the use of autocorrelated random effects. In particular, we propose that $u_{2,it}$'s follow a common first-order autoregressive [AR(1)] process for each small area:

$$u_{2,it} \;=\; \rho\, u_{2,i,t-1} + \xi_{it}\ , \qquad |\rho| < 1\ , \tag{2.4}$$

where ξ_{it}'s are the error terms satisfying $\xi_{it} \overset{iid}{\sim} N(0;\sigma^2)$ and ρ is a temporal autoregressive coefficient which measures the level of chronological autocorrelation.

Combining models (2.1)–(2.4), the proposed model involving spatially correlated and temporally autocorrelated random area effects may be written in matrix form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\upsilon} + \boldsymbol{\epsilon} \ , \tag{2.5}$$

where $\mathbf{Z} = [\mathbf{Z}_1 \ \mathbf{I}_{mT}]$, $\mathbf{Z}_1 = \mathbf{I}_m \otimes \mathbf{1}_T$ and $\boldsymbol{\upsilon} = [\mathbf{v}' \ \mathbf{u}_2']'$. Further, we assume that error terms $\mathbf{v} = (\mathbf{I}_m - \phi\,\mathbf{W})^{-1}\mathbf{u}_1$, $\mathbf{u}_2$ and $\boldsymbol{\epsilon}$ are mutually independent distributed with $\mathbf{u}_1 \overset{iid}{\sim} N(\mathbf{0};\sigma_u^2\mathbf{I}_m)$, $\mathbf{u}_2 \overset{iid}{\sim} N(\mathbf{0};\sigma^2\mathbf{I}_m \otimes \boldsymbol{\Gamma})$ and $\boldsymbol{\epsilon} \overset{iid}{\sim} N(\mathbf{0};\mathbf{R})$, where $\boldsymbol{\Gamma}(T\times T)$ is a matrix with elements $\rho^{|r-s|}/(1-\rho^2)$, $r,s = 1,...,T$ and $\mathbf{R} = \mathrm{diag}_{1\leq i\leq m;1\leq t\leq T}(\sigma_{it}^2)$. We can now see that model (2.5) is a special case of the general LMM with a block diagonal covariance matrix of $\boldsymbol{\upsilon}$, given by $\mathbf{G} = \mathrm{diag}_{1\leq k\leq 2}(\mathbf{G}_k)$, where $\mathbf{G}_1$ and $\mathbf{G}_2$ are the covariance matrices of $\mathbf{v}$ and $\mathbf{u}_2$, respectively. As showed by Salvati (2004) and by Rao & Yu (1994), these covariance structures are given by $\mathbf{G}_1 = E(\mathbf{v}\mathbf{v}') = \sigma_u^2\big[(\mathbf{I}_m - \phi\mathbf{W})'(\mathbf{I}_m - \phi\mathbf{W})\big]^{-1}$ and $\mathbf{G}_2 = E(\mathbf{u}_2\mathbf{u}_2') = \sigma^2\mathbf{I}_m \otimes \boldsymbol{\Gamma}$, respectively. It follows that the covariance matrix of $\mathbf{y}$ is:

$$\mathbf{V} \ = \ \mathrm{diag}_{1\leq i\leq m;1\leq t\leq T}(\sigma_{it}^2) + \mathbf{Z}_1\sigma_u^2\mathbf{B}^{-1}\mathbf{Z}_1' + \sigma^2\mathbf{I}_m \otimes \boldsymbol{\Gamma} \ , \tag{2.6}$$

where $\mathbf{B} = (\mathbf{I}_m - \phi\mathbf{W})'(\mathbf{I}_m - \phi\mathbf{W})$. Note that $\mathbf{V}$ is not a block diagonal covariance structure, like in the context of the well known Fay–Herriot and Rao–Yu models. Finally, note that the temporal model due to Rao & Yu (1994) can be obtained from model (2.5) setting $\phi = 0$, as well as the spatial model due to Salvati (2004) can be obtained from model (2.5) setting $T = 1$, $\rho = 0$ and $\sigma^2 = 0$. The model proposed by Fay & Herriot (1979) is also a particular case of model (2.5) since it can be obtained setting $T = 1$, $\phi = 0$, $\rho = 0$ and $\sigma^2 = 0$. However, if the spatial and temporal autocorrelation parameters are not equal to zero then the model proposed by Singh *et al.* (2005) cannot be obtained from model (2.5) because it is assumed in this model that the error terms are mutually independent.

2.2. The BLUP

The current small area parameter, $\theta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \nu_i + u_{2,it}$, is a special case of the linear combination $\tau = \mathbf{k}_{it}'\boldsymbol{\beta} + \mathbf{m}_{it}'\boldsymbol{\upsilon}$, where $\mathbf{k}_{it}' = \mathbf{x}_{it}'$ and $\mathbf{m}_{it}' = [\mathbf{m}_{1i}' \ \mathbf{m}_{2it}']$ in which $\mathbf{m}_{1i}' = (0,...,0,1,0,...,0)$ is a $1\times m$ vector with value 1 in the i^{th} position and 0 elsewhere, and $\mathbf{m}_{2it}' = (0,...,0,1,0,...,0)$ is a $1\times mT$ vector with value 1 in the $(it)^{\text{th}}$ position and 0 elsewhere. Noting that model (2.5) is a special case of the general LMM, thus the BLUP estimator of $\tau = \theta_{it}$ can be obtained from Henderson's general results (Henderson, 1975). Assuming first that $\boldsymbol{\psi} = (\sigma^2, \sigma_u^2, \phi, \rho)'$

is fully known, the BLUP of θ_{it} is given by:

$$\widetilde{\theta}_{it} = \widetilde{\theta}_{it}^{H}(\boldsymbol{\psi}) = \mathbf{x}_{it}'\widetilde{\boldsymbol{\beta}} + \mathbf{h}_{it}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \,, \tag{2.7}$$

where $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\boldsymbol{\psi}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$ and $\mathbf{h}_{it}'$ is a $1\times mT$ vector which captures the potential spatial and temporal autocorrelation present in the i^{th} small area. Further, $\mathbf{h}_{it}' = \sigma_u^2\boldsymbol{\varsigma}_i' \otimes \mathbf{1}_T' + \sigma^2\boldsymbol{\zeta}_{it}'$ where $\boldsymbol{\varsigma}_i' = \{\varsigma_{ii'}\}$ is the t^{th} row of the $\mathbf{B}^{-1}$ matrix and $\boldsymbol{\zeta}_{it}'$ is a $1\times mT$ vector with m T-dimensional blocks, with the t^{th} row of the $\boldsymbol{\Gamma}$ matrix, $\boldsymbol{\gamma}_t$, in the i^{th} block and null vectors, $\mathbf{0}_{1\times T}$, elsewhere, $i, i' = 1, ..., m$; $t = 1, ..., T$. This estimator can be classified as a combined estimator, since it can be decomposed into two components: a synthetic estimator, $\mathbf{x}_{it}'\widetilde{\boldsymbol{\beta}}$, and a correction factor, $\mathbf{h}_{it}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})$. We can say that the weights in $\mathbf{h}_{it}'\mathbf{V}^{-1}$ allow a correction of the synthetic part of the estimator (2.7) through the regression residuals from the small area that is the target of inference at t^{th} time point and from this area at previous time periods, but also from the regression residuals from other small areas at the t^{th} time point which are spatially correlated with the target small area. From expression (2.7) it is also possible to observe that when a particular small area is not represented in the sample of period t, it is still possible to estimate the correction factor through the spatial and/or temporal autocorrelation, if there are data collected for small area i in at least one of the previous samples and/or for one related small area at time period t. This is certainly a very appealing characteristic of the estimator (2.7): it is possible to avoid the reduction of the proposed estimator to a pure synthetic estimator, even when the sample size of period t in the i^{th} small area is null.

2.3. The two-stage estimator

The BLUP estimator depends on the parameters of the vector $\boldsymbol{\psi} = (\sigma^2, \sigma_u^2, \phi, \rho)'$, but in practice they are unknown and have to be estimated from the data. As far as the estimation of $\boldsymbol{\psi}$ is concerned, a number of methods have been proposed in the literature, such as the maximum likelihood method (Fisher, 1922; Hartley & Rao, 1967), the restricted maximum likelihood method (Thompson, 1962; Patterson & Thompson, 1971) and the analysis of variance (ANOVA) method (Henderson, 1953), among others. While the likelihood-based methods assume the normality of the error terms, the ANOVA method is free of this kind of assumptions. So, in the present work we have decided to estimate the variance components through an extension of Henderson method 3 (Henderson, 1953) to the model (2.5) with spatial correlated errors through a SAR process, ν_i, temporal autocorrelated errors through an AR(1) process, $u_{2,it}$, and independent sampling errors, ε_{it}. Furthermore, we have assumed that the autoregressive coefficients are known, due to difficulties on getting efficient and admissible estimators for ρ, as

was reported by Fuller (1987) and by Rao & Yu (1994). Thus, from this point forward we define the vector of variance components as $\boldsymbol{\psi} = (\sigma^2, \sigma_u^2)'$.

We first obtain an unbiased estimator of σ^2. For this purpose, model (2.5) is transformed in order to eliminate the vector of random area specific effects $\mathbf{v}$. First transform $\mathbf{y}_i$ to $\mathbf{z}_i = \mathbf{P}\mathbf{y}_i$ such that the covariance matrix $V(\mathbf{P}\mathbf{u}_{2i}) = \sigma^2\mathbf{I}_T$. In this Prais–Winsten transformation we use the decomposition $\boldsymbol{\Gamma} = \mathbf{P}^{-1}(\mathbf{P}^{-1})'$, where $\mathbf{P}(T\times T)$ has the following form: $p_{1,1} = (1-\rho^2)^{1/2}$, $p_{t,t'} = 1, \forall\, t = t'$ for $t, t' = 2, ..., T$, $p_{t+1,t} = -\rho$ for $t = 1, ..., T-1$ and remaining $p_{t,t'} = 0$ (Judge *et al.*, 1985). Pre-multiplying model (2.5) by $\mathrm{diag}_{1\leq i\leq m}(\mathbf{P})$, the transformed model is given by:

$$\mathbf{z}^* = \mathbf{H}^*\boldsymbol{\beta} + \mathrm{diag}_{1\leq i\leq m}(\mathbf{f})\,\mathbf{v} + \mathrm{col}_{1\leq i\leq m}(\mathbf{P}\mathbf{u}_{2i}) + \mathrm{col}_{1\leq i\leq m}(\mathbf{P}\boldsymbol{\epsilon}_i)\ , \tag{2.8}$$

where $\mathbf{z}^* = \mathrm{col}_{1\leq i\leq m}(\mathbf{z}_i)$, $\mathbf{H}^* = \mathrm{col}_{1\leq i\leq m}(\mathbf{H}_i)$, $\mathbf{z}_i = \mathbf{P}\mathbf{y}_i$, $\mathbf{H}_i = \mathbf{P}\mathbf{X}_i$ and $\mathbf{f} = \mathbf{P}\mathbf{1}_T = \mathrm{col}_{1\leq i\leq m}(f_t)$, with $f_1 = (1-\rho^2)^{1/2}$ and $f_t = 1-\rho$ for $2 \leq t \leq T$. Next we transform $\mathbf{z}_i$ to $\mathbf{z}_i^{(1)} = (\mathbf{I}_T - \mathbf{D})\,\mathbf{z}_i$, where $\mathbf{D} = (\mathbf{f}\mathbf{f}')/c$ is a $T\times T$ matrix with $c = \mathbf{f}'\mathbf{f} = (1-\rho)\big[T-(T-2)\rho\big]$. Pre-multiplying now model (2.8) by $\mathbf{D}^* = \mathrm{diag}_{1\leq i\leq m}(\mathbf{I}_T - \mathbf{D})$ and noting that $(\mathbf{I}_T - \mathbf{D})\mathbf{f} = \mathbf{0}_{T\times 1}$, then the transformed model reduces to:

$$\mathbf{z}^{(1)} = \mathbf{H}^{(1)}\boldsymbol{\beta} + \mathbf{e}^{(1)}\ , \tag{2.9}$$

where $\mathbf{z}^{(1)} = \mathrm{col}_{1\leq i\leq m}\big(\mathbf{z}_i^{(1)}\big)$, $\mathbf{H}^{(1)} = \mathrm{col}_{1\leq i\leq m}\big(\mathbf{H}_i^{(1)}\big)$, $\mathbf{H}_i^{(1)} = (\mathbf{I}_T - \mathbf{D})\mathbf{P}\mathbf{X}_i$ and $\mathbf{e}^{(1)} = \mathrm{col}_{1\leq i\leq m}\big[(\mathbf{I}_T - \mathbf{D})\,\mathbf{P}(\mathbf{u}_{2i} + \boldsymbol{\epsilon}_i)\big]$. Further, we can see at this moment that $E\big(\mathbf{e}^{(1)}\big) = \mathbf{0}_{mT\times 1}$ and $V\big(\mathbf{e}^{(1)}\big) = \mathrm{diag}_{1\leq i\leq m}\big[(\mathbf{I}_T - \mathbf{D})\,(\sigma^2\mathbf{I}_T + \mathbf{P}\mathbf{R}_i\mathbf{P}')\,(\mathbf{I}_T - \mathbf{D})'\big]$ do not involve σ_u^2, thus we can estimate σ^2 through the reduced model (2.9) using the residual sum of squares. Let $\widehat{\mathbf{e}}^{(1)\prime}\widehat{\mathbf{e}}^{(1)}$ be the residual sum of squares obtained by regressing $\mathbf{z}^{(1)}$ on $\mathbf{H}^{(1)}$ using ordinary least squares (OLS). An unbiased estimator of σ^2 is given by:

$$\widetilde{\sigma}^2 = \left(\widehat{\mathbf{e}}^{(1)\prime}\widehat{\mathbf{e}}^{(1)} - \mathrm{tr}\left\{\left[\mathbf{D}^* - \mathbf{H}^{(1)}\big(\mathbf{H}^{(1)\prime}\mathbf{H}^{(1)}\big)^{-}\mathbf{H}^{(1)\prime}\right]\mathbf{R}^{(1)}\right\}\right)\Big[m(T-1) - r(\mathbf{H}^{(1)})\Big]^{-1}, \tag{2.10}$$

where $\mathbf{R}^{(1)} = \mathrm{diag}_{1\leq i\leq m}(\mathbf{P}\mathbf{R}_i\mathbf{P})$ and $\mathbf{A}^-$ is the Moore–Penrose generalized inverse of $\mathbf{A}$. Although this estimator has been deducted in the context of the spatio-temporal proposed model, it is equal to the estimator proposed by Rao & Yu (1994) in the context of their temporal model. This happens due to the fact that we have transformed model (2.5) in order to eliminate the vector of random area specific effects $\mathbf{v}$, which accounts for the spatial dependence among small areas. Rao & Yu (1994) showed that estimator (2.10) is unbiased and asymptotically consistent.

Turning to the estimation of σ_u^2, we transform $\mathbf{z}_i$ to $z_i^{(2)} = c^{-1/2}\,\mathbf{f}'\mathbf{z}_i$ such that $u_{2i}^{(2)} = c^{-1/2}\,\mathbf{f}'\mathbf{P}\mathbf{u}_{2i}$ has mean 0 and variance σ^2. Pre-multiplying model (2.8) by $\mathrm{diag}_{1\leq i\leq m}(c^{-1/2}\,\mathbf{f}')$ and noting that $c^{-1/2}\,\mathbf{f}'\mathbf{f} = c^{1/2}$, we obtain the following

transformed model:

$$\mathbf{z}^{(2)} = \mathbf{H}^{(2)}\boldsymbol{\beta} + \mathbf{e}^{(2)} , \tag{2.11}$$

where $\mathbf{z}^{(2)} = \mathrm{col}_{1\leq i\leq m}(c^{-1/2}\mathbf{f}'\mathbf{z}_i)$, $\mathbf{H}^{(2)} = \mathrm{col}_{1\leq i\leq m}(c^{-1/2}\mathbf{f}'\mathbf{H}_i)$ and $\mathbf{e}^{(2)} = c^{1/2}\mathbf{v} + \mathbf{u}_2^{(2)} + \boldsymbol{\epsilon}^{(2)}$, in which $\mathbf{u}_2^{(2)} = \mathrm{col}_{1\leq i\leq m}(c^{-1/2}\mathbf{f}'\mathbf{P}\mathbf{u}_{2i})$ and $\boldsymbol{\epsilon}^{(2)} = \mathrm{col}_{1\leq i\leq m}(c^{-1/2}\mathbf{f}'\mathbf{P}\boldsymbol{\epsilon}_i)$. The error term of model (2.11) has $E(\mathbf{e}^{(2)}) = \mathbf{0}_{m\times 1}$ and $V(\mathbf{e}^{(2)}) = c\sigma_u^2\mathbf{B}^{-1} + \sigma^2\mathbf{I}_m + \mathbf{R}^{(2)}$, where $\mathbf{R}^{(2)} = \mathrm{diag}_{1\leq i\leq m}(c^{-1}\mathbf{f}'\mathbf{P}\mathbf{R}_i\mathbf{P}'\mathbf{f})$. Let $\widehat{\mathbf{e}}^{(2)\prime}\widehat{\mathbf{e}}^{(2)}$ be the residual sum of squares obtained by regressing $\mathbf{z}^{(2)}$ on $\mathbf{H}^{(2)}$ using OLS. An unbiased estimator of σ_u^2 is given by:

$$\widetilde{\sigma}_u^2 = \left\{\widehat{\mathbf{e}}^{(2)\prime}\widehat{\mathbf{e}}^{(2)} - \mathrm{tr}\big(\mathbf{P}_{H^{(2)}}\mathbf{R}^{(2)}\big) - \widetilde{\sigma}^2\Big[m - r\big(\mathbf{H}^{(2)}\big)\Big]\right\}\Big/\Big[c\times\mathrm{tr}\big(\mathbf{P}_{H^{(2)}}\mathbf{B}^{-1}\big)\Big] , \tag{2.12}$$

where $\mathbf{P}_{H^{(2)}} = \mathbf{I}_m - \mathbf{H}^{(2)}\big(\mathbf{H}^{(2)\prime}\mathbf{H}^{(2)}\big)^{-}\mathbf{H}^{(2)\prime}$ and $\widetilde{\sigma}^2$ is given by (2.10). The unbiasedness of $\widetilde{\sigma}_u^2$ follows by noting that $E(\widehat{\mathbf{e}}^{(2)\prime}\widehat{\mathbf{e}}^{(2)}) = c\sigma_u^2\,\mathrm{tr}(\mathbf{P}_{H^{(2)}}\mathbf{B}^{-1}) + \sigma^2\big[m - r(\mathbf{H}^{(2)})\big] + \mathrm{tr}(\mathbf{P}_{H^{(2)}}\mathbf{R}^{(2)})$. Since $\widetilde{\sigma}^2$ and $\widetilde{\sigma}_u^2$ can take negative values, we truncate them at zero and use $\widehat{\sigma}^2 = \max\{0, \widetilde{\sigma}^2\}$ and $\widehat{\sigma}_u^2 = \max\{0, \widetilde{\sigma}_u^2\}$. The truncated estimators are no longer unbiased but they are still asymptotically consistent.

A two-stage estimator of θ_{it} can now be obtained from (2.7) by replacing $\boldsymbol{\psi} = (\sigma^2, \sigma_u^2)'$ for $\widehat{\boldsymbol{\psi}} = (\widehat{\sigma}^2, \widehat{\sigma}_u^2)'$:

$$\widehat{\theta}_{it} = \widehat{\theta}_{it}^{H}(\widehat{\boldsymbol{\psi}}) = \mathbf{x}_{it}'\widehat{\boldsymbol{\beta}} + \big(\widehat{\sigma}_u^2\boldsymbol{\varsigma}_i' \otimes \mathbf{1}_T' + \widehat{\sigma}^2\boldsymbol{\zeta}_{it}'\big)\,\widehat{\mathbf{V}}^{-1}\big(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\big) . \tag{2.13}$$

This estimator is not a genuine EBLUP because we have assumed known autoregressive coefficients. An adequate name for estimator (2.13) would be partial EBLUP, because just the variance components were replaced by their estimators. However, for simplicity it is called EBLUP along this manuscript. Following Kackar & Harville (1984) and Rao & Yu (1994), we may note that estimator (2.13) is unbiased as estimators of variance components, (2.10) and (2.12), are even functions of $\mathbf{y}$ and translation-invariant.

3. ESTIMATION OF THE MSPE OF THE TWO-STAGE ESTIMATOR

Under the normality of the random effects and random errors, following the Kackar & Harville (1984) identity and using the Henderson's general result (Henderson, 1975), the MSE of an EBLUP can be decomposed as:

$$MSE\big[\widehat{\theta}_{it}(\widehat{\boldsymbol{\psi}})\big] = g_{1it}(\boldsymbol{\psi}) + g_{2it}(\boldsymbol{\psi}) + E\big[\widehat{\theta}_{it}(\widehat{\boldsymbol{\psi}}) - \widetilde{\theta}_{it}(\boldsymbol{\psi})\big]^2 , \tag{3.1}$$

where E means the expectation with respect to model (2.5), $g_{1it}(\boldsymbol{\psi})$ represents the uncertainty of the EBLUP due to the estimation of the random effects and

is of order $o(1)$, $g_{2it}(\boldsymbol{\psi})$ is due to the estimation of the fixed effects and is of order $o(m^{-1})$, and the last term measures the uncertainty due to the estimation of the variance components. The first two terms can be analytically evaluated, due to the linearity of the EBLUP in the data vector $\mathbf{y}$, from the following closed formulas,

$$g_{1it}(\boldsymbol{\psi}) = \sigma_u^2 \varsigma_{ii} + \frac{\sigma^2}{1-\rho^2} - \left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)' \mathbf{V}^{-1} \left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right) \tag{3.2}$$

and

$$\begin{aligned} g_{2it}(\boldsymbol{\psi}) = & \left[\mathbf{x}_{it} - \mathbf{X}'\mathbf{V}^{-1}\left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)\right]' (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ & \times \left[\mathbf{x}_{it} - \mathbf{X}'\mathbf{V}^{-1}\left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)\right]. \end{aligned} \tag{3.3}$$

However, the third term on the right of equation (3.1) does not have a closed-form expression, due to the non-linearity of the EBLUP in the data vector $\mathbf{y}$, and therefore an approximation is needed.

3.1. Analytical approximation of the MSPE estimator

Following Kackar & Harville (1984) Taylor series approximation and then the lines of Prasad & Rao (1990), we propose the following analytical approximation of the third term of (3.1):

$$E\left[\widehat{\theta}_{it}(\widehat{\boldsymbol{\psi}}) - \widetilde{\theta}_{it}(\boldsymbol{\psi})\right]^2 \approx \operatorname{tr}\left[\mathbf{L}_{it}(\boldsymbol{\psi}) \mathbf{V}(\boldsymbol{\psi}) \mathbf{L}'_{it}(\boldsymbol{\psi}) \overline{\mathbf{V}}(\widehat{\boldsymbol{\psi}})\right] = g_{3it}(\boldsymbol{\psi}) \,, \tag{3.4}$$

where $\mathbf{L}_{it}(\boldsymbol{\psi}) = \frac{\partial \mathbf{b}'_{it}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}$, $\mathbf{V}(\boldsymbol{\psi})$ is given by (2.6) and $\overline{\mathbf{V}}(\widehat{\boldsymbol{\psi}})$ denotes the asymptotic covariance matrix of $\widehat{\boldsymbol{\psi}}$. Using $\mathbf{b}'_{it}(\boldsymbol{\psi}) = \left(\sigma_u^2 \boldsymbol{\varsigma}'_i \otimes \mathbf{1}'_T + \sigma^2 \boldsymbol{\zeta}'_{it}\right)\mathbf{V}^{-1}$ it follows that $\mathbf{L}_{it}(\boldsymbol{\psi}) = \left(\frac{\partial \mathbf{b}_{it}}{\partial \sigma^2}, \frac{\partial \mathbf{b}_{it}}{\partial \sigma_u^2}\right)'$ is a $2 \times mT$ matrix with two blocks given by:

$$\frac{\partial \mathbf{b}'_{it}}{\partial \sigma^2} = \left[\boldsymbol{\zeta}'_{it} - \left(\sigma_u^2 \boldsymbol{\varsigma}'_i \otimes \mathbf{1}'_T + \sigma^2 \boldsymbol{\zeta}'_{it}\right) \mathbf{V}^{-1} \left(\mathbf{I}_m \otimes \boldsymbol{\Gamma}\right)\right] \mathbf{V}^{-1} \tag{3.5}$$

and

$$\frac{\partial \mathbf{b}'_{it}}{\partial \sigma_u^2} = \left[\boldsymbol{\varsigma}'_i \otimes \mathbf{1}'_T - \left(\sigma_u^2 \boldsymbol{\varsigma}'_i \otimes \mathbf{1}'_T + \sigma^2 \boldsymbol{\zeta}'_{it}\right) \mathbf{V}^{-1} \left(\mathbf{Z}_1 \mathbf{B}^{-1} \mathbf{Z}'_1\right)\right] \mathbf{V}^{-1}. \tag{3.6}$$

Let $\mathbf{A}_{it}(\boldsymbol{\psi}) = \mathbf{L}_{it}(\boldsymbol{\psi}) \mathbf{V}(\boldsymbol{\psi}) \mathbf{L}'_{it}(\boldsymbol{\psi}) = \{a_{kl}\}$, thus it follows from previous results that it is a 2×2 symmetric matrix with elements:

$$\begin{aligned} a_{11} = & \left[\boldsymbol{\zeta}_{it} - \left(\mathbf{I}_m \otimes \boldsymbol{\Gamma}\right) \mathbf{V}^{-1} \left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)\right]' \mathbf{V}^{-1} \\ & \times \left[\boldsymbol{\zeta}_{it} - \left(\mathbf{I}_m \otimes \boldsymbol{\Gamma}\right) \mathbf{V}^{-1} \left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)\right], \end{aligned} \tag{3.7}$$

$$\begin{aligned} a_{22} = & \left[\boldsymbol{\varsigma}_i \otimes \mathbf{1}_T - \left(\mathbf{Z}_1 \mathbf{B}^{-1} \mathbf{Z}'_1\right) \mathbf{V}^{-1} \left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)\right]' \mathbf{V}^{-1} \\ & \times \left[\boldsymbol{\varsigma}_i \otimes \mathbf{1}_T - \left(\mathbf{Z}_1 \mathbf{B}^{-1} \mathbf{Z}'_1\right) \mathbf{V}^{-1} \left(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\right)\right] \end{aligned} \tag{3.8}$$

and

$$a_{12} = a_{21} = \Big[\boldsymbol{\zeta}_{it} - (\mathbf{I}_m \otimes \boldsymbol{\Gamma})\,\mathbf{V}^{-1}\big(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\big)\Big]' \mathbf{V}^{-1} \times \Big[\boldsymbol{\varsigma}_i \otimes \mathbf{1}_T - (\mathbf{Z}_1 \mathbf{B}^{-1} \mathbf{Z}_1')\,\mathbf{V}^{-1}\big(\sigma_u^2 \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \sigma^2 \boldsymbol{\zeta}_{it}\big)\Big] . \tag{3.9}$$

It remains to obtain the elements of $\overline{\mathbf{V}}(\widehat{\boldsymbol{\psi}})$ in expression (3.4). Following the lines of Rao & Yu (1994), we propose the evaluation of those elements using a lemma on the covariance of two quadratic forms of normally distributed variables (Jiang, 2007, p. 238). For this purpose we have to write $\widetilde{\sigma}^2$ and $\widetilde{\sigma}_u^2$ as quadratic forms of normally distributed variables. These estimators can be written as:

$$\widetilde{\sigma}^2 = k_1\, \mathbf{a}'\mathbf{C}_1\mathbf{a} + k_2 \tag{3.10}$$

and

$$\widetilde{\sigma}_u^2 = k_3\, \mathbf{a}'\mathbf{C}_2\mathbf{a} + k_4\, \mathbf{a}'\mathbf{C}_1\mathbf{a} + k_5 \ , \tag{3.11}$$

where $k_1 = \big[m(T-1) - r(\mathbf{H}^{(1)})\big]^{-1}$, $k_2 = -\operatorname{tr}\big\{\big[\mathbf{D}^* - \mathbf{H}^{(1)}(\mathbf{H}^{(1)\prime}\mathbf{H}^{(1)})^-\,\mathbf{H}^{(1)\prime}\big]\,\mathbf{R}^{(1)}\big\}$ $\big[m(T-1) - r(\mathbf{H}^{(1)})\big]^{-1}$, $k_3 = \big[c \times \operatorname{tr}(\mathbf{P}_{H^{(2)}}\mathbf{B}^{-1})\big]^{-1}$, $k_4 = -k_1\big[m - r(\mathbf{H}^{(2)})\big]\big[c \times \operatorname{tr}(\mathbf{P}_{H^{(2)}}\mathbf{B}^{-1})\big]^{-1}$ and $k_5 = \big\{-k_2\big[m - r(\mathbf{H}^{(2)})\big] - \operatorname{tr}(\mathbf{P}_{H^{(2)}}\mathbf{R}^{(2)})\big\}\,\big[c \times \operatorname{tr}(\mathbf{P}_{H^{(2)}}\mathbf{B}^{-1})\big]^{-1}$ are constants. Furthermore $\mathbf{a} = \mathbf{Z}_1\mathbf{v} + \mathbf{u}_2 + \boldsymbol{\epsilon} \sim N(\mathbf{0}; \mathbf{V})$, $\mathbf{C}_1 = \mathbf{C}^{(1)\prime}\big[\mathbf{I}_{mT} - \mathbf{C}^{(1)}\mathbf{X}(\mathbf{X}'\mathbf{C}^{(1)\prime}\mathbf{C}^{(1)}\mathbf{X})^-\,\mathbf{X}'\mathbf{C}^{(1)\prime}\big]\mathbf{C}^{(1)}$ and $\mathbf{C}_2 = \mathbf{C}^{(2)\prime}\big[\mathbf{I}_m - \mathbf{C}^{(2)}\mathbf{X}(\mathbf{X}'\mathbf{C}^{(2)\prime}\mathbf{C}^{(2)}\mathbf{X})^-\,\mathbf{X}'\mathbf{C}^{(2)\prime}\big]\mathbf{C}^{(2)}$ are symmetric matrices with $\mathbf{C}^{(1)} = \operatorname{diag}_{1\le i\le m}\big[(\mathbf{I}_T - \mathbf{D})\mathbf{P}\big]$ and $\mathbf{C}^{(2)} = \operatorname{diag}_{1\le i\le m}(c^{-1/2}\mathbf{f}'\mathbf{P})$, respectively. Let $\overline{\mathbf{V}}(\widehat{\boldsymbol{\psi}}) \equiv \mathbf{D} = \{d_{kl}\}$, thus it follows from previous results that it is a 2×2 symmetric matrix with elements:

$$d_{11} = V(\widetilde{\sigma}^2) = 2\,k_1^2 \operatorname{tr}(\mathbf{C}_1\mathbf{V}\mathbf{C}_1\mathbf{V}) \ , \tag{3.12}$$

$$d_{22} = V(\widetilde{\sigma}_u^2) = 2\,k_3^2 \operatorname{tr}(\mathbf{C}_2\mathbf{V}\mathbf{C}_2\mathbf{V}) + 4k_3k_4 \operatorname{tr}(\mathbf{C}_1\mathbf{V}\mathbf{C}_2\mathbf{V}) + 2\,k_4^2 \operatorname{tr}(\mathbf{C}_1\mathbf{V}\mathbf{C}_1\mathbf{V}) \tag{3.13}$$

and

$$d_{12} = d_{21} = \operatorname{Cov}(\widetilde{\sigma}^2; \widetilde{\sigma}_u^2) = 2\,k_1k_3 \operatorname{tr}(\mathbf{C}_1\mathbf{V}\mathbf{C}_2\mathbf{V}) + 2\,k_1k_4 \operatorname{tr}(\mathbf{C}_1\mathbf{V}\mathbf{C}_1\mathbf{V}) \, . \tag{3.14}$$

Following Prasad & Rao (1990), it can be assumed that $E\big[g_{1it}(\widehat{\boldsymbol{\psi}}) + g_{2it}(\widehat{\boldsymbol{\psi}})\big] = g_{1it}(\boldsymbol{\psi}) + g_{2it}(\boldsymbol{\psi}) - g_{3it}(\boldsymbol{\psi})$. Thus it follows that a bias corrected analytical estimator of MSPE of two-stage estimator is given by:

$$mspe^A\big[\widehat{\theta}_{it}(\widehat{\boldsymbol{\psi}})\big] = g_{1it}(\widehat{\boldsymbol{\psi}}) + g_{2it}(\widehat{\boldsymbol{\psi}}) + 2\,g_{3it}(\widehat{\boldsymbol{\psi}}) \ . \tag{3.15}$$

3.2. Bootstrap approximation of the MSPE estimator

In this section we introduce an alternative way of approximating the MSPE of the EBLUP by a simple bootstrap procedure using similar ideas to Butar &

Lahiri (2003). Hereafter we describe a bootstrap procedure designed for estimating the MSPE under the spatio-temporal small area model (2.5), and using ANOVA estimates of variance components introduced in Section 2.3. In general, the parametric bootstrap method consists of generating parametrically a large number of area bootstrap samples from the model fitted to the original data, re-estimating the model parameters for each bootstrap sample and then estimating the separate components of the MSPE. Assuming that ρ and ϕ are known, the parametric bootstrap procedure follows the next steps:

1. Calculate estimates of the variance components, $\widehat{\sigma}^2$ and $\widehat{\sigma}_u^2$, from the initial data, $\mathbf{y}$, using the method of moments. Then fit model (2.5) in order to estimate the fixed effects, $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\mathbf{y}; \widehat{\boldsymbol{\psi}})$ with $\widehat{\boldsymbol{\psi}} = (\widehat{\sigma}^2, \widehat{\sigma}_u^2)'$.
2. Compute the EBLUP estimates of θ_{it}, $\widehat{\theta}_{it}(\widehat{\boldsymbol{\psi}})$, and the first two terms of MSPE, $g_{1it}(\widehat{\boldsymbol{\psi}})$ and $g_{2it}(\widehat{\boldsymbol{\psi}})$.
3. Generate m independent copies of a variable $\mathbf{u}_1^*$, with $\mathbf{u}_1^* \sim N(\mathbf{0}; \widehat{\sigma}_u^2 \mathbf{I}_m)$. From this values, construct the random vector $\mathbf{v}^* = (\mathbf{I}_m - \phi \mathbf{W})^{-1} \mathbf{u}_1^*$, assuming that ϕ is known.
4. Generate mT independent copies of a variable $\boldsymbol{\xi}^*$, with $\boldsymbol{\xi}^* \sim N(\mathbf{0}; \widehat{\sigma}^2 \mathbf{I}_{mT})$, independently of the generation of $\mathbf{u}_1^*$. From this values, construct the random vector $\mathbf{u}_2^*$, assuming that ρ is known.
5. Generate mT independent copies of a variable $\boldsymbol{\epsilon}^*$, with $\boldsymbol{\epsilon}^* \sim N(\mathbf{0}; \mathbf{R})$, independently of the generation of $\mathbf{u}_1^*$ and $\boldsymbol{\xi}^*$.
6. Construct the bootstrap data $\mathbf{y}^* = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\upsilon}^* + \boldsymbol{\epsilon}^*$ where $\boldsymbol{\upsilon}^* = [\mathbf{v}^{*\prime} \mathbf{u}_2^{*\prime}]'$.
7. Calculate bootstrap estimates of the variance components, $\widehat{\sigma}^{2*}$ and $\widehat{\sigma}_u^{2*}$, from the bootstrap data, $\mathbf{y}^*$, and then fit model (2.5) in order to obtain bootstrap estimates of the fixed effects, $\widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}}(\mathbf{y}; \widehat{\boldsymbol{\psi}}^*)$ with $\widehat{\boldsymbol{\psi}}^* = (\widehat{\sigma}^{2*}, \widehat{\sigma}_u^{2*})'$.
8. Calculate bootstrap estimates of the EBLUP, as well as estimates of the two components of the MSPE of BLUP, using bootstrap estimates of the variance components, $\widehat{\boldsymbol{\psi}}^*$:

$$\widehat{\theta}_{it}^* = \widehat{\theta}_{it}^*(\mathbf{y}; \widehat{\boldsymbol{\beta}}^*; \widehat{\boldsymbol{\psi}}^*) = \mathbf{x}_{it}' \widehat{\boldsymbol{\beta}}^* + \left(\widehat{\sigma}_u^{2*} \boldsymbol{\varsigma}_i' \otimes \mathbf{1}_T' + \widehat{\sigma}^{2*} \boldsymbol{\zeta}_{it}'\right) \left[\widehat{\mathbf{V}}(\widehat{\boldsymbol{\psi}}^*)\right]^{-1} \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^*\right) ,$$

$$\begin{aligned} g_{1it}^* = g_{1it}^*(\mathbf{y}; \widehat{\boldsymbol{\beta}}^*; \widehat{\boldsymbol{\psi}}^*) = {} & \widehat{\sigma}_u^{2*} \varsigma_{ii} + \frac{\widehat{\sigma}^{2*}}{1-\rho^2} \\ & - \left(\widehat{\sigma}_u^{2*} \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \widehat{\sigma}^{2*} \boldsymbol{\zeta}_{it}\right)' \left(\widehat{\mathbf{V}}^*\right)^{-1} \left(\widehat{\sigma}_u^{2*} \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \widehat{\sigma}^{2*} \boldsymbol{\zeta}_{it}\right) , \end{aligned}$$

$$\begin{aligned} g_{2it}^* = g_{2it}^*(\mathbf{y}; \widehat{\boldsymbol{\beta}}^*; \widehat{\boldsymbol{\psi}}^*) = {} & \left[\mathbf{x}_{it} - \mathbf{X}'(\widehat{\mathbf{V}}^*)^{-1} \left(\widehat{\sigma}_u^{2*} \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \widehat{\sigma}^{2*} \boldsymbol{\zeta}_{it}\right)\right]' \\ & \times \left[\mathbf{X}'(\widehat{\mathbf{V}}^*)^{-1} \mathbf{X}\right]^{-1} \left[\mathbf{x}_{it} - \mathbf{X}'(\widehat{\mathbf{V}}^*)^{-1} \left(\widehat{\sigma}_u^{2*} \boldsymbol{\varsigma}_i \otimes \mathbf{1}_T + \widehat{\sigma}^{2*} \boldsymbol{\zeta}_{it}\right)\right] . \end{aligned}$$

9. Repeat steps 3–8 B times. Let $\widehat{\sigma}^{2*(b)}$ and $\widehat{\sigma}_u^{2*(b)}$ be the bootstrap estimates of variance components, $\widehat{\boldsymbol{\psi}}^{*(b)} = (\widehat{\sigma}^{2*(b)}, \widehat{\sigma}_u^{2*(b)})'$; and $\widehat{\boldsymbol{\beta}}^{*(b)}$, $\widehat{\theta}_{it}^{*(b)}$, $g_{1it}^{*(b)}$ and $g_{2it}^{*(b)}$ the bootstrap estimates of $\boldsymbol{\beta}$, θ_{it}, g_{1it} and g_{2it}, respectively, all of them obtained in the b^{th} bootstrap replication, $b = 1, ..., B$.

10. Calculate a bootstrap estimate of g_{3it} using the following Monte Carlo approximation:

$$g_{3it}^{*} = B^{-1} \sum_{b=1}^{B} \big(\widehat{\theta}_{it}^{*(b)} - \widehat{\theta}_{it}\big)^2 .$$

Since it is known that the quantity $g_{1it}(\widehat{\boldsymbol{\psi}}) + g_{2it}(\widehat{\boldsymbol{\psi}})$ is a biased estimator of $g_{1it}(\boldsymbol{\psi}) + g_{2it}(\boldsymbol{\psi})$ (Prasad & Rao, 1990), thus following the lines of Butar & Lahiri (2003), a bias corrected bootstrap estimator of MSPE of the two-stage estimator can be defined as:

$$mspe^{B}\big[\widehat{\theta}_{it}(\widehat{\boldsymbol{\psi}})\big] = 2\big[g_{1it}(\widehat{\boldsymbol{\psi}}) + g_{2it}(\widehat{\boldsymbol{\psi}})\big] - B^{-1}\sum_{b=1}^{B}\big[g_{1it}^{*(b)} + g_{2it}^{*(b)}\big] + g_{3it}^{*} \ . \tag{3.16}$$

4. A MONTE CARLO SIMULATION STUDY

In order to assess the merits of our spatio-temporal estimator, in this section we present a simulation study designed for comparing the efficiency of the proposed EBLUP estimator (ST) against other well-known EBLUP estimators, such as the Fay–Herriot (FH), the Salvati (NS) and the Rao–Yu (RY) estimators. The first is one of the paradigms in small area estimation exploring neither spatial nor chronological similarities; the Salvati estimator is a well-known small-area estimator that explores spatial similarities in data; and the last one is a reference for small area estimation with chronological correlation. This simulation study also aims to study the accuracy of the proposed estimators of MSPE of the EBLUP.

In the simulated experiments we assume that the proposed EBLUP estimator performs better than the others when the spatio-temporal model provides a good fit. Thus we have decided to generate an artificial population of $\mathbf{y}$-values using model (2.5). We have considered model (2.5) with $p = 2$, that is, a constant and one explanatory variable, $\mathbf{x}_{it} = (1, x_{it})'$. The mT values of x_{it} were generated from a uniform distribution in the interval [0,1]. The true model coefficients were $\boldsymbol{\beta} = (1, 2)'$, the random area specific effects variance were $\sigma_u^2 \in \{0.5; 1.0\}$, the random area-by-time specific effects variance were $\sigma^2 \in \{0.25; 0.50; 1.00\}$, the temporal autoregressive coefficients were $\rho \in \{0.2; 0.4; 0.8\}$ and the spatial autoregressive coefficients were $\phi \in \{0.25; 0.50; 0.75\}$. The row standardized $\mathbf{W}$ matrix was kept fixed in all simulations and it corresponded to contiguous NUTSIII in

one European country (NUTSIII is a geocode standard for referencing the subdivisions of European countries for statistical purposes). We have selected $T = 7$ and $m = 28$ divided into four uniform groups. Further, we have considered three different σ^2_{it} patterns (see Table 1). Note that sampling variances are more dispersed in pattern (C) than in pattern (A).

Table 1: Values of the sampling variances for the simulation study.

i	**1–7**	**8–14**	**15–21**	**22–28**
Pattern (A)	1.0	1.0	1.0	1.0
Pattern (B)	0.6	0.8	1.1	1.5
Pattern (C)	0.3	0.6	1.0	2.0

In addition, we have assumed known ρ and ϕ, although the bootstrap method can also accommodate the situation of unknown coefficients. The random effects and errors were generated independently from a Normal distribution with zero mean. Finally, the vector of mT values of the interest variable, $\mathbf{y}$, was generated from the cross-sectional and time-series stationary small area model (2.5).

4.1. The efficiency of the EBLUP

The simulation study designed for comparing the efficiency of the proposed EBLUP estimator over other well-known EBLUP estimators has followed the algorithm described below:

1. Generate $L = 1{,}000$ samples of initial data, $\mathbf{y}^{(l)} = (y^{(l)}_{11}, ..., y^{(l)}_{it}, ..., y^{(l)}_{mT})$, as described above, $l = 1, ..., L$.
2. Fit the Fay–Herriot (FH), the Salvati (NS) and the Rao–Yu (RY) models, as well as the spatio-temporal (ST) model, to the initial data, $\mathbf{y}^{(l)}$, using method of moments estimates of those variance components, for each $l = 1, ..., L$.
3. Compute the EBLUP estimates under each of those models, $\widehat{\theta}^{a}_{it}$, $a \in$ {FH,NS, RY,ST}.
4. Calculate the Monte Carlo approximation of the relative efficiency (RE) of the proposed EBLUP over the other three EBLUP estimators as follows: $RE_{it} = \frac{MSE(\widehat{\theta}^{a}_{it})}{MSE(\widehat{\theta}^{ST}_{it})} \times 100$, where $MSE(\widehat{\theta}^{a}_{it}) = L^{-1} \sum_{l=1}^{L} \left(\widehat{\theta}^{a(l)}_{it} - y^{a(l)}_{it}\right)^2$ is the empirical MSE of each EBLUP, $a \in$ {FH,NS,RY,ST}. That measure of efficiency is calculated at area-by-time level. To summarize results, we have computed an average global measure over the mT small areas: $ARE = \frac{1}{mT} \sum_{i=1}^{m} \sum_{t=1}^{T} RE_{it}$.

These results are shown in Tables 2 to 4 for the sampling variance patterns (A), (B) and (C), respectively.

Table 2: Average relative efficiency of the proposed EBLUP over other EBLUP estimators — σ^2_{it} pattern (A).

σ^2	σ^2_u	$\phi = 0.25$			$\phi = 0.50$			$\phi = 0.75$		
		ρ=0.2	ρ=0.4	ρ=0.8	ρ=0.2	ρ=0.4	ρ=0.8	ρ=0.2	ρ=0.4	ρ=0.8
FH estimator										
0.25	0.50	130	133	151	136	138	156	156	159	177
	1.00	161	165	185	172	176	196	213	217	237
0.50	0.50	126	132	168	130	136	173	149	155	192
	1.00	153	160	199	162	169	209	197	205	246
1.00	0.50	120	129	186	124	133	190	138	147	206
	1.00	140	151	211	148	158	219	175	186	249
NS estimator										
0.25	0.5	130	133	150	135	138	156	153	156	176
	1.0	161	165	185	171	175	195	209	213	235
0.50	0.5	126	132	168	130	136	172	146	152	191
	1.0	153	160	199	161	169	209	194	202	245
1.00	0.5	120	129	186	123	132	190	135	145	206
	1.0	140	150	211	147	157	219	172	183	249
RY estimator										
0.25	0.5	101	101	101	105	105	104	122	121	119
	1.0	102	102	102	109	109	108	136	135	132
0.50	0.5	101	101	101	105	104	104	119	118	117
	1.0	102	102	102	108	108	108	132	131	129
1.00	0.5	101	101	101	104	104	105	116	116	117
	1.0	101	101	102	107	107	108	128	128	127

To begin with, it is interesting to note from Tables 2 to 4 that there is not much difference in efficiency among the three sampling variance patterns, which agrees with the results obtained by Pratesi & Salvati (2009) under a model with spatially correlated random effects. In addition, it should be emphasized that the gain in efficiency of the proposed EBLUP over other estimators slightly increases with the dispersion of the sampling variances. Note that the efficiency of the proposed estimator over the Fay–Herriot and the Salvati estimators increases mainly for higher values of the temporal autocorrelation coefficient, while the efficiency of that estimator over the Rao–Yu estimator tends to increase for higher values of the spatial and temporal autocorrelation coefficients.

Comparing the ARE, among the three sampling variance patterns (see Tables 2 to 4), we observe that substantial gains in efficiency are achieved when it is used the proposed EBLUP over other EBLUP estimators, especially over the

Table 3: Average relative efficiency of the proposed EBLUP over other EBLUP estimators — σ^2_{it} pattern (B).

σ^2	σ^2_u	$\phi = 0.25$			$\phi = 0.50$			$\phi = 0.75$		
		ρ=0.2	ρ=0.4	ρ=0.8	ρ=0.2	ρ=0.4	ρ=0.8	ρ=0.2	ρ=0.4	ρ=0.8
FH estimator										
0.25	0.5	130	133	151	135	138	157	157	161	180
	1.0	163	167	189	175	179	201	219	224	247
0.50	0.5	124	131	171	130	136	177	149	156	199
	1.0	154	162	206	164	172	217	201	210	258
1.00	0.5	119	129	191	123	132	195	137	147	211
	1.0	140	151	217	148	159	226	176	188	258
NS estimator										
0.25	0.5	129	133	151	134	138	157	154	158	179
	1.0	163	167	189	174	178	201	216	221	245
0.50	0.5	124	131	171	129	136	177	146	154	198
	1.0	153	161	206	163	171	217	198	207	257
1.00	0.5	119	128	191	122	132	195	135	146	211
	1.0	140	151	217	147	158	226	173	186	257
RY estimator										
0.25	0.5	101	101	101	106	106	105	124	124	121
	1.0	102	102	102	111	110	110	140	139	136
0.50	0.5	101	101	101	106	106	105	122	121	119
	1.0	102	102	102	110	109	109	135	135	132
1.00	0.5	101	101	101	105	105	106	118	118	118
	1.0	102	102	102	108	108	109	130	130	129

Fay–Herriot estimator and over the Salvati estimator. On average the gains in efficiency over these estimators are about 69% for pattern (A), 71% for pattern (B) and 78% for pattern (C). There are also some gains in efficiency of the proposed EBLUP over the Rao–Yu estimator, although they are smaller than the ones observed against the other two estimators (overall average gain in efficiency is equal to 11%, 12% and 13%, respectively, for patterns (A), (B) and (C). Note that the ARE values of the proposed EBLUP over the Rao–Yu estimator are negligible for $\phi = 0.25$ and $\phi = 0.50$. However, for $\phi = 0.75$ the proposed EBLUP performs clearly better than the Rao–Yu estimator in terms of efficiency. Let us also observe from Tables 2 to 4 that gains in efficiency tend to increase with σ^2_u, ρ and ϕ, i.e., the higher values of these three parameters the stronger gains in efficiency of the proposed estimator over the others. Nevertheless, there is a slight decrease on ARE values with the increase of σ^2. It is also worth noting that the proposed EBLUP shows significant gains in efficiency over the NS estimator even for small ρ and that these gains increase with the increase of ϕ. This means that the introduction of the chronological autocorrelation in small area estimation models has a better effect in the efficiency of the estimators than the

Table 4: Average relative efficiency of the proposed EBLUP over other EBLUP estimators — σ^2_{it} pattern (C).

σ^2	σ^2_u	$\phi = 0.25$			$\phi = 0.50$			$\phi = 0.75$		
		ρ=0.2	ρ=0.4	ρ=0.8	ρ=0.2	ρ=0.4	ρ=0.8	ρ=0.2	ρ=0.4	ρ=0.8
FH estimator										
0.25	0.5	123	127	152	129	134	160	156	161	191
	1.0	165	172	204	181	188	222	238	247	285
0.50	0.5	118	127	183	124	134	191	147	158	220
	1.0	154	166	230	166	179	246	211	227	302
1.00	0.5	113	125	203	118	130	209	134	147	229
	1.0	138	151	236	146	161	247	177	194	286
NS estimator										
0.25	0.5	123	127	152	129	133	160	153	159	190
	1.0	165	172	204	180	187	221	235	244	283
0.50	0.5	117	127	183	123	133	191	144	156	219
	1.0	154	166	230	165	179	246	208	224	301
1.00	0.5	113	125	203	117	129	209	131	145	229
	1.0	138	151	236	146	160	247	174	191	286
RY estimator										
0.25	0.5	98	98	98	105	105	104	129	129	127
	1.0	100	100	100	111	111	111	150	149	145
0.50	0.5	99	98	99	105	105	105	127	127	123
	1.0	100	100	100	110	110	110	144	143	138
1.00	0.5	99	99	100	105	105	106	123	122	121
	1.0	100	100	101	109	109	109	135	135	132

introduction of the spatial correlation. All these results confirm the superiority of the proposed spatio-temporal model in comparison to other well-known small area models when data exhibits both spatial and chronological correlations with the considered structure. Finally, it is important to note that our findings are consistent with the simulation results obtained by Singh *et al.* (2005) when it is assumed $\phi = 0.75$ and $\rho = 0.50$. Therefore, the spatial dependence and the temporal autocorrelation should be exploited to strengthen the small area estimates, no matter it is assumed or not an interaction between the autoregressive coefficients.

4.2. The accuracy of the MSPE of the EBLUP

The simulation study designed for studying the accuracy of the proposed estimators of the MSPE of the EBLUP has followed the algorithm described below:

1. Generate $L=1{,}000$ samples of initial data, $\mathbf{y}^{(l)}=(y_{11}^{(l)},...,y_{it}^{(l)},...,y_{mT}^{(l)})$, as described above, $l=1,...,L$.
2. From the initial data, $\mathbf{y}^{(l)}$, calculate estimates of the variance components, $\sigma^{2(l)}$ and $\sigma_u^{2(l)}$, using the method of moments, and then fit model (2.5) in order to estimate the fixed effects $\widehat{\boldsymbol{\beta}}^{(l)}=\widehat{\boldsymbol{\beta}}(\mathbf{y}^{(l)};\widehat{\boldsymbol{\psi}}^{(l)})$, where $\widehat{\boldsymbol{\psi}}^{(l)}=(\widehat{\sigma}^{2(l)},\widehat{\sigma}_u^{2(l)},\rho,\phi)'$, for each $l=1,...,L$.
3. Compute the EBLUP estimates, $\widehat{\theta}_{it}^{(l)}(\widehat{\boldsymbol{\psi}}^{(l)})$ and their analytic MSPE estimates (A-MSPE), $mspe^{A(l)}(\widehat{\theta}_{it}^{(l)})$, for each $l=1,...,L$.
4. Generate B bootstrap data sets as described in Section 3.2 from estimates $\sigma^{2(l)}$ and $\sigma_u^{2(l)}$, and then compute the bootstrap MSPE estimates: $mspe^{B(l)}$ $(\widehat{\theta}_{it}^{(l)})$ (B-MSPE), for each $l=1,...,L$.
5. Compute the empirical values of MSPE for each i^{th} small-area at t^{th} time point, $MSPE_{it}$, which are the benchmark values, with $R=5,000$ independent data sets in order to ensure high precision.
6. Calculate the Monte Carlo approximations of each MSPE estimative of the EBLUP ($mspe$), their relative bias (RB) and relative MSE (RMSE) as follows: $mspe_{it}^{b}=L^{-1}\sum_{l=1}^{L}mspe_{it}^{b(l)}$, $RB_{it}=L^{-1}\sum_{l=1}^{L}\frac{mspe_{it}^{b(l)}-MSPE_{it}}{MSPE_{it}}\times 100$ and $RMSE_{it}=L^{-1}\sum_{l=1}^{L}\frac{(mspe_{it}^{b(l)}-MSPE_{it})^2}{MSPE_{it}}\times 100$, where $b\in\{A,B\}$ denotes the different MSPE estimators. These measures are calculated at area-by-time level. To summarize results, we have produced three global measures over the mT small areas: the percentage of areas where the relative bias is negative (RBN), the average of absolute relative bias (AARB), $AARB=\frac{1}{mT}\sum_{i=1}^{m}\sum_{t=1}^{T}|RB_{it}|$ and the average relative MSE (ARMSE), $ARMSE=\frac{1}{mT}\sum_{i=1}^{m}\sum_{t=1}^{T}RMSE_{it}$.

Tables 5 to 7 report the percent RBN, percent AARB and percent ARMSE of the analytical and the bootstrap MSPE estimators of the EBLUP, for $\rho=0.2$, $\rho=0.4$ and $\rho=0.8$, respectively. The variance components patterns have a significant effect on the performance of both MSPE estimators, unlike the spatial autoregressive coefficient. For that reason, we have decided to report results only for $\phi=0.25$ and $\phi=0.50$. Note that Molina *et al.* (2009) showed, under a model with spatially correlated random effects, that the level of spatial dependence does not significantly affect the performance of the analytical and parametric bootstrap MSPE estimators.

The results in Table 5 suggest that the bootstrap MSPE estimator can compete with the analytical MSPE estimator in terms of bias and accuracy when $\rho=0.2$. From Table 5 we can see that the analytical estimator presents smaller percent AARB and ARMSE (although of the same order of magnitude) than the resampling-based estimator when $\sigma^2=0.25$ and $\sigma^2=1.00$. On the other hand, the results for those measures reveal that the bootstrap estimator is somewhat

better than the analytical estimator when $\sigma^2 = 0.50$. In Table 5 we can also see that both MSPE estimators tend to underestimate the true MSPE of the EBLUP for the majority of small areas, but this underestimation decreases as long as the variance, σ^2, increases. Furthermore, note that the analytical estimator has slight more negative bias than the other estimator.

Table 5: RBN, ARB and ARMSE of MSPE estimators, $\rho = 0.2$.

σ^2	σ_u^2	A-MSPE	B-MSPE	A-MSPE	B-MSPE
		$\phi = 0.25$		$\phi = 0.50$	
RBN (%)					
0.25	0.5	100.000	100.000	100.000	100.000
	1.0	100.000	100.000	100.000	100.000
0.50	0.5	99.908	99.867	99.908	99.878
	1.0	99.755	99.571	99.755	99.582
1.00	0.5	13.520	12.133	13.959	12.480
	1.0	9.153	8.051	9.296	8.245
AARB (%)					
0.25	0.5	58.674	59.872	58.753	59.995
	1.0	57.213	58.409	56.528	57.788
0.50	0.5	34.770	34.063	34.856	34.169
	1.0	31.991	31.246	32.121	31.421
1.00	0.5	17.287	18.090	17.294	18.124
	1.0	20.449	21.504	20.303	21.313
ARMSE (%)					
0.25	0.5	26.138	27.592	26.261	27.757
	1.0	24.899	26.317	23.984	25.434
0.50	0.5	8.637	8.501	8.696	8.568
	1.0	7.374	7.250	7.442	7.332
1.00	0.5	1.925	2.087	1.933	2.103
	1.0	2.537	2.775	2.516	2.741

Table 6 shows that the analytical MSPE estimator is slightly better than the bootstrap MSPE estimator when $\rho = 0.4$. In fact, the results reported in this table indicate again that there is not much difference on bias and accuracy between these MSPE estimators. Although the analytical estimator is always the best in terms on precision (it has the lowest percent ARMSE), the bootstrap estimator is somewhat better than the analytic estimator in terms of bias (according to AARB measure) when $\sigma^2 = 0.50$. The systematic underestimation of the true MSPE of the EBLUP is also revealed by Table 6.

The results reported in Table 7 suggest that the analytical MSPE estimator is the winner when $\rho = 0.8$. From this table we can see that gains on bias and accuracy are reached when it is used the analytical estimator instead of the

Table 6: RBN, ARB and ARMSE of MSPE estimators, $\rho = 0.4$.

σ^2	σ_u^2	A-MSPE	B-MSPE	A-MSPE	B-MSPE
		$\phi = 0.25$		$\phi = 0.50$	
RBN (%)					
0.25	0.5	100.000	100.000	100.000	100.000
	1.0	100.000	100.000	100.000	100.000
0.50	0.5	99.755	99.592	99.786	99.561
	1.0	99.429	99.010	99.408	99.051
1.00	0.5	26.316	22.939	26.347	22.908
	1.0	21.173	17.602	21.531	17.898
AARB (%)					
0.25	0.5	58.363	60.733	58.417	60.808
	1.0	56.305	58.634	56.385	58.757
0.50	0.5	36.561	36.026	36.595	36.037
	1.0	34.199	33.548	34.293	33.679
1.00	0.5	15.019	15.818	15.107	15.959
	1.0	16.875	18.092	16.853	18.052
ARMSE (%)					
0.25	0.5	26.036	28.725	26.128	28.852
	1.0	23.948	26.509	24.058	26.661
0.50	0.5	9.741	9.829	9.779	9.864
	1.0	8.592	8.658	8.649	8.727
1.00	0.5	1.645	1.812	1.665	1.846
	1.0	2.025	2.294	2.024	2.288

bootstrap estimator in measuring the uncertainty of the EBLUP, especially when $\sigma^2 = 0.25$. Furthermore, it should be noted that the underestimation of the true MSPE of the EBLUP is more pronounced when $\sigma^2 = 1.00$.

Comparing the results reported in Tables 5 to 7, among different levels of spatial and temporal autocorrelation, we can observe that both estimators show similar bias and accuracy in most cases, since they are very close on AARB and ARMSE measures. We can also observe for both estimators that: (i) the spatial autoregressive coefficient does not have a significant impact on the performance of the MSPE estimators; (ii) the performance of both MSPE estimators, in what concerns the AARB and the ARMSE, tends to improve for higher values of the variance components (mainly σ^2); (iii) the number of estimates with positive bias tends to increase for higher values of the variance components; (iv) the analytical estimator shows consistently negative bias in more small areas than the bootstrap estimator; and (v) the gain of accuracy of the analytical estimator over the bootstrap one tends to increase with the strength of chronological autocorrelation. Finally, our results reported in Tables 5 to 7 suggest that both MSPE estimators perform well, especially for higher values of the variance components

Table 7: RBN, ARB and ARMSE of MSPE estimators, $\rho = 0.8$.

σ^2	σ_u^2	A-MSPE	B-MSPE	A-MSPE	B-MSPE
		$\phi = 0.25$		$\phi = 0.50$	
RBN (%)					
0.25	0.5	99.990	99.898	99.980	99.878
	1.0	99.990	99.878	99.980	99.837
0.50	0.5	96.378	94.061	96.051	93.969
	1.0	96.163	93.469	95.990	93.184
1.00	0.5	59.765	54.153	59.571	54.122
	1.0	59.469	53.673	59.306	53.561
AARB (%)					
0.25	0.5	50.048	60.057	49.776	60.065
	1.0	49.143	58.869	49.062	58.844
0.50	0.5	36.395	38.984	36.257	38.966
	1.0	35.730	38.231	35.662	38.218
1.00	0.5	21.665	21.954	21.789	21.964
	1.0	21.666	22.110	21.773	22.173
ARMSE (%)					
0.25	0.5	19.944	29.258	19.774	29.329
	1.0	19.173	28.053	19.131	28.100
0.50	0.5	10.696	12.913	10.636	12.927
	1.0	10.350	12.460	10.320	12.467
1.00	0.5	3.502	3.883	3.555	3.889
	1.0	3.494	3.943	3.539	3.969

(for example, for $\sigma^2 = 1.00$ and $\sigma_u^2 = 0.5$ or $\sigma_u^2 = 1.0$) and for lower values of the temporal autoregressive coefficient, and can be in most cases used to adequately access the accuracy of the proposed EBLUP estimator. Analogous findings were reached by Pereira & Coelho (2010) when comparing the performance of MSPE estimators under a cross-sectional and time-series stationary model.

5. CONCLUSIONS

In this work we have studied the problem of "borrowing information" from related small areas and time periods in order to strengthen the estimators of the small area parameters of interest. In particular, we have proposed a spatio-temporal area level LMM involving spatially correlated and temporally autocorrelated random area effects, using both time-series and cross-sectional data. Under this model, we first obtained a partial EBLUP estimator. We then proposed two estimators of the MSPE of that EBLUP.

In the simulation study we have studied the efficiency of the proposed EBLUP estimator over other well-known EBLUP estimators and we have studied the accuracy of the proposed estimators of MSPE. Our empirical results based on simulated data have shown that the proposed EBLUP estimator can lead to remarkable efficiency gains. This is especially true over both the sectional FH and the spatial NS estimators and when the autocorrelation coefficients are high. It should also be noted that our results have shown mild gains in efficiency from the inclusion of a spatial structure into the Rao–Yu cross-sectional and time-series model, i.e. there are gains in efficiency of the proposed EBLUP over the temporal RY estimator.

Under several simulated scenarios for the variance components and autocorrelation coefficients, our empirical results have revealed that both MSPE estimators perform well. Furthermore, our results indicate that the analytical MSPE estimator performs slightly better than the bootstrap one on bias and precision, although its superiority is not uniform. In particular, it is to be noticed that this gain tend to be more conspicuous when chronological correlation is strong.

We believe that the proposed methodology can provide a useful tool for practitioners working with spatially correlated and temporally autocorrelated data in the context of small area estimation problems. Nonetheless, some of the issues mentioned in this paper require further theoretical work and/or more extensive simulation studies. For example, we have assumed known and positive autoregressive coefficients, but in practice these parameters are unknown and could be negative. A further issue relates to on what happens when random area effects do not follow approximately a SAR process or when random area-by-time effects do not follow an AR(1) process? Our expectation is that whenever spatial and chronological correlations exist, this can be a viable approach due to its simplicity of implementation and the fact of allowing to incorporate all the available information in the estimation process. Nevertheless, additional work is still needed to understand the properties of the partial EBLUP estimator and particularly of its MSPE approximations when spatial or chronological correlation processes significantly departure from the considered ones.

ACKNOWLEDGMENTS

The authors thank the Fundação para a Ciência e a Tecnologia (FCT).

REFERENCES

[1] Anselin, L. (1992). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Boston.

[2] Butar, F.B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small area estimators, *Journal of Statistical Planning and Inference*, **112**, 1–2, 63–76.

[3] Chandra, H.; Salvati, N. and Chambers, R. (2007). Small area estimation for spatially correlated populations - a comparison of direct and indirect model-based methods, *Statistics in Transition – new series*, **8**, 2, 331–350.

[4] Datta, G.S.; Lahiri, P. and Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data, *Journal of Statistical Planning and Inference*, **102**, 1, 83–97.

[5] Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data, *Journal of the American Statistical Association*, **74**, 366, 269–277.

[6] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society A*, **222**, 309–368.

[7] Fuller, W.A. (1987). *Measurement Error Models*, John Wiley & Sons, New York.

[8] Hartley, H.O. and Rao, J.N.K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model, *Biometrika*, **54**, 1–2, 93–108.

[9] Henderson, C.R. (1953). Estimation of variance and covariance components, *Biometrics*, **9**, 2, 226–252.

[10] Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model, *Biometrics*, **31**, 2, 423–447.

[11] Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*, Springer-Verlag, New York.

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

[13] Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models, *Journal of the American Statistical Association*, **79**, 388, 853–862.

[14] Molina, I.; Salvati, N. and Pratesi, M. (2009). Bootstrap for estimating the MSE of the Spatial EBLUP, *Computational Statistics*, **24**, 3, 441–458.

[15] Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal, *Biometrika*, **58**, 3, 545–554.

[16] Pereira, L.N. and Coelho, P.S. (2010). Assessing different uncertainty measures of EBLUP: a resampling-based approach, *Journal of Statistical Computation and Simulation*, **80**, 7, 713–727.

[17] Petrucci, A.; Pratesi, M. and Salvati, N. (2005). Geographic information in small area estimation: small area models and spatially correlated random area effects, *Statistics in Transition*, **7**, 3, 609–623.

[18] Petrucci, A. and Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment, *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 2, 169–182.

[19] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators, *Journal of the American Statistical Association*, **85**, 409, 163–171.

[20] Pratesi, M. and Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects, *Statistical Methods & Applications*, **17**, 1, 113–141.

[21] Pratesi, M. and Salvati, N. (2009). Small area estimation in the presence of correlated random area effects, *Journal of Official Statistics*, **25**, 1, 37–53.

[22] Rao, J.N.K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data, *The Canadian Journal of Statistics*, **22**, 4, 511–528.

[23] Saei, A. and Chambers, R. (2003). *Small area estimation under linear and generalized linear mixed models with time and area effects*, Working Paper M03/15, Southampton Statistical Sciences Research Institute, Southampton. Available from `http://eprints.soton.ac.uk/8165/`.

[24] Salvati, N. (2004). *Small area estimation by spatial models: the spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP)*, Working Paper No 2004/3, Dipartimento di Statistica "G. Parenti", Firenze. Available from `http://eprints.soton.ac.uk/8165/`.

[25] Singh, B.B.; Shukla, G.K. and Kundu, D. (2005). Spatio-temporal models in small area estimation, *Survey Methodology*, **31**, 2, 183–195.

[26] Thompson, W.A. (1962). The problem of negative estimates of variance components, *The Annals of Mathematical Statistics*, **33**, 1, 273–289.

[27] Wall, M.M. (2004). A close look at the spatial structure implied by the CAR and SAR models, *Journal of Statistical Planning and Inference*, **121**, 2, 311–324.

REVSTAT – Statistical Journal
Volume 10, Number 3, November 2012, 309–322

INFERENCES FOR THE CHANGE-POINT OF THE EXPONENTIATED WEIBULL HAZARD FUNCTION

Authors: Josmar Mazucheli
– Departamento de Estatística, Universidade Estadual de Maringá, Brazil
jmazucheli@uem.br

Emílio Augusto Coelho-Barros
– Departamento de Estatística, Universidade Estadual de Maringá, Brazil
eacbarros@uem.br

Jorge Alberto Achcar
– Departamento de Medicina Social – FMRP, Universidade de São Paulo, Brazil
achcar@fmrp.usp.br

Received: February 2011 Revised: April 2012 Accepted: May 2012

Abstract:

- In many applications of lifetime data analysis, it is important to perform inferences about the change-point of the hazard function. The change-point could be a maximum for unimodal hazard functions or a minimum for bathtub forms of hazard functions and is usually of great interest in medical or industrial applications. For lifetime distributions where this change-point of the hazard function can be analytically calculated, its maximum likelihood estimator is easily obtained from the invariance properties of the maximum likelihood estimators. From the asymptotical normality of the maximum likelihood estimators, confidence intervals can also be obtained. Considering the exponentiated Weibull distribution for the lifetime data, we have different forms for the hazard function: constant, increasing, unimodal, decreasing or bathtub forms. This model gives great flexibility of fit, but we do not have analytic expressions for the change-point of the hazard function. In this way, we consider the use of Markov Chain Monte Carlo methods to get posterior summaries for the change-point of the hazard function considering the exponentiated Weibull distribution.

Key-Words:

- *change-point; exponentiated Weibull distribution; hazard function; lifetime data analysis; Markov Chain Monte Carlo.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

Hazard function plays an important role in reliability and survival analysis and usually it can increase(decrease) up to a maximum(minimum) and then decrease(increase) after this change-point, also known as turning or critical point, [1, 12]. This is common in medical studies as in heart or kidney transplantation, where the patients have an increasing hazard during an adaptation period and a decreasing hazard after this adaptation period, [4]. In a study of recovery from breast cancer, it has been observed by [18] that the maximum mortality occurs after about three years and then it decreases slowly over a fixed period of time. In other situation, we have a bathtub shape for the hazard function where we have a decreasing hazard down to a minimum and then an increase after this change-point. In reliability, the turning point of a hazard function is useful in assessing the hazard in the useful life phase and this helps to determine and plan appropriate strategies for burn-in, maintenance and repair policies, [1].

Some common lifetime distributions like the exponential or Weibull are not appropriate to model non-monotonic hazard rate. Many existing probability distributions used to analyze lifetime data have unimodal hazard functions: the log-logistic distribution, [2]; the log-normal distribution, [17]; the inverse-Weibull distribution, [16]; the exponentiated Weibull distribution [20, 21] among many others. The exponentiated Weibull distribution is very flexible to be fitted by the data since it has constant, increasing, decreasing, unimodal and bathtub hazard functions. For situations where the hazard function is unimodal (bathtub) shaped, usually, we have interest in the estimation of the lifetime change-point that is, the point at which the hazard function reaches its maximum (minimum). In applications, the exponentiated Weibull distribution gives great flexibility of fit, but we do not have analytic expressions for the change-point of the hazard function.

In this paper, under the Bayesian point of view, we consider the use of Markov Chain Monte Carlo methods to get posterior summaries for the change-point of the hazard function. The maximum likelihood estimation procedure is also considered. It is important to point out that we do not have analytical expressions for this hazard change-point so we can not obtain classical asymptotic confidence intervals for the hazard change-point. The paper is organized as follows: in Section 2 we introduce some characteristics of the exponentiated Weibull distribution; in Section 3 we introduce the likelihood function in the presence of censored observations; in Section 4 we introduce some illustrative examples and finally, in Section 5 we present some conclusions.

2. THE EXPONENTIATED WEIBULL DISTRIBUTION

Let T be a non-negative random variable with an exponentiated Weibull distribution with hazard function given by:

$$h(t) = \theta \frac{h_1(t)\, S_1(t)\, F_1(t)^{\theta-1}}{1 - \big[F_1(t)\big]^{\theta}} \tag{2.1}$$

where $h_1(t)$, $S_1(t)$ and $F_1(t)$ are, respectively, the hazard function, the survival function and the accumulated distribution function of the Weibull distribution, [19]. For the Weibull distribution with scale parameter $\mu > 0$ and shape parameter $\beta > 0$, we have:

(2.2)

$$h_1(t) = \frac{\beta}{\mu^{\beta}}\, t^{\beta-1}, \qquad S_1(t) = \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right] \quad \text{and} \quad F_1(t) = 1 - \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right].$$

From the standard relations, [19], $S(t) = \exp\big[-\int_0^t h(u)\, du\big]$ and $f(t) = -\frac{d}{dt}S(t)$ and using the hazard function given in (2.1), we have the survival and density functions written as:

$$S(t) = 1 - \big[F_1(t)\big]^{\theta} \qquad \text{and} \qquad f(t) = \theta\, h_1(t)\, S_1(t)\, F_1(t)^{\theta-1} \tag{2.3}$$

respectively.

Explicitly we have that:

$$S(t) = 1 - \left\{1 - \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right]\right\}^{\theta} \tag{2.4}$$

and

$$f(t) = \theta \frac{\beta}{\mu^{\beta}}\, t^{\beta-1} \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right] \left\{1 - \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right]\right\}^{\theta-1} \tag{2.5}$$

where $\mu > 0$ is the scale parameter and $\theta > 0$ and $\beta > 0$ are the shape parameters. For $\theta = 1$ in (2.4) and (2.5) we have the survival and density function for the two parameter Weibull distribution. While, taking $\beta = 1$ we have the exponentiated exponential distribution, introduced by [11].

The great advantage of the exponentiated Weibull distribution in comparison to Weibull distribution is related to the behavior of the hazard function which depends on the values of θ and β. In Figure 1, we have the different shapes of the hazard function given by (2.1).

β	θ	Hazard
$= 1$	$= 1$	Constant (I)
> 1	> 1	Increasing (II)
< 1	> 1	Unimodal (III)
< 1	< 1	Decreasing (IV)
> 1	< 1	Bathtub (V)

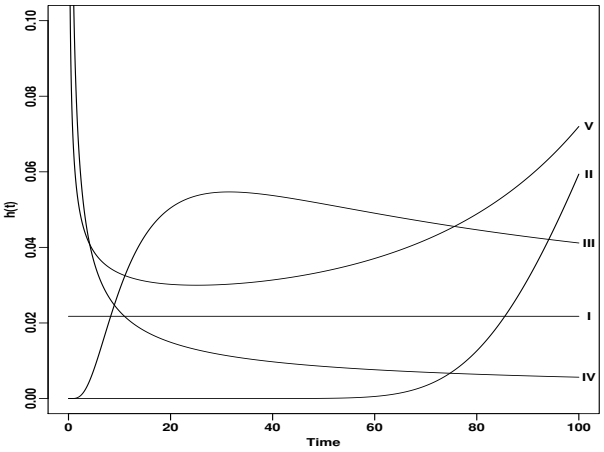

Figure 1: The exponentiated Weibull hazard function behavior.

3. HAZARD CHANGE POINT ESTIMATION — CLASSICAL APPROACH

Let us assume that $(t_1, \delta_1), ..., (t_n, \delta_n)$ is a random sample of size n of lifetimes generated by an exponentiated Weibull distribution with parameters μ, β and θ and that $(\delta_i = 1)$ if t_i is completely observed or $(\delta_i = 0)$ if t_i is a right censored observation $(i = 1, ..., n)$. Assuming a non-informative censoring mechanism, [19], the likelihood and log-likelihood functions are given, respectively, by:

$$L\big(\mu, \beta, \theta \mid \mathbf{t}, \boldsymbol{\delta}\big) = \prod_{i=1}^{n} \left(\theta \frac{h_1(t_i)\, S_1(t_i)\, F_1(t_i)^{\theta-1}}{1 - F_1(t_i)^{\theta}} \right)^{\delta_i} \big(1 - F_1(t_i)^{\theta}\big) \tag{3.1}$$

and:

$$\begin{aligned} l\big(\mu, \beta, \theta \mid \mathbf{t}, \boldsymbol{\delta}\big) \;=\; & R \log\theta + \sum_{i=1}^{n} \delta_i \log\big[f_1(t_i)\big] + (\theta - 1) \sum_{i=1}^{n} \delta_i \log\big[F_1(t_i)\big] \\ & + \sum_{i=1}^{n} (1 - \delta_i)\, \log\big[1 - F_1(t_i)^{\theta}\big] \end{aligned} \tag{3.2}$$

where $R = \sum_{i=1}^{n} \delta_i$, $h_1(t_i)$, $S_1(t_i)$ and $F_1(t_i)$ are defined in (2.2) and $f_1(t_i) = h_1(t_i)\, S_1(t_i)$.

Given a vector of observed lifetimes $(t_1, t_2, ..., t_n)$ and defining $l = l\big(\mu, \beta, \theta \mid \mathbf{t}, \boldsymbol{\delta}\big)$, the maximum likelihood estimates for μ, β and θ, denoted by $\hat{\mu}$, $\hat{\beta}$ and $\hat{\theta}$, are obtained solving, for example by Newton–Raphson, the following

likelihood equations:

$$\frac{\partial}{\partial\mu}l \;=\; -\beta R + \sum_{i=1}^{n}\delta_i t_i h_1(t_i) - (\theta-1)\sum_{i=1}^{n}\delta_i t_i\,\frac{f_1(t_i)}{F_1(t_i)} + \sum_{i=1}^{n}(1-\delta_i)\,t_i h(t_i) \;=\; 0\ ,$$

$$\begin{aligned}\frac{\partial}{\partial\beta}l \;=\; &\sum_{i=1}^{n}(1-\delta_i)\,t_i h(t_i)\log\!\left(\frac{t_i}{\mu}\right) + (\theta-1)\sum_{i=1}^{n}\delta_i t_i\,\frac{f_1(t_i)}{F_1(t_i)}\,\log\!\left(\frac{t_i}{\mu}\right)\\ &-\sum_{i=1}^{n}t_i\delta_i\,h_1(t_i)\log\!\left(\frac{t_i}{\mu}\right) + R\big[(1-\beta)\log(\mu)\big] + \beta\sum_{i=1}^{n}\delta_i\log(t_i) \;=\; 0\ ,\end{aligned}$$

$$\frac{\partial}{\partial\theta}l \;=\; R + \theta\sum_{i=1}^{n}\delta_i\log F_1(t_i) + \theta\sum_{i=1}^{n}\frac{(\delta_i-1)}{S(t)}\,F_1(t_i)^{\theta}\log F_1(t_i) \;=\; 0\ .$$

The $100\times(1-\alpha)\%$ confidence intervals for μ, β and θ can be obtained from the usual asymptotic normality of the maximum likelihood estimators with $\mathrm{Var}(\hat{\mu})$, $\mathrm{Var}(\hat{\beta})$ and $\mathrm{Var}(\hat{\theta})$ estimated from the inverse of the observed Fisher information matrix, that is, the inverse of the matrix of second derivatives of the log-likelihood function locally at $\hat{\mu}$, $\hat{\beta}$ and $\hat{\theta}$. From the invariance property of maximum likelihood estimators, we can obtain confidence intervals for functions of μ, β and θ. For $\phi = g(\mu,\beta,\theta)$, a differentiable function of μ, β and θ, we have $\hat{\phi} = g(\hat{\mu},\hat{\beta},\hat{\theta})$ and the variance of $\hat{\phi}$ is obtained using the delta method, [23].

Although the delta method is applied to estimate $\mathrm{Var}\big[g(\hat{\mu},\hat{\beta},\hat{\theta})\big]$, in some cases, it does not work, [5]. As a special situation, for $\beta<1$ and $\theta>1$ let us assume that we are interested in getting confidence intervals for the maximum of the exponentiated Weibull hazard function. Taking $\varphi = h(t)$, defined in (2.1), the maximum of the exponentiated Weibull hazard function, $T_{\max}$, is obtained as solution of the equation $\frac{d}{dt}\log(\varphi) = 0$ where, from (2.1):

$$\log(\varphi) \;\propto\; (\beta-1)\log(t) - \frac{t}{\beta}\,h_1(t) + (\theta-1)\log\big[F_1(t)\big] - \log\big[1-F_1(t)^{\theta}\big] \tag{3.3}$$

and:

$$\frac{d}{dt}\log(\varphi) \;=\; \frac{\beta-1}{t} - h_1(t) + (\theta-1)\,\frac{h_1(t)\,S_1(t)}{F_1(t)} + \theta\,\frac{h_1(t)\,S_1(t)\,F_1(t)^{\theta-1}}{\big[1-F_1(t)^{\theta}\big]}\ . \tag{3.4}$$

By the invariance principle of maximum likelihood estimator, the maximum likelihood estimator of the change point is the solution of (3.4) with μ, β and θ replaced by their maximum likelihood estimates. We observe that (3.4) is non-linear in t, so the maximum of the hazard function estimate $\hat{T}_{\max}$, should be obtained using some one dimensional root finding technique like Newton–Raphson. Since $\hat{T}_{\max}$ in not obtained from an analytical expression, it is not possible to estimate

$\mathrm{Var}\big[\hat{T}_{\max}\big]$ using the delta method. This fact shows the difficulty in applying the maximum likelihood methodology when the change point does not have closed form and this fact justify the application of the Bayesian methodology. Standard resampling procedures like the Bootstrap and the Jackknife are other alternatives but they will not be considered in this paper.

4. HAZARD CHANGE POINT ESTIMATION — BAYESIAN APPROACH

Under the Bayesian approach, assuming a joint prior distribution for $\boldsymbol{\nu} = (\mu,\beta,\theta)$ in the form $\pi(\boldsymbol{\nu}) = \pi(\theta)\,\pi(\beta)\,\pi(\theta)$, we get the joint posterior distribution given by:

$$\pi(\boldsymbol{\nu} \mid \mathbf{t},\boldsymbol{\delta}) \;\propto\; \pi(\boldsymbol{\nu}) \prod_{i=1}^{n} \left(\theta\, \frac{h_1(t_i)\, S_1(t_i)\, F_1(t_i)^{\theta-1}}{1 - F_1(t_i)^{\theta}} \right)^{\delta_i} \big(1 - F_1(t_i)^{\theta}\big)\ . \tag{4.1}$$

From (4.1) it is clear that is not possible to get explicit forms for the marginal posterior distributions for each parameter. In this way, we should use some approximation method to solve integrals as the Laplace method, [26], or some other numerical method, [22]. When models become too difficult to be analyze analytically, we have to use simulation algorithms, such as the Markov Chain Monte Carlo methods to obtain posterior estimates, [7, 14]. The Markov Chain Monte Carlo methods is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest. To simulate samples of the joint posterior distribution of interest, we need to sample successively from a target distribution. The Gibbs algorithm requires to decompose the joint posterior distribution into full conditional distributions for each parameter in the model and then sample from each one of these conditional distributions

For the exponentiated Weibull distribution, the conditional posterior densities for μ, β and θ show that standard sampling schemes are not feasible since the conditional distributions are not given in a known form. In this way, an alternative target distribution to the full conditional distributions should be used. The alternative proposal distribution should be a distribution from which it is easy to sample from it; in this way, we use Metropolis–Hastings algorithms, [3, 14]. Tierney (1994) suggested, when possible, use of the Metropolis–Hastings algorithm within Gibbs sampling to sample from full conditional distributions.

In our applications, to sample from the full conditional distributions for μ, β and θ, we have used the Adaptive Rejection Metropolis Sampling algorithm, *ARMS*, introduced by Gilks *et al.* (1995) also discussed in [10]. This algorithm is a generalization of the method of adaptive rejection sampling of Gilks (1992) which

includes a Metropolis step to accommodate non-log-concavity in the density that will be sampled. The *C* code, written by Gilks, and linked to the matrix language *Ox*, [6], was used to compute the posterior summaries of interest. The *ARMS* algorithm, to the best of our knowledge, also is available in the libraries *dlm*, *SamplerCompare* and *HI*, under *R,* [15]. *ARMS* is also available in *GENMOD*, *LIFEREG*, *PHREG* and *MCMC* procedures under *SAS* version 9.2. The MCMC procedure provides a flexible environment for fitting a wide range of models.

It is important to point out that [24] employed Bayesian and frequentist perspectives for estimating parameters for exponentiated Weibull distribution and showed a comprehensive and updating list of references.

5. SOME ILLUSTRATIVE EXAMPLES

5.1. An Example with a Simulated Data Set

In this subsection we introduce an example considering a simulated data set from an exponentiated Weibull distribution with parameters $\mu = 1$, $\beta = 0.8$ and $\theta = 4.0$. Since μ does not change the maximum of the hazard function, without loss of generality, we consider it known and equal to 1. Under this parameter configuration, the hazard function is of bathtub shape and we would like to estimate the parameter $T_{\max}$ (the "true value" of $T_{\max}$ is 3.9114). A total of $n = 50$ observations (see Table 1) were simulated using the inversion method considering $\delta_i = 1$ $(i = 1, ..., 50)$. Replacing the values of β and θ by their maximum likelihood or Bayesian estimates we can solve (3.4) for t to get the maximum of the hazard function.

Table 1: Simulated data set from an exponentiated Weibull distribution with $\mu = 1$, $\beta = 0.8$ and $\theta = 4.0$.

0.23	0.53	0.54	0.60	0.65	0.84	0.90	0.96	0.98	0.99
1.05	1.26	1.28	1.31	1.33	1.37	1.45	1.53	1.69	1.72
1.77	1.80	1.96	2.06	2.14	2.24	2.35	2.40	2.47	2.47
2.58	2.68	2.73	2.77	2.78	2.96	3.04	3.31	3.36	3.67
4.01	4.16	4.19	4.36	4.91	5.10	5.81	6.27	7.39	7.41

As observed in [1], the turning point of a hazard rate function is useful in assessing the hazard in the useful life phase and helps to determine and plan appropriate burn-in, maintenance, and repair policies and strategies. For many bathtub-shaped distributions, the turning point is unique, and the hazard varies little in the useful life phase.

In Table 2 we have the maximum likelihood estimates (standard-errors) and the posterior means of the parameters β, θ and $T_{\max}$ assuming non-informative gamma prior distribution for the parameters β and θ. The maximum likelihood estimates were obtained by the Newton–Raphson method available in the software *SAS/NLP* procedure, [13]. With the obtained maximum likelihood estimates $\hat{\beta}$ and $\hat{\theta}$ we estimate $T_{\max}$ maximizing (3.3). Under the Bayesian approach, the parameters β and θ were estimated using the *ARMS* algorithm in *Ox*. We simulated five separate chains using different overdispersed starting values for each run. The algorithm was run for 21 000 iterations and the starting values were based in previous runs of the *ARMS* algorithm for large intervals. This strategy follows the ideas discussed in Gilks *et al.* (1995, 1997). We considered five initial abscissae based on 5%, 40%, 50%, 60% and 95% of the envelope function after previous runs. In order to diminish the effect of the starting parameter values, we discarded the first 1000 elements of each chain. Convergence of the five combined simulated chain was observed using diagnostic procedures available in *BOA* library, [25], under *R*, [15]. For each parameter we considered every 5^{th} draw and stopped at a sample of size 20 000. The hyperparameters were set so that we had a proper but very non-informative prior. For all parameters we have adopted a gamma prior distribution with shape and scale parameters equal to 0.001. From each $\hat{\beta}$ and $\hat{\theta}$ the $T_{\max}$ estimates were obtained by maximization of (3.3). Again, the Newton–Raphson method under *SAS/NLP* procedure was used. Figure 2 shows the estimated marginal posterior distribution for β, θ and $T_{\max}$.

Table 2: Maximum likelihood estimates (standard error), 95% confidence intervals, posterior means (standard deviation) and 95% credible intervals.

Parameter	MLE[‡]	95% Confidence Interval
β	0.8152 (0.0610)	(0.6956; 0.9348)
θ	3.8845 (0.5570)	(2.7928; 4.9762)
$T_{\max}$	3.9329	—

[‡]maximum likelihood estimate.

Parameter	Posterior Mean	95% Credible Interval
β	0.8102 (0.0617)	(0.6891; 0.9295)
θ	3.8714 (0.5475)	(2.8790; 4.9981)
$T_{\max}$	3.9768 (0.6800)	(2.5564; 5.2575)

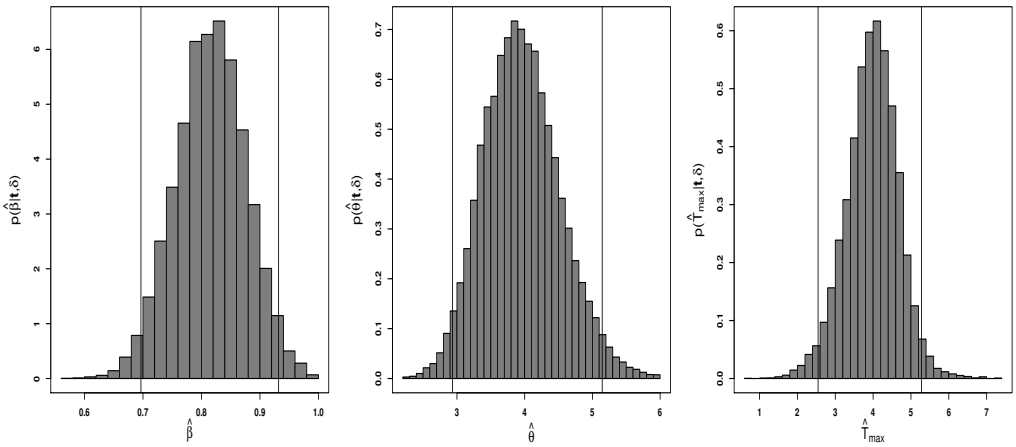

Figure 2: Estimated marginal posterior distribution for β, θ and $T_{\max}$.

5.2. An Example with a Real Data Set

As a second example, let us consider the failure data for a group of 60 electrical appliances in a life test ($1000s$ of cycles) extracted from Lawless (2003, p. 112). Figure 3 shows the Kaplan–Meier survival curve with fit for the exponentiated Weibull distribution. We observe close agreement between the Kaplan–Meier survival curve with the exponentiated Weibull distribution. The maximum likelihood and posterior means estimates are showed in Table 3 and obtained in a similar way as considered in the simulated example.

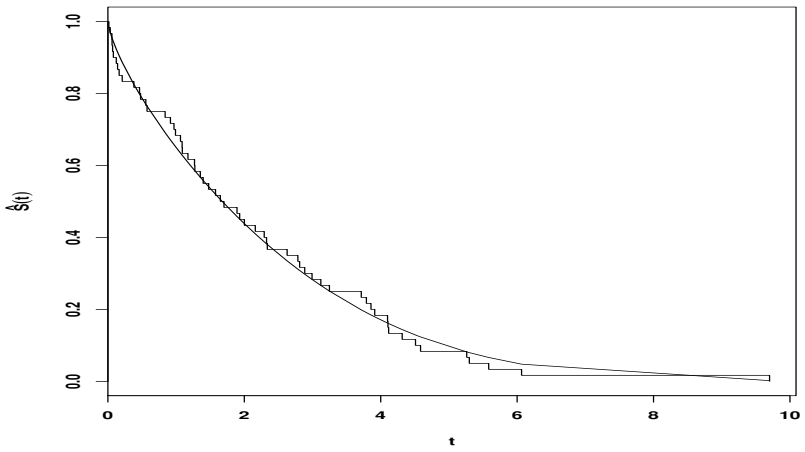

Figure 3: Fit of exponentiated Weibull distribution. (—) Kaplan–Meier survival curve and (···) exponentiated Weibull survival function.

From the results of Table 3, we observe that $\hat{\beta} > 1$ and $\hat{\theta} < 1$, that is, we have a bathtub form for the hazard function (see Figure 1). In this situation we have interest in getting inferences for the change-point of the hazard function. This change-point is given by the minimum $T_{\min}$ of the hazard function (2.1), obtained from equation (3.4).

Table 3: Maximum likelihood estimates (standard error), 95% confidence intervals, posterior means (standard deviation) and 95% credible intervals.

Parameter	MLE[‡]	95% Confidence Interval
μ	4.1595 (0.9794)	(2.2399; 6.0791)
β	1.9599 (0.6420)	(0.7017; 3.2181)
θ	0.3717 0.1657	(0.0470; 0.6964)
$T_{\min}$	0.8865	—

[‡]maximum likelihood estimate.

Parameter	Posterior Mean	95% Credible Interval
μ	4.0491 (0.6690)	(2.8853; 5.4856)
β	1.7862 (0.3635)	(1.2142; 2.6146)
θ	0.4360 (0.1164)	(0.2470; 0.6904)
$T_{\min}$	1.0316 (0.5217)	(0.1804; 2.3321)

In Figure 4, we have the plots for the approximated marginal posterior distributions for μ, β, θ and $T_{\min}$ based on the 20000 simulated samples.

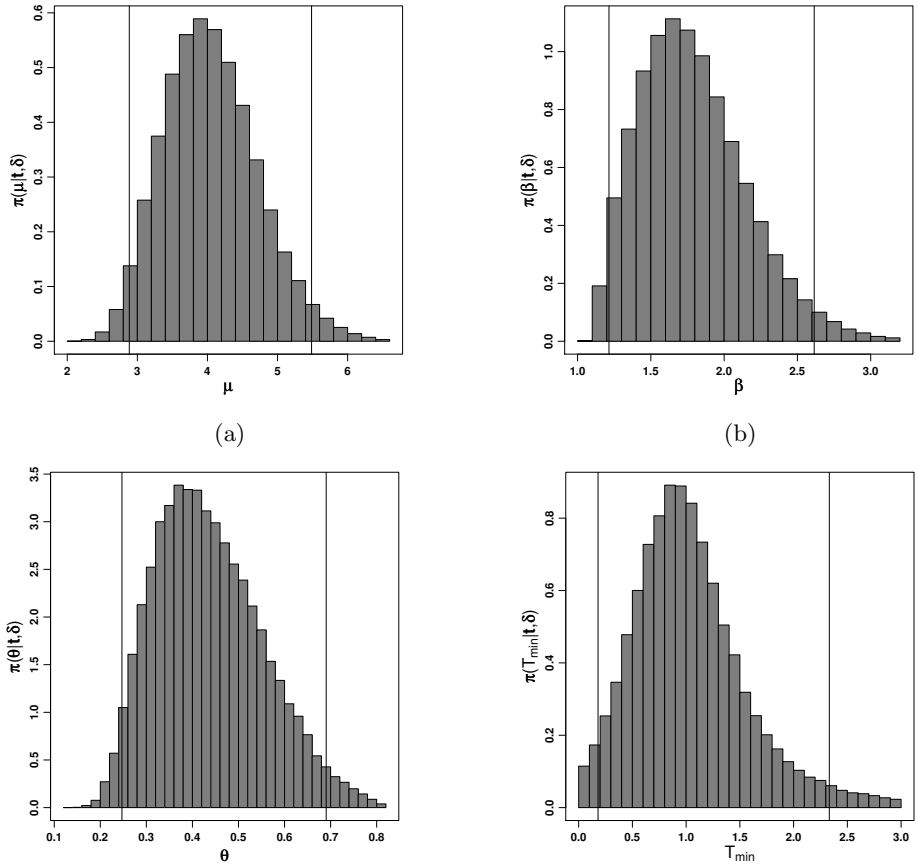

Figure 4: Marginal posterior distribution for μ, β, θ and $T_{\min}$.

6. CONCLUDING REMARKS

Inferences for the change-point of the hazard function are of great interest in lifetime studies, especially in medical or industrial applications. Assuming the exponentiated Weibull distribution we can have better fit for lifetime data since we have different shapes for the hazard function. In this situation, we do not have analytic expressions for the change-point of the hazard function (maximum if we have unimodal hazard function or minimum if we have bathtub hazard function) and we cannot use standard asymptotic classical inference methods to obtain inferences for the change-point. Using standard Markov Chain Monte Carlo simulation methods for a Bayesian analysis of the model, we get in a simple way, the posterior summaries of interest, like credible intervals for the change-points.

ACKNOWLEDGMENTS

The authors are gratefully to a referee for his valuable and carefully review of the paper and for his valuable comments.

REFERENCES

[1] Bebbington, M.; Lai, C. and Zitikis, R. (2008). Estimating the turning point of a bathtub-shaped failure distribution, *Journal of Statistical Planning and Inference*, **138**, 1157–1166.

[2] Bennet, S. (1983). Log-logistic regression models for survival data, *Applied Statistics*, **32**, 165–171.

[3] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm, *Journal of American Statistical Association*, **49**, 327–335.

[4] Collett, D. (1994). *Modelling Survival Data in Medical Research*, Chapman and Hall, New York.

[5] Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*, Chapman and Hall, London.

[6] Doornik, J. (1999). *An Object-Oriented Matrix Programming Language*, Timberlake, London.

[7] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal distributions, *Journal of the American Statistical Association*, **85**, 398–409.

[8] Gilks, W.R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4*, pp. 641–649, Clarendon Press, London.

[9] Gilks, W.R.; Best, N.G. and Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Applied Statistics*, **44**, 455–472.

[10] Gilks, W.R.; Neal, R.M.; Best, N.G. and Tan, K.K.C. (1997). Corrigendum: Adaptive rejection Metropolis sampling within Gibbs sampling, *Applied Statistics*, **46**, 541–542.

[11] Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions, *Australian & New Zealand Journal of Statistics*, **41**, 173–188.

[12] Gupta, R.C.; Lvin, S. and Peng, C. (2010). Estimating turning points of the failure rate of the extended Weibull distribution, *Computational Statistics and Data Analysis*, **54**, 924–934.

[13] Hartmann, W.M. (1994). *The NLP Procedure: Release 6.10 and 6.11, Extended User's Guide*, SAS Institute, Inc., SAS Campus Drive, Cary, NC 27513, USA.

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

[15] Ihaka, R. and Geltleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computation and Graphical Statistics*, **5**, 299–341.

[16] Jiang, R.; Murthy, D.N.P. and Ji, P. (2001). Models involving two inverse Weibull distributions, *Reliability Engineering & System Safety*, **73**, 73–81.

[17] Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*, Springer-Verlag, New York.

[18] Langlands, A.O.; Pocock, S.J.; Kerr, G.R. and Gore, S.M. (1979). Long term survival of patients with breast cancer: A study of the curability of the disease, *British Medical Journal*, **17**, 1247–1251.

[19] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd ed., John Wiley and Sons, New York.

[20] Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Transactions on Reliability*, **42**(2), 299–302.

[21] Mudholkar, G.S.; Srivastava, D.K. and Fraimer, M. (1995). The Exponentiated Weibull Family: A Reanalysis of the Bus-Motor-Failure Data, *Technometrics*, **37**, 436–445.

[22] Naylor, J.C. and Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distributions, *Journal of the Royal Statistical Society, C*, **31**, 214–225.

[23] Rao, C.R. and Toutenburg, H. (1999). *Linear Models*, Springer Series in Statistics, Springer-Verlag, New York, second Least squares and alternatives, With contributions by Andreas Fieger.

[24] Singh, U.; Gupta, P.K. and Upadhyay, S.K. (2005). Estimation of parameters for exponentiated-Weibull family under type-II censoring scheme, *Computational Statistics & Data Analysis*, **48**(3), 509–523.

[25] Smith, B. (2000). *Bayesian output analysis program (BOA), version 0.5.0 for S-PLUS and R*, Available on-line: `http://www.public-health.uiowa.edu/BOA`.

[26] Tierney, L.; Kass, R.E. and Kadane, J.B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions, *Journal of the American Statistical Association*, **84**(407), 710–716.

REVSTAT – Statistical Journal
Volume 10, Number 3, November 2012, 323–333

REGULAR A-OPTIMAL SPRING BALANCE WEIGHING DESIGNS

Author: Małgorzata Graczyk
– Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences,
Wojska Polskiego 28, 60-637 Poznań, Poland
magra@up.poznan.pl

Received: February 2011 Revised: June 2012 Accepted: June 2012

Abstract:

- The problem of indicating an A-optimal spring balance weighing design providing that the measurement errors have different variances and are uncorrelated is considered. The lowest bound of the trace of the inverse information matrix is given and the conditions determining the optimal design are also presented. The incidence matrices of balanced incomplete block designs and group divisible designs are used in constructions of the regular A-optimal spring balance weighing design.

Key-Words:

- *A-optimal design; balanced incomplete block design; group divisible design; spring balance weighing design.*

AMS Subject Classification:

- 62K05, 62K15.

1. INTRODUCTION

In several fields of the experiments, specially in the theory of spectroscopy, metrology, dynamical system theory, computational mechanics and 2^n fractional factorial designs we determine unknown measurements of p objects using n operations according to the linear model

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e} \,, \tag{1.1}$$

where $\mathbf{y}$ is an $n\times 1$ random vector of the observations. The design matrix $\mathbf{X} = (x_{ij})$ usually called weighing matrix belongs to the class $\mathbf{\Phi}_{n\times p}(0,1)$, which denotes the class of $n\times p$ matrices of known elements $x_{ij} = 0$ or 1 according as in the i^{th} weighing operation the j^{th} object is not placed on the pan or is placed. $\mathbf{w}$ is a $p\times 1$ vector of unknown weights of objects and $\mathbf{e}$ is an $n\times 1$ random vector of errors. We assume, that there are no systematic errors, the variances of errors are not equal and the errors are uncorrelated, i.e. $\mathbf{E}(\mathbf{e}) = \mathbf{0}_n$ and $\mathrm{Var}(\mathbf{e}) = \sigma^2\,\mathbf{G}$, where $\mathbf{0}_n$ denotes the $n\times 1$ vector with zero elements everywhere, $\mathbf{G}$ is the known $n\times n$ diagonal positive definite matrix.

For the estimation of individual unknown weights of objects we use normal equations $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}\mathbf{w} = \mathbf{X}'\mathbf{G}^{-1}\mathbf{y}$. Any spring balance weighing design is said to be singular or nonsingular, depending on whether the matrix $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}$ is singular or nonsingular, respectively. It is obvious, that if $\mathbf{G}$ is the known positive definite matrix then the matrix $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}$ is nonsingular if and only if the matrix $\mathbf{X}'\mathbf{X}$ is nonsingular, i.e. if and only if $\mathbf{X}$ is full column rank $\mathrm{r}(\mathbf{X}) = p$. However, if $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}$ is nonsingular, then the generalized least squares estimator of $\mathbf{w}$ is given by $\hat{\mathbf{w}} = (\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{G}^{-1}\mathbf{y}$ and the variance matrix of $\hat{\mathbf{w}}$ is $\mathrm{Var}(\hat{\mathbf{w}}) = \sigma^2(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}$. A more complete theory may be obtained in literature[1].

In many problems cases the weighing designs, the A-optimal design is considered. For given variance matrix of the errors $\sigma^2\mathbf{G}$, the A-optimal design is the design $\mathbf{X}$ for which, the sum of variances of estimators for unknown parameters is minimal, i.e. $\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}$ is minimal in $\mathbf{\Phi}_{n\times p}(0,1)$. Moreover, the design for which the sum of variances of estimators for parameters attains the lowest bound in $\mathbf{\Phi}_{n\times p}(0,1)$ is called the regular A-optimal design. Let note, in the set of design matrices $\mathbf{\Phi}_{n\times p}(0,1)$, the regular A-optimal design may not exist, whereas A-optimal design always exists. The concept of the A-optimality was shown in many papers[2].

[1] See, for instance, Raghavarao ([12]) and Banerjee ([1]).

[2] See, Jacroux and Notz ([8]), Shah and Sinha ([14]), Pukelsheim ([11]), Ceranka and Graczyk ([2]), Ceranka *et al.* ([3], [4]), Masaro and Wong ([10]) and Graczyk ([6], [7]).

2. REGULAR A-OPTIMAL DESIGNS

For any experimental setting, i.e. for fixed n, p and $\mathbf{G}$, there is always a number of designs available for using. In each class of available designs, the regular A-optimal design is considered. Furthermore, the main difficulty in carrying out the construction is that each form of $\mathbf{G}$ requires the specific investigations. That's why we consider the experimental situation we determine unknown measurements of p objects in $n = \sum_{i=1}^{h} n_s$ measurement operations under model 1.1. It is assumed that n_s measurements are taken in different h conditions or at different h installations. So, the variance matrix of errors $\sigma^2\mathbf{G}$ is given by the matrix $\mathbf{G}$

$$\mathbf{G} = \begin{bmatrix} g_1^{-1}\mathbf{I}_{n_1} & \mathbf{0}_{n_1}\mathbf{0}'_{n_2} & \cdots & \mathbf{0}_{n_1}\mathbf{0}'_{n_h} \\ \mathbf{0}_{n_1}\mathbf{0}'_{n_1} & g_2^{-1}\mathbf{I}_{n_2} & \cdots & \mathbf{0}_{n_2}\mathbf{0}'_{n_h} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{n_h}\mathbf{0}'_{n_1} & \mathbf{0}_{n_h}\mathbf{0}'_{n_2} & \cdots & g_h^{-1}\mathbf{I}_{n_h} \end{bmatrix}, \tag{2.1}$$

where $g_s > 0$ denotes the factor of precision, $s = 1, 2, ..., h$. Consequently, according to the form of $\mathbf{G}$ we write the design matrix $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0, 1)$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \cdots \\ \mathbf{X}_h \end{bmatrix} \tag{2.2}$$

where $\mathbf{X}_s$ is $n_s \times p$ design matrix of any spring balance weighing design.

The Lemma given below presented in [9] is required for determining the regular A-optimal design.

Lemma 2.1. *Let $\mathbf{\Pi}$ be the the set of all $p\times p$ permutation matrices and let $\mathbf{M}$ be a $p\times p$ matrix. If $\bar{\mathbf{M}} = \frac{1}{p!}\sum_{\mathbf{P}\in\mathbf{\Pi}} \mathbf{P}'\mathbf{M}\mathbf{P}$ then $\bar{\mathbf{M}} = \left(\frac{\mathrm{tr}(\mathbf{M})}{p} - \frac{Q(\mathbf{M})}{p(p-1)}\right)\mathbf{I}_p + \frac{Q(\mathbf{M})}{p(p-1)}\mathbf{1}_p\mathbf{1}'_p$, where $\mathrm{tr}(\mathbf{M})$ is the trace of $\mathbf{M}$, $Q(\mathbf{M})$ denotes the sum of the off-diagonal elements of $\mathbf{M}$ and $\mathbf{1}_p$ is the vector of ones. Moreover, $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\bar{\mathbf{M}})$ and $Q(\mathbf{M}) = Q(\bar{\mathbf{M}})$.*

From now on, we assume that $\mathbf{G}$ is taken into consideration in the form (2.1).

Theorem 2.1. *In any nonsingular spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0, 1)$ in (2.2) with the variance matrix of errors $\sigma^2\mathbf{G}$*

$$\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} \geq \begin{cases} \dfrac{4(p^2-2p+2)}{p\,\mathrm{tr}(\mathbf{G}^{-1})} & \text{if } p \text{ is even}, \\ \dfrac{4p^3}{(p+1)^2\,\mathrm{tr}(\mathbf{G}^{-1})} & \text{if } p \text{ is odd}. \end{cases} \tag{2.3}$$

Proof: For $\mathbf{X}$ in (2.2) and $\mathbf{G}$ in (2.1), we obtain $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \sum_{s=1}^{h} g_s \mathbf{X}_s'\mathbf{X}_s$ and moreover $\operatorname{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \sum_{z=1}^{p} \frac{1}{\mu_z}$, where μ_z is the eigenvalue of $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}$. Next, we consider the matrix $\bar{\mathbf{M}} = \alpha\,\mathbf{I}_p + \beta\,\mathbf{1}_p\mathbf{1}_p'$. From [8], $\bar{\mathbf{M}}$ has eigenvalues α with the multiplicity $p-1$ and $\alpha + p\beta$ with the multiplicity 1. Based on Lemma 2.1, $\bar{\mathbf{M}} = \frac{p\operatorname{tr}(\mathbf{M}) - \mathbf{1}_p'\mathbf{M}\mathbf{1}_p}{p(p-1)}\mathbf{I}_p + \frac{\mathbf{1}_p'\mathbf{M}\mathbf{1}_p - \operatorname{tr}(\mathbf{M})}{p(p-1)}\mathbf{1}_p\mathbf{1}_p'$. The eigenvalues of $\bar{\mathbf{M}}$ are $\mu_1 = \frac{1}{p(p-1)}\left(p\operatorname{tr}(\mathbf{M}) - \mathbf{1}_p'\mathbf{M}\mathbf{1}_p\right)$ with the multiplicity $p-1$ and $\mu_2 = \frac{1}{p}\mathbf{1}_p'\mathbf{M}\mathbf{1}_p$ with the multiplicity 1. Taking $\mathbf{M} = \mathbf{X}'\mathbf{G}^{-1}\mathbf{X}$ we have $\operatorname{tr}(\mathbf{M}) = \sum_{s=1}^{h} g_s \mathbf{k}_s'\mathbf{1}_{n_s}$ and $\mathbf{1}_p'\mathbf{M}\mathbf{1}_p = \sum_{s=1}^{h} g_s \mathbf{k}_s'\mathbf{k}_s$ so we obtain

$$
\begin{aligned}
\operatorname{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} &= \frac{p}{\mathbf{1}_p'\mathbf{M}\mathbf{1}_p} + \frac{p(p-1)^2}{p\operatorname{tr}(\mathbf{M}) - \mathbf{1}_p'\mathbf{M}\mathbf{1}_p} \\
&= \frac{p}{\sum_{s=1}^{h} g_s \mathbf{k}_s'\mathbf{k}_s} + \frac{p(p-1)^2}{\sum_{s=1}^{h} g_s (p\mathbf{1}_{n_s} - \mathbf{k}_s)'\mathbf{k}_s}\,,
\end{aligned} \tag{2.4}
$$

where $\mathbf{k}_s = \mathbf{X}_s\mathbf{1}_p$, $s = 1, 2, ..., h$. For even p, minimum of (2.4) is attained if and only if $\mathbf{k}_s = \frac{p}{2}\mathbf{1}_{n_s}$. For odd p, minimum of (2.4) is attained if and only if $\mathbf{k}_s = \frac{p+1}{2}\mathbf{1}_{n_s}$, $s = 1, 2, ..., h$. Hence, we obtain (2.3). □

Definition 2.1. Any $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0,1)$ given in (2.2) with the variance matrix of errors $\sigma^2\mathbf{G}$ is said to be the regular A-optimal if the equality in (2.3) is satisfied.

Theorem 2.2. *Any $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0,1)$ in (2.2) with the variance matrix of errors $\sigma^2\mathbf{G}$ is the regular A-optimal spring balance weighing design if and only if*

a) *for even p,* $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \frac{p}{4(p-1)}\operatorname{tr}(\mathbf{G}^{-1})\,\mathbf{I}_p + \frac{p-2}{4(p-1)}\operatorname{tr}(\mathbf{G}^{-1})\,\mathbf{1}_p\mathbf{1}_p'$,

or

b) *for odd p,* $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \frac{p+1}{4p}\operatorname{tr}(\mathbf{G}^{-1})\,(\mathbf{I}_p + \mathbf{1}_p\mathbf{1}_p')$.

Proof: The proof follows naturally into two parts. If p be odd, then from the Theorem 2.1, $\operatorname{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}$ takes minimum if and only if $\mathbf{k}_s = \frac{p}{2}\mathbf{1}_{n_s}$ for each s. Then it is easy to see that $\operatorname{tr}(\mathbf{M}) = \frac{p}{2}\operatorname{tr}(\mathbf{G}^{-1})$ and $\mathbf{1}_p'\mathbf{M}\mathbf{1}_p = \frac{p^2}{4}\operatorname{tr}(\mathbf{G}^{-1})$. So we have $\alpha = \frac{p\operatorname{tr}(\mathbf{G}^{-1})}{4(p-1)}$ and $\beta = \frac{(p-2)\operatorname{tr}(\mathbf{G}^{-1})}{4(p-1)}$ and we obtain a). The analogous consideration for odd p, imply that $\alpha = \beta = \frac{(p+1)\operatorname{tr}(\mathbf{G}^{-1})}{4p}$ and b) is true that finishes the proof. □

If $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0,1)$ satisfies the equalities given in Theorem 2.2 then $\mathbf{X}$ is the regular A-optimal design for any $\mathbf{G}$ in (2.1). Hence $\mathbf{X}$ is the regular A-optimal design in the special case when $\mathbf{G} = \mathbf{I}_n$ and

$$
\operatorname{tr}(\mathbf{X}'\mathbf{X})^{-1} \geq \begin{cases} \dfrac{4(p^2 - 2p + 2)}{np} & \text{if } p \text{ is even}\,, \\ \dfrac{4p^3}{n(p+1)^2} & \text{if } p \text{ is odd}\,. \end{cases} \tag{2.5}
$$

(2.5) is equivalent to the lowest bound of $\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}$ which follows from theorems given in [8]. On the other hand, we assume that $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0,1)$ is the regular A-optimal design for $\mathbf{G} = \mathbf{I}_n$. Then we can compare two traces $\frac{\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}}{\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}} = \frac{\sum_{s=1}^{h} g_s n_s}{n}$. We obtain the following Corollary.

Corollary 2.1. *Let* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0,1)$ *and* $\mathbf{G}$ *be of the form (2.1).*

a) *If* $\sum_{s=1}^{h} g_s n_s = n$, *then* $\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}$,

b) *If* $\sum_{s=1}^{h} g_s n_s > n$, *then* $\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} < \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}$,

c) *If* $\sum_{s=1}^{h} g_s n_s < n$, *then* $\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} > \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}$.

3. CONSTRUCTION OF THE REGULAR A-OPTIMAL DESIGNS

It is worth pointing out that the incidence matrices of the block designs may be used for the construction of the design matrix $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0,1)$, then we take $n = b$ and $p = v$. From all possible block designs, in this paper the construction of the regular A-optimal spring balance weighing design based on the incidence matrices of the balanced incomplete block designs is chosen.

Theorem 3.1. *Let* v *be even and let* $\mathbf{N}$ *be the incidence matrix of balanced incomplete block design with the parameters* $v = 2t$, $b = 2(2t-1)$, $r = 2t-1$, $k = t$, $\lambda = t-1$, $t = 2, 3, \ldots$. *Then any* $\mathbf{X} \in \mathbf{\Phi}_{hb\times 2t}(0,1)$ *in the form* $\mathbf{X} = \mathbf{1}_h \otimes \mathbf{N}'$ *is the regular A-optimal spring balance weighing design with the variance matrix of errors* $\sigma^2\mathbf{G}$.

Proof: An easy computation shows that the matrix $\mathbf{X} = \mathbf{1}_h \otimes \mathbf{N}'$ satisfies a) of Theorem 2.2. □

Theorem 3.2. *Let* v *be odd and let* $\mathbf{N}$ *be the incidence matrix of balanced incomplete block design with the parameters*

a) $v = 2t+1$, $b = 2(2t+1)$, $r = 2(t+1)$, $k = t+1$, $\lambda = t+1$,

or

b) $v = 4t-1$, $b = 4t-1$, $r = 2t$, $k = 2t$, $\lambda = t$,

$t = 1, 2, \ldots$. *Then any* $\mathbf{X} \in \mathbf{\Phi}_{hb\times v}(0,1)$ *in the form* $\mathbf{X} = \mathbf{1}_h \otimes \mathbf{N}'$ *is the regular A-optimal spring balance weighing design with the variance matrix of errors* $\sigma^2\mathbf{G}$.

Proof: This is proved by checking that the matrix $\mathbf{X} = \mathbf{1}_h \otimes \mathbf{N}'$ satisfies b) of Theorem 2.2. □

For an even v, we give the construction of the regular A-optimal spring balance weighing design based on the incidence matrices of group divisible design. Hence, we get

$$\mathbf{X}_s = \begin{bmatrix} \mathbf{N}'_{1s} \\ \mathbf{N}'_{2s} \end{bmatrix} , \qquad s = 1, 2, ..., h , \tag{3.1}$$

where $\mathbf{N}_{uh}$ is the incidence matrix of group divisible design with the parameters v, b_{us}, r_{us}, k_{us}, λ_{1us}, λ_{2us}, $u = 1, 2$, see [13]. Furthermore, let the condition

$$\lambda_{11s} + \lambda_{12s} = \lambda_{21s} + \lambda_{22s} = \lambda_s , \qquad u = 1, 2, \quad s = 1, 2, ..., h , \tag{3.2}$$

be satisfied. Below, we show the parameters of group divisible design which satisfy (3.1). Next, for $n = \sum_{s=1}^{h} \sum_{u=1}^{2} b_{us}$ bus measurements and v objects, based on the incidence matrices of the group divisible designs will be constructed $\mathbf{X} \in \mathbf{\Phi}_{n \times v}(0, 1)$. For the t, q, u given in Lemmas 3.1–3.5, some restrictions derive from the ones given in [5]: $r, k \leq 10$.

Lemma 3.1. *Let $v = 4$. If the parameters of group divisible designs are equal to*

a) $b_{1s} = 2(3t+1)$, $r_{1s} = 3t+1$, $k_{1s} = 2$, $\lambda_{11s} = t+1$, $\lambda_{21s} = t$, $t = 1, 2, 3$ *and* $b_{2s} = 2(3q+2)$, $r_{2s} = 3q+2$, $k_{2s} = 2$, $\lambda_{12s} = q$, $\lambda_{22s} = q+1$, $q = 0, 1, 2$,

b) $b_{1s} = 2(3t+2)$, $r_{1s} = 3t+2$, $k_{1s} = 2$, $\lambda_{11s} = t+2$, $\lambda_{21s} = t$, $t = 1, 2$ *and* $b_{2s} = 2(3q+4)$, $r_{2s} = 3q+4$, $k_{2s} = 2$, $\lambda_{12s} = q$, $\lambda_{22s} = q+2$, $q = 0, 1, 2$,

c) $b_{1s} = 2(u+3)$, $r_{1s} = u+3$, $k_{1s} = 2$, $\lambda_{11s} = u+1$, $\lambda_{21s} = 1$ *and* $b_{2s} = 4u$, $r_{2s} = 2u$, $k_{2s} = 2$, $\lambda_{12s} = 0$, $\lambda_{22s} = u$, $u = 1, 2, 3, 4, 5$,

d) $b_{1s} = 16$, $r_{1s} = 8$, $k_{1s} = 2$, $\lambda_{11s} = 0$, $\lambda_{21s} = 4$ *and* $b_{2s} = 2(3u+4)$, $r_{2s} = 3u+4$, $k_{2s} = 2$, $\lambda_{12s} = u+4$, $\lambda_{22s} = u$, $u = 1, 2$,

e) $b_{1s} = 18$, $r_{1s} = 9$, $k_{1s} = 2$, $\lambda_{11s} = 5$, $\lambda_{21s} = 2$ *and* $b_{2s} = 6(u+2)$, $r_{2s} = 3(u+2)$, $k_{2s} = 2$, $\lambda_{12s} = u$, $\lambda_{22s} = u+3$, $u = 0, 1$,

then for any matrix in (3.1), the condition (3.2) is satisfied.

Lemma 3.2. *Let $v = 6$. If the parameters of group divisible designs are equal to*

a) $b_{1s} = 4t$, $r_{1s} = 2t$, $k_{1s} = 3$, $\lambda_{11s} = 0$, $\lambda_{21s} = t$ *and* $b_{2s} = 6t$, $r_{2s} = 3t$, $k_{2s} = 3$, $\lambda_{12s} = 2t$, $\lambda_{22s} = t$, $t = 1, 2, 3$,

b) $b_{1s} = 2(2t+5)$, $r_{1s} = 2t+5$, $k_{1s} = 3$, $\lambda_{11s} = t+1$, $\lambda_{21s} = t+2$ *and* $b_{2s} = 6t$, $r_{2s} = 3t$, $k_{2s} = 3$, $\lambda_{12s} = t+1$, $\lambda_{22s} = t$, $t = 1, 2$,

c) $b_{1s} = 12$, $r_{1s} = 6$, $k_{1s} = 3$, $\lambda_{11s} = 4$, $\lambda_{21s} = 2$ *and* $b_{2s} = 2(5t+4)$, $r_{2s} = 5t+4$, $k_{2s} = 3$, $\lambda_{12s} = 2t$, $\lambda_{22s} = 2(t+1)$, $t = 0, 1$,

d) $b_{1s} = 16$, $r_{1s} = 8$, $k_{1s} = 3$, $\lambda_{11s} = 4$, $\lambda_{21s} = 3$ *and* $b_{2s} = 2(5t+2)$, $r_{2s} = 5t+2$, $k_{2s} = 3$, $\lambda_{12s} = t+2$, $\lambda_{22s} = 2t+1$, $t = 0, 1$,

then for any matrix in (3.1), the condition (3.2) is satisfied.

Lemma 3.3. *Let $v = 8$. If the parameters of group divisible designs are equal to*

a) $b_{1s} = 4(t+1)$, $r_{1s} = 2(t+1)$, $k_{1s} = 4$, $\lambda_{11s} = 0$, $\lambda_{21s} = t+1$ *and* $b_{2s} = 4(6-t)$, $r_{2s} = 2(6-t)$, $k_{2s} = 4$, $\lambda_{12s} = 6$, $\lambda_{22s} = 5-t$, $t = 1, 2, 3$,

b) $b_{1s} = 2(3t+2)$, $r_{1s} = 3t+2$, $k_{1s} = 4$, $\lambda_{11s} = t+1$, $\lambda_{21s} = t+2$ *and* $b_{2s} = 6(4-t)$, $r_{2s} = 3(4-t)$, $k_{2s} = 4$, $\lambda_{12s} = 4-t$, $\lambda_{22s} = 5-t$, $t = 1, 2$,

then for any matrix in (3.1), the condition (3.2) is satisfied.

Lemma 3.4. *Let $v = 10$. If the parameters of group divisible designs are equal to $b_{1s} = 8t$, $r_{1s} = 4t$, $k_{1s} = 5$, $\lambda_{11s} = 0$, $\lambda_{21s} = 2t$ and $b_{2s} = 10t$, $r_{2s} = 5t$, $k_{2s} = 5$, $\lambda_{12s} = 4t$, $\lambda_{22s} = 2t$, $t = 1, 2$, then for any matrix in (3.1), the condition (3.2) is fulfilled.*

Lemma 3.5. *If the parameters of group divisible designs are equal to*

a) $v = 2(2u+1)$, $b_1 = 4u$, $r_1 = 2u$, $k_1 = 2u+1$, $\lambda_{11} = 0$, $\lambda_{21} = u$ *and* $v = b_2 = 2(2u+1)$, $r_2 = k_2 = 2u+1$, $\lambda_{12} = 2u$, $\lambda_{22} = u$, $u = 1, 2, 3, 4$,

b) $v = 4(u+1)$, $b_1 = 2(2u+1)$, $r_1 = 2u+1$, $k_1 = 2(u+1)$, $\lambda_{11} = 2u+1$, $\lambda_{21} = u$ *and* $v = b_2 = 4(u+1)$, $r_2 = k_2 = 2(u+1)$, $\lambda_{12} = 0$, $\lambda_{22} = u+1$, $u = 1, 2, 3, 4$,

then for a matrix $\mathbf{X} = \mathbf{X}_s$ in (3.1), the condition (3.2) is true.

Lemmas given above are essential to construct the design $\mathbf{X}$. For a given number of objects $p = v$ and $n = \sum_{s=1}^{h}(b_{1s} + b_{2s})$ measurements, we choose appropriate number h of matrices $\mathbf{X}_s$ satisfying conditions in Lemmas 3.1–3.5 and in the result we form the design matrix $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0, 1)$. Thus we obtain the theorem.

Theorem 3.3. *Any $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(0, 1)$ in (2.2) for $\mathbf{X}_s$ in (3.1), where $\mathbf{N}_{uh}$ is the incidence matrix of group divisible design with the parameters given in Lemmas 3.1–3.5, $u = 1, 2$, with the variance matrix of errors $\sigma^2\mathbf{G}$ is the regular A-optimal spring balance weighing design.*

Proof: It is easy to verify that for the matrix $\mathbf{X}$ the condition a) of Theorem 2.2 is satisfied. □

Note 3.1. The criterion of A-optimality is interpreted as minimizing the sum of the variances of estimators of unknown measurements of objects. Some design matrices are better than others in the sense that the sum of the variances of estimators of unknown measurements of objects is smaller. The design $\mathbf{X}$ for which $\mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}$ attains the lower bound is called regular A-optimal. In any class $\mathbf{\Phi}_{n\times p}(0,1)$, regular A-optimal spring balance weighing design $\mathbf{X}_R$ may exist, whereas A-optimal spring balance weighing design $\mathbf{X}$ exists always. From statistical point of view, the sum of variances in regular A-optimal spring balance weighing design $\mathbf{X}_R$ is so small if it is possible and in this sense design $\mathbf{X}_R$ is the best one. Moreover, if in the class $\mathbf{\Phi}_{n\times p}(0,1)$, the regular A-optimal spring balance weighing design does not exist then determined lower bound of the variance of the sum of estimators may be used for indicating the design which is the closest to the best one $\mathbf{X}_R$.

4. EXAMPLES

Example 4.1. To present theory given above, let us consider $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ and let p be even. Among all possible variance matrices $\sigma^2\mathbf{G}$ we take matrix $\mathbf{G}$ in the form

a) $\mathbf{G}_1 = \begin{bmatrix} a\mathbf{I}_{\frac{n}{2}} & \mathbf{0}_{\frac{n}{2}}\mathbf{0}'_{\frac{n}{2}} \\ \mathbf{0}_{\frac{n}{2}}\mathbf{0}'_{\frac{n}{2}} & \frac{1}{a}\mathbf{I}_{\frac{n}{2}} \end{bmatrix}$, $a>0$. We have $\mathrm{tr}(\mathbf{X}'\mathbf{G}_1^{-1}\mathbf{X})^{-1} = \frac{8a(p^2-2p+2)}{np(a^2+1)}$ and $\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} - \mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \frac{4(p^2-2p+2)(a-1)^2}{np(a^2+1)} \geq 0$. Hence, for $a\neq 1$, $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with the variance matrix $\sigma^2\mathbf{G}$ is regular A-optimal spring balance weighing design, whereas $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with $\sigma^2\mathbf{I}_n$ is A-optimal design. For $a=1$, $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ is regular A-optimal spring balance weighing design with the variance matrix $\sigma^2\mathbf{G}_1$ and $\sigma^2\mathbf{I}_n$.

b) $\mathbf{G}_2 = \begin{bmatrix} \frac{1}{a}\mathbf{I}_{\frac{n}{2}} & \mathbf{0}_{\frac{n}{2}}\mathbf{0}'_{\frac{n}{2}} \\ \mathbf{0}_{\frac{n}{2}}\mathbf{0}'_{\frac{n}{2}} & \frac{1}{b}\mathbf{I}_{\frac{n}{2}} \end{bmatrix}$, $a,b>0$. We have $\mathrm{tr}(\mathbf{X}'\mathbf{G}_2^{-1}\mathbf{X})^{-1} = \frac{8(p^2-2p+2)}{np(a+b)}$ and

$$\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} - \mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \frac{4(p^2-2p+2)(a+b-2)}{np(a+b)}\,. \tag{4.1}$$

If $a+b-2>0$ then $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with the variance matrix $\sigma^2\mathbf{G}_2$ is regular A-optimal spring balance weighing design, whereas $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with $\sigma^2\mathbf{I}_n$ is A-optimal design. If $a+b-2=0$ then $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with the variance matrix $\sigma^2\mathbf{G}_2$ and with $\sigma^2\mathbf{I}_n$ is regular A-optimal design. If $a+b-2<0$ then $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with the variance matrix $\sigma^2\mathbf{I}_n$ is regular A-optimal spring balance weighing design, whereas $\mathbf{X}\in\mathbf{\Phi}_{n\times p}(0,1)$ with $\sigma^2\mathbf{G}_2$ is A-optimal design.

Example 4.2. As numerical example, let us consider $\mathbf{X} \in \mathbf{\Phi}_{12\times 4}(0,1)$. From Theorem 3.1, we construct the incidence matrix $\mathbf{N}$ of balanced incomplete block design with the parameters $v=4,\ b=6,\ r=3,\ k=2,\ \lambda=1$ as $\mathbf{N} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$ and next we form the design matrix $\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$. Hence $\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} = \frac{5}{6}$. Next, we consider possible forms of the matrix $\mathbf{G}$. For example

$$\text{if } \mathbf{G} = \begin{bmatrix} 2\mathbf{I}_6 & \mathbf{0}_6\mathbf{0}_6' \\ \mathbf{0}_6\mathbf{0}_6' & \frac{1}{2}\mathbf{I}_6 \end{bmatrix} \quad \text{then} \quad \mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \frac{4}{6} < \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} ,$$

$$\text{if } \mathbf{G} = \begin{bmatrix} 2\mathbf{I}_6 & \mathbf{0}_6\mathbf{0}_6' \\ \mathbf{0}_6\mathbf{0}_6' & \frac{2}{3}\mathbf{I}_6 \end{bmatrix} \quad \text{then} \quad \mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \frac{5}{6} = \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} ,$$

$$\text{if } \mathbf{G} = \begin{bmatrix} \frac{1}{3}\mathbf{I}_6 & \mathbf{0}_6\mathbf{0}_6' \\ \mathbf{0}_6\mathbf{0}_6' & \frac{1}{2}\mathbf{I}_6 \end{bmatrix} \quad \text{then} \quad \mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \frac{2}{6} < \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} ,$$

$$\text{if } \mathbf{G} = \begin{bmatrix} 2\mathbf{I}_6 & \mathbf{0}_6\mathbf{0}_6' \\ \mathbf{0}_6\mathbf{0}_6' & \frac{3}{2}\mathbf{I}_6 \end{bmatrix} \quad \text{then} \quad \mathrm{tr}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1} = \frac{10}{7} > \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} .$$

As you can see, in some cases the sum of the variances of estimators is smaller for the design $\mathbf{X}$ with $\mathbf{G}$ than for this design with $\mathbf{I}_n$. In the other ones it is inversely. Interestingly enough, it depends on the experimental conditions and the assumptions related to the variances of the errors: are they equal and $\mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{I}_n$ or are they different and $\mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{G}$. Into practice, the choice of the design $\mathbf{X}$ with $\mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{I}_n$ or $\sigma^2\mathbf{G}$ is conditional and depends on the experimental conditions. We choose one of these ones. Next, we can assess for which variance matrix of errors the sum of variances of errors is smaller. If the experimental conditions require the design with greater sum of variances of errors we can determine how far we are from the lowest bound considering the difference (4.1).

ACKNOWLEDGMENTS

The author wishes to express her gratitude to Reviewers for many constructive suggestions and comments which improved the paper.

REFERENCES

[1] Banerjee, K.S. (1975). *Weighing Designs for Chemistry, Medicine, Economics, Operations Research, Statistics*, Marcel Dekker Inc., New York.

[2] Ceranka, B. and Graczyk, M. (2004). A-optimal chemical balance weighing design, *Folia Facultatis Scientiarum Naturalium Universitatis Masarykianae Brunensis, Mathematica*, **15**, 41–54.

[3] Ceranka, B.; Graczyk, M. and Katulska, K. (2006). A-optimal chemical balance weighing design with nonhomogeneity of variances of errors, *Statistics and Probability Letters*, **76**, 653–665.

[4] Ceranka, B.; Graczyk, M. and Katulska, K. (2007). On certain A-optimal chemical balance weighing designs, *Computational Statistics and Data Analysis*, **51**, 5821–5827.

[5] Clatworthy, W.H. (1973). *Tables of Two-Associate-Class Partially Balanced Design*, NBS Applied Mathematics Series 63.

[6] Graczyk, M. (2011). A-optimal biased spring balance weighing design, *Kybernetika*, **47**, 893–901.

[7] Graczyk, M. (2012). Notes about A-optimal spring balance weighing design, *Journal of Statistical Planning and Inference*, **142**, 781–784.

[8] Jacroux, M. and Notz, W. (1983). On the optimality of spring balance weighing designs, *The Annals of Statistics*, **11**, 970–978.

[9] Katulska, K. and Rychlińska, E. (2010). On regular E-optimality of spring balance weighing designs, *Colloquium Biometricum*, **40**, 165–176.

[10] Masaro, J. and Wong, C.S. (2008). Robustness of A-optimal designs, *Linear Algebra and its Applications*, **429**, 1392–1408.

[11] Pukelsheim, F. (1983). *Optimal Design of Experiment*, John Wiley and Sons, New York.

[12] Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Designs of Experiment*, John Wiley Inc., New York.

[13] Raghavarao, D. and Padgett, L.V. (2005). *Block Designs, Analysis, Combinatorics and Applications*, Series of Applied Mathematics 17, Word Scientific Publishing Co. Pte. Ltd.

[14] Shah, K.R. and Sinha, B.K. (1989). *Theory of Optimal Designs*, Springer-Verlag, Berlin, Heidelberg.

REVSTAT – Statistical Journal
Volume 10, Number 3, November 2012, 335–349

ESTIMATING OF THE PROPORTIONAL HAZARD PREMIUM FOR HEAVY-TAILED CLAIM AMOUNTS WITH THE POT METHOD

Author: Rassoul Abdelaziz
– GEE Laboratory, National High School of Hydraulic, Blida, Algeria
rsl_aziz@yahoo.fr

Received: October 2011 Revised: May 2012 Accepted: July 2012

Abstract:

- In this paper we propose a new estimator of the proportional hazard premium for heavy-tailed claim amounts, with a help of the peak-over-threshold (POT) method. We establish the asymptotic normality of the new estimator, and its performance is illustrated in a simulation study. Moreover, we compare, in terms of bias and mean squared error, our estimator with the estimator of Necir and Meraghni (2009).

Key-Words:

- *distortion risk measures; proportional hazard premium; extreme values; GPD function; heavy tails; POT method.*

AMS Subject Classification:

- 62G32, 31B30.

1. INTRODUCTION AND MOTIVATION

A general class for constructing loaded pricing functional, introduced in the actuarial literature by Wang (1996), is namely the Distortion Risk Measures (DRM). For a given nondecreasing function $g\colon [0,1] \to [0,1]$ such that $g(0) = 0$ and $g(1) = 1$, for any nonnegative random variable X, where X is an insured risk with distribution function (df) F, and the tail of F will be denoted by $\overline{F} = 1 - F$. The distorted expectation is defined as follows:

$$\Pi_\rho = \int_0^\infty g\big(\overline{F}(x)\big)\, dx \ . \tag{1.1}$$

The function g is called a distortion function, if g is concave the DRM further satisfies the subadditivity and becomes coherent in the sense of Artzner *et al.* (1999); see, e.g., Wirch and Hardy (2000) and Dhaene *et al.* (2006). Some examples of continuous concave distortion functions corresponding to familiar risk measures are presented below, by choosing a suitable function g, one can easily express some popular risk measures:

The Tail-VaR: $g(x) = x/(1-q) \wedge 1,\ q \in (0,1)$,
Proportional Hazard Transform: $g(x) = x^{1/\rho},\ \rho \geq 1$,
Dual-Power Transform: $g(x) = 1 - (1-x)^\rho,\ \rho > 1$,
Wang Transform: $g(x) = \Phi\big(\Phi^{-1}(x) - \Phi^{-1}(\rho)\big)$,
where $\Phi(\cdot)$ is the df of the standard normal.

In this paper, we are interested by estimate the proportional hazard transform, that is

$$\Pi_\rho = \int_0^\infty \big(\overline{F}(x)\big)^{1/\rho}\, dx \ , \tag{1.2}$$

where $\rho \geq 1$ represents the distortion coefficient or the risk aversion index.

In practice the estimation of these risk measures from a sample of rv's i.i.d. $X_1, X_2, ..., X_n$, are based on the empirical distribution function F_n. The asymptotic behavior of this estimator has been studied by Jones and Zitikis (2003) provided that, the second moment is finite.

Now, assume that F is heavy tailed. This class includes popular distributions (such as Pareto, Burr, Student, Lévy-stable and log-gamma) known to be very appropriate models for fitting large insurance claims, large fluctuations of prices, log-returns and other data (see for instance, Beirlant *et al.*, 2001). In the remainder of the paper, we restrict ourselves to this class, more specifically, we deal within the class of regularly varying cdf's. For more details on this type of distributions, we refer to Bingham *et al.* (1987) and Rolski *et al.* (1999).

The tail of F is said to be with regulary varying at infinity, if

$$\overline{F}(x) \,=\, c\,x^{-1/\xi}\big(1+x^{-\delta}\,\mathbb{L}(x)\big)\,, \qquad \text{as} \quad x\to\infty\,, \tag{1.3}$$

for $\xi\in(0,1)$, $\delta>0$ and some real constant c, with $\mathbb{L}$ a slowly varying function, i.e. $\mathbb{L}(tx)/\mathbb{L}(x)\to 1$ as $x\to\infty$ for any $t>0$. For further properties of these functions, see chapter 0 in Resnick (1987) or Seneta (1976).

For example, when the tail of df is Pareto with $\xi=3/4$, we have $\mathbb{E}(X^2)=\infty$. To solve this problem, Necir and Meraghni (2009) used the extreme values approach and propose an asymptotically normal semiparametric estimator for Π_ρ. This estimator is based on the extreme quantile estimator of Weissman (1978). However this quantile is biased.

In this paper, we use the result of Balkema and de Haan (1974) and Pickands (1975), which states that for a certain class of distributions the Generalised Pareto Distribution (GPD) is the limiting for the distribution of the excesses F_u, as the threshold u tends to the right endpoint y_F. Formally, we can find a positive measurable function $\beta(u)$, such as

$$\lim_{u\to y_F}\,\sup_{0<y<y_F-u}\big|F_u(y)-\mathbb{G}_{\xi,\beta(u)}(y)\big| \,=\, O\big(u^{-\delta}\,\mathbb{L}(u)\big)\,, \tag{1.4}$$

where $u^{-\delta}\,\mathbb{L}(u)\to 0$ as $u\to\infty$, for any $\delta>0$.

We investigate this result for purpose a alternative estimator for the proportional hazard transform Π_ρ, as follows:

$$\widehat{\Pi}_{\rho,n} \,=\, \int_0^{u_n}\left(n^{-1}\sum_{j=1}^{n}\mathbf{1}\big(X_j\geq x\big)\right)^{1/\rho}dx \,+\, (\widehat{p}_n)^{1/\rho}\frac{\rho\,\widehat{\beta}_n}{1-\widehat{\xi}_n\,\rho}\,. \tag{1.5}$$

Under suitable assumptions, this estimator are asymptotically normal distributed and unbiased with an easily estimated variance.

The paper is organized as follows. In the second section of the paper, the new estimator of Π_ρ is introduced and its properties examined. This is followed by a simulation study of its behavior in comparison with the Necir and Meraghni estimator. Finally, the proofs of our result are postponed until the last section.

2. DEFINING THE ESTIMATOR AND THE MAIN RESULT

Let $X_1,...,X_n$ be an independent and identically distributed random variables, each with the same cdf F, and let u_n be some a large number, 'high level', which we later let tends to infinity when $n\to\infty$. With the notation

$$\overline{F}_{u_n}(y) \,=\, P\Big(X_1-u_n>y\mid X_1>u_n\Big)\,,$$

we have

$$\overline{F}_{u_n}(y) \;=\; \overline{F}(u_n+y)/\overline{F}(u_n)\;,$$

and thus

$$\overline{F}_{u_n}(y) \;=\; \left(1+\frac{y}{u_n}\right)^{-1/\xi}\,\frac{1+(u_n+y)^{-\delta}\mathbb{L}(u_n+y)}{1+u_n^{-\delta}\mathbb{L}(u_n)}\;,$$

and if $\beta = u_n\xi$, then $\overline{F}_{u_n}(y)$ is a GPD perturbed, where the df of the GPD has the form

$$\mathbb{G}_{\xi,\beta}(y) = \begin{cases} 1-\big(1+\xi\,\frac{y}{\beta}\big)^{-\frac{1}{\xi}}, & \xi\neq 0, \quad 0\le y<\infty \;\text{ if } \xi\ge 0\;,\\ 1-\exp\big(-y/\beta\big), & \xi=0, \quad 0\le y<-\beta/\xi \;\text{ if } \xi<0\;. \end{cases} \tag{2.1}$$

This means that, with the result (1.4) of Balkema and de Haan (1974) and Pickands (1975), for large values of u_n, we have

$$F_{u_n}(y) \;\approx\; \mathbb{G}_{\xi,\beta(u_n)}(y)\;. \tag{2.2}$$

By the definition of the excess distribution, we have

$$\overline{F}(u_n+t) \;=\; \overline{F}(u_n)\,\overline{F}_{u_n}(t)\;,$$

and, denote by

$$N = N_{u_n} = \;\text{card}\Big\{X_i > u_n\colon 1\le i\le n\Big\}\,,$$

the number of exceedance over u_n, we have $N \rightsquigarrow \mathcal{B}(p_n, n)$, where $p_n = P(X_1 > u_n)$. A natural estimator for $p_n = \overline{F}(u_n)$ is $\widehat{p}_n = N/n$. Let

$$Y_{i,n} = X_j - u_n\,, \qquad \text{provided} \quad X_j > u_n, \quad i = 1, ..., N\;,$$

(where j is the index of the i^{th} exceedance) are i.i.d. rv's with cdf F_{u_n} based on the sample $(Y_{1:n}, Y_{2:n}, ..., Y_{N_n,n})$, the approximation (2.2) motivates us to take an estimator for $\overline{F}_{u_n}(y)$ as follows:

$$\widehat{\overline{F}}_{u_n}(y) = \overline{\mathbb{G}}_{\widehat{\xi}_n,\widehat{\beta}_n}(y)\;, \qquad y>0\;. \tag{2.3}$$

Therefore, an estimator of $\overline{F}(u_n+y)$ is

$$\widehat{\overline{F}}(u_n+y) \;=\; \widehat{\overline{F}}(u_n)\,\widehat{\overline{F}}_{u_n}(y) \;=\; \widehat{p}_n\,\overline{\mathbb{G}}_{\widehat{\xi}_n,\widehat{\beta}_n}(y)\;, \tag{2.4}$$

where $\widehat{\xi}_N$ and $\widehat{\beta}_N$ are consistent estimators of ξ and β respectively. Moreover, these estimators are asymptotically normal provided that $\xi > -1/2$. Smith (1987) established in theorem (3.2), the asymptotic normality of $(\widehat{\xi}_N, \widehat{\beta}_N)$ as follows:

$$\sqrt{N}\begin{pmatrix} \widehat{\beta}_N/\beta_N - 1 \\ \widehat{\xi}_N - \xi \end{pmatrix} \;\xrightarrow{\mathcal{D}}\; \mathcal{N}_2(0, \mathbb{Q}^{-1}) \qquad \text{as} \quad N\to\infty\;, \tag{2.5}$$

where

$$\mathbb{Q}^{-1} = (1+\xi)\begin{pmatrix} 2 & -1 \\ -1 & 1+\xi \end{pmatrix}, \tag{2.6}$$

provided that $\sqrt{N}u_N^{-\delta}\,\mathbb{L}(u_N) \to 0$ as $N \to \infty$ and $x \to x^{-\delta}\mathbb{L}(x)$ is non-increasing near infinity. In the case $\sqrt{N}u_N^{-\delta}\,\mathbb{L}(u_N) \nrightarrow 0$, the limiting distribution in (2.5) is biased.
Here $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $\mathcal{N}_2(0,\epsilon^2)$ stands for the normal r.v. of mean 0 and variance ϵ^2.

We assume that the tail of the distribution start at the threshold u_n, then, we have

$$\Pi_\rho = \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho} dx + \int_{u_n}^{\infty} \big(\overline{F}(x)\big)^{1/\rho} dx\,, \qquad \rho > 1\,. \tag{2.7}$$

An estimator of Π_ρ is given by replacing (2.4) in equation (2.7), as follows:

$$\widehat{\Pi}_{\rho,n}(x) = \int_0^{u_n} \big(\overline{F}_n(x)\big)^{1/\rho} dx + (\widehat{p}_n)^{1/\rho} \int_0^{\infty} \big(\overline{\mathbb{G}}_{\widehat{\xi}_n,\widehat{\beta}_n}(y)\big)^{1/\rho} dy\,,$$

where F_n is the empirical distribution function pertaining to the sample $X_1, X_2, ..., X_n$. After Integration, we obtain the new estimator given by formula (1.5).

The asymptotic normality of $\widehat{\Pi}_{\rho,n}$ is established in the following theorem.

Theorem 2.1. *Let F be a distribution function fulfilling (1.3) with $\xi \in (1/2, 1)$. Suppose that the function $\mathbb{L}$ is locally bounded in $[x_0, +\infty)$ for $x_0 \geq 0$ and $x \to x^{-\delta}\mathbb{L}(x)$ is non-increasing near infinity, for some $\delta > 0$. For any $u_n = O(n^{\alpha\xi})$ with $\alpha \in (0,1)$, and $\rho > 1$ such that $4\alpha/\rho - 2\alpha\,\xi < 1$, we have*

$$\frac{\sqrt{n}}{\gamma_n \sigma_n}\big(\widehat{\Pi}_{\rho,n} - \Pi_\rho\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,, \qquad \text{as} \quad n \to \infty\,,$$

where

$$\sigma_n^2 := \frac{1}{\rho^2} + \frac{\theta_1^2}{\gamma_n^2}\,p_n(1-p_n) + \frac{2(1+\xi)\,\theta_2^2\,\beta_n^2}{p_n\gamma_n^2} + \frac{(1+\xi)^2\,\theta_3^2}{p_n\gamma_n^2} - \frac{(1+\xi)\,\beta_n\theta_2\,\theta_3}{p_n\gamma_n^2}\,,$$

$$\gamma_n^2 = \mathrm{var}\left(\int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho-1} \mathbf{1}_{\{X_1 \leq x\}}\, dx\right), \tag{2.8}$$

and

$$\theta_1 = \frac{\beta_n (p_n)^{1/\rho-1}}{1-\xi\rho}\,, \qquad \theta_2 = \frac{\rho (p_n)^{1/\rho}}{1-\xi\rho}\,, \qquad \theta_3 = \frac{\rho^2 \beta_n (p_n)^{1/\rho}}{(1-\xi\rho)^2}\,,$$

with $\beta_n = u_n\,\xi$.

3. SIMULATION STUDY

In this section, we carry out a simulation study (by means of the statistical software **R**, see Ihaka and Gentleman, 1996) to illustrate the performance of our estimation procedure and its comparison with the estimator of Necir and Meraghni (2009). We generate samples from Fréchet distributions with tail $\overline{F}(x) = 1 - \exp(-x^{-1/\xi})$, $x > 0$ (with tail index $\xi = 2/3$ and $\xi = 3/4$) and two distinct aversion index values $\rho = 1.1$ and $\rho = 1.2$.

In the first part, we evaluate the accuracy of the confidence intervals via their lengths and coverage probabilities (cov prob), we generate 200 independent replicates of sizes 1000 and 2000 from the selected parent distribution. For each simulated sample, we obtain a value of the estimators premium Π_ρ. The overall estimated premium Π_ρ is then taken as the empirical mean of the values in the 200 repetitions. We summarize the results in Table 1.

Table 1: Point estimates and 95%-confidence intervals for Π, based on 200 samples of Fréchet distributed rv's with tail index 2/3 and 3/4 with aversion index 1.1 and 1.2.

ξ	2/3				3/4			
ρ	1.1		1.2		1.1		1.2	
Π	3.439		4.699		5.351		9.645	
n	1000	2000	1000	2000	1000	2000	1000	2000
$\widehat{\Pi}_{\rho,n}$	3.775	3.571	4.517	4.619	5.450	5.441	9.439	9.692
rmse	0.570	0.515	0.781	0.639	0.187	0.127	0.109	0.107
lcb	2.661	2.896	2.166	3.803	3.567	3.744	5.921	6.909
ucb	4.482	4.655	6.867	6.434	7.333	7.139	12.957	12.476
length	1.821	1.759	4.701	2.631	3.765	3.395	7.036	5.567
cprob	0.785	0.815	0.75	0.821	0.975	0.98	0.85	0.85

In the second part in this study, we generate 200 independent replicates of sizes 1000 from a Fréchet distribution, we compare the bias and the root mean squared error (RMSE) of the two estimators of Π_ρ (our estimator $\widehat{\Pi}_{\rho,n}$ with the estimator of Necir and Meraghni $\widetilde{\Pi}_{\rho,n}$). The results are presented in Table 2.

Table 2: Analog between the new estimator and the estimator of Necir and Meraghni for the premium hazard proportional for two tail index and two risk aversions index.

ξ	2/3		3/4	
ρ	1.1	1.2	1.1	1.2
Π_ρ	3.44	4.699	5.350	9.645
$\widehat{\Pi}_{\rho,n}$	3.527	4.807	5.359	9.499
bias	0.087	0.108	0.009	0.142
RMSE	0.335	0.592	0.516	0.933
$\widetilde{\Pi}_{\rho,n}$	4.221	4.938	5.452	9.915
bias	0.781	0.238	0.136	0.262
RMSE	0.867	0.665	0.674	1.131

4. PROOF OF THE MAIN RESULT

The following proposition is instrumental for the proof of our result

Proposition 4.1. *Let F be a distribution function fulfilling (1.3) with $\xi \in (0,1)$, $\delta > 0$ and some real c. Suppose that $\mathbb{L}$ is locally bounded in $[x_0, +\infty)$ for $x_0 \geq 0$. Then, for n large enough, for any $u_n = O(n^{\alpha\xi})$, $\alpha \in (0,1)$, we have*

$$p_n = P\big(X_1 > u_n\big) = c\big(1 + o(1)\big)\, n^{-\alpha}\,,$$

$$\gamma_n^2 = \mathrm{var}\bigg(\int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} \mathbf{1}_{\{X_1 \leq x\}}\, dx\bigg) = O\big(n^{2\alpha(\xi - 1/\rho + 1)}\big)\,,$$

and

$$\sqrt{np_n}\, u_n^{-\delta}\, \mathbb{L}(u_n) = O\big(n^{-\alpha/2 - \alpha\xi\delta + 1/2}\big)\,.$$

Proof of the Theorem 2.1: Let us write

$$\sqrt{n}\,\big(\widehat{\Pi}_{\rho,n} - \Pi_\rho\big) = A_n + B_n\,, \tag{4.1}$$

where

$$A_n = \sqrt{n} \int_0^{u_n} \Big[\big(\overline{F}_n(x)\big)^{1/\rho} - \big(\overline{F}(x)\big)^{1/\rho}\Big]\, dx\,,$$

and

$$B_n = \sqrt{n}\left(\widehat{p}_n^{\,1/\rho} \frac{\rho\,\widehat{\beta}_n}{1 - \widehat{\xi}_n\,\rho} - \int_{u_n}^{\infty} \big(\overline{F}(x)\big)^{1/\rho}\, dx\right).$$

We begin by B_n, we may rewrite B_n as follows:

$$B_n = B_{n,1} + B_{n,2}\,,$$

where

$$B_{n,1} \;=\; (\widehat{p}_n)^{1/\rho}\frac{\rho\,\widehat{\beta}_n}{1-\widehat{\xi}_n\,\rho} \;-\; (p_n)^{1/\rho}\frac{\rho\,\beta_n}{1-\xi\,\rho}\;,$$

and

$$B_{n,2} \;=\; (p_n)^{1/\rho}\frac{\rho\,\beta_n}{1-\xi\,\rho} \;-\int_{u_n}^{\infty}\big(\overline{F}(s)\big)^{1/\rho}\,ds\;.$$

First, observe that $B_{n,1}$, may be rewrite into

$$\begin{aligned} B_{n,1} \;=\; & \frac{\rho\,\widehat{\beta}_n}{1-\widehat{\xi}_n\,\rho}\,\sqrt{n}\Big((\widehat{p}_n)^{1/\rho}-(p_n)^{1/\rho}\Big) \\ & +\,(p_n)^{1/\rho}\,\frac{\rho}{1-\widehat{\xi}_n\,\rho}\,\sqrt{n}\big(\widehat{\beta}_n-\beta_n\big) \\ & +\,\frac{\rho^2\,\beta_n(p_n)^{1/\rho}}{(1-\widehat{\xi}_n\,\rho)\,(1-\xi\,\rho)}\,\sqrt{n}\big(\widehat{\xi}_n-\xi\big)\;. \end{aligned}$$

From Smith (1987), we have, as $n\to\infty$

$$\widehat{\beta}_n/\beta_n-1=O_{\mathbb{P}}\big(u_n^{-\delta}\,\mathbb{L}(u_n)\big) \qquad\text{and}\qquad \widehat{\xi}_n-\xi=O_{\mathbb{P}}\big(u_n^{-\delta}\,\mathbb{L}(u_n)\big)\;. \tag{4.2}$$

On the other hand, by the central limit theorem, we have

$$\widehat{p}_n-p_n=O_{\mathbb{P}}\big(\sqrt{p_n/n}\,\big) \qquad\text{as}\quad n\to\infty\;. \tag{4.3}$$

Then, with the delta method, we obtain

$$\begin{aligned} B_{n,1} \;=\; & \theta_1\big(1+o_{\mathbb{P}}(1)\big)\,\sqrt{n}\,(\widehat{p}_n-p_n) \\ & +\,\theta_2\big(1+o_{\mathbb{P}}(1)\big)\,\sqrt{n}\,(\widehat{\beta}_n-\beta_n) \\ & +\,\theta_3(1+o_{\mathbb{P}}(1))\,\sqrt{n}\,(\widehat{\xi}_n-\xi)\;, \end{aligned}$$

where

$$\theta_1=\frac{\beta(p_n)^{1/\rho-1}}{1-\xi\,\rho}\;,\qquad \theta_2=\frac{\rho(p_n)^{1/\rho}}{1-\xi\,\rho}\;,\qquad \theta_3=\frac{\rho^2\,\beta(p_n)^{1/\rho}}{(1-\xi\rho)^2}\;.$$

Either, for $B_{n,2}$, we have

$$B_{n,2} \;=\; (p_n)^{1/\rho}\frac{\rho\,\beta_n}{\xi\,\rho-1} \;-\int_{u_n}^{\infty}\big(\overline{F}(s)\big)^{1/\rho}\,ds\;.$$

We may rewrite

$$\overline{F}_{u_n}(s) \;=\; \frac{\overline{F}(u_n+s)}{\overline{F}(u_n)} \;=\; \left(1+\frac{s}{u_n}\right)^{-1/\xi}\,\frac{1+(u_n+s)^{-\delta}\,\mathbb{L}(u_n+s)}{1+u_n^{-\delta}\,\mathbb{L}(u_n)}\;.$$

This allows us to rewrite

$$\begin{aligned}
\int_0^\infty \big(\overline{F}(s+u_n)\big)^{1/\rho}\,ds \;&=\\
&= \big(\overline{F}(u_n)\big)^{1/\rho}\int_0^\infty \big(\overline{F}_{u_n}(s)\big)^{1/\rho}\,ds\\
&= \big(\overline{F}(u_n)\big)^{1/\rho}\int_0^\infty \left[\left(1+\frac{s}{u_n}\right)^{-1/\xi}\frac{1+(u_n+s)^{-\delta}\,\mathbb{L}(u_n+s)}{1+u_n^{-\delta}\mathbb{L}(u_n)}\right]^{1/\rho} ds\\
&= p_n^{1/\rho}\left(\frac{1}{1+u_n^{-\delta}\mathbb{L}(u_n)}\right)^{1/\rho}\\
&\quad\times\int_0^\infty \left[\left(1+\frac{s}{u_n}\right)^{-1/\xi}\Big(1+(u_n+s)^{-\delta}\,\mathbb{L}(u_n+s)\Big)\right]^{1/\rho} ds\\
&= p_n^{1/\rho}\left(\frac{1}{1+u_n^{-\delta}\mathbb{L}(u_n)}\right)^{1/\rho} u_n^{1/\xi\rho}\int_{u_n}^\infty x^{-1/\xi\rho}\big(1+x^{-\delta}\mathbb{L}(x)\big)^{1/\rho}\,dx\\
&= p_n^{1/\rho}\left(\frac{1}{1+u_n^{-\delta}\mathbb{L}(u_n)}\right)^{1/\rho} u_n^{1/\xi\rho}\\
&\quad\times\left[\left(\frac{\xi\rho}{1-\xi\rho}\,u_n^{1-1/\xi\rho}\right)+\int_{u_n}^\infty x^{-1/\xi\rho-\delta}\,\mathbb{L}(x)^{1/\rho}\,dx\right].
\end{aligned}$$

Since function $\mathbb{L}$ is locally bounded in $[x_0,\infty)$ for $x_0\geq 0$ and $x^{-\delta}\mathbb{L}(x)$ is non-increasing near infinity, then for all large n, we have

$$u_n^{1/\xi\rho}\int_{u_n}^\infty x^{-1/\xi\rho-\delta}\,\mathbb{L}(x)^{1/\rho}\,dx \;=\; O\big(u_n^{-\delta}\big)\;,$$

and therefore, for all large n

$$\int_{u_n}^\infty \overline{F}(x)^{1/\rho}\,dx \;=\; p_n^{1/\rho}\frac{\beta_n\rho}{1-\xi\rho}\Big(1-u_n^{-\delta}\mathbb{L}(u_n)+O\big(u_n^{-\delta}\mathbb{L}(u_n)\big)\Big)^{1/\rho}.$$

Consequently

$$B_{n,2} = O\big(u_n^{1-1/\rho\xi-\delta/\rho}\big)\;,$$

which means, since $1-1/\rho\xi-\delta/\rho<0$, that $B_{n,2}\xrightarrow{P}0$ as $n\to\infty$.
For A_n, we have

$$A_n \;=\; \sqrt{n}\int_0^{u_n}\Big[\big(\overline{F}_n(x)\big)^{1/\rho}-\big(\overline{F}(x)\big)^{1/\rho}\Big]\,dx\;. \tag{4.4}$$

We next show that, the right-hand side of (4.4), converge to 0 in probability, by

the use of the Taylor formula, we have

$$\begin{aligned}
\int_0^{u_n} \Big[\big(\overline{F}_n(x)\big)^{1/\rho} - \big(\overline{F}(x)\big)^{1/\rho} \Big]\, dx \;&= \\
&= \frac{1}{\rho} \int_0^{u_n} \big(\overline{F}_n(x) - \overline{F}(x)\big) \big(\overline{F}(x)\big)^{1/\rho - 1}\, dx \\
&= -\frac{1}{\rho} \int_0^{u_n} \big(F_n(x) - F(x)\big) \big(\overline{F}(x)\big)^{1/\rho - 1}\, dx \\
&= -\frac{1}{\rho} \int_0^{u_n} \left(\frac{1}{n} \sum \mathbf{1}(X_i \le x) - F(x) \right) \big(\overline{F}(x)\big)^{1/\rho - 1}\, dx \\
&= -\frac{1}{\rho} \left[\frac{1}{n} \sum \int_0^{u_n} \mathbf{1}(X_i \le x) \big(\overline{F}(x)\big)^{1/\rho - 1}\, dx \;- \int_0^{u_n} F(x) \big(\overline{F}(x)\big)^{1/\rho - 1}\, dx \right] \\
&= -\frac{1}{\rho} \Big[\overline{Z} - \mathbb{E}[Z_1] \Big] \,,
\end{aligned}$$

where

$$Z_i := \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} \mathbf{1}(X_i \le x)\, dx \,.$$

We assume that

$$\gamma_n^2 = \operatorname{var}(Z_1) \,.$$

We are going to calculate γ_n. For $\overline{F}(x) = x^{-1/\xi} O(1)$ and $u_n = n^{\alpha\xi} O(1)$, we have

$$\begin{aligned}
\mathbb{E}[Z_i] \;&= \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1}\, \mathbb{E}\big[\mathbf{1}(X_i \le x)\big]\, dx \\
&= \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} \big(1 - \overline{F}(x)\big)\, dx \\
&= \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} dx \;- \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho} dx \\
&= \left(\int_0^{u_n} \big(x^{-1/\xi(1/\rho - 1)}\big)\, dx \;- \int_0^{u_n} \big(x^{-1/\xi\rho}\big)\, dx \right) O(1) \\
&= \left(\frac{\rho\, \xi \big(u_n^{1 - 1/\xi\rho + 1/\xi}\big)}{\rho\, \xi + \xi - 1} - \frac{\rho\, \xi \big(u_n^{1 - 1/\xi\rho}\big)}{\xi\rho - 1} \right) O(1)
\end{aligned}$$

and

$$\begin{aligned}
E(Z_i^2) \;&= E\left[\int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} \mathbf{1}(X_i \le x)\, dx \int_0^{u_n} \big(\overline{F}(y)\big)^{1/\rho - 1} \mathbf{1}(X_i \le y)\, dy \right] \\
&= \left[\int_0^{u_n} \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} \big(\overline{F}(y)\big)^{1/\rho - 1} E\Big[\mathbf{1}(X_i \le x)\, \mathbf{1}(X_i \le y)\Big]\, dx\, dy \right] \\
&= \left[\int_0^{u_n} \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho - 1} \big(\overline{F}(y)\big)^{1/\rho - 1} \min\big(F(x), F(y)\big)\, dx\, dy \right] =
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho-1} \left(\int_0^x \big(\overline{F}(y)\big)^{1/\rho-1} F(y)\, dy \right) dx \\
&\quad + \int_0^{u_n} \big(\overline{F}(y)\big)^{1/\rho-1} \left(\int_x^{u_n} \big(\overline{F}(x)\big)^{1/\rho-1} F(x)\, dx \right) dy \\
&= \left(\frac{\rho^2 \xi^2 \big(u_n^{2(1-1/\xi\rho+1/\xi)}\big)}{(\rho\,\xi+\rho-1)^2} - \frac{2\,\rho^2 \xi^2 \big(u_n^{2-2/\xi\rho+1/\xi}\big)}{(\xi\,\rho-1)\,(2\,\rho\,\xi+\rho-2)} \right) O(1)\ ,
\end{aligned}
$$

we conclude that

$$
\gamma_n = n^{\alpha(\xi-1/\rho)}\, O(1)\ .
$$

Now, we show that

$$
\frac{\sqrt{n}}{\gamma_n} \big(\overline{Z} - \mathbb{E}[Z_1]\big) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0,1)\,, \qquad \text{as} \quad n \to \infty\ .
$$

With Lindeberg–Feller Theorem (see e.g. Chapter 2 in Durrett (1996)), note that

$$
\begin{aligned}
\frac{\sqrt{n}}{\gamma_n} \big(\overline{Z} - \mathbb{E}[Z_1]\big) &= \sum_{k=1}^n \frac{\displaystyle\int_0^{u_n} \big(\overline{F}(x)\big)^{1/\rho-1} \mathbf{1}(X_k \le x)\, dx \;-\; \mathbb{E}[Z_1]}{\gamma_n \sqrt{n}} \\
&= \sum_{k=1}^n S_{k,n}\ ,
\end{aligned}
$$

where

$$
\mathbb{E}(S_{k,n}) = 0\,, \quad \mathbb{E}(S_{k,n}^2) = 1/n \quad \text{and} \quad \sum_{k=1}^n \mathbb{E}(S_{k,n}^2) = 1 \qquad \text{for all} \quad n \ge 1\ .
$$

We need to show that

$$
\sum_{k=1}^n \mathbb{E}\Big[|S_{k,n}|^2\, \mathbf{1}\big(|S_{k,n}| > \epsilon\big) \Big] \to 0\,, \qquad \text{as} \quad n \to \infty\ .
$$

Indeed, we have

$$
\sum_{k=1}^n \mathbb{E}\Big[|S_{k,n}|^2\, \mathbf{1}\big(|S_{k,n}| > \epsilon\big) \Big] \;=\; \frac{1}{\gamma_n^2}\, \mathbb{E}\Big[\big[Z_k - \mathbb{E}[Z_1]\big]^2 \mathbf{1}\Big(\big|Z_k - \mathbb{E}[Z_1]\big| > \epsilon\, \gamma_n \sqrt{n} \Big) \Big]\ .
$$

Since $\big|Z_k - \mathbb{E}[Z_1]\big| \le u_n$, then the right side of the previous inequality is less or equal than

$$
\frac{u_n^2}{\gamma_n^2}\, \mathbb{E}\Big[\mathbf{1}\big|Z_k - \mathbb{E}[Z_1]\big| > \epsilon\, \gamma_n \sqrt{n} \Big] \;=\; \frac{u_n^2}{\gamma_n^2}\, \mathbb{P}\Big[\big|Z_k - \mathbb{E}[Z_1]\big| > \epsilon\, \gamma_n \sqrt{n} \Big]\ .
$$

In view of Tchebychev's inequality, we get

$$
\frac{u_n^2}{\gamma_n^2}\, \mathbb{P}\Big[\big|Z_k - \mathbb{E}[Z_1]\big| > \epsilon\, \gamma_n \sqrt{n} \Big] \;\le\; \frac{u_n^2}{\gamma_n^2}\, \frac{1}{\big(\epsilon\, \gamma_n \sqrt{n}\big)^2}\ .
$$

Further, for all $\alpha \in (0,1)$, $\xi \in (0,1)$ and $\epsilon > 0$, with $u_n = O(n^{\alpha\xi})$ was used, then

$$\frac{u_n^2}{\epsilon\, n\, \gamma_n^4} \;=\; n^{-2\alpha\xi+4\alpha/\rho-1} O(1)\;.$$

We must assume: $4\alpha/\rho - 2\,\alpha\xi < 1$ for that $\sum_{k=1}^{n} \mathbb{E}\Big[|S_{k,n}|^2\, \mathbf{1}\big(|S_{k,n}| > \epsilon\big)\Big] \to 0$ as $n \to \infty$.

Finally, we obtain that

$$\begin{aligned}\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\Pi}_{\rho,n} - \Pi_\rho\right) \;\to\; & -\frac{1}{\rho}\frac{\sqrt{n}}{\gamma_n}\left(\overline{Z} - \mathbb{E}[Z_1]\right) + \theta_1 \frac{\sqrt{p_n(1-p_n)}}{\gamma_n}\frac{\sqrt{n}\,(\widehat{p}_n - p_n)}{\sqrt{p_n(1-p_n)}} \\ & + \frac{\theta_2\,\beta_n}{\sqrt{p_n}\,\gamma_n}\,\sqrt{np_n}\left(\widehat{\beta}_n/\beta_n - 1\right) + \frac{\theta_3}{\sqrt{p_n}\,\gamma_n}\sqrt{np_n}\left(\widehat{\xi}_n - \xi\right) + o_{\mathbb{P}}(1)\;,\end{aligned}$$

This enable us to rewrite into

$$\begin{aligned}\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\Pi}_{\rho,n} - \Pi_\rho\right) \;\to\; & -\frac{1}{\rho}\,\mathcal{W}_1 + \theta_1 \frac{\sqrt{p_n(1-p_n)}}{\gamma_n}\,\mathcal{W}_2 \\ & + \frac{\sqrt{2(1+\xi)}\,\theta_2\,\beta_n}{\sqrt{p_n}\,\gamma_n}\,\mathcal{W}_3 + \frac{(1+\xi)\,\theta_3}{\sqrt{p_n}\,\gamma_n}\,\mathcal{W}_4 + o_{\mathbb{P}}(1)\;,\end{aligned}$$

where $(\mathcal{W}_i)_{i=1,4}$ are standard normal rv's with $E[\mathcal{W}_i\mathcal{W}_j] = 0$ for every $i, j = 1, ..., 4$, except for

$$\begin{aligned}E[\mathcal{W}_3\mathcal{W}_4] \;&=\; E\left[\frac{1}{\sqrt{2\,(1+\xi)}}\,\sqrt{np_n}\left(\widehat{\beta}_n/\beta_n - 1\right)\,\frac{1}{(1+\xi)}\,\sqrt{np_n}\left(\widehat{\xi}_n - \xi\right)\right] \\ &=\; \frac{1}{(1+\xi)\,\sqrt{2\,(1+\xi)}}\;E\Big[\sqrt{np_n}\left(\widehat{\beta}_n/\beta_n - 1\right)\sqrt{np_n}\left(\widehat{\xi}_n - \xi\right)\Big] \\ &=\; -\frac{1}{\sqrt{2\,(1+\xi)}}\;.\end{aligned}$$

From **Lemma A-2** of Johansson 2003, under the assumptions of Theorem 2.1, we have, for any real numbers, t_1, t_2, t_3 and t_4,

$$\begin{aligned}\mathbb{E}\left[\exp\left\{ i\,t_1\,\frac{\sqrt{n}}{\gamma_n}\left(\overline{Z} - \mathbb{E}[Z_1]\right) + i\,\sqrt{np_n}\,(t_2, t_3)\begin{pmatrix}\widehat{\beta}_n/\beta - 1 \\ \widehat{\xi}_n - \xi\end{pmatrix} + i\,t_4\,\frac{\sqrt{n}\,(\widehat{p}_n - p_n)}{\sqrt{p_n(1-p_n)}}\right\}\right] \\ \to\; \exp\left\{-\frac{t_1^2}{2} - \frac{1}{2}\,(t_2, t_3)\,\mathbb{Q}^{-1}\begin{pmatrix}t_2 \\ t_3\end{pmatrix} - \frac{t_4^2}{2}\right\}\left(1 + o_{\mathbb{P}}(1)\right)\end{aligned}$$

as $n \to \infty$, where $\mathbb{Q}^{-1}$ is that in (2.6), $\gamma_n^2 = \mathrm{Var}(Z_1)$ and $i^2 = -1$.
It follows that, with this result that

$$\frac{\sqrt{n}}{\gamma_n\,\sigma_n}\left(\widehat{\Pi}_{\rho,n} - \Pi_\rho\right) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(0,1)\,, \qquad \text{as} \quad n \to \infty\;,$$

where

$$\begin{aligned}\sigma_n^2 \;:=\; & \frac{1}{\rho^2} + \frac{\theta_1^2}{\gamma_n^2}\,p_n(1-p_n) + \frac{2\,(1+\xi)\,\theta_2^2\,\beta_n^2}{p_n\gamma_n^2} \\ & + \frac{(1+\xi)^2\,\theta_3^2}{p_n\gamma_n^2} - 2\,\frac{(1+\xi)\,\beta_n\theta_2\,\theta_3}{p_n\gamma_n^2}\;.\end{aligned}$$

This complete the proof of Theorem (2.1). □

ACKNOWLEDGMENTS

This work has been supported by the National Agency of the University Development Research of Algeria (ANDRU) PNR project, code: 08/E09/5100. We are grateful to an anonymous referee for constructive criticism and a number of queries that helped us to produce a substantial revision of the paper. I also thank the Professor Necir who encouraged me to address this work.

REFERENCES

[1] Beirlant, J.; Matthys, G. and Dierckx, G. (2001). Heavy-tailed distributions and rating, *Astin Bull.*, **31**, 37–58.

[2] Bingham, N.; Goldie, C. and Teugels, J. (1987). Regular Variation, no. **27** in *Encyclopedia of Mathematics and its Applications*, Cambridge University Press.

[3] Buhlmann, H. (1980). An economic premium principle, *Astin Bull.*, **11**, 52–60.

[4] Cebrian, A.C.; Denuit, M. and Lambert, P. (2003). Generalized Pareto Fitto the Society of Actuaries. Large Claims Database, *NAAJ.*, **7**, 18–36.

[5] Centeno, M.L. and Andrade e Silva, J. (2005). Applying the proportional hazard premium calculation principle, *Astin Bull.*, **35**, 409–425.

[6] Embrechts, P.; Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Springer.

[7] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* 2, 2nd ed., Wiley, New York. MR 42:5292.

[8] Furman, E. and Zitikis, R. (2008). Weighted risk capital allocations, *Insurance Math. Econom.*, **43**, 263–269.

[9] Johansson, J. (2003) Estimating the mean of heavy-tailed distributions, *Extremes*, **6**, 91–131.

[10] Jones, B.L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities, *North American Actuarial Journal*, **7**, 44–54.

[11] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, Springer.

[12] Necir, A. and Meraghni, D. (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts, *Insurance Math. Econom.*, **45**, 49–58.

[13] Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag.

[14] Rolski, T.; Schimidli, H.; Schmidt, V. and Teugels, J.L. (1999). *Stochastic Processes for Insurance and Finance*, John Wiley & Sons, Chichester.

[15] Romano, J.P. and Wolf, M. (1999). Subsampling inference for the mean in the heavy-tailed case, *Metrika*, **50**, 55–69.

[16] Seneta, E. (1976). Regularly Varying Functions, *Lecture Notes in Mathematics*, no. **508**, Springer-Verlag.

[17] Smith, R. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, **15**, 1174–1207.

[18] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

[19] Wang, S.S. (1996). Premium calculation by transforming the layer premium density, *ASTIN Bull.*, **26**, 71–92.

[20] Yaari, M.E. (1987). The dual theory of choice under risk, *Econometrica*, **55**, 95–115.

REVSTAT – Statistical Journal

Background

Statistical Institute of 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.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract / indexed in

REVSTAT is expected to be abstracted / indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: 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.

Submitted 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 not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, 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), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's 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: liliana.martins@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.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright