ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
80 Anos years 1935 2015

REVSTAT

Statistical Journal

Volume 13, No.3
November 2015

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: +351 218 454 084
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

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 13, Number 3, November 2015, 193–206

FORECASTING MORTALITY RATE BY SINGULAR SPECTRUM ANALYSIS

Authors: Rahim Mahmoudvand
– Statistical Research and Training Center (SRTC), Tehran, Iran
r.mahmoudvand@basu.ac.ir

Fatemeh Alehosseini
– Department of Statistics, Shahid Beheshti University, Iran
f_alehosseini@yahoo.com

Paulo Canas Rodrigues
– Department of Statistics, Federal University of Bahia, Salvador, BA, Brazil
and
CAST — Center for Applied Statistics and Data Analytics, University of Tampere, Finland
paulocanas@gmail.com

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

Abstract:

- Singular spectrum analysis (SSA) is a relatively new and powerful non-parametric time series analysis technique that has demonstrated its capability in forecasting different time series in various disciplines. In this paper, we study the feasibility of using the SSA to perform mortality forecasts. Comparisons are made with the Hyndman–Ullah model, which is a new powerful tool in the field of mortality forecasting, and will be considered as a benchmark to evaluate the performance of the SSA for mortality forecasting. We use both SSA and Hyndman–Ullah models to obtain 10 forecasts for the period 2000–2009 in nine European countries including Belgium, Denmark, Finland, France, Italy, The Netherlands, Norway, Sweden and Switzerland. Computational results show a superior accuracy of the SSA forecasting algorithms, when compared with the Hyndman–Ullah approach.

Key-Words:

- *mortality rate; Singular Spectrum Analysis; Hyndman–Ullah model.*

AMS Subject Classification:

- 37M10, 15A18, 62M15.

1. INTRODUCTION

With the continuing increase in life expectancy, the mortality forecasting plays a major role to advice government policy and planning, and in decision making for pension and insurance industries. Lee and Carter (1992) proposed a new method which uses singular value decomposition to represent the logs of mortality rate in terms of two age-dependent factors and a single time-dependent factor. The time-dependent factor can be extracted and modelled using conventional time series methods so that forecasts could be made. The popular method of Lee and Carter (1992) to model and forecast mortality rate has undergone various extensions and modifications. For a review and recent developments, see Hyndman and Ullah (2007), Hyndman *et al.* (2011) and references therein. These methods exhibited a good performance of mortality rate forecasts. However producing more accurate forecasts can help, both pension and insurance companies and governments, to make better decisions.

Singular Spectrum Analysis (SSA) is a relatively new non-parametric approach for analysing time series data which incorporates elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing (Golyandina *et al.*, 2001). SSA has the ability to decompose the original time series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structureless noise. The literature review on SSA shows that there are more than hundred papers on the application of SSA in the different areas and, in the majority of them, superiority of SSA compared to other time series analysis techniques has been demonstrated (e.g. Hassani *et al.*, 2009; Hassani and Thomakos, 2010, and references therein). Most recent developments in the theory and methodology of SSA can be found in Zhigljavsky (2010) and Golyandina and Zhigljavsky (2013).

Mahmoudvand *et al.* (2013) compared the ability of SSA with the Hyndman–Ullah model for mortality forecast in France. In this paper we extend that study to nine European countries (Belgium, Denmark, Finland, France, Italy, The Netherlands, Norway, Sweden and Switzerland); consider two forecasting algorithms for SSA: Recurrent SSA (RSSA, Danilov, 1997a, b) and Vector SSA (VSSA, Nekrutkin, 1999); and consider the time series until 2009, in a new approach.

Since the proposal of Hyndman and Ullah (2007) can be seen as a benchmark because it achieves more accurate mortality forecasts than many other approaches, it will be used to compare with SSA forecasting results and, therefore, to evaluate SSA as a plausible alternative for mortality forecasting.

The rest of the paper is structured as follows: in Section 2 we give a brief description of Hyndman and Ullah (2007) model, and in Section 3 present the generic SSA methodology. The application is presented Section 4 and Section 5 gives some concluding remarks.

2. HYNDMAN–ULLAH APPROACH

The Hyndman–Ullah approach can be expressed using the equation (Hyndman and Ullah, 2007)

$$\log m_t(x) \;=\; a(x) + \sum_{j=1}^{K} k_{t,j}\, b_j(x) + e_t(x) + \sigma_t(x)\, \epsilon_t(x)\;, \tag{2.1}$$

where $m_t(x)$ denotes the mortality rate for age x at time t, $a(x)$ is the average pattern of mortality by age across years, $b_j(x)$ is a basis function and $k_{t,j}$ is a time series coefficient. The error term $\sigma_t(x)\,\epsilon_t(x)$ accounts for observational error that varies with age; i.e., it is the difference between the observed rates and the spline curves. The error term $e_t(x)$ is modelling error, i.e. the difference between the spline curves and the fitted curves from the model. By comparison, the Lee–Carter model (Lee and Carter, 1992)

$$\log m_t(x) \;=\; a(x) + k_t\, b(x) + \epsilon_t(x)\;, \tag{2.2}$$

has one set of $(k_t, b(x))$, while the Hyndman–Ullah model includes more than one set of components. This extension presented by Hyndman and Ullah (2007) gives more flexibility to the model because the additional components capture non-random patterns, which are not explained by the first principal component. Other extensions of the Lee–Carter model are discussed in Booth *et al.* (2006) and Shang *et al.* (2011).

3. SINGULAR SPECTRUM ANALYSIS

The basic SSA method consists of three complementary stages: decomposition, reconstruction and forecasting. In the first stage the time series is decomposed, in the second stage the noise free time series is reconstructed and in the third stage the reconstructed time series is used for forecasting new data points. A short description of the SSA technique is given below. More information can be found in Golyandina *et al.* (2001), Hassani (2007) and Golyandina and Zhigljavsky (2013).

3.1. Basic SSA

First Stage: Decomposition

1st step: Embedding. Let $x_1, ..., x_N$ be a time series of length N. Considering a window length L the result of this step is a $L \times K$ matrix $\mathbf{X} = [X_1 : ... : X_K]$, where $K = N - L + 1$ and $X_i = (x_i, ..., x_{i+L-1})^T,\ 1 \leq i \leq K$.

2nd step: Singular Value Decomposition (SVD). In this step, matrix $\mathbf{X}$ will be decomposed using SVD as $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d$, where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i{}^T$ and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ with $\lambda_1, ..., \lambda_L$, the eigenvalues of $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ and $U_1, ..., U_L$, the corresponding eigenvectors.

Second Stage: Reconstruction

3rd step: Grouping. The grouping step corresponds to splitting the elementary matrices into m disjunct subsets $I_1, ..., I_m$, and summing the matrices within each group. In our application we have $m = 2$, i.e. only two groups. $I_1 = \{1, ..., r\}$ and $I_2 = \{r + 1, ..., L\}$ are related to the single and noise components, respectively.

4th step: Diagonal averaging. The purpose of diagonal averaging is to transform each matrix $\mathbf{X}_{I_j}$ into a new series of length N. Using diagonal averaging we have that $\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \cdots + \widetilde{\mathbf{X}}_{I_m}$, where $\widetilde{\mathbf{X}}_{I_j}$ is the hankelized form of $\mathbf{X}_{I_j}$, $j = 1, ..., m$. Considering $\tilde{x}_{m,n}^{(I_j)}$ the $(m, n)^{th}$ entry of the estimated matrix $\widetilde{\mathbf{X}}_{I_j}$ and denoting by $\{\tilde{y}_{j_1}, ..., \tilde{y}_{j_T}\}$ the reconstructed components in the matrix $\widetilde{\mathbf{X}}_{I_j}$, $j = 1, ..., m$, applying diagonal averaging follows that

$$\tilde{y}_{j_l} = \begin{cases} \dfrac{1}{s-1} \displaystyle\sum_{n=1}^{s-1} \tilde{x}_{n,s-n}^{(I_j)} & 2 \leq s \leq L-1\ , \\ \dfrac{1}{L} \displaystyle\sum_{n=1}^{L} \tilde{x}_{n,s-n}^{(I_j)} & L \leq s \leq K+1\ , \\ \dfrac{1}{K+L-s+1} \displaystyle\sum_{n=n-K}^{L} \tilde{x}_{n,s-n}^{(I_j)} & K+2 \leq s \leq K+L\ . \end{cases}$$

Third Stage. Forecasting

The basic requirement to make SSA forecasting is that the time series satisfies a linear recurrent formula (LRF). A time series $Y_T = (y_1, ..., y_T)$ satisfies

LRF of order d if:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_d y_{t-d} , \qquad t = d+1, ..., T . \tag{3.1}$$

Although there are several versions of univariate SSA forecasting algorithms we consider here two of the mostly widely used: Recurrent SSA (RSSA, Danilov, 1997a, b) and Vector SSA (VSSA, Nekrutkin, 1999). In what follows, we give a brief description of these algorithms. Further details can be found in Golyandina *et al.* (2001).

Let us assume that U_j^{∇} is the vector of the first $L-1$ components of the eigenvector U_j and π_j is the last component of U_j $(j = 1, ..., r)$. Denoting $\upsilon^2 = \sum_{j=1}^{r} \pi_j^2$ we define the coefficient vector $\mathfrak{R}$ as:

$$\mathfrak{R} = \frac{1}{1-\upsilon^2} \sum_{j=1}^{r} \pi_j U_j^{\nabla} .$$

Recurrent SSA

Considering the above notation, the RSSA forecasts $(\hat{y}_{T+1}, ..., \hat{y}_{T+M})$ can be obtained by

$$\hat{y}_i = \begin{cases} \tilde{y}_i , & i = 1, ..., T , \\ \mathfrak{R}^T Z_i , & i = T+1, ..., T+M , \end{cases} \tag{3.2}$$

where, $Z_i = [\hat{y}_{i-L+1}, ..., \hat{y}_{i-1}]^T$ and $\tilde{y}_1, ..., \tilde{y}_T$, are the values for the reconstructed time series and can be obtained from 4th step in above.

Vector SSA

Define linear operator:

$$\mathcal{P}^{(\upsilon)} Y = \begin{pmatrix} \mathbf{\Pi} Y_{\triangle} \\ \mathfrak{R}^T Y_{\triangle} \end{pmatrix} , \qquad Y \in \text{span}\{U_1, ..., U_r\} , \tag{3.3}$$

where $\mathbf{\Pi} = \mathbf{U}^{\nabla} \mathbf{U}^{\nabla T} + (1-\upsilon^2) \mathfrak{R} \mathfrak{R}^T$ and $Y_{\triangle}$ denotes the last $L-1$ elements of Y. Suppose the vector Z_j is defined as follows

$$Z_j = \begin{cases} \widetilde{X}_j & \text{for } j = 1, ..., K , \\ \mathcal{P}^{(\upsilon)} Z_{j-1} & \text{for } j = K+1, ..., K+M+L-1 , \end{cases} \tag{3.4}$$

where $\widetilde{X}_j$ are the j^{th} reconstructed columns of the trajectory matrix of the time series after grouping and discarding noise components. Now, by constructing the matrix $\mathbf{Z} = [Z_1, ..., Z_{K+M+L-1}]$ and performing diagonal averaging, we obtain a new time series $\hat{y}_1, ..., \hat{y}_{T+M+L-1}$, where $\hat{y}_{T+1}, ..., \hat{y}_{T+M}$ form the M terms of the VSSA forecast.

3.2. Forecast accuracy

To evaluate the accuracy and reliability of forecasts in time series, one can use a suitable combination of the following three approaches: (i) construction of confidence intervals; (ii) assessment of retrospective forecasts; and (iii) checking the stability of forecasts. Although the three represent important approaches, in the present paper we will be interested only in (ii) assessment of retrospective forecasts. Further information about approaches (i) and (iii) can be found in Golyandina *et al.* (2001) and Pepelyshev *et al.* (2010), respectively.

Retrospective forecasts are usually performed by truncating the time series and by obtaining forecasts for points temporarily removed. These forecasts can then be compared with the observed values of the time series to asses their quality and reliability. Let $e_{T,h}(x) = y_{T+h}(x) - \hat{y}_{T,h}(x)$ denote the forecast error, where $\hat{y}_{T,h}(x)$ are the forecasts for $y_{T+h}(x)$ using RSSA or VSSA ($h = 1, ..., M$). Then, a measure of accuracy such as the Integrated Squared Error of forecast can be written as

$$ISE_{T,h} = \sum_x e^2_{T,h}(x) \ . \tag{3.5}$$

3.3. SSA parameter selection

The SSA calibration depends upon two basic, but very important, parameters: the window length L, and the number of eigentriples used for reconstruction r. The choice of improper values for the parameters L or r yield incomplete reconstruction and the forecasting results might be misleading. Despite the importance in choosing proper values for these parameters, no theoretical solution has been proposed to solve this problem. Some of the techniques to choose the appropriate value of L can be found in Golyandina (2010), Hassani *et al.* (2011), Mahmoudvand and Zokaei (2012) and Mahmoudvand *et al.* (2013). An overall agreeable suggestion to choose the window length is to have it close to the middle of the series and proportional to the number of observations per period (e.g. to 12 for monthly time series, to four for quarterly time series, etc.). However, this choice does not guarantee the best predictions (e.g. Mahmoudvand, et al, 2013).

For better results, the parameter choice should be made accordingly to available data and intended analysis.

In practice it is relatively rare that the number of singular values r, needed to be selected to reconstruct noise free series from a noisy time series, is known a priori. Among several ways to determine r described in the literature, the easiest way is done by checking breaks in the eigenvalues spectra. As a rule of thumb, a pure noise series produces a slowly decreasing sequences of singular values. Another useful insight is provided by considering separability between signal and noise components, which is a fundamental concept in studying SSA properties, by using w-correlations (Golyandina *et al.*, 2001) between two vectors $Y^{(1)} = [y_1^{(1)}, ..., y_T^{(1)}]^T$ and $Y^{(2)} = [y_1^{(2)}, ..., y_T^{(2)}]^T$:

$$\rho_w = \frac{\sum_{j=1}^{T} w_j^{L,T} y_j^{(1)} y_j^{(2)}}{\sqrt{\sum_{j=1}^{T} w_j^{L,T} \left(y_j^{(1)}\right)^2 \times \sum_{j=1}^{T} w_j^{L,T} \left(y_j^{(2)}\right)^2}} , \tag{3.6}$$

where, $w_j^{L,T} = \min\{j, L, T-j+1\}$ and $2 \leq L \leq T-1$. According to this measure, two series are separable if the absolute value of their w-correlation is small. Therefore, we determine r in such a way that the reconstructed series and residual have a small w-correlation. Another way to determine r is by examining the forecast accuracy, i.e. r is determined in such a way that the minimum error in forecasting will be obtained. Considering L fixed, the choice of r can be done as

$$r = \underset{r<L<T-1}{\operatorname{argmin}} \; ISE_{T,h}(x) \; . \tag{3.7}$$

In this study, we considered $L = 10$ and employed equation (3.7) to determine the number of singular values used for reconstruction, r.

4. RESULTS

Following a preliminary study by Mahmoudvand *et al.* (2013), we intend to demonstrate the feasibility of SSA for forecasting mortality rates using age-specific mortality rates from nine European countries: Belgium, Denmark, Finland, France, Italy, The Netherlands, Norway, Sweden and Switzerland. We have $y_t(x) = \log(m_t(x))$ where $m_t(x)$ denotes the mortality rate for age x in year t.

4.1. Empirical results: The case of nine European countries

Annual mortality rates of nine European countries for single years of age were obtained from the Human Mortality Database (http://www.mortality.org/). These mortality rates are the ratios of death counts to population exposure in the relevant interval of age and time. Figure 1 shows the typical patterns of log mortality rates for several ages and years in the considered countries.

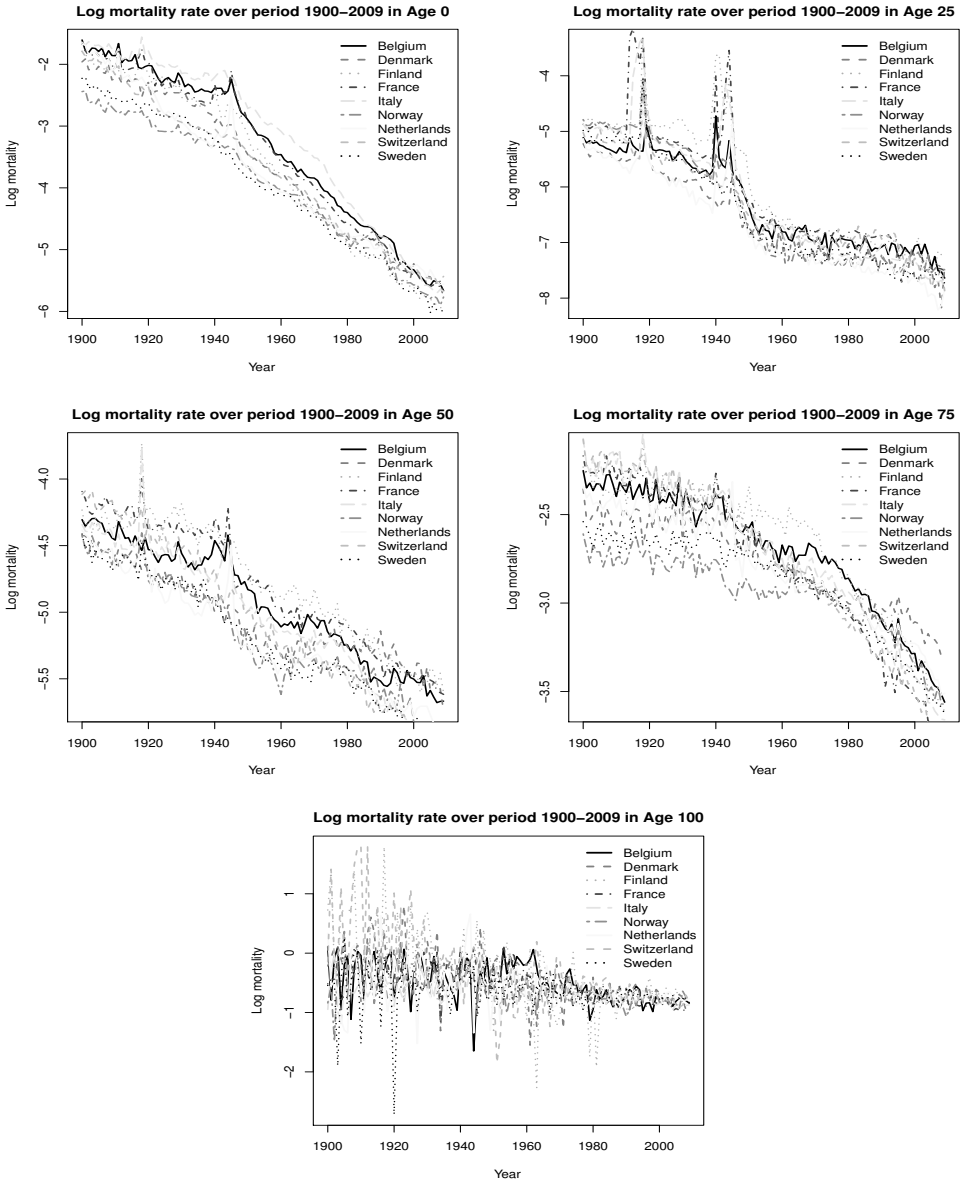

Figure 1: Changes in the total log mortality rates with respect to both age and year, over the period 1900–2009.

The plots depicted in Figure 1 show that, from 1900 to 2009, there was a general pattern of decline in mortality rates for all ages and all countries, as reported by Mahmoudvand *et al.* (2013) for the case of France. By analysing these plots, it can be seem that the decline for infant mortality is steeper than for adult mortality. The effects of the World War I (1914–1918) and World War II (1939–1945) are clearly visible in the top right plot of Figure 1, for the age of 25, being more dramatic for France and Italy, as expected. For the other ages the same effect is also visible but it not as extreme. Since the number of people with 100 years old is small, the bottom plot of Figure 1 shows a less clear pattern but a decrease is visible in terms of mortality rate and variability, with the time, for all countries.

Comparison

The results of our proposal were compared with the results obtained from the method of Hyndman and Ullah (2007). The comparison was made by considering the European mortality data between 1900 and 1999, with forecast for the years 2000–2009. Calculations for the Hyndman–Ullah model were made with the R package "demography" and for SSA we developed our own R code (available upon request). The forecasts were compared with the observed values of the time series, using the integrated squared error (3.5), where the squared errors that integrated by age, on the log scale.

Forecasts of log mortality rate for the period 2000–2009, using the time series 1900–1999, for all ages between 0 and 100 years, for both SSA and Hyndman–Ullah approaches, were computed and compared. This is, for each age, from 0 to 100 years old, and for each of the nine countries, the time series between 1900 and 1999 is used to forecast the next 10 values between 2000 and 2009, which result in the ten ISE of forecasting reported in Table 1. According to the ISE values, results in the mortality forecasts by RSSA and VSSA are significantly better than the results for the Hyndman–Ullah method. Ratios of ISE in the second and third rows for each country of Table 1 show that SSA provides more than 90% improvement in log mortality forecast for some country–year combinations. This confirms the superiority of SSA over the Hyndman–Ullah method. Moreover, the VSSA forecasting procedure is slightly better than the RSSA forecasting procedure, particularly for the long term forecasts. By comparing the results in Table 1 and the plots in Figure 1, it is clear that, because of its construction, HU procedure produces good results when the time series are smoother. However both RSSA and VSSA produce better results when forecasting the most of the mortality rates in these time series.

Table 1: ISE of forecasts for the considered countries.

Country	Model	Year									
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Belgium	HU[1]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	RSSA	0.65	0.65	0.38	0.60	0.51	0.82	1.44	0.49	0.38	0.99
	VSSA	0.90	0.59	0.37	0.79	0.39	0.53	1.00	0.38	0.52	1.08
Denmark	HU	4.70	1.68	3.31	2.92	2.29	5.48	3.57	7.80	6.72	6.96
	$\frac{\text{RSSA}}{\text{HU}}$	0.28	0.54	0.28	0.30	0.38	0.24	0.45	0.33	3.20	0.30
	$\frac{\text{VSSA}}{\text{HU}}$	0.30	0.43	0.39	0.24	0.45	0.23	0.39	0.28	3.33	0.26
Finland	HU	4.15	4.95	5.19	4.55	6.08	5.60	5.92	6.17	5.22	8.27
	$\frac{\text{RSSA}}{\text{HU}}$	0.21	0.11	0.20	0.20	0.44	0.89	0.15	0.51	0.41	0.32
	$\frac{\text{VSSA}}{\text{HU}}$	0.25	0.17	0.21	0.21	0.28	0.70	0.19	0.38	0.43	0.30
France	HU	0.50	0.74	0.61	0.93	1.83	1.61	2.13	2.44	2.50	2.15
	$\frac{\text{RSSA}}{\text{HU}}$	0.17	0.21	0.22	0.27	0.17	0.19	0.34	0.34	0.27	0.35
	$\frac{\text{VSSA}}{\text{HU}}$	0.26	0.21	0.23	0.40	0.14	0.16	0.28	0.26	0.21	0.33
Italy	HU	0.70	1.09	1.68	1.35	2.48	2.49	3.37	2.93	2.84	3.39
	$\frac{\text{RSSA}}{\text{HU}}$	0.16	0.10	0.13	0.14	0.21	0.17	0.27	0.29	0.19	0.27
	$\frac{\text{VSSA}}{\text{HU}}$	0.16	0.13	0.12	0.11	0.14	0.18	0.17	0.27	0.15	0.21
Netherlands	HU	1.00	0.73	1.05	1.10	1.83	2.44	2.78	3.41	4.39	3.94
	$\frac{\text{RSSA}}{\text{HU}}$	0.24	0.36	0.23	0.23	0.15	0.26	0.22	0.27	0.30	0.26
	$\frac{\text{VSSA}}{\text{HU}}$	0.23	0.30	0.10	0.32	0.14	0.23	0.19	0.23	0.29	0.22
Norway	HU	3.23	5.76	2.71	2.52	4.86	5.07	10.24	12.55	6.11	8.51
	$\frac{\text{RSSA}}{\text{HU}}$	0.21	0.09	0.58	0.35	0.16	0.21	0.43	0.64	0.20	0.32
	$\frac{\text{VSSA}}{\text{HU}}$	0.27	0.08	0.35	0.32	0.17	0.20	0.34	0.57	0.24	0.37
Sweden	HU	4.52	3.17	4.79	2.82	3.38	4.17	3.85	4.37	8.20	6.07
	$\frac{\text{RSSA}}{\text{HU}}$	0.30	0.17	0.16	0.36	0.23	0.28	0.32	0.27	0.43	0.40
	$\frac{\text{VSSA}}{\text{HU}}$	0.29	0.17	0.26	0.27	0.18	0.17	0.19	0.13	0.43	0.25
Switzerland	HU	1.54	6.46	4.94	3.33	2.72	4.05	5.01	4.19	5.48	5.46
	$\frac{\text{RSSA}}{\text{HU}}$	0.23	0.43	0.52	0.32	0.45	0.32	0.37	0.32	0.29	0.40
	$\frac{\text{VSSA}}{\text{HU}}$	0.31	0.51	0.46	0.31	0.40	0.27	0.40	0.27	0.29	0.41

[1]Due to a small amount of missing values the HU values were not possible to obtain. Data imputation techniques (e.g. Rodrigues and de Carvalho, 2013) can be used to fill in the missing values.

Figure 2 shows the results of a sensitivity analysis to choose the model parameters, window length L for SSA and K for the Hyndman–Ullah model, where the mean ISE (MISE) of forecasts over the period 2000–2009 is presented. Recall that MISE is provided by:

$$MISE = \frac{1}{M}\sum_{h=1}^{M} ISE_{T,h} \tag{4.1}$$

where M denotes the number of forecasts.

Although the model parameters, $L$for SSA and K for HU, are not directly comparable, it can be seen in the plots of Figure 2 that, for most of the cases, the

results of SSA, both RSSA and VSSA, are better than those of Hyndman–Ullah, in terms of MISE. This confirms the ability of SSA for mortality forecasting, being the RSSA slightly better than the VSSA, as visible in Table 1.

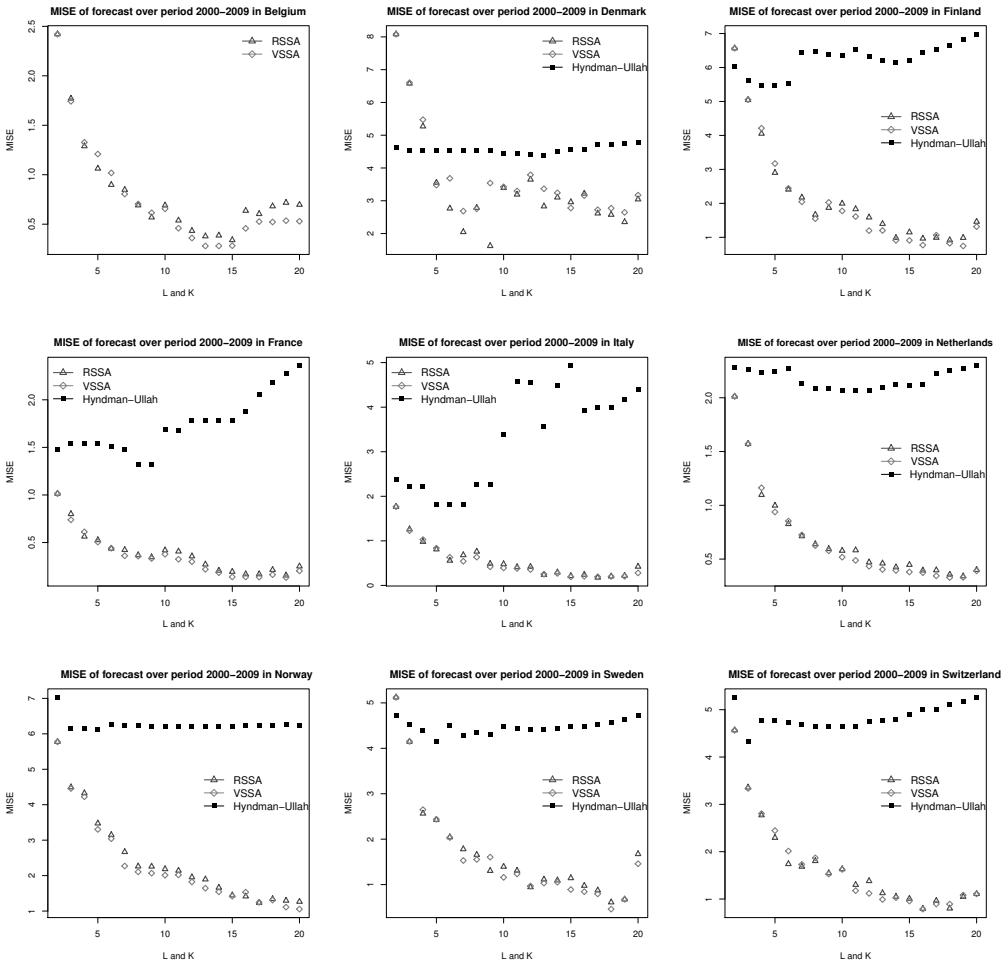

Figure 2: Mean integrated squared error (MISE) of total log mortality rates forecast by RSSA and Hyndman–Ullah model over the period 2000–2009 for different SSA and HU parameters.

5. CONCLUSION

In this paper, the usefulness and ability of Singular Spectrum Analysis (SSA) to forecast mortality rates was studied. The results of SSA based forecasting procedures were compared with those of Hyndman and Ullah method, which can be seen as a benchmark for mortality forecasting. As in the preliminary study

presented by Mahmoudvand *et al.* (2013) in this field of research, we can also conclude that the forecasting accuracy of SSA is higher than the forecasting accuracy of the Hyndman and Ullah method, for most of the cases. Within the two SSA based approaches, the RSSA shows slightly better results than the VSSA.

It should be noticed that our proposal does not take into consideration the correlations among ages, which certainly can add useful informations to the analyses and improve the forecast accuracy. Multivariate versions of SSA would be a valid alternative to deal with such correlations and should be considered in further studies. Other alternatives for further improvement of mortality forecasting might be achieved when considering other SSA based forecasting algorithms.

REFERENCES

[1] Booth, H.; Hyndman, R.; Tickle, L. and Jong, P.D. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions, *Demographic Research*, **15**, 289–310.

[2] Caselli, G.; Vallin, J.; Vaupel, J.W. and Yashin, A.I. (1987). Age-specific mortality trends in France and Italy since 1900: period and cohort effects, *European Journal of Population*, **3**, 33–60.

[3] Danilov, D. (1997a). Principal components in time series forecast, *Journal of Computational and Graphical Statistics*, **6**, 112–121.

[4] Danilov, D. (1997b). The Caterpillar method for time series forecasting. In "Principal Components of Time Series: the Caterpillar method" (D. Danilov and A. Zhigljavsky, Eds.), pages 73–104, University of St. Petersburg, St. Petersburg. (In Russian).

[5] Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison, *Journal of Data Science*, **5**, 239–257.

[6] Hassani, H.; Heravi, S. and Zhigljavsky, A. (2009). Forecasting European Industrial Production with Singular Spectrum Analysis, *International Journal of Forecasting*, **25**, 103–118.

[7] Hassani, H.; Mahmoudvand, R. and Zokaei, M. (2011). Separability and Window Length in Singular Spectrum Analysis, *Comptes Rendus Mathematique*, **349**, 987–990.

[8] Hassani, H. and Thomakos, D. (2010). A Review on Singular Spectrum Analysis for Economic and Financial Time Series, *Statistics and Its Interface*, **3**, 377–397.

[9] Hyndman, R.J.; Booth, H. and Yasmeen, F. (2011). *Coherent mortality forecasting: the product-ratio method with functional time series models*, Working paper 01/11, Monash University, Australia.

[10] Hyndman, R.J. and Ullah, S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach, *Computational Statistics and Data Analysis*, **51**, 4942–4956.

[11] Golyandina, N.; Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York – London.

[12] Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series*, Springer, New York – London.

[13] Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting U.S. mortality (discussion 671–675), *Journal of the American Statistical Association*, **87**, 659–671.

[14] Mahmoudvand, R.; Alehosseini, F. and Zokaei, M. (2013). Feasibility of Singular Spectrum Analysis in the Field of Forecasting Mortality Rate, *Journal of Data Science*, **11**, 851–866.

[15] Mahmoudvand, R.; Najari, N. and Zokaei, M. (2013). On the Parameters for Reconstruction and Forecasting in the Singular Spectrum Analysis, *Communication in Statistics: Simulations and Computations*, **42**, 860–870.

[16] Mahmoudvand, R. and Zokaei, M. (2012). On the singular values of the Hankel matrix with application in Singular Spectrum Analysis, *Chilean Journal of Statistics*, **3**, 43–56.

[17] Nekrutkin, V.V. (1999). Approximation spaces and continuation of time series. In "Statistical Models with Applications in Econometrics and Neibouring Fields" (S.M. Ermakovand and Yu.N. Kashtanov, Eds.), pages 3–32, University of St. Petersburg, St. Petersburg. (In Russian).

[18] Pepelyshev, A. and Zhigljavsky, A. (2010). Assessing the stability of long-horizon SSA forecasting, *Statistics and Its Interface*, **3**, 321–327.

[19] Rodrigues, P.C. and de Carvalho, M. (2013). Spectral modeling of time series with missing data, *Applied Mathematical Modelling*, **37**, 4676–4684.

[20] Shang, H.L.; Booth, H. and Hyndman, R. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods, *Demographic Research*, **25**, 173–214.

[21] Zhigljavsky, A. (2010). Singular Spectrum Analysis for time series: Introduction to this special issue, *Statistics and Its Interface*, **3**, 255–258.

REVSTAT – Statistical Journal
Volume 13, Number 3, November 2015, 207–226

THE TAYLOR PROPERTY IN BILINEAR MODELS

Authors: E. Gonçalves
– CMUC, Department of Mathematics, University of Coimbra, Portugal
esmerald@mat.uc.pt

C.M. Martins
– Department of Mathematics, University of Coimbra, Portugal
cmtm@mat.uc.pt

N. Mendes-Lopes
– CMUC, Department of Mathematics, University of Coimbra, Portugal
nazare@mat.uc.pt

Received: October 2013 Revised: April 2014 Accepted: May 2014

Abstract:

- The aim of this paper is to discuss the presence of the Taylor property in the class of simple bilinear models. Considering strictly and weakly stationary models, we deduce autocorrelations of the process and of its square and analyze the presence of the Taylor property in non-negative bilinear models considering several error process distributions, which are chosen according to the kurtosis value. For each one of these error process distributions, the class of parameterizations for the corresponding bilinear model satisfying Taylor property is obtained. The analysis of the relationship between the Taylor property and leptokurtosis in these bilinear processes allows to conclude that this property is a consequence of heavy tailed model distributions.

  With the goal of extending this research to real valued bilinear models, a simulation study is developed in a class of such models with symmetrical innovations.

Key-Words:

- *bilinear models; nonlinear time series; stationarity; Taylor property.*

AMS Subject Classification:

- 62M10.

1. INTRODUCTION

The search for non-trivial empirical regularities in time series, usually called stylized facts, has been the subject of several studies in order to identify classes of time series models that conveniently capture such empirical properties. A stylized fact detected by Taylor ([9]) when he analyzed 40 returns series is known as the Taylor effect. He observed that, for most of the returns series, denoted by X_t for instant t, the sample autocorrelations of the absolute returns, $\hat{\rho}_{|X|}(n) = \widehat{\text{corr}}(|X_t|, |X_{t-n}|)$, were larger than those of the squared returns, $\hat{\rho}_{X^2}(n) = \widehat{\text{corr}}(X_t^2, X_{t-n}^2)$, for $n \in \{1, ..., 30\}$. More recently, Gonçalves *et al.* ([2]) also recorded Taylor effect in the physical time series of plage region areas describing solar activity.

We point out that there is still little research on the theoretical counterpart on this empirical property due to the difficulty of handling the true autocorrelations of time series models. For example, this theoretical counterpart was studied by He and Teräsvirta ([5]) on conditionally Gaussian absolute value generalized ARCH (AVGARCH) models, assuring its presence for some of these models. More precisely, they called the theoretical relation $\rho_{|X|}(n) > \rho_{X^2}(n)$, $n \geq 1$, the Taylor property and concentrated their study on the autocorrelation of lag 1. Analogously, Gonçalves, Leite and Mendes-Lopes ([1]) studied the presence of the Taylor property in TARCH models, concluding that this property is satisfied when $n = 1$, for some first-order models. Generalizing these papers, Haas ([4]) proposed a methodology for identifying the Taylor property in AVGARCH$(1, 1)$ models at all lags.

The research of this property within heteroskedastic models is mainly related to the empirical facts observed and the good fit of those models to financial time series. The established results have shown a strong connection between the Taylor property and the kurtosis of the process; in fact, its presence seems to be more related to the leptokurtic character of those models than to its conditional heteroskedascity. This interpretation is consistent with the leptokurtic nature of the real series presenting such stylized fact. Thus, we believe that it is important to assess the presence of the Taylor property in other classes of processes with relevance in time series analysis as it is the case of bilinear ones, which have also been proven to be suitable in financial and physical time series modeling ([3], p. 181).

In this paper we consider the simple bilinear diagonal model

$$X_t = \beta X_{t-k}\varepsilon_{t-k} + \varepsilon_t \ , \qquad k > 0 \ , \tag{1.1}$$

where β is a real parameter and $(\varepsilon_t, t \in \mathbb{Z})$ a sequence of i.i.d. random variables, designated here by error process.

We state sufficient conditions for the strict and weak stationarity of the processes $X = (X_t, t \in \mathbb{Z})$ and $X^2 = (X_t^2, t \in \mathbb{Z})$, and we derive expressions for the moments of X up to the 4th order. We also consider the study of the Taylor property assuming that $\beta > 0$ and that the error process is non-negative. In fact, there has been considerable interest in non-negative bilinear models. For instance, Pereira and Scotto ([7]) studied some properties of the simple first-order bilinear diagonal model ($k = 1$) driven by exponentially distributed innovations. Also Zhang and Tong ([10]) have examined some distributional properties of a simple first-order non-negative bilinear model considering for the error process the uniform distribution in $(0, 1)$.

The remainder of the paper is organized as follows. In Section 2 we establish sufficient conditions under which X and X^2 are strictly and weakly stationary. Moreover, the moments of X up to 4th order are evaluated and a working example on this matter is presented in appendix. In Section 3, the Taylor property in first-order bilinear diagonal models with non-negative error process is analyzed. This study is developed considering several distributions for the error process with significantly different kurtosis values. A simulation study evaluating the Taylor property in other real-valued simple bilinear models is presented in Section 4. Some concluding remarks and future developments are given in Section 5.

2. STATIONARITY AND MOMENTS OF X AND X^2

In this section we consider the simple bilinear model defined by (1.1) and we denote $\mu_i = E(\varepsilon_t^i)$, $i \in \mathbb{N}$.

Proposition 2.1. *Suppose that μ_4 and $E(\ln|\varepsilon_t|)$ exist. If $\beta^2 \mu_2 < 1$ then the process X is strictly and weakly stationary.*

Proof: The strict stationarity of the process X is achieved by proving that $X_t = Y_t$, a.s., with

$$Y_t = \varepsilon_t + \sum_{n=1}^{+\infty} T_n \ ,$$

where, for each $n \in \mathbb{N}$, $T_n = T_n(t)$ is given by

$$T_n = \beta^n \varepsilon_{t-nk} \prod_{j=1}^{n} \varepsilon_{t-jk} \ .$$

The proof of this result is similar to that of Theorem 1 in Quinn ([8]), as the condition $\beta^2 \mu_2 < 1$ implies Quinn's condition $\ln|\beta| + E(\ln|\varepsilon_t|) < 0$ by applying Jensen's inequality to the random variable ε_t^2.

To prove the weak stationarity, we now verify that $E(Y_t^2) < +\infty$. We have

$$
\begin{aligned}
E(Y_t^2) \;&=\; E\left[\left(\varepsilon_t + \sum_{i=1}^{+\infty} T_i\right)^2\right] \\
&\leq\; E(\varepsilon_t^2) + 2\sum_{i=1}^{\infty} E(|\varepsilon_t|\,|T_i|) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E(|T_i T_j|)\;.
\end{aligned}
\tag{2.1}
$$

Under the given conditions, each series in (2.1) is convergent. In fact, let us consider, for example, the series $\displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E(|T_i T_j|)$.

For each $i, j \in \mathbb{N}$, we have

$$
\begin{aligned}
E(|T_i T_j|) \;&\leq\; |\beta|^{i+j}\left[E\Big(\varepsilon_{t-ik}^4\,\varepsilon_{t-k}^2\,\varepsilon_{t-2k}^2\cdots\varepsilon_{t-(i-1)k}^2\Big)\right]^{1/2} \\
&\qquad \left[E\Big(\varepsilon_{t-jk}^4\,\varepsilon_{t-k}^2\,\varepsilon_{t-2k}^2\cdots\varepsilon_{t-(j-1)k}^2\Big)\right]^{1/2} \\
&=\; \mu_4\,\mu_2^{-1}\Big[\big(\beta^2\mu_2\big)^{1/2}\Big]^{i+j},
\end{aligned}
$$

by Schwarz's inequality and the independence of the r.v.'s ε_t, $t \in \mathbb{Z}$. As $\big(\beta^2\mu_2\big)^{1/2} < 1$, the series is convergent.

Taking into account the equality $X_t = Y_t$, a.s., and the strict stationarity of the process X, we conclude that $E(X_t^2)$ exists and that X is weakly stationary. □

Proposition 2.2. *Suppose that $E(\ln|\varepsilon_t|)$ and μ_8 exist. If $\beta^4\mu_4 < 1$ then the process X^2 is strictly and weakly stationary.*

Proof: The condition $\beta^4\mu_4 < 1$ implies $\beta^2\mu_2 < 1$, by Schwarz's inequality, which implies in turn the strict stationarity of X and, consequently, of X^2. The proof of the weak stationarity of X^2 is analogous to the previous one. We have

$$
\begin{aligned}
E(Y_t^4) \;\leq\; E(\varepsilon_t^4) &+ \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty} E(|T_i T_j T_p T_q|) + 4\sum_{i=1}^{\infty} E(|\varepsilon_t^3|\,|T_i|) \\
&+ 4\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{p=1}^{\infty} E(|\varepsilon_t|\,|T_i T_j T_p|) + 6\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E(\varepsilon_t^2\,|T_i T_j|)\;.
\end{aligned}
$$

Let us consider, for example, the series $\displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty} E(|T_i T_j T_p T_q|)$, which is a sum of series of the types

$$\textbf{(i)} \quad \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=p+1}^{\infty} E(|T_iT_jT_pT_q|)\,,$$

$$\textbf{(ii)} \quad \sum_{i=1}^{\infty}\sum_{p=1}^{\infty} E(T_i^2T_p^2)\,,$$

$$\textbf{(iii)} \quad \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\sum_{p=1}^{\infty} E(|T_iT_j|T_p^2)\,.$$

Concerning (i), as $j>i$ and $q>p$, we have

$$\begin{aligned} E(|T_iT_jT_pT_q|) &= E\Big[(|T_iT_j|)\,(|T_pT_q|)\Big] \\ &\le \left[E\Big(\varepsilon_{t-ik}^4\varepsilon_{t-l}^4\varepsilon_{t-l-k}^4\cdots\varepsilon_{t-l-(i-1)k}^4\varepsilon_{t-jk}^2\varepsilon_{t-l-ik}^2\cdots\varepsilon_{t-l-(j-1)k}^2\Big)\right]^{1/2} \\ &\quad \left[E\Big(\varepsilon_{t-pk}^4\varepsilon_{t-l}^4\varepsilon_{t-l-k}^4\cdots\varepsilon_{t-l-(p-1)k}^4\varepsilon_{t-qk}^2\varepsilon_{t-l-pk}^2\cdots\varepsilon_{t-l-(q-1)k}^2\Big)\right]^{1/2}, \end{aligned}$$

by Schwarz's inequality.

Taking into account the independence of the random variables ε_t, we have, for $i,j\in\mathbb{N},\ j>i$,

$$E\Big(\varepsilon_{t-ik}^4\,\varepsilon_{t-l}^4\,\varepsilon_{t-l-k}^4\cdots\varepsilon_{t-l-(i-1)k}^4\,\varepsilon_{t-jk}^2\,\varepsilon_{t-l-ik}^2\cdots\varepsilon_{t-l-(j-1)k}^2\Big) = \mu_4^{i+1}\mu_2^{j-i+1}\,.$$

Then

$$\begin{aligned} &\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=p+1}^{\infty} E(|T_iT_jT_pT_q|) \\ &\qquad \le \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=p+1}^{\infty} |\beta|^{i+j+p+q}\left(\mu_4^{i+p+2}\mu_2^{j-i+q-p+2}\right)^{1/2} \\ &\qquad = \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\sum_{p=1}^{\infty}\sum_{q=p+1}^{\infty} \mu_2\,\mu_4\Big[\left(\beta^4\mu_4\right)^{1/2}\Big]^{i+p}\Big[\left(\beta^2\mu_2\right)^{1/2}\Big]^{[(j+q)-(i+p)]}. \end{aligned}$$

As $(\beta^4\mu_4)^{1/2}<1$ and $(\beta^2\mu_2)^{1/2}<1$, the series in (i) is convergent. The convergence of the series (ii) and (iii) is proved in a similar way. Then we conclude that $E(X_t^4)<+\infty,\ t\in\mathbb{Z}$. As the process X^2 is strictly stationary and $E(X_t^4)$ exists, then it is weakly stationary. □

Let us now evaluate the moments up to the 4th order of the process X given by (1.1).

Proposition 2.3. *If $\beta^4\mu_4 < 1$ and μ_8 exists then the nth moment of X_t, $n \leq 4$, can be expressed as*

$$E(X_t^n) = \sum_{i=0}^{n} \binom{n}{i} \beta^{n-i}\, \mu_i\, E(X_t^{n-i}\varepsilon_t^{n-i}) \;, \tag{2.2}$$

where

$$E(X_t^n\varepsilon_t^n) = \frac{1}{1-\beta^n\mu_n} \sum_{i=1}^{n} \binom{n}{i} \beta^{n-i}\, \mu_{n+i}\, E(X_t^{n-i}\varepsilon_t^{n-i}) \;, \qquad n \leq 4 \;. \tag{2.3}$$

Proof: For $n \leq 4$, we have

$$\begin{aligned} E(X_t^n) &= \sum_{i=0}^{n} \binom{n}{i} \beta^{n-i}\, E\Big[\varepsilon_t^i\, (X_{t-k}\varepsilon_{t-k})^{n-i}\Big] \\ &= \sum_{i=0}^{n} \binom{n}{i} \beta^{n-i}\, \mu_i\, E(X_t^{n-i}\varepsilon_t^{n-i}) \;, \end{aligned}$$

since the process $(X_t\varepsilon_t, t \in \mathbb{Z})$ is strictly stationary due to the fact that $X_t\varepsilon_t$ is a measurable function of $\varepsilon_t, \varepsilon_{t-1}, \ldots$. Now we need to evaluate $E(X_t^n\varepsilon_t^n)$, $n \leq 4$.

$$\begin{aligned} E(X_t^n\varepsilon_t^n) &= \sum_{i=0}^{n} \binom{n}{i} \beta^{n-i}\, E\Big[\varepsilon_t^i\, (X_{t-k}\varepsilon_{t-k})^{n-i}\, \varepsilon_t^n\Big] \\ &= \sum_{i=0}^{n} \binom{n}{i} \beta^{n-i}\, E(\varepsilon_t^{n+i})\, E(X_t^{n-i}\varepsilon_t^{n-i}) \\ &= \beta^n\, \mu_n\, E(X_t^n\varepsilon_t^n) + \sum_{i=1}^{n} \binom{n}{i} \beta^{n-i}\, \mu_{n+i}\, E(X_t^{n-i}\varepsilon_t^{n-i}) \end{aligned}$$

and the result follows. □

It is easy to verify that $E(X_t\varepsilon_t) = \mu_2/(1-\beta\mu_1)$. The values $E(X_t^n\varepsilon_t^n)$, $n = 1, 2, 3$, are obtained recursively by using the previous equation; and finally, we achieve $E(X_t^n)$, $n \leq 4$. A working example to illustrate these evaluations is developed in appendix for a first order bilinear model with exponentially-distributed error process.

We note that $\beta^4\mu_4 < 1$ implies $|\beta^n\mu_n| < 1$, $n = 1, 2, 3$, by Schwarz's inequality.

3. THE TAYLOR PROPERTY IN FIRST-ORDER NON-NEGATIVE BILINEAR MODELS

3.1. Preliminary results

In this section we consider the first-order non-negative bilinear model

$$X_t = \beta X_{t-1}\varepsilon_{t-1} + \varepsilon_t \ , \qquad t \in \mathbb{Z} \ , \tag{3.1}$$

where $\beta > 0$ and $(\varepsilon_t, t \in \mathbb{Z})$ is a sequence of non-negative i.i.d. random variables.

We assume that $E(\ln \varepsilon_t)$ and μ_8 exist and that $\beta^4\mu_4 < 1$ in order to guarantee that both processes, X and X^2, are strictly and weakly stationary.

In this context, the Taylor property for $n = 1$ establishes that $\rho_X(1) > \rho_{X^2}(1)$, where $\rho_X(1)$ and $\rho_{X^2}(1)$ denote, respectively, the autocorrelations of lag 1 of the processes X and X^2. It is enough to evaluate $E(X_tX_{t-1})$ and $E(X_t^2X_{t-1}^2)$ in order to obtain these autocorrelations since we derived $E(X_t^i)$, $i = 1, 2, 3, 4$, in the previous section. Using (3.1) and the stationarity of the involved processes, we have

$$\begin{aligned} E(X_tX_{t-1}) &= \beta E(X_t^2\varepsilon_t) + E(X_{t-1}\varepsilon_t) \\ &= \beta E\big(\beta^2 X_{t-1}^2\varepsilon_{t-1}^2\varepsilon_t + 2\beta X_{t-1}\varepsilon_{t-1}\varepsilon_t^2 + \varepsilon_t^3\big) + E(X_{t-1}\varepsilon_t) \ . \end{aligned}$$

Taking into account the independence of the random variables ε_t, $t \in \mathbb{Z}$, and the strict stationarity of the related processes, we have $E(X_{t-1}^2\varepsilon_{t-1}^2\varepsilon_t) = \mu_1 E(X_t^2\varepsilon_t^2)$ and $E(X_{t-1}\varepsilon_{t-1}\varepsilon_t^2) = \mu_2 E(X_t\varepsilon_t)$. Then

$$E(X_tX_{t-1}) = \beta^3\mu_1 E(X_t^2\varepsilon_t^2) + 2\beta^2\mu_2 E(X_t\varepsilon_t) + \mu_1 E(X_t) + \beta\mu_3 \ .$$

Using an analogous procedure, we obtain

$$\begin{aligned} E(X_t^2X_{t-1}^2) = {} & \beta^4 E_1 + 2\beta^3 E_2 + 2\beta^3\mu_1 E_3 + 4\beta^2\mu_1 E_4 + \beta^2 E_5 + 2\beta\mu_1 E_6 \\ & + \beta^2\mu_2 E(X_t^2\varepsilon_t^2) + 2\beta\mu_1\mu_2 E(X_t\varepsilon_t) + \mu_2^2 \ , \end{aligned}$$

where

$$\begin{aligned} E_1 &= E(X_t^2X_{t-1}^2\varepsilon_t^2\varepsilon_{t-1}^2) = \beta^2\mu_2 E(X_t^4\varepsilon_t^4) + 2\beta\mu_3 E(X_t^3\varepsilon_t^3) + \mu_4 E(X_t^2\varepsilon_t^2) \ , \\ E_2 &= E(X_t^2X_{t-1}\varepsilon_t^3\varepsilon_{t-1}) = \beta^2\mu_3 E(X_t^3\varepsilon_t^3) + 2\beta\mu_4 E(X_t^2\varepsilon_t^2) + \mu_5 E(X_t\varepsilon_t) \ , \\ E_3 &= E(X_tX_{t-1}^2\varepsilon_t\varepsilon_{t-1}^2) = \beta\mu_1 E(X_t^3\varepsilon_t^3) + \mu_2 E(X_t^2\varepsilon_t^2) \ , \\ E_4 &= E(X_tX_{t-1}\varepsilon_t^2\varepsilon_{t-1}) = \beta\mu_2 E(X_t^2\varepsilon_t^2) + \mu_3 E(X_t\varepsilon_t) \ , \\ E_5 &= E(X_t^2\varepsilon_t^4) = \beta^2\mu_4 E(X_t^2\varepsilon_t^2) + 2\beta\mu_5 E(X_t\varepsilon_t) + \mu_6 \ , \\ E_6 &= E(X_t\varepsilon_t^3) = \beta\mu_3 E(X_t\varepsilon_t) + \mu_4 \ . \end{aligned}$$

Finally, the results of the previous section allow us to obtain the values of $E(X_tX_{t-1})$ and $E(X_t^2X_{t-1}^2)$ in terms of the moments of ε_t.

3.2. The Taylor property and the error process

In the following, we investigate the presence of the Taylor property in Model (3.1), considering some non-negative distributions for the error process, namely, the uniform distribution in $]0, \alpha[$, the exponential distribution in $]0, +\infty[$ with mean α, and the Pareto distribution with density $f(x) = \dfrac{\nu\,\alpha^{\nu}}{x^{\nu+1}}\,\mathbb{I}_{]\alpha,+\infty[}(x)$, for $\nu = 12$ and $\nu = 9$. In all cases, α is a non-negative parameter and the condition $E(|\ln \varepsilon_t|) < +\infty$ is satisfied.

The choice of these distributions takes into account the fact that the Taylor property seems to be related with the kurtosis value of the process. In this paper, we consider that the kurtosis of a random variable Z is given by $K_Z = M_4/M_2^2 - 3$, where M_n is the n^{th} central moment of Z, $n = 2, 4$, providing that M_4 exists (K_Z is also called "excess kurtosis"). The uniform distribution is platykurtic with a constant kurtosis value equal to -1.2, while the exponential distribution is leptokurtic with a constant kurtosis value equal to 6. On the other hand, the kurtosis of the Pareto distribution depends on the parameter ν and it is given by $\frac{6(\nu^3+\nu^2-6\nu-2)}{\nu(\nu-3)(\nu-4)}$, $\nu > 4$. This is a decreasing function of ν that goes to 6 when ν tends to infinity, and to infinity when ν tends to 4. So, the Pareto distribution is leptokurtic, no matter what is the value of ν.

We also point out that, in all cases, the condition $\beta^4\mu_4 < 1$ and the values of $\rho_X(1)$ and $\rho_{X^2}(1)$ can be written in terms of $r = \alpha\beta$.

In each case, we also present the value of the kurtosis of the process X given by (3.1), which also depends on $r = \alpha\beta$, as well as the corresponding graphic representation as a function of r. We point out that, in all these models, the leptokurtosis of the error process implies the same property for the process X. In what concerns the Taylor property and kurtosis of X, comparisons are made separately between the first two distributions, uniform and exponential, and also between the two referred Pareto distributions.

3.2.1. Error process with uniform distribution in $]0, \alpha[$

In this case, the condition $\beta^4\mu_4 < 1$ is equivalent to $0 < r < \sqrt[4]{5} \simeq 1.495$ and we obtain

$$\rho_X(1) = \frac{r\left(-180 + 120r - 51r^2 - 4r^3 + r^4\right)}{-180 + 180r - 177r^2 + 12r^3 + 7r^4},$$

$$\rho_{X^2}(1) = -\frac{r}{12}\,\frac{N_U(r)}{D_U(r)},$$

with

$$\begin{aligned} N_U(r) &= -604\,800 - 480\,600\,r - 155\,700\,r^2 - 257\,400\,r^3 - 2\,490\,r^4 + 48\,525\,r^5 \\ &\quad - 6\,270\,r^6 + 6\,810\,r^7 + 10\,620\,r^8 + 11\,384\,r^9 + 4\,012\,r^{10} - 586\,r^{11} \\ &\quad + 94\,r^{12} - 53\,r^{13} + 6\,r^{14}\ , \\ D_U(r) &= 50\,400 + 12\,600\,r + 35\,700\,r^2 + 40\,200\,r^3 + 13\,490\,r^4 + 14\,015\,r^5 + 8\,360\,r^6 \\ &\quad - 5\,210\,r^7 - 5\,999\,r^8 - 2\,407\,r^9 - 720\,r^{10} + 114\,r^{11} + 177\,r^{12} - 8\,r^{13}\ . \end{aligned}$$

From Figure 1(a), we can see that the Taylor property is present for values of r in the interval[1] $\left]1.1868987, \sqrt[4]{5}\right[$. So, for a fixed α, the Taylor property is achieved for parameterizations of Model (3.1) such that

$$\beta \in \left]\frac{1.1868987}{\alpha}, \frac{\sqrt[4]{5}}{\alpha}\right[\ .$$

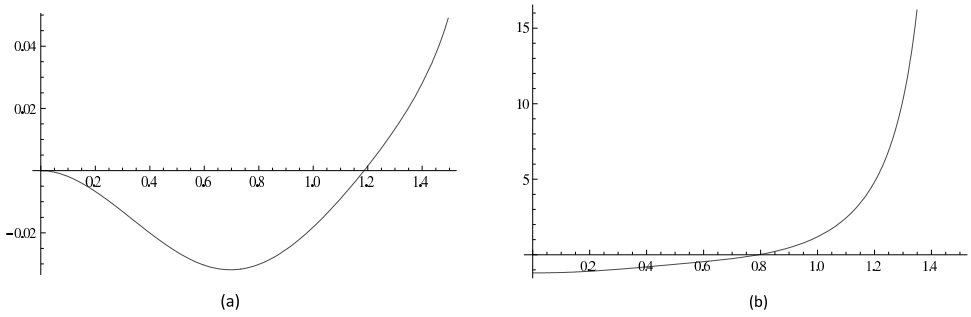

Figure 1: Graphs from $\rho_X(1) - \rho_{X^2}(1)$ (a) and $K_U(r)$ (b), $0 < r < \sqrt[4]{5}$.

For Model (3.1) with such an error process, the kurtosis is given by

$$K_U(r) = \frac{-3\,(-3+r^2)}{7\,(-4+r^3)\,(-5+r^4)}\,\frac{N_U^*(r)}{D_U^*(r)} - 3\ ,$$

where

$$\begin{aligned} N_U^*(r) &= 907\,200 - 1\,814\,400\,r + 4\,284\,000\,r^2 - 4\,510\,800\,r^3 + 3\,254\,460\,r^4 \\ &\quad - 2\,030\,520\,r^5 + 1\,973\,540\,r^6 - 617\,175\,r^7 - 185\,700\,r^8 + 371\,005\,r^9 \\ &\quad - 236\,308\,r^{10} + 78\,747\,r^{11} - 11\,496\,r^{12} + 511\,r^{13}\ , \\ D_U^*(r) &= \left(-180 + 180\,r - 177\,r^2 + 12\,r^3 + 7\,r^4\right)^2. \end{aligned}$$

From Figure 1(b), we observe that the kurtosis of this model is an increasing function of r and that the model is leptokurtic for $r > 0.8$ (approx.). We also observe that the Taylor property occurs for large values of the kurtosis, namely for $K_U(r) > 4.403$ ($\simeq K_U(1.1868987)$).

[1]The value 1.1868987 was obtained with an approximation error inferior to 5×10^{-9}.

3.2.2. Error process with exponential distribution with mean α (in $]0,+\infty[$)

The condition $\beta^4\mu_4 < 1$ is now equivalent to $0 < r < \frac{1}{\sqrt[4]{24}} \simeq 0.4518$. In this case,

$$\rho_X(1) = \frac{2r\,(2-3r+7r^2-6r^3+2r^4)}{1-2r+19r^2-20r^3+6r^4}\ ,$$

$$\rho_{X^2}(1) = 2r\,\frac{N_E(r)}{D_E(r)}\ .$$

with

$$\begin{aligned} N_E(r) = {} & -5-80\,r+65\,r^2-112\,r^3-1184\,r^4-5774\,r^5+10848\,r^6+12720\,r^7 \\ & -9408\,r^8-17880\,r^9-16272\,r^{10}+52992\,r^{11}+9216\,r^{12} \\ & -46656\,r^{13}+17280\,r^{14}\ , \end{aligned}$$

$$\begin{aligned} D_E(r) = {} & -5+2\,r-21\,r^2-602\,r^3-9060\,r^4+11126\,r^5+13252\,r^6-26448\,r^7 \\ & +16368\,r^8+13896\,r^9-12192\,r^{10}+13824\,r^{11}-12672\,r^{12}+4032\,r^{13}\,. \end{aligned}$$

So, when the errors are exponentially distributed with mean α, Model (3.1) presents the Taylor property for parameterizations such that[2]

$$\beta \in \left]0, \frac{0.0695566}{\alpha}\right[\cup \left]\frac{0.1437879}{\alpha}, \frac{1}{\sqrt[4]{24}\,\alpha}\right[\ .$$

This conclusion is illustrated in Figure 2(a). In Figure 2(b), we have the graphic representation of the kurtosis of Model (3.1) with exponential errors, which is given by

$$K_E(r) = \frac{-3\,(-1+2r^2)}{(-1+6r^3)\,(-1+24r^4)}\,\frac{N_E^*(r)}{D_E^*(r)} - 3\ ,$$

where

$$\begin{aligned} N_E^*(r) = {} & 3-12\,r+52\,r^2-134\,r^3+11815\,r^4-36752\,r^5+44802\,r^6+1062\,r^7 \\ & -42648\,r^8+17028\,r^9+12240\,r^{10}+5616\,r^{11}-17280\,r^{12}+6048\,r^{13}\,, \end{aligned}$$

$$D_E^*(r) = \left(1-2\,r+19\,r^2-20\,r^3+6\,r^4\right)^2\ .$$

As in the previous case, the kurtosis of Model (3.1) is an increasing function of r but the process X is always leptokurtic in this case. Again, we observe that large kurtosis values correspond to large values of the difference $\rho_X(1)-\rho_{X^2}(1)$. In fact, the Taylor property is clearly present in this model for kurtosis values greater than 13 ($\simeq K_E(0.16)$).

[2]The values 0.0695566 and 0.1437879 were obtained with an approximation error inferior to 5×10^{-8}.

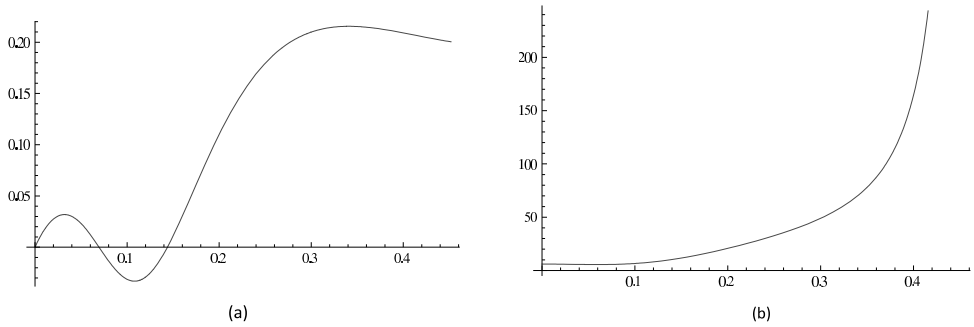

Figure 2: Graphs from $\rho_X(1) - \rho_{X^2}(1)$ (a) and $K_E(r)$ (b), $0 < r < \frac{1}{\sqrt[4]{24}}$.

We also observe that the kurtosis of the process X is larger when the errors are exponentially distributed than when they are uniformly distributed, corresponding to an analogous relation between the kurtosis of those error processes. The Taylor property seems to emerge in a relatively stronger way when the kurtosis of X increases.

3.2.3. Error process with Pareto density $f(x) = \dfrac{12\,\alpha^{12}}{x^{13}}\,\mathbb{I}_{]\alpha,+\infty[}(x)$

The region of existence of the autocorrelations in terms of $r = \alpha\beta$ is now defined by $0 < r < \sqrt[4]{\frac{2}{3}} \simeq 0.9036$. We have

$$\rho_X(1) = \frac{44\,r\,(6050 - 10230\,r + 13035\,r^2 - 7524\,r^3 + 1296\,r^4)}{3\,(36300 - 79200\,r + 219255\,r^2 - 171160\,r^3 + 29472\,r^4)}\,,$$

$$\rho_{X^2}(1) = \frac{r}{55}\,\frac{N_{P12}(r)}{D_{P12}(r)}\,,$$

with

$$\begin{aligned} N_{P12}(r) = {} & -7043652000 - 5638479000\,r - 1900483200\,r^2 - 6228372150\,r^3 \\ & - 3064649280\,r^4 + 2622844140\,r^5 + 24533447400\,r^6 \\ & + 19854650865\,r^7 + 11360213480\,r^8 - 16340416020\,r^9 \\ & - 30235824828\,r^{10} + 23037530976\,r^{11} + 7650162960\,r^{12} \\ & - 11215587456\,r^{13} + 2802615552\,r^{14}\,, \end{aligned}$$

$$\begin{aligned} D_{P12}(r) = {} & -58697100 + 14229600\,r - 142425360\,r^2 - 468153840\,r^3 \\ & - 218936564\,r^4 + 536116224\,r^5 + 616017864\,r^6 \\ & + 374454192\,r^7 + 130906149\,r^8 - 805701976\,r^9 \\ & - 15605040\,r^{10} + 401099652\,r^{11} \\ & - 245871648\,r^{12} + 48736320\,r^{13}\,. \end{aligned}$$

Concerning the kurtosis of this model, it is given by

$$K_{P12}(r) \;=\; \frac{-2(-5+6r^2)}{49(-3+4r^3)\,(-2+3r^4)}\,\frac{N^*_{P12}(r)}{D^*_{P12}(r)} - 3\;,$$

where

$$\begin{aligned}
N^*_{P12}(r) \;=\;& 599\,933\,276\,250 - 2\,617\,890\,660\,000\,r + 4\,970\,166\,270\,300\,r^2 \\
& - 5\,546\,727\,078\,200\,r^3 + 5\,904\,1720\,498\,845\,r^4 - 161\,234\,870\,633\,760\,r^5 \\
& + 126\,074\,334\,149\,694\,r^6 + 2\,238\,307\,939\,140\,r^7 + 25\,296\,348\,317\,400\,r^8 \\
& - 57\,875\,913\,071\,352\,r^9 - 89\,078\,826\,937\,116\,r^{10} + 180\,941\,306\,693\,040\,r^{11} \\
& - 102\,607\,682\,886\,720\,r^{12} + 19\,713\,391\,884\,288\,r^{13} \\
D^*_{P12}(r) \;=\;& \left(36\,300 - 79\,200\,r + 219\,255\,r^2 - 171\,160\,r^3 + 29\,472\,r^4\right)^2\;.
\end{aligned}$$

As can be seen in Figure 3(a), the Taylor property is now achieved for all considered parameterizations of Model (3.1). From Figure 3(b), we conclude that the process X is always leptokurtic.

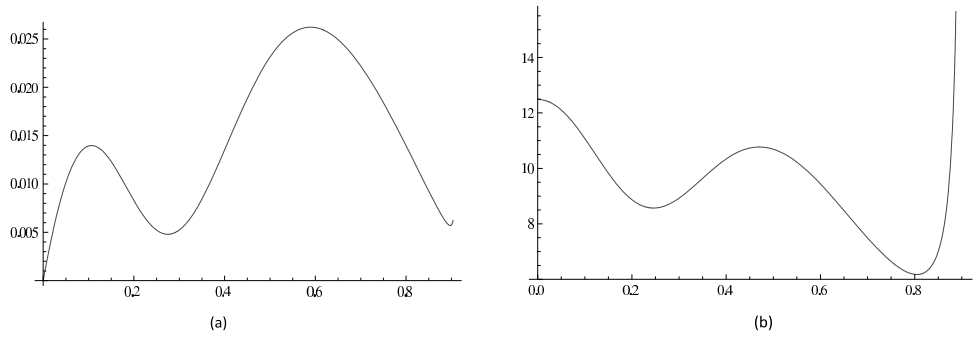

Figure 3: Graphs from $\rho_X(1)-\rho_{X^2}(1)$ (a) and $K_{P12}(r)$ (b), $0 < r < \sqrt[4]{\frac{2}{3}}$.

3.2.4. Error process with Pareto density $f(x) = \dfrac{9\alpha^9}{x^{10}}\,\mathbb{I}_{]\alpha,+\infty[}(x)$

We have

$$\beta^4\mu_4 < 1 \iff 0 < r < \sqrt[4]{\tfrac{5}{9}} \simeq 0.863 \quad \text{and}$$

$$\begin{aligned}
\rho_X(1) \;&=\; \frac{8r\,(15\,680 - 27\,720\,r + 39\,564\,r^2 - 27\,864\,r^3 + 6\,561\,r^4)}{47\,040 - 105\,840\,r + 343\,119\,r^2 - 315\,504\,r^3 + 73\,791\,r^4} \\
\rho_{X^2}(1) \;&=\; \frac{r}{48}\,\frac{N_{P9}(r)}{D_{P9}(r)}\;,
\end{aligned}$$

with

$$\begin{aligned} N_{P9}(r) \;=\; & -67737600 - 83339200\,r + 19038600\,r^2 - 88401600\,r^3 \\ & - 148138920\,r^4 - 511287075\,r^5 + 1466330040\,r^6 + 1499354145\,r^7 \\ & - 1537629480\,r^8 - 1966005837\,r^9 - 602608896\,r^{10} \\ & + 3869347563\,r^{11} - 61620912\,r^12 - 2818841796\,r^13 + 1179090432\,r^{14} \end{aligned}$$

$$\begin{aligned} D_{P9}(r) \;=\; & -627200 + 235200\,r - 1650600\,r^2 - 8601600\,r^3 - 13809280\,r^4 \\ & + 31729095\,r^5 + 27010080\,r^6 - 23002305\,r^7 - 21773448\,r^8 \\ & - 24182469\,r^9 + 58517640\,r^{10} + 9248823\,r^{11} \\ & - 50143536\,r^{12} + 19665504\,r^{13} \;. \end{aligned}$$

The Taylor property is also present for all considered parameterizations of Model (3.1), as it is illustrated in Figure 4(a), and we point out that the magnitude of the difference $\rho_X(1) - \rho_{X^2}(1)$ is greater in this case than in the case $\nu = 12$.

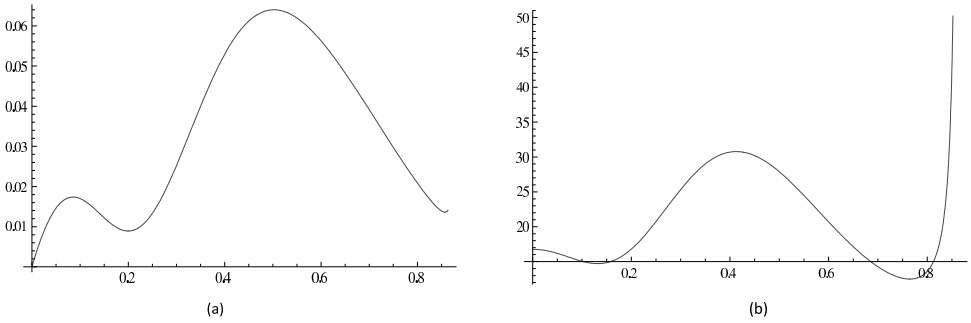

Figure 4: Graphs from $\rho_X(1) - \rho_{X^2}(1)$ (a) and $K_{P9}(r)$ (b), $0 < r < \sqrt[4]{\frac{5}{9}}$.

The kurtosis of Model (3.1) is now given by

$$K_{P9}(r) \;=\; \frac{7 - 9r^2}{9(-2+3r^3)\,(-5+9r^4)}\,\frac{N^*_{P9}(r)}{D^*_{P9}(r)} \;-\; 3 \;,$$

where

$$\begin{aligned} N^*_{P9}(r) \;=\; & 62449049600 - 281020723200\,r + 532657440000\,r^2 - 582241598400\,r^3 \\ & + 25718506014670\,r^4 - 92872063045440\,r^5 + 100396353649230\,r^6 \\ & - 6337711636725\,r^7 - 8536591340550\,r^8 - 41782534519365\,r^9 \\ & - 62336742758694\,r^{10} + 195729014255481\,r^{11} \\ & - 145385404543008\,r^{12} + 356648081091 93\,r^{13} \end{aligned}$$

$$D^*_{P9}(r) \;=\; \left(15680 - 35280\,r + 114373\,r^2 - 105168\,r^3 + 24597\,r^4\right)^2 \;.$$

The process X is also leptokurtic for all considered values of r. We observe that the kurtosis of the process X is greater when $\nu = 9$ than when $\nu = 12$, corresponding to an analogous relation between the kurtosis of the respective error processes. In these two examples, it is seen again how the Taylor property emerges when the process X is leptokurtic.

As regards the Pareto distribution, graphic representations for several values of ν also suggest that the presence of the Taylor property is stronger for higher values of the kurtosis of the process X. In fact, as functions of ν, the difference $\rho_X(1) - \rho_{X^2}(1)$ seems to increase when $K_{P\nu}(r)$ increases, for all values of r that satisfy the condition $\beta^4\mu_4 < 1$. This situation is illustrated in Figure 5 and strongly contributes to conjecture that the Taylor property and leptokurtosis are highly related in time series.

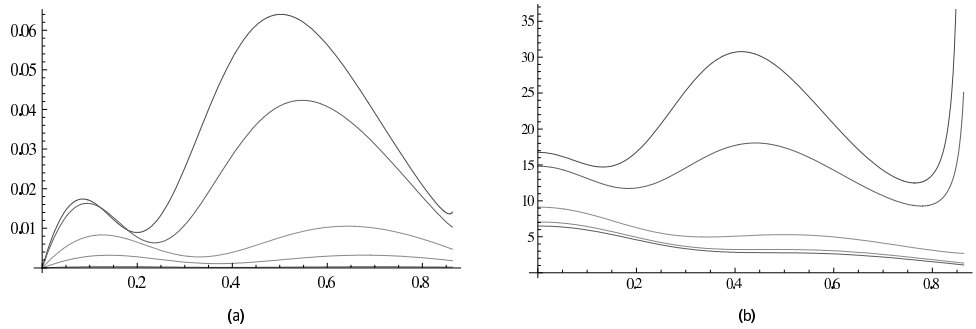

Figure 5: Graphs from $\rho_X(1) - \rho_{X^2}(1)$ (a) and $K_{P\nu}(r)$ (b), $\nu = 100, 50, 20, 10, 9$ (from bottom to top), $0 < r < \sqrt[4]{\frac{5}{9}}$.

4. THE TAYLOR PROPERTY IN THE CASE OF SYMMETRICALLY DISTRIBUTED ERRORS: SIMULATION STUDY

When the errors are symmetrically distributed, the autocorrelation function of X^2 for Model (1.1) verifies $\rho_{X^2}(1) = 0$, if $k > 1$ (Martins, [6]). So, in this case, the property $\rho_{|X|}(1) > \rho_{X^2}(1)$ is equivalent to $\rho_{|X|}(1) > 0$. However, the autocorrelation function of the process $(|X_t|,\, t \in \mathbb{Z})$ is not available when the error process is allowed to assume negative values. To investigate the presence of the Taylor property in Model (3.1) with symmetrically distributed errors, we perform a simulation study considering the simple first-order bilinear diagonal model with an i.i.d. error process $(\varepsilon_t,\, t \in \mathbb{Z})$ with four symmetrical distributions with unit variance, namely, the uniform distribution in $]-\sqrt{3}, \sqrt{3}[$, the standard normal distribution, and the distribution of a variable $\varepsilon = \sqrt{\frac{\nu-2}{\nu}}\, Y$, where Y has a Student distribution with ν degrees of freedom ($\nu = 30$ and $\nu = 9$). In each case,

the condition $E(|\ln|\varepsilon_t||) < +\infty$ is satisfied and parameterizations that satisfy $\beta^4\mu_4 < 1$ are considered in the simulations. For each value of the parameter β and each one of the considered distributions, we generate 500 observations according to the corresponding model and obtain the 95% confidence intervals for the probability that such a model satisfies the Taylor property. The results appear in Table 1 (where NA means "Not Applicable", due to the fact that the corresponding value of β does not satisfy the condition $\beta^4\mu_4 < 1$). The special values 0.69, 0.74, 0.75 and 0.863 are the greatest values of β such that $\beta^4\mu_4 < 1$ for each one of the considered distributions.

Table 1: 95% confidence intervals for the probability that the model with symmetrical innovations presents the Taylor property.

β	$U(]-\sqrt{3},\sqrt{3}[)$	$N(0,1)$	$\sqrt{\frac{14}{15}}\,Y,\ Y\sim T(30)$	$\sqrt{\frac{7}{9}}\,Y,\ Y\sim T(9)$
0.01	[0.373, 0.627]	[0.459, 0.708]	[0.459, 0.708]	[0.476, 0.724]
0.05	[0.357, 0.610]	[0.373, 0.627]	[0.373, 0.627]	[0.407, 0.660]
0.1	[0.140, 0.360]	[0.292, 0.541]	[0.214, 0.453]	[0.260, 0.506]
0.2	[0, 0]	[0, 0.105]	[0, 0.049]	[0, 0.049]
0.3	[0, 0]	[0, 0]	[0, 0]	[0, 0.079]
0.4	[0, 0]	[0, 0]	[0, 0.079]	[0.260, 0.506]
0.5	[0, 0]	[0.155, 0.379]	[0.292, 0.541]	[0.699, 0.901]
0.6	[0, 0]	[0.566, 0.801]	[0.603, 0.831]	[0.781, 0.953]
0.69	[0, 0]	[0.802, 0.965]	[0.802, 0.965]	[0.951, 1]
0.74	[0, 0.079]	[0.847, 0.987]	[0.870, 0.996]	NA
0.75	[0.004, 0.130]	[0.847, 0.987]	NA	NA
0.863	[0.566, 0.801]	NA	NA	NA

We can observe that the Taylor property seems to be present for high values of β and that this presence increases with the kurtosis of the error process, as we have established and observed in non-negative bilinear models.

The confidence intervals corresponding to small values of β do not allow us to infer about the presence of the Taylor property, as they certainly correspond to values of β for which the difference $\rho_X(1) - \rho_{X^2}(1)$ is close to zero.

5. CONCLUSIONS

In this paper, we analyze the presence of the Taylor property in first-order bilinear time series models. For this analysis we evaluate the autocorrelations of the process X and of X^2. Considering X non-negative, we discuss the presence of the Taylor property taking several distributions for the error process, chosen according to the kurtosis value as this property is strongly related with the value

of this parameter. More precisely, the Taylor property seems to emerge when the process X is leptokurtic.

Based on a simulation study, we also analyze the presence of the Taylor property in the class of real valued first-order bilinear diagonal models with symmetrical innovations.

The studies presented here show that bilinear models are able to reproduce the Taylor effect. They also reinforce the connection of the Taylor property to leptokurtic models which has been observed in the few theoretical studies developed until now. In fact, He and Teräsvirta ([5]), Gonçalves, Leite and Mendes-Lopes ([1]) and Haas ([4]) show the presence of this property in some conditional heteroskedastic models, which are leptokurtic processes. Moreover, all the cases considered in this paper also show that, when the Taylor property occurs, the model is leptokurtic.

We still observe that leptokurtosis is not enough to induce the Taylor property. Examples of bilinear models that are leptokurtic but do not have the Taylor property are $X_t = X_{t-1}\varepsilon_{t-1} + \varepsilon_t$, where ε_t is uniformly distributed in $[0, 1]$, and $X_t = 0.5\, X_{t-1}\varepsilon_{t-1} + \varepsilon_t$, where ε_t is exponentially distributed with mean 0.2. This is in line with the simulation results of He and Teräsvirta ([5]) suggesting that the Taylor property is not present for the standard GARCH$(1, 1)$ process with normal errors.

In conclusion, our study allows to conjecture that a general assessment of the Taylor property in the bilinear process is strongly dependent on the magnitude of its tails weight.

6. APPENDIX

A working example to illustrate the results of Section 2, namely evaluation of $E(X_t^n \varepsilon_t^n)$ and $E(X_t^n)$, $n \leq 4$, for a first-order bilinear process is now presented.

Let us suppose that ε_t, $t \in \mathbb{Z}$, is exponentially distributed with density $f(x) = \frac{1}{\alpha}\, e^{-x/\alpha}\, \mathbb{I}_{]0,+\infty[}(x)$. Then $\mu_n = n!\, \alpha^n$, $n \in \mathbb{N}$. In this case, the condition $\beta^4 \mu_4 < 1$ is equivalent to $0 < r < \frac{1}{\sqrt[4]{24}}$, where $r = \alpha\beta$. Under this hypothesis, and taking into account that ε_t is independent of $X_{t-1}^n \varepsilon_{t-1}^n$, $t \in \mathbb{Z}$, and that the process $(X_t\, \varepsilon_t,\, t \in \mathbb{Z})$ is strictly stationary, we have

$$E(X_t \varepsilon_t) \;=\; E(\beta X_{t-1}\varepsilon_{t-1}\varepsilon_t) + E(\varepsilon_t^2) \;=\; \beta E(X_t \varepsilon_t)\, \mu_1 + \mu_2$$

which is equivalent to

$$E(X_t\, \varepsilon_t) \;=\; \frac{2\alpha^2}{1 - r}\ . \tag{6.1}$$

Then, by (2.3), we have

$$
\begin{aligned}
E(X_t^2\varepsilon_t^2) &= \frac{1}{1-\beta^2\mu_2}\left(2\beta\mu_3 E(X_t\varepsilon_t)+\mu_4\right)\\
&= 24\,\alpha^4\,\frac{1}{(1-r)\,(1-2r^2)}\ .
\end{aligned}
\tag{6.2}
$$

Taking into account (2.3), (6.1) and (6.2), we now obtain

$$
\begin{aligned}
E(X_t^3\varepsilon_t^3) &= \frac{1}{1-\beta^3\mu_3}\left(3\beta^2\mu_4 E(X_t^2\varepsilon_t^2)+3\beta\mu_5 E(X_t\varepsilon_t)+\mu_6\right)\\
&= \frac{1}{1-6r^3}\left(\frac{1\,728\,\alpha^6 r^2}{(1-r)\,(1-r^2)}+\frac{720\,\alpha^6 r}{1-r}+720\,\alpha^6\right)\\
&= 144\,\alpha^6\,\frac{2r^2+5}{(1-r)\,(1-2r^2)\,(1-6r^3)}\ .
\end{aligned}
\tag{6.3}
$$

Finally, we evaluate $E(X_t^4\varepsilon_t^4)$ using (2.3), (6.1), (6.2) and (6.3).

$$
\begin{aligned}
E(X_t^4\varepsilon_t^4) &= \frac{1}{1-\beta^4\mu_4}\left(4\beta^3\mu_5 E(X_t^3\varepsilon_t^3)+6\beta^2\mu_6 E(X_t^2\varepsilon_t^2)+4\beta\mu_7 E(X_t\varepsilon_t)+\mu_8\right)\\
&= \frac{1}{1-24\,r^4}\left(\frac{69\,120\,\alpha^8 r^3\,(2r^2+5)}{(1-r)\,(1-r^2)\,(1-r^3)}\right.\\
&\quad\left.+\ \frac{103\,680\,\alpha^8 r^2}{(1-r)\,(1-r^2)}+\frac{40\,320\,\alpha^8 r}{1-r}+40\,320\,\alpha^8\right)\\
&= 5\,760\,\alpha^8\,\frac{18r^3+4r^2+7}{(1-r)\,(1-2r^2)\,(1-6r^3)\,(1-24r^4)}\ .
\end{aligned}
\tag{6.4}
$$

The values of $E(X_t^n)$, $n\le 4$, are then given by (2.2). More precisely,

$$
E(X_t) = \beta E(X_t\varepsilon_t)+\mu_1 = \alpha\,\frac{1+r}{1-r}\ ,
\tag{6.5}
$$

$$
\begin{aligned}
E(X_t^2) &= \beta^2 E(X_t^2\varepsilon_t^2)+2\mu_1 E(X_t\varepsilon_t)+\mu_2\\
&= \frac{24\,\alpha^4\beta^2}{(1-r)\,(1-2r^2)}+\frac{4\alpha^3\beta}{1-r}+2\alpha^2\\
&= 2\alpha^2\,\frac{1+r+10r^2-2r^3}{(1-r)\,(1-2r^2)}\ ,
\end{aligned}
\tag{6.6}
$$

$$
\begin{aligned}
E(X_t^3) &= \beta^3 E(X_t^3\varepsilon_t^3)+3\beta^2\mu_1 E(X_t^2\varepsilon_t^2)+3\beta\mu_2 E(X_t\varepsilon_t)+\mu_3\\
&= \frac{144\,\alpha^6\beta^3(2r^2+5)}{(1-r)\,(1-2r^2)\,(1-6r^3)}+\frac{72\,\alpha^5\beta^2}{(1-r)\,(1-2r^2)}+\frac{12\,\alpha^4\beta}{1-r}+6\alpha^3\\
&= 6\alpha^3\,\frac{1+r+10r^2+112r^3-6r^4-12r^5+12r^6}{(1-r)\,(1-2r^2)\,(1-6r^3)}
\end{aligned}
\tag{6.7}
$$

and

$$
\begin{aligned}
E(X_t^4) &= \beta^4 E(X_t^4\varepsilon_t^4) + 4\beta^3\mu_1 E(X_t^3\varepsilon_t^3) + 6\beta^2\mu_2 E(X_t^2\varepsilon_t^2) + 4\beta\mu_3 E(X_t\varepsilon_t) + \mu_4 \\
&= \frac{15\,760\,\alpha^8\beta^4(8r^3+4r^2+7)}{(1-r)(1-2r^2)(1-6r^3)(1-24r^4)} + \frac{576\,\alpha^7\beta^3(2r^2+5)}{(1-r)(1-2r^2)(1-6r^3)} \\
&\quad + \frac{288\,\alpha^6\beta^2}{(1-r)(1-2r^2)} + \frac{48\,\alpha^5\beta}{1-r} + 24\,\alpha^4 \\
&= \frac{24\,\alpha^4 D(r)}{(1-r)(1-2r^2)(1-6r^3)(1-24r^4)}\,,
\end{aligned}
\tag{6.8}
$$

with

$$
\begin{aligned}
D(r) &= 1 + r + 10\,r^2 + 112\,r^3 + 1\,650\,r^4 - 36\,r^5 + 732\,r^6 + 1\,632\,r^7 \\
&\quad + 144\,r^8 + 288\,r^9 - 288\,r^{10}\,.
\end{aligned}
$$

ACKNOWLEDGMENTS

The authors are grateful to the referee for his constructive comments and suggestions.

REFERENCES

[1] Gonçalves, E.; Leite, J. and Mendes-Lopes, N. (2009). A mathematical approach to detect the Taylor property in TARCH processes, *Statistics & Probability Letters*, **79**, 602–610.

[2] Gonçalves, E.; Mendes-Lopes, N.; Dorotovic, I.; Fernandes, J.M. and Garcia, A. (2014). North and South Hemispheric Solar Activity for Cycles 21–23: Asymmetry and Conditional Volatility of Plage Region Areas, *Solar Physics*, **289**(6), 2283–2296.

[3] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series, Nonparametric and Parametric Methods*, Springer Series in Statistics, Springer.

[4] Haas, M. (2009). Persistence in volatility, conditional kurtosis, and the Taylor property in absolute value GARCH processes, *Statistics & Probability Letters*, **79**, 1674–1683.

[5] He, C. and Teräsvirta, T. (1999). Properties of moments of a family of GARCH processes, *Journal of Econometrics*, **92**, 173–192.

[6] Martins, C.M. (1997). A note on the autocorrelations related to a bilinear model with non-independent shocks, *Statistics & Probability Letters*, **36**, 245–250.

[7] Pereira, I. and Scotto, M.G. (2006). On the non-negative first order exponential bilinear time series model, *Statistics & Probability Letters*, **76**, 931–938.

[8] Quinn, B.G. (1982). Stationarity and invertibility of simple bilinear models, *Stochastic Processes and their Applications*, **12**, 225–230.

[9] Taylor, S. (1986). *Modelling Financial Time Series*, Wiley.

[10] Zhang, Z. and Tong, H. (2001). On some distributional properties of a first order non-negative bilinear time series model, *Journal of Applied Probability*, **38**, 659–671.

REVSTAT – Statistical Journal
Volume 13, Number 3, November 2015, 227–231

FAILURE-TIME WITH DELAYED ONSET

Authors: Man Yu Wong
– Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China
mamywong@ust.hk

D.R. Cox
– Nuffield College, Oxford OX1 1NF, United Kingdom
david.cox@nuffield.ox.ac.uk

Received: February 2014 Accepted: May 2014

Abstract:

- A special distribution is suggested for the analysis of survival data in which there is a long random delay before the onset of the terminal process. Estimation by the method of moments and by maximum likelihood is compared.

Key-Words:

- *survival data; unobserved time origin; moment estimation; maximum likelihood estimation.*

AMS Subject Classification:

- 62F12, 62N02.

1. INTRODUCTION

Lai [1] has reviewed the rich variety of parametric families of distributions that have been suggested for the analysis of survival data. With the exception of the Weibull distribution and its connection with extreme value theory, most are essentially empirically-based flexible representations covering the rich variety of distributional shapes encountered in such contexts. Here we give a distribution extremely restricted in form but motivated by a very simple model of a data generating process.

A key element in the analysis of survival data is often the choice of time origin, for example, for patients, birth, entry into a study, first report of symptoms, etc. Appropriate choice may greatly clarify interpretation. Sometimes, however, the natural time origin is unobserved. In this paper we outline a very special model for such a situation. Each study individual has an observed time origin and we assume that for a long time following that there is no possibility of a critical event. Then an unobserved transition occurs and following that the critical event rate becomes high. If the processes of transition and occurrence arise in independent Poisson processes of respectively small rate ρ_1 and large rate ρ_2, it follows that the failure-time has the form $T = T_1 + T_2$, where T_1 and T_2 are independently exponentially distributed with rates ρ_1 and ρ_2, with $\rho_1 << \rho_2$. This inequality is crucial both for separate estimation of the two parameters and indeed for the interpretation of the model.

Then the probability density function of T can be written as

$$\frac{e^{-t/\mu_1} - e^{-t/\mu_2}}{\mu_1 - \mu_2} \ ,$$

where it is convenient to parameterize in terms of the means $\mu_i = 1/\rho_i$ for $i = 1, 2$.

2. METHODS OF ESTIMATION

One simple method of estimating the two parameters is from the first two moments. This requires solving the equations

$$\begin{aligned} \bar{t} &= \tilde{\mu}_1 + \tilde{\mu}_2 \ , \\ s_t^2 &= \tilde{\mu}_1^2 + \tilde{\mu}_2^2 \ , \end{aligned}$$

where $\bar{t}$ and s_t^2 are sample mean and sample variance of observed values of T and $\tilde{\mu}_1 > \tilde{\mu}_2$, thus defining the moment estimators. Provided that $1 \geq s_t^2/\bar{t}^2 \geq 1/2$,

as dictated by the special construction of the distribution,

$$\tilde{\mu}_1 = \frac{\bar{t}}{2} + \frac{1}{2}\sqrt{(2s_t^2 - \bar{t}^2)}\ ,$$
$$\tilde{\mu}_2 = \frac{\bar{t}}{2} - \frac{1}{2}\sqrt{(2s_t^2 - \bar{t}^2)}\ .$$

Notionally more efficient estimates will be given by the method of maximum likelihood. This involves numerical solution of the following two equations:

$$\sum_{i=1}^{n} \frac{t_i \exp(-t_i/\hat{\mu}_1)}{\{\exp(-t_i/\hat{\mu}_1) - \exp(-t_i/\hat{\mu}_2)\}} = \frac{n\hat{\mu}_1^2}{\hat{\mu}_1 - \hat{\mu}_2}\ ,$$
$$\sum_{i=1}^{n} \frac{t_i \exp(-t_i/\hat{\mu}_2)}{\{\exp(-t_i/\hat{\mu}_1) - \exp(-t_i/\hat{\mu}_2)\}} = \frac{n\hat{\mu}_2^2}{\hat{\mu}_1 - \hat{\mu}_2}\ ,$$

with $\hat{\mu}_1 > \hat{\mu}_2$. Both sets of estimates are sensible only if the data are consistent with the constraints on the squared coefficient of variation implied by the model and if one mean is substantially greater than the other.

3. COMPARISON OF ESTIMATORS

To compare the methods of estimation we compute asymptotic variances. For the moment estimates we have by local linearization, the delta method, that

$$\mathrm{var}(\tilde{\mu}_1) = \frac{\mu_2^2\left(2\lambda^4 - 2\lambda^3 + 2\lambda^2 + 1\right)}{n(\lambda-1)^2}\ ,$$
$$\mathrm{var}(\tilde{\mu}_2) = \frac{\mu_2^2\left(\lambda^4 + 2\lambda^2 - 2\lambda + 2\right)}{n(\lambda-1)^2}\ ,$$

where $\lambda = \mu_1/\mu_2$.

For the maximum likelihood estimates, the inverse of the Fisher information matrix gives

$$\mathrm{var}(\hat{\mu}_1) = \frac{\mu_2^2\lambda^4\left\{2\lambda^3\zeta - (\lambda-1)^2\right\}}{2n\left\{\lambda^3(\lambda^2+1)\zeta - (\lambda^2-\lambda+1)(\lambda-1)^2\right\}}\ ,$$
$$\mathrm{var}(\hat{\mu}_2) = \frac{\mu_2^2\left\{2\lambda^3\zeta + (\lambda-1)^2(\lambda^4 - 2\lambda^3 + 2\lambda - 2)\right\}}{2n\left\{\lambda^3(\lambda^2+1)\zeta - (\lambda^2-\lambda+1)(\lambda-1)^2\right\}}\ ,$$

where ζ is the generalized Riemann zeta function

$$\zeta\left[3,\ 2 + 1/(k-1)\right] = \sum_{k=0}^{\infty}\left\{k + 2 + 1/(k-1)\right\}^{-3}\ .$$

The efficiencies of the moment estimators for μ_1 and μ_2 depend only on the ratio of the two means, i.e., on λ. The efficiencies decrease as λ increases.

4. NUMERICAL COMPARISON

These results have been explored by simulation. Without loss of generality we set $\mu_2 = 1$ and generate 1000 sets of data for sample sizes 100 and 500 and for $\lambda = 2, 5, 10$. When $n = 100$ and $\lambda = 2$, approaching one-half the samples fall outside the range of validity of the above results, that is, are descriptively inconsistent with the model. This falls to about 15% when $n = 500$. In the more realistic case of larger λ, the incompatible samples are rare and theoretical and empirical variances agree reasonably closely, the variance of the maximum likelihood estimator being appreciably smaller than that of the moment estimate as λ becomes larger. When $\lambda = 10$, the asymptotic efficiency of $\tilde{\mu}_1$ is close to 50%, its limiting value for large λ, whereas that of $\tilde{\mu}_2$ is 25%, dropping slowly to its limiting value of zero.

5. DISCUSSION

The model could be generalized in various ways, for example to include uninformative censoring or a more complex transition process. With such generalizations simple estimation by the method of moments would typically not be possible. If the representation is plausible on general grounds and fits the data it would be very desirable to find a different type of observation predictive of the origin of the second component and study of such a marker would lead to a further generalization of the present model, which will, however, not be discussed here. If additional explanatory variables were available maximum likelihood estimation is likely to be needed. Finally note that the identification of the later stage parameter μ_2 with the smaller of the two estimates depends entirely on the prior specification.

A quite different approach to this kind of data is investigated in as yet unpublished paper by Peter McCullagh, University of Chicago to whom we are grateful for comments on an earlier version of the present note.

REFERENCES

[1] Lai, C.D. (2013). Constructions and applications of lifetime distributions, *Applied Stochastic Models in Business and Industry*, **29**, 127–140.

REVSTAT – Statistical Journal
Volume 13, Number 3, November 2015, 233–243

MARGINAL HOMOGENEITY MODEL FOR ORDERED CATEGORIES WITH OPEN ENDS IN SQUARE CONTINGENCY TABLES

Authors: Serpil Aktas
– Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey
spxl@hacettepe.edu.tr

Song Wu
– Department of Applied Mathematics and Statistics, Stony Brook University, NY, USA
song.wu@stonybrook.edu

Received: August 2013 Revised: December 2013 Accepted: May 2014

Abstract:

- A marginal homogeneity model tests whether the row and column distributions of a square contingency table have the same sample margins. However, for variables with ordered categories, the marginal homogeneity model does not take into account the ordering information, leading to significant loss of power. Score-based tests have been proposed for ordinal variables. In this paper, we extend the idea of scores and propose a new method based on the standardized scores to test the marginal homogeneity for ordered categories with open ends. Our simulation studies demonstrate that our proposed scores is more powerful than the usual scores.

Key-Words:

- *marginal homogeneity model; square contingency table; score.*

AMS Subject Classification:

- 49A05, 78B26.

1. MARGINAL HOMOGENEITY MODEL

Square contingency tables often arise in social, behavioral sciences and medical studies [1], and are used to display joint responses of two variables that have the same category levels. For example, to examine whether there exists any difference in unaided distance vision between right and left eyes in Royal Ordnance factories in Britain, 7477 employed women aged 30–39 were sampled between 1943–1946. Based on the data, a square contingency table can be constructed, with the row variable for the right eye vision grade and the column variable for the left eye vision grade. The right and left eyes have exactly the same vision grades, which are ordered based on the same criterion, e.g., from the best to the worst. Williams (1952) investigated the possibility of assessing association in a two-way table based on scores, which may be assigned to one or both of the row and column variables.

To study the symmetry of such square contingency tables, several models have been developed, including the complete symmetry model, the quasi-symmetry model, and the marginal homogeneity model [1, 3, 4, 5]. Consider a two-way $r \times r$ square contingency table with the same row and column classifications, let p_{ij} denote the cell probability that an observation falls in the ith row and jth column of the table ($i = 1, 2, ..., r; j = 1, 2, ..., r$). Stuart (1955) proposed the hypothesis of marginal homogeneity that can be expressed in terms of the marginal cell probabilities

$$H_0 : \; p_{i+} = p_{+i} \,, \quad i = 1, ..., r \;,$$

which is equivalent to the hypothesis that

$$H_0 : \text{ The two samples have the same marginal distribution} \,.$$

The p_{i+} and p_{+i} are the marginal probabilities of ith row and ith column, respectively. This hypothesis is tested with the test statistic

$$Q = n \, \mathbf{d}' \hat{V}^{-1} \mathbf{d} \;, \tag{1.1}$$

where $\mathbf{d}' = (d_1, ..., d_r)$ with $d_i = p_{i+} - p_{+i}$, n is sample size, and $\hat{V}$ is the maximum likelihood estimate of the covariance matrix. The elements of $\hat{V}$ are

$$\hat{v}_{ij} = -(p_{ij} + p_{ji}) \quad \text{for } i \neq j \qquad \text{and} \qquad \hat{v}_{ii} = p_{i+} + p_{+i} - 2p_{ii} \;. \tag{1.2}$$

Q has the chi-square distribution with $r - 1$ degrees of freedom. Later, Bhapkar (1979) proposed a similar type of test by taking the elements of covariance matrix as

$$\begin{aligned} \hat{v}_{ij} &= -(p_{ij} + p_{ji}) - (p_{i+} - p_{+i})\,(p_{j+} - p_{+j}) \qquad \text{for} \quad i \neq j \;, \\ \hat{v}_{ii} &= p_{i+} + p_{+i} - 2p_{ii} - (p_{i+} - p_{+i})^2 \;. \end{aligned} \tag{1.3}$$

A main disadvantage of the marginal homogeneity model is that it does not take into account the ordering information for ordinal variables, *i.e.* marginal homogeneity hypothesis is invariant to change of orders in the variable categories. However, one may be interested in whether one marginal distribution is stochastically larger or smaller than the other. If the distribution of cases across the row categories is the same as the distribution of cases across the column categories, these margins can be referred to as homogeneous and the distributions are the same; Otherwise, margins would be referred to as heterogeneous.

Agresti (1983) considered testing marginal homogeneity for ordinal categorical variables by using some fixed scores to weigh the marginal probability differences between corresponding row and column categories. The test statistics is:

$$d = \sum_{i=1}^{r} w_i(p_{i+} - p_{+i}) \tag{1.4}$$

for some fixed scores $\{w_i\}$. The estimated variance of d is

$$\hat{S}_d^2 = \frac{1}{n}\left(\sum_{i,j}(w_i - w_j)^2 p_{ij} - d^2\right). \tag{1.5}$$

Then for large n

$$z = \frac{d}{\hat{S}_d} \tag{1.6}$$

has approximately standard normal distribution under the null hypothesis [8]. Fleiss and Everitt (1971) also considered different forms of marginal homogeneity test. Relation of Marginal Homogeneity model with the other models were given in [1, 10, 11]. Caussinus' quasi-symmetry model can hold true in contingency tables in which the row and column marginals are not homogeneous [12].

2. STANDARDIZED SCORES FOR OPEN ENDED CATEGORIES

In the score-based methods, different score choices will lead to different test statistics and consequently provide different conclusions. Score can be assigned either based on distributional assumptions or based on prior knowledge [11]. Agresti (1983) defined some score sets, for example, for four categories, the scores $3, 1, -1, -3$) to detect differences in location and $1, -1, -1, 1$ to detect differences in dispersion. In well-balanced data, the impact of the choice of scores on the final inference is minimal [8]. However, when the data are very unbalanced or in an open-ended form, results may significantly change with respect to choice of scores [8].

Several scores have been suggested in the literatures that they could handle two-way contingency tables with open-ended categories, in which one or both of categories may take the form of "greater than" or "less than". For example, Graubard and Korn (1987) discussed the equally spaced scores for $2\times C$ contingency tables, and found that equally spaced scores might yield conservative results. In general, midrank is a very useful score choice when there are large differences in marginal counts. However, if the distribution is highly skewed within an interval, midpoints are also poor estimates of the true values. Particularly, midrank scores can be very unreasonable in applications when the marginals are far from uniform. Other ways to estimate scores include using latent root analysis that maximizes the correlation between the two sets [2], and optimizing conventional scores for a particular set of variables by comparing the squared correlation coefficients with the monotonic correlation ratios [14]. However, all these scores ignored the open-ended features of the categories.

To overcome the limitations of the equally spaced scores when variables have open ended categories, here we propose standardized z-scores based on semi-interquartile range for the row and column variables. The median is not only an appropriate measure for ordinal, interval, and ratio scale data, but also is well known to be the most convenient measure of location for the open ended categories. The standardized z-scores are defined as follows:

$$z_i = \frac{s_i - Q_2}{0.5\, IQR}\ , \tag{2.1}$$

$$z_j = \frac{s_j - Q_2}{0.5\, IQR}\ , \tag{2.2}$$

where,

IQR**:** interquartile range, $IQR = Q_3 - Q_1$

Q_1**:** first quartile

Q_2**:** second quartile (median)

Q_3**:** third quartile

z_i**:** ith row scores

z_j**:** jth row scores

s_i & s_j**:** midpoints of row and column categories calculated as

$$s_i = \frac{LL_i - UL_i}{2}\ , \qquad s_j = \frac{LL_j - UL_j}{2}$$

LL & UL**:** the lower and upper limits of a class, respectively.

Note that the semi-interquartile range is a good measure of spread for skewed distributions.

3. SIMULATION STUDY

Simulation studies have been carried out to investigate the statistical properties of our newly proposed standardized scores. We used two different schemes to simulate the data. The first one assumes that the row and column variables have the same marginal distribution, whereas the second one assumes that the row and column variables have different marginal distributions. The combinations of different sample size ($n = 50, 100, 500$), different table dimensions ($R = 5, 8$), and different levels of association between row and column variables ($\rho = 0.0, 0.5, 0.9$) were considered. For each simulation scenario, 1000 replications were performed. The simulated data were analyzed by both the standardized-score methods and the usual-score method defined in Agresti (1983), which allows the comparisons between the two methods. Matlab and SAS softwares were used for generating and analyzing data sets.

Table 1 shows the comparison of Type I errors between the usual and standardized scores, under the assumption that the row and column variables have the same marginal distribution. To construct the square contingency table, we applied the similar simulation strategy in Yang et al (2012). We first generated random numbers from a bivariate normal distribution with the same means of 25 and common variance of 36, and then the bivariate samples were cross tabulated into a two-way contingency table. Different correlation coefficient were assumed ($\rho = 0.0, 0.5, 0.9$) to evaluate how the association strength may affect the tests. Random samples were classified into 5×5 or 8×8 cross tables with the following categories:

$$X, Y(5\times5) \ : \ < 9.9, 10\text{–}14.9, 15\text{–}19.9, 20\text{–}24.9, > 25\,; \quad \text{or}\,,$$
$$X, Y(8\times8) \ : \ < 4.9, 5\text{–}9.9, 10\text{–}14.9, 15\text{–}19.9, 20\text{–}24.9, 25\text{–}29.9, 30\text{–}34.9, > 35\,.$$

Note that X and Y denote the row and column variables, and the first and last class intervals are open-ended classes. Hence, $R\times R$ contingency tables with open-ended ordinal variables were constructed. The row and column marginals are expected to be the same. In each simulation scenario, 1000 replication runs were performed to estimate the Type I error rates for both methods. Three different significance levels ($\alpha = 0.01, 0.05, 0.10$) were considered. As shown in Table 1, the actual Type I errors are very close to the nominal levels in each case, suggesting the validation of both score methods.

Next, we made the power comparison between the standardized and usual score methods. Similarly, we generated random numbers from a bivariate normal distribution with different means of 25 and 36, and common variance of 36, and the random samples were classified into square tables. In this way, the row and column variables have the different marginal distributions. Let $\alpha = 0.05$ denote the nominal level of significance of the tests, the empirical power of the tests can

be calculated as the proportion of the test statistic is greater than the critical value, which is given by $P(X^2 > C)/t$, where t is the number of replications in the simulation study, and C is the critical value of the chi-square distribution for $\alpha = 0.05$ with associated degrees of freedom $R - 1$.

Table 1: Comparison of Type I errors between the usual and standardized scores.

α	Usual Scores				Standardized Scores			
	ρ	n	$R=5$	$R=8$	ρ	n	$R=5$	$R=8$
0.01	0	50	0.0062	0.0105	0	50	0.0082	0.0088
		100	0.0106	0.0118		100	0.0098	0.0086
		500	0.008	0.0096		500	0.0101	0.0186
	0.5	50	0.0038	0.0048	0.5	50	0.008	0.0016
		100	0.0062	0.0054		100	0.0116	0.0101
		500	0.0084	0.0089		500	0.0094	0.0102
	0.9	50	0.0072	0.0104	0.9	50	0.0074	0.0112
		100	0.0076	0.0094		100	0.0988	0.0093
		500	0.0082	0.009		500	0.0099	0.0099
0.05	0	50	0.0352	0.0334	0	50	0.0398	0.0339
		100	0.0398	0.0395		100	0.0458	0.0444
		500	0.0421	0.044		500	0.0482	0.0468
	0.5	50	0.0383	0.0323	0.5	50	0.0418	0.0482
		100	0.0406	0.0456		100	0.0433	0.0506
		500	0.0456	0.0456		500	0.0439	0.0494
	0.9	50	0.0392	0.3297	0.9	50	0.0483	0.0456
		100	0.0438	0.0431		100	0.0489	0.0467
		500	0.0439	0.0437		500	0.0499	0.0494
0.10	0	50	0.0676	0.0668	0	50	0.0709	0.0701
		100	0.069	0.0687		100	0.0974	0.0974
		500	0.0768	0.0754		500	0.091	0.0905
	0.5	50	0.072	0.0651	0.5	50	0.0756	0.0756
		100	0.068	0.0954		100	0.0829	0.0829
		500	0.0826	0.1028		500	0.0965	0.0965
	0.9	50	0.0712	0.0808	0.9	50	0.0876	0.0877
		100	0.0823	0.0843		100	0.0943	0.0949
		500	0.0966	0.0985		500	0.0997	0.0996

As shown in Table 2, the proposed standardized scores has much higher power than the usual score procedures. For both methods, the power is substantially greater for larger sample size and correlation, that is, the power for

detecting the marginal heterogeneity is the lowest for $\rho = 0$, highest for $\rho = 0.90$, and lowest for $n = 50$, highest for $n = 500$.

Table 2: Empirical power comparison between the usual and standardized scores for $\alpha = 0.05$.

Usual Scores				Standardized Scores			
ρ	n	$R=5$	$R=8$	ρ	n	$R=5$	$R=8$
0	50	0.5754	0.5308	0	50	0.7782	0.7689
	100	0.6106	0.6326		100	0.7234	0.7316
	500	0.63	0.6596		500	0.7301	0.7886
0.5	50	0.6388	0.6768	0.5	50	0.7465	0.79
	100	0.7762	0.7754		100	0. 8065	0.8112
	500	0.8884	0.8109		500	0.9008	0.9102
0.9	50	0.7172	0.7103	0.9	50	0.8874	0.8812
	100	0.8876	0.8594		100	0.897	0.8711
	500	0.9092	0.9008		500	0.9976	0.9576

4. NUMERICAL EXAMPLES

A hypothetical 5×5 square contingency table with both row and column having open-ended categories is generated to illustrate the utilization and efficiency of the standardized scores (Table 3).

Table 3: A simulated 5×5 table.

R/C	≤ 9.9	10–14.9	15–19.9	20–24.9	≥ 25	Total
≤ 9.9	24	23	34	12	45	138
10–14.9	37	7	5	25	32	106
15–19.9	48	11	17	37	22	135
20–24.9	28	9	7	17	13	74
≥ 25	6	13	15	5	8	47
Total	143	63	78	96	120	500

The sample quartiles are calculated as:
For row variable: $Q_1 = 9.47$, $Q_2 = 15.17$, $Q_3 = 19.80$.
For column variable: $Q_1 = 9.32$, $Q_2 = 17.77$, $Q_3 = 24.68$.

Interquartile ranges for the row and column variables are $IQR = 19.80 - 9.47 = 10.33$; $IQR = 24.68 - 9.32 = 15.36$, respectively.
Using Equations (2.1) and (2.2), standardized scores are displayed in Table 4. The scores in Equation (1.4) would be $z_i \times z_j$ due to the nature of open ends.

Table 4: Standardized scores for row and column variables.

Standardized row scores (z_i)	Standardized column scores (z_j)
−1.49468	−1.34375
−0.52662	−0.69271
0.44143	−0.04167
1.40948	0.609375
2.377541	1.260417

Using the standardized scores we get,
$d = 0.085$, $\hat{S}_d^2 = 0.0053$, $\hat{S}_d = 0.0728$, $z = \frac{0.085}{0.0728} = 1.1675$.
Therefore, the null hypothesis of marginal homogeneity is not rejected.

When we utilize the usual scores as: $w_i = w_j = \{-3, -1, 0, 1, 3\}$, we get,
$d = 0.538$, $\hat{S}_d^2 = 0.01875$, $\hat{S}_d = 0.13693$, $z = \frac{0.538}{0.13693} = 3.929$.
The result indicates that null hypothesis of marginal homogeneity ($p_{i+} = p_{+i}$) is rejected at 0.05 significance level.

5. CONCLUSIONS

Ordinal variables are common in many research areas. Marginal homogeneity model tests that the marginal frequencies do not differ significantly between the row and column variables. Marginal homogeneity model requires assigning scores through row and column variables. The problem for open ended categories is to assign the proper scoring. The simplest scoring method is admittedly integer scoring. The standardized scores employing the marginal homogeneity test in the presence of an open-ended category is proposed in this paper. Ordinal models require assigning scores to levels of ordinal variables. When responses are ordered categories, it is usually important to test the hypotheses of marginal homogeneity using ordinal information. When the variation of the between variable levels in contingency tables are large, standardized scores will be appropriate. Different choices of the row and the column scores can lead to different conclusion concerning association of the rows and columns. When we employ different scores in the modeling, inferences derived from the analyses would be dependent on the scoring system.

The proposed scores give better results than the usual scores with respect to their statistical power values in detecting marginal heterogeneity. Results also show that the use of ordinal approach becomes relatively more efficient as correlation coefficient and the sample size increase. We showed that the usual score and standardized score methods can achieve similar type I errors when data were simulated under null hypothesis, while the standardized score method has larger power than usual score method when data were simulated under alternative hypothesis. When ordinal variables in a two contingency table are a discretized form of continuous variables, it is reasonable to use the standardized scores based on sample quartiles. Our simulation suggests that the proposed method competes well with alternative.

ACKNOWLEDGMENTS

We acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Bishop, Y.M.; Fienberg, S. and Holland, P.W. (1975). *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA.

[2] Williams, E.J. (1952). Use of scores for the analysis of association in contingency tables, *Biometrika*, **39**, 274–289.

[3] Ireland, C.T.; Ku, H.H. and Kullback, S. (1969). Symmetry and marginal homogeneity of r×r contingency tables, *J. Amer. Statist. Assoc.*, **64**, 1323–1341.

[4] Tomizawa, S.; Miyamoto, N. and Ouchi, M. (2006). Decompositions of symmetry model into marginal homogeneity and distance subsymmetry in square contingency tables with ordered categories, *REVSTAT – Statistical Journal*, **4**, 2.

[5] Yang Z.; Sun X. and Hardin, J.W. (2012). Testing marginal homogeneity in clustered matched-pair data, *J. Stat. Plan. Infer.*, **141**, 1313–1318.

[6] Stuart, A. (1955). A test for homogeneity of the marginal distribution in a two-way classification, *Biometrika*, **42**(11), 412–416.

[7] Bhapkar, V.P. (1979). On tests of symmetry when higher order interactions are absent, *J. Indian Statist. Assoc.*, **17**, 17–26.

[8] Agresti, A. (1983). Testing marginal homogeneity for ordinal categorical variables, *Biometrics*, **39**(2), 505–510.

[9] Fleiss, J.R. and Everitt, B.S. (1971). Comparing the marginal totals of square contingency tables, *British Journal of Mathematical and Statistical Psychology*, **24**, 117–123.

[10] Agresti, A. (2013). *Categorical Data Analysis*, Wiley, New York.

[11] Lawal, H.B. (2012). *Categorical Data Analysis with SAS and SPSS Applications*, LEA, Mahwah, New Jersey.

[12] Goodman, L.A. (2002). Contributions to the statistical analysis of contingency tables: Notes on quasi-symmetry, quasi-independence, log-linear models, log-bilinear models and correspondence analysis model, *Ann. Fac. Sci. Toulouse Math*, **6**(11), 525–540.

[13] Graubard, B.I. and Korn, E.L. (1987). Choice of column scores for testing independence in ordered 2×k contingency tables, *Biometrics*, **43**, 471–476.

[14] Allen, M.P. (1976). Conventional and optimal interval scores for ordinal variables, *Sociological Methods & Research*, **4**, 475–494.

REVSTAT – Statistical Journal
Volume 13, Number 3, November 2015, 245–262

BAYESIAN AND NON-BAYESIAN INTERVAL ESTIMATORS FOR THE POISSON MEAN

Authors: S. Nadarajah
– School of Mathematics, University of Manchester,
Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

M. Alizadeh
– Department of Statistics, Shahid Bahonar University of Kerman
Kerman, Iran

S.F. Bagheri
– Department of Statistics, Razi University of Kermanshah
Kermanshah, Iran

Received: December 2013 Revised: May 2014 Accepted: May 2014

Abstract:

- Seventeen different confidence intervals for the Poisson mean are compared using simulations and a real data application. The interval estimators include the Wald interval estimator, the score interval estimator, the exact interval estimator, the bootstrap interval estimator, the equal tails credible interval estimator, Jeffreys prior credible interval estimator, the HPD credible interval estimator and the relative surprise credible interval estimator. Recommendations for choosing among the seventeen intervals are given for different sample sizes and different Poisson means.

Key-Words:

- *coverage length; coverage probability; Poisson distribution.*

AMS Subject Classification:

- 62E20.

1. INTRODUCTION

Poisson distribution has wide spread applications in almost every area of the sciences, engineering and medicine. Hence, it is important that accurate estimators are available for its rate parameter.

Many authors have studied estimation of the Poisson mean. A comparison of nine interval estimators for a Poisson mean when the expected number of events ≤ 5 is given in Barker [2]. An easy to use method to approximate Poisson confidence limits is discussed by Bégaud *et al.* [4]. Three interval estimators for linear functions of Poisson rates are given in Stamey and Hamilton [31]. Asymptotic interval estimators for Poisson regression are studied by Michael and Adam [24]. Swift [32] gives recommendations for choosing between twelve different confidence intervals for the Poisson mean. Bayesian interval estimation for the difference of two independent Poisson rates for under reported data is considered in Greer *et al.* [18]. Improved prediction intervals for binomial and Poisson distributions are given in Krishnamoorthy and Jie [20]. Simple approximate procedures for constructing binomial and Poisson tolerance intervals are given in Krishnamoorthy *et al.* [21]. Interval estimators for the difference between two Poisson rates are given in Li *et al.* [22]. Interval estimation for misclassification rate parameters in a complementary Poisson model is described by Riggs *et al.* [28]. Patil and Kulkarni [27] compare nineteen confidence intervals for the Poisson mean. See also Byrne and Kabaila [7] and Ng *et al.* [26].

Most studies we are aware of have compared the performances of only classical interval estimators for the Poisson mean: all of the nine estimators considered in Barker [2] are classical interval estimators; only one of the estimators considered in Swift [32] is a Bayesian credible interval estimator; all of the nineteen estimators considered in Patil and Kulkarni [27] are classical interval estimators; and so on. Also none of these papers have used a real data set to compare the performance of the estimators.

The aim of this note is a comparison study of classical interval estimators as well as Bayesian credible interval estimators for the Poisson mean. We consider equal numbers of classical interval estimators and Bayesian credible interval estimators with a range of priors considered for the latter. In total, we compare seventeen different interval estimators for the Poisson mean. Our comparison is based on simulations as well as a real data set.

The contents of this note are organized as follows. In Sections 2 and 3, several interval estimators are described for the Poisson mean. Section 2 describes the following classical interval estimators: the Wald interval estimator, the score interval estimator, the exact interval estimator, and the bootstrap interval esti-

mator. Section 3 describes the following Bayesian credible estimators: the equal tails credible interval estimator, Jeffreys prior credible interval estimator, the HPD credible interval estimator and the relative surprise credible interval estimator. Each of the estimators in Section 3 was calculated under four different priors: uniform prior; exponential prior; gamma prior; chisquare prior. Section 4 performs a simulation study comparing the performance of the estimators of Sections 2 and 3. The performance is compared in terms of coverage probabilities and coverage lengths. A real data application is described in Section 5. Finally, some conclusions are noted in Section 6.

2. INTERVAL ESTIMATORS FOR POISSON MEAN

In this section, some methods to obtain interval estimators for the Poisson mean are described.

2.1. Approximate interval estimator

Here, we use some large sample methods for constructing interval estimators. Suppose $T(\underline{X})$ is an estimator based on sample mean such that

$$\sqrt{n}\,\frac{T(\underline{X})-\theta}{\sqrt{\nu(\theta)}} \xrightarrow{L} Z,$$

where $Z \sim N(0,1)$ and $\xrightarrow{L}$ means convergence in distribution (Chung [10]). Suppose further that there is a statistic $S(\underline{X})$ so that $\nu(\theta) \xrightarrow{p} S(\underline{X})$. Then, by Slutsky 's theorem,

$$\sqrt{n}\,\frac{T(\underline{X})-\theta}{\sqrt{S(\underline{X})}} \xrightarrow{L} Z\ .$$

We can obtain an approximate interval estimator for θ with confidence coefficient $1-\alpha$ by inverting the inequality (Rohatgi and Ehsanes Saleh [29]):

$$\left|\sqrt{n}\,\frac{T(\underline{X})-\theta}{\sqrt{S(\underline{X})}}\right| \leq z_{1-\alpha/2}\ .$$

In the following, we construct approximate interval estimators for the Poisson mean. Let $X_1, X_2, ..., X_n$ be a random sample from a Poisson distribution with mean λ. We consider two interval estimators.

a) The score interval. By using

$$Q = \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \sim N(0,1) \ , \qquad n \to \infty \ ,$$

we can write

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} < z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha \ .$$

So, we have

$$\left| \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \right| < z_{1-\frac{\alpha}{2}}$$

or

$$\lambda^2 - \lambda \left(2\overline{X} + \frac{z^2_{1-\frac{\alpha}{2}}}{n} \right) + \overline{X}^2 < 0 \ .$$

By solving this inequality, we see that

$$\left(\frac{\left(2\overline{X} + \frac{z^2_{1-\frac{\alpha}{2}}}{n} \right) - \sqrt{\Delta}}{2}, \frac{\left(2\overline{X} + \frac{z^2_{1-\frac{\alpha}{2}}}{n} \right) + \sqrt{\Delta}}{2} \right)$$

is an interval estimator for λ with confidence coefficient $1 - \alpha$, where

$$\Delta = \frac{z^2_{1-\frac{\alpha}{2}}}{n} \left(\frac{z^2_{1-\frac{\alpha}{2}}}{n} + 4\overline{X} \right),$$

see Shao [30].

b) The Wald interval. We know that $\widehat{\lambda} = \overline{X}$ is the maximum likelihood estimator for λ, so

$$Q = \frac{\overline{X} - \lambda}{\sqrt{\frac{\overline{X}}{n}}} \sim N(0,1) \ , \qquad n \to \infty \ .$$

So,

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}} \, , \ \overline{X} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}} \right)$$

is an interval estimator for λ with confidence coefficient $1 - \alpha$. Sometimes $\overline{X} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}}$ can be less than zero. In this case, we use the interval estimator

$$\left(\max \left(0 \, , \ \overline{X} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}} \right), \ \overline{X} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}}{n}} \right) .$$

2.2. Exact interval estimator

Let $X_1, X_2, ..., X_n$ be a random sample from a Poisson distribution with mean λ. Let $Y = \sum_{i=1}^{n} X_i$. We know that Y is a sufficient statistic for λ and $Y \sim$ Poisson$(n\lambda)$. Then, an exact interval estimator for λ with confidence coefficient $1-\alpha$ is given by

$$\left(\frac{1}{2n}\chi^2_{2y,\frac{\alpha}{2}}\,,\ \frac{1}{2n}\chi^2_{2(y+1),1-\frac{\alpha}{2}}\right).$$

For $y = 0$, we take $\chi^2_{0,1-\frac{\alpha}{2}} = 0$ (Casella and Berger [9]). Although an exact confidence interval estimator exists, it is still of interest to compare asymptotic and exact estimators. The readers are referred to Agresti and Coull [1]. They show that approximate approaches are better than the "exact" approach for interval estimation of the binomial distribution. Because Poisson distribution is a discrete distribution similar to the binomial distribution, it is of interest to investigate the performance of different interval estimators.

2.3. Bootstrap confidence intervals

Here, we use the percentile bootstrap method (see Davison and Hinkley [11] for details) to construct confidence intervals for λ. The percentile bootstrap method is popular: for example, Ibrahim and Kudus [19] used it to construct confidence intervals for the median of a three-parameter Weibull distribution.

The percentile bootstrap method can be applied as follows:

1. For a random sample $X_1, X_2, ..., X_n$ from a Poisson distribution with mean λ, compute the maximum likelihood estimate $\widehat{\lambda} = \overline{X}$.

2. Random select n observations from $X_1, X_2, ..., X_n$ with replacement.

3. Repeat step 2 B times to generate B bootstrap samples, say $X_1^j, X_2^j, ..., X_n^j$, $1 \le j \le B$.

4. Compute the maximum likelihood estimate $\widehat{\lambda}^j = \overline{X}^j$ of λ for each of the bootstrap samples in step 3.

5. Based on $\widehat{\lambda}^1, \widehat{\lambda}^2, ..., \widehat{\lambda}^B$, a $100(1-\alpha)$ percentile bootstrap confidence interval is

$$\left(2\widehat{\lambda} - \widehat{\lambda}^b,\ 2\widehat{\lambda} - \widehat{\lambda}^a\right),$$

where $a = (B+1)\frac{\alpha}{2}$ and $b = (B+1)\left(1-\frac{\alpha}{2}\right)$.

3. BAYESIAN CREDIBLE INTERVALS

In this section, we discuss four Bayesian credible intervals for the Poisson mean. Bayesian credible intervals incorporate problem-specific contextual information from the prior distribution into estimates, whereas classical interval estimators are based solely on the data. In real applications, we should employ Bayesian approaches whenever strong prior information exist. This could provide good coverage and relatively narrow intervals for the parameter. We consider Bayesian credible intervals for the Poisson distribution under different priors.

3.1. Posterior distributions under different priors

The efficiency of Bayesian framework is largely dependent upon the choice of an appropriate prior distribution. The prior information is combined to the current information to update the belief regarding a particular characteristic of the data. The prior information can be of two types; informative and non-informative priors. Though, the choice of a prior depends upon the circumstances of the study, the search for a suitable prior is always of interest. We utilize both informative and non-informative priors for our posterior analysis.

Let $X_1, X_2, ..., X_n$ be a random sample from Poisson(λ). The prior and posterior distributions considered are as follows:

(**a**) For the uniform prior,

$$\pi(\lambda) \propto 1 \ , \qquad \lambda > 0 \ , \tag{3.1}$$

the posterior distribution is

$$\pi\left(\lambda|\mathbf{x}\right) = \frac{n^{\sum_{i=1}^{n} x_i+1}}{\Gamma\left(\sum_{i=1}^{n} x_i\right)} \lambda^{\left(\sum_{i=1}^{n} x_i+1\right)-1} e^{-n\lambda} \ ,$$

$$\left[Gamma\left(\sum_{i=1}^{n} x_i + 1, n\right)\right]. \tag{3.2}$$

(**b**) For Jeffreys prior,

$$\pi(\lambda) \propto \lambda^{-\frac{1}{2}} \ , \qquad \lambda > 0 \ , \tag{3.3}$$

the posterior distribution is

$$\pi\left(\lambda|\mathbf{x}\right) = \frac{n^{\sum_{i=1}^{n} x_i+\frac{1}{2}}}{\Gamma\left(\sum_{i=1}^{n} x_i+\frac{1}{2}\right)} \lambda^{\left(\sum_{i=1}^{n} x_i+\frac{1}{2}\right)-1} e^{-n\lambda} ,$$

$$\left[Gamma\left(\sum_{i=1}^{n} x_i+\frac{1}{2}, n\right)\right]. \tag{3.4}$$

(**c**) For the exponential prior,

$$\pi(\lambda) = ae^{-a\lambda} , \qquad \lambda > 0 , \ \ a > 0 , \tag{3.5}$$

where a is a hyper parameter, the posterior distribution is

$$\pi\left(\lambda|\mathbf{x}\right) = \frac{(n+a)^{\sum_{i=1}^{n} x_i+1}}{\Gamma\left(\sum_{i=1}^{n} x_i\right)} \lambda^{\left(\sum_{i=1}^{n} x_i+1\right)-1} e^{-(n+a)\lambda} ,$$

$$\left[Gamma\left(\sum_{i=1}^{n} x_i+1, \, n+a\right)\right]. \tag{3.6}$$

(**d**) For the gamma prior,

$$\pi(\lambda) = \frac{a^b}{\Gamma(b)} \lambda^{b-1} e^{-a\lambda} , \qquad \lambda > 0 , \ \ a > 0 , \ \ b > 0 , \tag{3.7}$$

where a and b are hyper parameters, the posterior distribution is

$$\pi\left(\lambda|\mathbf{x}\right) = \frac{(n+a)^{\left(\sum_{i=1}^{n} x_i+b\right)}}{\Gamma\left(\sum_{i=1}^{n} x_i\right)} \lambda^{\left(\sum_{i=1}^{n} x_i+b\right)-1} e^{-(n+a)\lambda} ,$$

$$\left[Gamma\left(\sum_{i=1}^{n} x_i+b, \, n+a\right)\right]. \tag{3.8}$$

(**e**) For the chisquare prior,

$$\pi(\lambda) = \frac{\lambda^{\frac{b}{2}-1} e^{-\frac{\lambda}{2}}}{\Gamma\left(\frac{b}{2}\right) 2^{\frac{b}{2}}} , \qquad \lambda > 0 , \ \ b > 0 , \tag{3.9}$$

where b is a hyper parameter, the posterior distribution is

$$\pi\left(\lambda|\mathbf{x}\right) = \frac{\left(n+\frac{1}{2}\right)^{\left(\sum_{i=1}^{n} x_i+\frac{b}{2}\right)}}{\Gamma\left(\sum_{i=1}^{n} x_i\right)} \lambda^{\left(\sum_{i=1}^{n} x_i+\frac{b}{2}\right)-1} e^{-\left(n+\frac{1}{2}\right)\lambda} ,$$

$$\left[Gamma\left(\sum_{i=1}^{n} x_i+\frac{b}{2}, \, n+\frac{1}{2}\right)\right]. \tag{3.10}$$

For more discussion, see Feroze and Aslam [15].

In the following, we find Bayesian credible intervals based on the derived posteriors.

3.2. Equal tails credible intervals

Table 1 presents the $1-\alpha$ equal tails credible intervals.

Table 1: The $1-\alpha$ equal tails credible intervals under the different priors and posteriors.

Priors	Pivotal quantity	Lower bound	Upper bound
Uniform	$2n\lambda$	$\frac{1}{2n}\,\chi^2_{2\left(\sum_{i=1}^n x_i+1\right),\,1-\frac{\alpha}{2}}$	$\frac{1}{2n}\,\chi^2_{2\left(\sum_{i=1}^n x_i+1\right),\,\frac{\alpha}{2}}$
Jeffreys	$2n\lambda$	$\frac{1}{2n}\,\chi^2_{2\left(\sum_{i=1}^n x_i+\frac{1}{2}\right),\,1-\frac{\alpha}{2}}$	$\frac{1}{2n}\,\chi^2_{2\left(\sum_{i=1}^n x_i+\frac{1}{2}\right),\,\frac{\alpha}{2}}$
Exponential	$2(n+a)\,\lambda$	$\frac{1}{2(n+a)}\,\chi^2_{2\left(\sum_{i=1}^n x_i+1\right),\,1-\frac{\alpha}{2}}$	$\frac{1}{2(n+a)}\,\chi^2_{2\left(\sum_{i=1}^n x_i+1\right),\,\frac{\alpha}{2}}$
Gamma	$2(n+a)\,\lambda$	$\frac{1}{2(n+a)}\,\chi^2_{2\left(\sum_{i=1}^n x_i+b\right),\,1-\frac{\alpha}{2}}$	$\frac{1}{2(n+a)}\,\chi^2_{2\left(\sum_{i=1}^n x_i+b\right),\,\frac{\alpha}{2}}$
Chisquare	$2\left(n+\frac{1}{2}\right)\lambda$	$\frac{1}{2\left(n+\frac{1}{2}\right)}\,\chi^2_{2\left(\sum_{i=1}^n x_i+\frac{b}{2}\right),\,1-\frac{\alpha}{2}}$	$\frac{1}{2\left(n+\frac{1}{2}\right)}\,\chi^2_{2\left(\sum_{i=1}^n x_i+\frac{b}{2}\right),\,\frac{\alpha}{2}}$

3.3. Jeffreys prior credible intervals

The non-informative Jeffreys prior plays a special role in the Bayesian analysis, see, for example, Berger [5]. In particular, Jeffreys prior is the unique first-order probability matching prior for a real-valued parameter with no nuisance parameter, see Ghosh [16]. In our setting, simple calculations show that the Fisher information about μ is $I(\mu) = n\left(\mu + b\mu^2\right)^{-1}$ and thus Jeffreys prior is proportional to

$$I^{\frac{1}{2}}(\mu) = n^{\frac{1}{2}}\left(\mu + b\mu^2\right)^{-\frac{1}{2}}.$$

Denoting the posterior distribution by J, the $(1-\alpha)$ Jeffreys credible interval for μ can be written as

$$(J_\alpha,\, J_{1-\alpha})\,, \tag{3.11}$$

where $J_{1-\alpha}$ and J_α are, respectively, the $1-\alpha$ and α quantiles of the posterior distribution based on n observations (Cai [8]). By (3.3) and (3.4), we can rewrite

(3.11) as

$$\left(G_{\alpha/2,\sum_{i=1}^{n} x_i+\frac{1}{2},n}\,,\, G_{1-\alpha/2,\sum_{i=1}^{n} x_i+\frac{1}{2},n}\right).$$

For more discussion, see Brown *et al.* [6].

3.4. HPD credible intervals

The set $\{\theta : \pi(\theta|\underline{\mathbf{x}}) \geq k\}$ is called highest posterior density, where k is chosen so that

$$1-\alpha = \int_{\{\theta:\,\pi(\theta|\underline{\mathbf{x}})\geq k\}} \pi(\theta|\underline{\mathbf{x}})\, d\theta\ .$$

See Casella and Berger [9]. If the posterior pdf, $\pi(\theta|\underline{\mathbf{x}})$, is unimodal then the HPD set would be an interval, say $(\theta_{HL}, \theta_{HU})$ (Berger [5]). In this case, we construct HPD credible intervals for parameters of interest in the square:

$$\pi(\theta_{HL}|\underline{\mathbf{x}}) = \pi(\theta_{HU}|\underline{\mathbf{x}})\ , \qquad \int_{\theta_{HL}}^{\theta_{HU}} \pi(\theta|\underline{\mathbf{x}})\, d\theta = 1-\alpha\ .$$

3.5. Relative surprise credible intervals

Relative surprise credible intervals for θ, as discussed in Evans [12], are based on a particular approach to assessing the null hypothesis $H_0 : \theta = \theta_0$. For this, we compute the observed relative surprise (ORS) defined by

$$\pi\left(\frac{\pi(\theta|\underline{\mathbf{x}})}{\pi(\theta)} > \frac{\pi(\theta_0|\underline{\mathbf{x}})}{\pi(\theta_0)}\Big|\underline{\mathbf{x}}\right). \tag{3.12}$$

We see that (3.12) compares the relative increase in belief for θ, from a priori to a posteriori. Other approaches to measuring surprise are discussed in Good [17]. For estimation purposes, one may consider ORS in (3.12) as a function of θ_0 and select a value which minimizes this quantity as the estimator, called the least relative surprise estimator (LRSE). Moreover, to obtain a $1-\alpha$-credible region for θ, we simply invert (3.12) in the standard way to obtain the $(1-\alpha)$-relative surprise credible interval provided that

$$\pi\left(\frac{\pi(\theta|\underline{\mathbf{x}})}{\pi(\theta)} > \frac{\pi(\theta_0|\underline{\mathbf{x}})}{\pi(\theta_0)}\Big|\underline{\mathbf{x}}\right) \leq 1-\alpha\ .$$

It can be proved that if the posterior pdf $\pi(\theta|\underline{\mathbf{x}})$ is unimodal then the credible set is of the form $(\theta_{RL}, \theta_{RU})$ such that

$$\frac{\pi(\theta_{RL}|\underline{\mathbf{x}})}{\pi(\theta)} = \frac{\pi(\theta_{RU}|\underline{\mathbf{x}})}{\pi(\theta)}\ , \qquad \int_{\theta_{RL}}^{\theta_{RU}} \pi(\theta|\underline{\mathbf{x}})\, d\theta = 1-\alpha\ .$$

Relative surprise credible regions are shown to minimize, among Bayesian credible regions, the prior probability of covering a false value from the prior. Such regions are also shown to be unbiased in the sense that the prior probability of covering a false value is bounded above by the prior probability of covering the true value. Relative surprise credible regions are shown to maximize both the Bayes factor in favor of the region containing the true value and the relative belief ratio, among all credible regions with the same posterior content (Evans and Shakhatreh [13]).

3.6. Reparameterizations

A basic principle of inference is that inferences about a parameter of interest should be invariant under reparameterizations: for example, whatever rule we use to obtain a $(1-\alpha)$-credible region, $B_{1-\alpha}$, for a parameter of interest, θ, the rule should yield the region $\Psi(B_{1-\alpha})$ for any one-to-one, sufficiently smooth, reparameterization $\psi = \Psi(\theta)$. Relative surprise credible inferences satisfy this principle. For greater detail, see Evans and Shakhatreh [13] and Baskurt and Evans [3].

4. COMPARISON OF CONFIDENCE INTERVALS

In this section, we compare the interval estimators of Sections 2 and 3: the Wald (WA) interval estimator, the score (SC) interval estimator, the exact (EX) interval estimator, Jeffreys (Jef) prior credible estimator, the bootstrap (Boot) interval estimator, the HPD credible interval estimator, the relative surprise (RS) credible interval estimator and the equal tails (EQ) credible interval estimator. Note that the HPD, RS and the EQ credible interval estimators depend on the chosen prior. Others do not depend on the chosen prior.

The comparison is based on coverage probabilities and coverage lengths computed by simulation. Each coverage probability and coverage length was computed over ten thousand replications of the simulated sample. Throughout, the level of significance was taken to be five percent.

The parameters of the priors can be chosen either arbitrarily or using empirical Bayes (EB) estimation. EB estimation is discussed in the Appendix. But our simulations showed that both arbitrary choice and EB estimation gave the same results. So, we choose the prior parameters arbitrarily as $a = 3$ and $b = 2$.

4.1. Comparison based on coverage probability

Here, we compare the interval estimators based on their coverage probabilities. Figures 1 to 9 in Nadarajah *et al.* [25] show how the coverage probabilities vary with respect to sample size and λ for the classical interval estimators and for the priors and posteriors given by (3.1)–(3.10). The following observations can be drawn from the figures:

- among the classical interval estimators, the WA, SC and EX estimators have the coverage probabilities acceptably close to the nominal level;
- among the classical interval estimators, the Boot estimator has the coverage probabilities unacceptably further away from the nominal level;
- among the Bayesian credible estimators with the uniform prior, the Jef, HPD and EQ estimators have the coverage probabilities acceptably close to the nominal level;
- among the Bayesian credible estimators with the uniform prior, the RS estimator has the coverage probabilities unacceptably further away from the nominal level;
- among the Bayesian credible estimators with other priors, the Jef estimator has the coverage probabilities acceptably close to the nominal level;
- among the Bayesian credible estimators with other priors, the RS, EQ and HPD estimators have the coverage probabilities unacceptably further away from the nominal level;
- the Boot estimator and the EQ credible interval estimator generally underestimate the coverage probability;
- the RS credible interval estimator generally overestimates the coverage probability;
- the HPD credible interval estimator sometimes underestimates and sometimes overestimates the coverage probability.

Although these observations are limited to the ranges of λ and n specified by Figures 1 to 9 in Nadarajah *et al.* [25], they held for other values too.

4.2. Comparison based on coverage length

Here, we compare coverage lengths of the interval estimators. Figures 10 to 18 in Nadarajah *et al.* [25] show how the coverage lengths vary with respect to sample size and λ for the classical interval estimators and for the priors and

posteriors given by (3.1)–(3.10). The following observations can be drawn from the figures:

- the coverage lengths for each estimator generally increase with increasing λ;
- the coverage lengths generally decrease with increasing n except for the HPD and RS credible interval estimators;
- the coverage lengths for the HPD credible interval estimator sometime increase with n and sometimes decrease with n;
- also the coverage lengths for the RS credible interval estimator sometime increase with n and sometimes decrease with n;
- the coverage lengths appear largest for the HPD and RS credible interval estimators;
- the coverage lengths appear smallest for the WA, SC, EX, Jef, Boot and EQ estimators;
- among the Bayesian credible estimators, the coverage lengths appear largest for those with the exponential prior.

Although these observations are limited to the ranges of λ and n specified by Figures 10 to 18 in Nadarajah *et al.* [25], they held for other values too.

5. REAL DATA APPLICATIONS

Here, we present an analysis of the "Flying-bomb Hits in London During World War II" data reported by Feller [14]. The city was divided into five hundred and seventy six small areas of one-quarter square kilometers each, and the number of areas hit exactly k times was counted. There were a total of five hundred and thirty seven hits, so the average number of hits per area was 0.93. The observed frequencies in Table 2 are remarkably close to a Poisson distribution as we shall show now.

Table 2: Flying-bomb hits in London during World War II.

Hits	0	1	2	3	4	5^+
Observed	229	211	93	35	7	1

We fitted the Poisson, negative binomial and geometric distributions to the data in Table 2. The smallest chisquared statistic, the smallest Akaike informa-

tion criterion and the smallest Bayesian information criterion were obtained for the Poisson distribution. The quantile–quantile plots for the three fits shown in Figure 19 in Nadarajah *et al.* [25] show that the Poisson distribution has the points closest to the straight line.

For the fit of the Poisson distribution, $\widehat{\lambda} = 0.9288194$ with standard error 0.04015632. Using these estimates, the confidence intervals of Sections 2 and 3 can be computed. They are shown in Table 3.

Table 3: Bayesian and non-Bayesian confidence intervals for the mean number of hits.

Intervals	Lower bound	Upper bound	Upper − Lower
WA	0.85338	1.01093	0.15755
SC	0.85011	1.00752	0.15741
EX	0.85343	1.00916	0.15573
Jeffreys	0.8526	1.01006	0.15746
Bootstrap	0.84375	1.01215	0.1684
HPD.u	0.46441	1.39323	0.92882
RS.u	0.46441	1.39323	0.92882
EQ.u	0.85343	1.01097	0.15754
HPD.e	0.46441	1.39323	0.92882
RS.e	0.46441	1.39323	0.92882
EQ.e	0.85048	1.00747	0.15699
HPD.g	0.46441	1.39323	0.92882
RS.g	0.46441	1.39323	0.92882
EQ.g	0.85379	1.01107	0.15728
HPD.c	0.46441	1.39323	0.92882
RS.c	0.46441	1.39323	0.92882
EQ.c	0.85352	1.01099	0.15747

We see that the coverage length is smallest for the EX estimator, second smallest for the EQ credible interval estimator, third smallest for the SC estimator, fourth smallest for the Jef credible interval estimator, fifth smallest for the WA estimator, sixth smallest for the bootstrap estimator and the largest for the HPD and RS credible interval estimators. These observations are consistent with the results in Section 4.2.

6. CONCLUDING REMARKS

The estimation of Poisson mean is of great importance because of wide spread applications of the Poisson distribution. We have compared seventeen different interval estimators for the Poisson mean. They were compared in terms

of coverage probabilities and coverage lengths computed using simulations and a real data application. We have given various recommendations for choosing among the seventeen interval estimators. Some of them are: WA, SC and EX estimators are the best classical interval estimators in terms of coverage probabilities; Jef estimator is the best Bayesian credible interval estimator in terms of coverage probabilities; WA, SC, EX, Boot estimators are the best classical interval estimators in terms of coverage lengths; Jef and EQ estimators are the best Bayesian credible interval estimators in terms of coverage lengths.

ACKNOWLEDGMENTS

This research was partially supported by a research project grant from City University of Hong Kong (Project No. 9380058). An earlier version of this research was presented at APARM2012 in Nanjing, China. The authors would like to thank the Editor and the referee for careful reading and for their comments which greatly improved the paper.

APPENDIX: EMPIRICAL BAYES ESTIMATION

When the prior parameters are unknown, we may use another type of Bayesian estimation for estimating them without knowing or assessing the prior (subjective) distribution. The parameters of the (subjective) prior are estimated from the data. This method of estimation is called empirical Bayes (EB) estimation. For more details on EB estimation, see Maritz and Lwin [23].

In the following, we apply the EB method, in order to obtain an estimator for θ based on observed data. Suppose $X_1, X_2, ..., X_n \sim P(\lambda)$ is the observed data. Let $\lambda \sim \text{Gamma}(a, b)$ denote the prior distribution. Then, $\widehat{\lambda} = \overline{X}$. Using $S = n\widehat{\lambda} \sim P(n\lambda)$, we can write

$$f(s|\lambda) = \frac{e^{-n\lambda}(n\lambda)^s}{s!} \tag{6.1}$$

for $s = 0, 1, ...$ By using (6.1), we have

$$\begin{aligned} f(s) &= \int_0^\infty f(s|\lambda)\, g(\lambda)\, d\lambda \\ &= \int_0^\infty \frac{e^{-n\lambda}(n\lambda)^s}{s!} \frac{b^a}{\Gamma(a)}\, \lambda^{a-1} e^{-b\lambda}\, d\lambda \\ &= \frac{(s+a-1)!}{s!(a-1)!} \left(\frac{b}{b+n}\right)^a \left(\frac{n}{b+n}\right)^s, \end{aligned}$$

the probability mass function of a negative binomial random variable with parameters $p = \frac{b}{b+n}$ and $r = a$. The expectation and variance of a negative binomial random variable with parameters p and r are $\frac{rq}{p}$ and $\frac{rq}{p^2}$, respectively. Using these, the EB estimators of the prior parameters can be obtained as

- $\widehat{a} = \frac{\mu^2}{\sigma^2 - \mu}$ and $\widehat{b} = \frac{n\mu}{\sigma^2 - \mu}$ for the gamma prior;
- $\widehat{b} = \frac{n}{\mu}$ for the exponential prior;
- $\widehat{a} = \frac{\mu}{n}$ for the chisquare prior.

REFERENCES

[1] Agresti, A. and Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions, *The American Statistician*, **52**, 119–126.

[2] Barker, L. (2002). A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5, *The American Statistician*, **56**, 85–89.

[3] Baskurt, Z. and Evans, M. (2012). Hypothesis assessment and inequalities for Bayes factors and relative belief ratios, *Technical Report Number 1105*, 1 September 2011.

[4] Bégaud, B.; Martin, K.; Abouelfath, A.; Tubert-Bitter, P.; Moore, N. and Moride, Y. (2005). An easy to use method to approximate Poisson confidence limits, *European Journal of Epidemiology*, **20**, 213–216.

[5] Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition, Springer Verlag, New York.

[6] Brown, L.D.; Cai, T.T. and Das Gupta, A. (2002). Confidence interval for a binomial proportion and asymptotic expansions, *The Annals of Statistics*, **30**, 160–201.

[7] Byrne, J. and Kabaila, P. (2005). Comparison of Poisson confidence intervals, *Communication in Statistics-Theory and Methods*, **34**, 545–556.

[8] Cai, T.T. (2005). One-sided confidence intervals in discrete distributions, *Journal of Statistical Planning and Inference*, **131**, 63–88.

[9] Casella, G. and Berger, R.L. (2002). *Statistical Inference*, Duxbury, United Press, Belmont, California.

[10] Chung, K.L. (2001). *A Course in Probability Theory*, third edition, Academic Press, New York.

[11] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, New York.

[12] Evans, M. (1997). Bayesian inference procedures derived via the concept of relative surprise, *Communications in Statistics-Theory and Methods*, **26**, 1125–1143.

[13] Evans, M. and Shakhatreh, M. (2008). Optimal properties of some Bayesian inferences, *Electronic Journal of Statistics*, **2**, 1268–1280.

[14] Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*, second edition, John Wiley and Sons, New York.

[15] Feroze, N. and Aslam, M. (2012). Bayesian analysis of exponentiated gamma distribution under type II censored samples, *International Journal of Advanced Science and Technology*, **49**, 37–46.

[16] Ghosh, J.K. (1994). *Higher Order Asymptotics*, NSF-CBMS Regional Conference Series, Institute of Mathematical Statistics, Hayward, California.

[17] Good, I.J. (1988). Surprise index. In: "Encyclopaedia of Statistical Sciences" (S. Kotz, N.L. Johnson and C.B. Reid, Eds.), volume 7, John Wiley and Sons, New York.

[18] Greer, B.A.; Stamey, J.D. and Young, D.M. (2011). Bayesian interval estimation for the difference of two independent Poisson rates using data subject to under-reporting, *Statistica Neerlandica*, **65**, 259–274.

[19] Ibrahim, N.A. and Kudus, A. (2007). Bootstrap confidence interval for the median failure time of three-parameter Weibull distribution, *World Congress on Engineering*, **7**, 836–839.

[20] Krishnamoorthy, K. and Jie, P. (2011). Improved closed-form prediction intervals for binomial and Poisson distributions, *Journal of Statistical Planning and Inference*, **141**, 1709–1718.

[21] Krishnamoorthy, K.; Yanping, X. and Fang, X. (2011). A simple approximate procedure for constructing binomial and Poisson tolerance intervals, *Communication in Statistics-Theory and Methods*, **40**, 2243–2258.

[22] Li, H.Q.; Tang, M.L.; Poon, W.Y. and Tang, N.S. (2011). Confidence intervals for difference between two Poisson rates, *Communication in Statistics-Simulation and Computation*, **40**, 1478–1493.

[23] Maritz, J.L. and Lwin, T. (1989). *Empirical Bayes Methods*, second edition, Chapman and Hall, London.

[24] Michael, K. and Adam, K. (2007). Asymptotic confidence intervals for Poisson regression, *Journal of Multivariate Analysis*, **98**, 1072–1094.

[25] Nadarajah, S.; Alizadeh, M. and Bagheri, S.F. (2013). Bayesian and non-Bayesian interval estimators for the Poisson mean, *Technical Report*, School of Mathematics, University of Manchester, UK.

[26] Ng, H.K.T.; Filardo, G. and Zheng, G. (2008). Confidence interval estimating procedures for standardized incidence rates, *Computational Statistics and Data Analysis*, **52**, 3501–3516.

[27] Patil, V.V. and Kulkarni, H.V. (2012). Comparison of confidence intervals for the Poisson mean: Some new aspects, *REVSTAT*, **10**, 211–227.

[28] Riggs, K.; Young, D. and Stamey, J. (2011). Interval estimation for misclassification rate parameters in a complementary Poisson model, *Journal of Statistical Computation and Simulation*, **81**, 1145–1156.

[29] Rohatgi, V.K. and Ehsanes Saleh, A.K. (2001). *An Introduction to Probability and Statistics*, second edition, John Wiley and Sons, New York.

[30] Shao, J. (2003). *Mathematical Statistics*, second edition, Springer Verlag, New York.

[31] Stamey, J. and Hamilton, C. (2006). A note on confidence intervals for a linear function of Poisson rates, *Communications in Statistics-Simulation and Computation*, **35**, 849–856.

[32] Swift, M.B. (2009). Comparison of confidence intervals for a Poisson mean-further considerations, *Communication in Statistics-Theory and Methods*, **38**, 748–759.

REVSTAT – Statistical Journal
Volume 13, Number 3, November 2015, 263–282

GENERALIZED LEAST SQUARES AND WEIGHTED LEAST SQUARES ESTIMATION METHODS FOR DISTRIBUTIONAL PARAMETERS

Author: Yeliz Mert Kantar
– Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, Turkey
ymert@anadolu.edu.tr

Received: March 2014 Revised: August 2014 Accepted: August 2014

Abstract:

- Regression procedures are often used for estimating distributional parameters because of their computational simplicity and useful graphical presentation. However, the resulting regression model may have heteroscedasticity and/or correction problems and thus, weighted least squares estimation or alternative estimation methods should be used. In this study, we consider generalized least squares and weighted least squares estimation methods, based on an easily calculated approximation of the covariance matrix, for distributional parameters. The considered estimation methods are then applied to the estimation of parameters of different distributions, such as Weibull, log-logistic and Pareto. The results of the Monte Carlo simulation show that the generalized least squares method for the shape parameter of the considered distributions provides for most cases better performance than the maximum likelihood, least-squares and some alternative estimation methods. Certain real life examples are provided to further demonstrate the performance of the considered generalized least squares estimation method.

Key-Words:

- *probability plot; heteroscedasticity; autocorrelation; generalized least squares; weighted least squares; shape parameter.*

1. INTRODUCTION

Regression procedures are often used for estimating distributional parameters. In this procedure, the distribution function is transformed to a linear regression model. Thus, least squares (LS) estimation and other regression estimation methods can be employed to estimate parameters of a specified distribution. In the literature, the parameters of the Pareto and Weibull distributions, in particular, have been estimated by such methods, since these distributions have been commonly used in reliability and survival analysis as well as engineering (Abernethy, 1996; Boudt *et al.* 2011; Kantar and Usta, 2008; Genschel and Meeker, 2010; Hung, 2001; Lu *et al.* 2004; Baxter 1980). In addition to the Pareto and Weibull distributions, Burr-type, Gumbel, logistic and log-logistic distributions have been studied by regression estimation methods (Bergman, 1986; Hossain and Howlader, 1996; Wang and Cheng, 2010; Zhang *et al.* 2007; Zhang *et al.* 2008; Zyl, 2012; Zyl and Schall, 2012; Usta, 2013; Kantar and Arik, 2014; Kantar and Yildirim, 2015). One of the main advantages of using regression procedures for estimating parameters is that their implementation is simple in the case of complete data, censoring data or data with outliers. Nevertheless, as is well known, the resulting regression model may have unequal variance (heteroscedastic) (Engeman and Keefe, 1982; Boudt *et al.* 2011; Zyl, 2012; Zyl and Schall, 2012) and/or correction problems (Engeman and Keefe, 1982) and thus, the weighted least squares (WLS) estimation or alternative methods should be used (Engeman and Keefe, 1982; Lu and Tao, 2007; Zyl, 2012; Zyl and Schall, 2012). For example, Engeman and Keefe (1982) consider generalized least squares estimation of the Weibull distribution by means of a linear regression model. Hung (2001), Lu *et al.* (2004), Zyl and Schall (2012) emphasize that a weight function should be used when performing regression methods, and propose different weights using large sample properties of the empirical distribution function or order statistics, to stabilize the variance in order to perform the WLS estimation method for the Weibull distribution. Zhang *et al.* (2008) discuss alternative WLS estimation methods for the Weibull distribution. On the other hand, Malik (1970) studied the LS method, ridge regression and maximum product of spacing methods to estimate parameters for the Pareto distribution, while Zyl (2012) considered the Laplace distributed errors (LAD) (Koenker and Bassett, 1978) and Box–Cox regression to stabilize variance. It can be seen that Zyl's (2012) WLS and Lu and Tao' (2007) WLS perform almost as well as the maximum likelihood (ML) estimation method for the Pareto distribution.

In this article, we consider generalized least squares (GLS) and WLS estimation methods for distributional parameters by easily calculating an approximation of the variance-covariance matrix. GLS and WLS are then applied to the estimation of the parameters of the Weibull, Pareto and log-logistic distributions. The simulation results show that the proposed estimation methods, particularly

GLS for the shape parameter of the considered distributions, provide better performance than the ML, LS and some alternative WLS estimation methods, for most of the considered sample sizes.

The rest of this paper is organized as follows: Section 2 provides the process of estimation of distributional parameters via regression models for the Weibull, Pareto and log-logistic distributions. Section 3 introduces the GLS and WLS estimation methods and their application to each of these distributions. Alternative estimation methods for the Weibull, Pareto or log-logistic distributions are briefly discussed in Section 4. To show the performance of the considered GLS and WLS methods, a simulation study is presented in Section 5. A number of real-data examples are discussed in Section 6 and, finally, the last section summarizes the conclusions of the study.

2. ESTIMATION OF DISTRIBUTIONAL PARAMETERS VIA REGRESSION MODELS

Probability plots that use the quantile function of the random variable are often used for different objectives, such as, (i) to draw conclusions from data, (ii) to estimate the parameters of the considered distribution, (iii) to apply to both complete and censored data and (iv) to show graphical presentation. (For other advantages, see Nelson, 2004)

By taking into account the objective (ii) of probability plots, the distribution function is transformed into a linear regression model, so that various regression estimation methods can then be used to estimate the parameters of the specified distribution.

The probability density function (pdf) and cumulative distribution function (cdf) of the Weibull random variable are respectively given in the following equations:

$$f(x,\lambda,\alpha) = \frac{\alpha}{\lambda}\left(\frac{x}{\lambda}\right)^{\alpha-1} e^{-(\frac{x}{\lambda})^{\alpha}}, \qquad \text{for} \quad x>0\ , \tag{2.1}$$

$$F(x,\lambda,\alpha) = 1-e^{-(\frac{x}{\lambda})^{\alpha}}, \qquad \text{for} \quad x>0\ , \tag{2.2}$$

where λ is the scale parameter and α is the shape parameter. The Weibull distribution is a reversed J-shaped, bell shaped and exponential distribution for $\alpha<1$, $\alpha>1$ and $\alpha=1$, respectively. The Weibull distribution appears similar to a normal distribution for $\alpha=3.4$ (Kantar and Senoglu, 2008).

After some algebraic manipulation, equation (2.2) can be expressed as follows:

$$\ln\big[-\ln(1-F(x))\big] = \alpha \ln x - \alpha \ln \lambda \ , \tag{2.3}$$

$$\ln x = \ln \lambda + \frac{1}{\alpha} \ln\big[-\ln(1-F(x))\big] \ . \tag{2.4}$$

For a sample size of n and $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ the regression model is rewritten as:

$$\ln x_{(i)} = \ln \lambda + \frac{1}{\alpha} \ln\big[-\ln(1-F(x_{(i)}))\big] \ , \tag{2.5}$$

where $\ln x_{(i)}$ is the ith order statistics of the logarithm of the sample from the Weibull distribution. $\frac{i-a}{(n+b)}, (0 \leq a \leq 0.5, 0 \leq b \leq 1)$ is used as estimate of $F(x_{(i)})$ where i is the rank of the data point in the sample in ascending order. For complete samples, $\frac{i}{(n+1)}$ and $\frac{i-0.3}{(n+0.4)}$ are generally used (Tiryakioglu and Hudak, 2007; Zyl, 2012, Kantar and Yildirim, 2015).

If we replace $\ln x_{(i)}$ with Y_i, $\ln \lambda$ with β_0, $\frac{1}{\alpha}$ with β_1 and $\ln[-\ln(1-F(x_{(i)}))]$ with X_i, the regression model (2.5) occurs as:

$$Y_i = \beta_0 + \beta_1 X_i \ . \tag{2.6}$$

By using the regression model given in (2.5), the LS and other regression estimation methods can be easily employed to estimate the parameters of the Weibull distribution.

The Pareto cdf is given as follows:

$$F(x,k,\eta) = 1 - \left(\frac{k}{x}\right)^{\eta}, \tag{2.7}$$

where k is the scale parameter and η is the shape parameter. The Pareto distribution, which is generally used to model extreme values, is skewed and heavy-tailed.

After algebraic manipulation, equation (2.7) can be expressed as follows:

$$\ln x = \ln k - \frac{1}{\eta} \ln(1-F(x)) \ . \tag{2.8}$$

For the ordered sample, the regression model for the Pareto distribution is rewritten as:

$$\ln x_{(i)} = \ln k - \frac{1}{\eta} \ln(1-F(x_{(i)})) \ . \tag{2.9}$$

If we replace $\ln x_{(i)}$ with Y_i, $\ln k$ with β_0, $-\frac{1}{\eta}$ with β_1 and $\ln(1-F(x_{(i)}))$ with X_i, the linear regression model is obtained for the Pareto distribution.

The cdf of the log-logistic random variable is given as follows:

$$F(x) = 1 - \left(1 + \left(\frac{x}{\gamma}\right)^{\delta}\right)^{-1}, \qquad x \geq 0\,, \quad \delta, \gamma > 0\,, \tag{2.10}$$

where γ is the scale parameter and δ is the shape parameter. For $\delta > 1$, the log-logistic distribution is unimodal and its variance decreases as δ increases. The log-logistic distribution has been widely-used in hydrology to model stream flow (Ahmad *et al.* 1988; Ashkar and Mahdi, 2006; Chen, 2006).

Similar to the Weibull and Pareto distributions, the obtained regression model for the log-logistic distribution is presented as follows:

$$\ln(x) \;=\; \frac{1}{\delta}\ln\big((1 - F(x))^{-1} - 1\big) + \ln(\gamma)\,, \tag{2.11}$$

which may be written as:

$$\ln(x_{(i)}) \;=\; \frac{1}{\delta}\ln\big((1 - F(x_{(i)}))^{-1} - 1\big) + \ln(\gamma)\,. \tag{2.12}$$

In conclusion, the parameters of the Weibull, Pareto and log-logistic distributions can be estimated respectively using ordinary LS estimation for equations (2.5), (2.9) and (2.12).

It should be noted that the common LS and other regression estimation procedures applied in the literature for the Weibull and Pareto distributions use the least squares regression of X on Y, $(X_i = \beta_0 + \beta_1 Y_i)$ (Genschel and Meeker, 2010; Hossain and Howlader, 1996; Hung, 2001; Kantar and Arik, 2014; Kantar and Yildirim, 2015; Lu *et al.* 2004; Wang and Cheng, 2010; Zhang *et al.* 2008; Zyl, 2012; Zyl and Schall, 2012). To the best of our knowledge, only Zhang *et al.* (2007) compare these two LS estimation methods for the Weibull using intensive Monte Carlo simulations, finding that LS of Y on X provides better estimators than LS of X on Y.

3. GENERALIZED LEAST SQUARES ESTIMATION AND WEIGHTED LEAST SQUARES ESTIMATION METHODS FOR DISTRIBUTIONAL PARAMETERS: THE CASES OF WEIBULL, PARETO AND LOG-LOGISTIC DISTRIBUTIONS

The most obvious point to be noticed is that since the sample is ordered in the models (2.5), (2.9) and (2.12), $\ln x_{(i)}$ is also ordered. For this reason, the covariance matrices of the dependent variable of these models are not in the form $\sigma^2\mathbf{I}$, but of $\sigma^2\mathbf{V} = \sum$, where σ^2 is unknown and $\mathbf{V}$ is known (White, 1969; Engeman and Keefe, 1982). In this case, the LS estimates of the coefficients may

not have minimum variance. In such cases, alternative estimation approaches to stabilize variances can be used.

Generalized least squares (GLS) estimation is an efficient method for estimating the unknown coefficients of a linear regression model when the observations have unequal variance and there is a certain degree of correlation between the observations. In the linear regression model given in (2.6), if the form of the variance of $\mathbf{Y} = (Y_1, ..., Y_n)$ is $\sigma^2\mathbf{V} = \sum$, GLS minimizes

$$(\mathbf{Y} - \mathbf{X}\beta)' \, \mathbf{V^{-1}} (\mathbf{Y} - \mathbf{X}\beta) \ , \tag{3.1}$$

which is solved by

$$\hat{\beta}_{GLS} = (\mathbf{X'V^{-1}X}) \, \mathbf{X'V^{-1}Y} \ , \tag{3.2}$$

where $\hat{\beta}_{\mathbf{GLS}}$ is the vector of the GLS estimates of $\beta = (\beta_1, \beta_2)$ and $\mathbf{X}$ is the matrix of ones and x_i. In addition, the GLS estimates are equivalent to applying ordinary LS to a linearly transformed form of the data. That is, we can write $\mathbf{V} = \mathbf{SS'}$, where $\mathbf{S}$ is a triangular matrix, using Cholesky decomposition. The LS estimates obtained by regressing $\mathbf{S^{-1}Y}$ on $\mathbf{S^{-1}X}$ are equal to the GLS estimates. Thus, $\mathbf{Var(S^{-1}Y)} = \mathbf{S^{-1}Var(Y)(S^{-1})'} = \sigma^2\mathbf{S^{-1}V(S^{-1})'} = \sigma^2\mathbf{I}$. The transformed form of the data is uncorrelated, with constant variance.

On occasion, the observations are uncorrelated or have a small enough correlation to be ignored, but have unequal variance. That is, the covariance-matrix is diagonal, say $\mathbf{W}$, but does not have equal diagonal elements. WLS estimation can be used in this situation. WLS estimate is obtained as follows:

$$\hat{\beta}_{WLS} = (\mathbf{X'W^{-1}X}) \, \mathbf{X'W^{-1}Y} \ . \tag{3.3}$$

Now, the problem is to estimate the $\mathbf{V}$ matrix for the considered distributions. Taking into account equation (2.5) for the Weibull distribution, it is noted that the cumulative function F is transformed into $\ln[-\ln(1 - F(x))]$ and the random variable X is transformed into $\ln X$. If these transformations are respectively denoted by TF and TX, the regression model for the Weibull distribution given in (2.5) can be expressed as follows:

$$TX(X_i) = \beta_0 + \beta_1 TF(F_i) \ , \tag{3.4}$$

where $\beta_0 = \ln\lambda$ and $\beta_1 = \frac{1}{\alpha}$. Taking the expectation, variance and covariance of both sides yields:

$$E(TX(X_i)) = \beta_0 + \beta_1 E(TF(F_i)) \ , \tag{3.5}$$

$$Var(TX(X_i)) = \beta_1^2 Var(TF(F_i)) \ , \tag{3.6}$$

$$Cov(TX(X_i), TX(X_j)) = \beta_1^2 Cov(TF(F_i), TF(F_j)) \ . \tag{3.7}$$

By using Taylor expansion, it is possible to approximate the expectation of the observation and also variance and covariance between the observations. Taylor series expansion of $TF(F_i)$ about the value $F0_i = \frac{i}{n+1}$, $i = 1, 2, ..., n$, is given by:

$$TF(F_i) \approx TF(F0_i) + TF'(F0_i)(F - F0_i) + \frac{1}{2}TF''(F0_i)(F - F0_i)^2 \ . \tag{3.8}$$

Taking the expectation of both sides yields:

$$E(TF(F_i)) \approx E(TF(F0_i)) + \frac{1}{2}TF''(F0_i)E((F - F0_i)^2) \ . \tag{3.9}$$

Similarly, taking the variance of both sides yields:

$$Var(TF(F_i)) \approx (TF'(F0_i))^2 Var(F) = (TF'(F0_i))^2 \frac{F0_i(1 - F0_i)}{n + 2} \tag{3.10}$$

A similar closed-form approximate formula for the covariance between $TF(F_i)$ and $TF(F_j)$ is

$$Cov(TF(F_i), TF(F_j)) \approx TF'(F0_i)TF'(F0_j)\frac{F0_i(1 - F0_j)}{n + 2}, \quad i < j \ . \tag{3.11}$$

(See Blom, 1962; White, 1969; Engeman and Keefe, 1982). Thus,

$$E(TX(X_i)) \ \approx \ \beta_0 + \beta_1(TF(F0_i)) + \frac{1}{2}TF''(F0_i)\frac{F0_i(1 - F0_j)}{n + 2} \ , \tag{3.12}$$

$$Var(TX(X_i)) \ \approx \ \beta_1^2(TF'(F0_i))^2 \frac{F0_i(1 - F0_i)}{n + 2} \ , \tag{3.13}$$

$$Cov(TX(X_i), TX(X_j)) \ \approx \ \beta_1^2 TF'(F0_i)TF'(F0_j)\frac{F0_i(1 - F0_j)}{n + 2}, \quad i < j \ , \tag{3.14}$$

where $TF(F) = \ln(-\ln(1 - F))$, $TF'(F) = \frac{-1}{(1-F)\ln(1-F)}$ and $TF''(F) = \frac{\ln(1-F)+1}{((1-F)\ln(1-F))^2}$

(3.15)

$$Var(TX(X_i)) = Cov(TX(X_i), TX(X_j)) \approx \frac{\beta_1^2}{n + 2}\frac{i}{(n + 1 - i)}\frac{1}{\ln(\frac{n+1-i}{n+1})}\frac{1}{\ln(\frac{n+1-j}{n+1})} \ .$$

Thus, considering $\frac{\beta_1^2}{n+2}$ as $\sigma^2\mathbf{V}$ as in Engeman and Keefe (1982), the approximate formula for the $\mathbf{V}$ matrix can be expressed as follows:

$$v_{ij} = \frac{i}{(n + 1 - i)} \frac{1}{\ln(\frac{n+1-i}{n+1})} \frac{1}{\ln(\frac{n+1-j}{n+1})} \ , \qquad i \le j \ , \tag{3.16}$$

where v_{ij} is an element of the $\mathbf{V}$ matrix. Thus, in order to apply GLS and WLS estimation methods for the Weibull distribution, the covariance matrix and the matrix of the diagonal elements of the covariance matrix are expressed respectively as follows:

$$\mathbf{V} = \begin{pmatrix} v_{11} & ... & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & ... & v_{nn} \end{pmatrix} \ , \tag{3.17}$$

$$\mathbf{W} = \begin{pmatrix} v_{11} & ... & 0 \\ \vdots & \ddots & \vdots \\ 0 & ... & v_{nn} \end{pmatrix} \ . \tag{3.18}$$

For the Weibull distribution:

$$\hat{\beta}_{GLS1} = (\mathbf{X'V^{-1}X})^{-1}\mathbf{X'V^{-1}Y} ,$$
$$\hat{\beta}_{WLS} = (\mathbf{X'W^{-1}X})^{-1}\mathbf{X'W^{-1}Y} ,$$

where $\mathbf{Y} = [\ln(x_{(1)}), ..., \ln(x_{(n)})]$, $\mathbf{X} = \begin{pmatrix} 1 & \ln(-\ln(1-\hat{F}_1)) \\ \vdots & \vdots \\ 1 & \ln(-\ln(1-\hat{F}_n)) \end{pmatrix}$, $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$, where $\hat{\beta}_0 = \ln\hat{\lambda}$ and $\hat{\beta}_1 = \frac{1}{\hat{\alpha}}$, $\hat{\lambda} = \exp(\hat{\beta}_0)$ and $\hat{\alpha} = \frac{1}{\hat{\beta}_1}$.

It should be noted that independent variables in equations (2.5), (2.9) and (2.12) are unobserved, but are estimated differently from classical regression analysis. In order to estimate regression coefficients, we replaced the independent variable with its estimate.

Considering the expected value of $E(TF(F_i))$ given in (3.9) and (3.12), we can choose the independent variable as $TF(F0_i) + \frac{1}{2}TF''(F0_i)Var(F)$ to reduce bias in the GLS procedure. Nevertheless, unbiasedness of the resulting estimator is not claimed, since $TX(X_i)$, as the dependent variable, has an estimate of mean. Consequently, the considered generalized estimation procedure here is denoted by GLS2 for the purpose of distinction.

Thereby, the design matrix of the regression model for the Weibull distribution is given as follows:

$$\mathbf{Z} = \begin{pmatrix} 1 & \ln(-\ln(1-\hat{F}_1)) - 0.5 - \frac{\ln(1-\hat{F}_1)+1}{((1-\hat{F}_1)\ln(1-\hat{F}_1))^2} \\ \vdots & \vdots \\ 1 & \ln(-\ln(1-\hat{F}_n)) - 0.5 - \frac{\ln(1-\hat{F}_n)+1}{((1-\hat{F}_n)\ln(1-\hat{F}_n))^2} \end{pmatrix} . \tag{3.19}$$

Consequently,

$$\hat{\beta}_{GLS2} = (\mathbf{Z'V^{-1}Z})^{-1}\mathbf{Z'V^{-1}Y} .$$

Taken into account the model given in (2.9) for the Pareto distribution, $TF(F) = \ln(1-F)$, $TF'(F) = \frac{-1}{1-F}$ and $TF''(F) = \frac{-1}{(1-F)^2}$

The approximate formula for the $\mathbf{V}$ matrix for the Pareto distribution can be expressed as follows:

$$v_{ij} = \frac{1}{(n+1-i)} , \qquad i \leq j . \tag{3.20}$$

Similarly, the $\mathbf{V}$ matrix for the log-logistic distribution is:

$$v_{ij} = \frac{(n+1)^2}{i(n+1-i)} , \qquad i \leq j . \tag{3.21}$$

where $TF(F) = \ln(\frac{F}{1-F})$, $TF'(F) = \frac{1}{1-F}$ and $TF''(F) = \frac{2F-1}{(F(1-F))^2}$.

Similar to the Weibull distribution, WSL, GLS1 and GLS2 estimation methods can be applied to estimate the parameters of the Pareto and log-logistic distributions using the approximate covariance matrices. It should also be highlighted that the obtained covariance matrix and the proposed GLS1 estimation method coincide with GLS for the Weibull studied in (Engeman and Keefe, 1982). However, Engeman and Keefe (1982)

compare GLS estimation with ML for sample size $n = 25$ and find that GLS for the shape parameter of the Weibull distribution performs better than ML estimation. In this study, we compare GLS1 for the Weibull distribution with existing alternative WLS estimation methods for different sample sizes and shape parameter cases.

In conclusion, the considered WLS, GLS1 and GLS2 estimation methods are based on explicit functions of the sample observations and are therefore easy to compute, without the typical computational complexity of ML (Kantar and Senoglu, 2008; Gebizlioglu *et al.* 2011). Also, the standard error of the WLS, GLS1 and GLS2 estimates are easily calculated taking the square roots of the diagonal elements of $(\mathbf{X'W^{-1}X})^{-1}\sigma^2$, $(\mathbf{X'V^{-1}X})^{-1}\sigma^2$ and $(\mathbf{Z'V^{-1}Z})^{-1}\sigma^2$ respectively.

4. DISCUSSION OF OTHER ALTERNATIVE ESTIMATION METHODS

Many estimators have been proposed in the literature for the parameters of the Weibull distribution, and these estimators have been compared according to different criteria (Bergman, 1986; Gebizlioglu *et al.* 2011; Hassanein 1971; Hossain and Zimmer 2003; Hossain and Howlader, 1996; Hung, 2001; Marks, 2005; Kantar and Senoglu, 2008; Kantar and Usta, 2008; Zhang *et al.* 2008; Prakash and Singh, 2009; Zyl and Schall, 2012). The ML estimator, generally preferred due to its good theoretical properties for large sample sizes ($n > 100$), (Kantar and Senoglu, 2008), may have poor small sample performance (Kantar and Senoglu, 2008; Marks, 2005; Teimouri and S. Nadarajah, 2012). Moreover, ML requires an iterative numerical method for most distributions, such as Newton–Raphson. Among other alternative estimators, the most popular is LS estimation because of its computational simplicity in the case of complete data, censoring data and data with outliers (Genschel and Meeker, 2010; Hossain and Zimmer, 2003; Hung, 2001; Lu *et al.* 2004; Zhang *et al.* 2007). However, it is known that LS estimates for distributional parameters may give misleading inferences since there is a heteroscedasticity or correlation problem. With this in mind, GLS (Engeman and Keefe, 1982) and WLS are proposed in (Hung, 2001; Lu *et al.* 2004; Zyl and Schall, 2012), demonstrating that WLS totally outperforms the LS method. Also, Zyl and Schall's (2012) WLS performs almost as well as the ML estimation.

The Pareto distribution has been widely-studied in the literature (Quandt, 1966; Saksena and Johnson, 1984; Likes, 1969; Baxter, 1980). Since ML of its shape parameter is biased (Baxter, 1980; Saksena and Johnson, 1984), Malik (1970) studied the LS, ridge regression and maximum product of spacing methods to estimate the parameters for the Pareto distribution. Hossain and Zimmer (2000) consider LS for the Pareto distribution, showing the superiority of LS estimation over ML estimation. In addition, Lu and Tao (2007) provide a new weighting function for WLS, in order to estimate the parameters of the Pareto distribution. They show that their WLS method demonstrates better performance than classical LS estimation for the Pareto distribution.

On the other hand, the ML, LS, moment, generalized moment and probability weighted moment estimators are considered as estimation methods for log-logistic distribution in the literature (Ashkar and Mahdi, 2006; Chen, 2006; Kantar and Arik, 2014; Rao and Kantam, 2012). The moment estimator for the log-logistic distribution is not widely-used due to constraints $\delta > 1$ and $\delta > 2$. Among those mentioned, ML is the most-preferred estimation method.

5. MONTE CARLO SIMULATION

This section presents Monte Carlo simulation carried out to compare the performance of the proposed GLS1, GLS2 and WLS in comparison with ML, LS estimation methods, and also certain existing WLS estimation methods for the parameters of Weibull, Pareto and Log-logistic distributions.

Bias and RMSE for parameters are calculated using 20, 000 simulated samples. All computations for the simulation are performed using MATLAB 10.1. We consider sample sizes $n = 10, 20, 30, 50, 100$ and 250. While shape parameters are taken as $0.5, 1, 2, 4$ for the Pareto distribution, for log-logistic and Weibull distributions, shape parameters are taken as $1, 2, 3$ and 6 in common with previous studies. Also, without any loss of generality, the scale parameter is taken to be equal to 1.

Table 1 shows the RMSE and bias values for ML, LS estimation for regression of X on Y (LS1), known as classical LS in the literature, LS estimation for regression of Y on X (LS2), which is considered in this study, WLS (Zyl&Schall), WLS (Hung) and WLS (Lu *et al.*) and the considered WLS, GLS1 and GLS2 in this study, for the shape parameter of the Weibull distribution. From the simulation results presented in Table 1, the following conclusions may be summarized:

According to the RMSE criterion:

(**a**) GLS1 and GLS2 apparently show better performance than others for most considered sample sizes.

(**b**) While GLS1 provides less RMSE than others for $n = 10$, GLS1 and GLS2 show similar and best performance for $n = 20, 30, 50, 100$.

(**c**) WLS estimation performs better than LS1 for all sample sizes and shape parameter cases.

(**d**) LS2 performs better than LS1 for all considered sample sizes except $n = 10$. The same result is observed in the study of Zhang *et al.*, 2007.

According to bias criterion:

(**a**) GLS2 is clearly the best estimator in terms of bias. Particularly, a superior performance of GLS2 is observed for $n \geq 20$.

(**b**) Next to GLS2, LS2, GLS1 and WLS provide similar good performance for $n = 10, 20$ with ML being best for other $n = 50, 100, 250$.

From the simulation results presented in Table 2 for the shape parameter of the Pareto distribution, the following conclusions may be summarized:

According to RMSE:

(**a**) While the proposed GLS1 estimation shows better performance than ML, LS1, LS2, WLS (Lu&Tao) and ELS (Lu&Tao) for most of the considered sample sizes, GLS2 is the best for $n = 250$. Also, the considered WLS is the best performer next to GLS1 for $n = 10, 20, 30, 100$ with GLS2 presenting the second best performance after GLS1 for $n = 50$.

(**b**) The considered WLS estimation methods and WLS (Lu and Tao) show similar performance.

(**c**) ML shows the worst performance compared with others, in terms of RMSE for $n = 10$. This result matches the study of Lu and Tao (2007).

(**d**) LS2 has smaller RMSE than LS1 for all the considered shape parameter cases and sample sizes except $n = 10$.

Table 1: RMSE and Bias of the estimated shape parameters of the Weibull distribution.

Methods	Shape							
	1		2		3		6	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
	n=10							
ML	0.38863	−0.16511	0.76759	−0.33705	1.13419	−0.51566	2.31548	−0.98691
LS1	0.31833	0.13005	0.63634	0.26033	0.95289	0.39420	1.90491	0.78612
LS2	0.31990	0.05931	0.63967	0.11783	0.95616	0.18046	1.91233	0.36114
WLS (Z&S)	0.29821	0.13978	0.58720	0.27390	0.88842	0.40626	1.78828	0.84593
WLS (Hung)	0.30320	0.13897	0.60026	0.27159	0.90529	0.40397	1.81904	0.83832
WLS (Lu)	0.29903	0.14035	0.58920	0.27497	0.89115	0.40837	1.79431	0.84926
WLS	0.31800	0.06117	0.63271	0.11755	0.95065	0.17788	1.92669	0.37458
GLS1	0.29000	0.06362	0.56862	0.12068	0.86533	0.17573	1.73207	0.38308
GLS2	0.35385	−0.06119	0.56109	−0.04698	0.89027	−0.16520	1.96762	−0.38306
	n=20							
ML	0.22154	−0.07498	0.43758	−0.14731	0.64809	−0.21834	1.33219	−0.44944
LS1	0.23171	0.10663	0.46301	0.21645	0.69459	0.31827	1.38990	0.63916
LS2	0.21958	0.05772	0.43834	0.11909	0.65791	0.17143	1.31574	0.34487
WLS (Z&S)	0.20379	0.09617	0.40523	0.19618	0.60474	0.29597	1.22168	0.57716
WLS (Hung)	0.21225	0.08519	0.41520	0.17664	0.63069	0.26699	1.26565	0.50690
WLS (Lu)	0.20472	0.09610	0.40707	0.19616	0.60777	0.29605	1.22675	0.57687
WLS	0.21970	0.05956	0.43600	0.12022	0.65285	0.18496	1.31841	0.36573
GLS1	0.19204	0.05495	0.37992	0.11169	0.56485	0.17069	1.15498	0.33209
GLS2	0.19931	−0.02895	0.41816	−0.07534	0.55577	−0.05771	1.13999	−0.07700
	n=30							
ML	0.16694	−0.04893	0.33061	−0.09807	0.50327	−0.14007	0.99289	−0.27909
LS1	0.19433	0.09079	0.38810	0.18201	0.58324	0.27222	1.16341	0.53894
LS2	0.18047	0.05201	0.36069	0.10431	0.54226	0.15547	1.08346	0.30588
WLS (Z&S)	0.16227	0.07152	0.32184	0.14273	0.49135	0.21621	0.97494	0.44393
WLS (Hung)	0.17208	0.05960	0.34207	0.11918	0.52139	0.18016	1.03151	0.37621
WLS (Lu)	0.16303	0.07120	0.32345	0.14211	0.49372	0.21527	0.97947	0.44253
WLS	0.18075	0.05161	0.35854	0.10327	0.54220	0.15931	1.07667	0.33157
GLS1	0.15281	0.04580	0.30270	0.09133	0.46125	0.14013	0.91585	0.28804
GLS2	0.15024	−0.02217	0.30412	−0.01153	0.44822	−0.04302	0.90238	−0.02257
	n=50							
ML	0.11983	−0.02859	0.24544	−0.06272	0.36191	−0.08748	0.72335	−0.16474
LS1	0.15472	0.07106	0.30869	0.14283	0.46362	0.21366	0.92958	0.42794
LS2	0.14211	0.04228	0.28304	0.08543	0.42558	0.12742	0.85239	0.25580
WLS (Z&S)	0.12252	0.04702	0.24491	0.08928	0.36714	0.14028	0.74707	0.28503
WLS (Hung)	0.13334	0.03707	0.26475	0.07007	0.39962	0.10809	0.81009	0.22344
WLS (Lu)	0.12317	0.04661	0.24597	0.08849	0.36893	0.13909	0.75062	0.28238
WLS	0.14114	0.04233	0.28107	0.07777	0.42417	0.12536	0.85600	0.25945
GLS1	0.11502	0.03513	0.23203	0.06517	0.34612	0.10363	0.70692	0.21860
GLS2	0.11894	−0.01230	0.23650	−0.02100	0.34015	−0.03437	0.68911	−0.06076
	n=100							
ML	0.08173	−0.01503	0.16438	−0.02775	0.24102	−0.04263	0.47159	−0.07239
LS1	0.11206	0.04903	0.22417	0.09773	0.33142	0.14372	0.66981	0.29164
LS2	0.10253	0.03039	0.20509	0.06024	0.30433	0.08797	0.61356	0.18014
WLS (Z&S)	0.08532	0.02365	0.17234	0.05307	0.25084	0.07527	0.50474	0.16409
WLS (Hung)	0.09497	0.01702	0.18984	0.04145	0.27751	0.05683	0.56174	0.12774
WLS (Lu)	0.08567	0.02331	0.17300	0.05260	0.25179	0.07440	0.50692	0.16229
WLS	0.10173	0.02789	0.20430	0.06213	0.30194	0.09223	0.60435	0.18702
GLS1	0.08068	0.02223	0.16310	0.04667	0.23956	0.06941	0.47439	0.14998
GLS2	0.08112	−0.00479	0.15633	−0.01075	0.23748	−0.01722	0.47863	−0.03175
	n=250							
ML	0.05067	−0.00450	0.10093	−0.01063	0.15032	−0.01562	0.30313	−0.02655
LS1	0.07090	0.02789	0.14099	0.05622	0.21292	0.08332	0.42111	0.16382
LS2	0.06573	0.01797	0.13096	0.03675	0.19744	0.05341	0.39137	0.10486
WLS (Z&S)	0.05505	0.01139	0.10954	0.02045	0.16313	0.03184	0.32897	0.07036
WLS (Hung)	0.06107	0.00848	0.12296	0.01354	0.18259	0.02266	0.36949	0.05389
WLS (Lu)	0.05514	0.01126	0.10982	0.02018	0.16352	0.03146	0.32979	0.06964
WLS	0.06745	0.01893	0.13191	0.03565	0.19723	0.05478	0.39537	0.11392
GLS1	0.05235	0.01343	0.10281	0.02536	0.15375	0.03858	0.30788	0.08173
GLS2	0.04909	−0.00189	0.09938	−0.00330	0.14755	−0.00684	0.30200	−0.00780

Table 2: RMSE and Bias of the estimated shape parameters of the Pareto distribution.

Methods	Shape							
	0.5		1		2		4	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
	n=10							
ML	0.26682	−0.12505	0.53541	−0.25050	1.06415	−0.49718	2.13514	−1.00263
LS1	0.20968	0.04864	0.41600	0.09824	0.83662	0.19479	1.66619	0.39608
LS2	0.22352	0.01044	0.44241	0.02185	0.89118	0.04120	1.77565	0.08861
ELS	0.24558	−0.04463	0.49464	−0.09034	0.98164	−0.17592	1.96201	−0.36301
WLS (L&T)	0.19586	−0.00777	0.39277	−0.01558	0.78208	−0.02760	1.56403	−0.06245
WLS	0.19118	−0.00860	0.38300	−0.01950	0.76350	−0.00124	1.52641	−0.00795
GLS1	0.18815	0.00937	0.37398	0.01637	0.74324	0.03592	1.48892	0.06459
GLS2	0.21636	−0.04981	0.43160	−0.09986	0.86333	−0.20072	1.72788	−0.39918
	n=20							
ML	0.14642	−0.05530	0.29084	−0.11156	0.58265	−0.22108	1.16502	−0.44376
LS1	0.14772	0.05070	0.29604	0.09971	0.59634	0.19960	1.18038	0.39813
LS2	0.14621	0.02396	0.29318	0.04620	0.59092	0.09233	1.16898	0.18253
ELS	0.15748	−0.01357	0.31413	−0.02760	0.63213	−0.05491	1.25870	−0.11178
WLS (L&T)	0.12506	0.00255	0.24789	0.00402	0.49643	0.01074	0.99388	0.01982
WLS	0.12426	0.00388	0.24623	0.00693	0.49396	0.01578	0.98857	0.03018
GLS1	0.11975	0.01462	0.23707	0.02849	0.47619	0.05864	0.95094	0.11581
GLS2	0.12905	−0.02108	0.25999	−0.04308	0.51875	−0.09169	1.03973	−0.18273
	n=30							
ML	0.10913	−0.03567	0.21762	−0.07012	0.43559	−0.14240	0.87266	−0.28254
LS1	0.12375	0.04583	0.24788	0.09080	0.49645	0.18579	0.99177	0.36783
LS2	0.11925	0.02400	0.23871	0.04717	0.47713	0.09810	0.95442	0.19228
ELS	0.12570	−0.00582	0.25086	−0.00965	0.50424	−0.02141	1.01027	−0.04358
WLS (L&T)	0.09935	0.00262	0.19872	0.00624	0.39603	0.01096	0.79497	0.02404
WLS	0.09900	0.00322	0.19817	0.00742	0.39477	0.01342	0.79265	0.02855
GLS1	0.09482	0.01333	0.18955	0.0278	0.37870	0.05361	0.75963	0.10923
GLS2	0.10082	−0.01303	0.20118	−0.02577	0.40632	−0.05303	0.80194	−0.11129
	n=50							
ML	0.07906	−0.02046	0.15777	−0.04141	0.31351	−0.08160	0.62915	−0.16547
LS1	0.09967	0.03844	0.19854	0.07558	0.39837	0.15476	0.79066	0.30625
LS2	0.09372	0.02144	0.18723	0.04185	0.37493	0.08732	0.74424	0.17248
ELS	0.09798	0.00172	0.19465	−0.00282	0.38897	−0.00561	0.77529	−0.00025
WLS (L&T)	0.07629	0.00218	0.15245	0.00409	0.30287	0.00911	0.60790	0.01549
WLS	0.07616	0.00236	0.15212	0.00452	0.30241	0.00988	0.60706	0.01720
GLS1	0.07273	0.01106	0.14481	0.02164	0.28832	0.04439	0.57757	0.08671
GLS2	0.07446	−0.00845	0.15212	−0.01880	0.30202	−0.03505	0.60540	−0.06028
	n=100							
ML	0.05310	−0.01009	0.10578	−0.01982	0.21128	−0.04089	0.42800	−0.08124
LS1	0.07279	0.02822	0.14520	0.05607	0.29094	0.11272	0.57808	0.22687
LS2	0.06765	0.01673	0.13485	0.03315	0.26997	0.06692	0.53589	0.13466
ELS	0.06993	0.00239	0.13950	0.00553	0.27776	0.00878	0.55709	0.02359
WLS (L&T)	0.05376	0.00107	0.10691	0.00273	0.21361	0.00432	0.43429	0.00801
WLS	0.05367	0.00108	0.10686	0.00270	0.21335	0.00457	0.43377	0.00810
GLS1	0.05089	0.00726	0.10159	0.01491	0.20246	0.02852	0.41018	0.05789
GLS2	0.05186	−0.00413	0.10262	−0.00746	0.20545	−0.01435	0.41371	−0.03020
	n=250							
ML	0.03225	−0.00393	0.06527	−0.00813	0.13035	−0.01499	0.26125	−0.03125
LS1	0.04690	0.01701	0.09390	0.03438	0.18550	0.06628	0.37319	0.13153
LS2	0.04376	0.01060	0.08753	0.02151	0.17273	0.04044	0.34929	0.08020
ELS	0.04416	0.00226	0.08871	0.00508	0.17827	0.00943	0.35890	0.01919
WLS (L&T)	0.03376	0.00068	0.06812	0.00093	0.13531	0.00327	0.27164	0.00474
WLS	0.03376	0.00066	0.06804	0.00087	0.13535	0.00327	0.27137	0.00457
GLS1	0.03178	0.00390	0.06419	0.00758	0.12853	0.01631	0.25743	0.03136
GLS2	0.03107	−0.00171	0.06313	−0.00292	0.12684	−0.00505	0.25707	−0.01208

According to bias criterion:

(**a**) GLS2 estimation outperforms GLS1 for $n = 50, 100, 250$.

(**b**) WLS and GLS1 estimation show better performance than LS1 and LS2.

(**c**) WLS (Lu and Tao) and the considered WLS present similar bias and display the best performance of the considered other methods

(**d**) LS2 performs better than LS1 for all considered sample sizes and shape parameter cases.

Table 3: RMSE and Bias of the estimated shape parameters of the log-logistic distribution.

Methods	Shape							
	1		2		3		6	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
	n=10							
ML	0.38796	−0.15471	0.78149	−0.3109	1.16347	−0.45895	2.32594	−0.91451
LS1	0.30902	0.15929	0.62230	0.31840	0.92998	0.47446	1.85716	0.95062
LS2	0.29362	0.08694	0.59277	0.17352	0.88646	0.25647	1.76698	0.51585
WLS	0.29648	0.07146	0.59574	0.14163	0.89171	0.21789	1.78314	0.43795
GLS1	0.29472	0.08268	0.59331	0.16425	0.88798	0.25136	1.77551	0.50485
GLS2	0.33087	−0.05343	0.65867	−0.10590	0.97803	−0.14401	1.97124	−0.31417
	n=20							
ML	0.22390	−0.06861	0.45489	−0.13848	0.67829	−0.20789	1.34570	−0.41794
LS1	0.22448	0.12767	0.44931	0.25450	0.67293	0.38268	1.34322	0.76474
LS2	0.20281	0.07589	0.40629	0.15064	0.60819	0.22676	1.21250	0.45194
WLS	0.19860	0.05444	0.40275	0.10805	0.60008	0.16147	1.18873	0.32162
GLS1	0.19875	0.07033	0.40239	0.13916	0.60100	0.20922	1.19166	0.41605
GLS2	0.20748	−0.02236	0.41765	−0.04474	0.62353	−0.06622	1.24679	−0.13392
	n=30							
ML	0.17141	−0.04370	0.34395	−0.08501	0.52345	−0.14420	1.03422	−0.27163
LS1	0.18629	0.10648	0.36970	0.21193	0.55713	0.32123	1.11029	0.63844
LS2	0.16603	0.06451	0.32959	0.12839	0.49574	0.19560	0.98802	0.38663
WLS	0.16011	0.04180	0.32274	0.08621	0.48336	0.11330	0.96052	0.24106
GLS1	0.16092	0.05816	0.32357	0.11810	0.48363	0.16259	0.96633	0.34148
GLS2	0.16449	−0.01475	0.32823	−0.02932	0.48997	−0.03890	0.98450	−0.08392
	n=50							
ML	0.12621	−0.02668	0.25227	−0.04722	0.37653	−0.07609	0.76238	−0.16317
LS1	0.14506	0.08190	0.29183	0.16550	0.43646	0.24620	0.86976	0.48919
LS2	0.12857	0.05090	0.25852	0.10325	0.38689	0.15309	0.77050	0.30204
WLS	0.12240	0.02660	0.24763	0.05935	0.36726	0.08315	0.73900	0.15699
GLS1	0.12289	0.04194	0.24840	0.08941	0.36860	0.12961	0.73981	0.24918
GLS2	0.12421	−0.00941	0.24635	−0.01618	0.36919	−0.02661	0.73388	−0.04230
	n=100							
ML	0.08709	−0.01411	0.17256	−0.02669	0.25856	−0.04118	0.52055	−0.08304
LS1	0.10268	0.05523	0.20460	0.10987	0.30874	0.16597	0.61404	0.32834
LS2	0.09175	0.03532	0.18259	0.06989	0.27599	0.10623	0.54827	0.20855
WLS	0.08723	0.01382	0.17290	0.02787	0.25850	0.04125	0.52305	0.08135
GLS1	0.08647	0.02528	0.17223	0.05236	0.25759	0.07728	0.51743	0.15365
GLS2	0.08571	−0.00277	0.16993	−0.00706	0.25560	−0.01324	0.50743	−0.01613
	n=250							
ML	0.05291	−0.00510	0.10644	−0.01131	0.15879	−0.01616	0.32245	−0.03197
LS1	0.06344	0.03021	0.12690	0.06063	0.19083	0.09199	0.38095	0.18491
LS2	0.05809	0.01975	0.11620	0.03977	0.17459	0.06069	0.34850	0.12221
WLS	0.05426	0.00616	0.10873	0.01105	0.16357	0.01762	0.33027	0.03516
GLS1	0.05344	0.01334	0.10712	0.02560	0.15994	0.03929	0.32485	0.07890
GLS2	0.05431	−0.00109	0.10664	−0.00448	0.16027	−0.00499	0.32287	−0.00997

The results for the shape parameter of the log-logistic distribution are presented in Table 3.

According to RMSE:

(**a**) GLS1, GLS2 and WLS show better performance than ML and LS1 in terms of RMSE for all sample sizes except $n = 250$.

According to bias criterion:

(**a**) GLS2 apparently shows the best performance compared with others in terms of bias.

(**b**) WLS and GLS1 are the best performers next to GLS2 for all sample sizes.

In summary, it may be concluded that while the proposed WLS, GLS1 and GLS2 for the shape parameter of Pareto and log-logistic distributions are good alternatives to ML and LS1, the proposed GLS1 and GLS2 for the shape parameter of the Weibull distribution can be preferable estimators in terms of RMSE. If we only consider bias criterion, GLS2 is apparently the best alternative estimator for the shape parameter of the Weibull distribution. In other words, the bias reduction is achieved by GLS2.

Moreover, we found that the LS2 estimation for the shape parameter of Pareto and log-logistic distributions performs better than LS1 in terms of RMSE, similar to the result of the Weibull distribution (Zhang *et al.*, 2007).

Additionally, it can be deduced from all the simulation studies for scale parameters of the considered distributions that the considered GLS1, GLS2 and WLS are in competition with existing estimation methods. Simulation results for the scale parameter are available from the author upon request.

6. REAL LIFE EXAMPLES

In this section, we aim to show the performance of GLS by considering certain real applications.

Example 1

This example was studied with Pareto distribution in Clark, 2013. The sample consists of U.S. Weather/Climate Disasters, taken from the National Climatic Data Center and represents total economic damage from weather events in the U.S. for 1980–2011, adjusted to 2012 dollars. The sample size is 36. Using Q–Q plots, Kolmogorov–Smirnov and Chi-square tests, we show that the Pareto distribution can be used to model this data.

When we use LS1, LS2, GLS1 and GLS2 to estimate the parameters of the Pareto distribution, the descriptive statistics concerning their residuals are given in Table 4. Also, Jarque–Bera (JB) and Durbin–Watson (DW) tests are provided to test normality and autocorrelation, respectively.

The p-values, 0.075 and 0.080, for the Durbin–Watson (DW) test of null hypothesis that errors of the linear regression model are uncorrelated, show that there may be autocorrelation between residuals obtained from LS1 and LS2. On the other hand, the

p-values, 0.2855 and 0.2805 of the DW for the residuals of GLS1 and GLS2, respectively show that the null hypothesis cannot be rejected, that is, the residuals of GLS1 and GLS2 are not autocorrelated.

Table 4: Descriptive statistics and normality test results of regression residuals of the LS1, LS2, GLS1 and GLS2 for the Pareto distribution. (**Note:** Regression residuals' maximum (max), minimum (min), mean, variance (var), skewness (skew.), kurtosis (kurt.) values, also p-value and test value of Jarque Bera (JB) and Durbin Watson (DW) tests for the residuals are presented.)

Methods	Descriptive statistics				JB Test		DW Test	
	min	max	skew.	kurt.	p-val.	Test val.	p-val.	Test val.
LS1	−0.2638	0.1153	−1.6353	8.1862	0.0010	56.3904	0.0075	1.2658
LS2	−0.2430	0.1090	−1.5279	7.8415	0.0010	49.1681	0.0080	1.2723
GLS1	−0.9217	2.2652	1.1770	3.9269	0.0149	9.6002	0.2855	2.4067
GLS2	−1.1162	2.7486	1.2189	3.9606	0.0130	10.2977	0.2805	2.3987

We now calculate the estimates of the scale and shape parameters of the Pareto distribution using the estimation methods mentioned in this study. (See Table 7).

Table 5: Parameter estimates for the Pareto distribution.

Parameters	ML	LS1	LS2	ELS	WLS (Zyl&Schall)	WLS	GLS1	GLS2
Scale	5.3000	5.0354	5.0623	5.2445	5.0907	5.0891	5.1696	5.1744
Shape	1.1902	1.0680	1.0744	1.1754	1.1069	1.0824	1.0995	1.1577

Example 2

The considered data, taken from (Lawless, 2002), is analyzed by (Gupta and Kundu, 2001) with Gamma, Weibull and EE distributions. The data arose from results of tests on the endurance of deep groove ball bearings. We fit the Weibull distribution to this data set and observe that the Weibull distribution can be a plausible model.

The descriptive statistics of the resulting residuals from LS1, LS2, GLS1 and GLS2 are given in Table 6. As can be seen from Table 6, while autocorrelation among the residuals of LS1 and LS2 may be present according to the DW test, with p values less than 0.05, the p-values of 0.6004 and 0.6457 for the DW test suggests that the GLS1 and GLS2 residuals are not autocorrelated.

Table 6: Descriptive statistics and normality test results of regression residuals for the LS1, LS2 and GLS1 for the Weibull distribution.

Methods	Descriptive statistics				JB Test		DW Test	
	min	max	skew.	kurt.	p-val.	Test val.	p-val.	Test val.
LS1	−1.9634	1.8236	−0.0884	2.6113	0.5000	0.1747	0.0003	0.8454
LS2	−1.7711	2.2547	0.1630	2.8357	0.5000	0.1277	0.0004	0.8856
GLS1	−1.2475	2.1430	0.6942	2.4689	0.1289	2.1178	0.6004	2.3730
GLS2	−1.2786	2.1049	0.6090	2.3733	0.1708	1.7982	0.6457	2.3439

The estimates of the scale and shape parameters of the Weibull distribution, using estimation methods mentioned in this study, are provided in Table 7.

Table 7: Parameter estimates for the Weibull distribution.

Param.	ML	LS1	LS2	WLS (Z&S)	WLS (Hung)	WLS (Lu)	WLS	GLS1	GLS2
Scale	81.8958	82.2138	81.6037	81.1957	84.1501	81.0844	81.6037	82.8795	82.9189
Shape	2.1030	2.0430	2.1037	1.8748	1.8743	1.8863	2.1037	1.8756	2.0167

7. CONCLUSIONS

In this article, we consider generalized least squares (GLS1 and GLS2) and weighted least squares (WLS) estimation methods, based on an easily-calculated approximation of the covariance matrix, for estimating the parameters of a distribution that can be converted to a linear regression model. The considered GLS1, GLS2 and WLS methods, which are computationally easy and provide explicit estimators of the parameters, are then applied to the estimation of the parameters of different distributions, such as the Weibull, Pareto and log-logistic. The simulation results show that the considered GLS1, GLS2 and WLS estimation methods, for the shape parameters of Pareto and log-logistic distributions, show better performance than ML, LS and certain alternative estimation methods in terms of RMSE for most of the considered sample sizes and shape cases. In addition, GLS1 and GLS2 apparently provide an improvement over ML, LS and certain alternative WLS for the shape parameter of the Weibull distribution in terms of RMSE and bias. In conclusion, the results of the simulations and real life examples demonstrate that the considered GLS1 and GLS2 for the shape parameters of log-logistic, Pareto and Weibull distributions can be considered as good alternative estimators.

Moreover, it is also emphasized that the considered estimation methods can be applied to Gumbel, Burr XII, Fréchet and other distributions, which have explicit cumulative distribution functions, after calculation of the covariance matrix concerning them.

In a future study, we plan to investigate the performance of the GLS estimation method in the case of right censored data and contaminated data.

ACKNOWLEDGMENTS

I would like to thank the anonymous referees and particularly the Associate Editor for their helpful and constructive suggestions and particularly for the proposal for the covariance matrix.

REFERENCES

[1] Abernethy, R.B. (1996). *The New Weibull Handbook*, R.B. Abernethy North Palm Beach.

[2] Ahmad, M.I.; Sinclair, C.D. and Werrity, A. (1988). Log-logistic flood frequency analysis, *J. Hydrol.*, **98**, 205–224.

[3] Ashkar, F. and Mahdi, S. (2006). Fitting the log-logistic distribution by generalized moments, *Journal of Hydrology*, **3-4**, 328, 694–703.

[4] Baxter, M.A. (1980). Minimum variance unbiased estimation of the Pareto distribution, *Metrika*, **27**, 133–138.

[5] Bergman, B. (1986). Estimation of Weibull parameters using a weight function, *Journal of Materials Science Letters*, **5**, 611–614.

[6] Blom, G. (1962). *Nearly best linear estimators of scale and location parameters.* Contributions to Order Statistics. (A.E. Sarhan and B.G. Greenberg, Eds.), New York: John Wiley, Ch. 4B.

[7] Boudt, K.; Caliskan, D. and Croux, C. (2011). Robust explicit estimators of Weibull parameters, *Metrika*, **73**, 187–209.

[8] Chen, Z. (2006). Estimating the shape parameter of the log-logistic distribution, *Int. J. Rel., Qual. and Safety Eng.*, **13**, 257–266.

[9] Clark, D.R. (2013). A Note on the Upper-Truncated Pareto Distribution, *Actuarial Society E-Forum.*

[10] Engeman, R.M. and Keefe, T.J. (1982). On generalized least squares estimation of the Weibull distribution, *Communications in Statistics – Theory and Methods*, **19**, 11, 2181–2193.

[11] Gebizlioglu, O.L.; Senoglu, B. and Kantar, Y.M. (2011). Comparison of certain value-at-risk estimation methods for the two-parameter Weibull loss distribution, *Journal of Computational and Applied Mathematics*, **235**, 3304–3314.

[12] Genschel, U. and Meeker, W.Q. (2010). A Comparison of Maximum Likelihood and Median-Rank Regression for Weibull Estimation, *Quality engineering*, **4**, 22, 236–255.

[13] Gupta, R.D. and Kundu, D. (2001). Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions, *Biometrical Journal*, **43**, 117–130.

[14] Hassanein, K.M. (1971). Percentile estimators for the parameters of the Weibull distribution, *Biometrika*, **3**, 58, 673–676.

[15] Hossain, A. and Zimmer, W.J. (2000). Comparisons of methods of estimation for Pareto of the first kind, *Communications in Statistics – Theory and Methods*, **4**, 29, 859–878.

[16] Hossain, A. and Zimmer, W.J. (2003). Comparison of estimation methods for Weibull parameters: complete and censored samples, *Journal of Statistical Computation and Simulation*, **2**, 73, 145–153.

[17] Hossain, A. and Howlader, H.A. (1996). Unweighted least squares estimation of Weibull parameters, *Journal of Statistical Computation and Simulation*, **54**, 265–271.

[18] Hung, W.L. (2001). Weighted least squares estimation of the shape parameter of the Weibull distribution, *Qual. Reliab. Eng. Int.*, **6**, 17, 467–469.

[19] Kantar, Y.M. and Senoglu, B. (2008). A Comparative study for the location and scale parameters of the Weibull distribution with a given shape parameter, *Computers & Geosciences*, **12**, 34, 1900–1909.

[20] Kantar, Y.M. and Usta, I. (2008). Analysis of wind speed distributions: wind distribution function derived from minimum cross entropy principles as better alternative to Weibull function, *Energy Convers. Manage.*, **49**, 962–973.

[21] Kantar, Y.M. and Arik, I. (2014). M-estimation of log-logistic distribution parameters with outliers, *International Journal of Agricultural and Statistical Sciences*, Accepted manuscript.

[22] Kantar, Y.M. and Yildirim, V. (2015). Robust Estimation for Parameters of the Extended Burr Type III Distribution, *Communications in Statistics – Theory and Methods, Communications in Statistics – Simulation and Computation*, **44**, 7, 1901–1930.

[23] Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.

[24] Lawless, J.F. (2002). *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, New York.

[25] Likes, J. (1969). Minimum variance unbiased estimates of the Pareto distribution, *Metrika*, **15**, 126–132.

[26] Lu, H.L.; Chen, C.H. and Wu, J.W. (2004). A Note on weighted least-squares estimation of the shape parameter of the Weibull distribution, *Qual. Reliab. Eng. Int.*, **6**, 20, 579–586.

[27] Lu, H.L. and Tao, S.H. (2007). The Estimation of Pareto Distribution by a Weighted Least Square Method, *Quality & Quantity*, **41**, 913–926.

[28] Malik, H. (1970). Estimation of the parameters of the Pareto distribution, *Metrika*, **1**, 15, 126–132.

[29] Marks, N.B. (2005). Estimation of Weibull parameters from common percentiles, *J. Appl. Stat.*, **1**, 32, 17–24.

[30] Nelson, W. (2004). *Applied Life Data Analysis*, Wiley, USA.

[31] Quandt, R.E. (1966). Old and New methods of estimation and the Pareto distribution, *Metrika*, **1**, 10, 55–82.

[32] Prakash, G. and Singh, D.C. (2009). A Bayesian shrinkage approach in Weibull Type-II censored data using prior point information, *Revstat – Statistical Journal*, **2**, 7, 171–187.

[33] Saksena, S.K. and Johnson, A.M. (1984). Best unbiased estimators for the parameters of a two-parameter Pareto distribution, *Metrika*, **1**, 31, 77–83.

[34] Teimouri, M. and Nadarajah, S. (2012). A simple estimator for the Weibull shape parameter, *Int. J. Str. Stab. Dyn.*, **395**, DOI: 10.1142S0219455412500095.

[35] Tiryakioglu, M. and Hudak, D. (2007). On estimating Weibull modulus by the linear regression method, *Mater. Sci.*, **24**, 42, 10173–10179.

[36] Usta, I. (2013). Different estimation methods for the parameters of the extended Burr XII distribution, *J. App. Stat.*, **2**, 40, 397–414.

[37] Wang, F.K. and Cheng, Y.F. (2010). Robust regression for estimating the Burr XII parameters with outliers, *J. App. Stat.*, **5**, 37, 807–819.

[38] White, J.S. (1969). The Moments of Log-Weibull Order Statistics, *Technometrics*, **2**, 11, 373–386.

[39] Zhang, L.F.; Xie, M. and Tang, L.C. (2008). On weighted least squares estimation for the parameters of Weibull distribution, *Recent Advances in Reliability and Quality in Design Springer Series in Reliability Engineering*, **2**, 11, Chapter 57–84.

[40] Zhang, L.F.; Xie, M. and Tang, L.C. (2007). A study of two estimation approaches for parameters of Weibull distribution based on WPP, *Reliab. Eng. Syst. Safe.*, **92**, 360–368.

[41] Zyl, J.M.V. (2012). A Median Regression Model to Estimate the Parameters of the Three-Parameter Generalized Pareto Distribution, *Commun. Stat.-Simul. C.*, **41**, 544–553.

[42] Zyl, J.M.V. and Schall, R. (2012). Parameter Estimation through Weighted Least-Squares Rank Regression with Specific Reference to the Weibull and Gumbel Distributions, *Commun. Stat.-Simul. C.*, **41**, 1654–1666.

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