ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
PORTUGAL

REVSTAT

Statistical Journal

Volume 3, No.2
November 2005

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 (INE)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: (0351) 21 842 61 00
    Fax: (0351) 21 842 63 64
    Web site: http://www.ine.pt
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística*
- **EDITION**
  - *400 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue ..... **€ 12**
- Anual subscription ... **€ 19,20**

INDEX

REVSTAT – Statistical Journal
Volume 3, Number 2, November 2005, 99–111

ANALYSIS OF DYNAMIC PROTEIN EXPRESSION DATA

Authors: Klaus Jung
– Department of Statistics, University of Dortmund, Germany
(klaus.jung@uni-dortmund.de)

Ali Gannoun
– Equipe de Probabilités et Statistique, Université Montpellier, France

Barbara Sitek
– Medical Proteom-Center, Ruhr-University Bochum, Germany

Helmut E. Meyer
– Medical Proteom-Center, Ruhr-University Bochum, Germany

Kai Stühler
– Medical Proteom-Center, Ruhr-University Bochum, Germany

Wolfgang Urfer
– Department of Statistics, University of Dortmund, Germany
(urfer@statistik.uni-dortmund.de)

Received: May 2005 Accepted: June 2005

Abstract:

- Difference gel electrophoresis (DIGE) is the new gold standard analysing complex protein mixtures in proteomics. It is used for measuring the expression levels of proteins in different mixtures on the same two-dimensional electrophoresis (2-DE) gel. In this paper we review a method for the calibration and normalization of those protein expression measurements. Further we show how to find treatment effects and time-treatment-interactions in longitudinal data obtained from DIGE experiments. A problem in those data sets is the existence of a lot of missing values. Therefore, we propose a method for the estimation of missing data points.

Key-Words:

- *difference gel electrophoresis; data calibration; mixed linear model for longitudinal data; missing values; proteomics.*

1. INTRODUCTION

While the focus of biochemical research was addressed on the genome in the last decade the view is now turned onto the proteome. Big data sets of gene expression obtained from DNA-microarrays made the development of statistical methods necessary to make correct inferences from these measurements. For quantitative protein expression analysis either mass spectrometry (cf. Aebersold and Goodlett ([1]) and Gygi et al. ([7])) or two-dimensional gel electrophoresis (2-DE) (cf. Westermeier et al. ([14])) is applied. In this paper we focus on the analysis of protein expression data obtained from a new detection method (Difference Gel Electrophoresis, DIGE) based on fluorescence labelling before 2-DE. 2-DE separates the proteins of a mixture by their isoelectric point (pI) and molecular size to distinct spots. After separation the proteins are detected using a confocal fluorescence scanner whereas fluorescence intensity of a spot can be regarded as a measure of expression for its respective protein. DIGE enables the user to put up to three different mixtures of proteins on the same gel. The different mixtures are labelled by different fluorescence dyes (Cy2, Cy3 and Cy5). For quantitative proteome analysis image analysis software automatically determines the boundaries and sizes of the spots. Usually, a DIGE experiment is designed such that m independent replications of treatment and control mixtures are put on the same m gels. The internal standard, a mixture of same amounts of all m treatment and m control probes, is also put on each gel. This internal standard allows high accuracy calibration of the expression values. Calibration and normalization of protein expression data is reviewed in section 2. In order to obtain information about interactions of treatment and control with the time, DIGE experiments often include measurements over several time points. Known statistical methods for the analysis of longitudinal data can be used to analyze those experiments. One possible method for such an analysis is detailed in section 3. Often, 2-DE data contains up to 50% of missing values. The missing values occur because not each protein is visible on each gel when replicating probes on several gels. For example, on gel number one there are 1732 protein spots and 1967 spots are on gel number two, but only 1447 of these spots belong to proteins commonly represented on both gels. Some statistical methods, however, need complete data sets, for example, some methods for the detection of differentially expressed genes (cf. Gannoun et al. ([6])) or the correspondence analysis for microarray data (cf. Fellenberg et al. ([5])). These methods could also be applied to protein expression data if the data sets were complete. One possible method to overcome this problem is to estimate the missing values by using the available measurements from other proteins. In section 4, we investigate how the k nearest neighbor method behaves when being applied to DIGE data. This method was also applied for the estimation of missing values in gene expression data by Troyanskaya et al. ([13]). The idea of this method is that there are groups of proteins with similar expression profiles. A missing value of a protein can then be estimated by available values from the proteins of the same group.

2. CALIBRATION, NORMALIZATION AND STANDARDIZATION OF DIGE DATA

A usual DIGE experiment results in three values for each spot on a gel, i.e. treated, untreated and internal standard. From the DeCyder^{TM} software one can obtain the background subtracted spot volumes (cf. Amersham Biosciences ([2])). In this software, a borderline for each spot is automatically detected and the sum of the pixel intensities within the spot boundary is the spot volume. The background is subtracted by excluding the lowest 10th percentile pixel values on the spot boundary. As we will see in this section the statistical analysis cannot be done with this raw data material. Data obtained from analytical instruments are always affected by technical and biological variation. To make correct inferences on the biological variation preprocessing of data is necessary. In this section we discuss the features of the background subtracted spot volumes and describe how to calibrate and transform the values for further actual analysis. One source of technical variation comes from the different dyes. In figure 1 the Cy5 and Cy3 spot volumes of a DIGE gel are plotted against each other.

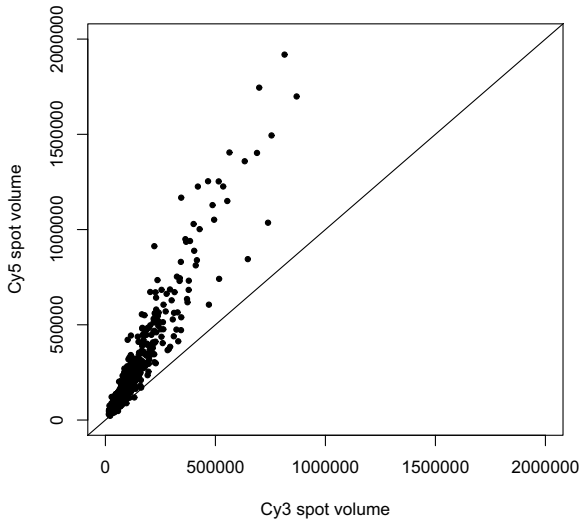


Figure 1: Scatterplot of the Cy5 versus the Cy3 spot volumes of a DIGE gel.

It can be seen that the Cy5 dye causes higher volume values than the Cy3 dye. To calibrate the spot volumes Karp et al. ([9]) proposed to use a scaling factor which adjusts for the dye-specific gain, and to use an additive offset which compensates for any constant additive bias present after background subtraction.

The additive offset is used because the different dyes result in different background fluorescence. This calibration method was originally introduced by Huber et al. ([8]) for the preprocessing of DNA-microarray data. Having n spots on a gel with three different mixtures (internal standard, treated, untreated) this calibration can be modelled by

$$\tilde{y}_{ih} = a_h + b_h y_{ih} \tag{2.1}$$

with $i = 1, ..., n$ and $h = 1, 2, 3$. For $h = 1$ we have the value for the treated probe, $h = 2$ for the untreated probe and $h = 3$ for the internal standard. In this model $\tilde{y}_{ih}$ are the measured background subtracted spot volumes, a_h are the additive offsets and b_h are the scaling factors. Hence, $2 * 3$ parameters have to be estimated. How to do this will be explained below. Some more features of the raw data require a second transformation. The scanning of the fluorescent gels results in lognormal distributions of the spot volumes. However, a normal distribution would be more appropriate for most statistical applications so the data has to be normalized. Furthermore, the variance of the spot volumes is dependent on the mean of the spot volumes. This is illustrated in figure 2.

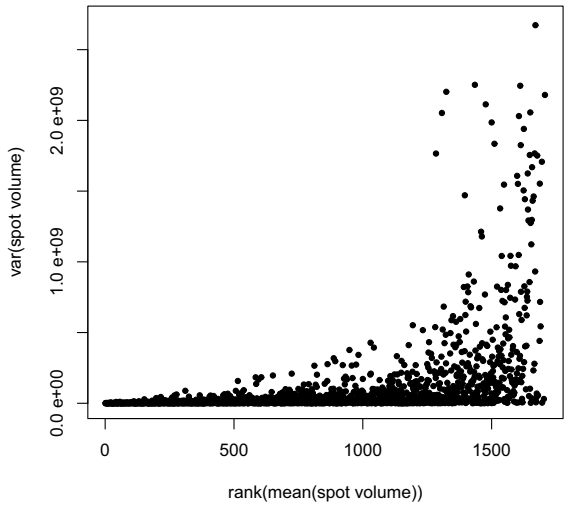


Figure 2: Scatterplot of the variance of the Cy3 and Cy5 spot volumes versus the rank their mean.

The variance of the spot volumes increases when the mean also increases. One possibility to normalize the data and to stabilize the variance would be to apply the logarithm on the data. But the logarithm results in a bias for low spot volumes as can be seen in in figure 3 where the Cy3 and Cy5 spot volumes with the logarithm applied on them are plotted against each other. Instead of using the logarithm we will use the arsinh for normalization and variance stabilization.

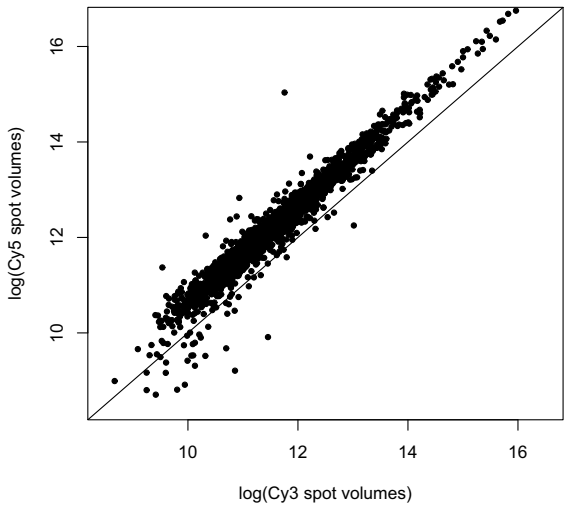


Figure 3: Scatterplot of the log-transformed Cy5 spot volumes versus the log-transformed Cy3 spot volumes.

The graphs of the logarithm and the arsinh are plotted in figure 4.

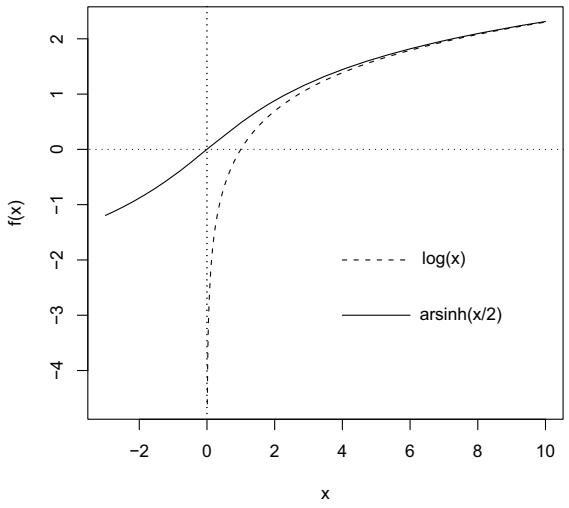


Figure 4: Graphs of the arsinh and the logarithm.

The relationship between the two functions can be expressed by

$$\lim_{\xi \to \infty} (\operatorname{arsinh} \xi - \log \xi - \log 2) = 0 \ .$$

Hence, for big values the arsinh is equivalent to the logarithm, but it has not a singularity at zero and it is smooth for small values. Now, using the calibration

transformation and the arsinh, the true protein abundance x_{ih} can be modelled by

$$\text{arsinh}\,\frac{\tilde{y}_{ih} - a_h}{b_h} = x_{ih} + \varepsilon_{ih} \tag{2.2}$$

where $\varepsilon_{ih} \sim N(0, \sigma_\varepsilon)$. To estimate $(a_1, a_2, a_3, b_1, b_2, b_3)$ Huber et al. ([8]) proposed a robust version of maximum likelihood estimation. The robust version is necessary because maximum likelihood estimation itself is very sensitive to deviations from the normal distribution and to the presence of differentially expressed proteins. The above estimation algorithm is implemented in the vsn-package for the software R (both free available at http://cran.r-project.org). The resulting benefits of calibration and normalization can be seen in the figures 5 and 6.

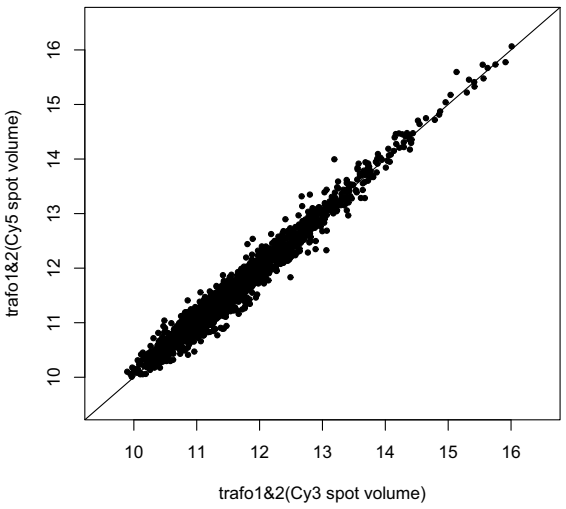


Figure 5: Calibrated and transformed Cy3 spot volumes versus calibrated and transformed Cy5 spot volumes.

In figure 5 it is shown that there is no more dye-specific gain for the calibrated and transformed spot volumes. Further, the bias for low spot volumes has disappeared. The variance of the calibrated and transformed volumes versus the rank of their mean is plotted in figure 6. It can be seen that there is no more dependence between variance and mean. Now, after calibration and normalization, we can use the benefit of the internal standard to reduce the gel-to-gel variation and bring all gels on the same level. This means we set the calibrated and arsinh-transformed treatment and control values in relation to the internal standard value. More precisely we have to subtract the internal standard from the treatment and control value, respectively, because ratios become differences when the logarithm or the arsinh is applied on them. Hence, the standardized treatment value is $x_{i1} - x_{i3}$ and the standardized control value is $x_{i2} - x_{i3}$.

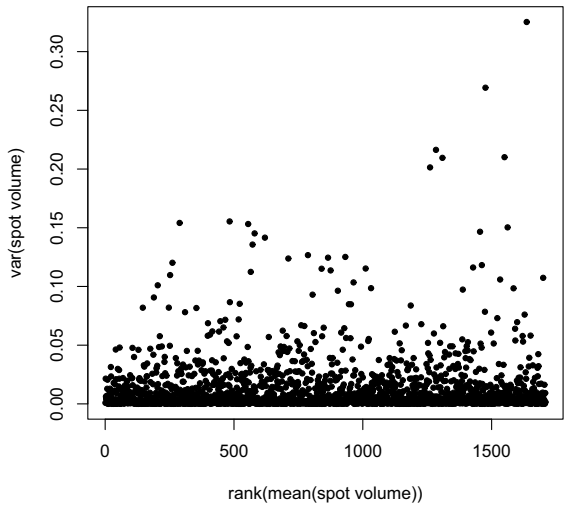


Figure 6: Variance of the calibrated and normalized spot volumes versus the rank of their mean.

3. ANALYSIS OF LONGITUDINAL DIGE DATA

A frequent subject of DIGE studies is the comparison of the temporal course of the protein expression in some treated probes to the temporal course of the protein expression in some untreated probes. Since there are only a few time points to be regarded such a study can be analyzed by using methods of longitudinal data analysis. Here, we adapt such a method, given in Diggle et al. ([4]), to the situation of a DIGE experiment. The design for a time dependent DIGE experiment is given in table 1. For each spot, which has been detected

Table 1: Design of a time dependent DIGE experiment.

	replication 1	replication 2	...	replication m
time 1	gel_{11}	gel_{12}	...	gel_{1m}
time 2	gel_{21}	gel_{22}	...	gel_{2m}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
time p	gel_{p1}	gel_{p2}	...	gel_{pm}

on each of the pm gels, the analysis is done separately. Recall, that for each spot and each gel we get a standardized volume value for the treated probe and a

standardized value for the untreated probe. We denote y_{hiq} as the standardized volume value for the spot in question on the jth gel at the qth time point within the hth group (treated, untreated), where $j = 1, ..., m$, $q = 1, ..., p$ and $h = 1, 2$. Since we analyze the same protein over the time we need a model which heeds the time-dependence of the values. Therefore, we regard the mixed linear model

$$y_{hjq} = \beta_h + \gamma_{hq} + U_{hj} + Z_{hjq} \tag{3.1}$$

where β_h is the main effect of the hth group, γ_{hq} is the interaction between group and time, $U_{hj} \sim N(0, \nu^2)$ is the random effect of the jth replication and $Z_{hjq} \sim N(0, \sigma^2)$ are the random errors. With the given distribution assumptions for the random effects the vector $Y_{hj} = (Y_{hj1}, Y_{hj2}, ..., Y_{hjp})$ is normally distributed with covariance matrix $V = \sigma^2 I + \nu^2 J$. That means that the correlation between two time points is given by $\rho = \nu^2/(\nu^2 + \sigma^2)$. At first we want to test the null hypothesis that there is no treatment effect, i.e. testing $\beta_h = \beta$ for $h = 1, 2$, meaning that the temporal course for the protein in the treated and untreated probe are on the same level. The F-statistic for testing this hypothesis is given by $F_1 = \{BTSS_1/(2-1)\}/\{RSS_1/(2m-2)\} \sim F_{(2-1),(2m-2)}$. The sums of squares are given in the corrected ANOVA table 2 below. We are further interested in the question if there is a treatment-time interaction, i.e. the temporal courses are not parallel. This can be answered by testing the null hypothesis $\gamma_{hq} = \gamma_q$ for $h = 1, 2$ and for $q = 1, ..., p$. This null hypothesis means that the response profiles of the group means are parallel. The according test statistic is given by $F_2 = \{ISS_2/[(2-1)(p-1)]\}/\{RSS_2/[(2m-2)(p-1)]\} \sim F_{(2-1)(p-1),(2m-2)(p-1)}$.

Table 2: ANOVA table for the Analysis of longitudinal DIGE data.

source of variance	sums of squares	d.o.f.
between treatment	$BTSS_1 = p\sum_{h=1}^{2} m(y_{h\cdot\cdot} - y_{\cdot\cdot\cdot})^2$	$2-1$
whole plot residual	$RSS_1 = TSS_1 - BTSS_1$	$2m-2$
whole plot total	$TSS_1 = p\sum_{h=1}^{2}\sum_{j=1}^{m}(y_{hj\cdot} - y_{\cdot\cdot\cdot})^2$	
between time	$BTSS_2 = 2m\sum_{q=1}^{p}(y_{\cdot\cdot q} - y_{\cdot\cdot\cdot})^2$	$p-1$
treatment-time interaction	$ISS_2 = \sum_{q=1}^{p}\sum_{h=1}^{2} m(y_{h\cdot q} - y_{\cdot\cdot\cdot})^2 - BTSS_1 - BTSS_2$	$(2-1)\times (p-1)$
split plot residual	$RSS_2 = TSS_2 - ISS_2 - BTSS_2 - TSS_1$	$(2m-2)\times (p-1)$
split plot total	$TSS_2 = \sum_{h=1}^{2}\sum_{j=1}^{m}\sum_{q=1}^{p}(y_{hjq} - y_{\cdot\cdot\cdot})^2$	$2pm-1$

4. MISSING VALUE ESTIMATION

As mentioned in the beginning missing values are a general problem in 2-DE data. In this section we present a method for the estimation of missing data, using the k nearest neighbor method. We begin with some notation. Let $E = (e_{ij})$ be the matrix of observations, where the rows are referred to protein spots and the columns are referred to replications (gels). Hence, e_{ij} is the expression value of protein i on gel j, with $i = 1, ..., n$ and $j = 1, ..., m$, as given below.

$$\begin{pmatrix} e_{11} & ... & e_{1m} \\ \vdots & & \vdots \\ e_{i1} & e_{ij} & e_{im} \\ \vdots & & \vdots \\ e_{n1} & ... & e_{nm} \end{pmatrix} \tag{4.1}$$

Now, we can define distances between each pair of rows of E ($E_i = (e_{i1}, ..., e_{im})'$, $E_{i'} = (e_{i'1}, ..., e_{i'm})'$). The Euclidean distance is given by

$$d_1(E_i, E_{i'}) = \sqrt{(e_{i1} - e_{i'1})^2 + (e_{i2} - e_{i'2})^2 + ...(e_{im} - e_{i'm})^2} , \tag{4.2}$$

the Tschebyscheff distance is given by

$$d_2(E_i, E_{i'}) = \sup |e_{ij} - e_{i'j}| , \tag{4.3}$$

$j = 1, ..., m,$ and the Mahalanobis distance is given by

$$d_3(E_i, E_{i'}) = \sqrt{(E_i - E_{i'})^T A^{-1} (E_i - E_{i'})} , \tag{4.4}$$

where A is the empirical covariance matrix of the m gels. The principle of the k nearest neighbor method is now the following. For the row E_i the k nearest neighbors are those rows of E with the k smallest distances to E_i. More details on the k nearest neighbor method can be found in Ripley ([11]). This method was used in nonparametric estimation of the density (see for example Rosenblatt ([12]) and regression (see for example Devroye ([3])) as well as in classification problems (see for example Ketskemety ([10])). With the above given notations missing protein measurements can be estimated as follows. Let E_i be the row where the value e_{ij} is missing. Let Q_i be the set of non missing values of E_i. We denote these values by e'_{ip}, $p = 1, ..., q$, and $E'_i = (e'_{i1}, ..., e'_{iq})^T$. Let E_s, $s \neq i$, be the row s of the Matrix E. We suppose that e_{sj} is available and at least q other e_{sp} are available, too, in the same columns as in E_i. Then we denote $E'_s = (e'_{s1}, ..., e'_{sq})^T$ and give the

Definition 4.1. E_i and E_s are neighbors if $d(E_i', E_s')$ is small.

and

Definition 4.2. The k rows E_s $(s \neq i)$ with the k smallest distances to E_i are the k nearest neighbors to E_i.

To estimate the missing value e_{ij} let $e_{s_1 j}, e_{s_2 j}, ..., e_{s_k j}$ be the e_{sj} such that E_s belongs to the k nearest neighbors of E_i. The missing value e_{ij} can now be estimated by

$$\hat{e}_{ij}^{\text{mean}} = \frac{1}{k} \sum_{l=1}^{k} e_{s_l j} \ , \tag{4.5}$$

$$\hat{e}_{ij}^{\text{wmean}} = \frac{1}{k} \sum_{l=1}^{k} w_{i s_l} e_{s_l j} \ , \tag{4.6}$$

with

$$w_{i s_l} = \frac{1}{d(E_i', E_{s_l}') \sum\limits_{t=1}^{k} \frac{1}{d(E_i', E_{s_t}')}} \ , \tag{4.7}$$

or by

$$\hat{e}_{ij}^{\text{median}} = \text{median}(e_{s_1 j}, e_{s_2 j}, ..., e_{s_k j}) \ . \tag{4.8}$$

We applied the k nearest neighbor algorithm to protein expression data from a neuroblastoma DIGE study. To get an idea of how good the method works, we took a complete matrix A from which we generated an incomplete matrix B with 40% of randomly chosen missing values. The missing values where estimated with the k nearest neighbor method by using different combinations of distances (d_1, d_2, d_3) and estimators $(\hat{e}_{ij}^{\text{mean}}, \hat{e}_{ij}^{\text{wmean}}, \hat{e}_{ij}^{\text{median}})$ as well as different ks. For each estimated matrix B we calculated the normalized root mean square (RMS) error

$$\frac{\sqrt{\sum\limits_{j=1}^{m} \sum\limits_{i=1}^{n} (A_{ij} - B_{ij})^2 / (n * m)}}{\text{mean}(A)} \ , \tag{4.9}$$

to compare it to the complete matrix A. By comparing the errors for the different ways of estimation we came to the result that $\hat{e}_{ij}^{\text{mean}}$, $\hat{e}_{ij}^{\text{wmean}}$ and $\hat{e}_{ij}^{\text{median}}$ have a similar performance. Further, we found out that the error is nearly the same when the Euclidean or Mahalanobis distance is used, but it is higher when the *sup*-distance is used. For the appropriate number of neighbors, we saw that the error was smallest between 5 and 20 neighbors. We applied this missing value estimation to get a balanced data structure for the analysis of the longitudinal DIGE data using the mixed linear model described in section 3.

ACKNOWLEDGMENTS

Klaus Jung is supported by a predoctoral fellowship of the German Research Foundation.

REFERENCES

[1] Aebersold, R. and Goodlett, D.R. (2001). Mass spectrometry in proteomics, *Chemical Reviews*, **101**, 269–295.

[2] Amersham Biosciences (2003). *DeCyder Differential Analysis Software, Version 5.0, User Manual*, Amersham Biosciences, Sweden.

[3] Devroye, L.P. (1978). The uniform convergence of nearest neighbor regression function estimators and their application in optimization, *IEEE Trans. Inf. Theory*, **24**, 142–151.

[4] Diggle, P.J.; Liang, K.Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Clarendon Press, Oxford.

[5] Fellenberg, K.; Hauser, N.C.; Brors, B.; Neutzner, A.; Hoheisel, J.D. and Vingron, M. (2001). Correspondence analysis applied to microarray data, *PNAS*, **98**, 10781–10786.

[6] Gannoun, A.; Saracco, J.; Urfer, W. and Bonney, G.E. (2004). Non-parametric analysis of replicated microarray experiments, *Statistical Modelling*, **4**, 195–209.

[7] Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nature Biotechnology*, **17**, 994–999.

[8] Huber, W.; Heydebreck, A. von; Sueltmann, H.; Poustka, A. and Vingron, M. (2002), *Bioinformatics*, **18**, S96–S104.

[9] Karp, A.N.; Kreil, D.P. and Lilley, K.S. (2004). Determining a significant change in protein expression with DeCyder during a pairwise comparison and to the quantification of differential expression, *Proteomics*, **4**, 1421–1432.

[10] Ketskemety, L. (2004). Effectiveness of nearest neighbor classification with optimal scaling, *Alkalmazott Mat. Lapok.*, **21**, 81–97.

[11] Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press.

[12] Rosenblatt, M. (1979). Global measures of deviation for kernel and nearest neighbor density estimates. Smoothing techniques for curve estimation. *Proc. Workshop, Heidelberg, Lect. Notes Math.*, **757**, 181–190.

[13] Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D. and Altman, R.B. (2001). Missing value estimation methods for DNA microarrays, *Bioinformatics*, **17**:6, 520–525.

[14] Westermeier, R.; Gronau, S. and Beckett, P. (2004). *Electrophoresis in Practice*, Wiley-VCH, Weinheim.

REVSTAT – Statistical Journal
Volume 3, Number 2, November 2005, 113–136

DIRECT REDUCTION OF BIAS OF THE CLASSICAL HILL ESTIMATOR *

Authors: Frederico Caeiro
– Universidade Nova de Lisboa, FCT(DM) and CEA, Portugal
(fac@fct.unl.pt)

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

Dinis Pestana
– Universidade de Lisboa, FCUL (DEIO) and CEAUL, Portugal
(dinis.pestana@fc.ul.pt)

Received: January 2005 Revised: June 2005 Accepted: July 2005

Abstract:

- In this paper we are interested in an adequate estimation of the dominant component of the bias of Hill's estimator of a positive tail index γ, in order to remove it from the classical Hill estimator in different asymptotically equivalent ways. If the second order parameters in the bias are computed at an adequate level k_1 of a larger order than that of the level k at which the Hill estimator is computed, there may be no change in the asymptotic variances of these reduced bias tail index estimators, which are kept equal to the asymptotic variance of the Hill estimator, i.e., equal to γ^2. The asymptotic distributional properties of the proposed estimators of γ are derived and the estimators are compared not only asymptotically, but also for finite samples through Monte Carlo techniques.

Key-Words:

- *Statistics of Extremes; semi-parametric estimation; bias estimation; heavy tails.*

AMS Subject Classification:

- 62G32, 65C05.

*Research partially supported by FCT / POCTI and POCI / FEDER.

1. INTRODUCTION AND MOTIVATION FOR THE NEW TAIL INDEX ESTIMATORS

In *Statistics of Extremes*, the tail index γ is the basic parameter of extreme events. Such a parameter plays a relevant role in other extreme events' parameters, like high quantiles and return periods of high levels, among others. The tail index is a real-valued parameter and the heavier the tail, the larger the tail index γ is. Heavy-tailed models have revealed to be quite useful in most diversified fields, like computer science, telecommunication networks, insurance and finance. In the field of Extremes, we usually say that a model F is *heavy-tailed* whenever the *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $\{-1/\gamma\}$, $\gamma > 0$, or equivalently, the quantile function $U(t) = F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, with $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, is of regular variation with index γ. This means that, for every $x > 0$,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma} \quad \Longleftrightarrow \quad \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma} \ . \tag{1.1}$$

We shall here concentrate on these Pareto-type distributions. Note that (1.1) is equivalent to saying that

$$1 - F(x) = x^{-1/\gamma}\, L_{_F}(x) \quad \Longleftrightarrow \quad U(x) = x^{\gamma}\, L_{_U}(x) \ , \tag{1.2}$$

with $L_{_F}$ and $L_{_U}$ slowly varying functions, i.e., functions $L_{\bullet}$ such that $L_{\bullet}(tx)/L_{\bullet}(t) \to 1$, as $t \to \infty$, for all $x > 0$.

The *second order parameter* ρ (≤ 0), rules the rate of convergence in the first order condition (1.1) (or equivalently, (1.2)), and is the non-positive parameter appearing in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \lim_{t\to\infty} \frac{\ln L_{_U}(tx) - \ln L_{_U}(t)}{A(t)} = \frac{x^{\rho} - 1}{\rho} \ , \tag{1.3}$$

which we assume to hold for all $x > 0$ and where $|A(t)|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). We shall assume everywhere that $\rho < 0$.

Remark 1.1. For the strict Pareto model, $F(x) = 1 - C\ x^{-1/\gamma}$, $x \geq C^{\gamma}$, and indeed only for this model, the numerator of the fraction in the left hand-side of (1.3) is null, i.e., $\ln U(tx) - \ln U(t) - \gamma \ln x \equiv 0$.

Remark 1.2. For Hall's class of Pareto-type models (Hall, 1982; Hall and Welsh, 1985), with a tail function

$$1 - F(x) \ = \ Cx^{-1/\gamma}\Big(1 + Dx^{\rho/\gamma} + o\big(x^{\rho/\gamma}\big)\Big) \ , \qquad \text{as} \ \ x \to \infty \ , \tag{1.4}$$

$C > 0$, $D \in \mathbb{R}_0$, $\rho < 0$, (1.3) holds and we may choose $A(t) = \gamma\rho DC^{\rho}t^{\rho}$.

This is a class of models where (1.2) (or equivalently, (1.1)) holds true, with an asymptotically constant slowly varying function $L_{_F}$ (or equivalently, $L_{_U}$).

To obtain information on the distributional behaviour of the second order parameters' estimators, we shall further assume that the rate of convergence in (1.3) is ruled by a function $B(t)$ such that $|B(t)|$ is also of regular variation with the same index ρ, i.e., we assume that

$$\lim_{t\to\infty} \frac{\frac{\ln U(tx)-\ln U(t)-\gamma\ln x}{A(t)} - \frac{x^\rho-1}{\rho}}{B(t)} = \frac{x^{2\rho}-1}{2\rho} \tag{1.5}$$

holds for all $x > 0$.

Remark 1.3. Condition (1.5) holds true for models with a tail function

$$1 - F(x) = Cx^{-1/\gamma}\Big(1 + D_1 x^{\rho/\gamma} + D_2 x^{2\rho/\gamma} + o\big(x^{2\rho/\gamma}\big)\Big), \quad \text{as } x \to \infty . \tag{1.6}$$

For the most common heavy-tailed models, like the Fréchet and the Student's t, condition (1.5) holds true, i.e., these models belong to the class in (1.6).

For intermediate k, i.e., a sequence of integers $k = k_n$, $1 \le k < n$, such that

$$k = k_n \to \infty , \qquad k_n = o(n) , \quad \text{as } n \to \infty , \tag{1.7}$$

we shall consider, as basic statistics, both the log-excesses over the random high level $\{\ln X_{n-k:n}\}$, i.e.,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n} , \qquad 1 \le i \le k < n , \tag{1.8}$$

and the scaled log-spacings,

$$U_i := i\left\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\right\} , \qquad 1 \le i \le k < n , \tag{1.9}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated to a random sample $(X_1, X_2, ..., X_n)$.

We may write $X_{i:n} \stackrel{d}{=} U(Y_{i:n})$, where $\{Y_i\}$ denotes a sequence of unit Pareto random variables (r.v.'s), i.e., $P(Y \le y) = 1 - 1/y$, $y \ge 1$. Also, for $j > i$, $Y_{j:n}/Y_{i:n} \stackrel{d}{=} Y_{j-i:n-i}$, $\ln Y_{i:n} \stackrel{d}{=} E_{i:n}$, where $\{E_i\}$ denotes a sequence of independent standard exponential r.v.'s, i.e., $P(E \le x) = 1 - \exp(-x)$, $x \ge 0$, and for intermediate k, $Y_{n-k:n} \sim n/k \to \infty$, as $n \to \infty$. Consequently, whenever we are under the first order framework in (1.1), we get

$$V_{ik} \stackrel{d}{=} \ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} = \ln \frac{U(Y_{n-k:n}\, Y_{k-i+1:k})}{U(Y_{n-k:n})} \sim \gamma\, E_{k-i+1:k} , \qquad 1 \le i \le k ,$$

i.e., $V_{ik},\ 1 \le i \le k$, are approximately the k o.s.'s from an exponential sample of size k and mean value γ. Also, since $\ln Y_{1:i} \stackrel{d}{=} E_{1:i} \stackrel{d}{=} E_i/i$,

$$U_i \stackrel{d}{=} i\left(\ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-i:n})}\right) = i\left(\ln \frac{U(Y_{n-i:n}\ Y_{1:i})}{U(Y_{n-i:n})}\right) \sim \gamma\, E_i\ , \qquad 1 \le i \le k\ ,$$

i.e., the U_i's, $1 \le i \le k$, are approximately independent and exponential with mean value γ. Then the Hill estimator of γ (Hill, 1975),

$$H(k) \equiv H_n(k) = \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \frac{1}{k}\sum_{i=1}^{k} U_i\ , \tag{1.10}$$

is consistent for the estimation of γ whenever (1.1) holds and k is intermediate, i.e., (1.7) holds.

Under the second order framework in (1.3) the asymptotic distributional representation

$$H(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{1}{1-\rho}\ A(n/k)\left(1 + o_p(1)\right) \tag{1.11}$$

holds true (de Haan and Peng, 1998), where $Z_k^{(1)} = \sqrt{k}\big(\sum_{i=1}^{k} E_i/k - 1\big)$ is an asymptotically standard normal r.v.

Remark 1.4. If the underlying model is the strict Pareto model in Remark 1.1, $\ln X_{i:n} = \gamma\, E_{i:n} + \gamma \ln C$, and the use of Rényi's representation of exponential order statistics, as a linear combination of independent unit exponential r.v.'s (Rényi, 1953), $E_{i:n} = \sum_{j=1}^{i} E_j/(n-j+1),\ 1 \le i \le n$, leads us to

$$H(k) \stackrel{d}{=} \frac{\gamma}{k}\sum_{i=1}^{k}\{E_{n-i+1:n} - E_{n-k:n}\} \stackrel{d}{=} \frac{\gamma}{k}\sum_{i=1}^{k}\sum_{j=i}^{k}\frac{E_j}{j} \stackrel{d}{=} \frac{\gamma}{k}\sum_{j=1}^{k} E_j \stackrel{d}{=} \frac{\gamma}{k}\, Ga(k)\ ,$$

where $Ga(k)$ denotes a Gamma r.v. with a shape parameter equal to k, i.e., a r.v. with probability density function (*p.d.f.*) $x^{k-1}\exp(-x)/\Gamma(k),\ x \ge 0$, with $\Gamma(t)$ denoting the complete gamma function, $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$. Then for every k, the Hill estimator in (1.10) is unbiased for the estimation of γ, i.e., $\mathbb{E}\,(H(k)) = \gamma$ for any k, and $\sqrt{k}\,(H(k)-\gamma)\,/\gamma$ is asymptotically standard normal, as $k \to \infty$.

We shall assume that we are in the class of models in (1.6). Consequently, we may choose

$$A(t) = \alpha\, t^\rho =: \gamma\,\beta\, t^\rho, \quad B(t) = \beta'\, t^\rho\ , \qquad \beta,\ \beta' \neq 0, \quad \rho < 0\ . \tag{1.12}$$

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years by several authors. The idea is to go further into the second order framework in (1.3). Then,

$$V_{ik} \stackrel{d}{=} \ln \frac{U(Y_{n-k:n}\, Y_{k-i+1:k})}{U(Y_{n-k:n})} \stackrel{d}{\approx} \gamma\, E_{k-i+1:k} + A(n/k)\,\frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}\ , \qquad 1 \le i \le k\,,$$

and

$$U_i \stackrel{d}{=} i \ln \frac{U(Y_{n-i:n}\, Y_{1:i})}{U(Y_{n-i:n})} \stackrel{d}{\approx} \gamma\left(1 + \frac{A(n/k)}{\gamma}\left(\frac{k}{i}\right)^{\rho}\right) E_i \ , \qquad 1 \le i \le k \ .$$

Beirlant *et al.* (1999) and Feuerverger and Hall (1999) work with the scaled log-spacings U_i, $1 \le i \le k$, in slightly different but equivalent ways, and consider the joint estimation of the first order parameter γ and the second order parameters at the same level k; in a similar set-up, Gomes and Martins (2002) advance with the "external" estimation of the second order parameter ρ, i.e., the estimation of ρ at a lower level (larger k) than the one used for the tail index estimation, being then able to reduce the asymptotic variance of the proposed tail index estimator, but they pay no special attention to the extra "scale" parameter $\beta \neq 0$ in the A function in (1.12). More recently, Gomes *et al.* (2004b) deal with a joint external estimation of both the "scale" and the "shape" parameters in the A function, being able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator, for an adequate choice of the level k. Such an estimator, also considered here for comparison with two new proposed estimators, is based on a linear combination of the excesses V_{ik} in (1.8), and is given by

$$WH_{\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{\widehat{\beta}\,(n/k)^{\widehat{\rho}}\,\left((i/k)^{-\widehat{\rho}}-1\right)/\left(\widehat{\rho}\ln(i/k)\right)}\, V_{ik} \ , \tag{1.13}$$

for adequate consistent estimators $\widehat{\beta}$ and $\widehat{\rho}$ of the second order parameters β and ρ, respectively, and with WH standing for *Weighted Hill*. In the same spirit, Gomes and Pestana (2004) study, mainly computationally, the estimator

$$\overline{H}_{\widehat{\beta},\widehat{\rho}}(k) := H(k) - \frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}} H\left(\left(\frac{(1-\widehat{\rho})^2\, n^{-2\widehat{\rho}}}{-2\,\widehat{\rho}\,\widehat{\beta}^2}\right)^{1/(1-2\widehat{\rho})}\right), \tag{1.14}$$

with H the Hill estimator in (1.10).

We shall here consider the estimator

$$\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k) := H(k)\left(1 - \frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right), \tag{1.15}$$

together with the asymptotically equivalent variant,

$$\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k) := H(k)\,\exp\left(-\frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right). \tag{1.16}$$

The dominant component of the bias of Hill's estimator, $A(n/k)/(1-\rho) = \gamma\,\beta\,(n/k)^{\rho}/(1-\rho)$, is thus estimated through $H(k)\,\widehat{\beta}\,(n/k)^{\widehat{\rho}}/(1-\widehat{\rho})$ and directly removed from Hill's classical tail index estimator, through two asymptotically equivalent expressions, provided that k is intermediate, i.e., provided that (1.7) holds true.

Remark 1.5. Note that the estimator $\overline{H}$ in (1.14) has been built in a way similar to the estimator $\widetilde{\overline{H}}$ in (1.15). The difference is that the bias $\gamma\,\beta\,(n/k)^{\rho}/(1-\rho)$ is estimated through $H\big(\widehat{k}_0\big)\,\widehat{\beta}\,(n/k)^{\widehat{\rho}}/(1-\widehat{\rho})$, with $\widehat{k}_0$ an estimate of the "optimal" level for the Hill estimator, in the sense of minimum mean squared error in Hall's class of models.

Remark 1.6. The reason to consider the two asymptotically equivalent estimators in (1.15) and (1.16) — also asymptotically equivalent to the estimator in (1.14) — lies in the fact that we have the clear experience that asymptotically equivalent estimators may exhibit quite different sample paths' properties. In practice, one never knows the peculiarities of the underlying models, and it is thus sensible to work with a set of a few estimators of the primary parameter of rare events, in order to take "the best decision".

1.1. A technical motivation

In the lines of Gomes and Martins (2004):

Lemma 1.1. *Under the second order framework in* (1.3)*, and for levels* k *such that* (1.7) *holds, the distributional representation*

$$\frac{\alpha}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}U_i \;\stackrel{d}{=}\; \gamma+\frac{\gamma\,\alpha}{\sqrt{(2\alpha-1)\,k}}\;Z_k^{(\alpha)}+\frac{\alpha\,A(n/k)}{\alpha-\rho}\,\big(1+o_p(1)\big) \tag{1.17}$$

holds true for any $\alpha\geq 1$*, where*

$$Z_k^{(\alpha)}=\sqrt{(2\alpha-1)\,k}\,\left(\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}E_i-\frac{1}{\alpha}\right) \tag{1.18}$$

are asymptotically standard normal r.v.'s. The asymptotic covariance structure between the r.v.'s in (1.18) *is given by*

$$\mathbb{C}ov_{\infty}\big(Z_k^{(\alpha)},\,Z_k^{(\beta)}\big)\;=\;\frac{\sqrt{(2\alpha-1)\,(2\beta-1)}}{\alpha+\beta-1}\;. \tag{1.19}$$

If we assume that only the tail index parameter γ in unknown, and similarly to the result in Gomes and Pestana (2004), we shall now state and prove a theorem that provides an obvious technical motivation for the estimator in (1.15) (or in (1.16)):

Theorem 1.1. *Under the second order framework in* (1.3)*, further assuming that* $A(t)$ *may be chosen as in* (1.12)*, and for levels* k *such that* (1.7) *holds, we get, for* $\widetilde{\overline{H}}_{\beta,\rho}(k)$ *in* (1.15) (*or for* $\overline{\overline{H}}_{\beta,\rho}(k)$ *in* (1.16))*, an asymptotic distributional representation of the type*

$$\gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + R_k\,, \qquad \textit{with} \;\; R_k = o_p\big(A(n/k)\big)\;, \tag{1.20}$$

where $Z_k^{(1)}$ *is the asymptotically standard normal r.v. in* (1.18) *for* $\alpha = 1$*. Consequently, both* $\sqrt{k}\,(\widetilde{\overline{H}}_{\beta,\rho}(k) - \gamma)$ *and* $\sqrt{k}\,(\overline{\overline{H}}_{\beta,\rho}(k) - \gamma)$ *are asymptotically normal with variance equal to* γ^2*, and with a null mean value not only when* $\sqrt{k}\;A(n/k) \longrightarrow 0$*, but also when* $\sqrt{k}\;A(n/k) \longrightarrow \lambda \neq 0$*, finite, as* $n \to \infty$*.*

Proof: The results related to the estimator in (1.15) come straighforwardly from the fact that if all parameters are known, apart from the tail index γ, we get from (1.11),

$$\begin{aligned}\widetilde{\overline{H}}_{\beta,\rho}(k) &\stackrel{d}{=} \left(\gamma + \frac{\gamma}{\sqrt{k}}\; Z_k^{(1)} + \frac{A(n/k)}{1-\rho}\,\big(1 + o_p(1)\big)\right) \times \left(1 - \frac{A(n/k)}{\gamma(1-\rho)}\right) \\ &\stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + o_p(A(n/k)\;,\end{aligned}$$

i.e., (1.20) holds. Since, for intermediate k,

$$\exp\!\left(-\frac{A(n/k)}{1-\rho}\right) = 1 - \frac{A(n/k)}{1-\rho} + o_p\big(A(n/k)\big)\;,$$

the same distributional representation in (1.20) holds true for the tail index estimator in (1.16). The remaining of the theorem follows then straightforwardly. □

1.2. A graphical motivation

For the second order parameters' estimators, discussed later on, in section 2, and as a supporting example of the technical motivation given in 1.1, we exhibit in Figure 1, the differences between the sample paths of the estimators $\widetilde{\overline{H}}_{\bullet}(k)$ in (1.15), for a sample of size $n = 10{,}000$ from a Fréchet model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, with $\gamma = 1$, when we compute $\widehat{\beta}$ and $\widehat{\rho}$ at the same level k used for the estimation of the tail index γ (*left*), when we compute only $\widehat{\beta}$ at that same level k, being $\widehat{\rho}$ computed at a larger k-value, let us say an intermediate level k_1 such that $\sqrt{k_1}\;A(n/k_1) \to \infty$, as $n \to \infty$ (*center*) and when both $\widehat{\rho}$ and $\widehat{\beta}$ are computed at that high level k_1 (*right*). For the notation used in Figure 1, see subsection 4.2. The high stability, around the target value

$\gamma = 1$, of the sample path in Figure 1 (*right*), is for sure related to the result in Lemma 1.1, and not purely coincidential. It thus seems sensible to compare asymptotically these estimation procedures, in order to detect the reasons for the differences in behaviour.

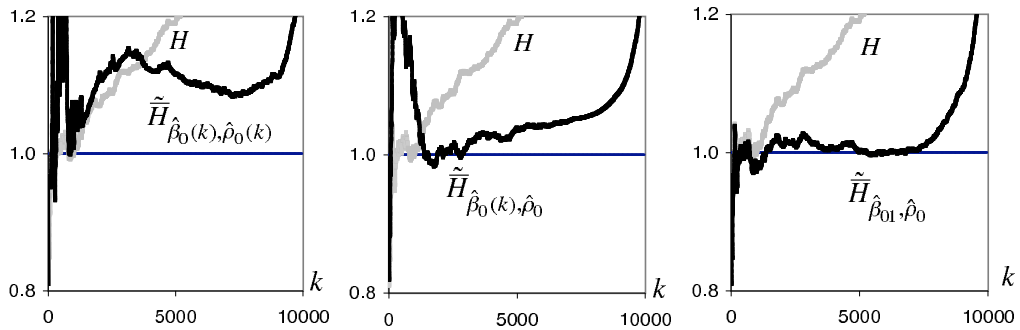


Figure 1: Sample paths of the Hill estimates H in (1.10) and the tail index estimates $\widetilde{\widetilde{H}}$ in (1.15), obtained through the estimation of (β, ρ) at the level $k_1 := \min(n-1,\ 2n^{0.995}/\ln\ln n)$ (*right*) versus the estimation at the same level both for β and ρ (*left*) and only for β (*center*).

1.3. Scope of the paper

When we look at the expression of the estimators from (1.13) till (1.16), we see that one of the topics to deal with is the adequate estimation of (β, ρ) in order to get the tail index estimators $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ and $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$. In section 2 of this paper, we shall thus briefly review the estimation of the two second order parameters β and ρ. Section 3 is devoted to the derivation of the asymptotic behaviour of the estimator $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ in (1.15) (equivalently, $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ in (1.16)), estimating β and ρ at a larger k value than the one used for the tail index estimation. We also do that only with the estimation of ρ, estimating β at the same level k used for the tail index estimation. In section 4, and through the use of simulation techniques, we shall exhibit the performance of the new estimators in (1.15) and (1.16), comparatively to the WH estimator in (1.13), to the classical Hill estimator and to $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$, for the β-estimator, $\widehat{\beta}_{\widehat{\rho}}(k)$, in Gomes and Martins (2002). We have here considered only an external estimation of the second order parameter ρ. Such a decision is related to the discussion in Gomes and Martins (2002) on the advantages of an external estimation of the second order parameter ρ (or even their misspecification, as in Gomes and Martins (2004)) versus an internal estimation at the same level k. Indeed, the estimation of γ, β and ρ at the same level k leads to very volatile mean values and mean squared error patterns. Finally, in section 5, some overall conclusions are drawn.

2. SECOND ORDER PARAMETER ESTIMATION

2.1. The estimation of ρ

We shall first address the estimation of ρ. We have nowadays two general classes of ρ-estimators, which work well in practice, the ones introduced in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003). We shall consider here particular members of the class of estimators of the second order parameter ρ proposed by Fraga Alves *et al.* (2003). Under adequate general conditions, they are semi-parametric asymptotically normal estimators of ρ, whenever $\rho < 0$. Moreover, for a large diversity of models, they give rise, for a wide range of large k-values, to highly stable sample paths, as functions of k, the number of top o.s.'s used. Such a class of estimators has been parameterised by a tuning parameter $\tau \geq 0$, and may be defined as

$$\widehat{\rho}_\tau(k) \equiv \widehat{\rho}_n^{(\tau)}(k) := -\left| \frac{3\left(T_n^{(\tau)}(k) - 1\right)}{T_n^{(\tau)}(k) - 3} \right|, \tag{2.1}$$

where

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^\tau - \left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}} & \text{if } \tau > 0 \\ \dfrac{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)} & \text{if } \tau = 0\,, \end{cases}$$

with

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\}^j, \quad j \geq 1 \qquad \left[M_n^{(1)} \equiv H \text{ in (1.10)}\right].$$

We shall here summarize a particular case of the results proved in Fraga Alves *et al.* (2003), now related to the asymptotic behaviour of the ρ-estimator in (2.1), under the second order framework in (1.3):

Proposition 2.1. *Under the second order framework in* (1.3)*, with* $\rho < 0$*, if* (1.7) *holds, and if* $\sqrt{k}\,A(n/k) \to \infty$*, as* $n \to \infty$*, the statistic* $\widehat{\rho}_n^{(\tau)}(k)$ *in* (2.1) *converges in probability towards* ρ*, as* $n \to \infty$*, for any* $\tau \in \mathbb{R}$*. More than this:* $\widehat{\rho}_n^{(\tau)}(k) - \rho = O_p\big(1/\big(\sqrt{k}\,A(n/k)\big)\big)$*, provided that, under the third order framework in* (1.5)*,* $\sqrt{k}\,A(n/k)\,B(n/k) \to \lambda_B$*, finite. If* $\sqrt{k}\,A(n/k)\,B(n/k) \to \infty$*, then* $\widehat{\rho}_n^{(\tau)}(k) - \rho = O_p\left(B(n/k)\right) = O_p\left(A(n/k)\right)$*.*

Remark 2.1. The theoretical and simulated results in Fraga Alves *et al.* (2003), together with the use of these estimators in the Generalized Jackknife statistics of Gomes *et al.* (2000), as done in Gomes and Martins (2002), as well as their use in the estimator in (1.13) (Gomes *et al.*, 2004*b*) and in the estimator in (1.14) (Gomes and Pestana, 2004), led us to advise in practice the consideration of the tuning parameters $\tau = 0$ for the region $\rho \in [-1, 0)$ and $\tau = 1$ for the region $\rho \in (-\infty, -1)$, together with the level $k_0 = \min(n-1, [2n/\ln\ln n])$. As done before, we however advise practitioners not to choose blindly the value of τ. It is sensible to draw a few sample paths of $\widehat{\rho}_\tau(k)$ in (2.1), as functions of k, electing the value of τ which provides higher stability for large k, by means of any stability criterion. For more details, see Gomes and Figueiredo (2003).

Remark 2.2. When we consider the level k_0 suggested in Remark 2.1, together with any of the ρ-estimators in this section, computed at that level k_0, $\{\widehat{\rho}(k_0) - \rho\}$ is of the order of $B(n/k_0) = (\ln\ln n)^\rho$, a very slow rate of convergence towards zero. We shall here work with a level

$$k_1 = \min\Big(n-1, \big[2n^{0.995}/\ln\ln n\big]\Big) \,. \tag{2.2}$$

Then $\widehat{\rho} - \rho = O_p((n^{0.005}\ln\ln n)^\rho)$ (provided that $\rho > -49.75$), and consequently, for any intermediate level k, $(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$, and $\sqrt{k}\,A(n/k)\,(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$ whenever $\sqrt{k}\,A(n/k) \to \lambda$, finite. This is going to be a fundamental result in the proof of Theorem 3.1, enabling the replacement of ρ by $\widehat{\rho}$, without disturbing the distributional result in Theorem 1.1, provided we estimate β adequately. In all the Monte Carlo simulations, we have considered the level k_1 in (2.2) and the following ρ-estimators in (2.1): $\widehat{\rho}_0 = \widehat{\rho}_0(k_1)$ if $\rho \geq -1$ and $\widehat{\rho}_1 = \widehat{\rho}_1(k_1)$ if $\rho < -1$.

2.2. Estimation of the second order parameter β

We have considered the estimator of β obtained in Gomes and Martins (2002) and based on the scaled log-spacings $U_i = i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$ in (1.9), $1 \leq i \leq k$. Let us denote $\widehat{\rho}$ any of the estimators in (2.1) computed at the level k_1 in (2.2). The β-estimator is given by

$$\widehat{\beta}_{\widehat{\rho}}(k) := \Big(\frac{k}{n}\Big)^{\widehat{\rho}} \frac{\Big(\frac{1}{k}\sum_{i=1}^{k}\big(\frac{i}{k}\big)^{-\widehat{\rho}}\Big)\Big(\frac{1}{k}\sum_{i=1}^{k}U_i\Big) - \Big(\frac{1}{k}\sum_{i=1}^{k}\big(\frac{i}{k}\big)^{-\widehat{\rho}}U_i\Big)}{\Big(\frac{1}{k}\sum_{i=1}^{k}\big(\frac{i}{k}\big)^{-\widehat{\rho}}\Big)\Big(\frac{1}{k}\sum_{i=1}^{k}\big(\frac{i}{k}\big)^{-\widehat{\rho}}U_i\Big) - \Big(\frac{1}{k}\sum_{i=1}^{k}\big(\frac{i}{k}\big)^{-2\widehat{\rho}}U_i\Big)} \,. \tag{2.3}$$

In Gomes and Martins (2002) and later in Gomes *et al.* (2004b), the following result has been proved:

Proposition 2.2. *If the second order condition* (1.3) *holds, with* $A(t) = \gamma\,\beta\,t^{\rho}$, $\rho < 0$, *if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.* (1.7) *holds, and if* $\sqrt{k}\,A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, *then* $\widehat{\beta}_{\widehat{\rho}}(k)$ *in* (2.3) *is consistent for the estimation of* β, *whenever* $\widehat{\rho} - \rho = o_p(1/\ln n)$. *Moreover, if* ρ *is known,*

$$\widehat{\beta}_{\rho}(k) \overset{d}{=} \beta + \frac{\gamma\,\beta\,(1-\rho)\,\sqrt{1-2\rho}}{\rho\,\sqrt{k}\,A(n/k)}\,W_k^B + R_k^B, \qquad \text{with } R_k^B = o_p(1) \tag{2.4}$$

and W_k^B *asymptotically standard normal. More precisely we may write*

$$W_k^B = \frac{(1-\rho)\,\sqrt{1-2\rho}}{|\rho|}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right), \tag{2.5}$$

with $Z_k^{(\alpha)}$, $\alpha \geq 1$, *given in* (1.18).

The asymptotic distributional representation (2.4) *holds true as well for* $\widehat{\beta}_{\widehat{\rho}}(k)$, *with* $\widehat{\rho}$ *any of the consistent* ρ*-estimators in* (2.1) *computed at the level* k_1 *in* (2.2). *If* $\sqrt{k}\,A(n/k)\,R_k^B \to \lambda_R^B$, *finite, we may further guarantee the asymptotic normality of* $\widehat{\beta}_{\widehat{\rho}}(k)$. *If we consider* $\widehat{\beta}_{\widehat{\rho}(k)}(k)$, *then*

$$\widehat{\beta}_{\widehat{\rho}(k)}(k) - \beta \overset{p}{\sim} -\beta\,\ln(n/k)\left(\widehat{\rho}(k) - \rho\right). \tag{2.6}$$

Remark 2.3. Note that when we consider the level k_1 in (2.2), the same restrictions for ρ as in Remark 2.2, and $\widehat{\beta} \equiv \widehat{\beta}_{\widehat{\rho}}(k_1)$, with $\widehat{\rho}$ any of the estimator in (2.1), computed also at the same level k_1, we may use (2.6) and derive that $\left\{\widehat{\beta} - \beta\right\}$ is of the order of $\ln(n/k_1)B(n/k_1) = O\left(\ln n\left(n^{0.005}\ln\ln n\right)^{\rho}\right)$. This result will also be needed in the proof of Theorem 3.1 and will enable us to keep the distributional result in Theorem 1.1.

Remark 2.4. Note also that if we estimate β through $\widehat{\beta}_{\widehat{\rho}}(k)$, since $\left\{\widehat{\beta}_{\widehat{\rho}}(k) - \beta\right\}$ is of the order of $1/\left(\sqrt{k}\,A(n/k)\right)$, we shall no longer be able to guarantee the distributional result in Theorem 1.1 (for details see Remark 3.2).

3. ASYMPTOTIC BEHAVIOUR OF THE ESTIMATORS

Let us assume first that we estimate both β and ρ externally at the level k_1 in (2.2). We may state the following:

Theorem 3.1. *Under the conditions of Theorem 1.1, let us consider the tail index estimators* $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ *and* $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ *in* (1.15) *and* (1.16)*, respectively, for any of the estimators* $\widehat{\rho}$ *and* $\widehat{\beta}$ *in* (2.1) *and* (2.3)*, respectively, both computed at the level* k_1 *in* (2.2) *and such that* $\widehat{\rho}-\rho = o_p(1/\ln n)$*. Then,* $\sqrt{k}\left\{\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)-\gamma\right\}$ *as well as* $\sqrt{k}\left\{\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)-\gamma\right\}$ *are asymptotically normal with variance equal to* γ^2 *and null mean value, not only when* $\sqrt{k}\ A(n/k)\to 0$*, but also whenever* $\sqrt{k}\ A(n/k)\to\lambda$*, finite, as* $n\to\infty$*.*

Proof: If we estimate consistently β and ρ through the estimators $\widehat{\beta}$ and $\widehat{\rho}$ in the conditions of the theorem, we may use Taylor's expansion series, and write,

$$\begin{aligned}\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k) \ &\stackrel{d}{=}\ H(k)\times\left(1-\frac{\beta}{1-\rho}\left(\frac{n}{k}\right)^{\rho}-\left(\widehat{\beta}-\beta\right)\frac{1}{1-\rho}\left(\frac{n}{k}\right)^{\rho}\left(1+o_p(1)\right)\right.\\ &\qquad\left.-\ \frac{\beta}{1-\rho}\left(\widehat{\rho}-\rho\right)\left(\frac{n}{k}\right)^{\rho}\left(\frac{1}{1-\rho}+\ln(n/k)\right)\left(1+o_p(1)\right)\right)\\ &\stackrel{d}{=}\ \widetilde{\overline{H}}_{\beta,\rho}(k)-\frac{A(n/k)}{1-\rho}\left(\frac{\widehat{\beta}-\beta}{\beta}+(\widehat{\rho}-\rho)\ \ln(n/k)\right)\left(1+o_p(1)\right)\ .\end{aligned}$$

Since $\widehat{\beta}$ and $\widehat{\rho}$ are consistent for the estimation of β and ρ, respectively, and $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$ (see Remark 2.2), the summands related to $\left(\widehat{\beta}-\beta\right)$ and $(\widehat{\rho}-\rho)$ are both $o_p(A(n/k))$, and the result in the theorem, related to the $\widetilde{\overline{H}}$-estimator, follows immediately, provided that $\sqrt{k}\ A(n/k)\to\lambda$, finite. The reasoning is exactly the same for the $\overline{\overline{H}}$-estimator. □

Remark 3.1. Note however that the levels k such that $\sqrt{k}\ A(n/k)\to\lambda$, finite, are sub-optimal for this type of estimators.

If we consider γ and β estimated at the same level, we are going to have an increase in the variance of our final tail index estimator $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ $\big($or equivalently, $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)\big)$. Similarly to Corollary 2.1 of Theorem 2.1 in Gomes and Martins (2002), there in connection with a ML-tail index estimator, as well as in Theorem 3.2 in Gomes *et al.* (2004b), in connection with the tail index estimator in (1.13), we may also get:

Theorem 3.2. *If the second order condition* (1.3) *holds, if* $k=k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and if* $\sqrt{k}\ A(n/k)\underset{n\to\infty}{\longrightarrow}\lambda$*, finite, non necessarily null, then*

$$\sqrt{k}\left(\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)-\gamma\right)\ \underset{n\to\infty}{\xrightarrow{d}}\ \mathit{Normal}\left(0,\ \sigma^2_{H_2}:=\gamma^2\left(\frac{1-\rho}{\rho}\right)^2\right)\ , \tag{3.1}$$

i.e., the asymptotic variance of $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ *increases of a factor* $((1-\rho)/\rho)^2$, *greater than one, for every* $\rho \leq 0$. *The same result holds obviously true for* $\widetilde{\overline{\overline{H}}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$.

Proof: If we consider

$$\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k) := H(k)\left(1-\frac{\widehat{\beta}_{\widehat{\rho}}(k)}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right),$$

we now get

$$\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k) = \widetilde{\overline{H}}_{\beta,\rho}(k) - \frac{A(n/k)}{1-\rho}\left(\frac{\widehat{\beta}_{\widehat{\rho}}(k)-\beta}{\beta}+(\widehat{\rho}-\rho)\ln(n/k)\right)\left(1+o_p(1)\right).$$

Since $\left(\widehat{\beta}_{\widehat{\rho}}(k)-\beta\right)/\beta$ is now of the order of $1/\left(\sqrt{k}\,A(n/k)\right)$, the term of the order of $1/\sqrt{k}$ is going to be, from (1.20), (2.4) and (2.5),

$$\frac{\gamma}{\sqrt{k}}\left(Z_k^{(1)}+\frac{(1-\rho)(1-2\rho)}{\rho^2}\left(\frac{Z_k^{(1)}}{1-\rho}-\frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right)\right),$$

which may be written as

$$\frac{\gamma}{\sqrt{k}}\left(\left(\frac{1-\rho}{\rho}\right)^2 Z_k^{(1)}-\frac{(1-\rho)\sqrt{1-2\rho}}{\rho^2}\,Z_k^{(1-\rho)}\right),$$

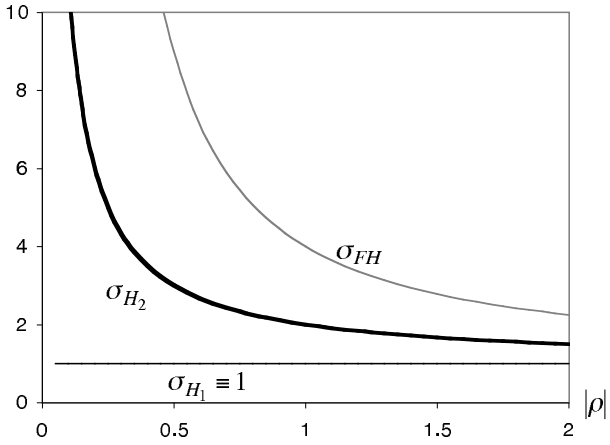
with $Z_k^{(\alpha)}$ the asymptotically standard normal r.v. in (1.18). Taking into account the fact that from (1.19), the asymptotic covariance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}\,/\,(1-\rho)$, together with the fact that $\sqrt{k}\;A(n/k)\,(\widehat{\rho}-\rho)\ln(n/k)\to 0$ (see Remark 2.2), (3.1) follows. □

Remark 3.2. If we compare Theorems 3.1 and 3.2 we see that the estimation of the two parameters γ and β at the same level k induces an increase in the asymptotic variance of the final γ-estimator of a factor given by $((1-\rho)/\rho)^2$, greater than 1 for all $\rho \leq 0$. As may be seen in Gomes and Martins (2002) the asymptotic variance of the estimator in Feuerverger and Hall (1999) (where the three parameters are computed at the same level k) is given by

$$\sigma_{FH}^2 := \gamma^2\left(\frac{1-\rho}{\rho}\right)^4.$$

In Figure 2 we provide both a picture and some values of $\sigma_{H_1}/\gamma \equiv 1$, σ_{H_2}/γ and σ_{FH}/γ, as functions of $|\rho|$.

It is obvious from Figure 2 that, whenever possible, it seems convenient to estimate both β and ρ "externally", at a k-value higher than the one used for the estimation of the tail index γ.



$\lvert\rho\rvert$	σ_{H_1}	σ_{H_2}	σ_{FH}
0.1	1.00	11.00	121.00
0.2	1.00	6.00	36.00
0.3	1.00	4.33	18.78
0.4	1.00	3.50	12.25
0.5	1.00	3.00	9.00
1.0	1.00	2.00	4.00
1.5	1.00	1.67	2.78
2.0	1.00	1.50	2.25
2.5	1.00	1.40	1.96
3.0	1.00	1.33	1.78

Figure 2: "Rulers" of the asymptotic standard deviations, σ_{H_1} and σ_{H_2} of the estimators under study, together with σ_{FH}, for $\gamma = 1$.

Remark 3.3. More generally, to obtain information on the asymptotic bias of $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$, $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ and $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\widehat{\rho}(k)}(k)$ — or equivalently, of $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$, $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\widehat{\rho}(k)}(k)$ — we need to go further into a third order framework, specifying, like has been done in (1.5), the rate of convergence in the second order condition in (1.3). This is however beyond the scope of this paper.

4. FINITE SAMPLE BEHAVIOUR OF THE ESTIMATORS

4.1. Underlying models

In this section we shall consider the following models in the class (1.6):

- the *Fréchet* model, with distribution function (d.f.) $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho = -1$;
- the *Generalized Pareto* (*GP*) model, with d.f. $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho = -\gamma$;
- the *Burr* model, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$, $\gamma > 0$, $\rho < 0$;
- the *Student*'s t_ν-model with ν degrees of freedom, with a probability density function (*p.d.f.*)

$$f_{t_\nu}(t) = \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\pi\nu}\ \Gamma(\nu/2)} \left[1 + \frac{t^2}{\nu}\right]^{-(\nu+1)/2}, \qquad t \in \mathbb{R} \quad (\nu > 0)\ ,$$

for which $\gamma = 1/\nu$ and $\rho = -2/\nu$.

4.2. The simulation design

We have here implemented multi-sample simulation experiments of size 50,000 = 5,000(*runs*) × 10(*replicates*), in order to obtain, for the above mentioned models, the distributional behaviour of the new estimators $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}$ and $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}$ in (1.15) and (1.16), respectively, based on the estimation of β at the level k_1 in (2.2), the same level we have used for the estimation of ρ, again not chosen in an optimal way. For details on multi-sample simulation, see Gomes and Oliveira (2001). We use the notation $\widehat{\beta}_{j1} = \beta_{\widehat{\rho}_j}(k_1)$, $j = 0, 1$, with $\widehat{\rho}_j$, $j = 0, 1$ and $\beta_{\widehat{\rho}}(k)$ given in (2.1) and (2.3), respectively. Similarly to what has been done in Gomes *et al.* (2004b) for the WH-estimator in (1.13), these estimators of ρ and β, for $j = 0, 1$, have been incorporated in the estimators under study, leading to $\widetilde{\overline{H}}_{\widehat{\beta}_{01},\widehat{\rho}_0}(k)$ / $\overline{\overline{H}}_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$ and to $\widetilde{\overline{H}}_{\widehat{\beta}_{11},\widehat{\rho}_1}(k) \,/\, \overline{\overline{H}}_{\widehat{\beta}_{11},\widehat{\rho}_1}(k)$, respectively. The simulations show that the tail index estimators $\widetilde{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}(k)$, j equal to either 0 or 1, according as $|\rho| \le 1$ or $|\rho| > 1$ seem to work reasonably well, as illustrated in the sequel. In the simulation we have also included the Hill estimator in (1.10), the "Weighted Hill" estimator in (1.13) and $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}$. The estimator $\overline{H}$ in (1.14) exhibits a behaviour quite similar to that of $\overline{\overline{H}}$ in (1.16), as may be seen from the results in Gomes and Pestana (2004), and was not pictured, for sake of simplicity.

We have simulated four different indicators. Let us denote generically $\widetilde{H}_n$ any of the estimators in (1.13), (1.15) and (1.16), and let

$$k_{0s}^{H}(n) \;:=\; \arg\min_{k} MSE_s\big[H_n(k)\big]$$

be the simulated optimal k (in the sense of minimum simulated mean squared error) for the Hill estimator $H_n(k) \equiv H(k)$ in (1.10). The two first indicators are related to the behaviour of the new estimators at Hill's optimal simulated level, i.e.,

$$REFF_n^{\widetilde{H}|H_0} \;:=\; \sqrt{\frac{MSE_s\Big[H_n\big(k_{0s}^{H}(n)\big)\Big]}{MSE_s\Big[\widetilde{H}_n\big(k_{0s}^{H}(n)\big)\Big]}} \tag{4.1}$$

and

$$BRI_n^{\widetilde{H}|H_0} \;:=\; \left|\frac{E_s\Big[H_n\big(k_{0s}^{H}(n)\big) - \gamma\Big]}{E_s\Big[\widetilde{H}_n\big(k_{0s}^{H}(n)\big) - \gamma\Big]}\right| . \tag{4.2}$$

The two additional indicators are related to the comparison of mean squared errors and bias of the new estimators with those of the Hill's estimators, when all the estimators are considered at their optimal levels. Denoting

$$H_{n0} := H_n\big(k_0^{H}(n)\big) \qquad \text{and} \qquad \widetilde{H}_{n0} := \widetilde{H}_n\big(k_0^{\widetilde{H}}(n)\big)\ ,$$

with the obvious meaning for $k_0^{\bullet}(n)$, the two extra simulated indicators are

$$REFF_{n0}^{\widetilde{H}|H} := \sqrt{\frac{MSE_s\big[H_{n0}\big]}{MSE_s\big[\widetilde{H}_{n0}\big]}} \tag{4.3}$$

and

$$BRI_{n0}^{\widetilde{H}|H} := \left| \frac{E_s\left[H_{n0} - \gamma\right]}{E_s\left[\widetilde{H}_{n0} - \gamma\right]} \right| . \tag{4.4}$$

Remark 4.1. Note that an indicator higher than one means a better performance than the Hill estimator. Consequently, the higher these indicators are, the better the new estimators perform, comparatively to the Hill estimator.

Remark 4.2. Note also that whereas we have appropriate techniques to deal with the estimation of the optimal level for Hill's estimator, in the sense of minimum mean squared error, we do not have yet equivalent techniques for the reduced bias' estimators. Consequently, the indicators in (4.3) and (4.4) are not useful in practice, but they give us an indication of the potentialities of this type of estimators.

4.3. Mean values and mean squared error patterns

In Figures from 3 till 9, and on the basis of the first replicate, with 5000 runs, we picture for the different underlying models considered, and a sample of size $n = 1000$, the mean values ($E[\bullet]$) and the mean squared errors ($MSE[\bullet]$) of the Hill estimator H, together with $\widetilde{H}_{\widehat{\beta}_{j1},\widehat{\rho}_j}$, $\overline{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}$, $WH_{\widehat{\beta}_{j1},\widehat{\rho}_j}$ and $\widetilde{\overline{H}}_{\widehat{\beta}_{j}(k),\widehat{\rho}_j}$, $j = 0$ or $j = 1$, according as $|\rho| \leq 1$ or $|\rho| > 1$. For comparison, we also picture the analogue behaviour of the r.v. $\widetilde{\overline{H}}_{\beta,\rho}$ for the models where there is a big discrepancy between the behaviour of the estimators and that of the r.v.'s. Such a discrepancy suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.

Remark 4.3. For a Burr model and for any of the estimators considered, $BIAS/\gamma$ and MSE/γ^2 are independent of γ, for every ρ. And we have seen no reason to picture the mean value and mean squared error patterns of the estimators for a GP underlying model because for all the estimators considered,

$$\mathbb{E}\,|\,GP(\gamma) \;=\; \gamma \times \mathbb{E}\,|\,BURR(\gamma = 1,\ \rho = -\gamma)$$

and

$$MSE\,|\,GP(\gamma) \;=\; \gamma^2 \times MSE\,|\,BURR(\gamma, \rho = -\gamma)\ .$$

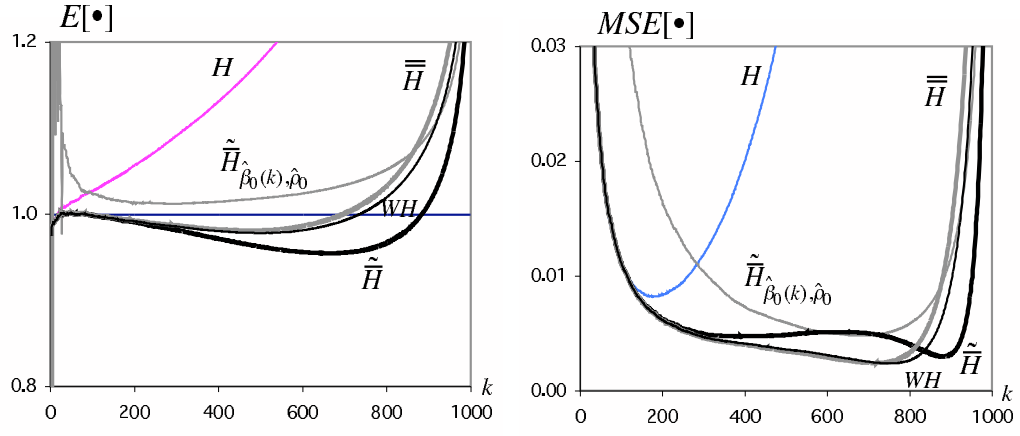


Figure 3: Underlying *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

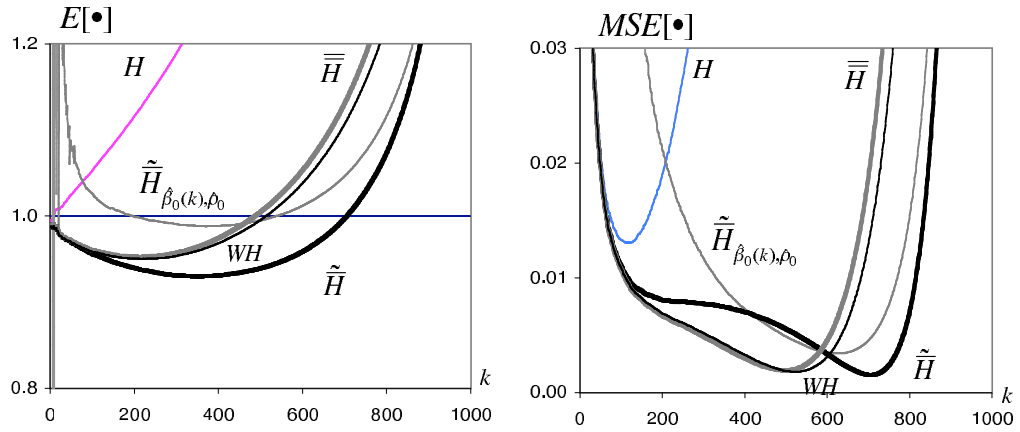


Figure 4: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -1$.

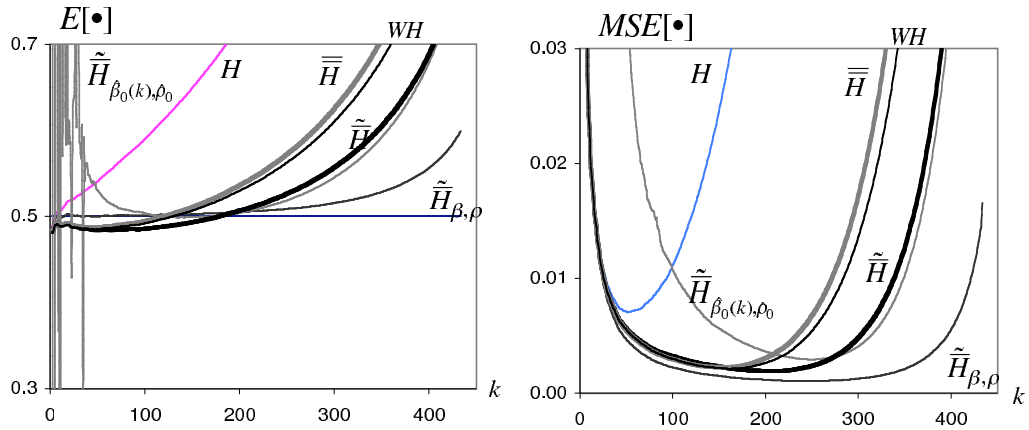


Figure 5: Underlying *Student* parent with $\nu = 2$ degrees of freedom ($\gamma = 0.5$ and $\rho = -1$).

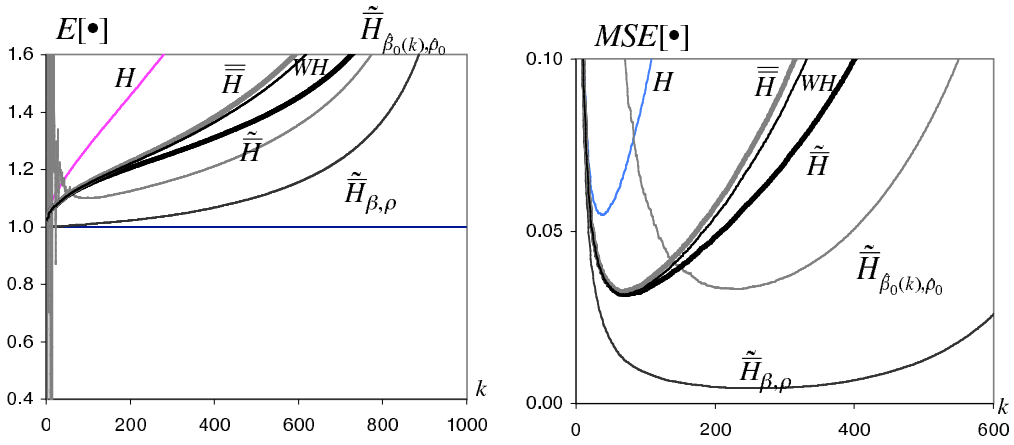


Figure 6: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

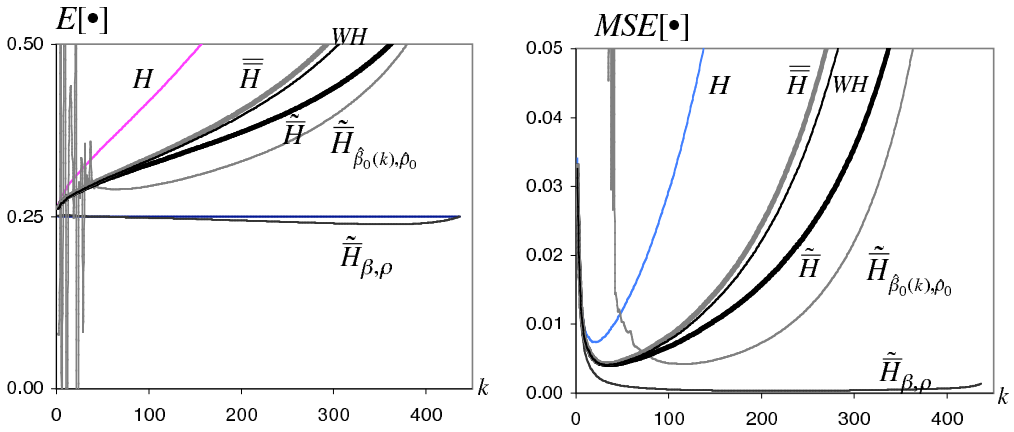


Figure 7: Underlying *Student* parent with $\nu = 4$ degrees of freedom ($\gamma = 0.25$ and $\rho = -0.5$).

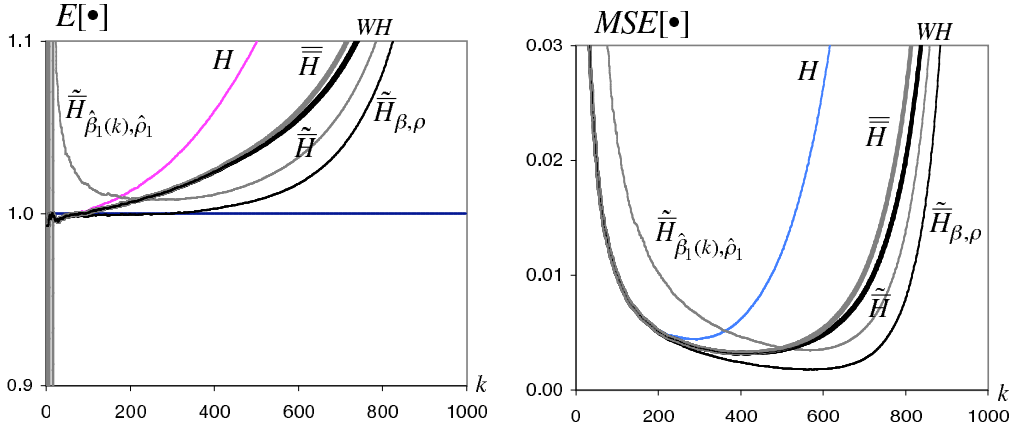


Figure 8: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -2$.

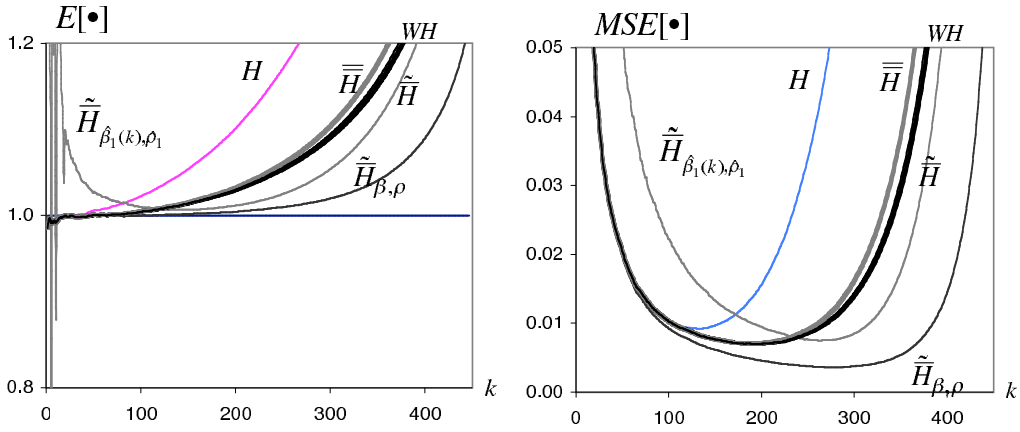


Figure 9: Underlying *Student* parent with with $\nu = 1$ degrees of freedom ($\gamma = 1$ and $\rho = -2$).

Remark 4.4. We may further draw the following specific comments:

- For a Fréchet model (with $\rho = -1$) (Figure 3) the bias of $\overline{\overline{H}}$ is the smallest one, being the one of $\widetilde{\overline{H}}$ the largest one, for small to moderate values of k. All the three reduced bias' statistics overestimate the true value of γ for very small as well as very large values of k, whereas, for moderate values of k, they all underestimate γ.

- For an underlying Burr parent with $\rho = -1$ (Figure 4), the three reduced bias' statistics are negatively biased for small values of k. Again, as for a Fréchet underlying parent, the $\overline{\overline{H}}$-statistic exhibits then the smallest bias and $\widetilde{\overline{H}}$ the largest one. The $\widetilde{\overline{H}}$ statistic is the best one regarding *MSE* at the optimal level, but the WH-statistic is the one with the smallest mean squared error for not too large values of k, followed by $\overline{\overline{H}}$. Quite similar results may be drawn for a Student model with $\rho = -1$ (Figure 5), but for this model, the mean squared error of $\widetilde{\overline{H}}$ is smaller than that of WH, which on its turn is smaller than that of $\overline{\overline{H}}$, for all values of k.

- For values of $\rho > -1$ (Figures 6 and 7), the three reduced bias' statistics are positively biased for all k. The $\widetilde{\overline{H}}$-statistic is better than the WH-statistic, which on its turn behaves slightly better than the $\overline{\overline{H}}$-statistic, both regarding bias and mean squared error.

- For $\rho < -1$ (Figures 8 and 9), we need to use $\widehat{\rho}_1$ (instead of $\widehat{\rho}_0$). In all the simulated cases the $\widetilde{\overline{H}}$ and the WH-statistics are the best ones and exhibit quite similar properties, but they are not a long way from the $\overline{\overline{H}}$-statistic.

Remark 4.5. For a Student model with ν degrees of freedom (Figures 5, 7 and 9), and whenever we assume β and ρ known, the most stable sample path around the target value γ is achieved by the statistic $\overline{\widetilde{H}}_{\beta,\rho}$, presented in the figures. And such a fact leads this statistic to have the smallest mean squared error, followed by the $\overline{\overline{H}}$ and next the WH statistics, for all values of ν. If we need to estimate β and ρ, the $\overline{\widetilde{H}}$-statistic is the one with the smallest mean squared error at the optimal level, also for every ν. Next comes the WH-statistic, quite close to the $\overline{\overline{H}}$-statistic when $\nu < 2$, i.e., when $\rho < -1$.

4.4. Relative efficiency and bias reduction indicators

In Table 1 we present the *REFF* indicators, in (4.1) and (4.3), and the *BRI* indicators in (4.2) and (4.4). For each model, the first, second and third rows are related to the WH-estimator in (1.13), the $\overline{\widetilde{H}}$-estimator in (1.15) and the $\overline{\overline{H}}$-estimator in (1.16), respectively. Each entry has two numbers: the first one is either the indicator in (4.1) or in (4.2) and the second one is either the indicator in (4.3) or in (4.4), according as we refer to the *REFF*-indicators (left hand-side table) or to the *BRI*-indicators (right hand-side table).

5. OVERALL CONCLUSIONS

- Generally, we may say that there is not a big difference between the estimators, WH, $\overline{H}$, $\overline{\widetilde{H}}$ and $\overline{\overline{H}}$ in (1.13), (1.14) (1.15) and (1.16), respectively. Anyway, whenever confronted with real data, the drawing of a few sample paths may help us in the choice of the most adequate estimate of the tail index γ.

- The $\overline{\widetilde{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}$ statistic may perhaps help us in the choice of the optimal sample fraction of Hill's estimator, and for some of the models exhibits sample paths more stable around the target value γ for a wider region of k-values. This is however a topic which deserves further investigation, being outside the scope of the present paper.

- The main advantage of these estimators lies on the fact that we may estimate β and ρ adequately through $\widehat{\beta}$ and $\widehat{\rho}$ so that the *MSE* of the new estimator is smaller than the *MSE* of Hill's estimator for all k, even when $|\rho| > 1$, a region where it has been difficult to find alternatives for the Hill estimator. And this happens together with a higher stability of the sample paths around the target value γ.

Table 1: *REFF* and *BRI* indicators.

***REFF* indicators**			***BRI* indicators**		
n			n		
200	500	1000	200	500	1000
Fréchet parent: $\rho = -1,\ \gamma = 1$					
1.07/1.53	1.11/1.69	1.12/1.86	4.56/25.40	7.16/13.12	11.50/10.22
1.06/1.39	1.10/1.53	1.12/1.67	4.18/45.93	6.33/31.87	9.99/69.15
1.08/1.52	1.12/1.67	1.12/1.85	5.85/5.56	8.92/47.70	14.28/22.89
Burr parent: $\rho = -0.5,\ \gamma = 1$					
1.20/1.39	1.26/1.35	1.23/1.33	1.83/1.32	1.76/1.23	1.67/1.24
1.19/1.39	1.26/1.35	1.22/1.33	1.73/1.25	1.73/1.22	1.64/1.24
1.18/1.35	1.25/1.32	1.22/1.31	1.69/1.30	1.69/1.20	1.62/1.22
Burr parent: $\rho = -1,\ \gamma = 1$					
1.19/2.09	1.23/2.42	1.23/2.69	2.25/11.26	1.91/10.31	1.70/30.21
1.18/2.27	1.22/2.63	1.21/2.94	2.14/306.29	1.74/22.55	1.55/28.80
1.20/2.02	1.24/2.34	1.24/2.61	2.67/9.35	2.07/24.51	1.79/13.63
Burr parent: $\rho = -2,\ \gamma = 1$					
1.05/1.18	1.08/1.18	1.11/1.21	2.75/1.08	2.29/1.12	2.09/1.10
1.05/1.17	1.08/1.18	1.10/1.21	2.60/1.09	2.25/1.10	2.07/1.09
1.05/1.16	1.08/1.17	1.10/1.20	2.54/1.10	2.22/1.10	2.05/1.06
Student parent: $\rho = -0.5,\ \gamma = 0.25$					
1.35/1.42	1.25/1.35	1.22/1.32	2.27/1.54	1.90/1.47	1.73/1.30
1.32/1.41	1.24/1.34	1.21/1.32	2.17/1.49	1.87/1.39	1.68/1.26
1.30/1.36	1.23/1.31	1.20/1.30	1.99/1.41	1.80/1.42	1.64/1.26
Student parent: $\rho = -1,\ \gamma = 0.5$					
1.07/1.56	1.02/1.70	1.15/1.86	3.17/3.40	1.68/3.94	1.25/4.85
1.03/1.51	1.02/1.75	1.14/1.97	2.77/7.88	2.03/7.15	1.25/8.71
1.08/1.51	1.02/1.65	1.16/1.82	4.58/2.84	2.26/3.31	1.24/4.65
Student parent: $\rho = -2,\ \gamma = 1$					
0.87/1.18	1.07/1.16	1.04/1.16	2.00/7.00	13.80/1.67	3.81/1.48
0.91/1.12	1.06/1.16	1.04/1.15	1.71/15.65	11.52/1.71	3.64/1.47
0.88/1.18	1.07/1.15	1.04/1.15	2.43/8.52	8.83/1.51	3.50/1.45

REFERENCES

[1] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, **2**, 177–200.

[2] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27**, 760–781.

[3] Fraga Alves, M.I.; Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, **60**:1, 193–213.

[4] Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

[5] Gomes, M.I.; Caeiro, F. and Figueiredo, F. (2004a). Bias reduction of a tail index estimator through an external estimation of the second order parameter. *Statistics*, **38**:6, 497–510.

[6] Gomes, M.I. and Figueiredo, F. (2003). Bias reduction in risk modelling: semi-parametric quantile estimation. Accepted at *Test*.

[7] Gomes, M.I.; Figueiredo, F. and Mendonça, S. (2005). Asymptotically best linear unbiased tail estimators under a second order regular variation condition. *J. Statist. Planning and Inference*, **134**:2, 409–433.

[8] Gomes, M.I.; Haan, L. de and Henriques, L. (2004b). *Tail index estimation through accommodation of bias in the weighted log-excesses*. Notas e Comunicações C.E.A.U.L. 14/2004. *Submitted*.

[9] Gomes, M.I.; de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour. *Extremes*, **5**:4, 387–414.

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

[11] Gomes, M.I. and Martins, M.J. (2004). Bias reduction and explicit estimation of the tail index. *J. Statist. Planning and Inference*, **124**, 361–378.

[12] Gomes, M.I.; Martins, M.J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology. *Extremes*, **3**:3, 207–229.

[13] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistical Extremes — choice of the optimal sample fraction. *Extremes*, **4**:4, 331–358.

[14] Gomes, M.I. and Pestana, D. (2004). A simple second order reduced bias' tail index estimator. Accepted at *J. Statist. Comp. and Simulation*.

[15] Haan, L. de and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, **52**, 60–70.

[16] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc.*, **B44**, 37–42.

[17] Hall, P. and Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.*, **13**, 331–341.

[18] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.

[19] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.*, **4**, 191–231.

REVSTAT – Statistical Journal
Volume 3, Number 2, November 2005, 137–153

MIXED EFFECTS IN STOCHASTIC DIFFERENTIAL EQUATION MODELS

Authors: Susanne Ditlevsen
– Department of Biostatistics, University of Copenhagen, Denmark
(sudi@pubhealth.ku.dk)

Andrea De Gaetano
– CNR IASI, Laboratorio di Biomatematica, Universita Cattolica del Sacro Cuore, Italy
(andrea.degaetano@biomatematica.it)

Received: January 2005 Revised: May 2005 Accepted: August 2005

Abstract:

- A class of statistical models is proposed where random effects are incorporated into a stochastic differential equations model, and an expression for the likelihood function is derived. In general, though, it is not possible to find an explicit expression for the likelihood function, but in a very simple example it is derived and explicit maximum likelihood estimators are found. The estimators are evaluated in a simulation study, and illustrated on dissolution data of metoprolol tartrate tablets.

Key-Words:

- *maximum likelihood; pharmacokinetics; population estimates; random effects; repeated measurements; stochastic processes.*

AMS Subject Classification:

- 60H10, 62M99.

1. INTRODUCTION

In biomedical research, studies in which repeated measurements are taken on a series of individuals or experimental animals play an important role. Models including random effects to model this kind of data enjoy an increasing popularity. In these models it is assumed that all responses follow a similar functional form, but with parameters that vary among individuals. The increasing popularity of mixed-effects models lies in the flexible modeling of correlation structures, where the total variation is specifically split in within-group and between-group variation. This will often lead to more precise estimation of population parameters. Especially in pharmacokinetic/pharmacodynamic (PK/PD) modeling most studies include random effects in the models, thereby improving population parameter estimation.

Continuous biological processes are often described by systems of ordinary differential equations (ODE), which unfortunately cannot account for noisy components often present in biological systems, representing the parts of the dynamics that we cannot predict or understand, or that we choose not to include in the explicit modeling. A natural extension is given by systems of stochastic differential equations (SDE), where system noise is modeled by including a diffusion term of some suitable form in the driving equations. In PK/PD modeling the focus is most often on the infinitesimal changes of substances, which naturally leads to a ODE-system. The inter-individual variability is modeled with the random effect, and the intra-individual variability with an additive noise term (possibly after some convenient transformation). However, noise in the differential equations describing the behavior of the system requires an extension of the model class to SDE models.

The theory for mixed-effects models is well developed for deterministic models (without system error), both linear and non-linear ([2, 3, 14, 25]), and standard software for model fitting is available, see e.g. ([18]) and references therein. Early and important references in the pharmacokinetic field are ([21, 22]). Estimating parameters in SDE models is not straightforward, except for simple cases. A natural approach would be likelihood inference, but the transition densities are rarely known, and thus it is usually not possible to write the likelihood function explicitly. A variety of methods for statistical inference in discretely observed diffusion processes has been developed during the past decades, see e.g. ([1, 4, 5, 6, 7, 9, 10, 13, 16, 17, 20, 23, 24]). However, to our knowledge there is practically no theory at present for SDE models with random effects. In ([15]) it is suggested to apply the Kalman filter to approximate the likelihood function for a SDE model with random effects, with a non-linear drift term and a constant diffusion term. Eventually, as SDE models will be more commonly applied to biomedical data, there will be an increasing need for developing a theory including mixed effects, and for results on the estimation of model parameters. In ([8]) methods for PK/PD population modeling are reviewed, but the authors regret

that system noise is not considered since it is difficult to estimate, and that there exists no software at present in the pharmacokinetic field.

In the present paper a class of statistical models is proposed where random effects are incorporated into a diffusion model, and an expression for the likelihood function is derived. In general, though, it is not possible to find an explicit expression for the likelihood function, but in a very simple example it is derived and explicit maximum likelihood estimators are found. The estimators are evaluated in a simulation study and illustrated on experimental data.

2. THE MODEL

Consider the one-dimensional SDE model for some continuous process evolving in M different subjects randomly chosen from a population:

$$\begin{aligned} dX_t^i &= g(X_t^i, \boldsymbol{\theta}, \mathbf{b}^i)\, dt + \sigma(X_t^i, \boldsymbol{\theta}, \mathbf{b}^i)\, dW_t^i \ ; \qquad i = 1, ..., M \\ \mathbf{b}^i &\sim N(0, \boldsymbol{\Sigma}) \\ X_0^i &= x_0^i \end{aligned} \tag{2.1}$$

where $\boldsymbol{\theta}$ is a p-dimensional *fixed effects* parameter (the same for the entire population) and $\mathbf{b}^i$ is a q-dimensional *random effects* parameter (subject specific), which is assumed to follow a normal distribution in the population, with covariance matrix $\boldsymbol{\Sigma}$ that is assumed known up to the parameter vector $\boldsymbol{\Psi}$. The W_t^i are standard Brownian motions. The W_t^i and $\mathbf{b}^j$ are assumed mutually independent for all $1 \le i, j \le M$, and independent of X_0^i. The drift and the diffusion coefficient functions $g(\cdot)$ and $\sigma(\cdot)$ are assumed known up to the parameters, and are assumed sufficiently regular to ensure a unique solution. Let $E \subseteq \mathbb{R}$ denote the state space of X_t^i. Assume that the distribution of X_t^i given $\mathbf{b}^i$ and $X_s^i = x, t > s$, has a strictly positive density w.r.t. the Lebesgue measure on E, which we denote by

$$y \mapsto p\big(y, x, t-s | \mathbf{b}^i, \boldsymbol{\theta}\big) > 0 \ , \qquad y \in E \ . \tag{2.2}$$

Assume the M subjects each are observed at the $(n_i + 1)$ discrete time points $(t_0^i, t_1^i, ..., t_{n_i}^i)$. Let $\mathbf{y}^i$ be the $(n_i + 1)$-dimensional response vector for the i'th subject: $\mathbf{y}^i = (y_0^i, ..., y_{n_i}^i)$, $y(t_j^i) = y_{t_j^i}^i = y_j^i$, and let $\mathbf{y}$ be the N-dimensional total response vector, $N = \sum_{i=1}^M (n_i+1)$. Write $t_j^i - t_{j-1}^i = \Delta_j^i$ for the distance between observation $j-1$ and j for subject i.

Parameters of the model are $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}$, which we wish to estimate.

3. MAXIMUM LIKELIHOOD ESTIMATION IN SDE MIXED EFFECTS MODELS

To obtain the marginal density, we integrate the conditional density of the data given the non-observable random effects $\mathbf{b}^i$ with respect to the marginal density of the random effects, using the fact that W_t^i and $\mathbf{b}^i$ are independent. This yields the likelihood

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) \;=\; \prod_{i=1}^{M} p(\mathbf{y}^i|\boldsymbol{\theta}, \boldsymbol{\Psi}) \;=\; \prod_{i=1}^{M} \int p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta})\, p(\mathbf{b}^i|\boldsymbol{\Psi})\, d\mathbf{b}^i \tag{3.1}$$

where $L(\cdot)$ is the likelihood and $p(\cdot)$ are densities. Now

$$p(\mathbf{y}^i|\mathbf{b}^i, \boldsymbol{\theta}) \;=\; \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\mathbf{b}^i, \boldsymbol{\theta}) \tag{3.2}$$

since X_t^i given $\mathbf{b}^i$ is Markov, where the transition densities are as in (2.2), and, by hypothesis,

$$p(\mathbf{b}^i|\boldsymbol{\Psi}) \;=\; \frac{\exp\left\{-(\mathbf{b}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{b}^i/2\right\}}{\sqrt{|\boldsymbol{\Psi}|}(2\pi)^{q/2}}\;, \tag{3.3}$$

where ^{T} denotes transposition. Substituting (3.2) and (3.3) into (3.1) we obtain

$$L(\boldsymbol{\theta}, \boldsymbol{\Psi}|\mathbf{y}) \;=\; \prod_{i=1}^{M} \int \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\mathbf{b}^i, \boldsymbol{\theta}) \frac{\exp\left\{-(\mathbf{b}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{b}^i/2\right\}}{\sqrt{|\boldsymbol{\Psi}|}(2\pi)^{q/2}}\, d\mathbf{b}^i\;. \tag{3.4}$$

Solving the integral yields the marginal likelihood of the parameters, independent of the random effects $\mathbf{b}^i$. Note how it is straightforward to generalize to other distributions for the random effects by letting $p(\mathbf{b}^i|\boldsymbol{\Psi})$ be any distribution depending on the parameter $\boldsymbol{\Psi}$. In general it will not be possible to find an explicit solution, but in simple cases we can find an explicit expression for the likelihood, and even find explicit estimating equations for the maximum likelihood estimators.

3.1. A random effect in Brownian motion with drift

In the simplest pharmacokinetic situation, the metabolism of a compound is modeled as a mono-exponential decay in the following way (first-order kinetics):

$$\frac{dC(t)}{dt} \;=\; -kC(t)\;; \qquad C(0) = D/V \tag{3.5}$$

with solution

$$C(t) \;=\; C(0)\, e^{-kt}$$

where $C(t)$ is the concentration of the compound in plasma at time t after a bolus injection, k is the (positive) rate elimination constant, D is the injected dose at time $t = 0$, and V the apparent volume of distribution of the compound. Now assume that we want to model the erratic behavior of the metabolic processes responsible for the removal of the compound from plasma, by allowing k to vary randomly as $k+\xi(t)$, where $\xi(t)$ is a white noise process. Then $\xi(t)\,dt = \sigma\,dW(t)$ where $W(t)$ is Brownian motion and σ a scaling parameter. Incorporating this into (3.5), writing $X_t = C(t)$ and $\beta = -k$, we obtain the equation

$$dX_t \;=\; \beta X_t\,dt + \sigma X_t\,dW_t \;,$$

which is the equation of geometric Brownian motion. The state space E is given by the positive real line. By applying Itô's formula to the transformation: $Y_t = \log X_t$, we obtain a Brownian motion with linear drift:

$$dY_t \;=\; \Big(\beta - \frac{1}{2}\sigma^2\Big)\,dt + \sigma\,dW_t$$

with solution

$$Y_t \;=\; Y_0 + \Big(\beta - \frac{1}{2}\sigma^2\Big)t + \sigma W_t \;.$$

Assume an experiment is conducted on different subjects where the concentration of a compound in plasma is measured at different time points after a bolus injection. We are interested in estimating the parameters in the population, but expect individual differences in the metabolic processes, and would therefore consider a random effect in β, which leads to the model:

$$\begin{aligned} Y_t^i \;&=\; Y_0^i + \Big(\beta + \beta^i - \frac{1}{2}\sigma^2\Big)t + \sigma W_t^i \\ \beta^i \;&\sim\; N(0,\sigma_\beta^2) \;. \end{aligned} \tag{3.6}$$

Another example where this model naturally arises is provided by the initial growth of bacterial or tumor cell populations, where we expect $\beta > 0$.

In this simple example we have $\boldsymbol{\theta} = (\beta, \sigma^2)$ and $\boldsymbol{\Psi} = \sigma_\beta^2$. We wish to estimate $\boldsymbol{\zeta} = (\beta, \sigma^2, \sigma_\beta^2)$. The conditional distribution $(Y_t^i | Y_0^i = y_0^i; \beta, \sigma^2, \beta^i)$ is Gaussian with

$$\begin{aligned} E\big[Y_t^i | Y_0^i = y_0^i; \beta, \sigma^2, \beta^i\big] \;&=\; y_0^i + \Big(\beta + \beta^i - \frac{1}{2}\sigma^2\Big)t \\ \mathrm{Var}\big[Y_t^i | Y_0^i = y_0^i; \beta, \sigma^2, \beta^i\big] \;&=\; \sigma^2 t \end{aligned}$$

so the conditional transition density is given by

$$p(y_j^i, y_{j-1}^i, \Delta_j^i; \beta, \sigma^2, \beta^i) = \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left\{ -\frac{\Big(y_j^i - y_{j-1}^i - \Big(\beta + \beta^i - \frac{1}{2}\sigma^2\Big)\Delta_j^i\Big)^2}{2\sigma^2\Delta_j^i} \right\}.$$

We will find the likelihood (3.4):

$$L(\boldsymbol{\zeta}|\mathbf{y}) \;=\; \prod_{i=1}^{M} \int p(\mathbf{y}^i | \beta^i, \beta, \sigma^2)\, p(\beta^i | \sigma_\beta^2)\; d\beta^i \;.$$

The computation would be much simplified if we assumed equidistant observations, that is $\Delta_j^i = \Delta$ for all times and subjects, but unfortunately this is rarely the case in real data. Not only will measurements often be taken with varying time gaps, but different subjects might be measured at different time points. In general $n_l \neq n_k$, and $\Delta_j^l \neq \Delta_j^k \neq \Delta_i^k$ (unbalanced data).

Due to the simplicity of the model, techniques adapted from linear regression with a random regression coefficient can be applied, (see e.g. [18]). Define the precision factor: $\eta^2 = \sigma^2/\sigma_\beta^2$. The conditional densities can be written as follows:

$$\begin{aligned} p(\mathbf{y}^i|\beta^i,\beta,\sigma^2) \;&=\; \prod_{j=1}^{n_i} p(y_j^i, y_{j-1}^i, \Delta_j^i|\beta^i,\beta,\sigma^2) \;= \\ &= \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2\Delta_j^i}} \exp\left\{-\frac{\left(y_j^i - y_{j-1}^i - \left(\beta+\beta^i-\tfrac{1}{2}\sigma^2\right)\Delta_j^i\right)^2}{2\sigma^2\Delta_j^i}\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \exp\left\{-\frac{\sum_j \frac{1}{\Delta_j^i}\left(y_j^i - y_{j-1}^i - \left(\beta+\beta^i-\tfrac{1}{2}\sigma^2\right)\Delta_j^i\right)^2}{2\sigma^2}\right\} \prod_{j=1}^{n_i} \frac{1}{\sqrt{\Delta_j^i}} \end{aligned}$$

and

$$p(\beta^i|\sigma_\beta^2) \;=\; \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left\{-\frac{(\beta^i)^2}{2\sigma_\beta^2}\right\} \;=\; \frac{(\eta^2)^{\frac{1}{2}}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\eta\beta^i)^2}{2\sigma^2}\right\}.$$

For ease of notation we define the parameter function $\alpha = \beta - \frac{1}{2}\sigma^2$ and the quantities $\Delta^i = \left(\prod_{j=1}^{n_i}\Delta_j^i\right)^{\frac{1}{n_i}}$ and $T^i = \sum_{j=1}^{n_i}\Delta_j^i$. The last sum is simply the length of the observation interval for the i'th subject; $t_{n_i} - t_0$. We obtain

$$L(\boldsymbol{\zeta}|\mathbf{y}) = \prod_{i=1}^{M} \frac{(\eta^2)^{\frac{1}{2}}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}} \int \frac{\exp\left\{-\dfrac{\sum_j \frac{1}{\Delta_j^i}\left(y_j^i - y_{j-1}^i - (\alpha+\beta^i)\Delta_j^i\right)^2 + (\eta\beta^i)^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}} \,\mathrm{d}\beta^i.$$

Solving the last integral yields the marginal likelihood of the parameters, independent of the random effects β^i. Define the vectors

$$\begin{aligned} \tilde{\mathbf{y}}^i \;&=\; \left((\Delta_1^i)^{-\frac{1}{2}}(y_1^i - y_0^i), ..., (\Delta_{n_i}^i)^{-\frac{1}{2}}(y_{n_i}^i - y_{n_i-1}^i), 0\right)^T \\ \tilde{\mathbf{x}}^i \;&=\; \left((\Delta_1^i)^{\frac{1}{2}}, ..., (\Delta_{n_i}^i)^{\frac{1}{2}}, 0\right)^T \\ \tilde{\mathbf{z}}^i \;&=\; \left((\Delta_1^i)^{\frac{1}{2}}, ..., (\Delta_{n_i}^i)^{\frac{1}{2}}, \eta\right)^T \end{aligned}$$

where ^{T} indicates transposition. Then

$$\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2 \;=\; \sum_{j=1}^{n_i} \frac{1}{\Delta_j^i}\left(y_j^i - y_{j-1}^i - (\alpha+\beta^i)\Delta_j^i\right)^2 + (\eta\beta^i)^2$$

such that, splitting the sum of squares into two parts that are independent of and dependent on the random effects, respectively, and noting that the integral of the dependent part is simply the integral of a normal density up to a constant, the likelihood function can be expressed as

$$
\begin{aligned}
L(\boldsymbol{\zeta}|\mathbf{y}) &= \prod_{i=1}^{M} \frac{(\eta^2)^{\frac{1}{2}}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}} \int \frac{\exp\left\{-\dfrac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}}\,\mathrm{d}\beta^i \\
&= \prod_{i=1}^{M} \frac{(\eta^2)^{\frac{1}{2}}}{(2\pi\sigma^2\Delta^i)^{\frac{n_i}{2}}}\ \exp\left\{-\frac{\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\hat{\beta}^i\|^2}{2\sigma^2}\right\} \frac{1}{\sqrt{T^i+\eta^2}} \\
&= \frac{(\eta^2)^{\frac{M}{2}}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^{M} \frac{1}{(\Delta^i)^{\frac{n_i}{2}}\sqrt{T^i+\eta^2}} \times \\
&\quad \exp\left\{-\frac{\sum_i \left(\sum_j \frac{1}{\Delta_j^i}\big(y_j^i - y_{j-1}^i - (\alpha+\hat{\beta}^i)\Delta_j^i\big)^2 + (\eta\hat{\beta}^i)^2\right)}{2\sigma^2}\right\} \\
&= \frac{(\eta^2)^{\frac{M}{2}}}{(2\pi\sigma^2)^{\frac{N-M}{2}}} \prod_{i=1}^{M} \frac{1}{(\Delta^i)^{\frac{n_i}{2}}\sqrt{T^i+\eta^2}} \times \\
&\quad \exp\left\{-\frac{\sum_{i,j}\frac{1}{\Delta_j^i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2 - \sum_i (y_{n_i}^i - y_0^i - \alpha T^i)^2 (T^i+\eta^2)^{-1}}{2\sigma^2}\right\}, \qquad (3.7)
\end{aligned}
$$

where $\hat{\beta}^i$ minimizes the sum of squares $\|\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha - \tilde{\mathbf{z}}^i\beta^i\|^2$ for fixed α, and is obtained from standard regression theory:

$$
\hat{\beta}^i = ((\tilde{\mathbf{z}}^i)^T\tilde{\mathbf{z}}^i)^{-1}(\tilde{\mathbf{z}}^i)^T(\tilde{\mathbf{y}}^i - \tilde{\mathbf{x}}^i\alpha) = \frac{\sum_{j=1}^{n_i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)}{\sum_{j=1}^{n_i}\Delta_j^i + \eta^2} = \frac{y_{n_i}^i - y_0^i - \alpha T^i}{T^i+\eta^2}.
$$

These directly provide predictors of the random effects given the parameters. The log-likelihood is

$$
\begin{aligned}
\log L(\boldsymbol{\zeta}|\mathbf{y}) &= \frac{M}{2}\log\eta^2 - \frac{N-M}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{M}\log\left((\Delta^i)^{n_i}(T^i+\eta^2)\right) \\
&\quad - \frac{\sum_{i,j}\frac{1}{\Delta_j^i}(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2 - \sum_i (y_{n_i}^i - y_0^i - \alpha T^i)^2 (T^i+\eta^2)^{-1}}{2\sigma^2}. \qquad (3.8)
\end{aligned}
$$

The derivatives of the log-likelihood function with respect to the parameters yield the *score functions* whose zeros will provide the maximum likelihood estimators of the parameters. Straightforward calculations yield the estimating equations,

where the estimator of a parameter is indicated with a hat, e.g. $\hat{\beta}$:

$$
\begin{aligned}
0 &= \sum_{i=1}^{M} \left(\frac{y_{n_i}^i - y_0^i - \hat{\alpha} T^i}{T^i + \hat{\eta}^2} \right) \\
0 &= \sum_{i=1}^{M} \left(\frac{\hat{\sigma_\beta}^2 T^i}{T^i + \hat{\eta}^2} - \frac{(y_{n_i}^i - y_0^i - \hat{\alpha} T^i)^2}{(T^i + \hat{\eta}^2)^2} \right) \\
0 &= \sum_{i=1}^{M} \left(\sum_{j=1}^{n_i} \left(\frac{(y_j^i - y_{j-1}^i - \hat{\alpha} \Delta_j^i)^2}{\Delta_j^i} \right) - \frac{(y_{n_i}^i - y_0^i - \hat{\alpha} T^i)^2}{T^i + \hat{\eta}^2} \right) - \hat{\sigma}^2 (N - M) \; .
\end{aligned}
$$

If we assume that each subject is observed in the same time interval, that is, we assume $T^i = T$ for all $1 \leq i \leq M$, this simplifies to the explicit estimators:

$$\hat{\beta} = \hat{\alpha} + \frac{\hat{\sigma}^2}{2} \tag{3.9}$$

$$\hat{\sigma}^2 = \frac{1}{N - 2M} \left((N - M) \mathrm{SSQ}_\Delta - M \mathrm{SSQ}_\mathrm{T} \right) \tag{3.10}$$

$$\hat{\sigma}_\beta^2 = \frac{N - M}{T(N - 2M)} \left(\mathrm{SSQ}_\mathrm{T} - \mathrm{SSQ}_\Delta \right) \tag{3.11}$$

where

$$\hat{\alpha} = \frac{1}{MT} \sum_{i=1}^{M} (y_{n_i}^i - y_0^i) \tag{3.12}$$

$$\mathrm{SSQ}_\mathrm{T} = \frac{1}{MT} \sum_{i=1}^{M} (y_{n_i}^i - y_0^i - \hat{\alpha} T)^2 \tag{3.13}$$

$$\mathrm{SSQ}_\Delta = \frac{1}{N - M} \sum_{i=1}^{M} \sum_{j=1}^{n_i} \left(\frac{\left(y_j^i - y_{j-1}^i - \hat{\alpha} \Delta_j^i \right)^2}{\Delta_j^i} \right) . \tag{3.14}$$

The asymptotic variances of the estimators estimated from the inverted Fisher information evaluated at the optimum is given by:

$$\hat{\mathrm{Var}}(\hat{\beta}) = \frac{\hat{\sigma}_\beta^2 T + \hat{\sigma}^2}{MT} + \frac{\hat{\sigma}^4}{2(N - 2M)} \tag{3.15}$$

$$\hat{\mathrm{Var}}(\hat{\sigma}) = \frac{\hat{\sigma}^2}{2(N - 2M)} \tag{3.16}$$

$$\hat{\mathrm{Var}}(\hat{\sigma}_\beta) = \frac{(\hat{\sigma}_\beta^2 T + \hat{\sigma}^2)^2}{2MT^2 \hat{\sigma}_\beta^2} + \frac{\hat{\sigma}^4}{2(N - 2M) T^2 \hat{\sigma}_\beta^2} \; . \tag{3.17}$$

There will only be positive solutions for the variance parameters in the data set if

$$\frac{M}{N - M} \mathrm{SSQ}_\mathrm{T} < \mathrm{SSQ}_\Delta < \mathrm{SSQ}_\mathrm{T} \; . \tag{3.18}$$

The last inequality ensures existence of the estimator of the random effect variance parameter σ_β^2, and can be interpreted in the following way: For simplicity assume $\Delta_j^i = \Delta$ for all i, j. Define $a_j^i = (y_j^i - y_{j-1}^i - \hat{\alpha}\Delta)$, the increment for subject i from observation $j-1$ to observation j subtracted the expected increment in the population. Then

$$\mathrm{SSQ}_\mathrm{T} = \frac{1}{MT}\sum_{i=1}^{M}\left(\sum_{j=1}^{n_i} a_j^i\right)^2 \quad \text{and} \quad \mathrm{SSQ}_\Delta = \frac{1}{MT}\sum_{i=1}^{M}\sum_{j=1}^{n_i}(a_j^i)^2 \ .$$

For SSQ_Δ to be smaller than SSQ_T, it is required that at least for one i, $\sum_{j=1}^{n_i}(a_j^i)^2 < (\sum_{j=1}^{n_i} a_j^i)^2$, which e.g. will be the case if all a_j^i are of the same sign. If this is the case it means that all observed increments are either above or under the expected increments for the population, which indicates that the decay rate for this specific subject most probably is different from the general population decay rate β, that is $\beta^i \neq 0$. On the other hand, to estimate the system noise parameter σ^2, we require $\frac{\Delta}{T}\mathrm{SSQ}_\mathrm{T} < \mathrm{SSQ}_\Delta$. The left hand side increases when the number of measured points for each subject decreases. In this case it is natural that we have more information on variation between subjects than variation within subjects.

Considering model (3.6) with $\sigma_\beta^2 = 0$, such that $\beta^i = 0$ for all i (no random effects), leads to the log-likelihood function

$$\log L(\beta, \sigma^2|\mathbf{y}) \ = \ -\frac{N-M}{2}\log(2\pi\sigma^2) - \sum_i \frac{n_i}{2}\log(\Delta^i) - \sum_{i,j}\frac{(y_j^i - y_{j-1}^i - \alpha\Delta_j^i)^2}{2\sigma^2\Delta_j^i}$$

which could also be derived from (3.8) by letting $\eta^2 \to \infty$. This leads to the maximum likelihood estimators

$$\hat{\beta} \ = \ \hat{\alpha} + \hat{\sigma}^2/2 \tag{3.19}$$

$$\hat{\sigma}^2 \ = \ \mathrm{SSQ}_\Delta \ . \tag{3.20}$$

The asymptotic variances of the estimators estimated from the inverted Fisher information evaluated at the optimum is given by:

$$\hat{\mathrm{Var}}(\hat{\beta}) \ = \ \frac{\hat{\sigma}^2}{MT} + \frac{\hat{\sigma}^4}{2(N-M)} \tag{3.21}$$

$$\hat{\mathrm{Var}}(\hat{\sigma}) \ = \ \frac{\hat{\sigma}^4}{2(N-M)} \ . \tag{3.22}$$

3.2. Simulation results

To check the estimators a simulation study was performed. Six sets of parameter values were used to investigate the behavior of the estimators for different relations among the variance components, namely for $\sigma^2 \gg \sigma_\beta^2$, $\sigma^2 \approx \sigma_\beta^2$, and $\sigma^2 \ll \sigma_\beta^2$, respectively, and for two different values of β, consistent with physiologically observed decay rate values. Moreover, two sets of values for the experimental designs were investigated, namely for $M \gg n$ and $M \ll n$, respectively. The values used in the different simulations are reported in Table 1.

Table 1: Values used in the different simulations.

	Parameter values used in simulations				
	β	σ^2	σ_β^2	M	n
1	-0.02	0.02	0.02	10	50
2	-0.02	0.2	0.02	10	50
3	-0.02	0.02	0.2	10	50
4	-0.02	0.02	0.02	50	10
5	-0.02	0.2	0.02	50	10
6	-0.02	0.02	0.2	50	10
7	-0.2	0.02	0.02	10	50
8	-0.2	0.2	0.02	10	50
9	-0.2	0.02	0.2	10	50
10	-0.2	0.02	0.02	50	10
11	-0.2	0.2	0.02	50	10
12	-0.2	0.02	0.2	50	10

For each of these 12 sets of values, 1.000 data sets were generated from model (3.6), by simulating trajectories according to the Milstein scheme with a step size of 0.01, see Kloeden and Platen (1999), and retaining the observation points at equidistant time points depending on the chosen n. For all simulations the total length of the simulation interval was 100, and the initial value was $\log(100)$. On the simulated data sets, parameters were estimated using Equations (3.9) to (3.14). Parameters were also estimated assuming (wrongly) the model with no random effects by Equations (3.19) and (3.20). Results are reported in Table 2, where the 95% confidence intervals are the 2.5% and 97.5% empirical quantiles of estimates, and are given in brackets.

In all 12.000 simulations the estimators existed ($\hat{\sigma}^2, \hat{\sigma}_\beta^2 > 0$), but for $\beta = -0.02$ a considerable part of the estimates were positive, reflected in the large 97.5% quantiles for $\hat{\beta}$. Not surprisingly, β is more difficult to estimate when σ_β^2 is large. The diffusion parameter σ^2 is well determined with 95% of estimates lying

no more than 11% from the true value, whereas σ_β^2 is more difficult to estimate and depends on the size of M: for small M, the distribution of estimates is right-skewed with wide confidence limits; for larger M, σ_β^2 is better determined.

Table 2: Mean of estimates (95% CI) from simulations of model (3.6). For each set of values, 1.000 data sets were generated and the parameters were estimated using (3.9) to (3.11) (assuming random effects) and (3.19) and (3.20) (assuming no random effects). For all simulations $T = 100$ and $Y_0^i = \log(100)$. See also main text.

	Assuming random effects		
	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_\beta^2$
1	-0.018 (-0.094;0.060)	0.020 (0.018;0.022)	0.018 (0.006;0.035)
2	-0.019 (-0.097;0.056)	0.200 (0.178;0.222)	0.018 (0.006;0.034)
3	-0.027 (-0.263;0.210)	0.020 (0.018;0.022)	0.178 (0.063;0.317)
4	-0.021 (-0.054;0.012)	0.020 (0.018;0.022)	0.020 (0.013;0.027)
5	-0.021 (-0.057;0.017)	0.200 (0.178;0.221)	0.020 (0.013;0.028)
6	-0.020 (-0.119;0.080)	0.020 (0.018;0.022)	0.197 (0.137;0.267)
7	-0.199 (-0.276;-0.126)	0.020 (0.018;0.022)	0.017 (0.006;0.034)
8	-0.198 (-0.281;-0.123)	0.201 (0.180;0.223)	0.017 (0.005;0.035)
9	-0.198 (-0.440;0.029)	0.020 (0.018;0.022)	0.175 (0.066;0.337)
10	-0.200 (-0.233;-0.167)	0.020 (0.018;0.022)	0.020 (0.013;0.026)
11	-0.201 (-0.236;-0.163)	0.200 (0.178;0.222)	0.020 (0.013;0.027)
12	-0.202 (-0.302;-0.096)	0.020 (0.018;0.022)	0.195 (0.135;0.260)

	Assuming no random effects (wrong model)		
	$\hat{\beta}$	$\hat{\sigma}^2$	-
1	0.000 (-0.076;0.079)	0.057 (0.033;0.091)	
2	-0.001 (-0.082;0.076)	0.236 (0.203;0.275)	
3	0.155 (-0.089;0.404)	0.384 (0.151;0.667)	
4	0.087 (0.039;0.141)	0.237 (0.166;0.315)	
5	0.088 (0.033;0.144)	0.418 (0.339;0.510)	
6	1.075 (0.729;1.458)	2.209 (1.543;2.985)	
7	-0.181 (-0.260;-0.107)	0.056 (0.033;0.089)	
8	-0.181 (-0.264;-0.105)	0.236 (0.202;0.277)	
9	-0.020 (-0.292;0.257)	0.377 (0.156;0.708)	
10	-0.092 (-0.138;-0.042)	0.237 (0.170;0.310)	
11	-0.092 (-0.142;-0.039)	0.418 (0.343;0.499)	
12	0.884 (0.547;1.273)	2.192 (1.519;2.907)	

If a model with no random effects is wrongly assumed, both β and σ^2 are poorly estimated. The estimates are worse for large σ_β^2 and large M, as expected. This illustrates the need to include random effects in the modelling process if they are present in the data.

3.3. Application to Metoprolol Tartrate dissolution data

The method was applied on Metoprolol Tartrate dissolution data taken from [19], where the percentage of released drug of four types of tablet formulations of 100-mg Metoprolol Tartrate are tabulated at 5-min intervals up to 30 minutes and at 45 minutes after the onset of the experiments, except for the Slow Dissolving Test Formulation, where measurements were taken up to two hours, for details see [19]. Each experiment was repeated six times. The data were also analysed in [12, 13]. In [12], they found that the formulation closest to an exponential behavior was the Slow Dissolving Test Formulation, which is used here to illustrate the methods. Only data up to 45 minutes are used.

The data are illustrated in Figure 1. The percentage of Metoprolol not yet dissolved is modeled as (3.6), where y_j^i are the log-transformed measured percentages for experiment i at time point j. Moreover, the measurement at 30 minutes for experiment four was removed in the analysis since the dissolution process cannot go backwards, see Figure 1. Finally $M = 6$ and $N = 41$.

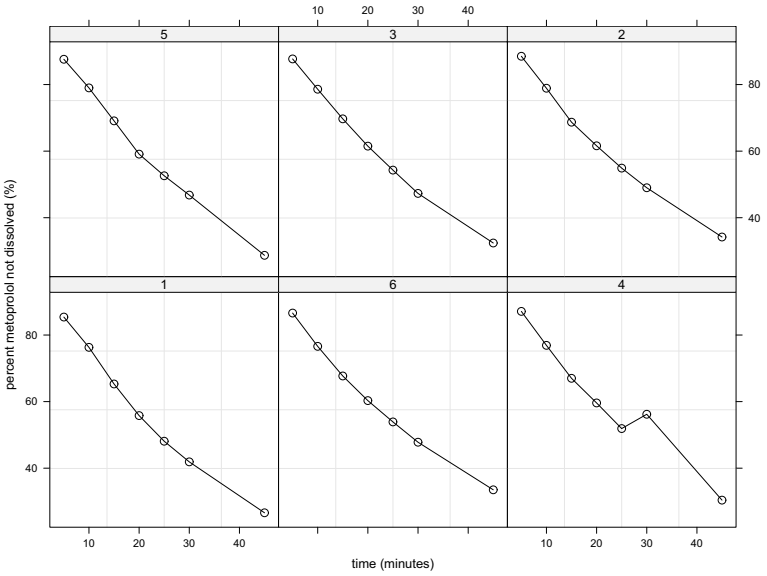


Figure 1: Dissolution profiles of Metoprolol Tartrate tablets. Data is taken from [19].

The data set yields the following quantities: $\hat{\alpha} = -0.026, \mathrm{SSQ}_{\mathrm{T}} = 0.000166$ and $\mathrm{SSQ}_{\Delta} = 0.0000699$ such that condition (3.18) is fulfilled. Estimates and their standard errors are reported in Table 3. The estimates of β are in agreement with comparable values found in [12, 13, 19]. Since $\hat{\sigma}_\beta^2$ is small compared to $\hat{\sigma}^2$, the estimates in the model without random effects only change slightly.

Table 3: Metoprolol data estimates using Equations (3.9) to (3.11) (assuming random effects) and (3.19) and (3.20) (assuming no random effects). The standard errors were estimated using Equations (3.15), (3.16), (3.17), (3.21) and (3.22).

	Assuming random effects		**Assuming no random effects**	
	estimate	std error	estimate	std error
$\hat{\beta}$	-0.02594	0.00083	-0.02593	0.00054
$\hat{\sigma}$	0.00707	0.00093	0.00836	0.00001
$\hat{\sigma}_\beta$	0.00171	0.00071	-	-

Figure 2 shows simulated trajectories from the random effects model with the estimated parameters, and the observed points from two of the six dissolution profiles.

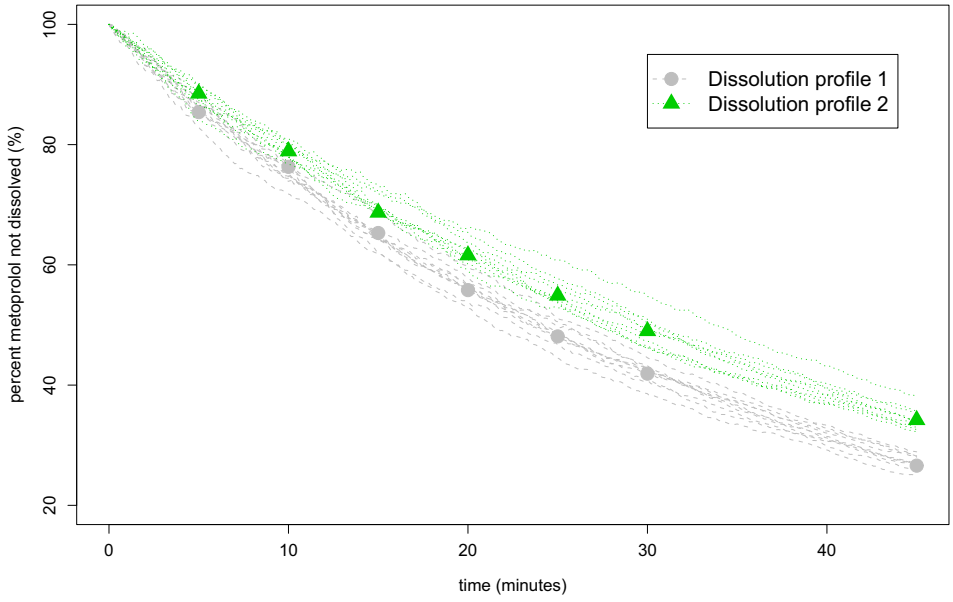


Figure 2: Simulated trajectories from model (3.6), incorporating the estimated parameters and random effect estimates for two of the dissolution profiles. The points are the observed data for the same two dissolution profiles.

4. SUMMARY

In the present paper we propose to extend random effects techniques to the estimation of parameters in SDE models. We believe this extension to be both relevant and needed. It is relevant because as the sophistication of builders and users of mathematical models of biological processes increases, there will be a progressive growth of the use of stochastic differential equations to represent noisy processes. When only few observations can be collected from any given human or animal experimental subject, as is usually the case, recourse to random or mixed effects models will be necessary.

Statistical inference for this class of models is not straightforward. In the present work, a very simple model gave rise to explicit expressions for the likelihood function and for the maximum likelihood estimators. This model is in its deterministic version frequently employed in pharmacokinetics (e.g. to represent drug elimination from plasma or initial tumor cell population growth), and the proposed development is therefore not only of academic interest. However, it is often the case that more complicated models with nonlinearities and/or several compartments are necessary to plausibly represent the system under observation.

Unfortunately, in general it will not be possible to find an explicit expression for the likelihood function (3.4) since the transition densities are rarely known. One possibility could be to approximate the likelihood function numerically, and then optimize the approximated likelihood function directly. It is obviously necessary to find other estimation procedures if the proposed model class is to be of interest to a wider audience.

ACKNOWLEDGMENTS

Work supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2000-00100, DYNSTOCH.

REFERENCES

[1] Bibby, B.M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 1/2, 017–039.

[2] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *JASA*, **88**, 421, 9–25.

[3] Diggle, P.J.; Heagerty, P.; Liang, K.-Y. and Zeger, S.L. (2002). *Analysis of Longtitudinal Data*, Oxford University Press, 2nd edition.

[4] Ditlevsen, S. and Sørensen, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist*, **31**, 3, 417–429.

[5] Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bulletin of Mathematical Biology*, **67**, 547–561.

[6] Ditlevsen, S.; Yip, K.P. and Holstein-Rathlou, N.H. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Mathematical Biosciences*, **194**, 49–69.

[7] Elerain, O.; Chib, S. and Shepard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, **69**, 959–993.

[8] Jelliffe, R.; Schumitzky, A. and Van Guilder, M. (2000). Population pharmacokinetics/pharmacodynamics modeling: Parametric and nonparametric methods. *Therapeutic Drug Monitoring*, **22**, 354–365.

[9] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist*, **24**, 211–229.

[10] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 2, 299–314.

[11] Kloeden, P.E. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin and Heidelberg GmbH & Co. K.

[12] Lansky, P. and Weiss, M. (2003). Classification of dissolution profiles in terms of fractional dissolution rate and a novel measure of heterogeneity. *J. Pharmaceutical Sciences*, **92**, 1632–1647.

[13] Lansky, P.; Lanska, V. and Weiss, M. (2004). A stochastic differential equation model for drug dissolution and its parameters. *J. Controlled Release*, **100**, 267–274.

[14] Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673–687.

[15] Overgaard, R.V.; Jonsson, N.; Tornøe, C.W. and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations. Implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, **32**, 85–107.

[16] Pedersen, A.R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**, 1, 55–71.

[17] Pedersen, A.R. (2001). *Likelihood inference by Monte Carlo methods for incompletely discretely observed diffusion processes*, Technical Report 1, Department of Biostatistics, Univ. of Aarhus.

[18] Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York.

[19] Polli, J.E.; Singh Rekhi, G.; Augsburger, L.L. and Shah, V.P. (1997). Methods to compare dissolution profiles and a rationales for wide dissolution specifications for metoprolol tartrate tablets. *J. Pharmaceutical Sciences*, **86**, 6, 690–700.

[20] Prakasa Rao, B.L.S. (1999). *Statistical Inference for Diffusion Type Processes*, Arnold Publishers.

[21] Sheiner, L.B. and Beal, S.L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, **8**(6), 553–571.

[22] Sheiner, L.B. and Beal, S.L. (1981). Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, **9**(5), 635–651.

[23] Sørensen, H. (2003). Simulated likelihood approximations for stochastic volatility models. *Scand. J. Statist.*, **30**, 257–276.

[24] Sørensen, M. (2000). Prediction-based estimating functions. *Econometrics J.*, **3**, 123–147.

[25] Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.

REVSTAT – Statistical Journal
Volume 3, Number 2, November 2005, 155–170

A NON-PARAMETRIC TEST FOR NON-INDEPENDENT NOISES AGAINST A BILINEAR DEPENDENCE

Authors: E. Gonçalves
– Departamento de Matemática, Universidade de Coimbra, Portugal
(esmerald@mat.uc.pt)

P. Jacob
– Université de Montpellier II, France
(jacob@math.univ-montp2.fr)

N. Mendes-Lopes
– Departamento de Matemática, Universidade de Coimbra, Portugal
(nazare@mat.uc.pt)

Received: March 2005 Revised: June 2005 Accepted: August 2005

Abstract:

- A new methodology, based on the asymptotic separation of probability laws, was introduced by Gonçalves, Jacob and Mendes-Lopes (2000) in the development of the statistical inference of bilinear models, namely in the construction of a consistent decision procedure for the simple bilinear ones.

  This paper presents a generalisation of that study by introducing in the procedure a smoother decision statistics.

  The aim of this decision method is to discriminate between an error process and a simple bilinear model. So, we use it as a consistent test, its consistence being obtained by establishing the asymptotic separation of the sequences of probability laws defined by each hypothesis.

  The convergence rate of the procedure is studied under the truthfulness of the error process hypothesis. An exponential decay is obtained.

Key-Words:

- *time series; asymptotic separation; bilinear models; test.*

AMS Subject Classification:

- 62M10, 62F03.

1. INTRODUCTION

In Gonçalves, Jacob and Mendes-Lopes (2000) a new methodology of statistical decision to discriminate between an error process and a diagonal simple bilinear model was presented. This methodology was inspired in an asymptotic separation result obtained, in 1976, by Geffroy, which appeared particularly useful to construct consistent tests and estimators for detecting a signal in a white noise (Pieczinsky, 1986, Moché, 1989).

Let $X = (X_t, t \in \mathbb{Z})$ be a real stochastic process whose law belongs to a set of parametric laws $(P_\theta, \theta \in \Theta)$, with $\Theta = \{\theta_1, \theta_2\}$. Following Geffroy (1976), we say that the two laws P_{θ_1} and P_{θ_2} are asymptotically separated if there exists a sequence of Borel sets of $\mathbb{R}^T$, $(A_T, T \in \mathbb{N})$, such that

$$\begin{cases} P_{\theta_0}^T(A_T) \underset{T\to+\infty}{\longrightarrow} 1 \\ P_{\theta_1}^T(A_T) \underset{T\to+\infty}{\longrightarrow} 0 \ , \end{cases}$$

where P_θ^T denotes the probability law of $(X_1, X_2, ..., X_T)$.

In this way, a consistent decision rule was defined and studied in Gonçalves, Jacob and Mendes-Lopes (2000) to separate the hypothesis "H_0 : X follows an error process" against "H_1 : X follows a diagonal bilinear model".

With the aim of improving the rate of convergence of the decision procedure we present, in this paper, a generalisation of that study in which a smoother statistics is considered in the definition of the sequence of acceptance regions $(A_T)_{T\in\mathbb{N}.}$ In fact, unlike what we have considered in that pioneer study, the statistics here considered is, in general, a continuous function of the sample.

2. GENERAL PROPERTIES AND HYPOTHESES

Let us consider the diagonal bilinear model $X = (X_t, t \in \mathbb{Z})$ defined by

$$X_t = \varphi X_{t-1}\,\varepsilon_{t-1} + \varepsilon_t \ , \tag{1}$$

where φ is a real number and $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ a real stochastic process.

We are going to construct a decision procedure to discriminate between the hypotheses $H_0 : \varphi = 0$ against $H_1 : \varphi = \beta$ $(\beta > 0, \text{fixed})$.

Let us denote the process $X = (X_t, t \in \mathbb{Z})$ distribution and the corresponding expectation by P_φ and E_φ respectively, when the parameter of the model is equal to φ.

We suppose that

$\mathcal{C}_1$: $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ is a strictly stationary and ergodic process.

$\mathcal{C}_2$: $E\big|\log|\varepsilon_t|\big| < +\infty$ and $E\big(\log|\varepsilon_t|\big) + \log|\varphi| < 0$.

Under these conditions, model (1) has a strictly stationary and ergodic solution, P_φ-a.s. unique, given by

$$X_t \;=\; \varepsilon_t + \sum_{n=1}^{+\infty} \varphi^n \varepsilon_{t-n} \prod_{j=0}^{n-1} \varepsilon_{t-1-j} \quad \text{(a.s.)} \;, \qquad t \in \mathbb{Z} \;.$$

If, in adition, we have

$\mathcal{C}_3$: $E\big|\log|X_t|\big| < +\infty$ and $E\big(\log|X_t|\big) + \log|\varphi| < 0$,

then model (1) is invertible and

$$\varepsilon_t \;=\; X_t + \sum_{n=1}^{+\infty} (-\varphi)^n X_{t-n} \prod_{j=0}^{n-1} X_{t-1-j} \quad \text{(a.s.)} \;, \qquad t \in \mathbb{Z} \;.$$

Under conditions $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ we deduce, in view of the two equalities above, that $\underline{X}_t = \underline{\varepsilon}_t$, $\underline{X}_t$ and $\underline{\varepsilon}_t$ denoting the σ-fields generated by $(X_t, X_{t-1}, ...)$ and $(\varepsilon_t, \varepsilon_{t-1}, ...)$ respectively.

Hereafter we assume these general hypotheses concerning the stationarity, ergodicity, and invertibility of model (1). We also define the process $Y = (Y_t, t \in \mathbb{Z})$ by

$$Y_t \;=\; X_t \left(X_t + \sum_{n=1}^{\infty} (-\varphi)^n X_{t-n} \prod_{j=0}^{n-1} X_{t-1-j} \right) \quad \text{(a.s.)} \;.$$

This process is also strictly stationary and ergodic. We will denote it by $Y_t(\varphi)$, if its dependence on the parameter φ is to be emphazised.

We note that $X_t = \varphi Y_{t-1} + \varepsilon_t$, according to (1). Otherwise, taking into account that $E|\log|\varepsilon_t|| < +\infty$ and $E|\log|X_t|| < +\infty$, we have $Y_t(\varphi) \neq 0$, a.s., $\forall \varphi$.

3. A CONSISTENT TEST

We are going to construct a decision procedure to distinguish, in model (1), the hypotheses

$$H_0 : \varphi = 0 \quad \text{against} \quad H_1 : \varphi = \beta \;\; (\beta > 0)$$

from T observations of the process X, denoted by $x_1, x_2, ..., x_T$.

The procedure we are proposing is based, as referred above, on the notion of asymptotic separation of two families of probability laws (Geffroy (1980), Moché (1989)) and it generalises recent works as, for instance, Gonçalves, Jacob, Mendes-Lopes (2000), Gonçalves, Martins, Mendes-Lopes (2001).

First of all, we establish the asymptotic separation of the families of probability laws associated to the hypotheses under investigation by presenting a sequence of Borel sets of $\mathbb{R}^T$, $(A_T, T \in \mathbb{N})$, called separation regions, such that

$$\begin{cases} P_0^T(A_T) \underset{T\to+\infty}{\longrightarrow} 1 \\ P_\beta^T(A_T) \underset{T\longrightarrow+\infty}{\longrightarrow} 0 \end{cases}$$

denoting by P_φ^T the probability law of $(X_1, ..., X_T)$ when the parameter is equal to φ.

We will accept $H_0 : \varphi = 0$ against $H_1 : \varphi = \beta > 0$ if $(x_1, ..., x_T) \in A_T$.

The separation regions that we are going to propose are inspired in previous works. In those papers the test takes into account the number of times that $u\left(\frac{\beta}{2}u - v\right) > 0$ when $(u, v) = (y_{t-1}, x_t)$, y_t denoting the particular value of Y_t, $t = 1, ..., T$. So, the set

$$\begin{aligned} D &= \left\{(u,v) \in \mathbb{R}^2 :\ u > 0,\ v < \frac{\beta}{2}u\right\} \cup \left\{(u,v) \in \mathbb{R}^2 :\ u < 0,\ v > \frac{\beta}{2}u\right\} \\ &= \left\{(u,v) \in \mathbb{R}^2 :\ u\left(\frac{\beta}{2}u - v\right) > 0\right\} \end{aligned}$$

is very important in the construction of a convergent test for the same hypotheses.

The generalization here studied consider a test statistic which is defined following the same basical idea but using a smoother function, eventually a continuous one.

From the definition of D we have

$$(y_{t-1}, x_t) \in D \iff \left(y_{t-1} > 0,\ \frac{\beta}{2}y_{t-1} - x_t > 0\right) \text{ or } \left(y_{t-1} < 0,\ \frac{\beta}{2}y_{t-1} - x_t < 0\right).$$

So, if we consider a distribution function F of a symmetrical law we have

$$\begin{aligned} (y_{t-1}, x_t) \in D \implies & \left(2F(y_{t-1}) - 1 \geq 0,\ 2F\left(\frac{\beta}{2}y_{t-1} - x_t\right) - 1 \geq 0\right) \\ & \text{or} \quad \left(2F(y_{t-1}) - 1 \leq 0,\ 2F\left(\frac{\beta}{2}y_{t-1} - x_t\right) - 1 \leq 0\right) \\ \implies & \left[2F(y_{t-1}) - 1\right]\left[2F\left(\frac{\beta}{2}y_{t-1} - x_t\right) - 1\right] \geq 0\ . \end{aligned}$$

The study here presented takes into account this product. Moreover, a great degree of generality is achieved as the distribution function considered in

the first factor may be different from that appearing in the second one. So, let us define

$$g(u,v) \;=\; \left[2G(u)-1\right]\left[2F\left(\frac{\beta}{2}u-v\right)-1\right], \qquad (u,v)\in\mathbb{R}^2\,,$$

where F and G are distribution functions of symmetrical laws with decreasing densities on $\mathbb{R}^+$.

Let us consider the following regions

$$A_T=\left\{(x_1,x_2,...,x_T)\in\mathbb{R}^T:\sum_{t=2}^{T} g\big(y_{t-1}(\beta),x_t\big)\geq 0\right\}.$$

In what follows, we take $g_t=g(y_{t-1}(\beta),x_t)$ and $\overline{g}_T=\frac{1}{T}\sum_{t=2}^{T} g_t$, and we assume the hypothesis:

$\mathcal{C}_4$: the conditional distribution of ε_t given $\underline{\varepsilon}_{t-1}$ is symmetrical.

We have the following result:

Lemma 3.1.

(i) *Under the hypothesis* $\varphi=0$, $\lim\limits_T \overline{g}_T=E_0(g_2)>0$.
(ii) *Under the hypothesis* $\varphi=\beta>0$, $\lim\limits_T \overline{g}_T=E_\beta(g_2)<0$.

Proof: By the ergodic theorem we have

$$\lim_T\ \overline{g}_T=E_\varphi(g_2)\quad\text{(a.s.)},$$

with

$$\begin{aligned}E_\varphi(g_2) \;&=\; E_\varphi\Big(g\big(Y_1(\beta),X_2\big)\Big)\\ &=\; E_\varphi\left(\left[2G\left(Y_1(\beta)\right)-1\right]\left[2F\left(\frac{\beta}{2}Y_1(\beta)-X_2\right)-1\right]\right).\end{aligned}$$

Let us now study the sign of the limit under each one of the hypotheses H_0 and H_1. In what follows, we take $Y_1(\beta)=Y_1$.

Under $\varphi=0$ we have $X_2=\varepsilon_2$ and so

$$\begin{aligned}E_0(g_2) \;&=\; E_0\left(\left[2G(Y_1)-1\right]\left[2F\left(\frac{\beta}{2}Y_1-\varepsilon_2\right)-1\right]\right)\\ &=\; E_0\left(\left[2G(Y_1)-1\right]E_0\left\{\left[2F\left(\frac{\beta}{2}Y_1-\varepsilon_2\right)-1\right]\middle|\,\underline{\varepsilon}_1\right\}\right)\\ &=\; E_0\left(\left[2G(Y_1)-1\right]\mathbb{I}_{\{Y_1>0\}}E_0\left\{\left[2F\left(\frac{\beta}{2}Y_1-\varepsilon_2\right)-1\right]\middle|\,\underline{\varepsilon}_1\right\}\right)+\\ &\quad+E_0\left(\left[2G(Y_1)-1\right]\mathbb{I}_{\{Y_1<0\}}E_0\left\{\left[2F\left(\frac{\beta}{2}Y_1-\varepsilon_2\right)-1\right]\middle|\,\underline{\varepsilon}_1\right\}\right).\end{aligned}$$

When $Y_1 > 0$, we have $2G(Y_1) - 1 > 0$ and $E_0\left\{\left[2F\left(\frac{\beta}{2}Y_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\} > 0$ using the symmetry of the law of $-\varepsilon_t$ given $\underline{\varepsilon}_{t-1}$; if $Y_1 < 0$ then $2G(Y_1) - 1 < 0$ and $E_0\left\{\left[2F\left(\frac{\beta}{2}Y_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\} < 0$. So, $E_0(g_2) > 0$.

Under $\varphi = \beta > 0$ we have $Y_1 = X_1\varepsilon_1$, $X_2 = \beta X_1\varepsilon_1 + \varepsilon_2$ and then

$$\begin{aligned} E_\beta(g_2) &= E_\beta\left([2G(X_1\varepsilon_1) - 1]\left[2F\left(\frac{\beta}{2}X_1\varepsilon_1 - \beta X_1\varepsilon_1 - \varepsilon_2\right) - 1\right]\right) \\ &= E_\beta\left([2G(X_1\varepsilon_1) - 1]\, E_\beta\left\{\left[2F\left(-\frac{\beta}{2}X_1\varepsilon_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\}\right) \end{aligned}$$

$$\begin{aligned} = {} & E_\beta\left([2G(X_1\varepsilon_1) - 1]\,\mathbb{I}_{\{X_1\varepsilon_1 > 0\}} E_\beta\left\{\left[2F\left(-\frac{\beta}{2}X_1\varepsilon_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\}\right) \\ & + E_\beta\left([2G(X_1\varepsilon_1) - 1]\,\mathbb{I}_{\{X_1\varepsilon_1 < 0\}} E_\beta\left\{\left[2F\left(-\frac{\beta}{2}X_1\varepsilon_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\}\right). \end{aligned}$$

As previously, $2G(X_1\varepsilon_1) - 1 > 0$ and $E_\beta\left\{\left[2F\left(-\frac{\beta}{2}X_1\varepsilon_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\} < 0$, when $X_1\varepsilon_1 > 0$; on the other hand, if $X_1\varepsilon_1 < 0$, $2G(X_1\varepsilon_1) - 1 < 0$ and $E_\beta\left\{\left[2F\left(-\frac{\beta}{2}X_1\varepsilon_1 - \varepsilon_2\right) - 1\right]\middle|\underline{\varepsilon}_1\right\} > 0$. Then $E_\beta(g_2) < 0$. □

We immediately deduce, by the bounded convergence theorem, the following result:

Corollary 3.1.

(i) *If* $\varphi = 0$, $P_0(\overline{g}_T \geq 0) \longrightarrow 1$, *as* $T \longrightarrow +\infty$.
(ii) *If* $\varphi = \beta > 0$, $P_\beta(\overline{g}_T \geq 0) \longrightarrow 0$, *as* $T \longrightarrow +\infty$.

Taking into account the definition of regions A_T, we conclude that the probability laws of process $(X_t, t \in \mathbb{Z})$ defined by the hypotheses $H_0 : \varphi = 0$ and $H_1 : \varphi = \beta > 0$ are asymptotically separated.

So, A_T is the acceptance region of a consistent test for these hypotheses.

4. CONVERGENCE RATE OF THE DECISION PROCEDURE

The convergence rate of the decision procedure, presented in the previous paragraph as a test, may be evaluated when we consider, in the acceptance regions A_T, the true value of Y_t, i.e., $Y_t(\varphi)$, and we assume that the null hypothesis is true. Let us denote these borelians by $A_T(\varphi)$.

We are going to evaluate the convergence rate of $P_0\left(\overline{A}_T(\varphi)\right)$. We have

$$\begin{aligned} P_0\left(\overline{A}_T(\varphi)\right) &= P_0\left(\sum_{t=2}^{T} g_t < 0\right) \\ &\leq E_0\left[\exp\left(-\sum_{t=2}^{T} g_t\right)\right] \\ &= E_0\left\{E_0\left[\exp\left(-\sum_{t=2}^{T} g_t\right)\middle|\, \underline{\varepsilon}_{T-1}\right]\right\} \\ &= E_0\left\{\exp\left(-\sum_{t=2}^{T-1} g_t\right) E_0\left[\exp\left(-g_T\right)|\, \underline{\varepsilon}_{T-1}\right]\right\} \end{aligned}$$

Firstly, we study $E_0\left[g_t|\, \underline{\varepsilon}_{t-1}\right], t \in \mathbb{Z}$.

$$E_0\left[g_t|\, \underline{\varepsilon}_{t-1}\right] = \left[2G\left(\varepsilon_{t-1}^2\right) - 1\right]\left\{2E_0\left[F\left(\frac{\beta}{2}\varepsilon_{t-1}^2 - \varepsilon_t\right)\middle|\, \underline{\varepsilon}_{t-1}\right] - 1\right\}.$$

Let us suppose that ε verifies the following condition

$\mathcal{C}_5$: $\quad \varepsilon_t = \eta_{t-1} Z_t, \quad t \in \mathbb{Z}$

where η_t is a measurable and strictly positive function of $\varepsilon_t, \varepsilon_{t-1}, ...$ with $0 < m \leq \eta_t \leq M$ and $(Z_t, t \in \mathbb{Z})$ is a sequence of independent and identically distributed real random variables, with distribution function F and density f that we suppose symmetrical and decreasing on $\mathbb{R}^+$. We also assume that Z_t is independent of $\underline{\varepsilon}_{t-1}$.

So,

$$\begin{aligned} E_0\left[F\left(\frac{\beta}{2}\varepsilon_{t-1}^2 - \varepsilon_t\right)\middle|\, \underline{\varepsilon}_{t-1}\right] &= E_0\left[F\left(\frac{\beta}{2}\varepsilon_{t-1}^2 - \eta_{t-1} Z_t\right)\middle|\, \underline{\varepsilon}_{t-1}\right] \\ &= \int_{-\infty}^{+\infty} F\left(\frac{\beta}{2}\varepsilon_{t-1}^2 - \eta_{t-1} u\right) f(u)\, du \\ &\geq \int_{-\infty}^{+\infty} F\left(\frac{\beta}{2}\varepsilon_{t-1}^2 - M u\right) f(u)\, du\ . \end{aligned}$$

Choosing the function

$$G(v) = \int_{-\infty}^{+\infty} F\left(\frac{\beta}{2} v - M u\right) f(u)\, du\ ,$$

we note that, by lemma 5.1 (in the appendix), G is the distribution function of a law with a symmetrical density, decreasing on $\mathbb{R}^+$. Moreover, we obtain

$$E_0\left[F\left(\frac{\beta}{2}\varepsilon_{t-1}^2 - \varepsilon_t\right)\middle|\, \underline{\varepsilon}_{t-1}\right] \geq G\left(\varepsilon_{t-1}^2\right)$$

and so

$$E_0\left[g_t|\underline{\varepsilon}_{t-1}\right] \geq \left[2G\left(\varepsilon_{t-1}^2\right)-1\right]^2 .$$

From Hoeffding inequality (Hoeffding (1953)),

$$\begin{aligned} E_0\left[e^{-g_t}\middle|\underline{\varepsilon}_{t-1}\right] &\leq e^{-E_0\left[g_t|\underline{\varepsilon}_{t-1}\right]+\frac{1}{2}\left[2G\left(\varepsilon_{t-1}^2\right)-1\right]^2} \\ &\leq e^{-\frac{1}{2}\left[2G\left(\varepsilon_{t-1}^2\right)-1\right]^2}. \end{aligned}$$

Then

$$\begin{aligned} P_0(\overline{A}_T) &\leq E_0\left\{\exp\left(-\sum_{t=2}^{T-1} g_t\right)\exp\left[-\frac{1}{2}\left(2G\left(\varepsilon_{T-1}^2\right)-1\right)^2\right]\right\} \\ &= E_0\left\{\exp\left(-\sum_{t=2}^{T-2} g_t\right)E_0\left[\exp(-g_{T-1})\exp\left[-\frac{1}{2}\left(2G(\varepsilon_{T-1}^2)-1\right)^2\right]\middle|\underline{\varepsilon}_{T-2}\right]\right\}. \end{aligned}$$

From lemma 5.5 (see appendix), we have the following inequality, for every $t \in \mathbb{Z}$,

$$\begin{aligned} (2)\qquad & E_0\left\{\exp\left(-g_{t-1}\right)\exp\left[-\frac{1}{2}\left(2G\left(\varepsilon_{t-1}^2\right)-1\right)^2\right]\middle|\underline{\varepsilon}_{t-2}\right\} \leq \\ & \leq E_0\left[\exp\left(-g_{t-1}\right)|\underline{\varepsilon}_{t-2}\right]E_0\left[\exp\left[-\frac{1}{2}\left(2G\left(\varepsilon_{t-1}^2\right)-1\right)^2\right]\middle|\underline{\varepsilon}_{t-2}\right] . \end{aligned}$$

In fact,

i) given $\underline{\varepsilon}_{t-2}$, $g_{t-1} = \left[2G\left(\varepsilon_{t-2}^2\right)-1\right]\left[2F\left(\frac{\beta}{2}\varepsilon_{t-2}^2 - x_{t-1}\right)-1\right]$ has the form of the function $h(x) = c\left[2R\left(a-dx\right)-1\right]$ presented in lemma 5.2 (see the appendix), as $x_{t-1} = \eta_{t-2}Z_{t-1}$ under H_0 and $c = 2G\left(\varepsilon_{t-1}^2\right)-1 > 0$, $R = F$, $a = \frac{\beta}{2}\varepsilon_{t-1}^2\left(>0\right)$, and $d = \eta_{t-2}\left(>0\right)$.

ii) On the other hand, $\frac{1}{2}\left[2G\left(d^2x^2\right)-1\right]^2 = \frac{1}{2}\left[G\left(d^2x^2\right)-G\left(-d^2x^2\right)\right]^2$ is a symmetrical function, increasing on $\mathbb{R}^+$, null in the origin and bounded.

As Z_{t-1} is independent of $\underline{\varepsilon}_{t-2}$, the inequality (2) takes the form

$$E_0\left[\exp\left(-h\left(Z_{t-1}\right)-g\left(Z_{t-1}\right)\right)\right] \leq E_0\left[\exp\left(-h\left(Z_{t-1}\right)\right)\right]E_0\left[\exp\left(-g\left(Z_{t-1}\right)\right)\right] .$$

We can then write, with $u_T = \exp\left(-\sum_{t=2}^{T-2} g_t\right)$,

$$\begin{aligned} & P_0\left(\overline{A}_T\left(\varphi\right)\right) \leq \\ & \leq E_0\left\{u_T E_0\left[\exp\left(-g_{T-1}\right)|\underline{\varepsilon}_{T-2}\right]E_0\left[\exp\left[-\tfrac{1}{2}\left(2G\left(\varepsilon_{T-1}^2\right)-1\right)^2\right]\middle|\underline{\varepsilon}_{T-2}\right]\right\} \end{aligned}$$

$$= E_0\left\{E_0\left[\exp\left(-\sum_{t=2}^{T-1} g_t \middle| \underline{\varepsilon}_{T-2}\right) E_0\left[\exp\left[-\tfrac{1}{2}\left(2G\left(\varepsilon_{T-1}^2\right)-1\right)^2\right]\middle|\underline{\varepsilon}_{T-2}\right]\right]\right\}$$

$$\leq E_0\left\{E_0\left[\exp\left(-\sum_{t=2}^{T-1} g_t \mid \underline{\varepsilon}_{T-2}\right) E_0\left[\exp\left[-\tfrac{1}{2}\left(2G\left(m^2 Z_{T-1}^2\right)-1\right)^2\right]\middle|\underline{\varepsilon}_{T-2}\right]\right]\right\}.$$

But $E_0\left[\exp\left[-\frac{1}{2}\left(2G\left(m^2 Z_{T-1}^2\right)-1\right)^2\right]\middle|\underline{\varepsilon}_{T-2}\right]$ is constant as Z_{t-1} is independent of $\underline{\varepsilon}_{T-2}, \forall t \in \mathbb{Z}$. So,

$$P_0\left(\overline{A}_T(\varphi)\right) \leq E_0\left[\exp\left[-\frac{1}{2}\left(2G\left(m^2 Z_{T-1}^2\right)-1\right)^2\right]\right] E_0\left[\exp\left(-\sum_{t=2}^{T-1} g_t \middle| \underline{\varepsilon}_{T-2}\right)\right].$$

Using recursively the procedure leading to

$$E_0\left[\exp\left(-\sum_{t=2}^{T} g_t\right)\right] \leq cE_0\left[\exp\left(-\sum_{t=2}^{T-1} g_t\right)\right]$$

we obtain

$$P_0\left(\overline{A}_T(\varphi)\right) \leq \left\{E_0\left[\exp\left[-\frac{1}{2}\left(2G\left(m^2 Z^2\right)-1\right)^2\right]\right]\right\}^{T-1}$$

where Z is a random variable with the same law of Z_t.

Finally, we may state the following result:

Theorem 4.1. *Let $X = (X_t, t \in \mathbb{Z})$ be a real stochastic process satisfying the model (1) subject to the general conditions C_1, C_2 and C_3.*

If the error process satisfies condition C_5 and the function G is defined by $G(v) = \int_{-\infty}^{+\infty} F\left(\frac{\beta}{2}v - Mu\right) f(u)du$ then the proposed decision rule satisfies

$$P_0\left(A_T(\varphi)\right) \geq 1 - \left\{E_0\left[\exp\left[-\frac{1}{2}\left(2G\left(m^2 Z^2\right)-1\right)^2\right]\right]\right\}^{T-1}, \qquad \forall T \in \mathbb{N}\,.$$

5. APPENDIX

The convergence rate study has been developped assuming absolute continuity and symmetry of the distribution laws involved. So, in this appendix we establish several lemmas concerning distribution functions of symmetrical densities.

Lemma 5.1. *Let f be a symmetrical density decreasing on $\mathbb{R}^+$ with distribution function F. Let a and b be fixed real numbers, with $a > 0$. Then the function $\widetilde{G}$ defined by*

$$\widetilde{G}(v) = \int_{-\infty}^{+\infty} F(av - bu) f(u)\, du$$

is the distribution function of a law with symmetrical density g decreasing on $\mathbb{R}^+$.

Proof: As we can differentiate under the integral sign (Métivier, 1972, p. 156) we obtain

$$\frac{d}{dv}\widetilde{G}(v) = \int_{-\infty}^{+\infty} af(av - bu) f(u)\, du\ .$$

Then, as f is symmetrical,

$$\begin{aligned}
\frac{d}{dv}\widetilde{G}(-v) &= \int_{-\infty}^{+\infty} af(-av - bu) f(u)\, du \\
&= \int_{-\infty}^{+\infty} af(av + bu) f(u)\, du \\
&= \int_{-\infty}^{+\infty} af(av - by) f(y)\, dy \\
&= \int_{-\infty}^{+\infty} af(av - bu) f(u)\, du \\
&= \frac{d}{dv}\widetilde{G}(v)\ .
\end{aligned}$$

Denoting $\frac{d}{dv}\widetilde{G} = g$, g is a symmetrical function. Let us prove that g is a density function and $\widetilde{G}$ the distribution function of density g.

From Fubini, we obtain

$$\int_{-\infty}^{+\infty} dv \int_{-\infty}^{+\infty} af(av - bu) f(u)\, du = \int_{-\infty}^{+\infty} f(u) \left(\int_{-\infty}^{+\infty} af(av - bu)\, dv \right) du\ .$$

But

$$\int_{-\infty}^{+\infty} af(av - bu)\, dv = \int_{-\infty}^{+\infty} af(z) \frac{1}{a} dz = 1\ .$$

Then

$$\int_{-\infty}^{+\infty} dv \int_{-\infty}^{+\infty} af(av - bu) f(u)\, du = 1\ .$$

On the other hand, again from Fubini,

$$\begin{aligned}\int_{-\infty}^{x} g(v)dv &= \int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty} af\left(av-bu\right)du\right)dv\\ &= \int_{-\infty}^{+\infty} f(u)\left[\int_{-\infty}^{x} af\left(av-bu\right)dv\right]du\\ &= \int_{-\infty}^{+\infty} f(u)\left[\int_{-\infty}^{ax-bu} af\left(z\right)\frac{1}{a}dz\right]du\\ &= \int_{-\infty}^{+\infty} F\left(ax-bu\right)f(u)du\\ &= \widetilde{G}\left(v\right).\end{aligned}$$

From the definition of g and as f is decreasing on $\mathbb{R}^+$, it is obvious that g is decreasing on $\mathbb{R}^+$. □

Lemma 5.2. *Let $h\left(x\right)=c\left[2R\left(a-dx\right)-1\right], x\in\mathbb{R}$, where c,a,d are positive numbers and R is the distribution function of a symmetrical and decreasing on $\mathbb{R}^+$ density, r. Let $H(x)=e^{-h(x)}$. Then $H(x)+H(-x)$ is increasing on $\mathbb{R}^+$.*

Proof: We have

$$\begin{aligned}\frac{d}{dx}\left[H(x)+H(-x)\right] &= \frac{d}{dx}\left[e^{-h(x)}+e^{-h(-x)}\right]\\ &= \left[-h'(x)e^{-h(x)}+h'(-x)e^{-h(-x)}\right]\\ &= 2cdr\left(a-dx\right)e^{-h(x)}-2cdr\left(a+dx\right)e^{-h(-x)}.\end{aligned}$$

Let us show that this derivative is non negative. As c and d are positive, it is enough to show that

$$\begin{cases} r\left(a-dx\right)\geq r\left(a+dx\right), & \forall x\geq 0\\ e^{-h(x)}\geq e^{-h(-x)}, & \forall x\geq 0\,.\end{cases}$$

As $a>0$ and $d>0$ and r is decreasing on $\mathbb{R}^+$, we have $r\left(a-dx\right)\geq r\left(a+dx\right)$, for every $x\geq 0$ such that $a-dx>0$.

But, as r is symmetrical, r is increasing on $\mathbb{R}^-$ and if $a-dx<0$ we have

$$r\left(a-dx\right)=r\left(dx-a\right)\geq r\left(a+dx\right),$$

as $0\leq dx-a<dx+a$.

Moreover, as r is a symmetrical density, the function

$$2R\left(x\right)-1=R\left(x\right)-R\left(-x\right)$$

is odd and obviously increasing on $\mathbb{R}^+$.

As c and d are positive we can conclude, by an analogous way, that for every $x \geq 0$

$$c\big[2R\,(a - dx) - 1\big] \;\leq\; c\big[2R\,(a + dx) - 1\big]$$

that is

$$h(x) \leq h(-x)$$

and, in consequence,

$$e^{-h(x)} \geq e^{-h(-x)} \;.$$

Lemma 5.3. *Let φ and f be two symmetrical densities and $a > 0$ such that $\varphi > f$ on $[0, a[$ and $\varphi < f$ on $]a, +\infty[$. Let T be a positive and increasing function, defined on $\mathbb{R}^+$. Then*

$$\int_0^{+\infty} \varphi\,(x)\; T\,(x)\, dx \;<\; \int_0^{+\infty} f\,(x)\; T\,(x)\, dx \;.$$

Proof: We have

$$\begin{aligned}
\int_0^{+\infty} [\varphi\,(x) - f\,(x)]\, T\,(x)\, dx \;&=\; \\
&= \int_{[0,a[} [\varphi\,(x) - f\,(x)]\, T\,(x)\, dx \;+\; \int_{]a,+\infty[} [\varphi\,(x) - f\,(x)]\, T\,(x)\, dx \\
&< \; T\left(a^-\right) \int_{[0,a[} [\varphi\,(x) - f\,(x)]\, dx \;+\; T\left(a^+\right) \int_{]a,+\infty[} [\varphi\,(x) - f\,(x)]\, dx
\end{aligned}$$

as T is an increasing function and where $T\,(a^-)$ denotes the left limit and $T\,(a^+)$ the right limit on a.

As the first quantity is positive, we have

$$\int_0^{+\infty} [\varphi\,(x) - f\,(x)]\, T\,(x)\, dx \;<\; T\left(a^+\right) \int_{]0,+\infty[} [\varphi\,(x) - f\,(x)]\, dx \;=\; 0 \;,$$

taking into account that φ and f are symmetrical densities. □

Lemma 5.4. *Let h be the function of lemma 5.2, φ and f the probability densities of lemma 5.3 and Y and Z real random variables with densities f and φ, respectively. Then*

$$E\left[e^{-h(Z)}\right] \leq E\left[e^{-h(Y)}\right].$$

Proof: We have

$$\begin{aligned} E\left[e^{-h(Z)}\right] &= \int_{-\infty}^{+\infty} e^{-h(z)}\varphi(z)\,dz \\ &= \int_{-\infty}^{+\infty} e^{-h(z)}\varphi(-z)\,dz \\ &= \int_{-\infty}^{+\infty} e^{-h(-u)}\varphi(u)\,du \\ &= E\left[e^{-h(-Z)}\right]. \end{aligned}$$

Then, with $H(x) = e^{-h(x)}$,

$$\begin{aligned} \int_{-\infty}^{+\infty} H(x)\,\varphi(x)\,dx &= \int_{-\infty}^{+\infty} H(-x)\,\varphi(x)\,dx \\ &= \int_{-\infty}^{+\infty} \frac{H(x)+H(-x)}{2}\varphi(x)\,dx \\ &= \int_{-\infty}^{0} \frac{H(x)+H(-x)}{2}\varphi(x)\,dx + \int_{0}^{+\infty} \frac{H(x)+H(-x)}{2}\varphi(x)\,dx \\ &= \int_{0}^{+\infty} (H(x)+H(-x))\,\varphi(x)\,dx, \end{aligned}$$

as φ is symmetrical.

In the same way, we have

$$E\left[e^{-h(Y)}\right] = \int_{0}^{+\infty} [H(x)+H(-x)]\,f(x)\,dx\ .$$

As, by lemma 5.2, the function $H(x)+H(-x)$ is increasing on $\mathbb{R}^+$, we can apply lemma 5.3 to obtain

$$\int_{0}^{+\infty} \varphi(x)\,[H(x)+H(-x)]\,dx < \int_{0}^{+\infty} f(x)\,[H(x)+H(-x)]\,dx\ ,$$

that's to say,

$$E\left[e^{-h(Z)}\right] < E\left[e^{-h(Y)}\right]. \qquad \square$$

Lemma 5.5. *Let g be a symmetrical function, increasing on $\mathbb{R}^+$, equal to zero in the origin and bounded. Let Y be a real random variable with a symmetrical and decreasing on $\mathbb{R}^+$ density f. Let h be the function of lemma 5.2. Then*

$$E\left[e^{-g(Y)-h(Y)}\right] < E\left[e^{-g(Y)}\right] E\left[e^{-h(Y)}\right].$$

Proof: Let us take

$$\frac{1}{b} = \int_{-\infty}^{+\infty} e^{-g(x)} f(x)\, dx \ .$$

We note that $b > 1$, as $e^{-g} < 1$ almost everywhere.

We consider

$$\varphi(x) = b e^{-g(x)} f(x) \ .$$

Then φ is a symmetrical density.

On the other hand, as $b > 1$ and $g(0) = 0$, we obtain

$$\varphi(0) = b e^{-g(0)} f(0) > f(0) \ .$$

Moreover

$$\varphi(x) = f(x) \iff b e^{-g(x)} = 1 \ .$$

As g is monotone increasing, there is a unique root $a > 0$ such that $\varphi > f$ in $[0, a[$ and $\varphi < f$ in $]a, +\infty[$.

Let Z be a real random variable with density φ. From lemma 5.4 we have

$$E\left[e^{-h(Z)}\right] \leq E\left[e^{-h(Y)}\right] \iff \int_{-\infty}^{+\infty} e^{-h(x)} \varphi(x)\, dx \leq \int_{-\infty}^{+\infty} e^{-h(x)} f(x)\, dx \ .$$

As $\varphi(x) = b e^{-g(x)} f(x)$, we obtain

$$b \int_{-\infty}^{+\infty} e^{-h(x)} e^{-g(x)} f(x)\, dx \leq \int_{-\infty}^{+\infty} e^{-h(x)} f(x)\, dx$$

or, using the b definition,

$$\int_{-\infty}^{+\infty} e^{-h(x)} e^{-g(x)} f(x)\, dx \leq \int_{-\infty}^{+\infty} e^{-h(x)} f(x)\, dx \int_{-\infty}^{+\infty} e^{-g(x)} f(x)\, dx$$

which is equivalent to

$$E\left[e^{-g(Y)-h(Y)}\right] \leq E\left[e^{-g(Y)}\right] E\left[e^{-h(Y)}\right] . \qquad \square$$

ACKNOWLEDGMENTS

The authors thank the referee for providing useful comments and suggestions.

REFERENCES

[1] Geffroy, J. (1976). Inégalités pour le niveau de signification et la puissance de certains tests reposant sur des données quelconques, *C. R. Acad. Sci. Paris*, Ser. A, **282**, 1299–1301.

[2] Geffroy, J. (1980). *Asymptotic separation of distributions and convergence properties of tests and estimators*. In "Asymptotic Theory of Statistical Tests and Estimation" (I.M. Chakravarti, Ed.), Academic Press, 159–177.

[3] Gonçalves, E.; Jacob, P. and Mendes-Lopes, N. (1996). A new test for ARMA models with a general white noise process, *Test*, **5**(1), 187–202.

[4] Gonçalves, E.; Jacob, P. and Mendes-Lopes, N. (2000). A decision procedure for bilinear time series based on the asymptotic separation, *Statistics*, **33**, 333–348.

[5] Gonçalves, E.; Martins, C.M. and Mendes-Lopes, N. (2001). Asymptotic separation-based tests for noise processes against first order diagonal bilinear dependence, *Annales de l'ISUP*, **45**(2–3), 9–27.

[6] Granger, C.W.J. and Andersen, A. (1978). *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

[7] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Stat. Assoc.*, **58**(1), 13–30.

[8] Métivier, M. (1972). *Notions Fondamentales de la Théorie des Probabilités*, 2nd ed., Dunod, Paris.

[9] Moché, R. (1989). Quelques tests relatifs à la detection d'un signal dans un bruit Gaussian, *Publications de l'IRMA*, Univ. Lille, France, **19** (III), 1–32.

[10] Pieczinsky, W. (1986). *Sur diverses applications de la décantation des lois de probabilité dans la théorie générale de l'estimation statistique*, Thèse de Doctorat d'État, Univ. Paris VI.

[11] Pham, T.D. and Tran, L.T. (1981). On the first order bilinear time series model, *J. Appl. Prob.*, **18**, 617–627.

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

REVSTAT – STATISTICAL JOURNAL

Background

Statistical Institute of Portugal (INE), 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.
- Two volumes are scheduled for publication, 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*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts*, and *Zentralblatt für Mathematic*.

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.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
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.