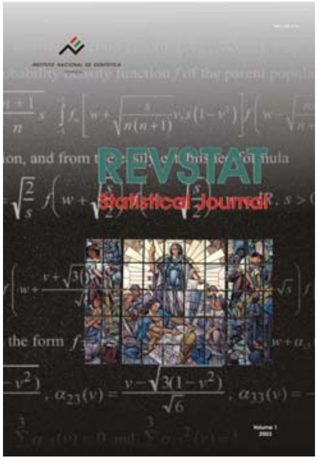
30th March 2007

Estatísticas Gerais

REVSTAT - STATISTICAL JOURNAL

REVSTAT - STATISTICAL JOURNAL, Volume 5, No. 1 – March 2007



I

In 2003 the National Statistical Institute launched the scientific statistical journal **REVSTAT-STATISTICAL JOURNAL**, published in English two times a year, with a prestigious international Editorial Board, which came to substitute the *Revista de Estatística* [Statistical Review], published in Portuguese between 1996 and 2002.

The aim of the Editorial Board of **REVSTAT** is to publish articles of high scientific content, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems, covering all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

REVSTAT hopes to become a place where scientists may feel proud of publishing their research results changing the character of the previous *Revista de Estatística* from a national to an international scientific journal.

For more information about **REVSTAT**, namely on-line articles, subscription of the publication, and submission of papers, please visit the link of the National Statistical Institute's website: http://www.ine.pt/revstat.html

This Volume of **REVSTAT: Volume 5, No. 1 - March 2007**, is about **"Robust Statistics"** and includes seven articles. Their abstracts are presented below:

THE BREAKDOWN POINT— EXAMPLES ANDCOUNTEREXAMPLES

Authors: *P.L. Davies* and *U. Gather*

The breakdown point plays an important though at times controversial role in statistics. In situations in which it has proved most successful there is a group of transformations which act on the sample space and which give rise to an equivariance structure. For equivariant functionals, that is those functionals which respect the group structure, a non-trivial upper bound for the breakdown point was derived in Davies and Gather (2005). The present paper briefly repeats the main results of Davies and Gather (2005) but is mainly concerned with giving additional insight into the concept of breakdown point. In particular, we discuss the attainability of the bound and the dependence of the breakdown point on the sample or distribution and on the metrics used in its definition.

SOME THOUGHTS ABOUT THE DEIGN OF LOSS FUNCTIONS

Authors: *Christian Hennig* and *Mahmut Kutlukaya*

The choice and design of loss functions is discussed. Particularly when computational methods like cross-validation are applied, there is no need to stick to “standard” loss functions such as the L_2-loss (squared loss). Our main message is that the choice of a loss function in a practical situation is the translation of an informal aim or interest that a researcher may have into the formal language of mathematics. The choice of a loss function cannot be formalized as a solution of a mathematical decision problem in itself. An illustrative case study about the location of branches of a chain of restaurants is given. Statistical aspects of loss functions are treated, such as the distinction between applications of loss functions to prediction and estimation problems and the direct definition of estimators to minimize loss functions. The impact of subjective decisions to the design of loss functions is also emphasized and discussed

ESTIMATING SPECTRAL DENSITY FUNCTIONS ROBUSTLY

Authors: *Bernhard Spangl* and *Rudolf Dutter*

We consider in the following the problem of robust spectral density estimation.
Unfortunately, conventional spectral density estimators are not robust in the presence of additive outliers (Martin and Thomson (1982). In order to get a robust estimate of the spectral density function, it turned out that cleaning the time series in a robust way first and calculating the spectral density function afterwards leads to encouraging results. To meet these needs of cleaning the data we use a robust version of the Kalman filter which was proposed by Ruckdeschel (2001). Similar ideas were proposed by Martin and Thomson (1982).
Both methods were implemented in R Development core Team (2005) and compared by extensive simulation experiments. The competitive method is also applied to real data. As a special practical application we focus on actual heart rate variability measurements of diabetes patients.

COMPARATIVE PERFOMANCE OF SEVERAL ROBUST LINEAR DISCRIMINATE ANALYSIS METHODS

Authors: *Valentin Todorov* and *Ana M. Pires*

The problem of the non-robustness of the classical estimates in the setting of the quadratic and linear discriminant analysis has been addressed by many authors: Todorov *et al.* (1990 and 1994), Chork and Rousseeuw (1992), Hawkins and McLachlan (1997), He and Fung (2000), Croux and Dehon (2001), Hubert and Van Driessen (2004). To obtain high breakdown these methods are based on high breakdown point estimators of location and covariance matrix like MVE, MCD and S. Most of the authors use also one step re-weighting after the high breakdown point estimation in order to obtain increased efficiency. We propose to use M-iteration as described by Woodruff and Rocke (1994) instead, since this is the preferred means of achieving efficiency with high breakdown. Further we experiment with the pairwise class of algorithms proposed by Maronna and Zamar (2002) which were not used up to now in the context of discriminant analysis. The available methods for robust linear discriminant analysis are compared on two real data sets and on a large scale simulation study. These methods are implemented as R functions in the package for robust multivariate analysis *rrcov*.

De-biasing Weighted MLE Via Indirect Inference:The Case Of Generalized Linear Latent Variable Models

Authors: *Maria-Pia Victoria-Feser*

In this paper we study bias-corrections to the weighted MLE (Dupuis and Morgenthaler, 2002), a robust estimator simply defined through a weighted score function. Indeed, although the WMLE is relatively simple to compute, for most models it is not consistent and hence not very helpful. For example, the model we consider in this paper is the generalized linear latent variable model (GLLVM) proposed in Moustaki and Knott (2000) (see also Moustaki, 1996, Sammel, Ryan, and Legler, 1997 and Bartholomew and Knott, 1999). The score functions of this model are very complicated. They contain integrals that need to be evaluated. Moreover, they are highly nonlinear in the parameters which makes the use of complicated robust estimator quite impossible in practice. Moustaki and Victoria-Feser (2006) propose to use a weighted MLE and develop indirect inference (Gouriéroux, Monfort, and Renault, 1993, Gallant and Tauchen, 1996 and also Genton and de Luna, 2000, Genton and Ronchetti, 2003) to remove the bias. It can be computed in a simple iterative fashion. In this paper, we actually focus on indirect inference for bias correction in general. We rely heavily on the findings of Moustaki and Victoria-Feser (2006).

Application of Robust Statistics to Asset Allocation Models

Authors: *Roy E. Welsch* and *Xinfeng Zhou*

Many strategies for asset allocation involve the computation of the expected value and the covariance matrix of the returns of financial instruments. How much of each instrument to own is determined by an attempt to minimize risk — the variance of linear combinations of investments in these financial assets — subject to various constraints such as a given level of return, concentration limits, etc. The covariance matrix contains many parameters to estimate and two main problems arise. First, the data will very likely have outliers that will seriously affect the covariance matrix. Second, with so many parameters to estimate, a large number of return observations are required and the nature of markets may change substantially over such a long period. In this paper we discuss using robust covariance procedures, FAST-MCD, Iterated Bivariate Winsorization and Fast 2-D Winsorization, to address the first problem and penalization methods for the second. When back-tested on market data, these methods are shown to be effective in improving portfolio performance. Robust asset allocation methods have great potential to improve risk-adjusted portfolio returns and therefore deserve further exploration in investment management research.

Penalized Trimmed Squares and A Modification of Support Vectors for Unmasking Outliers In Linear Regression

Authors: *G. Zioutas; A. Avramidis* and *L. Pitsoulis*

We consider the problem of identifying multiple outliers in linear regression models. We propose a penalized trimmed squares (PTS) estimator, where penalty costs for discarding outliers are inserted into the loss function. We propose suitable penalties for unmasking the multiple high-leverage outliers. The robust procedure is formulated as a Quadratic Mixed Integer Programming (QMIP) problem, computationally suitable for small sample data. The computational load and the effectiveness of the new procedure are improved by using the idea of *ε-insensitive* loss function from support vector machines regression. The small errors are ignored, and the mathematical formula gains the sparseness property. The good performance of the PTS estimator allows identification of multiple outliers avoiding masking effects.