Final Report for project AKTION 50p14

General Report

Within the framework of project Action 50p14, Dr. Z. Fabián has accomplished four trips from Prague to Linz. Unfortunately, Dr. I.Vajda was ill the whole last three months of the year so that he was not able to accomplish the planned visit. Visits of Dr. Z. Fabián: (1) 12.5.-16.5. The work during visit consisted of discussion the results and revising the paper [1], which we sent to the special issue of CDQM. We have formulated our next aim: to study the Hill estimator. (2) 27.10.- 1.11. Revision of paper [7]. We discussed results of our studies on the Hill estimator and formulated the concrete aim: a distribution sensitive Hill-like method based on generalized moment estimators established, used also in [1]. (3) 18.11.-20.11. Simulations of the Hill-like estimator for data from Pareto distribution. (4) 30.11. - 4.12. Discussion of results and establishing the framework of paper [5]. Simulations for real data.

Synergies of the project: Dr. Milan Stehlík has visited České Budějovice on 24.10.2008 and on 5.12.2008 to find a new contacts for further applications of derived methods. This finally led to the accepted project Project 53p19 "Aktion Czech Republic-Austria":Spatial mapping of ticksand tick-borne infectious diseases of the region of South Bohemia and Upper Austria. We will apply some results from project 50p14, especially those on homogeneity testing and t-estimation also in the project 53p19.

Planned activities

In the literature up to now only partial answers can be found on the question on favorable estimation for heavy tailed distributions. The maximal likelihood estimation is very sensitive to deviations from the theoretical distributions, namely in the heavy-tailed class of distributions. Robust methods considering contaminated distributions work much better, but sweep out the differences among distributions so that do not fully use the prior information. A new method based on t-estimation was proposed by Fabián in 2006. It was based on idea, that data are taken from a non-contaminated heavy-tailed distribution, and are treated by means of a scalar inference function T characterizing the given distribution. It appeared that functions T of heavy-tailed distributions are bounded. In contrast to usual moments, the moments $E(T^k)$ exist and their finite versions give simple robust estimates of parameters of heavy-tailed distributions. The aim of the project 50p14 was further investigation of this inference method together with partial tasks in computing and testing competitive risk models.

Obtained Results We have finalized the following tasks:

In [1] we describe an alternative method of estimation of severities in risk assessment through the t-score moment method when the data are contaminated. The method is a bit less efficient than the standard maximum likelihood method, however it is more robust. Simulation study emphasize the fact that crucial role is played by level of contamination. In cases of very small contamination, both methods give similar results. The larger contamination, the better are

the t-moment estimates. The reason for it is the bounded t-score of heavy-tailed distributions (and, consequently, bounded influence functions of the estimators).

In [2] we have provided a thorough theoretical and simulation study of the testing procedure for homogeneity against mixture alternative (competitive risk model). We have found that for the lower contamination (in the sense of Mosler) the ELR2 and ELRH procedures are outperforming the competing testing procedures. This partially solves the problem of accuracy of maximal likelihood based estimation through the approximation of the exact mixture model with the subpopulation model. Sill, a lot of work remains, especially comparison of widely used Bayesian approach to the likelihood ratio based approach.

In [3] further aspects on favorable estimation for heavy tailed data are given. A mixture of gamma and log gamma frequently used for claims modeling by actuaries, is used as an example illustrating how non robust maximum likelihood is, when the underlying distribution is misspecified. Also real data from non-life insurance are processed by various estimation techniques.

A commonly used approach for modeling operational risk is the loss distribution approach, where separate models for frequency and severity of individual losses are combined to a model for aggregate loss. [4] gives an introduction to Bayesian analysis of these type of models. The Bayesian approach offers several advantages over the classical approach to model and analyze operational risk: First it allows to combine loss data with prior information provided by experts. As Bayesian estimation of risk models typically relies on MCMC methods, estimates of arbitrary parameters of the total loss distribution, e.g. quantiles are readily available. In contrast to classical methods where prediction is accomplished by plugging in parameter estimates the Bayesian predictive distribution of aggregate loss fully takes into account parameter uncertainty.In [4] a detailed account on Bayesian analysis for a compound Poisson-Pareto model is given and illustrated on simulated data.

In [5] we have introduce a novel, robust and distribution sensitive Hill-like method. The method is based on t-estimation and leads to the consistent t-Hill estimator of the Pareto distribution. This paper is a step towards the new approach for heavy -tailed data estimation which give a potential to develop the distribution-sensitive approach. In many areas, like telecommunications, it has been recognized already, that Pareto approximation is too vague in some situations. Of course, still much of work remains.

In [6] we show that commonly used approaches for calculating extreme upper quantiles of these distributions fail in some cases and thus alternative methods should be used. In the second part a specific problem is discussed. It is widely known that for extremal events the following situation occurs very often: 20% of the individual claims are responsible for more than 80% of the total (aggregated) claim amount in a portfolio of policies.The latter is called the 20-80 rule. Of course, characterizing distributions which obey the rule is of great importance for insurance companies. The real data example illustrates the problem. Finally, in discussion a possible answer to the problem of catastrophic events is outlined. It is using financial securities such as catastrophe index options, futures and cat bonds which brings at least partial solutions.

In [7] new estimators based on t-score moments are introduced and compared with the classical estimators (maximum likelihood and moment estimators) and with recently introduced robust estimators of "generalized median" and "trimmed mean" type. We derive the exact distribution of the likelihood ratio tests of homogeneity and tail index of the two-parameter Pareto model which support the assessment of performance of estimators. In particular, we discuss some problems that one can encounter when misemploying the log-normal assumption based methods supported by the Basel II framework. The real data example illustrates the

methods.

Publications of the project There are 7 publications which have been made with the help of the project 50p14. They are listed in the following References section.

References

[1] Fabián Z. and Stehlík M. (2008). A note on favorable estimation when data is contaminated, CDQM, Volume 11, Number 4, pp. 36-43

[2] Stehlík M. and Wagner H. (2009) Exact likelihood ratio testing for homogeneity of exponential distribution, *IFAS Research Report Nr. 39*, http://www.ifas.jku.at/e2550/e2756/index_ger.html (submitted to Compuatational Statistics and Data Analysis)

[3] Waldl H. (2008) On parameter estimation for heavy tailed data CDQM, Volume 11, Number 4, pp. 101-114

[4] Helga Wagner, Bayesian analysis of operational risk models based on loss distributions CDQM, Volume 11, Number 4, 2008, pp. 87-1003.

[5] Fabián Z. and Stehlík M. (2009). On robust and distribution sensitive Hill like method, technical report, in preparation for submission to Statistics and Probability Letters

[6] Rastislav Potocký, Milan Stehlík and Helmut Waldl, Catastrophic claims: new chalenge for insurance companies, submitted to Ekonomie+Management

[7] Stehlík M, Potocký R, Fabián Z, Waldl H. (2008). Some notes on the favourable estimation of fitting heavy tailed data, *IFAS Research Report Nr. 32*,http://www.ifas.jku.at/e2550/e2756/index_ger.html, submitted to Computational Statistics

Linz, 16.2.09

Dr. Milan Stehlík
Principal investigator