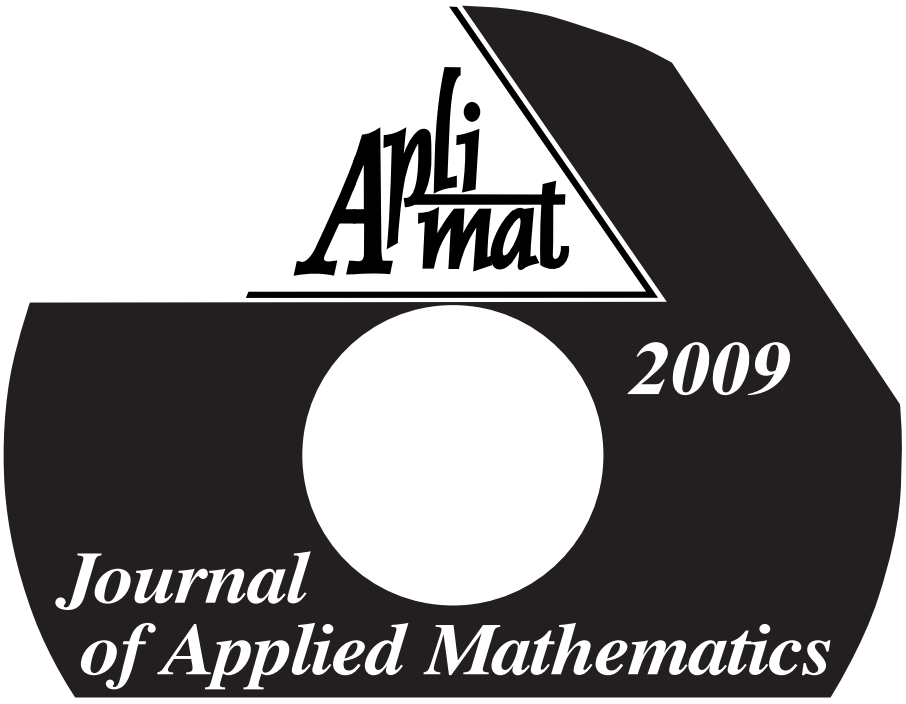


ISSN: 1337-6365

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

VOLUME 2 (2009), NUMBER 3

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

VOLUME 2 (2009), NUMBER 3

Edited by: Slovak University of Technology in Bratislava

Editor - in - Chief: KOVÁČOVÁ Monika (Slovak Republic)

Editorial Office: Institute of natural sciences, humanities and social sciences
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava
Námestie slobody 17
812 31 Bratislava

Correspodence concerning subscriptions, claims and distribution:
F.X. spol s.r.o
Azalková 21
821 00 Bratislava
journal@aplimat.com

Frequency: One volume per year consisting of two issues at price of 120 EUR, per volume, including surface mail shipment abroad.
Registration number EV 2540/08

Information and instructions for authors are available on the address: www.aplimat.com

Printed by: F.X. spol s.r.o, Azalková 21, 821 00 Bratislava

APLIMAT - JOURNAL OF APPLIED MATHEMATICS
VOLUME 2 (2009), NUMBER 3

STATISTICAL METHODS IN TECHNICAL AND ECONOMIC SCIENCES AND PRACTICE

APLIMAT - JOURNAL OF APPLIED MATHEMATICS
VOLUME 2 (2009), NUMBER 3

FINANCIAL AND ACTUARY MATHEMATICS

APLIMAT - JOURNAL OF APPLIED MATHEMATICS
VOLUME 2 (2009), NUMBER 3

ENGINEERING APPLICATIONS AND SCIENTIFIC COMPUTATIONS

APLIMAT - JOURNAL OF APPLIED MATHEMATICS

LIST OF REVIEWERS

Abderraman Jesus C., PhD.	UPM - Technological University of Madrid, Madrid, Spain
Andrade Marina, Dr., PhD.	ISCTE Business School, Lisbon, Portugal
Baranová Eva, RNDr.	Technical university Košice, Košice, Slovak Republic
Beránek Jaroslav, doc. RNDr. CSc.	Masaryk University, Brno, Czech Republic
Daveau Christian, Dr. of Mathematics	Universit'e de Cergy-Pontoise, Cergy-Pontoise Cedex, France
Diblík Josef, Prof. RNDr., DrSc.	Brno University of Technology, Czech Republic
Dobrucky Branislav, Prof.	University of Zilina, Zilina, Slovak Republic
Dorociaková Božena, RNDr., PhD.	University of Zilina, Zilina, Slovak Republic
Došlá Zuzana, Prof. RNDr. DrSc.	Masarykova univerzita, Brno, Czech Republic
Fajmon Břetislav, RNDr., PhD.	Brno University of Technology, Brno, Czech Republic
Ferreira Manuel Alberto M., Full Prof.	ISCTE, Lisboa, Portugal
Filipe Jose Antonio, Assistant Prof.	ISCTE, Lisboa, Portugal
Grodzki Zdzisław, Dr. hab.	Techinical University of Lublin, Lublin, Poland
Habiballa Hashim, RNDr. PaedDr., PhD.	University of Ostrava, Ostrava, Czech Republic
Heckenbergerova Jana, Mgr.	University of Pardubice, Pardubice, Czech Republic
Hinterleitner Irena, Mgr.	Brno University of Technology, Brno, Czech Republic
Hošková Šárka, RNDr., PhD.	University of Defence, Brno, Czech Republic
Hřebíček Jiří, Prof. RNDr., CSc.	Masaryk University, Brno, Czech Republic
Hrnčiarová Ľubica, doc.Ing., PhD.	University of Economics in Bratislava, Bratislava, Slovak Republic

Huňka František, doc. Ing., CSc.	Ostravská univerzita in Ostrava, Czech Republic
Chocholatá Michaela, Ing., PhD.	University of Economics, Bratislava, Slovak Republic
Chvalina Jan, Prof. RNDr., DrSc.	Brno University of Technology, Brno, Czech Republic
Chvátalová Zuzana, RNDr., PhD.	Brno University of Technology, Brno, Czech Republic
Jancarik Antonin, PhD.	Charles University, Prague 1, Czech Republic
Kalina Martin, doc. RNDr., CSc.	Slovak University of Technology, Bratislava, Slovak Republic
Kapička Vratislav, Prof. RNDr., DrSc.	Masarykova univerzita, Brno, Czech Republic
Khelifi Abdessatar, PhD.	Sciences of Bizerte, Zarzouna, Tunisia
Kováčová Monika, Mgr., PhD.	Slovak University of Technology, Bratislava, Slovak Republic
Kriz Jan, RNDr., PhD.	University of Hradec Kralove, Hradec Kralove, Czech Republic
Kureš Miroslav, Assoc. Prof.	Brno University of Technlogy, Brno, Czech Republic
Kvasz Ladislav, doc. Dr., PhD.	Charles University, Prague, Czech Republic
Lovisek Jan, Prof.	Slovak University of Technology, Bratislava, Slovak Republic
Lungu Nicolaie, Prof.	Technical University of Cluj-Napoca, Romania
Malacká Zuzana, RNDr., PhD.	University of Žilina, Slovak Republic
Mamrilla Dušan, doc. RNDr., CSc.	University of Prešov in Prešov, Slovak Republic
Marček Dušan	University of Žilina, Žilina, Slovak Republic
Marčoková Mariana, doc. RNDr., CSc.	University of Žilina, Žilina, Slovak Republic
Maroš Bohumil, doc. RNDr., CSc.	Brno University of Technology, Brno, Czech Republic
Martinek Pavel, Ing., PhD.	Palacky University, Olomouc, Czech Republic
Mikeš Josef, Prof. RNDr., DrSc.	Palacký University, Olomouc, Czech Republic
Mišútová Mária, doc. RNDr., PhD.	Slovak University of Technology, Trnava, Slovak Republic
Moučka Jiří, doc. PhDr., PhD.	University of Defence, Brno, Czech Republic

Pavlačka Ondřej, Mgr., PhD.	Palacký University, Olomouc, Czech Republic
Plháková Alena, doc., PhD.	Palacký University, Olomouc, Czech Republic
Pokorný Milan, PaedDr. PhD.	Trnava University, Trnava, Slovak Republic
Pokorny Michal, Prof.	University of Zilina, Zilina, Slovak Republic
Pribullová Anna	Gephysical Insitute SAS, Bratislava, Slovak Republic
Půlpán Zdeněk, Prof. RNDr., PhD.	University Hradec Králové, Hradec Králové, Czech Republic
Rus Ioan A., Prof.	Babes-Bolyai University of Cluj-Napoca, Cluj-Napoca, Romania
Slapal Josef, Prof. RNDr., CSc.	Brno University of Technology, Brno, Czech Republic
Svoboda Zdeněk, RNDr., CSc.	Brno University of Technology, Brno, Czech Republic
Šmarda Zdeněk, doc. RNDr., CSc.	Brno University of Technology, Brno, Czech Republic
Talašová Jana, doc. RNDr., CSc.	Palacký University Olomouc, Olomouc, Czech Republic
Vacek Vladimír, RNDr.	Technická univerzita, Zvolen, Slovak Republic
Vajsablova Margita, PhD.	Slovak University of Technology, Bratislava, Slovak Republic
Vanžurová Alena, doc. RNDr., CSc.	Palacky University., Olomouc, Czech Republic
Vávra František, doc. Ing., CSc.	University of West Bohemia, Plzeň, Czech Republic
Velichová Daniela, doc. RNDr., CSc.	Slovak University of Technology, Bratislava, Slovak Republic
Volna Eva, doc. RNDr. PaedDr., PhD.	University of Ostrava, Ostrava 1, Czech Republic
Witkovsky Viktor, doc. RNDr., CSc.	SlovakAcademy of Sciences, Bratislava, Slovak Republic
Žáčková Petra, Mgr.	Technical university of Liberec, Liberec, Czech Republic
Žák Libor	University of Technlogy, Brno, Czech Republic

Journal of Applied Mathematics

volume II (2009), number III

BAYESIAN NETWORKS IN FORENSIC IDENTIFICATION PROBLEMS

ANDRADE Marina, (P), FERREIRA Manuel Alberto M., (P)

Abstract. Paternity dispute and criminal identification problems are examples of situations in which forensic approach the DNA profiles study is a common procedure. In order to deal with the problems mentioned it is needed an introduction to present and explain the various concepts involved, since distinct areas must be considered. In the second paragraph some problems are presented. Here it is exhibited an algebraic treatment, for the simpler problems and with those the use of the object-oriented Bayesian networks is shown. Then the most complex kind of problems that may occur is presented. In the last paragraph some comments are added.

Key words: Bayesian networks, DNA profiles, identification problems.

Mathematics Subject Classification: Primary 62C10; Secondary 62P99.

1 Introduction

The use of networks transporting probabilities began with the geneticist Sewall Wright in the beginning of the 20th century (1921). Since then their use had different forms in several areas like social sciences and economy – in which the used models are, in general, linear named Path Diagrams or Structural Equations Models (SEM), and in artificial intelligence – usually non-linear models named Bayesian networks also called Probabilistic Expert Systems (PES).

Bayesian networks are graphical structures for representing the probabilistic relationships among a large number of variables and for doing probabilistic inference with those variables, Neapolitan (2004). Before we approach the use of Bayesian networks to our interest problems we briefly discuss some aspects of PES in connection with uncertainty problems.

1.1 Probability concept

The interpretation of probability has been and still is a subject of intense debate. It has important implications for the practice of probability modelling and statistical inference, both in general and

in expert systems applications. We believe that the main division may be stated between objective and epistemological, Gillies (1994), understandings of *P(A)*, the probability of the event *A*; or more generally of *P(A|B)*, the probability of A conditional on the happening of the event *B*.

Objective theories consider such probabilities as real world attributes of the events they refer to, and are not affected by or related to our perception of them. The most influent objective interpretation has been the frequentist interpretation (Venn; von Mises, Reichenbach, etc.), to which probability is defined as the limit of the proportion of successes in an infinite sequence of experiments. It only allows the approach of repeatable events. Despite this important limitation this interpretation has been the dominant one and was the basis of Neyman and Pearson's frequentist approach to statistical inference.

Epistemological theories see *P(A|B)* as a state of mental uncertainty about *A*, in the knowledge of *B* – where *A* and *B* may be singular propositions and not necessarily repeatable events. These theories can be divided into logical and subjectivists theories. Logical theories suppose the existence of a single rational degree of uncertainty about A, in the knowledge of B. However, the problem is that it is not yet known a method for the evaluation of logical probabilities. The subjectivist interpretation has become more popular in the last years. Subjectivists regard probability as a degree of reasonable belief in a certain event, from an individual viewpoint; therefore probability is a numeric subjective measure of a particular person according his/her degree of belief, as long as it is 'coherent'[1].

Obviously, from the objective part the critics can claim that it is an extremely vulnerable assertion. However, experience shows that distinct people, with different degrees of knowledge or information with respect to certain events, have different quantifications of the associated uncertainty.

From a subjectivist perspective it is possible to specify probabilities of individual propositions, and even to treat unknown constants or parameters as random variables. Being unknown it is possible to assign them probabilities, under a coherent structure. The subjectivist interpretation is the one we follow here.

1.2 Expert systems

Expert systems are attempts to crystallize and codify the knowledge and skills of one or more experts into a tool that can be used by non-specialists, Cowell *et al*. (1999). An expert system can be decomposed as follows:

Expert system = knowledge base + Inference engine.

The first term on the right-hand side of the equation, *knowledge base*, refers the specific knowledge domain of the problem. The *inference engine* is given by a set of algorithms, which process the codification of the *knowledge base* jointly with any specific information known for the application in study.

[1]The principle of coherence requires that an individual should not make a collection of probability assessments that could put him in the position of suffering a sure loss, no matter how the relevant uncertain events turn out, Cowell e*t al*. (1999).

Usually it is presented in a software program, as the one we are going to show hereafter, but such is not an imperative rule. Each of those parts is important for the inferences, but *knowledge base* is crucial. The inferences obtained depend naturally on the quality of the *knowledge base*, of course in association with a sophisticated *inference engine*. The better those parts are the best results we can get.

A PES is a representation of a complex probability structure by means of a directed acyclic graph, having a node for each variable, and directed links describing probabilistic causal relationships between variables, Dawid *et al*. (2002). Bayesian approach is the adequate for making inferences in probabilistic expert systems.

1.2.1 Bayesian networks

Bayesian networks are graphical representations expressing qualitative relationships of dependence and independence between variables. A Bayesian network is a directed acyclic graph $\mathcal{G}$ (DAG) having a set of V vertices or nodes and directed arrows. Each node $v \in V$ represents a random variable X_v with a set of possible values or states. The arrows connecting the nodes describe conditional probability dependencies between the variables.

The set of parents, $pa(v)$, of a node v comprises all the nodes in the graph with arrows ending in v. The probability structure is completed by specifying the conditional probability distributions for each random variable X_v and each possible configuration of variables associated with its parent nodes $x_{\mathrm{pa}(v)}$. The conditional distribution of X_v is expressed given $X_{pa(v)} = x_{\mathrm{pa}(v)}$. The joint distribution is $p(x) = \Pi_{v \in V} p(x_v | x_{\mathrm{pa}(v)})$. There are algorithms to transform the network into a new graphical representation, named junction tree of cliques, so that the conditional probability $p(x_v | x_A)$ can be efficiently computed, for all $v \in V$, any set of nodes $A \subseteq V$, and any configuration x_A of the nodes X_A. The nodes in the conditioning set A are generally nodes of observation and input of evidence $X_A = x_A$, or they may specify hypotheses being assumed.

Software such as Hugin[2] can be used to build the Bayesian network through the graph $\mathcal{G}$. That can be done by specifying the graph nodes, their space of states and the conditional probabilities $p(x_v | x_{\mathrm{pa}(v)})$. In the compiling process the software will construct its internal *junction tree* representation. Then, by entering the evidence $X_A = x_A$ at the nodes in A, and requesting its propagation to the remaining nodes in the network, the conditional probabilities $p(x_v | x_A)$ are obtained.

OOBN are one example of the general class of Bayesian networks. An instance or object is a regular network possessing input and output nodes as well as ordinary internal nodes. The interface nodes have grey fringes, with the input nodes exhibiting a dotted line and the output nodes a solid line. The instances of a given class have identical conditional probability tables for non-input nodes. The objects are connected by directed links from output to input nodes. The links represent identification of nodes. We use **bold face** to refer the object classes and *math mode* to refer the nodes. The modular flexibility structure of the OOBN is of great advantage in complex cases

[2] http://www.hugin.com - OOBN a resource available in the Hugin 6.4 software.

1.2.2 STR markers and DNA profiles

The development of the molecular biology, since the decade of 60, allowed the knowledge of the DNA structure and its implementation as a genetic information vehicular, so that it can also be used in the clarification of judicial forensic problems.

Every human being has 23 pairs of chromosomes in the nuclear of human cell. One of those pairs determines the gender – XY for male, XX for female. The other 22 pairs are said homologous pairs. All of them are DNA molecules. A DNA molecule is a double helix composed by four different nucleotides: C, A, G and T, binding in pairs C-G and A-T.

A locus, sometimes also named a gene for simplification, is an area on a chromosome and the DNA composition on that area is an allele. Thus, a locus corresponds to a random variable and the allele is its realized state.

A DNA marker is a known locus where the allele can be measured in the laboratory, by the use of appropriate techniques. More recently, the techniques provide the use of Short Tandem Repeats (STR) markers, which avoid the possibility of measurement errors. STR markers are given by integers, but they can be codified even to protect the process or case. If an STR allele exhibits a value of 5, a certain expression (e.g. GTCCAG) is repeated exactly five times at that locus.

A DNA profile for an individual is a measurement on several markers to which a genotype is observed. The genotype is an unordered pair of alleles, one inherited from the individual's father and the other from the mother, although it is not possible to distinguish which is which. In this work we implement the product rule that is Hardy-Weinberg and linkage equilibrium assumptions; in practice it assumes the independence of the individual's alleles both within and across markers. If a more complex genetic model was desired it could be implemented by introducing dependencies between founder nodes.

1.3 Forensic identification

The use of DNA profiles in forensic identification problems has become, in the last years, an almost regular procedure in many and different situations. Among those are: 1) disputed paternity problems, in which it is necessary to determine if the putative father of a child is or is not the true father; 2) criminal cases as if a certain individual A was the origin of a stain found in the scene of a crime; or 3) in more complex cases to determine if an individual or more did contribute to a mixture trace found. In criminal cases it is common to find traces with more than one single contributor. As it is known a person has at most two different alleles for each marker. If a trace exhibits more than two alleles to one or more markers then it is certainly a mixture trace.

Mixture traces can happen in rape cases, where the vaginal swab typically will contain DNA from the victim as well as the perpetrator, and also from a consensual partner or several perpetrators. Homicides or robberies are other possible origin for mixture traces, where we can admit a fight that produces some material.

There are still some other forensic identification problems, however not too frequent ones. That is the case of an identification of a body found, together with is information of a missing person belonging to a known family, or the identification of more than one body resultant of a disaster or an attempt. And even immigration cases in which it is important to establish family relations.

We can say that the use of Bayesian arguments in forensic problems begun with a Dennis Lindley work in 1977. Since then there is a huge amount of published works in this area, in great part due to the evolution of the DNA profiling techniques. The interest in forensic identification problems was not exclusively of forensic scientists as it can be seen by innumerable articles made with the contribution of statisticians.

2 Using Bayesian networks

Dawid *et al*.'s (2002) work describes a new approach to the problems mentioned above. The construction and use of Bayesian networks to analyse complex problems of forensic identification inference was initially done there followed by Evett *et al*. (2002), Dawid *et al*. (2002), Mortera (2003) and Mortera *et al*. (2003) among others.

Here we start with a simple graphical and numerical representation and extend our analysis to more complex problems, such as DNA mixtures and cases where the evidence is composed with more than one trace.

2.1 Disputed paternity

In a case of disputed paternity the genetic information of the child can be seen as partial information about the true father. In a simple case the paternity is imputed to a certain individual who rejects it. DNA profiles of the mother *m*, the child *c*, and the putative father *pf* are available.

Becoming the paternity assumption litigious we can say that, in formal terms, two hypotheses are established, which for simplification we will name the prosecution and the defense hypotheses, i.e.,

H_P: The true father is the putative father.

vs

H_D: The true father is another individual randomly drawn from the population, and not genetically related with the mother or the putative father.

We need to assess the likelihood function over the hypotheses as to the true father. If we denote the data (*mgt*, *cgt*, *pfgt*) as the evidence *E*, then we want to evaluate the likelihood ratio:

$$LR = \frac{P(E \mid H_P)}{P(E \mid H_D)}.$$

Naturally the court has to answer to the truly paternity of the child. If we want the court has to evaluate the ratio of the hypotheses in dispute. That is

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(E \mid H_P)}{P(E \mid H_D)} \times \frac{P(H_P)}{P(H_D)}.$$

If we admit that $P(H_P) = P(H_D)$ then

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(E \mid H_P)}{P(E \mid H_D)}.$$

Before continuing let us briefly explain the equations above.
Being the markers in different chromosomes (*linkage equilibrium*) and assuming random mating (*Hardy-Weinberg equilibrium*) we have independence between and within markers. Therefore we can obtain the *LR* for each marker separately and multiply the values to determine the overall likelihood ratio based on the data available for all markers.

We want to determine the probability of the triplet *E*, under the two hypotheses. We can agree that before knowing any data on the child it is reasonable to assume that the identity of the true father is independent of the mother's and the putative father's. And supported on that, it is easily seen that we can determine the conditional probability of the child's genotype, given the other two available genotypes. Thus, to determine $P(E \mid H_P)$ we simply have to apply the Mendel's laws. But the calculus of $P(E \mid H_D)$ necessarily demands the knowledge of the population allele frequencies for the considered markers.

Let us admit that for a certain marker the triplet *E* = (*mgt*, gt*c*, *pfgt*) is the following $E = ((A,B);(B,B);(A,B))$, and p_A and p_B are the population allele frequencies for the considered marker.

$$\begin{aligned} P(E \mid H_P) &= P[(mgt; cgt; pfgt) \mid (mgt; pfgt)] \\ &= P[cgt \mid (mgt; pfgt)] \\ &= 0.5 \times 0.5 \end{aligned}$$

and

$$\begin{aligned} P(E \mid H_P) &= P[(mgt; cgt; pfgt) \mid (mgt; rgt)] \\ &= P[cgt \mid (mgt; rgt)] \\ &= 0.5 \times p_B \end{aligned}$$

where *rgt* assigns the genotype of a random individual of the population, not related to the mother or the putative father.
Therefore,

$$LR = \frac{P(E \mid H_P)}{P(E \mid H_D)} = \frac{0.5}{p_B}.$$

The considered problem is, as shown, easily algebraically solved. However we will use it to illustrate the simplicity and the advantages of using this tool in more complex situations. Given the freedom of choice for the variables to include in the graphical representation, different representations can be obtained. Some of them simpler than others. To get a 'good' representation is very important to the efficiency and the viability of the computational routines. These are extremely sensible to the organization of the graphical structure. The first step consists on the identification and definition of the nodes for all the variables of interest to the problem.

After that we are able to build the graph representation. In accordance with Dawid *et al*. (2002), *in order to maximize the efficiency of the calculations as well as the logical clarity of the representation we chose to disaggregate each individual's genotype into its constituent, unobserved, paternally and maternally inherited genes*.

Thus, in Fig. 1 we have the OOBN for the paternity case discussed in Dawid *et al*. (2002), considering a single marker. Each node (instance) in the network represents itself a Bayesian network. In this simple paternity case instances **pfmg**, **pfpg**, **mpg** and **mmg** are all of class **founder**, Fig. 2, and represent the 'putative father's maternal and paternal gene', and similarly for the mother. Instances **mgt**, **cgt** and **pfgt** are of class **genotype**, Fig. 3, and consider the observed genotype. The instances **tfmg** and **tfpg** are of class **whom**, Fig. 4, and specify whether the correspondent allele is or is not from the putative father. And **cpg** and **cmg** of class **inherit**, Fig. 5 represent the allele transmission through meiosis. The node *tf=pf?* represents the binary query 'Is the true father the putative father?'

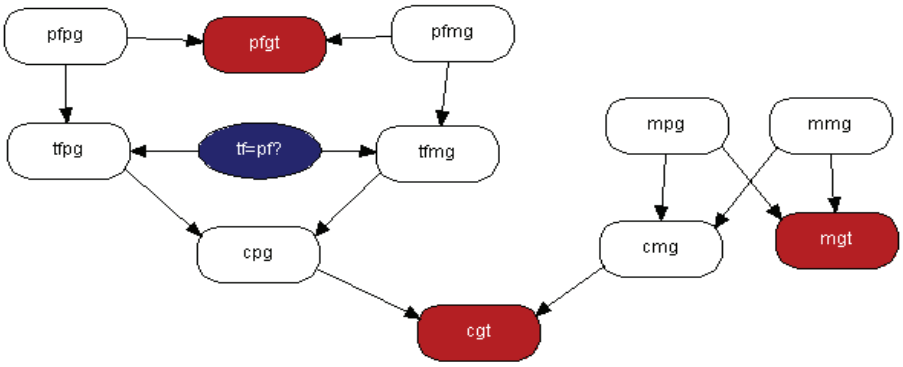


Figure 1: Simple paternity network.

The instance **founder** contains a single node *gene,* having for its space of states all the possible alleles that can be presented for the specific case, and the correspondent population gene frequencies.

Figure 2: founder network.

The genotype of an individual is an unordered pair of alleles inherited from paternal, *pg*, and maternal, *mg*, genes, here represented by *gtmin:=min*{*pg*, *mg*} and *gtmax:=max*{*pg*, *mg*}, where *pg* and *mg* are input nodes identical to the *gene* node of **founder**.

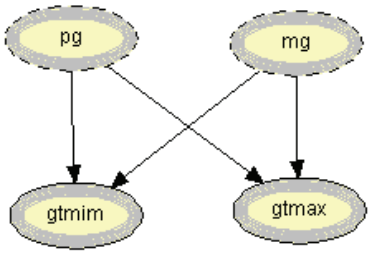


Figure 3: genotype network.

The instance **whom** describes the true father's allele origin. If *tf=pf?* has true for value then the true father's allele, *tfg*, will be identical with the putative father's, *pfg*, otherwise the true father's allele is chosen randomly from another man in the population, with **otherg** an instance of the class **founder**, and *tfg:=if*(*tf=pf?= = true*, *pfg*, *otherg.gene*).

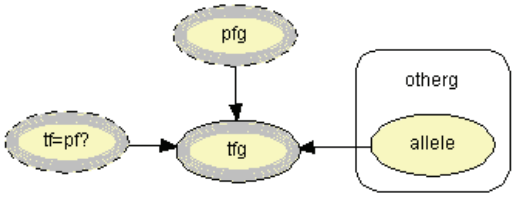


Figure 4: whom network.

The network models the Mendel's inheritance in which the child's allele is chosen at random from the two parents, *pg* and *mg*, here as the sequence of the observed outcome of a fair coin toss. The node *coin* is modeled as a Binomial(1, 0.5), therefore *cg:= if(fcoin.coin = =1, pg, mg)*.

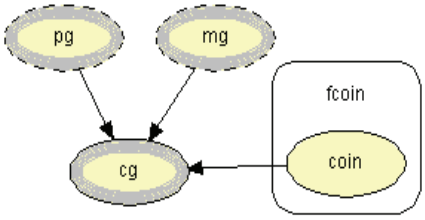


Figure 5: inherit network.

Following Dawid *et al*. (2002), the data for marker FES are child genotype *cgt* = {12, 12}, mother's genotype *mgt* = {10, 12} and putative father's genotype = {10, 12}. The population allele frequencies are p_{10} = 0.28425 and p_{12} = 0.25942. As the authors point out, this simple problem can be easily handled by an algebraic approach. But, the interest in it is to illustrate the simplicity and advantages of using this tool, and to extend its use to more complex problems.

After specifying the network we can put it to run and then insert the evidence. Considering equal prior probabilities for the query node representing the hypotheses, we get the likelihood after inserting the evidence. The likelihood ratio, based on the data for this marker, is obtained from the marginal posterior distribution of the query node. Thus, $P(tf = pf\,?:= true \mid E) = 0.6584$ and $P(tf = pf\,?:= false \mid E) = 0.3416$, and $LR = 1.9274$, being these results in agreement with the algebraic approach.

BN for more complex problems can be built out of the same fundamental local modules that we have already described for the simple problem above, Dawid et al. (2002).

Connect with more complex paternity cases – indirect evidence: only one brother of the putative father available or a brother and another child (with a different mother) of the putative father; or admitting the possibility of mutation in transmission of the putative father's alleles.

2.2 Mixtures

The advances achieved in the forensic biology have certainly encouraged the interest in problems of forensic identification also allowing a much more rigorous treatment of the problems in analysis. That is the case of problems of DNA mixtures - Mortera (2003) and Mortera *et al*. (2003).
One of the complexities in the interpretation of the mixture traces is assigning the number of contributors to the mixture. In general, the trace suggests a lower bound for the total number of contributors but no upper bound. Lauritzen and Mortera (2002) gave a useful low upper bound on the number of contributors worth considering.

In what follows we describe a complex mixture case and present the data to be considered in the analysis. After formulating the hypotheses we perform the analysis for one marker considering the information from one trace. Then we consider the two traces and finally we generalize the analysis considering two mixture traces and the three markers.

The case considered

A crime has been committed, and two persons were murdered, V_1 and V_2. At the scene of the crime two different mixture traces were found: T_1 in the toilet and T_2 in the victims' car. S_2 is a potential suspect. S_2's DNA profile was measured and found to be compatible with the mixture traces.

If we accept that there was a fight during the assault and that produced some material, it is obvious that the individual who perpetrated the crime could have left some of his/her material in some but not in all traces. The non-DNA evidence indicates the possibility that two people were involved in the crime.

Excerpt of data

In order to summarize the evidence we present in Table 1 the DNA profiles of the victims' and the suspect, S_2. In Table 2 we present the profiling results for the mixtures traces (T_1 and T_2), for the STR markers studied, respectively, and the allele frequencies for each marker.

Marker	V_1 (f)	V_2 (m)	S_2
TH01	*D;E*	*D;E*	*B;C*
FES	*A;C*	*C;C*	*B;B*
FGA	*B;E*	*B;C*	*A;C*

Table 1: DNA profiles of the two victims and the suspect

	TH01	FES	FGA
T_1	*B; C; D; E*	*A; B; C*	*A; B; C; E*
T_2	*B; C; D; E*	*B; C*	*A; B; C*
p_A	*	0.0129	0.0684
p_B	0.1696	0.3287	0.1740
p_C	0.1386	0.3664	0.1606
p_D	0.1984	*	*
p_E	0.2748	*	0.0321

Table 2: DNA mixture traces and allele frequencies and allele frequencies[3]

In the traces there is biological material that must belong to some person other than the two victims. The allele frequencies used in this work are the Portuguese population frequencies collected in the worldwide database 'The Distribution of Human DNA-PCR Polymorphisms, since the case mentioned took place in Portugal.

Here we consider that the crime traces can contain DNA from up to three unknown contributors, in addition to the victims and/or the suspect. In what follows we will explain how this is implemented.

If the DNA from S_2 is present in at least one of the traces this will place him at the scene of the crime and consequently as one of the possible perpetrators. Consideration of whether or not the suspect was a contributor to any of the mixture traces will give a measure of the strength.

Hypotheses

The court has to determine if the suspect is or is not guilty. These are described as the level III, or offence, propositions, Cook *et al*. (1998). However the forensic scientist does not typically address such propositions. In this case it appears more appropriate to address source level propositions. Hypotheses to be addressed:

H_1: S_2 is one of the contributors to T_1 but not T_2.

[3] we use * to refer values that are of no concern in the analysis.

H_2: S_2 is one of the contributors to T_2 but not T_1.

H_3: S_2 is one of the contributors to both T_1 and T_2.

H_4: S_2 did not contribute to trace T_1 or T_2.

We are interested in measuring:

$P(S_2 \textit{contributed to at least one of the traces} \mid \xi)$, where ξ is the vector comprising the profiles observed of the traces found at the crime scene, the victims' and the suspect profiles. This is equivalent to

$$P(H_1 \cap H_2 \cap H_3 \mid \xi) = 1 - P(H_4 \mid \xi).$$

2.1.1 One mixture trace and a single marker

The network for one trace and a single marker follows Mortera *et al*. (2003), Fig. 4 section 3.2, an OOBN version considering up to three unknown contributors Fig. 6, **marker** network. Here we present the networks for the marker, FES[4].

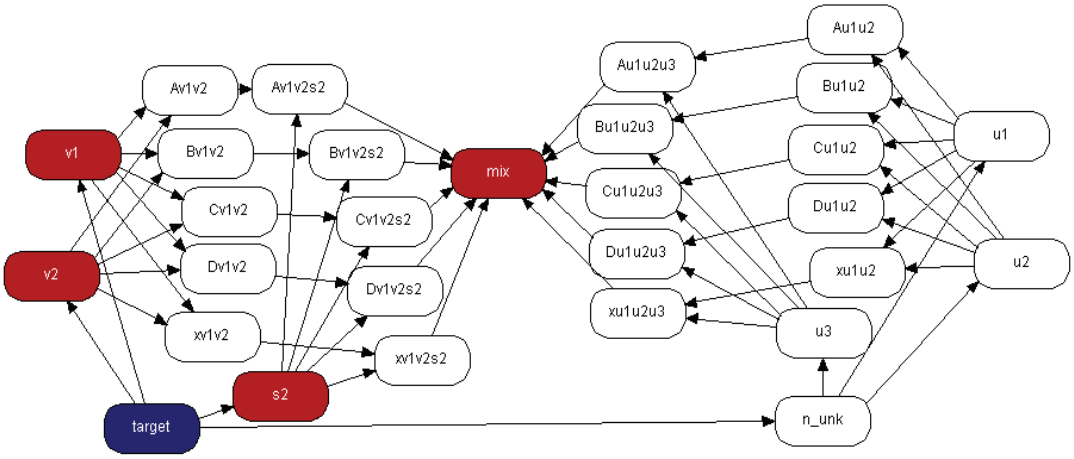


Figure 6: marker network.

The instance **target** follows the reformulation of the query presented by Mortera *et al*. (2003) in order to use simple arithmetic expressions avoiding the tedious construction of the states and tables for the nodes; this is presented in more detail in Fig. 7.

[4]The marker networks differ only in the number of alleles to consider, whether it is the space of states of the nodes referring the alleles or in the presence of one more allele to consider in the network. Since Hugin does not allow modification of the state of a node in order to reuse a network, for markers TH01 and FGA we started with a codification in the space of states of the node *gene* and put it in accordance with the alleles of each marker under consideration so that we could use the same network.

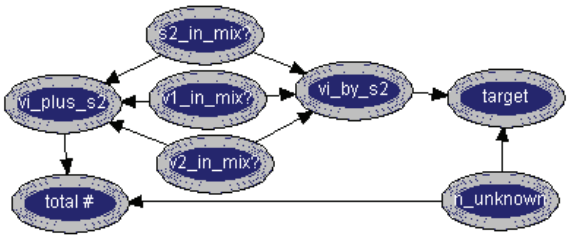


Figure 7: target network.

The node *vi_plus_s*2 takes values 0,1,2,3 according to the number of true states in its parent nodes *v*1*_in_mix*?, *v*2*_in_mix*? and *s*2*_in_mix*?. Consequently the node *total#* takes values from 0 to 6 being given by *vi_plus_s*2 + *n_unknown*. The node *n_unknown* accounts the number of possible unknown contributors to the mixture, between 0 and 3. Node *vi_by_s*2 takes values from 0 to 7, expressing the result values of the one-to-one correspondence with the eight joint configurations of its parent nodes *v*1*_in_mix*?, *v*2*_in_mix*? and *s*2*_in_mix*?. The *target* node has 32 states and is given by *v_ by_s*2 + 8 **n_ unknown*. These 32 states of *target* node describe all the possibilities for contributors to the mixture, i.e., *target* has all states from *v*1&*v*2&*s*2&3*u*, *v*1&*s*2&3*u*,..., *v*1&*v*2&*s*2, ..., *null*. Naturally, the unrealistic hypotheses (those incompatible with the minimum number of contributors) are excluded when the evidence is inserted.
The nodes *v*1*_in_mix*?, *v*2*_in_mix*?, *s*2*_in_mix*?, *target*, *vi_by_s*2, and *n unknown* are given uniform prior distributions. The true or false states of the **ui**'s are indirectly given from the value of *n_unknown* and through the instance **n_unk**, Fig. 8. When *n_ unknown* is 0 then all the **ui**'s, in **marker**, are false, so none of this possible contributors is included in the mixture, when *n_unknown* = 1 then **u1** is included in the mixture and similarly for states 2 and 3 of *n_ unknown*. This information is passed to the **ui**'s through **n_unk** and its respective nodes. Therefore, for example **u1** is considered in the mixture when *n_ unknown* is more or equal to 1, i.e., *n_unk* >= 1 is true if *n_ unknown* is more or equal to 1 else is false. Similarly for **u2** and **u3**.

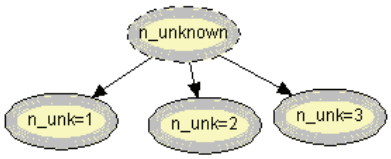


Figure 9: n_unk network.

In the **marker** network we defined a new instance for each **individual**, Fig. 10.

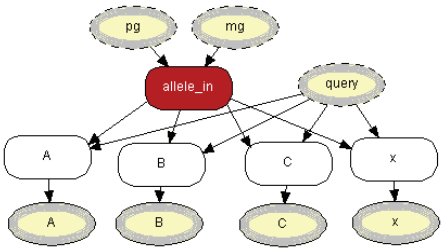


Figure 10: individual network.

For each person, **v1**, **v2**, **s2**, **u1**, **u2**, **u3**, we have a repeated structure in which we consider the genetic background information – the paternal and maternal inheritance, **pg** and **mg**. These are instances of a class named **founder**, a network constituted by a simple *allele* node in which the population allele frequencies are used to specify the unconditional distribution. By taking this approach we implement the product rule that is Hardy-Weinberg and linkage equilibrium assumptions. If a more complex genetic model was desired it could be implemented by introducing dependencies between founder nodes.

The individual's genotype, known for **v1**, **v2** and **s2** and unknown for **u1**, **u2** and **u3**, are indirectly inserted, for the known persons, through the instance **allele_in** shown in Fig. 11.

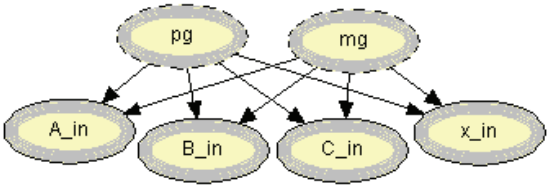


Figure 11: allele_in network.

The instances **A**, **B**, **C** and **x**[5], instances of the class allele, Fig. 12, are expressing the logical conjunction between the *query* node and the presence or absence of the allele in the considered individual, given through **allele_in**. The node *query* represents a binary query mentioning if an individual (for example **v1**) is or is not in the mixture.

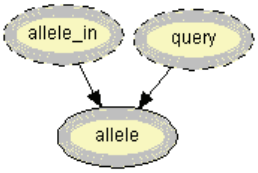


Figure 12: allele network.

The instance **marker** has also an instance named **mix**, Fig. 13, which for each allele, expresses the logical disjunction of the parent instances, e.g., allele A is in the mixture if **Av1v2s2** is true or **Au1u2u3** is true. Here we use *k* to refer that the allele came from the known individual's **v1**, **v2** or **s2**, in the same way that *u* is used for the unknowns.

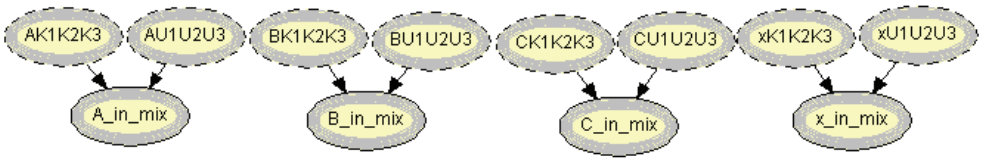


Figure 13: mix network.

[5]For marker FES A, B and C (possibly translate to 8, 9, 11, etc) are the alleles present in the mixture and we use x to represent all the alleles not observed for the marker.

In **marker** the instances **Av1v2**, ..., **xu1u2u3** are expressing the possession of an allele by at least one of the individuals **v1**, **v2**, s**2**, **u1**, **u2**, **u3**, instances of logical disjunction, Fig. 14, i.e, **Av1v2** is true if at least one of **v1** or **v2** has allele *A*. The node *vi_by_s*2 is identical to the same named output node of the instance target, and it refers the probabilities of the state given the evidence. For each trace *vi_by_s*2 is the node measuring the presence of the suspect at the scene of the crime.

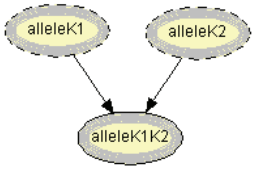


Figure 14: disjunction network.

We can put the network **marker** to run and obtain the results for one of the traces.

2.1.2 Two mixture traces and a single marker

In the case mentioned there were two different traces found at the scene of the crime. So it is necessary to combine the information from both traces. To do so we defined an instance **combine**, Fig. 15. This instance has as parents the output nodes *vi_by_s*2 of the instance **marker** for trace T_1 and trace T_2. The node $T_1\_T_2$ combines the results obtained in the parent instances for node *vi_by_s*2 expressing the result values of the one-to-one correspondence with the eight joint configurations of its parent nodes *v*1*_in_mix*?, *v*2*_in_mix*?, *s*2*_in_mix*? in each trace (*vi_by_s*2*_t*1, *vi_by_s*2*_t*2) for the considered marker.

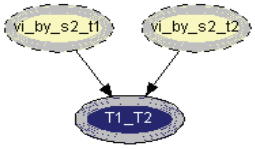


Figure 15: combine network.

Therefore, the node $T_1\_T_2$ takes values 0, 1, 2, 3 corresponding to the hypothesis H_4, H_1, H_2 and H_3, respectively. $T_1\_T_2$ is 0 if *vi_by_s*2 is less than 4 in T_1 and T_2; assumes value 1 if *vi_by_s*2 is equal to 4 or more in T_1 and less than 4 in T_2; takes value 2 if *vi_by_s*2 is less than 4 in T_1 and equal to 4 or more in T_2; and is 3 if *vi_by_s*2 is equal to 4 or more in both T_1 and T_2. We start with a uniform prior distribution for node $T_1\_T_2$.

We are now able to put the networks for each trace together and compute the information in which we are interested, Fig. 16. The instances **FES trace_t1** and **FES trace_t2** are of class **marker** in which all the individuals in any of the networks have the same structure (**individual**). Its differentiation is made when the evidence is inserted.

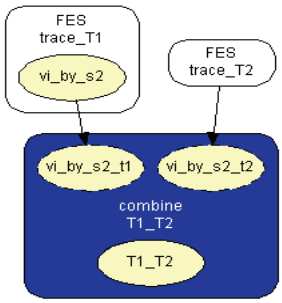


Figure 16: combine_T_1_T_2 network.

When we combine the two traces in order to obtain a measure of the evidential weight associated with the possible presence of genetical material from the suspect in the traces found at the crime scene we get the results listed in the Tables below. For marker FES with different mixture traces we obtain:

S_2, V_2, V_1	trace T_1	trace T_2
0 (FFF)	0.0048	**0.1470**
1 (FFT)	0.1334	**0.0000**
2 (FTF)	0.0068	**0.1791**
3 (FTT)	0.1334	**0.0000**
4 (TFF)	0.0072	**0.1881**
5 (TFT)	0.3526	**0.0000**
6 (TTF)	0.0092	**0.4857**
7 (TTT)	**0.3526**	**0.0000**

Table 3: results of the node *vi_by_s2*

Where the state 0 corresponds to *s2_in_mix*? = *False*, *v2_in_mix*? =*False* and *v1_in_mix*? = *False* (FFF), and for simplicity the state 0 is read as S_2; V_2; V_1 = *FFF*.

In Table 4 it is shown the combined information for the two traces for marker FES.

H_1	**0.2353**
H_2	**0.1876**
H_3	**0.4862**
H_4	**0.0908**

Table 4: results for the node *T1_T2*.

Thus,

$$P(S_2 \text{ contributed to at least one of the traces} \mid \xi) = 0.9092 .$$

2.2 Generalizing two mixture traces and three markers

Given the results obtained for one marker it is necessary to extend the reasoning in order to consider the information for the three markers, FES, TH01 and FGA.

The instances **combine_T_1_T_2** express the results for each marker accounting for the information for the two traces. The node $T_1\_T_2$ in each of these instances computes the results for each marker. Therefore we can extract the respective tables, similar to Table 4, for the other two markers.

The instance **accumulate** having as inputs the output nodes of the instances combine $T_1\_T_2$, with the results of each marker, incorporates the information for the two traces obtained separately, Fig. 17. The node *multi_markers* combines the information from the different instances **combine_T_1_T_2**, i.e., *multi_markers* gives the results synthesizing the results of $T_1\_T_2$ for the three markers. The node *multi_markers* with states 0, 1, 2 and 3 assumes the state 0 if all the input nodes are 0. Takes value 1 if all the input nodes are 1 or at least one of the input nodes has state 1 and the others have the state 0[6]. The node *multi markers* is 2 if all the input nodes have state 2 or this state 2 is combined between the states 0 and 2 of the input nodes. The node assumes state 3 if all the input nodes have state 3 or if the inputs are combining state 0, state 1 and state 2.

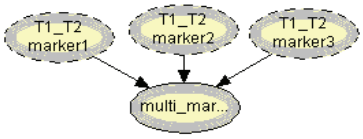


Figure 17: accumulate network.

When we join the networks for the three markers, each of which accounts for the two traces, we obtain the **accumulate_three_markers** network, Fig. 19.

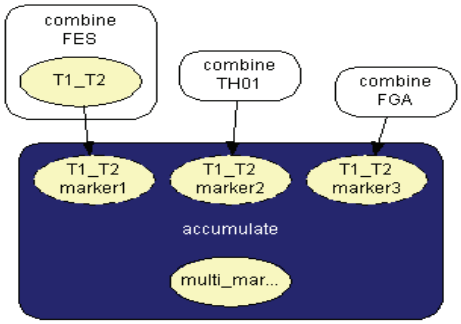


Figure 19: accumulate three markers network.

Tables 5 and 6 display the results for the marker FGA and TH01 and the cumulative result for all three markers, rescaled to sum up to 1. This aims at the question of interest.

S_2, V_2, V_1	**trace T_1**	**trace T_2**	**trace T_1**	**trace T_2**
0 (FFF)	0.0010	0.0084	0.0134	**0.0134**
1 (FFT)	0.0150	0.0000	0.0342	**0.0342**
2 (FTF)	0.0037	0.0476	0.0342	**0.0342**

[6] e.g., *multi markers*=1 if
$T_1\_T_2$ =1 for *marker*1, *marker*2 and *marker*3; or $T_1\_T_2$ =1 for *marker*1 and *marker*2 and $T_1\_T_2$ =0 for *marker*3; or $T_1\_T_2$ =1 for *marker*1 and *marker*3 and $T_1\_T_2$ =0 for *marker*2; or $T_1\_T_2$ =1 for *marker*2 and *marker*3 and $T_1\_T_2$ =0 for *marker*1; or $T_1\_T_2$ =1 for *marker*1 and $T_1\_T_2$ =0 for *marker*2 and *marker*3; or $T_1\_T_2$ =1 for *marker*2 and $T_1\_T_2$ =0 for *marker*1 and *marker*3; or $T_1\_T_2$ =1 for *marker*3 and $T_1\_T_2$ =0 for *marker*1 and *marker*2.

3 (FTT)	0.0290	0.0000	0.0342	**0.0342**
4 (TFF)	0.0079	0.0977	0.0599	**0.0599**
5 (TFT)	0.4644	0.0000	0.2748	**0.2748**
6 (TTF)	0.0146	0.8463	0.2748	**0.2748**
7 (TTT)	**0.4644**	**0.0000**	**0.2748**	0.2748

Table 5: results for the eight configurations for markers FGA and TH01.

$\mathbf{H_1}$	**0.002114**
$\mathbf{H_2}$	**0.001568**
$\mathbf{H_3}$	**0.996313**
$\mathbf{H_4}$	**0.000003**

Table 6: results for the node *T1_T2* for markers FGA and TH01.

Therefore, we can say that,

$$P\left(S_2 \textit{contributed to at least one of the traces} \mid \xi\right) = 0.999997$$

When all the information for the two traces on the three markers is taken into account we get a very significant value for the quantity in which we are interested.

3 Comments

The use of DNA evidence analysis is commonly accepted nowadays in all courts. However, the presentation, interpretation and evaluation of this type of evidence sometimes raise some problems. And we are still far from a total incorporation of this kind of evidence, although in some cases it has been decisive for the conviction or absolution of the individuals. This is already a good support for justice, specially in disputed paternity cases.
The statistical treatment of criminal evidence has raised new challenges to those that have to decide, in the basis of the presented results. Independently of the methodology used, the great difficulty inhabits in the interpretation of the evidence, which is summarized in a number – what does that value means?

In the most complex problems, as the mentioned ones, the use of Bayesian networks for the analysis and interpretation of the evidence can be of great help. In a Bayesian network the complex inter-relations between the variables are transformed into modular units.

This tool – whose use is everyday more common in different areas – supplies, as a support to the decision, a number. It does not give the decision; it is a decision support instrument. Consequently it is important that the legal system knows how to evaluate and interpret correctly the information contained in it. However, there is still much to do.

Acknowledgement

The authors are members of StatMath/UNIDE Research Centre which support they gratefully thank.

References

[1] COWELL, R. G.. DAWID, A. P., LAURITZEN, S. L., SPIEGELHALTER, D. J. (1999). Probabilistic expert systems. Springer, New York.
[2] DAWID, A. P., MORTERA, J., PASCALI, V. L., van BOXEL, D. W. (2002). Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*, **29**, 577-595.
[3] EVETT, I. W., GILL, P. D., JACKSON, G., WHITAKER, J. CHAMPOD, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*, **47**, 520-530.
[4] MORTERA, J. (2003). Analysis of DNA mixtures using probabilistic expert systems. In: Green, P. J., Hjort, N. L., Richardson, S. (Eds.), *Highly Structured Stochastic Systems. Oxford University Press*
[5] MORTERA, J., DAWID, A. P., LAURITZEN, S. L. (2003). Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, **63**, 191-205.

Current address

Marina Alexandra Pedro Andrade, Professor Auxiliar
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa UNIDE - Unidade de Investigação e Desenvolvimento Empresarial, Av. das Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel. +351 217 903 000
e-mail: marina.andrade@iscte.pt

Manuel Alberto Martins Ferreira, Professor Catedrático
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa UNIDE - Unidade de Investigação e Desenvolvimento Empresarial, Av. das Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel. +351 217 903 000
e-mail: manuel.ferreira@iscte.pt

Journal of Applied Mathematics

volume II (2009), number III

ANALYSIS AND MODELLING OF FINANCIAL POWER OF CZECH HOUSEHOLDS

BARTOŠOVÁ Jitka, (CZ)

Abstract. The process of transformation and subsequent process of globalization in the frames of EU resulted in many changes of economic, technical, political, legislative and legal spheres. These significantly (positively and even negatively) influenced the structure of economy and financial power of citizens. The transformation to market economic system, mainly the formation of new income sources and the process of significant differentiation of wages, has caused crucial changes to the income distribution. This paper concentrates on verification of validity of the statistical model of income distribution presently used in the Czech Republic.

Key words. financial power, income distribution, probability model

Mathematics Subject Classification: Primary 46N30; Secondary 62E15

1 Introduction

The knowledge of financial power of the population and the possibility of its observation in various social-economic and spatiotemporal viewpoints is the demand for a well decision making in the sphere of health and social services, tax burden etc. Analysis of financial power of citizens is a datum for decision making in sphere of budget and social policy of the state. Observing the level, structure and progress of financial power of citizens is on the forefront also in a monitoring of expenditures and households amenities, buying intentions, abilities to service mortgages, debts etc.

The speed and the quality of obtaining relevant knowledge is therefore important part of planning on the microeconomic level. There are models for these purposes which allow us to make simple withal sufficiently accurate approximations of very complex situations.

2 Methods of analysis and modelling of financial power

Up to date research focuses on analysis of financial power dynamics and its stability and on revealing factors that significantly affect incomes. This trend shows up in both theoretical and

applied contributions and papers presented at international level. Professor N. T. Longford's paper (Longford – Pittau, 2006) deals with its stability in EU. Detailed analyses and modelling of the state and development of financial power in transitional economics are the main goal of PhD theses of two project participators, J. Bartošová (Bartošová, 2006a) and Ľ. Sipková (Sipková, 2005a).

Among popular methods of income distribution modelling, there are besides traditional methods of modelling (see e.g. Bartošová, 2007, Bartošová – Bína, 2007), also quintile modelling (e.g. Pacáková, 2005, Sipková, 2005). This progressive way of modelling is frequently based on application of properties of a quintile function of generalized lambda distribution (RS GLD), see Ramberg – Schmeiser, 1974), or Paret's generalized distribution (see Luceno, 2006).

2.1 Construction of income models

Economic quantities, such as income, wages, turn-out, profits, expenses etc., are bounded below by nonnegative values. In past the three-parameter logarithm-normal distribution with parameters μ, σ^2 and γ, where γ is the theoretical minimum, represented a good approximation of income distribution. Therefore, the probability distribution function of the chosen model is determined by the following relation

$$f(x;\mu,\sigma^2,\gamma) = \begin{cases} \left(2\pi\sigma^2\right)^{-\frac{1}{2}}(x-\gamma)^{-1} e^{-\frac{(\ln(x-\gamma)-\mu)^2}{2\sigma^2}} & \text{if } x \in (\gamma,\infty) \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

Where $f(x;\mu,\sigma^2,\gamma)$ is probability distribution function and μ,σ^2,γ are parameters of the model.

The basic aim for construction of the theoretic model is its maximum correspondence to the empirical distribution (Bartošová, 2006a, Bílková, 2008). Because of the fact that the sample file of household incomes in year 2005 is sufficiently large for the construction of logarithm-normal models with parameters μ, σ^2 and γ the maximal likelihood method was applied. It is based on the search for argument of the likelihood function supreme.

$$\arg\sup_{(\vec{\theta})} L\left(\vec{\theta}\middle|x_1,...,x_n\right), \tag{2.2}$$

Where $\{x_1,...,x_n\}$ are values of net annual financial incomes of households in particular groups and n is sample size.

The system of likelihood equations for estimation of parameters vector is derived by maximization of the respective likelihood function and could be solved only numerically.

The maximal likelihood estimate of parameter γ of three parameter logarithmic-normal distribution is numerically calculated by the search of the minimum of likelihood ratio (see Anděl, 1985, Bartošová, 2006a)

$$LR(\mu,\sigma^2,\gamma|n) = 2[\ell(\vec{p}|n) - \ell(\vec{\pi}(\mu,\sigma^2,\gamma)|n)], \tag{2.3}$$

Where $\vec{p}$ is vector of income empirical probability, $\vec{\pi}(\mu,\sigma^2,\gamma)$ vector of income theoretical probability and $\ell\left(\vec{p}|n\right)$, $\ell(\vec{\pi}(\mu,\sigma^2,\gamma)|n)$ are corresponding log-likelihood functions.

According to the fact that we treat only finite samples, the maximum likelihood estimates are not guaranteed to have sufficient quality. More detailed information about accuracy of the maximum likelihood estimates of parameter vector $\hat{\vec{\theta}} = (\hat{\mu}, \hat{\sigma}^2, \hat{\gamma})$ in logarithm-normal model of income distribution is presented in dissertation thesis (Bartošová, 2006a).

3 Modelling of income distribution of Czech households in 2005

After the Czech Republic accession to EU the former Microcenzus was replaced by the SILC survey. In 2005, a sample survey SILC was made in 0.15% of Czech households, which represented about 4 000 households. Complete non-aggregated sample set enabled us to gain quality estimates of parameters for distribution models. For the purposes of this research, following data were chosen

- social class of the head of household,
- net income of household (CZK per year).

In connection with the economical transformation in progress, new sources of income arose and social structure of sample sets was changing accordingly. Prior to the "Velvet Revolution", households were divided into classes of workers, cooperative farmers, employees and retired. Social structure of the sample sets of household income from year 2005 is depicted in the Table 1.

Table 1. Structure of data sets. (Data source: SILC 2005).

Social class	Size	%
Employees	2148	49.37
Self-employed	391	8.99
Retired with economically active members	178	4.09
Retired without economically active members	1425	32.75
Unemployed	131	3.01
Others	78	1.79
All	4351	100.00

It is apparent from the Table 1 that two biggest social classes, i.e. employees and retired without economically active members, formed 82.12% of all households in the Czech Republic in 2005. Consequently, the character of income distribution all households will be mainly determined by the manner of income distribution in these three major classes.

3.1 The results obtained

In the Table 2 the values of likelihood ratio LR are written for the constructed logarithm-normal models with two and three parameters. In all social groups greater agreement of empirical distribution with model was achieved for three parametric logarithm-normal models. The use of three parametric logarithm-normal models led to the improvement of models validity. In the construction the method of minimization of likelihood ratio was used in order to estimate the value of parameter γ.

Table 2. Comparison of conformity of empirical distribution with two- and three-parametrical lognormal models. (Data source: SILC 2005).

Social class	Statistic LR		Quantile
	$\gamma = 0$	γ_LR	$\chi^2_{0,95}$
Employees	59.723929	**51.369312**	81.381015
Self-employed	27.306059	**27.260379**	43.772972
Retired with economically active members	21.809623	**21.642089**	30.143527
Retired without economically active members	565.20513	535.01713	68.669294
Unemployed	7.3784923	**7.2391322**	24.995790
Others	13.784844	**12.339053**	22.362032
All	398.93645	397.66954	104.13874

For the consideration of the models validity are the results of the construction supplemented by the values of LR statistics and 95% quantiles of χ^2 distribution. From the Table 2 we could infer that in most cases the values of LR and $\chi^2_{0,95}$ are comparable. Only in the case of retired without economically active members and all households the strong discrepancy between the empirical distribution and the model appears (see Bartošová, 2006b). In both above mentioned cases the LR statistics significantly exceed the value of the corresponding χ^2 quantile. For the retired without economically active members is $LR \approx 10 \cdot \chi^2_{0,95}$ and for all households is $LR \approx 3 \cdot \chi^2_{0,95}$. Those income sets have bimodal distribution (see Figure 1) and could not be modelled using simple parametrical models. Modelling of such mixtures is the topic of previous paper (see Bartošová – Bína, 2007).

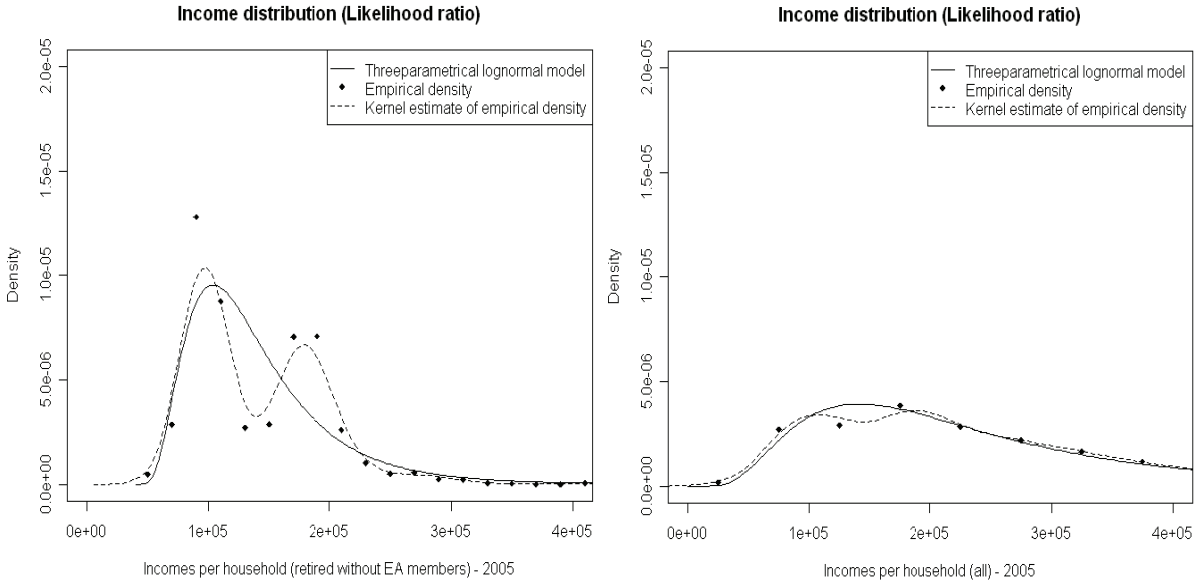


Figure 1. Three-parametrical models, empirical densities and kernel estimates for retired without EA members and all households. (Data source: SILC 2005)

Figure 1 shows the three-parametrical lognormal models, the empirical densities and the kernel estimates of theoretical density of income distribution of retired without economically active members and all households in year 2005. On the mentioned figure we can see that the kernel estimates of theoretical density of the households of retired without economically active members are obviously bimodal, in the case of all households is this effect less significant. Three-parametrical lognormal models, the empirical densities and the kernel estimates of theoretical density of income distribution in the other social classes (employees, self-employed, retired with economically active members and unemployed) are shown in Figure 2.

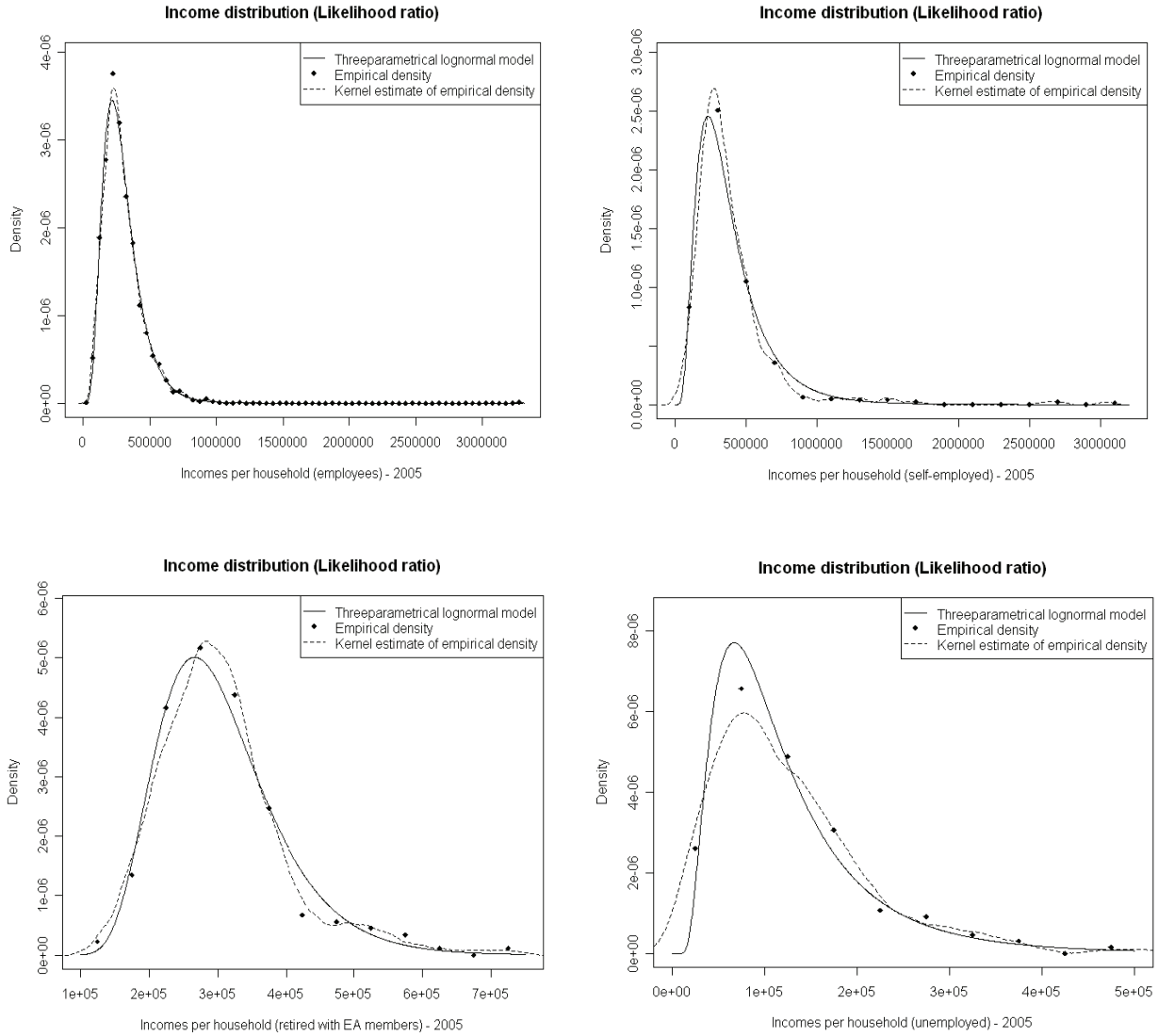


Figure 2. Three-parametrical models, empirical densities and kernel estimates for employees, self-employed, retired with economically active members and unemployed. (Data source: SILC 2005)

4 Conclusions

It can be seen that for the majority of social classes the logarithm-normal distribution can be considered as suitable model of household income distribution. We can observe in the analysis aging of the population and increasing influence of income distribution of retired without economically active members. In this class, there's a significant discrepancy between the empirical distribution and this model. The same situation appears in the case of all households, which is mainly influenced by this class. These income sets have bimodal distribution and could not be modelled using simple parametrical models. Bimodality of income distribution could be in both cases overcome by splitting the files into two subgroups – households with one member and households with more members.

Acknowledgement
The paper was supported by grant from Grant Agency of the Czech Republic no. 420/09/0515 with title "Analysis and modelling of financial power of Czech and Slovak Households".

References

[1.] ANDĚL, J.: *Statistické metody*. Matfyzpres, 1985, Praha.
[2.] BARTOŠOVÁ, J.: *Volba a aplikace metod analýzy stavu rozdělení příjmů domácností v České republice po roce 1990*. Ph.D. Thesis. University of Economics in Prague. Faculty of Informatics and Statistics, 144 p., 2006a.
[3.] BARTOŠOVÁ, J.: *Logarithmic-Normal Model of Income Distribution in the Czech Republic*. In Austrian Journal of Statistics 35 (2&3), pp. 215-222, 2006b.
[4.] BARTOŠOVÁ, J.: *Pravděpodobnostní model rozdělení příjmů v České republice*. Acta Oeconomica Pragensia Vol.15, Nr.1, Statistické a matematické metody v ekonomii, pp.7-12, 2007.
[5.] BARTOŠOVÁ, J., BÍNA, V.): *Mixture models of household income distribution in the Czech Republic*. In: M. Kováčová (Ed.): 6th International Conference APLIMAT 2007, Part I. Slovak Univ. of Technology, Bratislava, pp. 307-316, 2007.
[6.] BÍLKOVÁ, D.: *Application of Lognormal Curves in Modeling of Wage Distributions*. In: 7th International Conference APLIMAT 2008 [CD-ROM]. Universita of Technology, Bratislava, pp. 341–352, 2008.
[7.] LONGFORD, N.T., PITTAU, M.G.: *Stability of household income in European countries in the 1990s*. Computational Statistics & Data Analysis 51, pp. 1364-1383, 2006.
[8.] LUCENO, A.: Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. Computational Statistics & Data Analysis 51, pp. 904-917, 2006.
[9.] RAMBERG, J., SCHMEISER, B.: *An approximate method for generating asymmetric random variables*. Communications of the ACM 17 (2), pp. 78-82, (1974).

Current address

Jitka Bartošová, RNDr., PhD.
University of Economics Prague, Faculty of Management, Jarošovská 1117/II, Jindřichův Hradec, 377 01, Czech Republic, tel. +420 384 417 221,
e-mail:bartosov@fm.vse.cz

Journal of Applied Mathematics

volume II (2009), number III

PARETO DISTRIBUTION AND WAGE MODELS

BÍLKOVÁ Diana, (CZ)

Abstract. Pareto distribution is usually used as a model of the distribution of the largest wages, not for the whole wage distribution. The parameter b of the Pareto distribution is usually called the Pareto coefficient. It can be used as a characteristic of differentiation of 50 % highest wages. The Pareto distribution will be a good model of the wage distribution if the empirical differentiation of 50 % of the highest wages is similar to the differentiation that has the Pareto distribution i.e. the differentiation for which the following ratios are equal: Ratio of the upper quartile to the median; ratio of the eight decile to the sixth decile; ratio of the ninth decile to the eight decile. This property can be used as one of the criterion to measure the quality of fit of the Pareto distribution to some empirical wage distribution. If in a particular case the observed differences of the rates of the above mentioned quantiles are negligible, Pareto distribution will be an appropriate model of the considered wage distribution. In the case the differences are quite material, the approximation of the considered wage distribution with Pareto distribution will be more or less inappropriate. If the Pareto distribution is chosen as a model for a particular distribution we have to keep in mind that this model is only an approximation. The wage distribution will be only approximated and the relations derived from the model will also hold for the "true distribution" only approximately.

Key words. Pareto distribution, Pareto coefficient, estimation methods for parameters, least squares method, wage distributions

Mathematics Subject Classification: Primary 62H10, 62H12; Secondary 62H15.

1 Pareto Distribution

Pareto distribution is usually used as a model of the distribution of the largest wages, not for the whole wage distribution. In this article we will consider using the Pareto distribution to model wages higher than median.

The $100 \cdot P\%$ quantile of the wage distribution will be denoted by x_P, $0 < P < 1$. This value represents the upper bound of $100 \cdot P$ % lowest wages and also the lower bound of $100 \cdot (1 - P)$ % highest wages. A particular quantile (denoted as x_{P0}) which will be the lower bound of some small

number of the highest wages is usually set to be the maximum wage. If the following formula (1) holds for any quantile x_P, the wage distribution is Pareto distribution

$$\frac{x_{P0}}{x_P} = \left(\frac{1-P}{1-P_0}\right)^b. \tag{1}$$

The parameter b of the Pareto distribution (1) is called the Pareto coefficient. It can be used as a characteristic of differentiation of 50 % highest wages.

We will now consider a pair of quantiles x_{P1} and x_{P2}, $P_1 < P_2$. It follows from (1) that

$$\frac{x_{P0}}{x_{P1}} = \left(\frac{1-P_1}{1-P_0}\right)^b \tag{2.1}$$

and

$$\frac{x_{P0}}{x_{P2}} = \left(\frac{1-P_2}{1-P_0}\right)^b. \tag{2.2}$$

From what we can derive for the rate of x_{P2} to x_{P1} that

$$\frac{x_{P2}}{x_{P1}} = \left(\frac{1-P_1}{1-P_2}\right)^b. \tag{3}$$

The rate

$$\frac{x_{P2}}{x_{P1}}$$

is an increasing function of the Pareto coefficient b. If the rate of quantiles increases, the relative differentiation of wages increases too. If only absolute differences between quantiles increase, only the absolute differentiation of wages increases.

It follows from the equation (1) that once the values x_{P0} and b are chosen we can determine the quantile x_P for any chosen P or the other way around for any value x_P we can find the corresponding value of P. In the first case it is advantageous to write the equation (1) as

$$x_P = \frac{x_{P0}}{\left(\frac{1-P}{1-P_0}\right)^b} \tag{4.1}$$

or after logarithmic transformation as

$$\log x_P = \log x_{P_0} - b[\log(1 - P) - \log(1 - P_0)], \tag{4.2}$$

in the second case

$$1 - P = (1 - P_0)\sqrt[b]{\frac{x_{P0}}{x_P}} \tag{5.1}$$

or after logarithmic transformation as

$$\log(1-P) = \log(1-P_0) + \frac{1}{b}(\log x_{P_0} - \log x_P). \tag{5.2}$$

The equations (2) and (3) will after logarithmic transformation have the following form

$$b = \frac{\log \dfrac{x_{P0}}{x_{P1}}}{\log \dfrac{1-P_1}{1-P_0}}, \tag{6.1}$$

$$b = \frac{\log \dfrac{x_{P2}}{x_{P1}}}{\log \dfrac{1-P_1}{1-P_2}}. \tag{6.2}$$

It follows from the equation (6.1) that instead of the Pareto coefficient b we can use any other quantile x_{P1} of the Pareto distribution and it follows from the equation (6.2) that the Pareto coefficient b can be calculated using any known quantiles x_{P1} and x_{P2}. Then we can also determine the value x_{P0} using the formulas

$$x_{P0} = x_{P1}\left(\frac{1-P_1}{1-P_0}\right)^b, \tag{7.1}$$

$$x_{P0} = x_{P2}\left(\frac{1-P_2}{1-P_0}\right)^b. \tag{7.2}$$

The model characterized with the relationship (1) will be practically applicable if the following is known:

- The value of the quantile that characterizes the assumed wage maximum and the value of the Pareto coefficient b;
- The value of the quantile that characterizes the assumed wage maximum and the value of any other quantile;
- The values of any two quantiles of the Pareto distribution.

Any two quantiles can be written as x_P and x_{P+k}, where $0 < k < 1 - P$. Using the equation (3), we can derive for the rate of these two quantiles

$$\frac{x_{P+k}}{x_P} = \left(\frac{1-P}{1-P-k}\right)^b. \tag{8}$$

The rate (8) will be equal for such pairs of quantiles for which the following formula holds

$$\frac{1-P}{1-P-k} = c, \tag{9.1}$$

where c is a constant, i. e. the rate will be the same for all pairs of quantiles for which

$$k = \frac{c-1}{c}(1-P). \tag{9.2}$$

We will use the constant $c = 2$ in (9.2) and we will choose gradually $P = 0{,}5;\ 0{,}6;\ 0{,}8$. Then using the equation (8) we can show the equality of rates of some frequently used quantiles

$$\frac{x_{0,75}}{x_{0,5}} = \frac{x_{0,8}}{x_{0,6}} = \frac{x_{0,9}}{x_{0,8}}. \tag{10}$$

From the relationship (10) we can conclude that Pareto distribution assumes such a wage differentiation for which the rate of the upper quartile to median is the same as:

- The rate of the 8th to the 6th decile;
- And as the rate of the 9th to the 8th decile.

If in a particular case the observed differences of the rates of the above mentioned quantiles are negligible, Pareto distribution will be an appropriate model of the considered wage distribution. In the case the differences are quite material, the approximation of the considered wage distribution with Pareto distribution will be more or less inappropriate.

2 Parameter estimates

If the Pareto distribution is chosen as a model for a particular distribution we have to keep in mind that this model is only an approximation. The wage distribution will be only approximated and the relations derived from the model will also hold for the "true distribution" only approximately. Which relations will hold more precisely and for which the precision will be lower will be mostly dependent on the method of parameter estimates.

There are many possibilities to choose from. In the following text the quantiles of Pareto distribution will be denoted as x_P and the quantiles of the observed wage distribution will be denoted as y_P.

First we need to decide which quantile to choose as x_{P0} . It this article we will assume that $x_{P0} = x_{0,99}$. From the equation (1) we can see that the considered Pareto distribution will be defined by the equation

$$\frac{x_{0,99}}{x_P} = \left(\frac{1-P}{0{,}01}\right)^b. \tag{11}$$

Then we need to determine the value $x_{0,99}$ and the value of the Pareto coefficient b. Because it is necessary to estimate the values of two parameters we need to choose two equations to estimate from.

A natural choice is the equation $x_{P0} = y_{P0}$; that is in our case $x_{0,99} = y_{0,99}$. As the other equation we set a quantile x_{P1} equal to the corresponding observed quantile, i.e. $x_{P1} = y_{P1}$. In this case, the parameters of the model will be

$$x_{P0} = y_{P0} \tag{12.1}$$

and using (6.1)

$$b = \frac{\log \dfrac{y_{P0}}{y_{P1}}}{\log \dfrac{1-P_1}{1-P_0}}. \tag{12.2}$$

We can get different modifications using different choice of the maximum wage and the second quantile. If we use equation $x_{0,99} = y_{0,99}$ and we use the median in the second equation, i.e. $x_{0,5} = y_{0,5}$ we get a model with parameters

$$x_{0,99} = y_{0,99}, \tag{13.1}$$

$$b = \frac{\log \frac{y_{0,99}}{y_{0,5}}}{\log \frac{0,5}{0,01}}. \tag{13.2}$$

Another possibility is setting any two quantiles of the model equal to the quantiles of the observed distribution

$$x_{P1} = y_{P1}, \tag{14.1}$$

$$x_{P2} = y_{P2}. \tag{14.2}$$

Using the formula (6.2) we get the following parameter estimates

$$b = \frac{\log \frac{y_{P2}}{y_{P1}}}{\log \frac{1-P_1}{1-P_2}} \tag{15}$$

and from (7.1) and (7.2) we get

$$x_{P0} = y_{P1}\left(\frac{1-P_1}{1-P_0}\right)^b = y_{P2}\left(\frac{1-P_2}{1-P_0}\right)^b. \tag{16}$$

With this alternative we can also get numerous modifications depending on the choice of quantiles y_{P1} and y_{P2} that are used.

The third possibility is based on the request that $x_{P0} = y_{P0}$ and that the rate of some other two quantiles of the Pareto distribution x_{P2}/x_{P1} is equal to the rate y_{P2}/y_{P1} of correspoding quantiles of the wage distribution observed. In this case we will estimate the parameters using (see (6.2))

$$x_{P0} = y_{P0}, \tag{17.1}$$

$$b = \frac{\log \frac{y_{P2}}{y_{P1}}}{\log \frac{1-P_1}{1-P_2}}. \tag{17.2}$$

In this case notwithstanding that $x_{P2}/x_{P1} = y_{P2}/y_{P1}$ hold, the equality of quantiles itself, $x_{P1} \neq y_{P1}$ and $x_{P2} \neq y_{P2}$, does not hold. In this case we can also arrive to numerous modifications depending on what maximum wage is chosen and what quantiles y_{P1} and y_{P2} are chosen.

For all of the above methods the equality of two characteristics of the model and the observed distribution was required. There are also different approaches to the parameter estimates.

The least squares method is frequently used for the Pareto distribution parameter estimates as well. We will consider the following quantiles of the observed wage distribution $y_{P1}, y_{P2}, \ldots, y_{Pk}$ and corresponding quantiles of the Pareto distribution $x_{P1}, x_{P2}, \ldots, x_{Pk}$. The model distribution will be most precise when the sum of squared differences

$$\sum_{i=1}^{k}(y_{P_i} - x_{P_i})^2 \tag{18}$$

is minimized. In this case closed formula solution does not exist. Therefore sum of squared differences of logarithms of quantiles is often considered

$$\sum_{i=1}^{k}(\log y_{P_i} - \log x_{P_i})^2. \tag{19}$$

Minimizing the objective function (19), it is possible to derive the following estimates

$$b = \frac{k\sum_{i=1}^{k}\log y_{P_i}\ \log\frac{1-P_0}{1-P_i} - \sum_{i=1}^{k}\log y_{P_i}\ \sum_{i=1}^{k}\log\frac{1-P_0}{1-P_i}}{k\sum_{i=1}^{k}\log^2\frac{1-P_0}{1-P_i} - \left(\sum_{i=1}^{k}\log\frac{1-P_0}{1-P_i}\right)^2}, \tag{20.1}$$

$$\log x_{P_0} = \frac{\sum_{i=1}^{k}\log y_{P_i}}{k} - b\frac{\sum_{i=1}^{k}\log\frac{1-P_0}{1-P_i}}{k}. \tag{20.2}$$

In the case we use this estimating method, it is needed to keep in mind that the equality of model quantiles and observed quantiles is not guaranteed for any P. Again we can arrive to different results depending of what quantiles $y_{P1}, y_{P2}, \ldots, y_{Pk}$ are used for the calculations. Furthermore the parameter estimates also depend on the choice of the maximum wage.

3 Characteristics of the appropriateness of the Pareto distribution

For the application of Pareto distribution as a model of the wage distribution, it is crucial that the model fits the observed distribution as close as possible. It is important that the observed relative frequencies in particular wage intervals are as close to the theoretical probabilities assigned to these intervals by the model as possible.

It is needed to note that the same parameter estimation method does not always lead to the best results. It is of particular importance in "what direction" is the observed wage distribution different from Pareto distribution. Pareto distribution assumes such wage differentiation that the relations (10) hold. With real data we can encounter many different situations

$$\frac{y_{0,75}}{y_{0,5}} < \frac{y_{0,8}}{y_{0,6}} < \frac{y_{0,9}}{y_{0,8}}, \tag{21.1}$$

$$\frac{y_{0,75}}{y_{0,5}} > \frac{y_{0,8}}{y_{0,6}} > \frac{y_{0,9}}{y_{0,8}}, \tag{21.2}$$

$$\frac{y_{0,75}}{y_{0,5}} < \frac{y_{0,9}}{y_{0,8}} < \frac{y_{0,8}}{y_{0,6}}, \tag{21.3}$$

$$\frac{y_{0,75}}{y_{0,5}} > \frac{y_{0,9}}{y_{0,8}} > \frac{y_{0,8}}{y_{0,6}}, \tag{21.4}$$

$$\frac{y_{0,8}}{y_{0,6}} < \frac{y_{0,75}}{y_{0,5}} < \frac{y_{0,9}}{y_{0,8}}, \tag{21.5}$$

$$\frac{y_{0,8}}{y_{0,6}} > \frac{y_{0,75}}{y_{0,5}} > \frac{y_{0,9}}{y_{0,8}}. \tag{21.6}$$

It follows from (21) that the observed distributions will more or less systematically differ from the Pareto distribution. In the case of (21.1) the differentiation of the observed wage distribution is higher; in the case of (21.2) the differentiation will be lower than in the case of Pareto distribution. Some bias occurs in cases (21.3), (21.4), (21.5) and (21.6) as well (but cannot be so specified). Systematical bias should be a signal for potential adjustment of the model which could be based for example on adding one or more parameters into the model. These adjustments usually lead to more complicated models. Therefore, the above mentioned bias is often neglected and simple models are preferred even though they lead to some bias.

4 Wage distributions of male and female in Czech Republic in the years 2001 – 2006

The data used in this article is the gross monthly wage of male and female in CZK in the Czech Republic in the years 2001 – 2006. Data were sorted in the table of interval distribution with opened lower and upper bound in the lowest and highest interval respectively. The source is the web page of the Czech statistical office. We were calculated the following quantiles: median $y_{0,50}$, 6th decile $y_{0,60}$, upper quartile $y_{0,75}$, 8th decile $y_{0,80}$, 9th decile $y_{0,90}$ and 99th precentile $y_{0,99}$ of gross monthly wages in the Czech Republic in the years 2001 – 2006 (total and split up to male and female separated). We have found that, with the exception of male in the year 2003, all other wage distributions have lower differentiation than Pareto distribution. The systematical error occurred also in the case of male in the year 2003. It follows from the empirical criterion (10) that in all cases the differences between the rates of the considered quantiles are negligible and therefore Pareto distribution can be used as the model of the distribution.

The 99th percentile will be considered as a characteristic of the maximum wage. The parameters of the Pareto distribution are estimated using the above described methods.

First we consider the conditions $x_{P0} = y_{P0}$ a $x_{P1} = y_{P1}$ and we chose median as the second quantile, i.e. $x_{0,99} = y_{0,99}$ a $x_{0,5} = y_{0,5}$. We estimate the parameter b using the formula (13.2). The summary of the parameter estimates is in the table 1.

Next we apply the conditions $x_{P1} = y_{P1}$ and $x_{P2} = y_{P2}$ and we choose 6th and 9th decile for y_{P1} and y_{P2}. We use the formulas (15) and (16) to estimate the parameters. The summary of the parameter estimates is in the table 1.

Parameters of the Pareto distribution can also be estimated using the equations $x_{P0} = y_{P0}$ and $x_{P2}/x_{P1} = y_{P2}/y_{P1}$. We choose the 9th and 6th decile in the rate y_{P2}/y_{P1}. In this case we use the relations (17) to estimate the parameters. The summary of the parameter estimates is also in the table 1.

In the end we also estimate the parameters of the Pareto distribution using the least squares method. We use the relations (20). In this method we choose 5th, 6th, 7th, 8th and 9th deciles of the observed wage distribution, i.e. $k = 5$. Parameters estimated using the least squares method are summarized in the table 2. The values of the sum of absolute differences of observed and theoretical absolute frequencies of all intervals calculated for all cases considered wage distributions are in the table 3. In the case of the theoretical frequencies at first we determined theoretical probabilities using the formula (5.2). From these we determined theoretical absolute frequencies.

Table 1: Estimated parameters of Pareto distribution for different choice of the estimation equations

		Equations used					
		$x_{0,99} = y_{0,99}, x_{0,5} = y_{0,5}$		$x_{0,6} = y_{0,6}\ , x_{0,9} = y_{0,9}$		$x_{0,99} = y_{0,99}, \frac{x_{0,9}}{x_{0,6}} = \frac{y_{0,9}}{y_{0,6}}$	
		Parameter estimates		Parameter estimates		Parameter estimates	
Total	Year	x_{P0}	b	x_{P0}	b	x_{P0}	b
	2001	44 921	0,326 952	54 143	0,365 843	44 921	0,365 843
	2002	47 172	0,283 758	61 890	0,348 293	47 172	0,348 293
	2003	47 719	0,267 846	64 800	0,340 425	47 719	0,340 425
	2004	56 369	0,295 969	67 096	0,334 192	56 369	0,334 192
	2005	56 852	0,299 456	74 095	0,347 455	56 852	0,347 455
	2006	57 326	0,275 468	79 614	0,354 083	57 326	0,354 083
		Equations used					
		$x_{0,99} = y_{0,99}, x_{0,5} = y_{0,5}$		$x_{0,6} = y_{0,6}\ , x_{0,9} = y_{0,9}$		$x_{0,99} = y_{0,99}, \frac{x_{0,9}}{x_{0,6}} = \frac{y_{0,9}}{y_{0,6}}$	
		Parameter estimates		Parameter estimates		Parameter estimates	
Male	Year	x_{P0}	b	x_{P0}	b	x_{P0}	b
	2001	46 781	0,305 624	61 207	0,367 449	46 781	0,367 449
	2002	48 047	0,265 814	72 613	0,368 246	48 047	0,368 246
	2003	48 417	0,249 540	84 934	0,390 464	48 417	0,390 464
	2004	57 514	0,278 536	78 632	0,353 784	57 514	0,353 784
	2005	57 808	0,267 749	86 165	0,364 658	57 808	0,364 658
	2006	58 104	0,257 739	93 098	0,373 653	58 104	0,373 653
		Equations used					
		$x_{0,99} = y_{0,99}, x_{0,5} = y_{0,5}$		$x_{0,6} = y_{0,6}\ , x_{0,9} = y_{0,9}$		$x_{0,99} = y_{0,99}, \frac{x_{0,9}}{x_{0,6}} = \frac{y_{0,9}}{y_{0,6}}$	
		Parameter estimates		Parameter estimates		Parameter estimates	
Female	Year	x_{P0}	b	x_{P0}	b	x_{P0}	b
	2001	37 526	0,319 087	39 196	0,316 679	37 526	0,316 679
	2002	43 339	0,293 539	47 418	0,308 749	43 339	0,308 749
	2003	44 883	0,283 055	48 172	0,291 217	44 883	0,291 217
	2004	50 776	0,300 989	49 971	0,287 505	50 776	0,287 505
	2005	52 508	0,296 625	54 551	0,297 414	52 508	0,297 414
	2006	54 054	0,291 062	57 954	0,299 456	54 054	0,299 456

Table 2: Parameters estimated using the least squares method

Total	Year					
Parameter estimates	2001	2002	2003	2004	2005	2006
x_{P0}	56 562	64 026	67 219	69 311	76 310	81 721
b	0,379 911	0,358 469	0,351 034	0,344 615	0,356 935	0,362 626
Male	Year					
Parameter estimates	2001	2002	2003	2004	2005	2006
x_{P0}	63 774	73 770	85 080	80 310	88 251	95 225
b	0,379 912	0,372 825	0,391 617	0,360 986	0,372 535	0,381 012
Female	Year					
Parameter estimates	2001	2002	2003	2004	2005	2006
x_{P0}	42 520	49 188	51 125	52 763	57 413	60 917
b	0,341 047	0,320 682	0,309 187	0,303 849	0,312 826	0,315 022

Table 3: Sums of the absolute diferences of the observed and theoretical frequencies

Total	Year	Equations used			
		$x_{0,99} = y_{0,99}$ $x_{0,5} = y_{0,5}$	$x_{0,6} = y_{0,6}$ $x_{0,9} = y_{0,9}$	$x_{0,99} = y_{0,99}$ $\frac{x_{0,9}}{x_{0,6}} = \frac{y_{0,9}}{y_{0,6}}$	Least squares method
	2001	37 459	23 255	85 795	23 859
	2002	51 358	27 327	171 404	31 658
	2003	73 388	36 520	204 535	39 722
	2004	103 625	64 422	249 348	66 249
	2005	167 946	69 930	353 661	68 679
	2006	157 094	68 849	426 442	69 104
Male	Year	Equations used			
		$x_{0,99} = y_{0,99}$ $x_{0,5} = y_{0,5}$	$x_{0,6} = y_{0,6}$ $x_{0,9} = y_{0,9}$	$x_{0,99} = y_{0,99}$ $\frac{x_{0,9}}{x_{0,6}} = \frac{y_{0,9}}{y_{0,6}}$	Least squares method
	2001	20 603	10 089	56 291	9 959
	2002	33 576	19 711	111 796	20 298
	2003	47 909	23 576	96 863	23 747
	2004	60 241	32 457	178 858	33 076
	2005	81 505	35 349	220 276	36 321
	2006	96 789	37 737	250 764	37 653
Female	Year	Equations used			
		$x_{0,99} = y_{0,99}$ $x_{0,5} = y_{0,5}$	$x_{0,6} = y_{0,6}$ $x_{0,9} = y_{0,9}$	$x_{0,99} = y_{0,99,}$ $\frac{x_{0,9}}{x_{0,6}} = \frac{y_{0,9}}{y_{0,6}}$	Least squares method
	2001	24 256	23 926	23 687	21 270
	2002	23 697	16 716	42 148	18 595
	2003	37 215	30 902	40 237	30 011
	2004	45 429	41 416	45 460	40 957
	2005	51 793	41 615	52 493	41 449
	2006	58 014	41 137	74 302	41 812

5 Conclusions

The appropriateness of particular modifications of the Pareto distribution can be evaluated comparing the theoretic and empirical frequencies. It is possible to compare both the absolute and relative differences between the theoretic and observed empirical distributions. In this article we used the absolute differences. The values of sums these differences are in the table 3. The values seem to be relatively high. The question of appropriateness of a given theoretic wage distribution in the case of large samples was described for example in [1]. Some more general conclusions can be made from the values of the absolute differences of observed and theoretic distributions.

With the exception of the wage distribution of women in the year 2001 the worst results are achieved using the equation $x_{0,99} = y_{0,99}$ and setting the ratio of other two quantiles of the Pareto distribution $x_{0,9}/x_{0,6}$ equal to the ratio $y_{0,9}/y_{0,6}$ of the corresponding empirical quantiles. This fact is less obvious for female distribution and most obvious for total distribution. This is also due to the larger sample size of the total sample (in comparison with the sample size of the sub groups of male and female). Again with the exception of the wage distribution of women in the year 2001 the second worst model is the estimate based on the equations $x_{0,99} = y_{0,99}$ and $x_{0,5} = y_{0,5}$. This fact is again less obvious for female distribution and most obvious for total distribution. In the case of the wage distribution of women in the year 2001 the worst estimate is based on the equations $x_{0,99} = y_{0,99}$ and $x_{0,5} = y_{0,5.}$ In the case of the total group is the third worst (second best) method the least squares method (with the exception of the year 2005). The best results are achieved with the method based on the equations $x_{0,6} = y_{0,6}$ and $x_{0,9} = y_{0,9}$. In the case of the total wage distribution in the year 2005 is the third worst method based on the equations $x_{0,6} = y_{0,6}$ and $x_{0,9} = y_{0,9}$ and the best method is the least squares method. In the case of the wage distribution of male (with the exception of the years 2001 and 2006) the third worst (second best) results are again achieved using the least squares method. The best results are achieved with the method based on the equations $x_{0,6} = y_{0,6}$ and $x_{0,9} = y_{0,9}$. In the years 2001 and 2006 (set of men) is the third worst method the method based on the equations $x_{0,6} = y_{0,6}$ a $x_{0,9} = y_{0,9}$ and the best is the least squares method. In the case of the female group (with the exception of the years 2001, 2002 and 2006) is the third worst (second best) method based on the equations $x_{0,6} = y_{0,6}$ and $x_{0,9} = y_{0,9}$ and the most precise results are achieved with the least squares method. In the years 2001, 2003, 2004 and 2005 was for the group of women the most precise the least squares method. The very best method for the group of male in the year 2001 was the least squares method. In this case other methods had much higher values of the above mentioned sum of absolute differences.

From the above described comparison, it is obvious that the simplest parameter estimating methods can be in the case of the Pareto distribution competing with more advanced methods.

References

[1] BÍLKOVÁ, D.: *Application of Lognormal Curves in Modeling of Wage Distributions*. In Journal of Applied Mathematics, No. 7, pp. 341 − 351 + CD, 2008.

[2] NOVÁK, I.: *Paretovo rozdělení jako model mzdových rozdělení (výzkumná práce)*. VŠE, Praha, 1966.

Current address

Diana Bílková, Ing., Dr.
Prague University of Ecomomics, sq. W. Churchill 4, 130 67 Prague3,
Phone: +420 224 095 484, e-mail: bilkova@vse.cz

Journal of Applied Mathematics

volume II (2009), number III

THE EUROPEAN COUNTRIES PRICE LEVEL ANALYSIS USING ROBUST REGRESSION

BLATNÁ Dagmar, (CZ)

Abstract. .A price level in the European countries depends on many indicators of general economic background, employment, innovation and research, science and technology. The values of these indicators vary among the European countries and, consequently the occurrence of outliers can be supposed in the analysis of the price level in the European countries. In such case, the classical statistical approach – the least squares method (LS) may be highly unreliable and the robust regression methods represent an acceptable and useful tool. The analysis performed is based on the data of the year 2006.

Key words. robust regression, *LS* regression , outliers' detection, criteria of robust regression model selection

Mathematics Subject Classification: 62J05; 62F35

1 Regression analysis

The classical statistical approach – the least squares method (*LS*) may be highly unsatisfactory due to the presence of outliers which can be supposed in the analysis of the European countries data. Robust regression techniques are an important complement to the classical least squares (*LS*) regression and are acceptable and useful tools because they provide a good fit to the bulk of the data and exposes the outliers quite clearly.

Robust techniques provide results similar to *LS* regression when the data are linear with normally distributed errors. However, the results can differ significantly when the errors do not satisfy the normality conditions or when the data contain significant outliers.

We can distinguish between different types of observations in regression. Thus, a regression dataset can include four types of points: regular observations, vertical outliers, good leverage points and bad leverage points. Leverage points are observations $(\boldsymbol{x}_i, y_i)$ whose $\boldsymbol{x}_i$ are outlying; that is, $\boldsymbol{x}_i$ deviates from the majority in x – space. We call such an observation $(\boldsymbol{x}_i, y_i)$ a good leverage point if $(\boldsymbol{x}_i, y_i)$ follows the linear pattern of the majority. On the other hand, if $(\boldsymbol{x}_i, y_i)$ does not follow this linear pattern, we call it a bad leverage point. An observation whose $\boldsymbol{x}_i$ belongs to the majority in x –

space but where $(\boldsymbol{x}_i, y_i)$ deviates from the linear pattern is called a vertical outlier. Leverage points attract the *LS* solution toward them, so bad leverage points are often not apparent in classical regression analysis.

First, let us briefly mention the principles of selected robust methods.

The least trimmed squares (*LTS*) estimator is obtained by minimizing $\sum_{i=1}^{h} r_{(i)}^2$, where $r_{(i)}$ is the *i*-th order statistic among the squared residuals written in the ascending order, $h = \left\lfloor n/2 \right\rfloor + \left\lfloor (p+1/2) \right\rfloor$ and [*x*] denotes the largest integer which is less or equal to *x*.

The *MM*-estimates are defined by a three-stage procedure. At the first stage an initial regression estimate is computed – it is consistent, robust, with high breakdown-point but is not necessarily efficient. At the second stage, an *M*-estimate of the errors scale is computed using residuals based on the initial estimate. Finally, at the third stage, an *M*-estimate of *MM* estimates is a combination of high breakdown value estimation and efficient estimate of the regression parameters based on a proper redescending ψ-function is computed.

Reweighted least squares (*RWLS*) regression minimizes the sum of the squared residuals multiplied by weights w_i, which are determined from the *LTS* solution. The effect of the weights, which can only take values 0 or 1, is the same as deleting the cases for which w_i equals zero. Therefore, the *RLS* can be seen as ordinary *LS* on a "reduced" data set consisting of only those observations that received nonzero weights. The *RWLS* estimates have the same breakdown value as the initial *LTS* estimates and a much better statistical efficiency and all the usual inferential output such *t*-statistics, *F*-statistics and R^2 and the corresponding *p*-values can be obtained.

A robust regression with high breakdown point *LTS* can be used to detect outliers, leverage points and influence points (the observations whose inclusion or exclusion result in substantial changes in the fitted model (coefficients, fitted values).

2 Identification of outliers and leverage points

To detect leverage points in higher dimensions we must detect outlying $\boldsymbol{x}_i$ in *x* – space. For this purpose we can use the robust distances RD_i defined as

$$RD(x_i) = \sqrt{[x_i - \mathbf{T(X)}]^T \mathbf{C(X)}^{-1} [X_i - \mathbf{T(X)}]} \tag{1}$$

where **T(X)** and **C(X)** are the robust location and scatter matrix for the multivariates.

On the other hand, we can see whether a point $(\boldsymbol{x}_i, y_i)$ lies near the majority pattern by looking at its standardized *LTS* residual $r_i / \hat{\sigma}$.

Diagnostic plots are provided as a fundamental data mining graphical tools for quick identifying of an outlier and determining whether or not outliers have influence on the classical estimate. A regression diagnostic plot (a plot of the standardized residuals of robust regression versus the robust distances $RD(x_i)$ proposed by Rousseeuw and Zomeren (1990)) indicates the corresponding cutoffs by horizontal and vertical lines. Points for which the standardized *LTS*

residuals exceed the cutoff $\sqrt{\chi^2_{1;1-\alpha/2}}$ are considered as regression outliers, whereas observations for which $RD(x_i)$ exceed the cutoff $\sqrt{\chi^2_{p;1-\alpha/2}}$ are considered as leverage points.

3 Methods of model selection

In the case of the classical *LS* regression, the classical *R*-square and the results of the significance of *t*-tests and *F*-tests are used. In the case of the robust regression, the decision which of candidate model may be preferred is based on the following robust diagnostic selection criteria:
- *Robust index of determination R-squared* for *M* (or *MM*) regression defined as

$$R^2 = \frac{\sum_{i=1}^{n}\rho\left(\frac{y_i-\hat{\mu}}{\hat{s}}\right)-\sum_{i=1}^{n}\rho\left(\frac{y_i-x_i^T\hat{\beta}}{\hat{s}}\right)}{\sum_{i=1}^{n}\rho\left(\frac{y_i-\hat{\mu}}{\hat{s}}\right)} = 1-\frac{\sum_{i=1}^{n}\rho\left(\frac{y_i-x_i^T\hat{\beta}}{\hat{s}}\right)}{\sum_{i=1}^{n}\rho\left(\frac{y_i-\hat{\mu}}{\hat{s}}\right)} \tag{2}$$

- *Robust deviance* defined as the optimal value of the objective function on the σ^2-scale:

$$D = 2\left(\hat{s}\right)^2\sum\rho\left(\frac{y_i-x_i^T\hat{\beta}}{\hat{s}}\right) \tag{3}$$

in formulas (2) and (3) $\hat{\beta}$ is the *MM*-estimator of β, $\hat{\mu}$ is the *MM* estimator of location, and $\hat{s}$ is the *MM*-estimator of the scale parameter in the full model.
- *Significance robust tests of variables* for determining which of two candidate models is preferred: *robust t-test* (*t*-statistics and *p*-value of the robust coefficient estimates for the robust fit computed using a robust covariance matrix for the parameter estimates), *robust F-tests* and *robust Wald test.*
- *Robust selection information criteria* (the best model has the lowest value)
  - *Robust Akaike's Information Criterion* (*AICR*) defined as

$$AICR(p;\alpha,\rho) = 2\sum_{i=1}^{n}\rho(r_{i;p})+\alpha p = 2\sum_{i=1}^{n}\rho\left(\frac{y_i-x_i^T\hat{\beta}}{\hat{\sigma}}\right)+\alpha p \tag{5}$$

$r_{i;p}$ are regression residuals connected with *MM*-estimate of parameters, $\hat{\sigma}$ is robust estimate of σ, p is the number of parameters.

- *Robust Bayesian information criterion* (*BICR*) (sometimes also named the Schwarz information criterion) defined as

$$BICR = 2\sum_{i=1}^{n}\rho\left(\frac{y_i-x_i^T\hat{\beta}}{\hat{\sigma}}\right)+p\ln(n) \tag{6}$$

- *Robust Final Prediction Error* (*RFPE*) is generalized the *AIC* to a robust model. For a *p*-dimensional model of *p* predictor variables, *RFPE* is calculated as

$$RFPE = \sum_{i=1}^{n} \rho\left(\frac{y_i - x_{p,i}^T \hat{\beta}_p}{\hat{s}_p}\right) + p\frac{\frac{1}{n}\sum_{i=1}^{n} \psi^2\left(\frac{r_i}{\hat{s}}\right)}{\frac{1}{n}\sum_{i=1}^{n} \psi'\left(\frac{r_i}{\hat{s}}\right)} \tag{7}$$

where $r_i = y_i - x_{p,i}^T \hat{\beta}_p$ and $\psi = \rho'$ is the derivative of the loss function. When considering a variety of model choices with respect to several different choices of predictor variables, the model with the smallest value of *RFPE* is preferred.

4 Data set and variables analyzed

The data set analyzed contains the information about 29 European countries (27 members of the European Union, one candidate country (Turkey) and one non EU member (Norway). The data were obtained from Eurostat. The analysis is based on the data of the year 2006 (the last year for which the data for analyzed countries were available).

The indicator analyzed (a dependent variable) is the Comparative Price Level (*PL*) which is indicator of price level differences across countries. The price level of each country in the comparison is expressed in relation to a group of countries (EU-27 =100%). However, the *PL* refer to the aggregate "Household Final Consumption expenditure", which includes expenditure on goods and services purchased by households. The *PL* for a given country is calculated as the ratio between that country's Purchasing Power Parity (PPP) and the nominal exchange rate of its currency against the euro.

The following indicators from different economic fields as explanatory variables have been taken into account.

- Gross domestic product per capita in purchasing power standards (*GDP*)
- Real GDP growth rate - Percentage Change (*gGDP*)
- Labour productivity per person employed (GDP in PPS per person employed) (*LP*)
- Total investment (% of GDP) (*INV*)
- Total state aid - % of GDP (*SA*)
- Inflation rate (annual average rate of chance) (*IR*)
- Total employment rate - % (*EM*)
- Unemployment rate – total (as a share of the total active population) (*UN*)
- Electricity prices [Euro per kWh] (*EP*)
- Energy intensity of the economy-gross inland consumption of energy divided by GDP (*EN*]
- Gross domestic expenditure on R&D - % of GDP (*GE*)
- ICT expenditure - % of GDP – communication expenditure (*ICT*)
- ICT expenditure - % of GDP- information technology expenditure (*IT*)
- Turnover from innovation (% of innovative enterprises) (*IN*)
- Export of high technology product as a share of total exports (*HT)*
- Percentage of households having access to the internet at home (*INT*)
- Total population having completed at least upper secondary education (*ED*)
- Early school leavers – total (*EX*)

5 Regression analysis results of dependent variable the Comparative Price Level

The following regression methods have been applied in our analysis: least squares methods (*LS*), least trimmed squares regression (*LTS*), *MM* – regression, Reweighted least squares method (*RWLS*).

Software SAS 9.1 and S-PLUS 6.2. have been used. In the case of classical *LS* regression, the classical *R*-square and the results of significance of *t*-tests and *F*-tests were applied to selection acceptable model. In the case of robust regression, the decision which of candidate models may be preferred is based on the above mentioned robust diagnostic selection criteria.

As an example, the results of dependence of the Comparative Price Level (*PL*) on the combination of the explanatory variables *GDP* and *GE* are demonstrated. This model belongs to acceptable ones from all points of view and satisfies the recommended ways for model selection. In this case two outliers and eight leverage points were detected by using *LTS* regression (the summary of the robust diagnostic is shown in Table 1, another diagnostic criteria see Table 4).

Table 1 Robust diagnostics

Observation	Mahalanobis distance	Robust MCD Distance	Leverage	Stand. Rob. Residual	Outlier
4 Dennmark	1.1053	1.6135		2.6269	*
7 Ireland	0.9951	3.2144	*	0.6232	
8 Greece	1.0383	2.5526	*	0.3953	
11Italy	0.8551	2.7716	*	-0.3560	
12 Cyprus	1.1914	2.6570	*	8.8735	
15Luxembourg	3.8622	10.5086	*	-9.4022	*
24 Finland	2.3779	3.9083	*	0.9963	
25 Sveden	2.6722	4.2748	*	0.0096	
27 Norway	1.7905	5.1902	*	-0.4152	

As we can see, one observation (15 Luxembourg) are identified both as a leverage point and an outlier. Similar result can also be seen from the graphical outlier detection tool - Standardized Residuals vs. Robust Distances Plot. (see Figure 1).

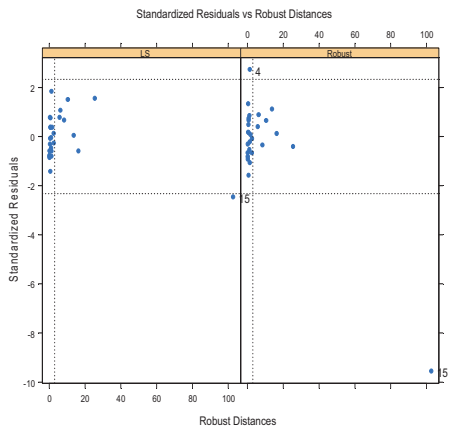


Figure 1

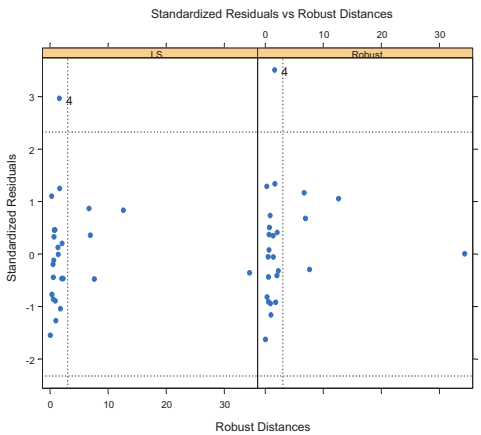


Figure 2

Points outside the horizontal lines are regarded as residual outliers, and points to the right of the vertical line are leverage points. In our case, the *LS* fit produces one residual outlier, whereas the robust fit produces two outliers and eight leverage points. One point (15) is both outlier and leverage point. The interpretation is that this point has substantial influence on the *LS* fit. In such case when outliers are identified, the difference between the *LS* fit and the robust fit can be anticipated. In Table 2, the model fitting for example above-mentioned (*PL ~ GDP + GE*) is presented.

Table 2 Model fitting results

Method	Coefficients	Value	Std.error	t-value	Pr(>|t|)	Chi-sq.	P(>Chi)
Robust	Intercept	22.2101	4.6349	4.7920	0.0001		
RWLS	Intercept	22.2821	4.5857			23.61	<0.0001
LS	Intercept	40.4888	7.6599	5.2858	0.0000		
Robust	*GDP*	0.6181	0.0549	11.2510	0.0000		
RWLS	*GDP*	0.6177	0.0544			129.15	<0.0001
LS	*GDP*	0.2944	0.0689	4.2721	0.0003		
Robust	*GE*	4.5407	2.2249	2.0409	0.0524		
RWLS	*GE*	4.4774	2.2013			4.14	0.0420
LS	*GE*	13.3538	3.8762	3.4451	0.0021		

R-squares are 0.6822 for the *LS* fit and 0.7157 for the robust fit. However, with respect to existing bad leverage points and outliers, the use of a robust model is recommended.

If no outliers and bad leverages are identified, *LS* and robust regressions should provide similar results. As an example, we present the dependence of *PL ~ ICT + GE.* In these cases, results of the classical *LS* regression are quite satisfactory. The results obtained by robust regression are very close to the fit obtained by the classical *LS* regression.

Further, very close coincidence in results obtained with the use of robust and classical regressions arises even in cases, when the same outliers are identified both by robust and classical diagnostic tools. This case can be demonstrated on the example of dependence of *PL ~ GDP+ IN* where one outlier 4 (Dennmark) has been identified (but this point was not a bad leverage point). The results of this case are compiled in Table 3. The graphical diagnostic can be seen from Figure 2.

Table 3 Model fitting results

Method	Coefficients	Value	Std.error	t-value	Pr(>|t|)	Chi-sq.	P(>Chi)
Robust	Intercept	-7.5337	12.5463	-0.6005	0.5543		
RWLS	Intercept	-7.5337	11.6073			0.42	0.5163
LS	Intercept	-9.1872	14.9773	-0.6134	0.5459		
Robust	*GDP*	0.5204	0.0777	6.6931	0.000		
RWLS	*GDP*	0.5204	0.0719			52.34	<.0001
LS	*GDP*	0.5457	0.0925	5.8989	0.0000		
Robust	*IN*	1.9930	0.7779	2.5621	0.0178		
RWLS	*IN*	1.9930	0.7197			7.67	0.0056
LS	*IN*	2.0113	0.9292	2.1646	0.0415		

To describe the dependence of the Comparative Price Level (*PL*) in the European countries, other acceptable regression models can be applied and considered as convenient. Some of them are compiled in Table 4, where the results of both the *LS* and robust fit and goodness-of-fit tests for the robust models are presented.

In Table 4, only regression models suitable both in term of goodness-of-fit tests and satisfying *t* and chi-square tests for individual parameters are presented. More complicated models (with at least four parameters) didn't provide significantly better results in the most examples. The last two models obtained using backward stepwise selection with *RFPE* contain some parameters with non-significant tests. Points which are both outliers and leverage points are marked with bold type.

Table 4 Some of final competitive models

Outliers Leverage points	Robust *MM* model *LS* model	*R*-sq.	*AICR*	*BICR*	*RFPE*
*4,**15*** 7,11,**15**, 24,25,27	22.210 + 0.618 *GDP* + 4.541 *GE* 40.489 + 0.294 *GDP* + 13.354 *GE*	0.7157 0.6822	22.145	29.111	14.123
*4,**15*** 2,3,6,13.14,**15**,20,21,23,27	39.189 +0.570 *GDP* – 0.012 *EN* 88.746 + 0.188 *GDP* – 0.039 *EN*	0.7088 0.7296	22.573	29.661	14.544
*4,**15*** 2,12,**15**,16,21,24,25	11.776 + 0.742 *GDP* + 7.266 *SA* 43.870 + 0.392 *GDP* + 7.118 *SA*	0.7016 0.4890	21.248	28.227	14.096
*4,**15*** 2,12,**15**,16,21,24,25	8.352 + 0.783 *LP* + 11648 *SA* 22.383 + 0.646 *LP* + 7.892 *SA*	0.6366 0.6715	21.664	29.136	17.480
*4,**25*** 2,3,6,8,12,13,17,19,21,22, **25**	8.3784 -0.023 *EN*+ 5.932 *IN* 9.462 – 0.0274 *EN* + 4.105 *IN*	0.6799 0.7910	20.439	27.374	17.970
4 2,6,8,12,19,28	34.779-7.677 *ICT* +3.698 *IN* 36.198-8.116 *ICT* +3.762 *IN*	0.7229 0.8006	16.967	23.683	15.298
4 7,8,11,22,25	-7.533+0.520 *GDP* +1.993 *IN* -9.187 +0.546 *GDP* +2.011 *IN*	0.7658 0.8896	19.667	26.159	12.887
4 8,22,25	-11.038 + 0.586 *LP* + 1.974 *IN* -16.222 + 0.578 *LP* + 2.296 *IN*	0.7779 0.8549	20.471	26.797	12.153
0 2,6,12,22,23	121.314– 11.693 *ICT* + 10.047 *GE* 121.314 11.693 *ICT* + 10.047 *GE*	0.6752	19.999	25.995	12.895
*4,**15*** 9,12,**15**,16,20,21,24	36.708 +0.735 *GDP* +7.652 *SA*- 0.302 *ED* 60.951 +0.383 *GDP* +7.377 *SA* - 0.203 *ED*	0.7323 0.4946	20.973	30.581	13.655
*4,7,**12,21,24*** 2,3,**12**,16.**21**,**24**,25	95.109– 0.035 *EN* + 2.156 *GE* - 18.244 *SA* 93.376 -0.037 *EN* + 11.558 *GE* - 5.144 *SA*	0.5339 0.8042	17.681	27.084	24.364
*4,7,**19*** 2,6,12,14,16,**19**,21,22,24	51.325 -0.319 *EM*-11.497 *SA* - 0.024 *EN* +2.739 *IN* + 7.417 *GE* -13.614+0.701*EM*– 3.619 *SA* - 0.028 *EN* + 2.968 *IN* + 3.455 *GE*	0.7763 0.8555	13.237	28.013	15.446
4,7,12,21,24 2,**4**,6,**7,12**,14,15,17,18,**21**, 2223,**24**,25	167.054-3.334 *gGDP* + 7.985 *INF*- 2.264 *UN* -20.324*ICT* +0.240 *HTE* 163.92 +0.168 *gGDP*-0.685 *INF*- 3.093 *UN* -13.019 *ICT* +0.130 *HTE*	0.7161 0.7836	17.432	31.450	21.119

6 Conclusion

To select an acceptable regression model, the following way recommended in literature can be applied: compare the *LS* and robust *MM*-estimate, if there is a significant difference, use results of robust regression method with high breakdown point (*MM*) especially when outliers and bad leverage points have been identified. Another applicable recommendation is to use backward stepwise variable selection with *RFPE* for selecting variables included in the final model. *RFPE* is computed at each step, and a variable is eliminated only if *RFPE* goes down. But the results obtained by this last approach can provide also nonsignificant tests of parameters in some cases. When no outliers and bad leverage points are identified, the LS fits are satisfactory.

With a view to existing outliers and bad leverages, it is recommended to prefer a robust regression model against the classical *LS* one in the most cases. Unambiguous selection of suitable model describing the dependence of Comparative Price Level in European countries on selected set of explanatory variables is impossible unless we prefer only one criterion for the selection of suitable robust model. All resulting competitive robust regression models presented can be considered as satisfactory from statistical point of view. But the following economic stand-point is necessary for the selection of the final model to describe the dependence of Comparative Price Level in European countries on selected set of explanatory variables

References

[1] BLATNÁ, D. *Robust model selection criteria.* In: Applications of Mathematics and Statistics in Economy. Banská Bystrica: Univerzita M.Béla, 2007, Občianske združenie Financ, 2007, p. 16–22. ISBN 978-80-969535-7-8.

[2] BLATNÁ, D. *Outliers in Regression*. Trutnov 30.08.2006 – 03.09.2006. In: *AMSE 2006 [CD-ROM]*. Praha : KSTP VŠE, 2006, p. 1–6.

[3] BLATNÁ, D. *Robust Regression*. Statystyka i ryzyko, 2007, roč. 1, č. 13, p. 19–29. ISSN 0324-8445.

[4] BLATNÁ, D. *Robust Regression in Analysis of Internet Access in European Countries*. Bratislava, Aplimat 2008, Slovak University of Technology. p.1053-1061. ISBN 978-80-89313-03-7.

[5] HUBERT, M., ROUSSEEUW, P.J., Van AELST. *High-Breakdown Robust Multivariate Methods.* Statistical Science 2008, Vol 23, No.1 ,p. 92-119. ISSN0883-4237

[6] OLIVE,D.: *Applied Robust Statistics*. Preprint M-02-006., http://www.math.siu.edu/

[7] *Regression with SAS*. http://www.ats.ucla.edu/stat/sas

[8] *Robust regression*. http://en.wikipedia.org/wiki/Robust_regression.

[9] ROUSSEEUW, P.J., LEROY,A.M. *Robust Regression and Outlier Detection*. J.Wiley, New Jersey 2003. ISBN 0-471-48855-0.

[10] SAS 9.1.3 Help and Documentation.

[11] S-PLUS 6 Robust Library. User's Guide.

[12] The home of the S-PLUS statistical software package. http://www.insightful.com/

Current address

doc. Ing. Dagmar Blatná, CSc.
Department of Statistics and Probability, University of Economics, Prague
W.Churchill sq.4, Prague 3, Czech Republic
e-mail:blatna@vse.cz

Journal of Applied Mathematics

volume II (2009), number III

ESTIMATION OF THE PARAMETRS IN THE MIXED LINEAR MODEL WITH TYPE II CONSTRAINTS

BOHÁČOVÁ Hana, (CZ)

Abstract. This paper deals with the estimation of the fixed effects parameters and the variance-covariance matrix parameters in the mixed linear model with type II constraints. These estimates are based on the estimates of the parameters of the mixed linear model without constraints. The derived estimators are used in the numerical study of an uniformly accelerated movement.

Key words. Linear regression model with type II constraints, fixed effects parameters, variance components, uniformly accelerated movement

Mathematics Subject Classification: Primary 62J10, 62J12; Secondary 62P30.

1 Basic symbols

I	identity matrix
A^+	Moore-Penrose generalized inverse of the matrix A
$\mathcal{M}(A)$	column space of the matrix A
M_A	projection matrix (in the Euclidean norm) on a space orthogonal to $\mathcal{M}(A)$, ($M_A = I\text{-}AA^+$)
$r(A)$	rank of the matrix A
$tr(A)$	trace of the matrix A, it is defined for square matrices as a sum of its diagonal elements
$Y \sim (X\beta, \Sigma)$	the mean value of the n-dimensional random vector Y is $X\beta$ and its covariance matrix is Σ
$Var_{\theta 0} [\hat{\beta}(\theta)]$	the variance-covariance matrix of the estimator $\hat{\beta}$ when the parameter θ_0 is under consideration
θ_i	the i-th component of the vector θ

2 Mixed linear model with type II constraints

We can meet a situation when we need to estimate a value of some variable that cannot be measured but depends on some other variables, values of which can be obtained. One of such cases is when we need to determine the starting acceleration which is achieved by the car take-off. This acceleration can be estimated by the help of the measured values of time and distance.

Such situations can be described by the mixed linear model with type II constraints:

$$Y \sim_n (X\beta_1, \Sigma_\theta) , \tag{2.1}$$

$$B_1\beta_1 + B_2\beta_2 + b = 0 . \tag{2.2}$$

This linear model is usually a result of linearization as we will see in the numerical study. Parameter β_1 is indirectly measurable, parameter β_2 is to be determined from the condition (2.2) after we estimate β_1. Let's assume X is a $n \times k$ matrix of full column rank k, B_1 is of $q \times k_1$ dimension, dimension of B_2 is $q \times k_2$ and $r(B_1, B_2) = q$, $r(B_2) = k_2$, $k_2 < q < k_1 + k_2$.

The variance-covariance matrix Σ_θ is supposed to be of a form

$$\operatorname{var} Y = \Sigma_\theta = \sum_{i=1}^{r} \theta_i V_i , \tag{2.3}$$

here $\theta_1,...,\theta_r$ are unknown parameters – so-called variance components and $V_1,...,V_r$ are known symmetrical matrices. Whereas Σ_θ is supposed to be at lest positive semidefinite. The task is to estimate $\beta_1, \beta_2, \theta_1,...,\theta_r$.

3 Estimators of the parameters of the mixed linear model without constraints

Let us consider a mixed linear model without constraints

$$Y \sim_n (X\beta, \Sigma_\theta) . \tag{3.1}$$

X is a known matrix of a $n \times k$ dimension which is of a full column rank k. $\beta = (\beta_1, \ldots, \beta_k)'$ is a vector of unknown fixed effects parameters. The variance-covariance matrix is of a form

$$\operatorname{Var} Y = \Sigma_\theta = \sum_{i=1}^{r} \theta_i V_i , \tag{3.2}$$

where $V_1, \ldots, V_r$ are known symmetrical matrices and $\theta_1, \ldots, \theta_r > 0$ are unknown variance components. As we consider a regular model we take into account only $\theta = (\theta_1, \ldots, \theta_r)' \in \mathcal{S}$, where $\mathcal{S}$ is a set of such θ for which Σ_θ is positive definite.

Let us use a given $\theta_0 \in \mathcal{S}$ as a starting value of the variance components, it is usually received from the realization of the observation vector Y. We know that the θ_0–locally best linear unbiased estimator of the fixed effects parameters is (see [1])

$$\hat{\beta}(\theta_0) = [X'(\Sigma_{\theta 0})^{-1}X]^{-1}X'(\Sigma_{\theta 0})^{-1}Y. \tag{3.3}$$

Here $\Sigma_{\theta 0}$ denotes a matrix of type (3.2) with θ_0 instead of θ. The variance-covariance matrix of this estimator (when we consider θ_0 to be the real value of θ) is

$$\operatorname{Var}_{\theta 0}[\hat{\beta}(\theta_0)] = [X'(\Sigma_{\theta 0})^{-1}X]^{-1}. \tag{3.4}$$

The variance components can be estimated via MINQE (MInimum Norm Quadratic Estimator) method. According to [3] the θ_0-MINQE of the variance components in model (3.1) is (according to [3])

$$\hat{\theta} = S^{-1}_{(M_X\Sigma_{\theta 0}M_X)^+}\begin{pmatrix} Y'\left(M_X\Sigma_{\theta 0}M_X\right)^+ V_1\left(M_X\Sigma_{\theta 0}M_X\right)^+ Y \\ \cdots \\ Y'\left(M_X\Sigma_{\theta 0}M_X\right)^+ V_r\left(M_X\Sigma_{\theta 0}M_X\right)^+ Y \end{pmatrix}. \tag{3.5}$$

Here $S_{(M_X\Sigma_{\theta 0}M_X)^+}$ denotes a matrix with

$$\left\{S_{(M_X\Sigma_{\theta 0}M_X)^+}\right\}_{i,j} = tr\left[V_i\left(M_X\Sigma_{\theta 0}M_X\right)^+ V_j\left(M_X\Sigma_{\theta 0}M_X\right)^+\right] \tag{3.6}$$

on its *(i,j)*-th position, for $i, j = 1,...,r$.

The variance-covariance matrix of estimator (3.5) is (when we consider θ_0 to be the real value of the variance components)

$$\text{Var}_{\theta 0}[\hat{\beta}(\theta_0)] = 2S^{-1}_{(M_X\Sigma_{\theta 0}M_X)^+}. \tag{3.7}$$

When computing the estimates we usually use the Moore-Penrose generalized inverse $\left(M_X\Sigma_{\theta 0}M_X\right)^+$ in this form (cf. [2]):

$$\left(M_X\Sigma_{\theta 0}M_X\right)^+ = \Sigma^{-1}_{\theta 0} - \Sigma^{-1}_{\theta 0}X\left(X'\Sigma^{-1}_{\theta 0}X\right)^{-1}X'\Sigma^{-1}_{\theta 0}. \tag{3.8}$$

4 Estimators of the parameters of the mixed linear model with type II constraints

Estimators of the unknown parameters in model (2.1), (2.2) can be derived from (3.3) and (3.5) when we transform model with type II constraints to an equivalent model without constraints (see [2] for more details on this transformation). Assume that K_1 is a $\left(k_1, k_1 + k_2 - q\right)$ matrix and K_2 is a $\left(k_2, k_1 + k_2 - q\right)$ matrix meeting condition

$$\left(B_1, B_2\right)\begin{pmatrix} K_1 \\ K_2 \end{pmatrix} = 0\ . \tag{4.1}$$

Let $\begin{pmatrix} \beta_1^{(0)} \\ \beta_2^{(0)} \end{pmatrix}$ be any vector satisfying $B_1\beta_1^{(0)} + B_2\beta_2^{(0)} + b = 0$. Then we can express:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1^{(0)} \\ \beta_2^{(0)} \end{pmatrix} + \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}\gamma, \tag{4.2}$$

for $\gamma \in R^{k_1+k_2-q}$. Model (2.1), (2.2) is equivalent to a model

$$Y - X\beta_1^{(0)} \sim_n \left(XK_1\gamma, \Sigma_\theta\right) \tag{4.3}$$

without constraints. What we need to do is to estimate parameters γ and $\theta_1,...,\theta_r$ of model (4.3) according to (3.3) and (3.5). We get

$$\hat{\gamma} = (K_1'X'\Sigma^{-1}_{\theta 0}XK_1)^{-1}K_1'X'\Sigma^{-1}_{\theta 0}(Y - X\beta_1^{(0)}) \tag{4.4}$$

and

$$\hat{\theta} = S^{-1}_{\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+}\begin{pmatrix} \left(Y - X\beta_1^{(0)}\right)'\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+ V_1\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+\left(Y - X\beta_1^{(0)}\right) \\ \cdots \\ \left(Y - X\beta_1^{(0)}\right)'\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+ V_r\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+\left(Y - X\beta_1^{(0)}\right) \end{pmatrix}, \tag{4.5}$$

where

$$\left\{S_{\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+}\right\}_{i,j} = tr\left[V_i\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+ V_j\left(M_{XK_1}\Sigma_{\theta 0}M_{XK_1}\right)^+\right] \; i,j = 1,\ldots,r. \tag{4.6}$$

(4.5) can be directly used as an estimator of variance components also in model (2.1), (2.2). Estimators for β_1 and β_2 will be based on (4.2) and (4.4). These estimators must satisfy

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \beta_1^{(0)} \\ \beta_2^{(0)} \end{pmatrix} + \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}\hat{\gamma}\,. \tag{4.7}$$

According to (4.7) we have for $X\hat{\beta}_1$:

$$\begin{aligned} X\hat{\beta}_1 &= X\beta_1^{(0)} + XK_1\hat{\gamma} = X\beta_1^{(0)} + XK_1(K_1'X'\Sigma_{\hat{\theta}}^{-1}XK_1)^{-1}K_1'X'\Sigma_{\hat{\theta}}^{-1}(Y - X\beta_1^{(0)}) \\ &= X\beta_1^{(0)} + XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}(Y - X\beta_1^{(0)}) \\ &= X\beta_1^{(0)} + XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}(Y - X(M_{B_1'M_{B_2}} + P_{B_1'M_{B_2}})\beta_1^{(0)}) \\ &= X\beta_1^{(0)} + XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}Y \\ &\quad - XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}X(M_{B_1'M_{B_2}} + P_{B_1'M_{B_2}})\beta_1^{(0)} \\ &= X\beta_1^{(0)} + XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}Y \\ &\quad - XM_{B_1'M_{B_2}}\beta_1^{(0)} - XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XB_1'M_{B_2}(M_{B_2}B_1B_1'M_{B_2})^+ \\ &\quad \times M_{B_2}B_1\beta_1^{(0)} = XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}Y + XP_{B_1'M_{B_2}}\beta_1^{(0)} \\ &\quad - XM_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XP_{B_1'M_{B_2}}\beta_1^{(0)} \\ &= P^{\Sigma_{\hat{\theta}}^{-1}}_{XM_{B_1'M_{B_2}}}Y + M^{\Sigma_{\hat{\theta}}^{-1}}_{XM_{B_1'M_{B_2}}}XB_1'M_{B_2}(M_{B_2}B_1B_1'M_{B_2})^+ M_{B_2}B_1\beta_1^{(0)} \\ &= P^{\Sigma_{\hat{\theta}}^{-1}}_{XM_{B_1'M_{B_2}}}Y - M^{\Sigma_{\hat{\theta}}^{-1}}_{XM_{B_1'M_{B_2}}}XB_1'M_{B_2}(M_{B_2}B_1B_1'M_{B_2})^+ M_{B_2}B_1(b + B_2\beta_2^{(0)}) \\ &= P^{\Sigma_{\hat{\theta}}^{-1}}_{XM_{B_1'M_{B_2}}}Y - M^{\Sigma_{\hat{\theta}}^{-1}}_{XM_{B_1'M_{B_2}}}XB_1'M_{B_2}(M_{B_2}B_1B_1'M_{B_2})^+ M_{B_2}B_1b. \end{aligned}$$

For the estimator of β_1 we have

$$\begin{aligned} \hat{\beta}_1 &= M_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}Y \\ &\quad -[I - M_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}X]B_1'(M_{B_2}B_1B_1'M_{B_2})^+ b \\ &= M_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ X'\Sigma_{\hat{\theta}}^{-1}Y \\ &\quad -[I - M_{B_1'M_{B_2}}(M_{B_1'M_{B_2}}X'\Sigma_{\hat{\theta}}^{-1}XM_{B_1'M_{B_2}})^+ X'\Sigma_{\hat{\theta}}^{-1}X]B_1'(M_{B_2}B_1B_1'M_{B_2})^+ b. \end{aligned} \tag{4.8}$$

Estimator of β_2 is

$$\hat{\hat{\beta}}_2 = \beta_2^{(0)} + K_2\hat{\gamma} = \beta_2^{(0)} - \left(B_2'B_2\right)^{-1}B_2'B_1K_1\hat{\gamma} = \beta_2^{(0)} - \left(B_2'B_2\right)^{-1}B_2'B_1\left(\hat{\beta}_1 - \beta_1^{(0)}\right)$$

$$= \beta_2^{(0)} - \left(B_2'B_2\right)^{-1}B_2'B_1\hat{\hat{\beta}}_1 + \left(B_2'B_2\right)^{-1}B_2'B_1\beta_1^{(0)} = -(B_2'B_2)^{-1}B_2'\left(B_1\hat{\beta}_1 + b\right). \quad (4.9)$$

5 Numerical study – uniformly accelerated movement

Let's investigate following situation. We are interested in a time necessary for a given car to accelerate from *0* to *100 km/h*. Let's suppose the acceleration is constant during the whole time to simplify the mathematical model of such a situation. This means we can use formulas for uniformly accelerated movement. We could measure in four points of a takeoff runway, the distance between the starting line and that point (marked as $s_1,...,s_4$) and the time necessary to reach the point ($t_1,...,t_4$). This measurement was replicated ten times. Measured data are in *Table 1*.

Table 1. Input data

	Distance from the starting line [m]				Time [s]			
Measurement	s_1	s_2	s_3	s_4	t_1	t_2	t_3	t_4
1	30,0001	60,0010	89,9997	119,9997	4,15	5,91	7,14	8,27
2	30,0003	60,0004	89,9999	120,0004	4,18	5,87	7,10	8,33
3	30,0000	60,0006	89,9995	120,0007	4,18	5,90	7,21	8,34
4	29,9997	60,0004	89,9995	119,9995	4,18	5,83	7,16	8,30
5	30,0002	60,0006	90,0005	120,0004	4,16	5,93	7,27	8,30
6	30,0000	60,0003	89,9998	119,9995	4,20	5,86	7,13	8,33
7	30,0009	60,0003	90,0010	120,0006	4,15	5,87	7,14	8,38
8	30,0000	59,9995	90,0005	119,9998	4,16	5,87	7,23	8,32
9	30,0000	60,0009	90,0001	120,0005	4,28	5,84	7,17	8,38
10	29,9992	60,0002	90,0002	119,9999	4,27	5,94	7,15	8,35

5.1 Appropriate mixed linear model

The main task is to estimate the acceleration during the take-off and dispersions of distance and time measurement. The acceleration estimation will be based on a distance formula for the uniformly accelerated movement $s = \frac{1}{2}at^2$, thus

$$2s_i - at_i^2 = 0,\ i = 1,...,4 \quad (5.1)$$

The next step is to compose an appropriate regression model. We have direct measurements of s_i and t_i, $i = 1,...,4$, these will be elements of vector β_1. Acceleration *a* is a non-measurable parameter β_2, it will act in the constraints only. Let's suppose the distance was measured with dispersion σ_s^2, the time measurement dispersion was σ_t^2, it means the wanted model has two variance components σ_s^2 and σ_t^2. Our situation can be described as follows:

$$Y = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix} + \varepsilon\,, \tag{5.2}$$

$$\operatorname{var} Y = \sigma_s^2 \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \sigma_t^2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.3}$$

$$\begin{pmatrix} 2 & 0 & 0 & 0 & -2a^{(0)}t_1^{(0)} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & -2a^{(0)}t_2^{(0)} & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & -2a^{(0)}t_3^{(0)} & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & -2a^{(0)}t_4^{(0)} \end{pmatrix} \delta\beta_1 + \begin{pmatrix} -t_1^{(0)2} \\ -t_2^{(0)2} \\ -t_3^{(0)2} \\ -t_4^{(0)2} \end{pmatrix} \delta\beta_2 + \begin{pmatrix} 2s_1^{(0)} - a^{(0)}t_1^{(0)2} \\ 2s_2^{(0)} - a^{(0)}t_2^{(0)2} \\ 2s_3^{(0)} - a^{(0)}t_3^{(0)2} \\ 2s_4^{(0)} - a^{(0)}t_4^{(0)2} \end{pmatrix} = 0 \tag{5.4}$$

Constraints (5.4) arise from (5.1) by linearization, upper index (0) denotes a prior value.

5.2 Acceleration and variance components estimates

In case of car take-off following prior values were chosen: $s_1^{(0)} = 30m$, $s_2^{(0)} = 60m$, $s_3^{(0)} = 90m$, $s_4^{(0)} = 120m$, $a^{(0)} = 3{,}46m \cdot s^{-2}$ (this value corresponds approximately with acceleration from *0* to *100 km/h* in *8s* according to a car producer statement), $t_1^{(0)} = 4{,}16s$, $t_2^{(0)} = 5{,}89s$, $t_3^{(0)} = 7{,}21s$, $t_4^{(0)} = 8{,}33s$ ($t_i^{(0)} = \sqrt{\dfrac{2s_i^{(0)}}{a^{(0)}}}$, $i = 1,...,4$). Final estimates of a (β_2), σ_s^2 and σ_t^2 (variance components) are listed in *Table 2*.

Table 2. Estimates – attempt 1

Measurement	$\hat{\hat{a}}$	$\hat{\sigma}_s^2$	$\hat{\sigma}_t^2$
1	3,420	$2{,}26 \cdot 10^{-7}$	$2{,}81 \cdot 10^{-3}$
2	3,426	$7{,}69 \cdot 10^{-7}$	$3{,}36 \cdot 10^{-3}$
3	3,470	$3{,}69 \cdot 10^{-8}$	$1{,}03 \cdot 10^{-4}$
4	3,422	$2{,}75 \cdot 10^{-7}$	$6{,}98 \cdot 10^{-4}$
5	3,475	$2{,}73 \cdot 10^{-7}$	$1{,}47 \cdot 10^{-3}$
6	3,433	$6{,}39 \cdot 10^{-7}$	$2{,}11 \cdot 10^{-3}$
7	3,446	$4{,}28 \cdot 10^{-7}$	$2{,}69 \cdot 10^{-3}$
8	3,458	$5{,}02 \cdot 10^{-8}$	$2{,}20 \cdot 10^{-4}$
9	3,473	$2{,}76 \cdot 10^{-6}$	$6{,}68 \cdot 10^{-3}$
10	3,478	$2{,}07 \cdot 10^{-6}$	$5{,}50 \cdot 10^{-3}$

It's obvious that the acceleration estimates are relatively stable but the variance components estimates differ even in decimal positions of valid numbers. We can expect less variability when using a bigger number of measurements. On this account another attempt was done. There were twelve points on the take-off runway and the distance and time were measured in each of them. This measurement was replicated five times and a mixed linear regression model with constraints analogous to (5.2), (5.3) was put together. The estimates of the variance components from this sekond attempt are in table 3.

Table 3. Variance components estimates – attempt 2

Measurement	$\hat{\sigma}_s^2$	$\hat{\sigma}_t^2$
1	$2{,}52 \cdot 10^{-7}$	$1{,}79 \cdot 10^{-3}$
2	$5{,}56 \cdot 10^{-7}$	$4{,}80 \cdot 10^{-4}$
3	$6{,}83 \cdot 10^{-8}$	$6{,}31 \cdot 10^{-4}$
4	$1{,}54 \cdot 10^{-7}$	$1{,}06 \cdot 10^{-4}$
5	$5{,}65 \cdot 10^{-7}$	$5{,}92 \cdot 10^{-4}$

We can see that the situation with the variance components estimates is similar to the one in the first attempt.

5.3 Conclusions

Input data were simulated from the normal distribution, distances with a standard deviation $\sigma_s = 0{,}5 \cdot 10^{-3}$ and times with a standard deviation $\sigma_t = 0{,}05$. Although the variance components vary significantly when we compare them to the squares of standard deviations σ_s and σ_t used for input data simulation we find that the estimates $\hat{\sigma}_s^2$ and $\hat{\sigma}_t^2$ don't deflect from the expected

interval. However the accuracy of their determination is not too high. (This accuracy is characterized by the variance-covariance matrix of the variance components estimates $2S^{-1}_{(M_{XK_1}\Sigma_{\theta 0}M_{XK_1})^+}$.)

References

[1.] ANDĚL, J.: *Statistical Methods*. (in Czech), MATFYZPRESS, Praha, 2003. ISBN 80-86732-08-8.

[2.] KUBÁČEK, L., KUBÁČKOVÁ, L.: *Statistika a metrologie*. Univerzita Palackého v Olomouci - vydavatelství, Olomouc, 2000.

[3.] RAO, C.R., KLEFFE, J.: *Estimation of variance components and applications*. North-Holland, Amsterdam – New York – Oxford – Tokyo, 1988.

Current address

Mgr. Hana Boháčová
Institute of Mathematics
Faculty of Economics and Administration
University of Pardubice, Studentská 95, 532 10 Pardubice
Phone: +420 466 036 170
e- mail: Hana.Bohacova@upce.cz

Journal of Applied Mathematics

volume II (2009), number III

ON THE INSENSITIVITY REGION FOR FIXED EFFECTS PARAMETERS AND ITS RELATIVE POSITION TO THE CONFIDENCE REGION FOR VARIANCE COMPONENTS

BOHÁČOVÁ Hana, (CZ)

Abstract. The maximum likelihood estimators of the fixed effects parameters $\boldsymbol{\beta}$ need a suitable choice of a starting value $\boldsymbol{\theta}_0$ of variance components in the mixed linear regression models. The question is how to choose these starting values and what happens if we change them with some small drift $\delta\boldsymbol{\theta}$. The task is to find a set of the admissible input variance components values $\boldsymbol{\theta}_0 + \delta\boldsymbol{\theta}$ (for given $\boldsymbol{\theta}_0$), it means a set of input values which cause ε-multiple increase (for some small $\varepsilon > 0$ which is specified in advance) of the standard deviation of the estimator of $\mathbf{h}'\boldsymbol{\beta}$ at most ($\mathbf{h}$ is a given vector of the same dimension as $\boldsymbol{\beta}$. Such a set is called an insensitivity region for fixed effects parameters. This paper should show an explicit formulation of this set, its properties and possible utilization.

Key words and phrases. Mixed linear regression model, maximum likelihood estimator, fixed effect parameters, variance components, insensitivity region.

Mathematics Subject Classification. Primary 62J10, 62J12; Secondary 62P30.

1 Denotation

$r(\mathbf{A})$	rank of matrix $\mathbf{A}$
$\mathbf{A}^+$	Moore-Penrose generalized inverse of matrix $\mathbf{A}$
$\mathcal{M}(\mathbf{A})$	column space of matrix $\mathbf{A}$
$\mathbf{M}_\mathbf{A}$	projection matrix on an orthogonal complement of the column space of matrix $\mathbf{A}$ (according to the Euclidean norm)
$\mathbf{Y}_n \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$	random vector $\mathbf{Y}$ has the n-dimensional normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and the variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$
$\mathrm{Var}_{\boldsymbol{\theta}_0}(\mathbf{Y})$	variance-covariance matrix of $\mathbf{Y}$ when we suppose the real value of $\boldsymbol{\theta}$ to be $\boldsymbol{\theta}_0$

2 Maximum likelihood estimators in the mixed linear regression model

Let us consider following linear regression model:

$$\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right), \tag{1}$$

$\mathbf{X}$ is a known matrix of $n \times k$ dimension which is of a full column rank $r(\mathbf{X}) = k$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a vector of unknown fixed effects parameters. Let the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ be of a form

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sum_{i=1}^{r} \theta_i \mathbf{V}_i. \tag{2}$$

$\theta_1, \theta_2, \ldots, \theta_r > 0$ are unknown variance components and $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_r$ are known symmetrical positive semideninite matrices, whereas the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ has to be positive definite. The aim is to estimate the fixed effects parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ and the unknown parameters of the covariance matrix - the variance components $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$.
Maximum likelihood method will be used to determine the estimators. Let us denote $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The likelihood equations for this model are (cf. [9]):

$$\mathbf{X}'{\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}}^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'{\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}}^{-1}\mathbf{Y} \tag{3}$$

$$\mathrm{Tr}({\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}}^{-1}\mathbf{V}_i) = \mathbf{Y}'\left(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}\mathbf{M}_{\mathbf{X}}\right)^{+}\mathbf{V}_i\left(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}\mathbf{M}_{\mathbf{X}}\right)^{+}\mathbf{Y}, i = 1, \ldots, r. \tag{4}$$

As we suppose $\mathbf{X}$ to have a full column rank and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ to be regular (as it is positive definite), matrix $\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}^{-1}\mathbf{X}$ is also a regular matrix. In case the variance components included in $\boldsymbol{\theta}$ are known we can get the $\boldsymbol{\beta}$ - estimator directly from (3):

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}}^{-1}\mathbf{Y}. \tag{5}$$

This estimator is identical with the $\boldsymbol{\theta}$-locally best linear unbiased estimator of the fixed effects parameters in model (1) (cf. [1]). As the variance components are supposed to be unknown we have to follow a different procedure. We need to choose some starting value of the variance components $\boldsymbol{\theta}_0$. Then we can start an iterative process based on (4):

$$\mathrm{Tr}({\boldsymbol{\Sigma}_{\boldsymbol{\theta}_{k+1}}}^{-1}\mathbf{V}_i) = \mathbf{Y}'\left(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_k}\mathbf{M}_{\mathbf{X}}\right)^{+}\mathbf{V}_i\left(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_k}\mathbf{M}_{\mathbf{X}}\right)^{+}\mathbf{Y}, i = 1, \ldots, r. \tag{6}$$

for $k = 0, 1, \ldots, p$. The iterations stop when $\boldsymbol{\theta}_p = \boldsymbol{\theta}_{p+1}$ for the first time and we consider $\boldsymbol{\theta}_p$ to be an estimate of $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_p.$$

Finally we use $\boldsymbol{\theta}_p$ as an input value for (5) and count the estimate $\widehat{\boldsymbol{\beta}}$.

3 Insensitivity region for a variance of an estimator of a linear function of the fixed effects parameters

As described at the end of the first section we should estimate the variance components at first (by means of the iterative procedure) and then use the final estimate of $\boldsymbol{\theta}$ as an input

value for the fixed effects parameters estimation. In some cases (especially when the variance components are not the object of our interest) we do not need to come trough the iterations. It should be enough when we base the estimate of $\boldsymbol{\beta}$ on some suitably chosen initial value $\boldsymbol{\theta}_0$. We just need to verify whether this simplification does influence the quality of the estimate. This question can be answered by the help of the conception of insensitivity stated below.

When looking at the estimator (5) of the fixed effects parameters it is obvious that the estimator is a function of variance components $\boldsymbol{\theta}$. We will denote it with $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ instead of $\widehat{\boldsymbol{\beta}}$ in the following text. Let's say we will use an initial value $\boldsymbol{\theta}_0$ of the variance components instead of $\boldsymbol{\theta}$, we will get estimate $\widehat{\boldsymbol{\beta}}\,(\boldsymbol{\theta}_0)$. We need to find a criterion to decide whether the choice of the prior value $\boldsymbol{\theta}_0$ enables to determine trustfulness estimates. To investigate this let us see what happens when we change $\boldsymbol{\theta}_0$ with some small $\delta\boldsymbol{\theta} = (\delta\theta_1, \ldots, \delta\theta_k)$. It seems to be reasonable not to allow the dispersions of the componets of the estimate of vector $\boldsymbol{\beta}$ to increase too much when using different prior value. Following two definitions are based on this restriction.

Definition 3.1 *(Cf. [6]) Let* $\mathbf{h} \in \mathbf{R}^k$*. Within a mixed linear model (1) the expression*

$$\left.\frac{\partial \widehat{\mathbf{h}'\boldsymbol{\beta}}\,(\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

is called the sensitiveness of a linear unbiasedly estimable function $\mathbf{h}'\boldsymbol{\beta}$ *with respect to* θ_j*,* $j = 1, 2, \ldots, r$*. Here* $\boldsymbol{\theta}^*$ *is an actual value of the parameter* $\boldsymbol{\theta}$*.*

Definition 3.2 *(Cf. [6]) Within a mixed linear model (1) the set* $\mathcal{N}_{\mathbf{h}'\boldsymbol{\beta},\boldsymbol{\theta}_0}$ *that fulfills:*

$$\boldsymbol{\theta}_0 + \delta\boldsymbol{\theta} \in \mathcal{N}_{\mathbf{h},\boldsymbol{\beta},\boldsymbol{\theta}_0} \Longrightarrow$$

$$\Longrightarrow \sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0 + \delta\boldsymbol{\theta})\right]} \leq (1+\varepsilon)\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]}, \tag{7}$$

is called the insensitivity region for the variance of the estimate of a linear function $\mathbf{h}'\boldsymbol{\beta}$*. Here* $\varepsilon > 0$ *is a given number.*

Remark 3.3 *When we choose* $\mathbf{h}$ *(a vector that designs a linear combination of the components of vector* $\widehat{\boldsymbol{\beta}}$*) to be an i-th unit vector* $\mathbf{e}_i$ *(*$\mathbf{e}_1$ *denotes a vector having all the components zero except the i-th one which is 1),* $\mathcal{N}_{\mathbf{h}'\boldsymbol{\beta},\boldsymbol{\theta}_0}$ *is (for given prior value* $\boldsymbol{\theta}_0$*) a set of all admissible prior values which do cause only* ε*-multiple increase of the standard deviation of the estimator of* β_i *at most when compared to the standard deviation of the same estimator based on the prior value* $\boldsymbol{\theta}_0$*. When we find* $\mathcal{N}_{\mathbf{e}_i'\boldsymbol{\beta},\boldsymbol{\theta}_0}$ *for all* $\mathbf{e}_i$*,* $i = 1, \ldots, k$ *and make their intersection*

$$\bigcap_{i=1}^{k} \mathcal{N}_{\mathbf{e}_i'\boldsymbol{\beta},\boldsymbol{\theta}_0},$$

we get a set of those variance components whose utilization doesn't cause too large increase of a standard deviation of an estimator any component of $\boldsymbol{\beta}$*.*

According to (5) we have

$$\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] = \left(\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\right)^{-1}. \tag{8}$$

Consequently

$$\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\mathbf{h}\right] = \mathbf{h}'\left(\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\right)^{-1}\mathbf{h}. \tag{9}$$

Further we need $\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta})\right]$. Trough the use of the differential we can approximate

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta}) \approx \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0) + \left.\frac{\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\delta\boldsymbol{\theta} \tag{10}$$

As the derivative is

$$\left.\frac{\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial\theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -(\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X})^{-1}\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{V}_i{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right],$$

we can approximately write

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta}) \approx \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0) - \sum_{i=i}^{r}(\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X})^{-1}\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{V}_i{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\delta\theta_i. \tag{11}$$

Let us dentote $\mathbf{C}(\boldsymbol{\theta}_0) = \mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}$ in what follows. As $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$ and $\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]$ are uncorrelated we can approximate $\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta})\right]$ using (11):

$$\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta})\right] \approx \mathrm{Var}_{\boldsymbol{\theta}_0}\left\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] + \mathrm{Var}_{\boldsymbol{\theta}_0}\left[\sum_{i=i}^{r}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{V}_i{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\delta\theta_i\right\}$$

$$= \mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] + \mathrm{Var}_{\boldsymbol{\theta}_0}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\delta\theta}}{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\right\} = \mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]$$

$$+\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\delta\theta}}{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}-\mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}'\right]{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\delta\theta}}{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)$$

$$= \mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] + \mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\delta\theta}}(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}\mathbf{M}_{\mathbf{X}})^{+}\boldsymbol{\Sigma}_{\boldsymbol{\delta\theta}}{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0). \tag{12}$$

Let us consider an arbitrary vector $\mathbf{h}\in\mathbf{R}^k$ and denote $\mathbf{L_h}' = \mathbf{h}'\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}$. For the appropriate linear combination of the components of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta})$ we have

$$\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0+\delta\boldsymbol{\theta}) \approx \mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0) - \sum_{i=1}^{r}\mathbf{L_h}'\mathbf{V}_i{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\delta\theta_i = \mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$$

$$-\left(\mathbf{L_h}'\mathbf{V}_1{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right],...,\mathbf{L_h}'\mathbf{V}_r{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\right)\delta\boldsymbol{\theta} = \mathbf{L_h}'\mathbf{Y}$$

$$-\left(\mathbf{L_h}'\mathbf{V}_1{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right],...,\mathbf{L_h}'\mathbf{V}_r{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\left[\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\right)\delta\boldsymbol{\theta}.$$

Denote further $\boldsymbol{\xi} = \begin{pmatrix} \mathbf{L_h}'\mathbf{V}_1\boldsymbol{\Sigma_{\theta_0}}^{-1}\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] \\ \dots \\ \mathbf{L_h}'\mathbf{V}_r\boldsymbol{\Sigma_{\theta_0}}^{-1}\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] \end{pmatrix}$, then $\boldsymbol{\xi} \sim_r (\mathbf{0}, \mathrm{Var}(\boldsymbol{\xi}))$. Let $\mathrm{Var}_{\boldsymbol{\theta}_0}(\boldsymbol{\xi}) = \mathbf{W_h}$. Then we have

$$\{\mathbf{W_h}\}_{i,j} = \mathrm{cov}\left(\mathbf{L_h}'\mathbf{V}_i\boldsymbol{\Sigma_{\theta_0}}^{-1}\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right], \mathbf{L_h}'\mathbf{V}_j\boldsymbol{\Sigma_{\theta_0}}^{-1}\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]\right) = \mathbf{L_h}'\mathbf{V}_i\boldsymbol{\Sigma_{\theta_0}}^{-1}$$

$$\cdot[\boldsymbol{\Sigma_{\theta_0}} - \mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{X}']\boldsymbol{\Sigma_{\theta_0}}^{-1}\mathbf{V}_j\mathbf{L_h} = \mathbf{L_h}'\mathbf{V}_i(\mathbf{M_X}\boldsymbol{\Sigma_{\theta_0}}\mathbf{M_X})^{+}\mathbf{V}_j\mathbf{L_h}. \tag{13}$$

Lemma 3.4 *(See [7])*
Let us consider a given point $\boldsymbol{\theta}_0 \in \mathbb{R}^r$ and the matrix

$$\mathbf{W}_h = \begin{pmatrix} \mathbf{L_h}'\mathbf{V}_1 \\ \vdots \\ \mathbf{L_h}'\mathbf{V}_r \end{pmatrix}[\mathbf{M_X}\boldsymbol{\Sigma_{\theta_0}}\mathbf{M_X}]^{+}(\mathbf{V}_1\mathbf{L_h}, \dots, \mathbf{V}_r\mathbf{L_h}),$$

which is assigned to the function $h(\boldsymbol{\beta}) = \mathbf{h}'\boldsymbol{\beta}$ with $\mathbf{h} \in \mathbb{R}^k$. Then

$$\mathcal{M}(\mathbf{W_h}) \perp \boldsymbol{\theta}_0.$$

Proof. $[\mathbf{M_X}\boldsymbol{\Sigma_{\theta_0}}\mathbf{M_X}]^{+}(\mathbf{V}_1\mathbf{L_h}, \dots, \mathbf{V}_r\mathbf{L_h})\boldsymbol{\theta}_0$
$= [\boldsymbol{\Sigma_{\theta_0}}^{-1} - \boldsymbol{\Sigma_{\theta_0}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma_{\theta_0}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma_{\theta_0}}^{-1}]\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma_{\theta_0}}^{-1}\mathbf{X})^{-1}\mathbf{h} = \mathbf{0}$.

Corollary 3.5 *According to lemma 2.4.* $\mathbf{W_h}$ *is a singular matrix.*

When we compare (12) and (13) we can write

$$\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0 + \delta\boldsymbol{\theta})\right] = \mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] + \delta\boldsymbol{\theta}'\mathbf{W_h}\delta\boldsymbol{\theta}. \tag{14}$$

The resulting set $\mathcal{N}_{\mathbf{h}'\boldsymbol{\beta},\boldsymbol{\theta}_0}$ is an insensitivity region for the variance of the estimate of a linear function $\mathbf{h}'\boldsymbol{\beta}$ from the definition 2.2. As we know from the mentioned definition

$$\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0 + \delta\boldsymbol{\theta})\right]} \leq (1+\varepsilon)\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]}$$

should be satisfied. According to (13)

$$\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right] + \delta\boldsymbol{\theta}'\mathbf{W_h}\delta\boldsymbol{\theta}} \leq (1+\varepsilon)\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]},$$

$$\sqrt{1 + \frac{\delta\boldsymbol{\theta}'\mathbf{W_h}\delta\boldsymbol{\theta}}{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]}} \leq 1 + \varepsilon,$$

$$\frac{\delta\boldsymbol{\theta}'\mathbf{W}_{\mathbf{h}}\delta\boldsymbol{\theta}}{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]} \leq (1+\varepsilon)^2 - 1,$$

$$\frac{\delta\boldsymbol{\theta}'\mathbf{W}_{\mathbf{h}}\delta\boldsymbol{\theta}}{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)\right]} \leq 2\varepsilon + \varepsilon^2.$$

$$\delta\boldsymbol{\theta}'\mathbf{W}_{\mathbf{h}}\delta\boldsymbol{\theta} \leq \left(2\varepsilon+\varepsilon^2\right)\mathrm{Var}[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)]. \tag{15}$$

Theorem 3.6 *The insensitivity region for the variance of the estimate of the linear function* $\mathbf{h}'\boldsymbol{\beta}$ *for given* $\mathbf{h}\in\mathbb{R}^k$*, given* $\boldsymbol{\theta}_0\in\mathbb{R}^r$ *and given* $\varepsilon>0$ *is the following set*

$$\mathcal{N}_{\mathbf{h}'\boldsymbol{\beta},\boldsymbol{\theta}_0} = \{\boldsymbol{\theta}_0+\delta\boldsymbol{\theta} : \delta\boldsymbol{\theta}'\mathbf{W}_{\mathbf{h}}\delta\boldsymbol{\theta} \leq \left(2\varepsilon+\varepsilon^2\right)\mathrm{Var}_{\boldsymbol{\theta}}[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)]\}. \tag{16}$$

Proof. This assertion is an immediate consequence of the defintion 2.2 and (15).

Lemma 3.7 $\mathcal{N}_{\mathbf{h}'\boldsymbol{\beta},\boldsymbol{\theta}_0}$ *is influenced by the direction of the vector* $\mathbf{h}$ *but not with is norm* $||\mathbf{h}||$.

Proof. According to (13) and (15) we have

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\delta\theta_i\delta\theta_j\mathbf{h}'\mathbf{C}(\boldsymbol{\theta}_0)^{-1}\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{V}_i(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}\mathbf{M}_{\mathbf{X}})^{+}\mathbf{V}_j{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{h}$$

$$\leq \left(2\varepsilon+\varepsilon^2\right)\mathrm{Var}_{\boldsymbol{\theta}_0}[\mathbf{h}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)] = 2\varepsilon\mathbf{h}'\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{h}. \tag{17}$$

Let us choose a fixed vector $\mathbf{h}_0\in\mathbf{R}^k$. Then for an arbitrary vector $\mathbf{h}\in\mathbb{R}^k$ with the same direction as $\mathbf{h}_0$ such $k\in\mathbb{R}$ exists that $\mathbf{h}=k\mathbf{h}_0$. When we establish this into (17) we get

$$k^2\sum_{i=1}^{r}\sum_{j=1}^{r}\delta\theta_i\delta\theta_j\mathbf{h}_0'\mathbf{C}(\boldsymbol{\theta}_0)^{-1}\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{V}_i(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}\mathbf{M}_{\mathbf{X}})^{+}\mathbf{V}_j{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{h}_0$$

$$\leq \left(2\varepsilon+\varepsilon^2\right)k^2\mathbf{h}_0'\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{h}_0.$$

As we can divide both the sides of the inequality above by k^2 we have finally

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\delta\theta_i\delta\theta_j\mathbf{h}_0'\mathbf{C}(\boldsymbol{\theta}_0)^{-1}\mathbf{X}'{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{V}_i(\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}\mathbf{M}_{\mathbf{X}})^{+}\mathbf{V}_j{\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}}^{-1}\mathbf{X}\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{h}_0$$

$$\leq \left(2\varepsilon+\varepsilon^2\right)\mathbf{h}_0'\mathbf{C}^{-1}(\boldsymbol{\theta}_0)\mathbf{h}_0.$$

4 Utilization of the insensitivity region for a variance of an estimator of a linear function of the fixed effects parameters

As we saw in the previous section the insensitivity regions are sets of such admissible input values of the variance components that do not cause an undesirable increase of the variance of the estimates. To be able to review the quality of the estimates $\widehat{\boldsymbol{\beta}}$ we will compare them to the confidence region for the variance components. For a special case of $r = 2$ (when we have two variance components) we will determine the confidence region to be a rectangle which covers the real value of the variance components with the given probability $1 - \alpha$ and centre of which is the estimate $\widehat{\boldsymbol{\beta}}$. According to the Chebyschev inequality (cf. [10])

$$P\left\{|\widehat{\theta}_1(\boldsymbol{\theta}_0) - \theta_1| \leq k\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\theta}_1(\boldsymbol{\theta}_0)\right]}\right\} \geq 1 - \frac{1}{k^2}$$

and in a similar way

$$P\left\{|\widehat{\theta}_2(\boldsymbol{\theta}_0) - \theta_2| \leq k\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\theta}_2(\boldsymbol{\theta}_0)\right]}\right\} \geq 1 - \frac{1}{k^2}.$$

Both these inequalities are included in the Bonferroni inequality (cf [5]):

$$P\left\{|\widehat{\theta}_1(\boldsymbol{\theta}_0) - \theta_1| \leq k\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\theta}_1(\boldsymbol{\theta}_0)\right]} \wedge |\widehat{\theta}_2(\boldsymbol{\theta}_0) - \theta_2| \leq k\sqrt{\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\theta}_2(\boldsymbol{\theta}_0)\right]}\right\}$$

$$\geq 1 - \frac{2}{k^2}. \tag{18}$$

The Bonferroni inequality is the fundamental of the confidence region we are looking for. We need $\left(1 - \frac{2}{k^2}\right) = (1 - \alpha)$, which means $k = \sqrt{\frac{2}{\alpha}}$. The resulting confidence region (according to (18)) for the variance components is a set which will be denoted by $\mathcal{E}_{\boldsymbol{\theta},\boldsymbol{\theta}_0}$, it's explicit for is following:

$$\mathcal{E}_{\boldsymbol{\theta},\boldsymbol{\theta}_0} = \left\{\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} : |\widehat{\theta}_1(\boldsymbol{\theta}_0) - \theta_1| \leq \sqrt{\frac{2}{\alpha}\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\theta}_1(\boldsymbol{\theta}_0)\right]}\right.$$

$$\left. \wedge |\widehat{\theta}_2(\boldsymbol{\theta}_0) - \theta_2| \leq \sqrt{\frac{2}{\alpha}\mathrm{Var}_{\boldsymbol{\theta}_0}\left[\widehat{\theta}_2(\boldsymbol{\theta}_0)\right]}\right\} \tag{19}$$

Here $\widehat{\theta}_1(\boldsymbol{\theta}_0)$ and $\widehat{\theta}_2(\boldsymbol{\theta}_0)$ denote the estimates of particular variance componets θ_1 and θ_2 based on the prior value of $\boldsymbol{\theta}_0$ (they are the result of the first iteration (6)). It means we have $\widehat{\boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \begin{pmatrix} \widehat{\theta}_1(\boldsymbol{\theta}_0) \\ \widehat{\theta}_2(\boldsymbol{\theta}_0) \end{pmatrix}$.

As mentioned above (19) gives a confidence region when $\boldsymbol{\theta}$ is two-dimensional. According to lemma 2.4 and its corollary 2.5 when $r = 2$ then $\mathbf{W_h}$ is of rank 1. The boundary of set (16) is then a singular conic - two parallel lines and the insensitivity region is a zone in between these

two lines. Looking at lemma 2.4 once more we can say that the direction of $\boldsymbol{\theta}_0$ is orthogonal to the lines specifying the insensitivity region.

In this special case we can compare the relative position of the insensitivity region and the confidence region as they are both sets of some values of the variance components. When the confidence region is a subset of the insensitivity region then the uncertainty arising when we determine the estimate $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0)$ (this uncertainty is expressed by the variance of the estimate) is not too large so that we can trust the estimate and stop the iterations after the first step already.

When the confidence region is not a subset of the insensitivity region then the uncertainty is larger, the variance of the estimate could be large. We have to be careful when we decide to use such an estimate.

However we still have some possibilities how to improve the estimate. We can continue with the iterations and stop them after the results stabilize (if they do so) and then use an iterated estimate of the variance components instead of the starting value $\boldsymbol{\theta}_0$ and check the relative position of appropriate insensitivity region and confidence region again. If it still doesn't help we could try to change the design of the experiment (if the situations allows to do so).

The theoretic background of the insensitivity region issue is in more detail described in [8] and [7]. A numerical study focused on the relative position of the insensitivity region and the confidence region for the variance components can be found in [4]. A further analysis of the insensitivity region properties is in [3]. [2] is focused on the insensitivity region in case of the mixed linear model with type I constraints and possible computational problems. Paper [4] contains the derivation of the insensitivity region for an estimate of a linear function of the variance components, further description and possible properties of the insensitivity region for an estimate of a linear function of the variance components are the aim of future research.

Reference

[1] ANDĚL, J.: *Statistical methods.* (in Czech). MATFYZPRESS, Praha, 2003

[2] BOHÁČOVÁ, H., HECKENBERGEROVÁ, J.: *Insensitivity regions for fixed effects parameters in mixed linear model with type I constraints and calculation problems linked with them.*(in Czech). In Forum Statisticum Slovacum, Vol. 3, No. 6, pp. 31-35, 2007.

[3] BOHÁČOVÁ, H., HECKENBERGEROVÁ, J.: *Properties of the insensitivity regions for an estimate of a linear function of the fixed effects parameters in the mixed linear model.* In Forum Statisticum Slovacum, Vol. 7, No. 7, 2008.

[4] BOHÁČOVÁ, H.: *Insensitivity regions for variance components in general linear model.* In Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Mathematica, Vol. 49, pp. 7-22, 2008

[5] HUMAK, K. M. S.: *Statistischen Methoden der Modellbildung Band I, Statistische Inferenz für lineare Parameter.* (in German). Akademie - Verlag, Berlin, 1977

[6] KUBÁČEK, L., KUBÁČKOVÁ, L.: *Unified Approach to determining Nonsensitiveness Regions.* In Tatra Mountains Mathematical Publications, Vol. 17, pp. 121-128, 1999

[7] KUBÁČEK, L., KUBÁČKOVÁ, L.: *Statistics and metrology.* (in Czech). Palacký University, Olomouc, 2000

[8] KUBÁČEK, L., KUBÁČKOVÁ, L., VOLAUFOVÁ, J.: *Statistical models with lineare structures*. Veda, Bratislava, 1995

[9] RAO, C. R., KLEFFE, J.: *Estimation of Variance Components and Application*. North-Holland, Amsterdam - New York - Oxford - Tokyo, 1988

[10] RÉNYI, A.: *Probability theory*. Akadémiai Kiadó, Budapest, 1970

Current address

Mgr. Hana Boháčová
Univerzita Pardubice, Fakulta ekonomicko-správní, Ústav matematiky, Studentská 84, 532 10 Pardubice
tel. number: +420 466 036 170
e-mail: hana.bohacova@upce.cz

Journal of Applied Mathematics

volume II (2009), number III

GNSS TRAIN POSITION INTEGRITY MONITORING BY THE HELP OF DISCRETE PIM ALGORITHMS

HECKENBERGEROVA Jana, (CZ)

Abstract. Currently there are several methods of train position determination, safety verification of train position and view of driven trace on given time interval. Research in the area of Global Navigation Satellite Systems (GNSS) shows, that it is advantageous to use satellite navigation for these purposes.
Described algorithms monitor the position integrity of the GNSS receiver. The aim of these algorithms is to verify that the GNSS position corresponds to the given train track and to find out that obtained GNSS position complies with safety requirements – Safety Integrity Level (SIL) – for given probability of undetected failure. The train track can be defined by a parametric equation, analytical function or discrete timely equidistant set of position coordinates. This paper is mainly devoted to introduction into the discrete algorithms.

Key words. GNSS, Integrity Monitoring, Multivariate Statistical Analysis, Wishart matrix, Hotteling's test statistics, Threshold Domain, Protection Level (PL)

Mathematics Subject Classification: Primary 62J05, 62F10, 62F03

1 Introduction

1.1 Global Navigation Satellite Systems

The name Global Navigation Satellite System (GNSS) covers all existing satellite systems (e.g. GPS-USA, GLONASS-Russia, GALILEO-EU). GNSS receivers of satellite signal provide information about position and its precision to their users. Some GNSS receivers are able to collect signals from different satellite systems. More information about operation GNSS can be found in (Mervart, 1993), where algorithms for (x,y)-position of GNSS receiver determination are also described. GNSS have been successfully used in the automotive field (GPS navigation), geodesy, and modern telecommunication technology. In air transport, elaborate methods for safety position determination by the help of GNSS have been developed and used.

1.2 GNSS in railway transport

Czech Railways perform tests and experiments with this new technology. GNSS receivers are installed on several selected locomotives and obtained GNSS data is compared with the map of train tracks. Trains can move only on tracks, so they can move only on the route, which has been strictly defined. This is the greatest advantage of GNSS usage for railway transport. For safety determination of locomotive position, it is enough to recognize in which track the reference point of the locomotive is situated. Then the determination of train position on this track is quite easy. Reasons for usage of this modern technology are evident – to reduce running costs and to increase effectiveness, safety and capacity of train operation.

In Laboratory of Intelligent Systems (LIS) the new Train Position Locator (TPL) is developed and tested. TPL consists of the GNSS receiver and additional inertial on-board sensors (INS) such as an odometer, accelerometer, gyroscope and microwave Doppler speedometer. Acquired data is fused by a Kalman filter and projected to the track map. It is expected that GNSS/INS TPL will play an important role in future railway safety related applications. First TPL has been installed on selected locomotive type 130-023-3 and it is used for online tests. For laboratory tests of TPL Test Bed and simulator have been developed. More information about TPL could be found in (Filip et al., 2004).

Nowadays GNSS database system for monitoring and prediction of locomotive position is also developed. Locomotive equipped with GNSS receiver send own positions every 5sec to central computer. All GNSS positions (with other information like train identification number, conductor number and track number) are saved to GNSS database. By the help of database it could by possible to describe (analytically or discretely) all used tracks, view driven trace of some train, safety verify current position of given locomotive. Algorithms which worked above GNSS database are named as GNSS Simplification (GNSS-SIM) algorithms.

1.3 GNSS-PIM algorithms

The GNSS Position Integrity Monitoring (GNSS-PIM) algorithms monitor the position integrity of the GNSS receiver. Let GNSS position of locomotive reference point be given by (x,y)-coordinates. The aim of these algorithms is to verify that GNSS position corresponds to the given train track and to find out that obtained GNSS position complies with safety requirements – Safety Integrity Level (SIL) – for given probability of missed detection (P_{MD}). For every GNSS position the Protection Level (PL) is computed. In statistical point of view the PL is the threshold domain, which is limited by probability of the second kind error. GNSS position integrity monitoring is provided by comparing of obtained PL with Horizontal Alert Limit (HAL), which is for given SIL already defined by user. If PL value is same or lower than HAL, then GNSS position has enough integrity and it could be regarded as safe. If PL is greater than HAL, the GNSS position doesn't satisfy safety requirements and mustn't be used for following safety related applications.

For correct operation of GNSS-PIM algorithms it is necessary to ensure sufficient integrity of the GNSS signal. The GNSS signal integrity can be obtained by Ground Integrity Channel (GIC) report and by usage of Receiver Autonomous Integrity Monitoring (RAIM). Another necessary demand for correct operation of monitoring algorithms is the usage of safety track map. For this reason the track map must satisfy safety requirements and it is necessary to ensure sufficient integrity of this map. If the map contains position errors and hasn't demanded precision, it mustn't be used for safety related applications.

The GNSS-PIM algorithms could be divided in analytical, parametrical and discrete according to the train track definition. Another division of GNSS-PIM algorithms could be done by a character of the track. Tracks specification for railway traffic is linear, arch and cubic. Algorithms development begins from linear character of the track, arch and cubic algorithms are explored after. They are solved by linearization of gain statistical models.

Analytical algorithms are used when the train track is defined by analytical function $f(x, y) = 0$. Incepted problems could be described by statistical model with constraints. Analytical GNSS-PIM algorithms for linear and arch track are solved in M.S. theses (Dvorakova, 2004).

Parametrical algorithms are used when the train track is defined by parametrical equations $\begin{cases} x = \varphi(t) \\ y = \psi(t) \end{cases}$ with parameter *t*. Statistical description of these algorithms appears in (Heckenbergerova, 2007) and (Heckenbergerova, 2008).

If the train track is defined by discrete timely equidistant set of position coordinates $Y = \left\{ \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} \right\}_{i=1}^{n}$, then GNSS-PIM algorithms are called discrete. **Discrete algorithms** have great advantage against others, because knowledge of track character (line, arch, cubic) isn't demanded. Discrete problems could be described by multivariate statistical model. This is the main disadvantage. Features of multivariate models like confidence area and insensitivity region hasn't been described. Incepted problems in discrete GNSS-PIM algorithms are main aim of current research. Some basic discrete algorithms are described in conference proceedings (Dvorakova et al., 2005).

2 Discrete GNSS-PIM algorithms

2.1 Theoretical basement for discrete problem

Definition 1:
Let $X_1 \sim N_p(\mu_1, \Sigma), ..., X_n \sim N_p(\mu_n, \Sigma)$ be independent stochastic vectors, μ_i be p-dimensional columned vector for $\forall i = 1, ..., n$ and Σ be matrix of type $(p \times p)$.

Let $\mathbf{X}_{n\times p} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}$ and $\mathbf{M}_{n\times p} = \begin{pmatrix} \mu_1^T \\ \vdots \\ \mu_n^T \end{pmatrix}$.

Conjugate distribution of the elements of matrix $\mathbf{Y} = \mathbf{X}^T\mathbf{X}$ is called **p-dimensional Wishart distribution with *n* degrees of freedom** and with parameters Σ, $\mathbf{M}$:

$\mathbf{Y} \sim W_p(n, \Sigma)$ – central Wishart distribution ($\mathbf{M} = \mathbf{0}$),

$\mathbf{Y} \sim W_p(n, \Sigma, \mathbf{M})$ – noncentral Wishart distribution ($\mathbf{M} \neq \mathbf{0}$).

Definition 2:
Let $\mathbf{y} \sim N_p(\mu, \frac{1}{c}\Sigma)$, $\mathbf{Y} \sim W_p(k, \Sigma)$ be independent and Σ be positive definite matrix.

Test statistics $T^2 = c \cdot k \cdot \mathbf{y}^T \mathbf{Y}^{-1} \mathbf{y}$
is called **Hotelling t-square test statistics**.

Theorem 1:

If $k > p\text{-}1$ then $F = \frac{k-p+1}{p} \cdot \frac{T^2}{k} \sim F_{p,k-p+1}(\delta),$

where $\delta = c \cdot \mu^T \Sigma^{-1} \mu$ is noncentrality parameter of the *F*-distribution.

2.2 Formulation of discrete problem

The GNSS Position Integrity Monitoring (GNSS-PIM) algorithms monitor the position integrity of the GNSS receiver. Let **Z** be vector of GNSS position of locomotive reference point be given by (*x*,*y*)-coordinates. The aim of Discrete GNSS-PIM algorithms is to verify that GNSS position corresponds to the given train track defined by polygon **Y**. And then find out that obtained GNSS position complies with safety requirements – Safety Integrity Level (SIL) – for given probability of missed detection (P_{MD}). For every GNSS position the Protection Level (PL) is computed. GNSS position integrity monitoring is provided by comparing of obtained PL with Horizontal Alert Limit (HAL). If PL value is same or lower than HAL, then GNSS position has enough integrity and it could be regarded as safe. If PL is greater than HAL, the GNSS position doesn't satisfy safety requirements and mustn't be used for following safety related applications.

Let $\mathbf{Y} = \left\{ \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} \right\}_{i=1}^{n}$ be vector of the (x, y)-coordinates vector track axis,

$\mathbf{Z} = \left\{ \begin{pmatrix} Z_{j,1} \\ Z_{j,2} \end{pmatrix} \right\}_{j=1}^{m}$ be vector of the GNSS train positions,

$\widetilde{\mathbf{Z}} = \left\{ \begin{pmatrix} \widetilde{Z}_{j,1} \\ \widetilde{Z}_{j,2} \end{pmatrix} \right\}_{j=1}^{m}$ be vector of the orthogonal projections **Z** to **Y** based polygon and

$\Delta = \left\{ \begin{pmatrix} \Delta_{j,1} \\ \Delta_{j,2} \end{pmatrix} \right\}_{j=1}^{m}$ be vector of the projection residual vectors $\Delta_j = \begin{pmatrix} \Delta_{j,1} \\ \Delta_{j,2} \end{pmatrix} = \begin{pmatrix} Z_{j,1} - \widetilde{Z}_{j,1} \\ Z_{j,2} - \widetilde{Z}_{j,2} \end{pmatrix}.$

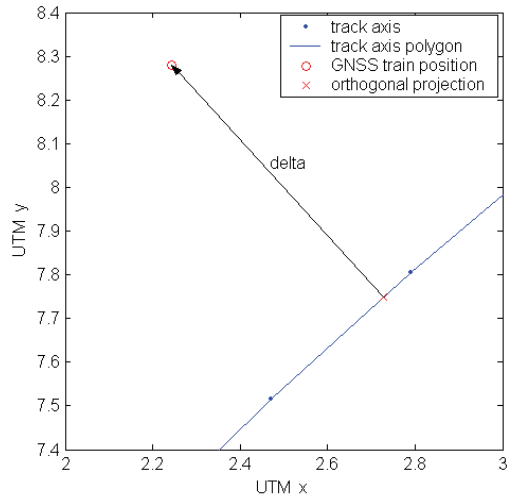


Fig 1. Graphical representation of described vectors

First aim of the lower described algorithm is testing hypotheses, whether true positions of given locomotive lie on the supposed track given by discrete timely equidistant set of position coordinates Y.

So null hypotheses: $$H_0 : E(\Delta_j) = 0$$

is tested against alternative hypotheses: $$H_a : E(\Delta_j) \neq 0$$

for every $j=1,\ldots m$ so for every measured GNSS train position.

Test statistics properties are introduced in following theorem. This theorem can be proved by Theorem 1 and Definitions 1, 2 using.

Theorem 2: Test of fit

Let Δ_j for $j=1,\ldots ,m$ be a sequence of the projection residuals vectors and $W_{2,m} = \sum_{j=1}^{m} \Delta_j \Delta_j^T$ be Wishart matrix type (2x2). Let Δ_{m+1} be the projection residual vector correspondent with $(m+1)$ GNSS position, then test statistics

$$T = \Delta_{m+1}^T W_{2,m}^{-1} \Delta_{m+1} \sim \frac{\chi_2^2}{\chi_{m-2+1}^2} = \frac{2}{m-1} F_{2,m-1} .$$

It is obvious that minimal three GNSS positions are essential for algorithm first using. Let first three projection residual vectors $\Delta_1, \Delta_2, \Delta_3$ be evaluate and test significance level α (probability of the first order error, false alarm probability) be select. Now, perform test of fit for third GNSS train position $\begin{pmatrix} Z_{3,1} \\ Z_{3,2} \end{pmatrix}$, where the test statistics $T = \Delta_3^T W_{2,2}^{-1} \Delta_3 \sim \frac{1}{2} F_{2,1}$.

If $T \leq 2 \cdot F_{2,1}(1-\alpha)$, hypothesis H_0 can not be refused on the given test significance level α, if $T > 2 \cdot F_{2,1}(1-\alpha)$, hypothesis H_0 is refused in behalf of alternative hypothesis H_a on the given test significance level α.

In case of hypothesis H_0 refusal, Δ_3 is significantly differ from other projection residual vectors, so it's necessary to generated error report and delete given GNSS position from safety related applications.. Then algorithm restart must be carry out.

GNSS position seems to be right, when hypothesis H_0 isn't refused. Then PL for this measurement can be determinate and algorithm can pass on the next iteration. If obtained PL value is same or less than HAL value, which is previously defined by user, then the GNSS position has enough integrity and it can be regard as safety. If PL is greater than HAL, the GNSS position doesn't satisfy safety requirements.

Protection Level is the threshold domain, which is limited by the probability of the second kind error (probability of undetected failure, P_{MD}). The threshold domain determinate how big difference between projection residual vector and mean-value of these vectors can be detected with given probability of undetected failure, P_{MD} value is defined by safety requirements (SIL).

Let Ω be the set of all plane vectors and let $\delta_{\max}$ be solution of equation

then
$$P\left\{F_{2,m-1}(\delta) \geq F_{2,m-1}(0,1-\alpha)\right\} = 1 - P_{MD},$$

$$\mathrm{PL} = \mathrm{T}_{\kappa_\alpha} = \left\{u : u \in \Omega,\ \tfrac{m-1}{2} \cdot u^T W_{2,m}^{-1} u \leq \delta_{\max}\right\} \quad .$$

2.3 Discrete GNSS-PIM algorithm

Now the algorithm m-iteration description can be introduced:

- Evaluation of Wishart matrix $W_{2,m}$

  $$W_{2,m} = \sum_{j=1}^{m} \Delta_j \Delta_j^T$$

- Determination of the projection residual vector Δ_{m+1}
- Evaluation of the test criteria value T

  $$T = \Delta_{m+1}^T W_{2,m}^{-1} \Delta_{m+1}$$

- Comparing of values $\frac{m-1}{2} \cdot T$ and $F_{2,m-1}(1-\alpha)$
  - $\frac{m-1}{2} \cdot T \leq F_{2,m-1}(1-\alpha)$ – PL determination and passing on the next iteration $m+1$
  - $\frac{m-1}{2} \cdot T > F_{2,m-1}(1-\alpha)$ – generation of error report and algorithm restart.

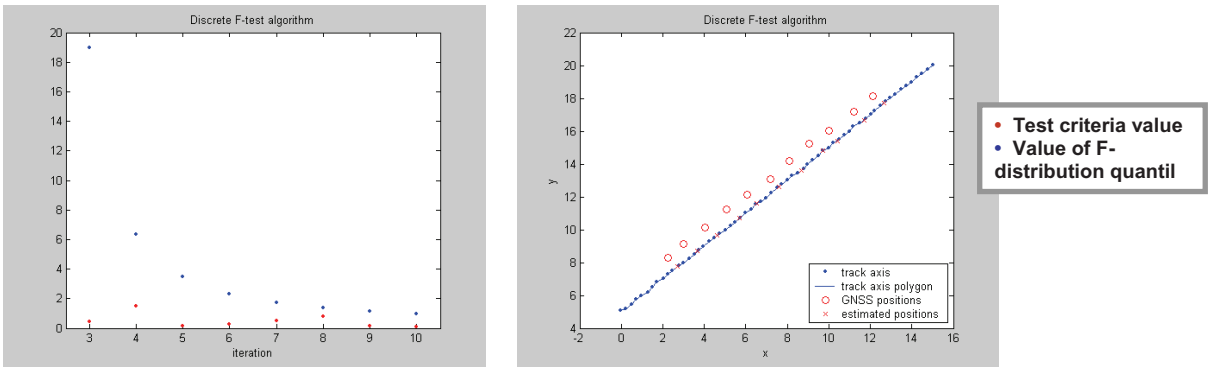


Fig 2. Numerical results of Discrete GNSS-PIM algorithm

Discrete GNSS-PIM algorithm can be negatively influenced by dependence of GNSS positions and their orthogonal projection, Wishart matrix become singular and can not be inverted. Because of singularity, it is necessary to decrease degree of freedom of F-distribution and generalized inverse matrix must be used for the test criteria. This problem increases time costingness of the whole algorithm.

3 Conclusion

During the realization of the mentioned research many difficult mathematical problems can be expected. For example up to now normal distribution of GNSS position errors is assumed and due to this assumption estimation procedures and testing of statistical hypotheses have been used in standard way. Usually variances of the actual errors in determination of train position are unknown and must be estimated. This fact influences heavily the procedure of confidence domain determination and power functions of statistical tests. Utilization of other measure techniques e.g. odometers, gyroscopes and Doppler radar, make the set of results heterogeneous, what need development of special numerical algorithms. Only satisfactory solution of mentioned problems has chance to utilize in the practice successfully and this way to increase effectiveness, safety and capacity of train operation.

Aim of future research is integration PIM algorithms into existing software for train position determination, laboratory and online tests of whole positioning system. Confidence area, threshold domains and insensitivity regions in discrete PIM algorithms are essential part of future statistical problem. PIM algorithms could be used for car position verification too. Evolution of automobile safety position monitoring is also direction of future development.

4 References

[1] KUBÁČEK, L., KUBÁČKOVÁ, L.: *Statistika a metrologie*. Univerzita Palackého v Olomouci 2000, ISBN 80-244-0093-6.

[2] RAO, R.C.: *Lineární metody statistické indukce a jejich aplikace.* Academia Praha (1978)

[3] MERVART, L.: *Základy GPS*. Vydavatelství ČVUT, Praha, 1993.

[4] DVOŘÁKOVÁ, J.: *Statistické metody pro identifikaci polohy vlaku.* Olomouc 2004.

[5] FILIP, A., MOCEK, H., BAZANT, L., MAIXNER, V.: *The High Integrity GNSS/INS Based Train Position Locator.* COMPRAIL'04, 2004.

[6] DVOŘÁKOVÁ, J., MOCEK, H., MAIXNER, V.: *Statistical Approach to the Train Position Integrity Monitoring*. Sborník „Reliability, safety and diagnostics of transport structures and means 2005", Pardubice, ISBN 80-7194-769-5.

[7] HECKENBERGEROVÁ, J.: *Parametrické algoritmy pro ověření GNSS polohy vlaku.* Sborník Infotrans 2007, Pardubice, ISBN 978-80-7194-989-3.

[8] HECKENBERGEROVA, J.: *GNSS train position verification by the help of parametrical PIM algorithms.* 7th International Conference „Aplimat 2008", Bratislava 2008, ISBN: 978-80-89313-03-7.

Current address

Heckenbergerová Jana, Mgr.
University of Pardubice,
Faculty of Electrical Engineering and Informatics, Department of Information Technologies,
Studentská 95, 532 10 Pardubice, phone: +420 466 036 726
e-mail: jana.heckenbergerova@upce.cz

Journal of Applied Mathematics

volume II (2009), number III

EVALUATION OF GROWTH CURVES VIA LINEAR MIXED EFFECTS MODEL

JAROŠOVÁ Eva, (CZ)

Abstract. The linear mixed effects model was applied to data representing the growth of yeast colonies. The growth was observed under optimal and stress cultivation conditions for fourteen days. Two temperatures (10 and 20°C) and three NaCl concentrations (0, 1 and 2%) were used according to the full factorial design. The colony diameter obtained by means of the method of image analysis was a response variable. S-PLUS program was used to analyze repeated measurements. Two-phase kinetics was confirmed and the mean growth rates in the second linear phase under various stress conditions were estimated. Confidence and prediction intervals for the time after which a colony reaches the specific diameter were calculated.

Key words. experimental design, repeated measures, confidence and prediction intervals

Mathematics Subject Classification: 62P10, 62-07.

1 Introduction

The aim of the study was to evaluate the growth of yeast under optimal and stress cultivation conditions. Giant colonies of *Rhodotorula mucilaginosa* were cultivated at different temperatures and in the presence of different NaCl concentrations. The effect of temperature and NaCl concentration on the colony growth was examined. The temperature, concentration and time formed the full factorial design. As repeated observations of the specimens were made over time, the observations of the same specimen were correlated and, in addition, a higher correlation of adjacent observations than of those more distant in time could be expected.

To take into account the special covariance structure of the data, the linear mixed effects model was used. A detailed description of mixed effects models (both linear and nonlinear) can be found in [4], covariance structure modelling is described e.g. in [2]. The SAS Help includes a good overview, too. The usual application of the mixed effects model consisting in prediction of the response of individual specimens was not of any use here. The mixed effects model was, however, needed to make a correct statistical inference such as hypothesis tests or confidence intervals. By means of the mixed effects model behaviour of specimens could be explored in more detail and the main sources of response variance could be identified.

Beside the analysis of temperature and NaCl concentration effects, confidence and prediction intervals for the response in the given time were constructed and by means of them the time it takes the response to reach the specific value was estimated.

Although the paper deals with the data from food microbiology, similar problems can be found in various industrial branches. For example some degradation or deterioration processes are monitored and the time to a failure is predicted. The mixed effects model can be a suitable tool in such cases. The use of the model is not restricted to the functions linear in parameters; non-linear mixed effects models are applied in a similar way. Both linear and non-linear mixed effects models are implemented in most known commercial software products such as S-PLUS or SAS.

2 Mixed effects model

The linear mixed effects model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e} \tag{1}$$

was used, where $\mathbf{y}$ is the Nx1 vector of responses, $\mathbf{X}$ is the known Nxp design matrix linking $\boldsymbol{\beta}$ to $\mathbf{y}$, $\boldsymbol{\beta}$ is the px1 vector of unknown parameters (fixed effects), $\mathbf{Z}$ is the Nxq design matrix linking $\mathbf{b}$ to $\mathbf{y}$, $\mathbf{b}$ is the qx1 vector of unknown random effects and $\mathbf{e}$ is the Nx1 vector of random errors. Assuming $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, and $\mathbf{b}$ and $\mathbf{e}$ independent, the mean profile is given by $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and the covariance structure depends on the matrices $\mathbf{G}$ and $\mathbf{R}$, namely $\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$. $\mathbf{G}$ and $\mathbf{R}$ are determined by a set of h parameters ($\boldsymbol{\theta}$) that are estimated either by the maximum likelihood method or by the restricted maximum likelihood method. The corresponding likelihood functions are

$$l(\boldsymbol{\theta}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{r}^T\mathbf{V}^{-1}\mathbf{r} - \frac{N}{2}\log(2\pi), \tag{2}$$

$$l_R(\boldsymbol{\theta}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{r}^T\mathbf{V}^{-1}\mathbf{r} - \frac{N-p}{2}\log(2\pi), \tag{3}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of $\mathbf{X}$.

Solving the mixed model equations where the unknown $\mathbf{G}$ and $\mathbf{R}$ are replaced by their estimates

$$\begin{bmatrix} \mathbf{X}^T\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}^T\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}^T\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}^T\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}, \tag{4}$$

the empirical best linear unbiased estimator (EBLUE) and empirical best linear unbiased predictor (EBLUP) are obtained [1]

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y} \tag{5}$$

$$\tilde{\mathbf{b}} = \hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{6}$$

According to [3] the approximate covariance matrix of $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \tilde{\mathbf{b}} - \mathbf{b})$ is

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{21}^T \\ \hat{\mathbf{C}}_{21} & \hat{\mathbf{C}}_{22} \end{bmatrix}, \tag{7}$$

where

$$\hat{\mathbf{C}}_{11} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^-$$

$$\hat{\mathbf{C}}_{21} = -\hat{\mathbf{G}} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^-$$

$$\hat{\mathbf{C}}_{22} = (\mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} + \hat{\mathbf{G}}^{-1})^{-1} - \hat{\mathbf{C}}_{21} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{Z} \hat{\mathbf{G}} .$$

No account is made for the uncertainty in estimating **G** and **R** and so these covariance matrices tend to underestimate the true variability. The bias, however, should be small for fairly well balanced data [5].

The confidence interval for $\mathbf{k}^T \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix}$, where **k** is a ($p$+$q$)x1 vector, is given by

$$\mathbf{k}^T \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\mathbf{b}} \end{bmatrix} \pm t_{1-\alpha/2}(\nu) \sqrt{\mathbf{k}^T \hat{\mathbf{C}} \mathbf{k}} , \tag{8}$$

where $t_{1-\alpha/2,\nu}$ is $1-\alpha/2$ quantile of the t-distribution with ν degrees of freedom, $\nu = n - p - h$.

3 Model of colony diameter growth

Giant colonies were cultivated at 10 or 20°C for 14 days. Different NaCl concentrations (0, 1, 2%) were added to the media as a stress factor. Six specimens were observed under the same treatment conditions. The area of the colony was measured by the method of image analysis and the equivalent diameter was derived (the diameter of a circle having the same area as the colony). The camera and illuminating system used in the experiment did not enable to monitor the initial stages of colony growth due to a low contrast between the colony and the background. The first results were obtained after five days (t_0 = 5). From then on, most growth curves exhibited linear dependence of equivalent diameter on time (Figure 1).

The equivalent colony diameter was a response variable. With time taken as continuous and on the assumption that our measurements record the period of linear growth, the diameter of the i-th colony at the j-th level of temperature, the k-th level of NaCl concentration and time t_l can be expressed in the form

$$y_{ijkl} = \beta_{0,jk} + b_{0,i(jk)} + (\beta_{1,jk} + b_{1,i(jk)})(t_l - t_0) + e_{ijkl} , \tag{9}$$

where $\beta_{0,jk}$ corresponds to the mean diameter at $t = t_0$, $\beta_{1,jk}$ denotes the mean growth rate (in cm per day) in the "linear" period, random effects $b_{0,i(jk)}$ represent variation of line intercepts around $\beta_{0,jk}$ and random effects $b_{1,i(jk)}$ correspond to variation of line slopes around $\beta_{1,jk}$. Random errors e_{ijkl} denote departures of observations from the model.

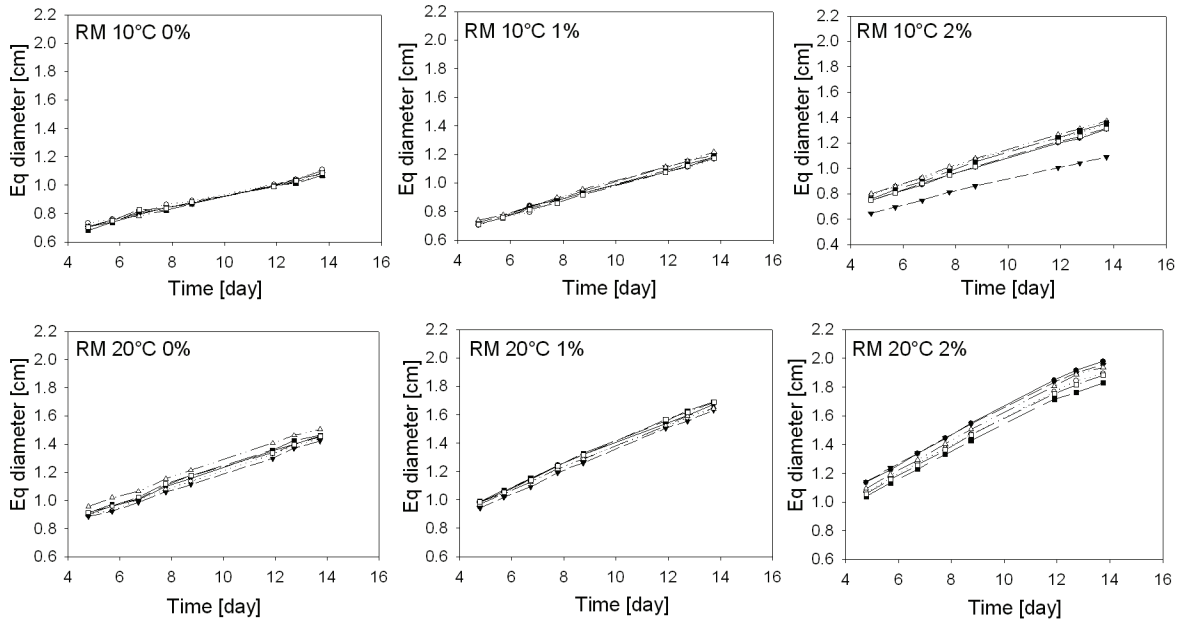


Figure 1 Growth curves of giant colonies.

Both parameters $\beta_{0,jk}$ and $\beta_{1,jk}$ are supposed to be affected by a treatment (note that $\beta_{0,jk}$ does not represent the intercept at t = 0); the indices j and k stand for various treatment conditions. Columns of **X** in Eq. (1) correspond to the fixed part of the model (in notation used in S Plus)

$$T_j + NaCl_k + T_j * NaCl_k + time + (T_j + NaCl_k + T_j * NaCl_k) * time, \tag{10}$$

where T is a two-level factor (j = 1, 2) and $NaCl$ is a three-level factor (k = 1, 2, 3). *Time* is a continuous variable.

Variances $\sigma^2_{0,jk}$ and $\sigma^2_{1,jk}$ of random effects $b_{0,i(jk)}$ and $b_{1,i(jk)}$, respectively, lying on the main diagonal of **G** may or may not differ for various treatments. In case of equal variances we shall denote $b_{0,i}$ ($b_{1,i}$) and σ^2_0 (σ^2_1).

For $t \geq t_0$ the mean profile at T_j and $NaCl_k$ has the form

$$E(Y(t) \mid T_j, NaCl_k) = \beta_{0,jk} + \beta_{1,jk}(t - t_0), \tag{11}$$

The confidence interval can be obtained following Eq. (8), i.e.

$$\mathbf{x}^T \hat{\boldsymbol{\beta}} \pm t_{1-\alpha/2}(\nu)\sqrt{\mathbf{x}^T \hat{\mathbf{C}}_{11} \mathbf{x}}, \tag{12}$$

where $\mathbf{x}^T$ denotes the row vector corresponding to Eq. (10) at time t. It can be expressed simply as

$$\hat{\beta}_{0,jk} + \hat{\beta}_{1,jk}(t - t_0) \pm t_{1-\alpha/2}(\nu)\sqrt{\hat{D}(\hat{\beta}_{0,jk}) + 2\hat{C}(\hat{\beta}_{0,jk}, \hat{\beta}_{1,jk}) + \hat{D}(\hat{\beta}_{1,jk})(t - t_0)^2}. \tag{13}$$

From Eq. (13) the confidence interval for the time t_* it takes the colony to reach the specific diameter Y_* can be obtained. The confidence limits are determined by solving the equation

$$\left(\frac{\hat{\beta}_1^2}{t_{1-\alpha/2,\nu}^2}-\hat{D}(\hat{\beta}_1)\right)(t_*-t_0)^2-2\left(\frac{(Y_*-\hat{\beta}_0)\hat{\beta}_1}{t_{1-\alpha/2,\nu}^2}+\hat{C}(\hat{\beta}_0,\hat{\beta}_1)\right)(t_*-t_0)+\frac{(Y_*-\hat{\beta}_0)^2}{t_{1-\alpha/2,\nu}^2}-\hat{D}(\hat{\beta}_0)=0 \quad (14)$$

where indices j and k were omitted for clearness, $\hat{D}$ and $\hat{C}$ denote estimated variance and covariance, respectively.

Further the prediction interval for the diameter at time t was considered. It is given by

$$\mathbf{x}^T\hat{\boldsymbol{\beta}}\pm t_{1-\alpha/2}(\nu)\sqrt{\mathbf{x}^T\hat{\mathbf{C}}_{11}\mathbf{x}+\mathbf{z}^T\hat{\mathbf{G}}\mathbf{z}+\hat{\sigma}^2}\,, \quad (15)$$

or in a more transparent form

$$\hat{\beta}_{0,jk}+\hat{\beta}_{1,jk}(t-t_0)\pm t_{1-\alpha/2}(\nu)\sqrt{V_{jk}(t)}\,, \quad (16)$$

where

$$V_{jk}(t)=\hat{D}(\hat{\beta}_{0,jk})+\hat{\sigma}_{0,jk}^2+2\hat{C}(\hat{\beta}_{0,jk},\hat{\beta}_{1,jk})(t-t_0)+[\hat{D}(\hat{\beta}_{1,jk})+\hat{\sigma}_{1,jk}^2](t-t_0)^2+\hat{\sigma}^2\,.$$

Similarly, the prediction limits for the time to reach the diameter Y_* can be determined by solving the equation

$$\left(\frac{\hat{\beta}_1^2}{t_{1-\alpha/2,\nu}^2}-\hat{D}(\hat{\beta}_1)-\hat{\sigma}_{1,jk}^2\right)(t_*-t_0)^2-2\left(\frac{(Y_*-\hat{\beta}_0)\hat{\beta}_1}{t_{1-\alpha/2,\nu}^2}+\hat{C}(\hat{\beta}_0,\hat{\beta}_1)\right)(t_*-t_0)+$$

$$+\frac{(Y_*-\hat{\beta}_0)^2}{t_{1-\alpha/2,\nu}^2}-\hat{D}(\hat{\beta}_0)-\hat{\sigma}_{0,jk}^2-\hat{\sigma}^2=0\,. \quad (17)$$

4 Results

Firstly, the structure of **G** and **R** was examined. Differences among $\sigma_{0,jk}$ (standard deviations of random effects $b_{0,i(jk)}$) under different treatment conditions were apparent. As the systematic effect of *NaCl* on $\sigma_{0,jk}$ was distinguishable, only three parameters corresponding to different concentrations and a single parameter for the random effects of slopes were included so that the overparametrization, manifesting itself by boundless confidence intervals for $\sigma_{0,jk}$, should be avoided. The variance matrix of random effects had the form $\mathbf{G}=diag\left\{\sigma_{0,1}^2,\sigma_{0,2}^2,\sigma_{0,3}^2,\sigma_1^2\right\}$. The matrix **R** corresponded to AR(1) scheme. The estimates of the standard deviations supplemented by the estimates of the matrix **R** parameters are in Table 1, where $\sigma_{0,jk}$ represents the standard deviation of $b_{0,i(jk)}$ dependent on *NaCl*, σ_1 represents the standard deviation of $b_{1,i}$, the same for all treatment conditions, σ denotes the error standard deviation, Φ is the autocorrelation coefficient. According to the model of the covariance structure the response variance was composed of three components and varied with time, i.e. $V(y)=\sigma_{0,k}^2+\sigma_1^2(t-t_0)^2+\sigma^2$. Increasing differences between

specimens that exceeded random error variation were evident especially at the highest NaCl concentration. It could be concluded that this pattern existed also before $t = 5$.

Table 1 Estimated parameters of the variance structure model

Matrix **G**					Matrix **R**	
NaCl [%]	0	1	2	$\hat{\sigma}_1$	$\hat{\sigma}$	0.0211
$\hat{\sigma}_{0,jk}$	0.0142	0.0054	0.0593	0.0015	Φ	0.7569

As for the fixed part of the model (Eq. (10)), all main effects and interactions were significant (all p-values < 0.0001). Considering the same initial mean diameter regardless of the treatment as a reasonable assumption, substantial differences among the estimates $\hat{\beta}_{0,jk}$ (Table 2) reflected different mean rates under various treatment conditions in the previous growth period. In addition to the effect of temperature a positive effect of higher NaCl concentration was observed both at lower and higher temperature. The effect was more distinguishable at the higher temperature. The effects of temperature and NaCl concentration on $\beta_{1,jk}$ were similar (Table 2).

Table 2 Estimated parameters of the mean structure model with 95% confidence limits

T [°C]	NaCl[%]	$\hat{\beta}_0$	95% lcl	95% ucl	$\hat{\beta}_1$	95% lcl	95% ucl
10	0	0.7088	0.6895	0.7282	0.0414	0.0388	0.0440
10	1	0.7163	0.7001	0.7325	0.0523	0.0498	0.0549
10	2	0.7540	0.7038	0.8041	0.0607	0.0581	0.0633
20	0	0.9101	0.8908	0.9295	0.0611	0.0585	0.0637
20	1	0.9787	0.9625	0.9949	0.0768	0.0742	0.0794
20	2	1.0942	1.0440	1.1444	0.0932	0.0906	0.0958

The estimated times to reach a double diameter ($\hat{t}_* - t_0$) under different treatment conditions are displayed in Table 3. It is obvious that the time becomes shorter with higher temperature and higher NaCl concentration.

Table 3 Confidence and prediction limits for time to reach double diameter

T [°C]	NaCl [%]	$\hat{t}_* - t_0$	95% lclm	95% uclm	95% lcl	95% ucl
10	0	17.12	16.22	18.14	15.28	19.25
10	1	13.68	13.13	14.28	12.39	15.12
10	2	12.42	11.53	13.34	11.17	13.76
20	0	14.90	14.35	15.48	13.80	16.11
20	1	12.75	12.40	13.13	11.05	14.51
20	2	11.74	11.17	12.33	10.26	13.26

5 Model validation

Suitable covariance models were chosen based on the information criteria AIC and BIC (the lower the values of AIC and BIC the better).

$$AIC = -2LL_R + 2(p+h)$$

$$BIC = -2LL_R + (p+h)\log N,$$

where LL_R is the restricted log-likelihood at convergence and h is the number of parameters in **G** and **R**. Various forms of the matrix **G** determined by the random effects considered and two forms of the matrix **R**, i.e. AR(1) scheme or independent errors were examined (**I** is the identity matrix). Both criteria indicated RM3 as the best (Table 4).

Table 4 Covariance structure models

Model	**G**	**R**	AIC	BIC
RM1	$b_{0,i}, b_{1,i}$	AR(1)	-1416.333	-1358.406
RM2	$b_{0,i}$	AR(1)	-1415.867	-1361.561
RM3	$b_{0,i(k)}, b_{1,i}$	AR(1)	-1431.083	-1365.916
RM4	$b_{0,i(k)}, b_{1,i}$	$\sigma^2\mathbf{I}$	-1362.538	-1300.991

Q-Q plot (Figure 2a) is nearly straight, indicating no serious evidence against the assumption of normality. Plot in Figure 2b shows a very good agreement between the observed data and the fitted model including random effects and may serve as a further confirmation of the adequacy of both the mean and the covariance structure models.

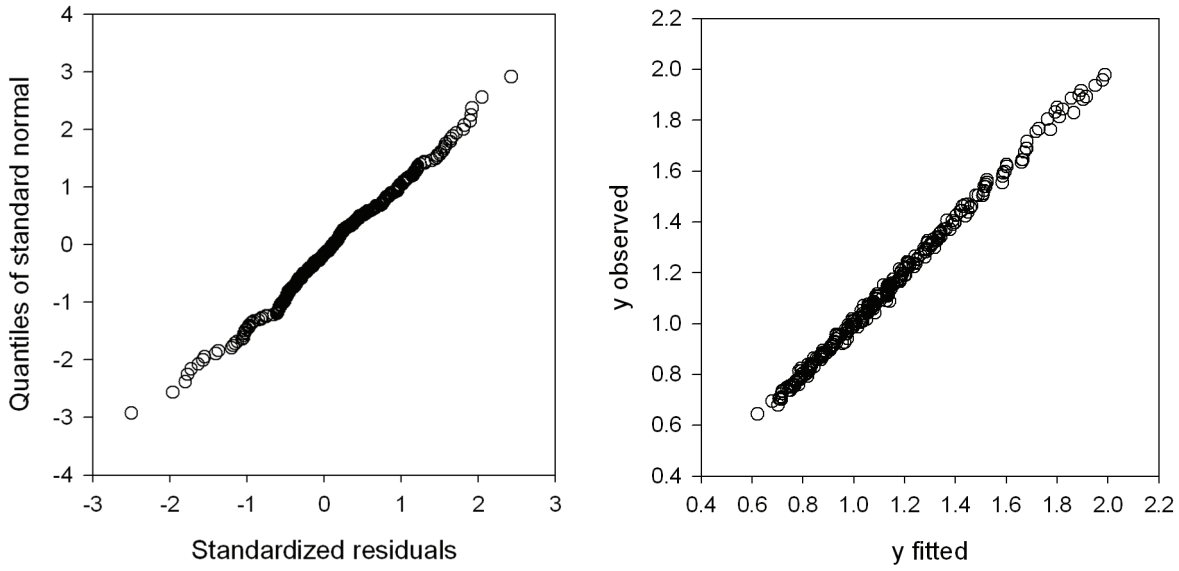


Figure 2 a) Q-Q plot, b) Comparison of fitted and observed values

6 Discussion

Although it was impossible to measure the area at the beginning of the process in our experiment the data analysis indicated a two-phase kinetic profile. Intercepts of growth lines obtained by extrapolation for $t = 0$ differed across the experimental treatments and this phenomenon contradicted the reasonable assumption of only random variation of specimen sizes at $t = 0$. It followed that the extrapolation was not valid and in reality a distinct form of growth or at least a different growth rate occurred in the previous period.

The mean growth rates in the observed linear phase under various stress conditions were estimated using the linear mixed effects model. The positive effects of temperature and NaCl concentration were obvious. The highest growth rate was observed in media with 2% of NaCl at 20°C. A positive effect of temperature on the growth rate in the initial phase was recognized through the level of intercepts at $t = 5$.

The growth rate of individual colonies varied and the most pronounced difference was observed at the highest NaCl concentration. Through the variation of intercepts at $t = 5$ the effect of NaCl concentration on the growth rate variation in the initial phase was recognized, which was in accordance with the variation in the observable linear phase.

Although the data-based approach was used, the covariance structure model is reasonably interpretable. Variation of the higher growth rates is greater. Growth rates in both phases depend on temperature and NaCl concentration. It is not surprising that variation in the first period projects itself to different values of $\sigma_{0,k}$ depending on the NaCl concentration.

In case of linear dependence the time t_* after which a colony reaches the specific value of the diameter is inversely proportional to the corresponding growth rate. The estimate of the mean time t_* obtained from Eq. (11) is biased. The confidence or prediction limits obtained from Eq. (14) and (17), however, are true limits satisfying given probability.

References:

[1.] HENDERSON, C.R.: Applications of Linear Models in Animal Breeding. University of Guelph,1984.
[2.] LITTEL, R.C., PENDERGAST, J., NATARAJAN, R.: Tutorial in biostatistics, Modelling covariance structure in the analysis of repeated measures data. Statistics in medicine 19, 1793-1819, 2000.
[3.] McLEAN, R.A., SANDERS, W.L., and STROUP, W.W.: "A Unified Approach to Mixed Linear Models." The American Statistician, 45, 54 – 64, 1991.
[4.] PINHEIRO, J.C., BATES, D.M.: Mixed-Effects Models in S and S-PLUS. Springer, New York, 2000.
[5.] SAS 9.1. Help and Documentation. SAS Institute Inc., Cary, NC, USA, 2002-2003

Current address

doc. Ing. Eva Jarošová, CSc.
Skoda Auto University, Tr. Vaclava Klementa 864
293 60 Mlada Boleslav, Czech Republic
e-mail: jarosova@vse.cz

Journal of Applied Mathematics

volume II (2009), number III

STATISTICAL ANALYSIS OF SUBCONSCIOUS HUMAN BEHAVIOUR

KORDEK David, (CZ)

Abstract. Statistical analysis is frequently used for studying people's behaviour in normal as well as critical situations. It is present in many works dealing with many different situations in which human behaviour is observed

The aim of this work is to demonstrate that there is a mathematical dependence between the subconscious behaviour of two or more persons. There are numerous possibilities to study subconscious behaviour (yawning, etc.). However, it is necessary to discover a possibility that can be observed and measured objectively.

Therefore, after accepting Professor Šeba's advice, we suggested the following experiment: to observe the behaviour of people consuming beverages at restaurants and to record the drinking times of persons sitting at the same table. We assumed that the subconscious behaviour of persons sitting at the same table would influence one another. We recorded all the persons' individual drinking times on a laptop computer using a special programme we had prepared beforehand.

The output of the measurements was a multidimensional sequence of temporal data. These sequences consisted of individual persons' drinking times. We chose a correlation coefficient as the indicator of the level of dependence between the behaviour of each pair of persons at a table. If we prove mutual dependence between two persons' drinking times, we can be further interested in the average delay in the drinking of the person who drinks under the influence of another person. We can discover this by means of the cross correlation function.

As the results stated in the article reveal, we have proved that people sitting at the same table synchronise their consumption of beverages on an utterly subconscious basis, i.e. there is an interdependence of their subconscious behaviour.

Key words. Statistical analysis, Gaussian curve, correlation coefficient

1 Introduction

Statistical analysis is a frequently used method of studying human behaviour. It is applied in papers dealing with a large number of situations in which human behaviour is studied, ranging from rather entertaining studies, such as the analysis of changes in human behaviour in the United Kingdom on Friday the thirteenth [1], to much more serious ones describing such things as the correlation between suicidal ideation and attempt among Chinese prostitutes [2].

The objective of this paper is to prove the existence of a correlation between the subconscious behaviour of two and more individuals. There are certainly many possible phenomena on which the correlation can be studied (e.g., yawning, winking), but it is crucial to find such a phenomenon that is observable and measurable objectively. We have proposed the following experiment: to study the behaviour of public house guests when having drinks. In all cases, we were interested in a group of guests sharing a table; we recorded the drinking times of each individual. We presumed that the subconscious behaviour patterns of the people in a group (sharing a table) would influence each other. The individuals within a group would then receive yes-no information (have a drink – do not have a drink), which we considered to be of identical probability. Based on that, yes-no behaviour (have a drink – do not have a drink) would be exhibited in response to the received information. We also presumed that the surrounding ambience would be identical for all the individuals sharing a table and without any influence on their drinking behaviour.

2 The Experiment

We performed three measurements in various public houses in Hradec Králové. To the measurement we used the following programme; see Fig. 1.

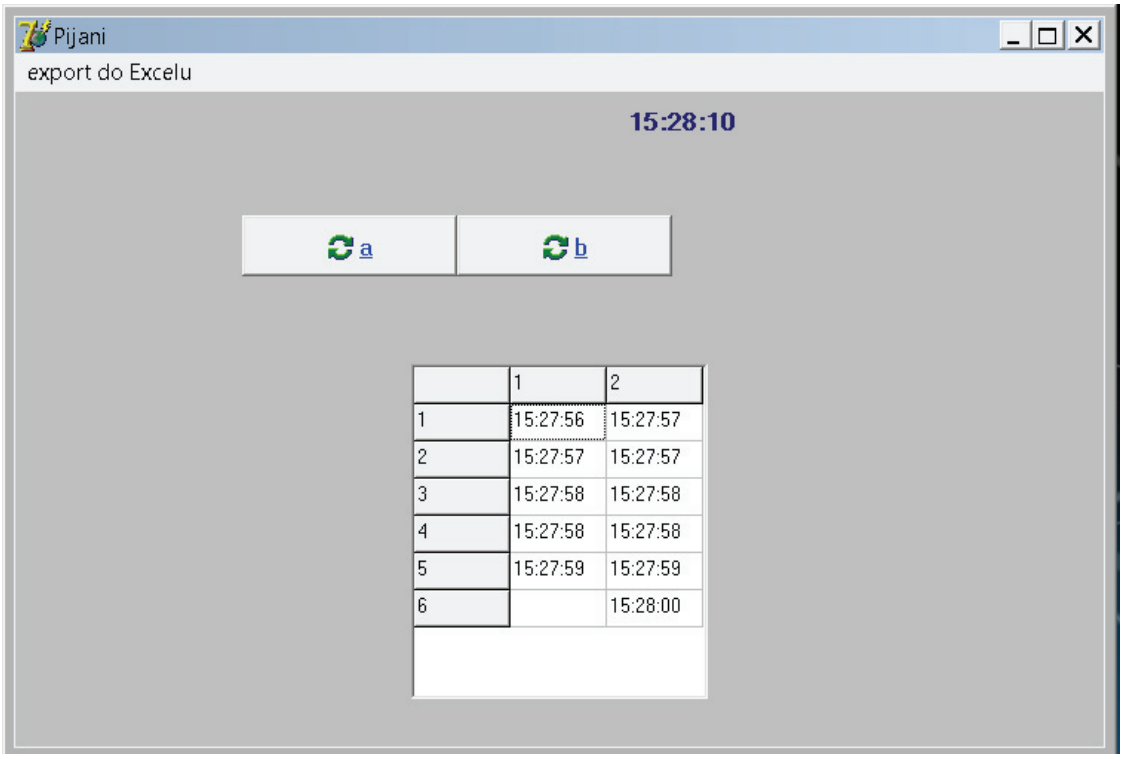


Fig. 1: Programme, which is used to measure. (Programme notes times of drinking)

Two pairs and a group of three were studied. The measurements resulted in finite multidimensional time-related data series: two two-dimensional ones and a three-dimensional one, specifically. The series elements represent the drinking times of the studied persons. If the time of the *i*-th drink by the *j*-th person is labelled $\tau_j(i)$, the time series is then shown as $\{\tau_j(i)\}_{j=1,\ldots,N}^{i=1,\ldots,T_j}$, where T_j is the total number of drinks by the *j*-th person and *N* is the number of persons.

3 Data Handling Method

We reason that if the time difference between two persons' drinks is greater than h_0, there is no correlation between the two persons' drinks. For each group member, i.e., $\forall j = 1,\ldots,N$, the following function is defined on the set of real numbers:

$$\widetilde{x}_j(t) = \sum_{i=1}^{T_j} \exp\left(-\frac{\left(t-\tau_j(i)\right)^2}{2\sigma^2}\right),$$

where $\sigma = \dfrac{h_0}{4\sqrt{2\ln 2}}$. Thus, we placed the maxima of Gaussian curves with a half-width of $h = h_0/2$ to each drinking moment.

Restriction of the functions to finite time interval and their discretization (with the step of 1 second) yields an *N*-dimensional time series $\left\{x_j(i)\right\}_{j=1,\ldots,N}^{i=0,\ldots,T}$, where the discrete time $i = 0$ corresponds, for example, to the time $t = \min\limits_{j\in\{1,\ldots,N\}} \tau_j(1) - 3h$ and $i = T$ corresponds, for example, to the time $t = \max\limits_{j\in\{1,\ldots,N\}} \tau_j(1) - 3h$, so far as *h* is an integer (the stored values were one-second integers).

The correlation coefficient of the functions $\widetilde{x}_j$ and $\widetilde{x}_k$, or the time series x_j and x_k was chosen as the measure of the dependence of the conscious behaviour patterns of the *j*-th and *k*-th individuals within the studied group:

$$\widetilde{C}_{jk} = \frac{\int_{-\infty}^{\infty} \widetilde{x}_j(t)\,\widetilde{x}_k(t)\,dt}{\sqrt{\int_{-\infty}^{\infty}\left(\widetilde{x}_j(t)\right)^2 dt \int_{-\infty}^{\infty}\left(\widetilde{x}_k(t)\right)^2 dt}}, \text{ or } C_{jk} = \frac{\sum_{i=0}^{T} x_j(i)\,x_k(i)}{\sqrt{\sum_{i=0}^{T}\left(x_j(i)\right)^2 \sum_{i=0}^{T}\left(x_k(i)\right)^2}}.$$

If both studied persons drank synchronously, i.e. $\widetilde{x}_j \equiv \widetilde{x}_k$, the correlation coefficient $\widetilde{C}_{jk}$ would equal one. If taking into account, for instance, only one drink by each person in a studied group, delayed by the semi-width of the Gaussian curve *h*, a direct calculation (using the substitution $t = x + h/2$) yields $\widetilde{C}_{jk} = 1/4$.

Let us now assume, to the contrary, that $\left|\tau_j(i) - \tau_k(l)\right| \geq h_0$. Under the assumption, the drinking of one individual is then not influenced by the drinking of another at all, and vice versa. Let us estimate the value of $\widetilde{C}_{jk}$ from above. Let us denote

$$I = \int_{-\infty}^{\infty}\left(\exp\left(-\frac{t^2}{2\sigma^2}\right)\right)^2 dt.$$

The denominator of the correlation coefficient is estimated from below, using the inequality

$$\int_{-\infty}^{\infty}\left(\sum_{i=1}^{T_j}\exp\left(-\frac{\left(t-\tau_j(i)\right)^2}{2\sigma^2}\right)\right)^2 dt > \sum_{i=1}^{T_j}\int_{-\infty}^{\infty}\left(\exp\left(-\frac{\left(t-\tau_j(i)\right)^2}{2\sigma^2}\right)\right)^2 dt = T_j I\,.$$

The numerator is estimated from above:

$$\int_{-\infty}^{\infty}\left(\sum_{i=1}^{T_j}\exp\left(-\frac{\left(t-\tau_j(i)\right)^2}{2\sigma^2}\right)\sum_{l=1}^{T_k}\exp\left(-\frac{\left(t-\tau_k(l)\right)^2}{2\sigma^2}\right)\right)dt =$$

$$= \sum_{i,l}\int_{-\infty}^{\infty}\exp\left(-\frac{\left(t-\tau_j(i)\right)^2+\left(t-\tau_k(l)\right)^2}{2\sigma^2}\right)dt = I\sum_{i,l}\exp\left(-\frac{\left(\tau_j(i)-\tau_k(l)\right)^2}{4\sigma^2}\right) <$$

$$< I\sum_{i,l}\exp\left(-\frac{h_0^2}{4\sigma^2}\right) = IT_jT_k\,2^{-8},$$

where the first equation uses the substitution $t = x + \left(\tau_j(i)+\tau_k(l)\right)/2$. In all, then, we have the estimate $\widetilde{C}_{jk} < \sqrt{T_jT_k}\,2^{-8}$. Therefore, if the correlation coefficient value is greater than this estimate, there is a correlation between two persons' drinking behaviour.

Let us note that the aforementioned estimate of the correlation coefficient from above is a very rough one. In reality, the time difference between only two closest drinks will be approximately h_0; the estimate uses this limit for all pairs of drinks.

If we show a correlation between the drinking behaviour of two persons in the above way, we may further be interested in the delay with which a person under the drinking influence of another will drink. This may be established by means of a standard cross-correlation function (see, for instance, [3] section 9.1):

$$\widetilde{K}_{jk}(t') = \frac{\int_{-\infty}^{\infty}\widetilde{x}_j(t)\,\widetilde{x}_k(t-t')\,dt}{\sqrt{\int_{=\infty}^{\infty}\left(\widetilde{x}_j(t)\right)^2 dt\int_{-\infty}^{\infty}\left(\widetilde{x}_k(t)\right)^2 dt}}, \text{ or } K_{jk}(l) = \frac{\sum_{i=0}^{T}x_j(i)\,x_k(i-l)}{\sqrt{\sum_{i=0}^{T}\left(x_j(i)\right)^2\sum_{i=0}^{T}\left(x_k(i)\right)^2}}.$$

It is obvious that $\widetilde{K}_{jk}(0) = \widetilde{C}_{jk}$, or that $K_{jk}(0) = C_{jk}$. The positions of the local maximums of the function $\widetilde{K}_{jk}$ correspond to the average delays with which a person under the drinking influence of another will drink. Specifically, the local maximum corresponding to a positive time *t'* shows the delay with which the *k*-th person drinks after the *j*-th; the local maximum corresponding to a negative time *t'* shows the delay with which the *j*-th person drinks after the *k*-th.

4 The Results

The limit time difference having a zero influence on the drinking behaviour was chosen to be h_0=20 sec. The values of the correlation coefficient C_{jk} are shown in the 6th column of Table 1. For

purposes of comparison, the fifth column shows the above derived limit of the drinking influence $\sqrt{T_j T_k}\, 2^{-8}$. The values of the correlation coefficient C_{jk} can also be seen in the graphs in Fig. 2, which plot out the time dependence of the cross-correlation function K_{jk} ($K_{jk}(0) = C_{jk}$, as stated above).

All the graphs in Fig. 2 exhibit a noticeable local maximum near the time t=0 sec. The position of that maximum corresponds to the most likely delay with which a person subconsciously drinking as a consequence of another's drinking will drink. In some cases, the maximum will decompose into two if Gaussian curves with smaller semi-widths (e.g., h=2 sec) are used in the definition $\tilde{x}_j$; see Fig. 3.

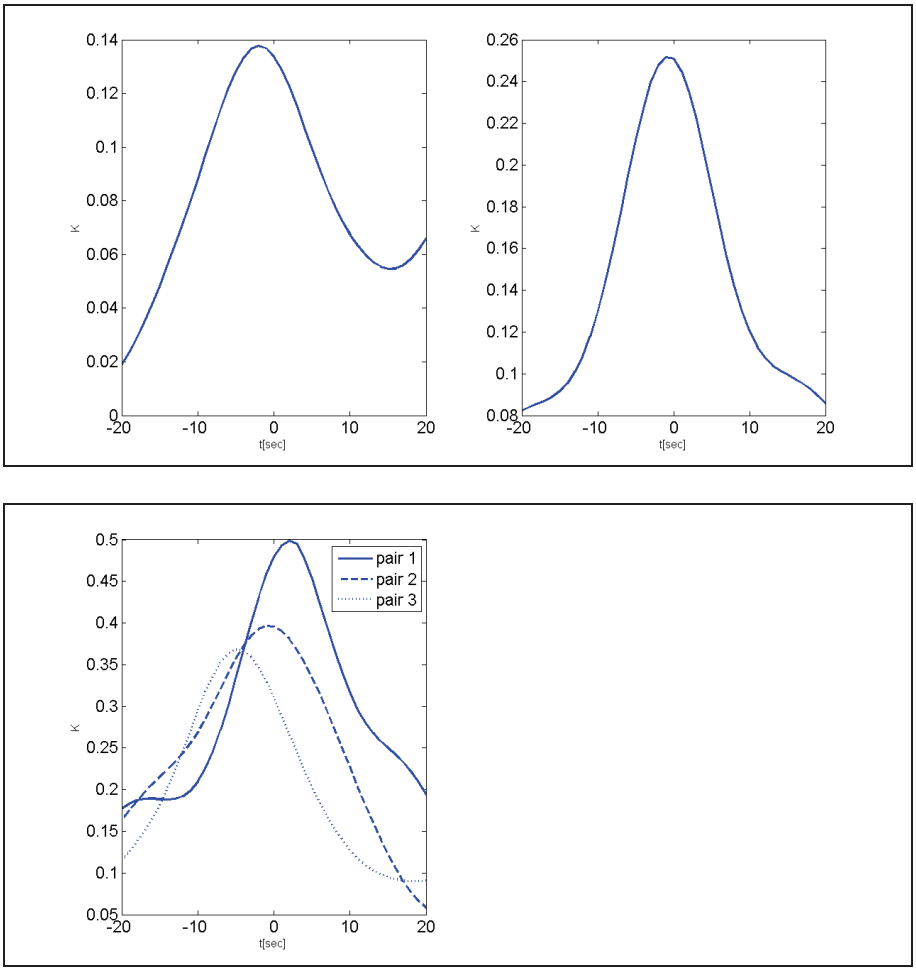


Fig. 2: Cross-correlation function for all the measured pairs.

If the maximum does not decompose, it can be said that one of the persons drank first in most cases (the j-th person in the group mostly drank before the k-th person if the time co-ordinate of the maximum is negative and, contrariwise, the k-th person in the group mostly drank before the j-th person if the time co-ordinate of the maximum is positive). If the maximum does decompose, it yields the most likely delay with which the j-th person drinks after the k-th as well as the most likely delay with which the k-th person drinks after the j-th. The positions of the maximums (following a decomposition in two cases) are shown in the 7th column in Table 1.

Measurement	Pair number (Fig. 2)	Amount of drinks T_j	Amount of drinks T_k	Limit value $\sqrt{T_j T_k}\,2^{-8}$	Correl. coeff. C_{jk}	Position of maximum $t_{\max}$(sec)
1	1	22	29	0,0987	0,1337	-2
2	2	24	16	0,0765	0,2509	-3 1
3	3	19	13	0,0614	0,4793	2
	4	19	19	0,0742	0,3955	-3 5
	5	13	19	0,0614	0,3097	-5

Table 1: Correlation coefficient values for all measured pairs compared to the limit values and positions of the cross-correlation maximums.

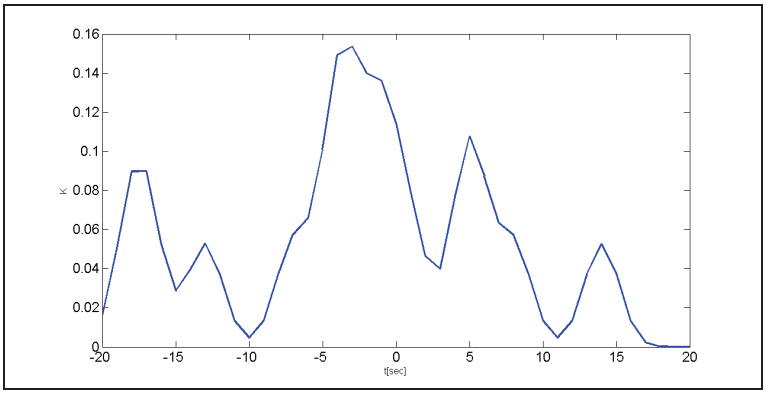


Fig. 3: Decomposition of the cross-correlation function maximum into two (Pair 4 in Fig. 2) at h=2 sec.

5 Discussion and Conclusions

As evident from the results shown in Table 1, we have identified correlation coefficient values that were higher than the theoretically calculated limit value (if the correlation coefficient assumed the limit or lower value, the drinking behaviour correlation would not be provable) and even several times higher in all cases except one (Pair 1). It can therefore be maintained that people drink synchronously in groups, that is, there is a correlation between their subconscious behaviour. The most likely delay between two persons' drinks was (in absolute values) between 1 sec and 5 sec. The chosen value h_0=20 sec – the time difference which does not allow any drinking behaviour influence – was therefore reasonable.

To improve the persuasiveness of our conclusions, we would need more experimental data. As is clear from the description of our experiment, however, the data collection is very time-consuming.

References

[1] SCANLON, T. J., LUBEN, R. N., SCANLON, F. L. and SINGLETON, N., Is Friday 13th bad for your health?, *BMJ* **307** (1993), 1584-1586.

[2] HONG, Y., Li, X., FANG, X. and ZHAO, R., Correlates of Suicidal Ideation and Attempt Among Female Sex Workers in China, *Health Care Women Int.* **28** (2007), 490–505.

[3] PRIESTLY, M. B., *Spectral Analysis and Time Series*, Elsevier Academic Press, London, 1981.

Current address

Mgr. David Kordek
Department of Physics and Informatics, University Hradec Králové,
500 03 Hradec Králové, Rokitansky Str. 62, CZ,
Tel: 493 331 518,
email: david.kordek@uhk.cz

Journal of Applied Mathematics

volume II (2009), number III

MODIFICATION OF OBJECTS SIMILARITY FOR SEARCHINGF T-CLUSTERS

ŽÁK Libor, (CZ)

Abstract: This article deals with searching T continuous clusters. To find them we can use classical clustering methods (based on similarity and dissimilarity of objects) and the transitivity of closure of fuzzy relations. Thus we can use classical clustering methods to search T continuous clusters.

Keywords: similarity, dissimilarity, clusters, T clusters.

1 Introduction

We often need to be able to recognize various shapes, especially contours, written and printed text etc. Generally we need identify $n-k$-dimensional objects in n-dimensional space ($k > 0$), it means we must cluster points describing such object. the cluster analysis can be used with success. The classical cluster analysis based on similarity (dissimilarity - distance) enables us to find objects of the same dimension as the dimension of the space is. To find objects of lesser dimension we can use a modification of clusters – so called T-clusters. There are many methods using classical similarity but more less methods to find T-clusters. This article deals with using standard clustering methods to search T-clusters.

2 Cluster

Let us have n objects and each object characterized by m parameters:

$$\boldsymbol{O} = \{O_1,..., O_n\},\ h\text{-th object } O_h = (x_{h1},..., x_{hm}),\ \text{where } x_{hj} \in \mathbf{R} \text{ for } h \in \{1,..., n\}.$$

An idea of „closeness“ of two objects is very important for clustering process. Such a „closeness“ can be described by Π - similarity function. The function assigns to two objects a nonnegative number $\Pi : \boldsymbol{O} \times \boldsymbol{O} \rightarrow \mathbf{R}^+_0$, for which holds:

$$\Pi(O_h, O_s) \geq 0,$$

$$\Pi(O_h, O_s) = \Pi(O_s, O_h).$$

There is often used a **dissimilarity measure** of objects instead of a similarity in clustering methods. The dissimilarity of objects is indicated **d**: $\boldsymbol{O}\times\boldsymbol{O} \rightarrow \mathbf{R}^+_0$ and it must satisfy:

$$\mathbf{d}(O_h, O_s) = 0 \Leftrightarrow O_h = O_s,$$
$$\mathbf{d}(O_h, O_s) \geq 0,$$
$$\mathbf{d}(O_h, O_s) = \mathbf{d}(O_s, O_h).$$

d is often equal to any metric on $\mathbf{R}^m$ in real situations.

We try to divide objects into clusters. We call **cluster** such subset A of set $\boldsymbol{O}$, which satisfies:

$$\min_{O_i, O_j \in A} \Pi(O_i, O_j) > \max_{O_k \in A, O_l \notin A} \Pi(O_k, O_l) \quad \text{or} \quad \max_{O_i, O_j \in A} \mathbf{d}(O_i, O_j) < \min_{O_k \in A,\, O_l \notin A} \mathbf{d}(O_k, O_l).$$

3 T-cluster

Object O_h is T-continuous with object O_s for given threshold T if there exists sequence of objects $O_h = O_1, O_2, O_3, ..., O_k = O_s$, $k > 1$, that $\mathbf{d}(O_i, O_{i+1}) \leq$ T for $i = 1, 2, ..., k-1$, where $\mathbf{d}(O_i, O_{i+1})$ is the coefficient of dissimilarity of objects O_i a O_{i+1}. T-continuous cluster (T-cluster) is a subset $A \subset \boldsymbol{O}$, fitting following conditions:
a) each couple of objects from A is T-continuous,
b) no object from $\boldsymbol{O} - A$ is not T-continuous with any object from A.

It is useful to define T-cluster using the similarity:
Object O_h is T-continuous with object O_s for given threshold T, if there exists sequence of objects $O_h = O_1, O_2, ..., O_k = O_s$, $k > 1$, that $\Pi(O_i, O_{i+1}) \geq$ T for $i = 1, ..., k-1$, where $\Pi(O_i, O_{i+1})$ is the coefficient of similarity of objects O_i and O_{i+1}. **T-continuous cluster** (**T-cluster**) is a subset $A \subset \boldsymbol{O}$, that holds:
a) each couple of objects from A is T-continuous,
b) no object from $\boldsymbol{O} - A$ is not T-continuous with any object from A.

We define a new similarity $\Pi_T(O_h, O_s)$, to describe object O_h is T-continuous with object O_s for given limit T:

$$\Pi_T(O_h, O_s) = \max_{O_{i_1}, O_{i_2}, ..., O_{i_k}} \min\{\Pi(O_h, O_{i_1}), \Pi(O_{i_1}, O_{i_2}), ..., \Pi(O_{i_k}, O_s)\},\ k \in \{1, 2. \ldots, n\text{-}2\}.$$

Hence:

$$\Pi_T(O_h, O_s) = \max_{O_k \in O} \min\{\Pi_T(O_h, O_k), \Pi_T(O_k, O_s)\}.$$

The similarity holding above mentioned relation we call T-similarity.

A dissimilarity is preferred in some clustering methods. Then we define also T-dissimilarity $\mathbf{d}_T$:

$$\mathrm{d}_T(O_h, O_s) = \min_{O_k \in O} \max\{\mathrm{d}_T(O_h, O_k), \mathrm{d}_T(O_k, O_s)\}.$$

Then we call T-cluster such a subset A of set of objects $\boldsymbol{O}$, that holds:

$$\min_{O_i,O_j \in A} \Pi_T(O_i,O_j) > \max_{O_k \in A,\, O_l \notin A} \Pi_T(O_k,O_l) \text{ or } \max_{O_i,O_j \in A} \mathbf{d}_T(O_i,O_j) < \min_{O_k \in A, O_l \notin A} \mathbf{d}_T(O_k,O_l).$$

Usually we know the similarity Π or dissimilarity **d.** We need to transform it congruously to T-similarity. To find T-clusters we then may use clustering methods based on the „classical" similarity.

4 Modification of objects similarity Π for finding T-similarity

Let us have a set of n objects $\boldsymbol{O} = \{O_1,..., O_n\}$ and their similarity $\Pi(O_h, O_s)$. We can suppose:

$$\Pi(O_h, O_h) \geq \Pi(O_h, O_s), \forall O_h, O_s \in \boldsymbol{O}.$$

Let $\Pi_1(O_h, O_s) = \Pi(O_h, O_s)$ and a new similarity $\Pi_2(O_h, O_s)$ be defined as

$$\Pi_2(O_h,O_s) = \max_{O_k \in O} \min\{\Pi_1(O_h,O_k), \Pi_1(O_k,O_s)\}.$$

Let us have n objects $\boldsymbol{O} = \{O_1,..., O_n\}$ and similarity $\boldsymbol{\Pi}(O_h, O_s)$. of them.

Holds:
a) $\Pi_2(O_h, O_s) \geq 0 \ \forall O_h, O_s \in \boldsymbol{O}$,
b) $\Pi_2(O_h, O_s) = \Pi_2(O_s, O_h) \ \forall O_h, O_s \in \boldsymbol{O}$,
c) $\Pi_2(O_h, O_h) = \Pi_1(O_h, O_h) \ \forall O_h \in \boldsymbol{O}$,
d) $\Pi_2(O_h, O_h) \geq \Pi_2(O_h, O_s) \ \forall O_h, O_s \in \boldsymbol{O}$.
Proof:

a) $\Pi_2(O_h,O_s) = \max_{O_k \in O} \min\{\Pi_1(O_h,O_k), \Pi_1(O_k,O_s)\}$,

$\Pi_1(O_h, O_k) \geq 0 \ \forall O_h, O_k \in \boldsymbol{O}$ a $\Pi_1(O_k, O_s) \geq 0 \ \forall O_k, O_s \in \boldsymbol{O}$, then
$\Pi_2(O_h, O_s) \geq 0 \ \forall O_h, O_s \in \boldsymbol{O}$.

b) $\Pi_2(O_h,O_s) = \max_{O_k \in O} \min\{\Pi_1(O_h,O_k), \Pi_1(O_k,O_s)\} =$

$= \max_{O_k \in O} \min\{\Pi_1(O_k,O_h), \Pi_1(O_s,O_k)\} = \max_{O_k \in O} \min\{\Pi_1(O_s,O_k), \Pi_1(O_k,O_h)\} = \Pi_2(O_s, O_h)$.

c) $\Pi_2(O_h, O_h) = \max_{O_k \in O} \min\{\Pi_1(O_h,O_k), \Pi_1(O_k,O_h)\} \geq$

$\geq \min\{\Pi_1(O_h,O_h), \Pi_1(O_h,O_h)\} = \Pi_1(O_h, O_h)$.

$\Pi_1(O_h, O_h) \geq \Pi_1(O_h, O_k), \forall O_k \in \boldsymbol{O}$,

$\Pi_1(O_h, O_h) \geq \max_{O_k \in O} \{\Pi_1(O_h,O_k)\} \ \forall O_k \in \boldsymbol{O}$,

$\Pi_1(O_h, O_h) \geq \max_{O_k \in O} \min\{\Pi_1(O_h,O_k), \Pi_1(O_k,O_h)\} = \Pi_2(O_h, O_h)$,

then $\Pi_2(O_h, O_h) = \Pi_1(O_h, O_s) \ \forall O_h \in \boldsymbol{O}$.

d) $\Pi_2(O_h, O_h) = \Pi_1(O_h, O_h)$.

$\Pi_1(O_h, O_h) \geq \Pi_1(O_h, O_k), \forall O_k \in \boldsymbol{O}$,

$\Pi_1(O_h, O_h) \geq \max_{O_k \in O} \{\Pi_1(O_h,O_k)\} \ \forall O_k \in \boldsymbol{O}$,

$\Pi_1(O_h, O_h) \geq \max_{O_k \in O} \min\{\Pi_1(O_h, O_k), \Pi_1(O_k, O_s)\} = \Pi_2(O_h, O_s)\ \forall O_h, O_s \in \boldsymbol{O}$,

then $\Pi_2(O_h, O_h) \geq \Pi_2(O_h, O_s)\ \forall O_h, O_s \in \boldsymbol{O}$.■

Let a similarity $\Pi_q(O_h, O_s)$ be defined as:

$$\Pi_q(O_h, O_s) = \max_{O_k \in O} \min\{\Pi_{q-1}(O_h, O_k), \Pi_1(O_k, O_s)\}.$$

Holds:
a) $\Pi_q(O_h, O_s) \geq 0\ \ \forall O_h, O_s \in \boldsymbol{O}$,
b) $\Pi_q(O_h, O_s) = \Pi_q(O_s, O_h)\ \forall O_h, O_s \in \boldsymbol{O}$,
c) $\Pi_q(O_h, O_h) = \Pi_{q-1}(O_h, O_h)\ \forall O_h \in \boldsymbol{O}$,
d) $\Pi_q(O_h, O_h) \geq \Pi_q(O_h, O_s)\ \forall O_h, O_s \in \boldsymbol{O}$,
The proof is of the same way as for $\Pi_2(O_h, O_s)$.
Holds: There exist a number m, $m \leq n$-1, $\Pi_m(O_h, O_s) = \Pi_{m+1}(O_s, O_h)\ \forall O_h, O_s \in \boldsymbol{O}$.
Proof: The similarity $\Pi_q(O_h, O_s)$ can be expressed this way:

$$\Pi_q(O_h, O_s) = \max_{O_{k_1}, O_{k_2}, \ldots, O_{k_{q-1}} \in O} \min\{\Pi_1(O_h, O_{k_1}), \Pi_1(O_{k_1}, O_{k_2}), \ldots, \Pi_1(O_{k_{q-1}}, O_s)\}.$$

If $q \geq n$-1, then some similarities will recur. Then for $r = q$+1 holds:

$$\max_{O_{k_1}, O_{k_2}, \ldots, O_{k_{q-1}} \in O} \min\{\Pi_1(O_h, O_{k_1}), \ldots, \Pi_1(O_{k_{q-1}}, O_s)\} =$$

$$= \max_{O_{k_1}, O_{k_2}, \ldots, O_{k_{r-1}} \in O} \min\{\Pi_1(O_h, O_{k_1}), \ldots, \Pi_1(O_{k_{r-1}}, O_s)\}.■$$

Holds:
The similarity $\Pi_m(O_h, O_s)$ complies: $\Pi_m(O_h, O_s) = \max_{O_k \in O} \min\{\Pi_m(O_h, O_k), \Pi_m(O_k, O_s)\}$.

Proof:
$\Pi_m(O_h, O_s) = \Pi_{m+1}(O_s, O_h) = \Pi_{m+2}(O_s, O_h) = \ldots = \Pi_{m+m}(O_s, O_h)$, where
$\Pi_{m+m}(O_h, O_s) = \max_{O_k \in O} \min\{\Pi_m(O_h, O_k), \Pi_m(O_k, O_s)\}$.■

Then $\Pi_m(O_h, O_s)$ is the searched $\Pi_T(O_h, O_s)$.

5 Algorithm for finding $\Pi_T(O_h, O_s)$

1) Let Π be a similarity of objects and P^1 a matrix of similarity of objects.
$P^1 = (p^1{}_{h,s})_{n,n}$, where $p^1{}_{h,s} = \Pi(O_h, O_s)$.
2) We find matrix P^{q+1}:
$p^{q+1}{}_{h,s} = \max_{k=1..n} \min\{p^q{}_{h,k}, p^1{}_{k,s}\}$.
3) We compare P^q a P^{q+1}.
If $P^q = P^{q+1}$, the algorithm stops and $P^q = P^T$ is the matrix of similarities.
If $P^q \neq P^{q+1}$, then q:=q+1 and algorithm continues with step 2.

6 Conclusion

We can obtain the T-clusters using the similarity matrix P^T and standard methods of hierarchical and nonhierarchical clustering.

Acknowledgements

The paper was supported by project from MSMT of the Czech Republic no. 1M06047 Centre for Quality and Reliability of Production.

References

[1] LUKASOVÁ, A.- ŠARMANOVÁ, J. Methods of Clustering Analysis, SNTL, Praha 1985 (in Czech).
[2] ZADEH, L. A. Fuzzy Sets and Their Application to Pattern Classification and Cluster Analysis. In Classification and Clustering . Academic Press, New York 1977
[3] ŽÁK, L. Clustering of vaguely defined objects, PhD Thesis, Technical University, Brno, 2002 (in Czech).

Current address

RNDr. Libor Žák Ph.D.
Mathematical Institute, Faculty of Mechanical Engineering
Technical University, Technická 2, 616 69 Brno, Czech Republic
tel.: +420-05-41142550
e-mail: zak.l@fme.vutbr.cz

Journal of Applied Mathematics

volume II (2009), number III

STRESS TESTING OF INTEREST RATE RISK

KLACSO Ján, (SK)

Abstract: Interest rate risk refers to the risk of the decline of a bank's interest rate income and market value of the security portfolio and interest rate derivatives portfolio because of unexpected changes in interest rates. This risk results from differences in the price sensitivities of assets and liabilities caused by maturity mismatch and duration mismatch in the banks balance sheet. In this paper, we examine what is the exposure of the Slovak banking sector to unexpected changes in the Basic interest rate of the National bank of Slovakia. We use error correction to model the transmission of the changes in the Basic interest rate to the inter-bank market rates, zero-coupon bond rates and client rates (interest rates for non-banks loans and deposits) and calculate how these changes affect the interest rate income, and the value of the securities and derivatives portfolio of the banking sector.

Key words: interest rate risk, stress test, interest rate transmission, cointegration, error correction

1. Introduction

Commercial banks serving as financial intermediaries use different products to meet the demand of the customers and to execute business strategies. They take deposits, grant loans, purchase securities and bonds with different maturities and interest rates. Hence, they face different kind of risks resulting from their activities. These risks can be separated into few main categories:

- credit risk
- interest rate risk
- operational risk
- market risk
- liquidity risk
- other risks

In this paper, we focus on the interest rate risk. Briefly said, this is the risk of the decline of a bank's interest rate income and market value of the security portfolio and interest rate derivatives portfolio because of unexpected changes in interest rates.

From the year 2000 the National Bank of Slovakia uses qualitative approach for conducting its monetary policy. It means that NBS controls the level of inflation by setting the value of the so

called Basic interest rate or the Policy rate. When changing the value of this rate these changes are gradually transmitted into the market rates – the inter-bank rates, zero-coupon bond rates and client rates. In this paper, we will assume that unexpected changes in the interest rates are the result of unexpected changes in the Policy rate.

The decline of the interest rate income is the result of the maturity mismatch, where typically loans are granted for longer and deposits are taken for shorter periods. From the economic theory, we assume that changes in the Policy rate are transmitted faster into the short-term deposit rates than into the long-term loan rates. It means that the increase of the Policy rate causes decrease of the net interest income. In case of the bond portfolio, we have to differentiate between fixed rate bonds and floating rate bonds. For floating rate bonds, changes of interest rates will affect the level of the coupon payment, where increase of the interest rates will increase the value of the coupon. The increase of the interest rates causes losses resulting from revaluation of bonds to fir value; these losses gradually decrease and are zero when the new value of the coupon is stated. For fixed rate bonds, changes of the interest rates don't affect the value of the coupon directly; losses from reinvestment can be calculated in case interest rates increase. Losses from revaluation are increasing during the whole life of the bond. The portfolio of interest rate derivatives is usually used to hedge the interest rate risk of the securities portfolio; we assume that this portfolio consists of interest rate swaps. These instruments can be treated as a pair of bonds, one with fixed, one with floating coupon rate.

When stress testing the interest rate risk, we test the robustness of the banking sector for sufficiently high and unexpected changes in the Policy rate. The scenario consists of an increase/decrease of the Policy rate by 2 percentage points as of 1. January 2008 and then remains unchanged for the next twelve months. We calculate the affect of this change on the value of the security portfolio, derivatives portfolio and the portfolio of loans and deposits.

The paper is organized as follows. In section two, we describe the data and the models used. In section three, we discuss the results and conclude.

2. Data set and modeling

Our data set includes the Policy rate, inter-bank market rates of all maturities (overnight, 1-, 2-weeks, 1-, 2-, 3-, 6-, 9-, and 12-months rates), zero coupon bond rates of longer maturities (from 2- to 8-years rates), client rates (corporate and household loan and deposit rates) and volumes of client loans and deposits. All of the time series consists of monthly data from January 2004 to December 2007. Unit root tests (ADF) confirmed that time series of interest rates are non-stationary series integrated of order one (**Table 1**). The portfolio of dept securities and interest rate derivatives as of December 2007 was included.

From economic theory, if Policy rate changes, the change is transmitted first into the inter-bank market rates and the coupon rates. These changes than affect the client rates and so the net interest income from loans and deposits; the coupon rates of the bond portfolio and so the net income from the coupon payments; and the real value of the bonds and interest rate derivatives. Based on this theory, first we modeled the behavior of the inter-bank market rates. We expected long-run relationship among these rates and the Policy rate, the existence of which was confirmed by Johansens cointegration tests (**Table 2**). Therefore we used error-correction model of the form:

$$\Delta r_t = -\alpha(r_{t-1} - \beta_1 rp_{t-1} - \beta_2) + \delta_u \Delta rp_t^u + \delta_d \Delta rp_t^d + \sum_{i=1}^{p} \gamma_i \Delta r_{t-i} + \sum_{i=1}^{p} \delta_i \Delta rp_{t-i} + \varepsilon_t , \qquad (2.1.)$$

where r_t is the specific inter-bank market rate, rp_t is the Policy rate, β_1 and β_2 are long run coefficients, where β_1 is the relation among the inter-bank rate and the Policy rate in the long run (it explains, if there is a change in the Policy rate, to what extent is this change transmitted into the inter-bank market rate in he long run), β_2 is the long run spread among the rates, α is the speed of reversion (it explains, how long does it take to the inter-bank rate to revert to the long run "equilibrium" in case there is a deviation from this equilibrium), δ_u and δ_v explain the direct effect of an upward or downward movement of the Policy rate (this allows us to catch the possible asymmetry in the reaction of the inter-bank rate), γ_i and δ_i are short run coefficients, p is the number of lags included and ε_t are stochastic perturbations, which are considered to be i.i.d. processes.

This equation is also used for the modeling of the behavior of the zero coupon bond rates, where we also confirmed long run relationship among the Policy rate and the bond rates (**Table 2**). Here, r_t refers to the zero coupon bond rates instead of the inter-bank rate.

The fair value or the net present value of the dept securities can be calculated by the formula

$$FV = \sum_{i=1}^{T} \frac{r_t P}{y^i} + \frac{P}{y^T}, \tag{2.2.}$$

where P is the principal paid at the maturity of the bond, r_t is the coupon rate at time t, so $r_t P$ is the coupon paid, y is the yield rate and T is the time to maturity (in months in our case). If the coupon rate is floating, the rate r_t is changed at the date of the next fixation, if the bond has fixed coupon rate, $r_t \equiv r$ is constant.

There are four categories of the impact of the changes of the interest rates on the bond portfolio. In case of floating coupon rate, increase of the interest rates means increase of the coupon paid (after the refixation is done). In case of fixed coupon rates, there is a reinvestment loss when interest rates increase, because the coupon paid does not increase with the interest rates. Losses from revaluation to fair value in case of the increase of the interest rates (increase of y) are decreasing for bonds with floating coupon rates, and are zero when the coupon rate is changed (increase of r_t). For fixed rate bonds these losses are accumulating during the whole life of the bond, because r_t is not shifted in case of changes in interest rates.

In case of the portfolio of interest rate derivatives, we assumed that all instruments are interest rate swaps (this assumption is based on banks reports about this portfolio). This derivative is used to change the type of the coupon paid, where two financial institutions can agree on this change. It means, that if there are two institutions, one paying fixed coupon, other paying floating coupon, and they want to pay the other type, they can agree on changing the payments. Therefore interest rate swaps can be separated into two parts; to a bond with fixed coupon rate and to a bond with floating coupon rate with the same principal, maturity and coupon period. This assumption allows us to calculate the profit/loss for this portfolio as for the portfolio of dept securities.

The third portfolio is that of the loans and deposits. Here, we expected that loan and deposit rates are in long run relationship with one of the inter-bank market rates or with the policy rate, therefore here we also used error-correction for modeling the behavior of these rates. We used Johansens cointegration tests to confirm the existence of long run relationship (**Table 2**). In case results confirmed relationship between specific client rate and more inter-bank rates and/or the Policy rate, we made our choice among the rates based on Akaike and Schwartz information

criterion. In some cases, no cointegration was confirmed. We described the movements of these rates by simple autoregressive equations with constant and/or trend.

In case of the existence of cointegration between specific client rate and the Policy rate, we used error-correction model as described in equation (2.1.), here, r_t refers to the client rate. If cointegration exists between the client rate and one of the inter-bank rates, we used error correction model of the form:

$$\Delta r_t = -\alpha\left(r_{t-1} - \beta_1 r_{t-1}^C - \beta_2\right) + \sum_{i=1}^{p} \gamma_i \Delta r_{t-i} + \sum_{i=0}^{p} \delta_i \Delta r_{t-i}^C + \varepsilon_t, \tag{2.3.}$$

where r_t^C is the inter-bank rate. The difference between equation (2.1.) and (2.3.) is that in the latter we did not separate the direct effect of the upward and downward movement of the inter-bank rate. This is because the potential asymmetric reaction in this case is built in the reaction of the inter-bank rate to the changes in the Policy rate.

Volumes of loans and deposits were modeled as simple autoregressive processes with constant and/or trend:

$$v_t = c_1 + c_2 t + AR(p) + \varepsilon_t, \tag{2.4.}$$

where v_t is the volume of the specific loan or deposit, t refers to trend and p is the number of lags.

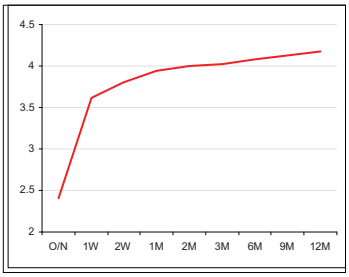
The interest rate risk refers to losses which are the result of unexpected changes of the Policy rate; therefore we need to estimate the expected changes of the inter-bank market for the year 2008 of this rate. First conclusions can be drawn from the yield curve.

The normal shape of the yield curve indicates that the market does not expect movements of the Policy rate, or expects just slight increase. To confirm this, we tried to create a simple equation. First, we calculated by interpolation the expected values of the 2 weak BRIBOR rate for the year 2008, based on no arbitrage assumption. Than, values of expected Policy rates were calculated using the equation

$$rp_t^e = c_1 + c_2 r_t^{2w-e} + AR(2) + \varepsilon_t, \tag{2.5.}$$

where rp_t^e is the expected value of the policy rate and r_t^{2w-e} is the expected value of the 2 week BRIBOR rate. Results of the calculation confirmed the conclusions drawn from the shape of the yield curve.

Figure 1 Yield curve as of December 2007



3. Results and conclusion

The scenario for stress testing was created based on the calculated values of the expected Policy rate. We set up a 2 percentage point decrease/increase as of 1. January 2008 and expected no changes the rest of the year.

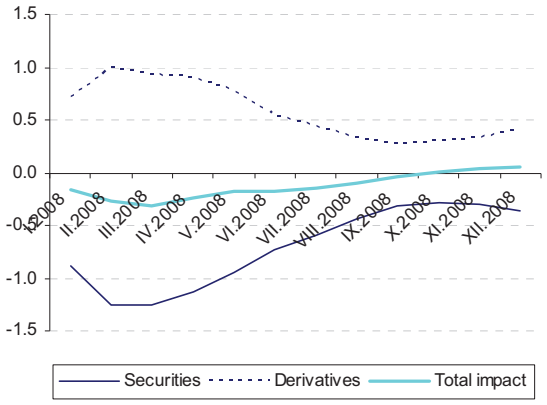
First, we estimated the movements of inter-bank market rates. Results show that rates of shorter maturities would react symmetrically to the increase/decrease of the Policy rate. After increase in the first half of the year 2008 slight decrease would be observable, changes of the Policy rate would be transmitted to full extent for rates of shorter maturities, but just partially for rates of longer maturities (**Table 3**).

Zero coupon bond rates would react like the inter-bank rates, after increase in the first half of the year there would be a slight decrease, the transmission would not be complete. Also asymmetric reactions would be expected with higher movements in case of increase. (**Table 3**)

Changes would not be fully transmitted into client rates; also the asymmetry would be smaller than in previous cases. (**Table 4**, **Table 5**)

For the calculation of the profit/loss stemming from the movements stated in the scenarios, we used two approaches based on that the portfolio of banks can be separated into two parts. The first part consists of bonds and derivatives revaluated to fair value. For bonds, it is the Fair value and Trade portfolio (FV and TRD). We expected that interest rate derivatives are not used to hedge the portfolio of bonds which are not revaluated to fair value. The second part is the portfolio of bonds not revaluated to fair value and the client loans and deposits. The bonds not revaluated to fair value are bonds in the Held to maturity portfolio, and also bonds in the Available for sale portfolio are not revaluated against profit and loss (HTM and AFS).

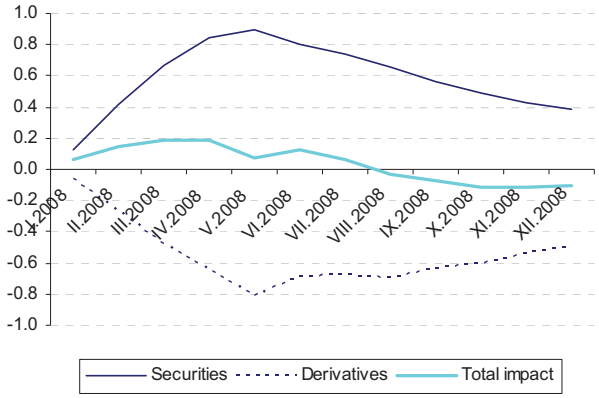
Figure 2 Impact of the increase of the Policy rate by 2 p.p., 1. approach



In the first approach, we assumed that only bonds in FV and TRD portfolio and interest rate derivatives are revaluated to real value. Results show that after the increase of the interest rate losses from the securities portfolio would be culminating during the first three months, slight decrease would be observable after this period. This decrease of losses is because the profit from coupon payment and also because of the movements of inter-bank rates. Results also show that reactions of the derivative portfolio would be the opposite of the reactions of the securities portfolio. This confirms the fact that interest rate derivatives are used for hedging the securities portfolio against interest rate risk. The total impact of the interest rate movements would be nearly zero, which means that the FV and TRD portfolio is completely hedged against interest rate risks on the aggregate level.

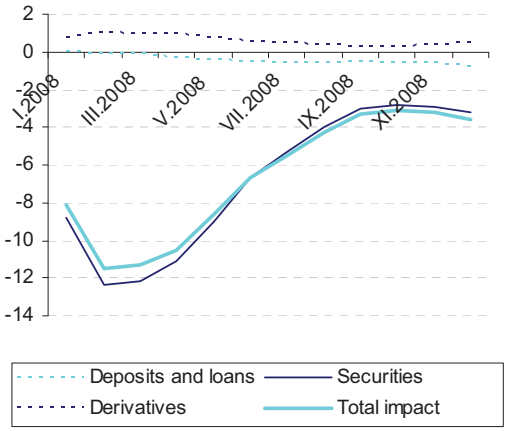
The scenario of the decrease of the Policy rates showed the same results. Profit from the securities portfolio would be balanced by losses from derivative portfolio; total impact would be nearly zero.

Figure 3 Impact of the decrease of the Policy rate by 2 p.p. , 1. approach



In the second approach we expected that the entire securities portfolio is revaluated to real value. This is a reasonable assumption because in case of trouble bank may sell instruments from HTM portfolio, too and in this case the bank has to revaluate also this portfolio. Results show that in case of increase of the Policy rate, the security and the derivatives portfolio would react like in the previous case. However, losses from the security portfolio would be much greater than profit from derivatives portfolio, which means that the HTM and AFS portfolio is not hedged fully against the interest rate risk on the aggregate level. Losses from the portfolio of loans and deposits would be accumulating during the whole period. As in case of increasing Policy rate there are large losses for the banking sector, the case of the decrease of the Policy rate is not as interesting.

Figure 4 Impact of the increase of the Policy rate by 2 p.p., 2. approach



In sum, we may conclude that interest rate risk is an issue for banks, as the portfolio of securities, interest rate derivatives and loans and deposits would react to changes of the Policy rate. However, banks seem to be hedged against this risk, as results do not show large impact and losses in the first approach, losses are calculated using the second approach, but these are more theoretical and not so realistic.

References

[1.] SVERIGES RISKBANK, 'The transmission mechanism', Available on internet: < http://www.riksbank.com/templates/Page.aspx?id=10547 >
[2.] BANK FOR INTERNATIONAL SETTLEMENTS, 'The transmission of monetary policy in emerging market economies', BIS policy papers No. 3. January 1998
[3.] HEFFERNAN, Shelagh, FUERTES, Ana – Maria, 'Bank Heterogeneities in the Interest Rate Transmission Mechanism', Cass Business School Research Paper, July 2006, Available on internet: < http://papers.ssrn.com/sol3/papers.cfm?abstract_id=903348 >
[4.] LEON, Costas, CHIONIS, Dionysos, 'Modeling Interest Rate Transmission Dynamics in Greece – Is There Any Structural Brake After EMU?', Working Paper series, May 2005, Available on internet: < http://papers.ssrn.com/sol3/papers.cfm?abstract_id=815584 >
[5.] BURGSTALLER, Johann, 'Interest Rate Transmission to Commercial Credit Rates in Austria', Working Paper No. 0306, May 2003, Available on internet: < http://ideas.repec.org/p/jku/econwp/2003_06.html >
[6.] ENDERS, Walter. 'Applied econometric time series', 1995, John Wiley and Sons, Inc.

Current address

Ján Klacso, Mgr.
Comenius University, Faculty of Mathematics, Physics and Informatics,
Department of Applied Mathematics and Statistics
Mlynská dolina
842 48 Bratislava
tel.: +421 2 654 24 000
mail.: jan.klaso@fmph.uniba.sk, klacso@gmail.com

Appendix

Table 1 Results of ADF unit root tests

Interest rate	Level		1st difference	
	t-statistics	p-value	t-statistics	p-value
Policy rate *	-2.46	0.13	-6.94	0.00
O/N BRIBOR *	-1.76	0.40	-12.05	0.00
1W BRIBOR *	-1.60	0.47	-8.69	0.00
2W BRIBOR *	-1.55	0.50	-7.47	0.00
1M BRIBOR *	-1.74	0.41	-5.76	0.00
2M BRIBOR *	-1.82	0.37	-5.08	0.00
3M BRIBOR *	-1.88	0.34	-4.88	0.00
6M BRIBOR *	-1.90	0.33	-4.56	0.00
9M BRIBOR *	-1.96	0.30	-4.48	0.00
12M BRIBOR *	-1.92	0.32	-4.56	0.00
Bond 2y *	-1.51	0.52	-5.07	0.00
Bond 3y *	-2.06	0.26	-3.89	0.00
Bond 4y *	-1.98	0.29	-4.29	0.00
Bond 5y *	-2.04	0.27	-4.31	0.00
Bond 6y *	-2.01	0.28	-4.53	0.00
Bond 7y *	-1.96	0.30	-4.82	0.00
Bond 8y *	-1.95	0.31	-4.96	0.00
Corporate loans and deposits rates				
Interest rate	Level		1st difference	
	t-statistics	p-value	t-statistics	p-value
Loans with maturity up to 1 year *	-2.64	0.09	-7.50	0.00
Loans with maturity 1 to 5 years **	-2.26	0.19	-3.48	0.01
Loans with maturity more than 5 years **	-1.87	0.34	-3.35	0.02
Sight deposits *	-1.91	0.33	-10.90	0.00
Saving deposits *	-1.99	0.29	-13.78	0.00
o/n deposits *	-2.15	0.23	-14.98	0.00
Deposits with maturity up to 7 days *	-1.78	0.39	-8.94	0.00
Deposits with maturity 7 days to 1 month *	-1.56	0.50	-9.77	0.00
Deposits with maturity 1 to 3 months *	-1.54	0.51	-5.07	0.00
Deposits with maturity 3 to 6 months *	-2.08	0.25	-6.24	0.00
Deposits with maturity 6 to 12 months *	-1.60	0.48	-7.44	0.00
Deposits with maturity 1 to 2 years *	-1.72	0.42	-7.89	0.00
Deposits with maturity 2 to 5 years *	-2.41	0.14	-6.08	0.00
Deposits with maturity more than 5 years *	-0.74	0.83	-6.69	0.00

Households loans and deposits rates				
Interest rate	Level		1st difference	
	t-statistics	p-value	t-statistics	p-value
Loans with maturity up to 1 year *	-0.76	0.82	-8.71	0.00
Loans with maturity 1 to 5 years *	-2.46	0.13	-7.56	0.00
Loans with maturity more than 5 years *	-0.91	0.78	-7.99	0.00
Sight deposits *	-1.92	0.32	-12.01	0.00
Saving deposits *	-1.63	0.46	-5.87	0.00
o/n deposits **	-2.36	0.16	-3.32	0.02
Deposits with maturity up to 7 days *	-2.03	0.28	-8.64	0.00
Deposits with maturity 7 days to 1 month *	-1.58	0.49	-5.27	0.00
Deposits with maturity 1 to 3 months *	-1.17	0.68	-4.57	0.00
Deposits with maturity 3 to 6 months **	-1.65	0.45	-2.92	0.05
Deposits with maturity 6 to 12 months **	-2.72	0.08	-2.82	0.01
Deposits with maturity 1 to 2 years **	-2.19	0.21	-2.02	0.04
Deposits with maturity 2 to 5 years *	-1.73	0.41	-8.23	0.00
Deposits with maturity more than 5 years *	-1.31	0.62	-7.62	0.00

* 1% significance level
** 5% significance level

Table 2 Results of tests of cointegration

Cointegrating time series	Number of CE(s)	Eigenvalue	Trace Statistic	p-value
O/N BRIBOR, Policy rate	None	0.45	37.99	0.00
	At most 1	0.05	2.94	0.09
1W BRIBOR, Policy rate	None	0.33	25.89	0.00
	At most 1	0.05	3.06	0.08
1M BRIBOR, Policy rate	None	0.30	24.02	0.00
	At most 1	0.05	3.18	0.07
2M BRIBOR, Policy rate	None	0.31	24.78	0.00
	At most 1	0.06	3.29	0.07
3M BRIBOR, Policy rate	None	0.31	24.75	0.00
	At most 1	0.06	3.42	0.06
6M BRIBOR, Policy rate	None	0.28	22.25	0.00
	At most 1	0.06	3.29	0.07
9M BRIBOR, Policy rate	None	0.27	21.88	0.00
	At most 1	0.06	3.48	0.06
12M BRIBOR, Policy rate	None	0.27	21.71	0.01
	At most 1	0.06	3.34	0.07
Bond 2y, Policy rate	None	0.34	26.45	0.00
	At most 1	0.12	6.24	0.01
Bond 3y, Policy rate	None	0.33	26.02	0.00
	At most 1	0.13	6.60	0.01
Bond 4y, Policy rate	None	0.31	24.49	0.00
	At most 1	0.12	6.17	0.01
Bond 5y, Policy rate	None	0.27	21.89	0.00
	At most 1	0.12	6.20	0.01
Bond 6y, Policy rate	None	0.25	20.60	0.01
	At most 1	0.13	6.57	0.01
Bond 7y, Policy rate	None	0.22	19.42	0.01
	At most 1	0.13	6.98	0.01
Bond 8y, Policy rate	None	0.20	18.18	0.02
	At most 1	0.14	7.30	0.01

Corporate loans and Inter-bank rates/Policy rate				
Cointegrating time series	Number of CE(s)	Eigenvalue	Trace Statistic	p-value
Loans with maturity up to 1 year , 3M BRIBOR	None	0.30	23.98	0.00
	At most 1	0.06	3.62	0.06
Loans with maturity 1 to 5 years, 3M BRIBOR	None	0.37	31.14	0.00
	At most 1	0.07	4.09	0.04
Loans with maturity more than 5 years, 6M BRIBOR	None	0.29	22.89	0.00
	At most 1	0.06	3.40	0.07
Sight deposits, 2W BRIBOR	None	0.29	22.36	0.00
	At most 1	0.05	2.83	0.09
Saving deposits, 3M BRIBOR	None	0.33	26.98	0.00
	At most 1	0.06	3.84	0.05
o/n deposits, Policy rate	None	0.32	25.08	0.00
	At most 1	0.05	2.96	0.09
Deposits with maturity up to 7 days, 2W BRIBOR	None	0.30	23.96	0.00
	At most 1	0.05	2.89	0.09
Deposits with maturity 7 days to 1 month, 1M BRIBOR	None	0.24	20.51	0.01
	At most 1	0.08	4.76	0.03
Deposits with maturity 1 to 3 months, 2M BRIBOR	None	0.40	32.21	0.00
	At most 1	0.05	3.02	0.08
Deposits with maturity 3 to 6 months, 12M BRIBOR	None	0.42	34.90	0.00
	At most 1	0.06	3.67	0.06
Deposits with maturity 6 to 12 months, 3M BRIBOR	None	0.18	12.62	0.04
	At most 1	0.01	0.85	0.41
Deposits with maturity 1 to 2 years, 1M BRIBOR	None	0.25	20.07	0.03
	At most 1	0.06	3.48	0.06
Deposits with maturity 2 to 5 years, 1M BRIBOR	None	0.17	12.83	0.04
	At most 1	0.04	2.07	0.18
Deposits with maturity more than 5 years, Policy rate	None	0.41	34.29	0.00
	At most 1	0.06	3.47	0.82

Households loans and Inter-bank rates/Policy rate				
Cointegrating time series	Number of CE(s)	Eigenvalue	Trace Statistic	p-value
Saving deposits, 2M BRIBOR	None	0.22	19.38	0.01
	At most 1	0.08	4.91	0.03
o/n deposits, 9M BRIBOR	None	0.44	37.16	0.00
	At most 1	0.06	3.85	0.05
Deposits with maturity up to 7 days, 2W BRIBOR	None	0.18	14.25	0.08
	At most 1	0.05	2.87	0.09
Deposits with maturity 7 days to 1 month, 1M BRIBOR	None	0.12	9.03	0.36
	At most 1	0.03	1.81	0.18
Deposits with maturity 1 to 3 months, 1M BRIBOR	None	0.23	18.37	0.02
	At most 1	0.06	3.30	0.07
Deposits with maturity 3 to 6 months, 12M BRIBOR	None	0.32	26.21	0.00
	At most 1	0.07	3.92	0.05
Deposits with maturity 6 to 12 months, 12M BRIBOR	None	0.21	14.40	0.02
	At most 1	0.01	0.80	0.43
Deposits with maturity 1 to 2 years, Policy rate	None	0.26	20.31	0.00
	At most 1	0.04	2.26	0.16
Deposits with maturity 2 to 5 years, Policy rate	None	0.25	19.87	0.01
	At most 1	0.05	3.23	0.07

Table 3 Estimated short term coefficients, number of lags and the adjusted R^2 for the inter-bank market rates and zero coupon bond rates

	α	β_1	β_2	p	R^2
o/n BRIBOR	1.7	1.0	-0.4	2	64%
1W BRIBOR	1.0	1.0	-0.2	3	46%
1M BRIBOR	0.7	1.0	-0.1	2	40%
2M BRIBOR	0.5	1.0	-0.0	1	42%
3M BRIBOR	0.4	1.0	0.1	1	46%
6M BRIBOR	0.3	0.9	0.5	1	46%
9M BRIBOR	0.2	0.8	0.7	1	47%
12M BRIBOR	0.2	0.7	1.0	2	47%
Bond 2y	0.3	0.7	1.5	2	51%
Bond 3y	0.2	0.4	2.4	2	48%
Bond 4y	0.2	0.5	2.2	2	40%
Bond 5y	0.6	0.7	1.2	7	37%
Bond 6y	0.5	0.6	1.6	5	38%
Bond 7y	0.5	0.6	1.7	5	37%
Bond 8y	0.5	0.6	1.9	5	35%

Table 4 Estimated short term coefficients, number of lags and the adjusted R^2 for the corporate loan and deposit rates

	α	β_1	β_2	p	R^2	r^C
Loans with maturity up to 1 year	0.2	0.9	2.0	2	73%	3M BRIBOR
Loans with maturity 1 to 5 years	0.1	0.9	2.2	3	83%	3M BRIBOR
Loans with maturity more than 5 years	0.2	1.2	0.6	2	84%	6M BRIBOR
Sight deposits	1.0	0.3	-0.5	1	63%	2W BRIBOR
Saving deposits	0.3	0.7	-0.7	1	58%	3M BRIBOR
o/n deposits	1.4	1.0	0.9	1	66%	Policy rate
Deposits with maturity up to 7 days	1.3	0.9	-0.3	1	68%	2W BRIBOR
Deposits with maturity 7 days to 1 month	0.4	0.9	0.0	2	89%	1M BRIBOR
Deposits with maturity 1 to 3 months	0.5	1.0	-1.0	1	74%	2M BRIBOR
Deposits with maturity 3 to 6 months	0.4	1.0	-0.7	1	47%	12M BRIBOR
Deposits with maturity 6 to 12 months	0.4	0.7	-0.3	1	59%	3M BRIBOR
Deposits with maturity 1 to 2 years	0.7	0.8	-0.9	1	63%	1M BRIBOR
Deposits with maturity 2 to 5 years	0.4	0.8	-0.2	2	41%	1M BRIBOR
Deposits with maturity more than 5 years	0.3	0.1	2.2	3	33%	Policy rate

Table 5 Estimated short term coefficients, number of lags and the adjusted R^2 for the households loan and deposit rates

	α	β_1	β_2	n	R^2	r^C
Loans with maturity up to 1 year						
Loans with maturity 1 to 5 years						
Loans with maturity more than 5 years						
Sight deposits						
Saving deposits	0.5	0.6	-0.7	4	60%	2M BRIBOR
o/n deposits	1.3	1.0	-1.4	1	64%	9M BRIBOR
Deposits with maturity up to 7 days	0.4	0.7	-0.7	3	69%	2W BRIBOR
Deposits with maturity 7 days to 1 month	0.3	0.7	-0.7	0	69%	1M BRIBOR
Deposits with maturity 1 to 3 months	0.2	0.6	-0.1	2	76%	1M BRIBOR
Deposits with maturity 3 to 6 months	0.2	0.8	-0.9	2	84%	12M BRIBOR
Deposits with maturity 6 to 12 months	0.1	0.7	-0.2	2	73%	12M BRIBOR
Deposits with maturity 1 to 2 years	0.2	0.6	0.3	3	47%	Policy rate
Deposits with maturity 2 to 5 years	0.1	1.1	-2.4	1	17%	Policy rate
Deposits with maturity more than 5 years						

Journal of Applied Mathematics

volume II (2009), number III

MODELS OF INTEREST RATE EVOLUTION - VAŠÍČEK AND CIR MODELS

OTTA Josef, (CZ)

Abstract. This paper is devoted to one-dimensional stochastic models of interest rates. We focused to Vašíček model and Cox-Ingersoll-Ross model which are still quite popular and provide their analysis and improved parameter estimation, based on exact integration of interpolated interest rates, with respect to real LIBOR data.

Key words and phrases. Interest rates, parameter estimation, Vašíček model, CIR model.

Mathematics Subject Classification. Primary 60A05, 08A72; Secondary 28E10.

1 Introduction

The interest rates play an important role in all transactions based on lending and borrowing. Their importance has risen with developing of World economics. In this paper we focus to one dimensional mathematical models of interest rates which are used for predicting of future rates to estimate that the lending and borrowing operations will be profitable for us. Let us aim to well-known Vašíček and CIR models.

2 Vašíček Model

Vašíček model is a mathematical one-factor model describing an evolution of interest rates. It describes interest rate movements as driven by only one source of market risk. The model can be used in the evaluation of interest rate derivatives and it has been also adapted for credit markets. It was introduced in 1977 by Oldřich Vašíček, [7]. The equation has following form

$$\mathrm{d}r(t) = a(b - r(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{1}$$

where $W(t)$ is a Wiener process which models the random market risk factor. Parameter σ is a standard deviation parameter and it determines the volatility of the interest rate. Value of parameter a stands for rate of convergence to the mean, $a > 0$, and b is long-term expected interest rate.

The model has property *mean reversion*. This property is typical for behavior of interest rates, in the case that rate is higher than value b the value of rate is forced down. On the other hand, if rate $r(t)$ is lower than b its drift is positive and leads $r(t)$ to expected interest rate b.

2.1 Simulation of Vašíček Model

The simulation of Vašíček interest rate model can be performed by discretization of basic Vašíček equation (1) in the form

$$r(t + \Delta t) = r(t) + \Delta r \simeq r(t) + a(b - r(t))\Delta t + \sigma \Delta W(t) \tag{2}$$

The random process $\Delta W(t)$ can be simulated by $\mu_t \sqrt{\Delta t}$, where μ_t is generated random number from $\mathcal{N}(0, 1)$, cf. [5]. Simulations on Figures 1, 2 illustrate the behavior of this model. One

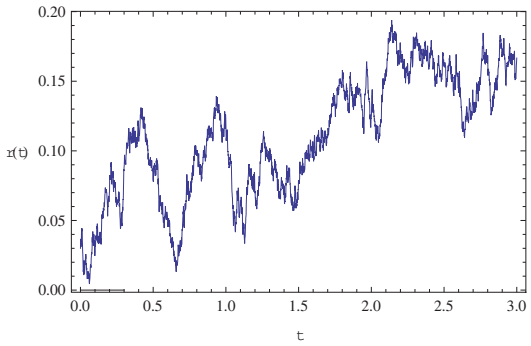


Figure 1: Vašíček model for $a = 0.02, b = 0.04, \sigma = 0.1, \Delta t = 0.001$ with positive spot rates

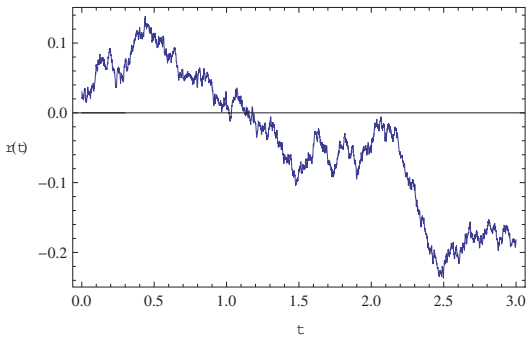


Figure 2: Vašíček model for $a = 0.02, b = 0.04, \sigma = 0.1, \Delta t = 0.001$ with negative spot rates

can observe the most problematic point of Vašíček model – the spot rates can become negative. For parameters $\sigma = 0.1$, long term interest rate $b = 0.04$ and starting value $r(0) = 0.03$ we can get many negative values $r(t)$. Moreover Vašíček model can have negative expected value. In the case that $r(t)$ is for some t negative, as it was illustrated in the Figure 2, the conditional expected value of interest rate

$$\mathbb{E}[r(T)|r(t)] = b + (r(t) - b)e^{-a(T-t)} \tag{3}$$

can be negative too. This is an inconvenient property of this model. On the other hand, if we consider real interest rates with respect to inflation, the negative rates can occur, [5].

2.2 Yield Curve

The yield curve is a representation of the dependence of interest rates $R(t,T)$ on the time t to the maturity T. Yield curve, sometimes called term structure, is well studied object because it helps to give an idea of future interest rates by measuring expectations based on current market conditions. There exist three main types of these curves - normal, inverted and flat shaped curve. The shape of these curves basically identifies the market behavior, cf. [2].

Let us determine the value of *zero bond* with risk neutral approach, cf. [5, p. 92]. By integration of Vašíček equation (1) from t to T and using variation variable method we get

$$r(T) = r(t)e^{-a(T-t)} + b(1 - e^{-a(T-t)}) + \int_t^T e^{-a(T-t-s)}\mathrm{d}W \tag{4}$$

One can see that $r(T)$ has normal distribution and it can be proven that the conditional expected value is defined by (3) and variance has following form

$$\mathrm{var}[r(T)|r(t)] = \frac{\sigma^2}{2a}\left(1 - e^{-2a(T-t)}\right), \tag{5}$$

cf. [6, Theorem 1.69, p. 286].

The price of zero bond at the time t with nominal value 1 is equal to

$$P(t,T) = \mathbb{E}[e^{-\int_t^T r(s)\mathrm{d}s}]. \tag{6}$$

After several steps, cf. [6, p. 294], we get the formula for zero bond in the form

$$P(t,T) = e^{A(t,T)-B(t,T)r(t)} \quad \text{where} \quad A(t,T) = \frac{a^2b - \sigma^2/2}{a^2}\left(B(t,T) - (T-t)\right) - \frac{\sigma^2 B^2(t,T)}{4a} \tag{7}$$

$$\text{and} \quad B(t,T) = \frac{1}{a}\left(1 - e^{-a(T-t)}\right).$$

To get the interest rate $R(t,T)$ from time t to T we use

$$R(t,T) = -\frac{1}{T-t}\ln\left(P(t,T)\right).$$

Thus

$$R(t,T) = R + \frac{1}{aT}(r(t) - R)\left(1 - e^{-aT}\right) + \frac{\sigma^2}{4a^3T}\left(1 - e^{-aT}\right)^2, \quad \text{where} \quad R = b - \frac{\sigma^2}{2a^2}. \tag{8}$$

If we pass $T \to \infty$ we get $\lim_{T\to\infty} R(t,T) = R$ which is a yield of the bond. Therefore we can see that the yield does not depend on the initial value of interest rate $r(t)$. Figures 3, 4 and 5 illustrates the behavior of yield curves for Vašíček model. It was shown that the curve has inverted shape for $r(t) > R + \frac{\sigma^2}{4a}$; it is increasing(normal shaped) for $r(t) \leq R - \frac{\sigma^2}{4a}$ and it has humped structure for $r(t) \in [R - \frac{\sigma^2}{4a}, R + \frac{\sigma^2}{4a})$, cf. [7]. Let us remark the difficulty to get humped yield curve for real data due to the narrowness of the interval $[R - \frac{\sigma^2}{4a}, R + \frac{\sigma^2}{4a})$.

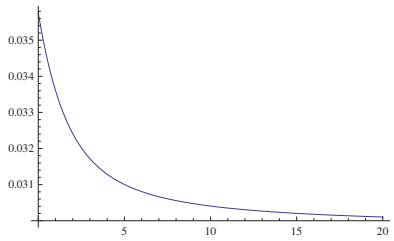


Figure 3: Inverted shaped yield curve for (1).

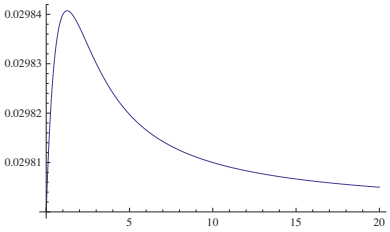


Figure 4: Humped structure of yield curve for (1).

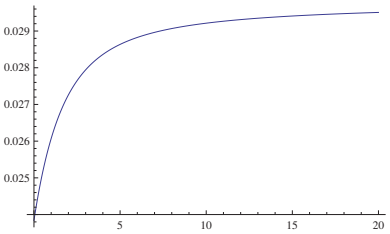


Figure 5: Normal shaped yield curve for (1).

2.3 Parameter Identification

Now, let us focus to the finding of the values of parameters a, b and σ to determine yield curve for real interest rates in the time interval [0,T]. Classical discretization scheme for stochastic equation which has been used for simulation of Vašíček model (2) is not one of the best.

Let us recall equation (1) and integrate it from t_k to t_{k+1}, where t_k are quoted values, i.e. $t_k \in [0, T], k = 0, \ldots, n$. Hence we obtain

$$r_{k+1} = r_k + \int_{t_k}^{t_{k+1}} a(b - r(t))\mathrm{d}t + \int_{t_k}^{t_{k+1}} \sigma \mathrm{d}W, \quad \text{where} \quad r_k = r(t_k),\ r_{k+1} = r(t_{k+1}). \tag{9}$$

A classical choice of the approximation of $\int_{t_k}^{t_{k+1}}(\cdot)\mathrm{d}t$ is a substitution of the integrand by constant function $a(b - r_k)$ which leads to the Euler scheme, cf. [3, p. 183],[5].

But the economists usually work with values r_k and they interpolate values between quoted data linearly, cf. [4]. Let us take advantage of this approach and compute the integral $\int_{t_k}^{t_{k+1}}(\cdot)\mathrm{d}t$ exactly for interpolating function between neighboring quoted rates r_k and r_{k+1} in the form

$$\int_{t_k}^{t_{k+1}} a(b - r(t))\mathrm{d}t = ab\Delta t - \frac{a\Delta t}{2}(r_k + r_{k+1}). \tag{10}$$

This approach is more natural than standard Euler discretization scheme and it is suitable for parameter identification method. The stochastic integral can be computed as

$$\int_{t_k}^{t_{k+1}} \sigma \mathrm{d}W = \sigma(W(t_{k+1}) - W(t_k)) \tag{11}$$

where the difference $\Delta W_k = W(t_{k+1}) - W(t_k)$ has normal distribution $\mathcal{N}(0, \Delta t)$, Therefore we can model this process by some randomly chosen values with normal distribution $\mathcal{N}(0, \Delta t)$.

If we employ (10), (11) to identity (9) we obtain system of n linear algebraic equations

$$r_{k+1} = \alpha r_k + \beta + \gamma \Delta W_k, \quad k = 0, \ldots, n-1 \tag{12}$$

with

$$\alpha = \frac{1 - \frac{a}{2}\Delta t}{\nu}, \quad \beta = \frac{ab\Delta t}{\nu}, \quad \gamma = \frac{\sigma}{\nu} \quad \text{and} \quad \nu = 1 + \frac{a}{2}\Delta t.$$

The system of equations (12) can be solved for unknown α, β and γ by *least square method*.

Now, let us put $\gamma := 0$ and solve the system (12) to get values of α and β. Let us define function $\tilde{r}(t_k + 1) = \alpha\tilde{r}(t_k) + \beta$, for $k = 0, \ldots, n-1$ and investigate the difference $g(t_k) := r(t_k) - \tilde{r}(t_k)$. Due to using least square method the set $\{g(t_k)\}_{k=0}^{n}$ has zero mean value and some positive variance $v(g)$. Since the interest rates r_k are real, we cannot expect "pure" normal distribution of g. Thus we recommend to compare *Cumulative distribution function* of g with normal distribution function $\mathcal{N}(0, v(g))$.

If we solve the system (12) with $\Delta W_k := g(t_k)$ we get the same values of parameters α, β and $\gamma = \sqrt{v(g)}$. These values are then considered as a solution of (12). By substitution back to (10) and (11), we get values of original parameters a, b and σ in the form

$$a = \frac{2}{\Delta t}\frac{1-\alpha}{1+\alpha}, \quad b = \frac{\beta}{1-\alpha} \quad \text{and} \quad \sigma = \frac{2\gamma}{1+\alpha}. \tag{13}$$

Thanks to this values we can find upper bound r^u and lower bound r^l estimate for future interest rate using 3σ-rule. Let us recall condotional expected value of r (3) and variance (5) then upper bound estimate is defined by

$$r^u(t,T) = b + (r(t) - b)e^{-a(T-t)} + 3\sigma\sqrt{\frac{1 - e^{-2a(T-t)}}{2a}}, \tag{14}$$

the lower estimate in the form

$$r^u(t,T) = b + (r(t) - b)e^{-a(T-t)} - 3\sigma\sqrt{\frac{1 - e^{-2a(T-t)}}{2a}} \tag{15}$$

and the main function r^m of $r(t)$ is defined by

$$r^m(t,T) = b + (r(t) - b)e^{-a(T-t)}. \tag{16}$$

The main function corresponds to the theoretical Vašíček model for $\sigma = 0$. Thanks to these bounds we justify the assumption $a \geq 0$. In the case that $a < 0$ the model is unstable and bounds defined above diverges with increasing T.

3 CIR Model

The Cox-Ingersoll-Ross model (CIR) in also well-known mathematical model which describes interest rate movements as driven by only one source of market risk. It was introduced in 1985 by John C. Cox, Jonathan E. Ingersoll and Stephen A. Ross as an extension of the Vašíček model. It has following form

$$\mathrm{d}r(t) = a(b - r(t)) + \sigma\sqrt{r(t)}\mathrm{d}W(t), \tag{17}$$

where $W(t)$ is a Wiener process which models the random market risk factor. Parameter σ is a normalized standard deviation parameter and it determines the volatility of the interest rate.

The difference of CIR model and Vašíček model is in the fact that CIR model does not admit negative values of interest rates, cf. [5]. If $r(0) > 0$ then $r(t) \geq 0$ for all $t > 0$.

	Data					
	EUR 3M			USD 3M		
Parameters	a	b	σ	a	b	σ
VAS	0.015013	2.1372	0.00729136	0.00860017	2.46422	0.0676337
CIR	0.0150526	2.11576	0.00506814	0.00941037	2.24241	0.0349574
VAS-CIR	−0.00003961	0.021506	−0.00222322	−0.0008102	0.221804	−0.0326763

Table 1: Values of parameters for Vašíček model (VAS) and CIR model for given data with their comparison (absolute and relative difference).

The analysis of bond prices is similar to Vašíček model. For more details we refer to [5] and [6]. The parameter estimation procedure is based on the same idea as used in Vašíček method therefore we omit the implementation to the reader.

4 Application to LIBOR data

Let us deploy these models to the real *London Interbank Offered Rate data*(LIBOR) provided by British Bankers Association, [1]. In this paper we present analysis of interest rates for LIBOR for USD from 1.11.2007 to 31.10.2008, illustrated on the Figure 7, and for 3-month EUR data from 16.11.2003 to 27.5.2005 its history of interest rates is illustrated on the Figure 8.

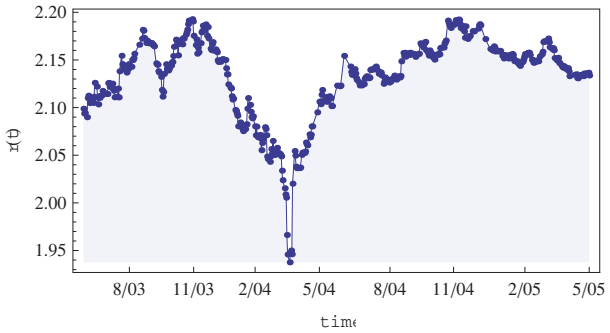


Figure 6: EUR LIBOR 3M data from 16.11.2003 to 27.5.2005.

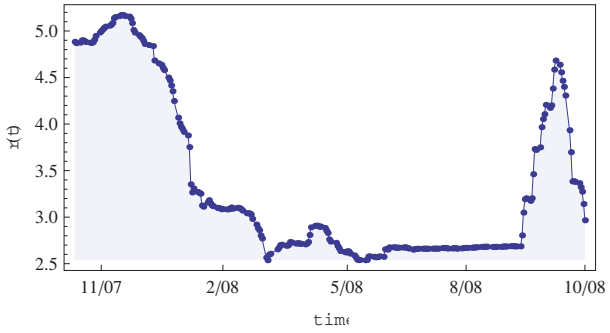


Figure 7: USD LIBOR 3m data from 1.11.2007 to 31.10.2008.

Now we try to fit the Vašíček and CIR model to this LIBOR data. We solve system of equations (12) to get parameters α, β and γ. Then we substitute these values back to (13) and get values of original parameters a, b and σ for Vašíček model. Parameters for CIR model are computed analogously. The values of parameters for given interest rates are presented in Table 1.

One can observe that we get similar values for Vašíček and CIR model which is in correspondence with the fact that these models has quite similar behavior, [5], [6].

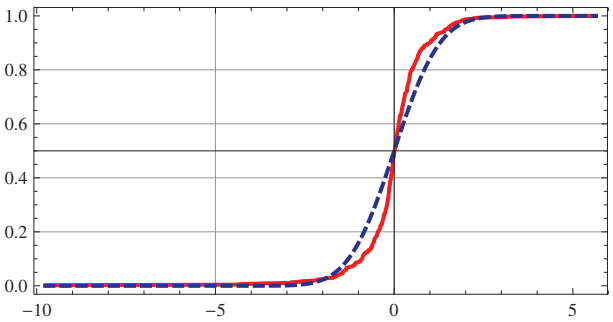


Figure 8: CDF for ΔW_t for data EUR LIBOR 3M with respect to CDF of $\mathcal{N}(0, \Delta t)$.

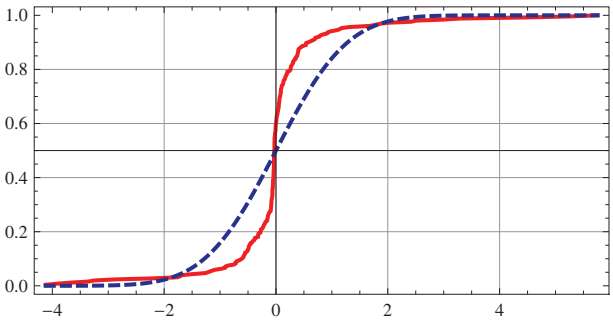


Figure 9: CDF for ΔW_t for data USD LIBOR 3M with respect to CDF of $\mathcal{N}(0, \Delta t)$.

Thanks to this values, we are able to determine the term structure for some given spot interest rate $r(t)$ by formula (8). The illustration of term structures $R(t, T)$ for Vašíček model applied to EUR LIBOR 3M data is presented on Figure 10 and its contours are shown on Figure 11.

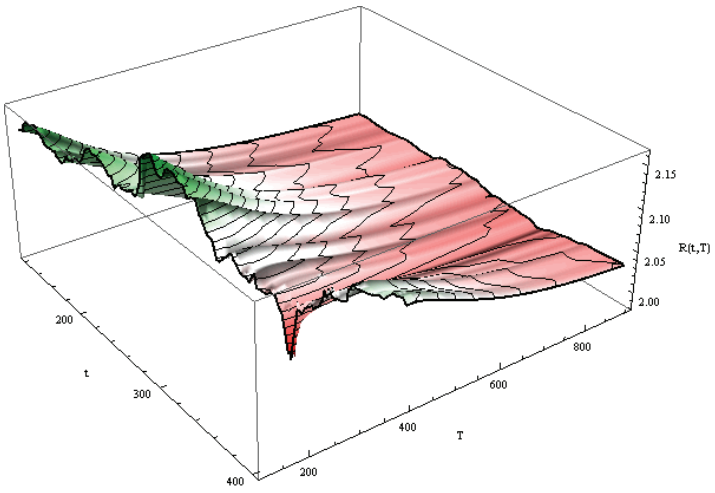


Figure 10: $R(t, T)$ for EUR data.

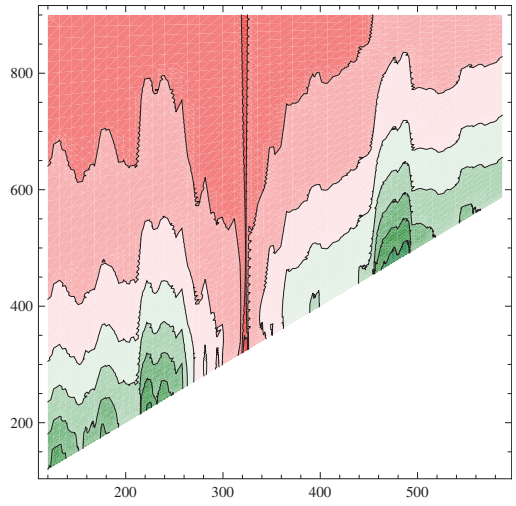


Figure 11: Contours of $R(t, T)$ for EUR data.

We are also able to compute expectation of interest rate by using 3σ rule. By definition of upper bounds (14) and lower bounds (15) we can predict the behavior of interest rates by values of current and past rates. On Figures 12 and 13 one can see the upper bounds displayed by blue curves and lower bounds by brown curves. The main function r_m is illustrated by thick orange curve. The blue area above dashed line enclose values of $r(t)$ lying in the interval $[\text{mean}[r(t)], \text{mean}[r(t)] + \text{std}[r(t)]]$ and red area enclose values in $[\text{mean}[r(t)] - \text{std}[r(t)], \text{mean}[r(t)]]$.

The behavior of CIR model is illustrated on the Figures 14 and 15.

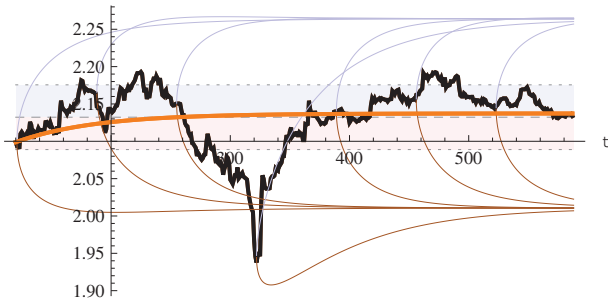


Figure 12: EUR LIBOR 3M data with expected value and upper and lower bounds for Vašiček model.

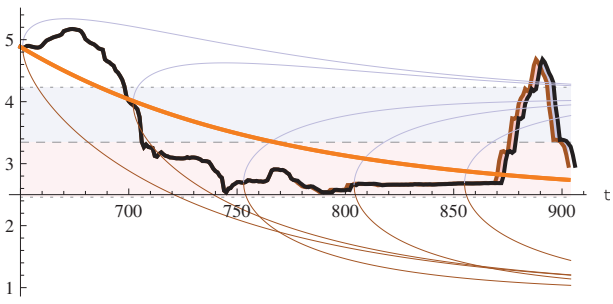


Figure 13: USD LIBOR 3M data with expected value and upper and lower bounds for Vašiček model.

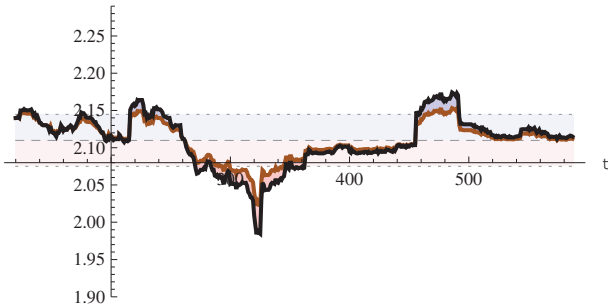


Figure 14: EUR LIBOR 3M data with expected value and upper and lower bounds for CIR model.

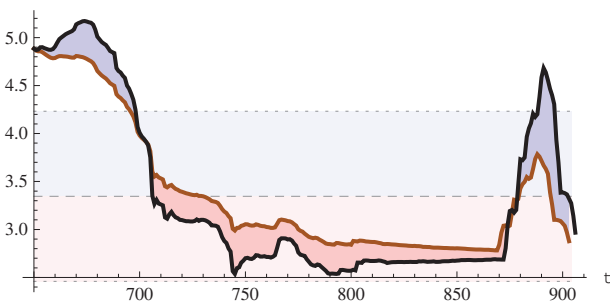


Figure 15: USD LIBOR 3M data with expected value and upper and lower bounds for CIR model.

Acknowledgement

This work was made within project Realization of interactive-informational portal for scientific-technical applications, number 1N04078, financed by Ministry of Education of The Czech Republic and solved in University of West Bohemia in Plzeň, Czech Republic.

References

[1] Data of London Interbank Offered Rate, *British Bankers Association* `http://www.bba.org.uk/bba/jsp/polopoly.jsp?d=141&a=627`.

[2] INVESTOPEDIA: *Advanced bond concepts.* `http://www.investopedia.com/university/advancedbond`

[3] FRIES, O.:*An Equilibrium Characterization of the Term Structure.* Journal of Financial Economics, No. 5, 1977.

[4] HURT, J.:*Yield Curves with Mathematica 6.0.* Proceedings of Wolfram Technology Conference 2007, 2007.

[5] MÁLEK, J.:*Dynamika úrokových měr.* Ekopress, 2005

[6] SHREVE, S. CHALASANI, P., JHA, S.: *Stochastic Calculus and Finance*. Lecture notes, 1997.

[7] Vašíček, O.:*An Equilibrium Characterization of the Term Structure*. Journal of Financial Economics, No. 5, 1977.

Current address

Josef Otta, Ing.
Department of Mathematics, Faculty of Applied Sciences
University of West Bohemia in Pilsen, Univerzitní 22, CZ-300 00 Plzen
tel: +420-377632649
e- mail: jotta@kma.zcu.cz

Journal of Applied Mathematics

volume II (2009), number III

A STRATEGIC PROFIT MODEL TO MEASURE INDIAN APPAREL RETAIL PERFORMANCE

SELVARASU A., (I), FILIPE J. A., (P), FERREIRA M. A. M., (P), PEDRO M. I., (P)

Abstract. The strategic profit model (SPM) has been used to measure the performance of three Indian retail companies that focus in apparel. In the study three companies have been chosen to represent three groups of companies viz., large, medium and small corporate. In the case of large size, medium size and small size corporate, Pantaloon retail is considered with a turnover of Rs.3,031 crores, Shoppers' stop retail with a turnover of Rs.849 crores and Provogue with a turnover of Rs.228 for the year 2007 have been included for the study, respectively. In order to find out the suitable strategies for the companies in the categories, the comparison has been done with its best performance as well as the indicators in relation to peer averages. The ROA has been noticed around 5 percent and the study attempt to reveal the appropriate strategy to increase it by 10 percent. The study has been carried out to indicate suitable strategies to maximize the performance in the future years for Indian apparel retail sector in line with the measure of RONW.

Key words: Retail strategy, competitiveness, Strategic Profit Model

1 Introduction

Indian textile industry is one of the largest sectors in Indian economy in terms of its contribution to foreign exchange earnings and employment. It has been estimated that one out of every six households in India is directly or indirectly dependant on this sector. It contributes about 14 percent to industrial production, 9 percent to excise collections, 18 percent of employment in industrial sector around 16.63 percent to foreign exchange earnings and 4 percent to GDP of India. This sector employs more than 35 million people and is the second largest employer in the country after agriculture. The per capita consumption of cloth in India is Rs. 1,391 for the year 2006.

In India, the retail industry contributes about 10% to national GDP and 8% towards employment. The share of rural market accounted for 55% of total retail market in 2006. India is the fourth largest economy in terms of Purchase Power Parity. It is expected to be the third largest

economy in terms of PPP behind U.S.A. and China by 2010. According to the reports of Images Retail Study 2007 and FICCI Research 2007, Indian retail market caters to be 2nd largest markets in India. The total size of Indian retail market is about Rs 10, 63,800 crore out of which organised retail sector enjoys a small share of around 4.6% which is Rs 48,934.8 crore in value terms. It is estimated that organised retail sector is expected to contribute about 22% to total retail market over Rs 3, 70,500 crore by 2010. As per the recent study on city wise size of retail market, top 6 cities in India accounted for 66% of the total organised retail sales. As per Merrill Lynch India Retail Report March 2007, textile and apparel sector enjoys a major share in Indian organized retail industry which is around 39%.

1.1 Need for the study

Increasing organized apparel retail market results in more competitive situation among Indian companies, which makes top management of the respective companies to think how to curb upon increasing excessive expenditure without disturbing the rising sales for enhancing profit curve. At the same time, they have to keep in mind about the interest of shareholders also. A company performance is measured in terms of their effective allocation of available resources for increasing the value of their shareholders. In other words, measurement of company's Return on Asset (ROA), cash flow and Return on Networth (RONW) reflects financial soundness and overall performance of the company. These acts a indicator upon which top management generally considers for drafting futures strategies. A company's profit performance is determined by the profit earned in relative to the capital invested. In this direction, there is need to study Indian apparel retail sector at three levels spread across small, medium and large companies.

1.2 Objective of the study

In accordance with strategies of the Indian apparel retail companies in the categories of small, medium and large, four objectives have been set as below;

1. To draw the interrelationship of various retail financial indicators of companies for the year 2007
2. To compare the retail financial ratio used for establishing SPM for three years from 2005 to 2007
3. To analyze the deviation of COGS, total expenses and current assets from its peer average for the year 2005(bench mark-the best performance so far) and its peer average for the year 2007
4. To suggest measures to maximize ROA by 10% as target upon ROA for the year 2007 focusing change in sales, COGS, expenses, inventory etc.

2 Review of literature

A company's which keeps in mind the interest of their shareholders and perform accordingly, exercise a huge inflow of capital from the investors. A value created by the company towards its share holders helps decision makers in drafting and implementing certain strategies as it is considered as one of the best metric of performance. To measure value created by the company for

there share holders and stake holders along with performance of the company certain methods are used .They are as Customer Satisfaction and customer value added (CVA), Strategic Profit Model (SPM), Economic value –added (EVA), Profitability Analysis, Total cost analysis, Market value-added (MVA) and Balance Scorecard. The review of studies pertinent to application of SPM model has been presented. Mackay (1992) has studied firm's financial road map which leads to profitability. In this paper, SPM has been applied to study the profit performance of the firm. Koenig, Harold F. (1994), has taken SPM as a example regarding presenting in a easy alternative form using Hypermedia to make marketing class an attractive one. Lambert, Douglas M, Renan, Burduroglo (2000) has discussed the significance of logistics in measuring and selling the value provided to customers along with the impact assessment of customer satisfaction and customer value –added for achievement of higher shareholder value using SPM. Andrew Stapleton, Hanna, Joe B , Steve Yagla,Jay Johnson , Dan Markussen (2002) has applied SPM to six different firm (Nike , Adidas , Fila , Reebok , Converse & K-Swiss) in the footwear industry to provide an insight by offering a predictive ability to the logistics manager of each firm to improve upon ROA using SPM. Evas, R.Joel (2005) has studied the performance of large retailers of U.S. from 1982-2001 using SPM. N. Viswanadham and Poornima Luthra (2005) have studied the performance in terms of measuring and creation of shareholder value of four players in the IT industry using SPM.

3 Research Methodology

Decision without use of any proper technique can result to the adaptation of irrelevant approach to a particular strategy. Thus, a suitable tool of retail finance viz., SPM is used study the indicators affecting company's growth and reputation. Our study is focused on measuring the performance of retail companies focusing on apparel. For this purpose, three companies have been selected representing three different levels as small, medium and large based on the turnover. These companies occupy a unique position in their categories.

In case of large category of companies, Pantaloon India (Retail) Ltd. founded by Mr. Kishoreji Biyani in 1987 has been selected for the study. It is the flagship company of the future group headquartered at Mumbai. In the year 1992, IPO (Initial Public Offering) was made. Its 1st outlet was opened at Kolkata in 1997. It operates with multiple formats i.e. value and lifestyle segments of the Indian consumer market. Company has over 7 million sq. ft. retail spaces comprising of more than 1000 stores across 51 cities providing employment to over 25000 people. Company has been awarded with many distinguished awards and honors. It was awarded as the International Retailer of the Year 2007 by the US-based National Retail Federation (NRF) and the Emerging Market Retailer of the Year 2007 at the World Retail Congress held in Barcelona. Some of its leading retail formats includes Pantaloons, Big Bazar, Central, Fashion station, aLL, Depot, Brand Factory, an online portal futurebazzar.com etc. Future Group's vision is to, "deliver Everything, Everywhere, Every time to Every Indian Consumer in the most profitable manner." The group considers 'Indian-ness' as a core value and its corporate credo is - Rewrite rules, Retain values. Turnover of the company in 2007 was about 3,031.44 crores

In case of medium category of companies, Shopper's Stop Ltd. founded by K.Raheja Corp. Group (L.Charu Chandra Group) in 1991 has been identified for the study. It is headquartered at Mumbai. Shopper's Stop Ltd. is India's largest retail chain of departmental stores having 20 Shoppers' stop and 2 Home Stop covering an area over 1.1 million sq. feet across 11 cities. These stores offer more than 200 different and finest national and international brands. They attract more than 19.95 million shoppers every year with more than 54,000 footfalls everyday. Company made

its IPO (Initial Public Offering) in 2005. It is the only Indian member of the Inter Continental Group of Department Stores (IGDS) along with 29 other experienced retailers from all over the world. Company has been awarded with many distinguished awards and honour. In 2006, it was awarded as "Retailer of the Year – Department Store" at Images India Retail Forum (2006).Turnover of the company in 2007 was about 846.31 crores.

In case of small category, Provogue (India) Ltd. formerly Acme Clothing Private Ltd. founded by Nikhil Chaturvedi and Deep Gupta in 1997 has been included as representative sample of study. It is headquartered at Mumbai. The Company launched the fashion brand 'Provogue' in March 1998. It operates over 100 Provogue stores and shop in shop across 56 cities and Promart stores which offers more than 100 brands in apparel as well as home furnishings and accessories at discounted price round the year. In 2007, first Promart store was opened in Ahmadabad covering an area of around 40,000 sq. feet. Provogue brand was ranked 5^{th} amongst brands in all categories in a national customer loyalty survey conducted by Business World in 2006. Company's vision is to evolve Provogue as retail-centric group of branded business focused on customer needs. Turnover of the company in 2007 was 228.27 crores.

In order to draw the interrelationship of various retail financial indicators of companies for the year 2007, the values have been extracted from company's published income and expenditure statement, balance sheet, schedules of income and expenditure and annual reports for the year 2005-06 and 2006-07. In addition, the retail financial values have been obtained for three years from 2005 to 2007. It is understood from the annual report of the companies in this sector that the performance of all the companies have been reported as exemplary in the year 2005 and hence the values of the same year is considered as benchmark for the study in terms its actual and peer average. The retail financial ratios have been used for establishing SPM for three years from 2005 to 2007 in all three categories. In the pursuit of identifying strategies, the deviation financial values of COGS, total expenses and current assets have been analyzed for all three companies from its peer average for the year 2005 and its peer average for the year 2007. As a measure of prediction to maximize ROA by 10% as target upon ROA for the year 2007 'what if' analysis goal seek option have been used focusing change in sales and its corresponding change in COGS, expenses, inventory etc.

3.1 Strategic Profit Model (SPM)

The strategic profit model is used to study the performance of the Indian apparel retail companies with peer averages as benchmarks. According to Oxford Dictionary benchmarking refers to setting a standard point of reference with which other points can be compared and thus evaluated. Benchmarking is systematic process of searching for innovative, effective and operative ideas and procedures to fuel the growth line improving upon performance capability. It enables an organization in readdressing their strategies through comparison of their key performance with their competitors and other industry leaders. Generally companies take financial performance measures as their key performance. It helps top management in taking right decision at the right time to improve their bottom line. These financial performances have been estimated from balance sheet, Profit and loss a/c and cash flow statements.

Performance of retail companies is measured by the use of Strategic Profit Model (SPM). Strategic profit model is the result of subsequent research formularization of DuPont model, useful system of analysis of which considers important inter-relationships based on the information found in financial statements. It explains how RONW is dependent on other functions factors .i.e. financial leverage, ROA (net profit margin (%) x asset turnover ratio). Financial leverage provides a

relationship between total equity of the firm with the total investment made by the shareholders. Financial leverage measures how effectively a firm is using an outside fund for its operation to increase firm's RONW whereas ROA measures how much profit is being returned by the investment made in the asset.

$$\text{RONW} = \frac{\text{Total asset}}{\text{Net worth}} \times \frac{\text{Net Profit}}{\text{Net sales}} \times \frac{\text{Net sales}}{\text{Total asset}}$$

Or

RONW = Financial Leverage x Net Profit Margin (%) x Asset Turnover ratio

SPM involves the application of ratio analysis method for the calculation of Net profit margin, asset turnover, financial leverage, ROA and RONW for which required data is obtained from financial statement of the firm available publicly. In other words

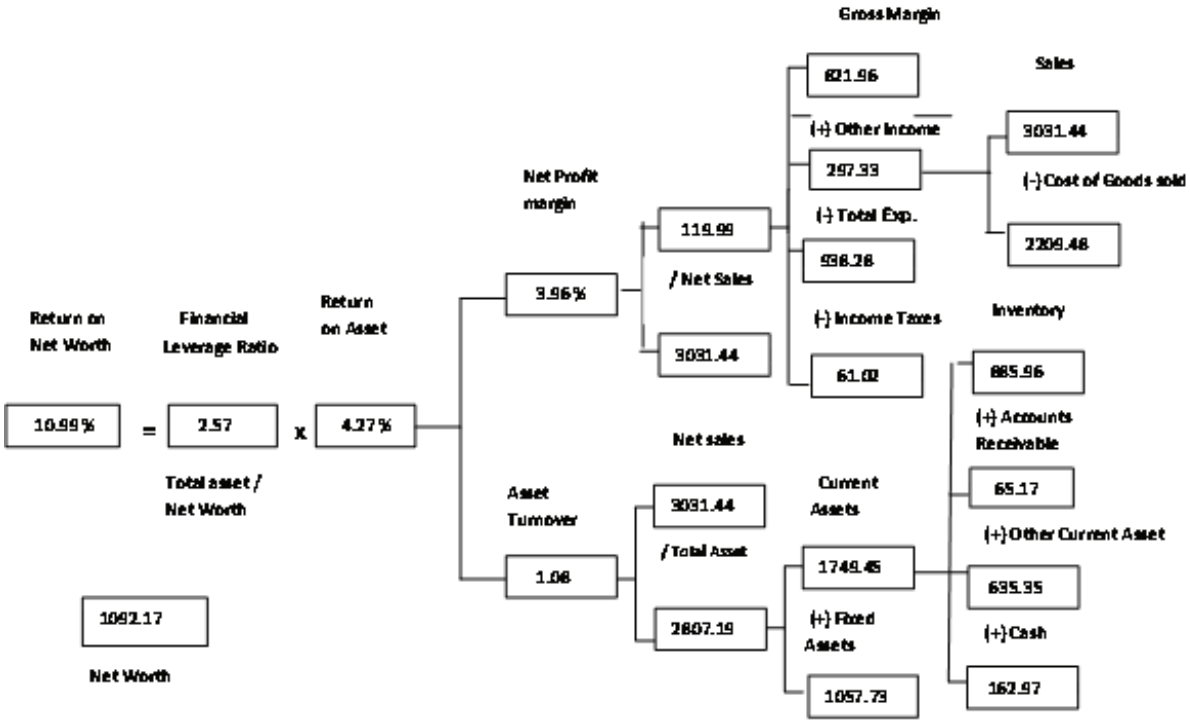


Figure No-1:Pantaloon Retail (India) Ltd. (Rs. Cr.)

Individual firm performance is measured in terms of ROA and RONW by comparing it with other firm's performance.RONW measures how effectively a firm is using shareholders investment. In other terms, it measures the value which is being created by the firm for its shareholder. **Figure 1** provides an outline of the available SPM model which consist of financial leverage and ROA which further consist of two parts .i.e. net profit margin (%) and asset turnover (%). Net profit margin tells at what rate a firm is earning a profit on the sales and also how effectively a company is manufacturing and selling its products and services. Net profit is calculated after deduction total operating expenses and income tax from gross sales which is obtained after deducting cost of goods sold from the net sales. Net sales means sales obtained after the payment of excise duty and other taxes such as Sales tax and VAT. This information has been obtained from the company's Profit and Loss a/c. While asset turnover ratio indicates number of times total asset is being used for

generating sales in a year. It is calculated by dividing net sales by total asset. Further total asset is divided into fixed asset and current asset. Fixed asset is a tangible long lived resources used for producing goods and services whereas current asset is generally converted into cash during the operating cycle of the firm. These information can easily be obtained from the company's balance sheet. SPM is used for comparing firm's performance with the peer average .Thus helping in taking strategic decisions for improving upon required component/ areas to achieve desired ROA and RONW. RONW of the company can be increased by applying certain changes. The strategies are focused in such way to attempt decreasing financial leverage and increasing ROA. Similarly ROA can be increased through implementation of increasing sales and other income sources on one hand and decreasing cost of goods sold, operating expenses and income taxes, inventory, accounts receivable, cash and other current asset on the other.

3.2 Assumptions

The study encompasses various levels of apparel retail in the context of Indian market in which there are value segment and lifestyle segments promoting apparel as separate entity and also as assortments. In the process of extracting data for the study, there are specific methods of adjusting the data have been done. The ways and means of obtaining data from financial statements is altered based on the type of auditing and reporting formats of annual reports. The following are the specific assumptions proposed for all the three companies in Indian apparel business.

1. In case of Pantaloon Retail India Ltd., financial data has been extracted from company's annual report 2005-06 (Pg. No. 86, 87, 93, 94 & 95) and 2006-07 (Pg.No.62, 63, 70 & 71) respectively. The net sales is obtained after adjusting for excise duty, sales tax and VAT and the other operating income is included for obtaining total value of other income. In the case of total expenses, the personal costs, manufacturing & other costs, financial charges and depreciation have been included.
2. In case of Shopper's Stop Ltd., financial data has been extracted from company's annual report 2005-06 (Pg. No. 60, 61, 71 & 72) and 2006-07 (Pg. No. 68, 69, 79 & 80). The net sales has been obtained after adjusting the own merchandise, consignment merchandise for VAT and sales tax. The other income is adjusted for other retail operating income as in the case of large companies in apparel as above. The COGS has been adjusted with the cost of consignment merchandise. The total expenses have been adjusted for employee cost, operating and administrative expenses, interest, financial charge, depreciation and amortisation.
3. In case of Provogue (India) Ltd., financial data has been extracted from company's annual report 2006-07(Pg. No.62, 63, 71 & 72) and 2005-06 (Pg. No. 60, 61, 68 & 69). The net sale has been obtained after adjusting the gross sales for excise duty and sales tax. The other income has been arrived after adjusting export benefits, incentives, gain on foreign exchange fluctuation, royalty income and other income. The cost of goods sold has been adjusted for increase in stock.
4. All the Investments are considered as a long term investment. It is also assumed that income taxes and fixed asset remains that same.
5. Since cost of goods sold, total expenses, account receivable and inventory are independent variable where as sales is a dependent variable, an increase in sales will result in increase in these variables also. So, to calculate effect on these variables, it is assumed that this variable remains constant percentage of the sales through out the period.

6. It is also assumed that all variable will remain constant, when a variable other than sales is calculated to achieve the desired ROA % using what if analysis tool.

7. It is assumed for that total expenses include both variable and fixed expenses. As it is difficult to differentiate among them due to terminology used to define them differs from company to company.

8. Peer average performance is considered as the average industry performance to facilitate research process.

3.3 Limitations

1. General economic conditions, competition, local factors and policy adopted by the management affect financial results of the company or business. Thus, any change in ratio or other financial indicators must be considered with the above factors while taking any decision.
2. Sample size for the study is restricted with three companies in their respective categories .i.e. large, medium and small due to non availability of valid financial data source as there is limited no. of players in Indian organised apparel retail sector.
3. Any interpretation about the financial conditions and performance of the firm is a result of correct diagnosis of the study of combined effect of the various financial indicators in the SPM not alone.
4. Financial indicators analysed in the SPM are only the preliminary steps in interpretation which helps a business in drawing attention towards the area which requires further analysis and investigation.
5. This model does not tell how to increase or decrease various financial factors like COGS, sales, expenses, cash, account receivable etc.
6. Research provides only the suggestions to improve company performance.

4 Analysis and interpretation

The retail financial indicators have been used for the purpose of understanding the strategic focus of the companies in apparel business in India. The retail financial performance of the companies has been compared for the period of three years from 2005 to 2007as intra company comparison. Similarly, the retail financial performance has been compared with its peer average as best performance in the history of their business which is for the year 2005 and the same is also compared for the current year peer average for the year 2007. The results of the retail financial performance have been presented for Pantaloon, Shoppers' stop and Provogue under separate sections.

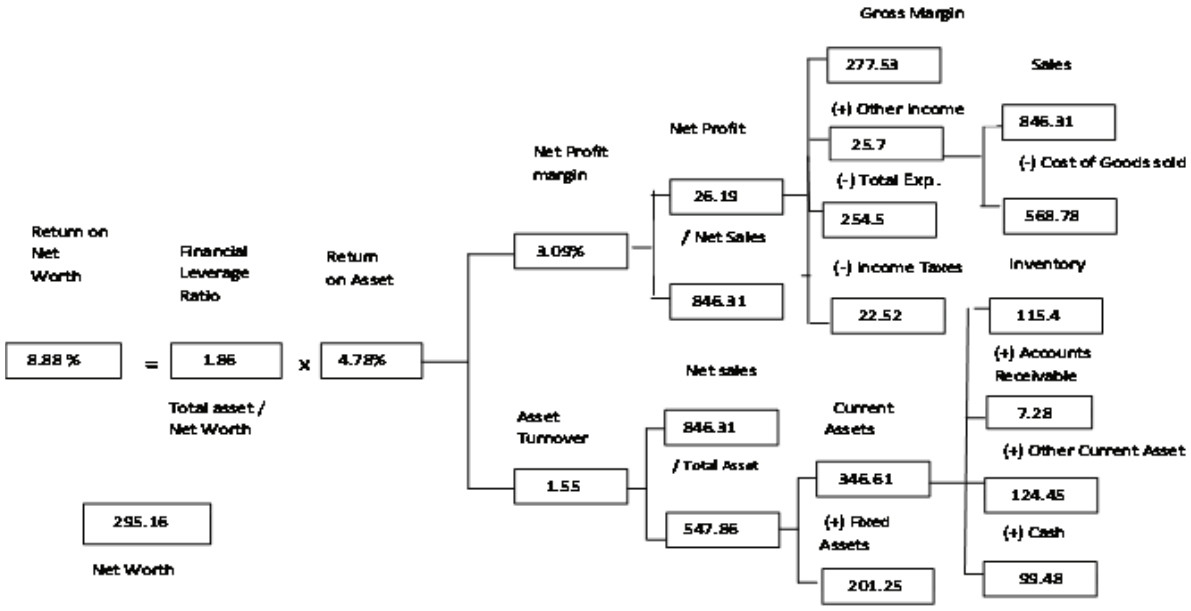


Figure No – 2: Shopper's Stop Ltd. (Rs. Cr.)

4.1 Interrelationship of various financial indicators of Indian apparel companies for the year 2007 (Table No.2)

The financial indicators of the companies in large, medium and small categories have been drawn for Pantaloon (large), shoppers' stop (medium) and Provogue (small). Profit margin of Provogue (8.59%) is much higher in comparison to Pantaloon (3.96%) and Shopper's Stop (3.09%) indicating Provgue management effectiveness to curb upon the manufacturing and other operating expenses to increase their bottom line. But at the same time Provogue is facing a problem in generating enough sales from the investment made in their asset. Asset turnover of Provogue (0.62) which tells that firm is able to generate only 62 paise from the Re 1 investment made in the asset is much below accepted level indicating that firm is facing problems like excess stock, high account receivable due to lenient credit policy , unused cash , idle or improperly used assets etc. But Shopper's Stop is the leader in generating sales from the all the financial resources committed by it indicating high asset turnover ratio (1.55) result of better asset management by the company. Pantaloon is also generating enough sales to cross the accepted level of asset turnover ratio (1.08). As, high profit margin able Provogue to generate enough profit from the investment made in the total asset. Lowering the effect of their low asset turnover ratio which acts as a retarding force for accelerating ROA. Provogue is the leader in earning profit from their asset .i.e. ROA (5.33%) followed by Shopper's Stop (4.78%) whose better sales generation from the asset acts as an energizer for strengthening profit margin. Thus ROA while on the other side Pantaloon in spite of having decent profit margin is not able to generate better sales from their asset all though above accepted level resulting in lowering ROA (4.27%) . Higher leverage ratio indicates that more asset generation or acquisition through debt .Company's aggressiveness in utilizing debt for the generating asset.. Pantaloon with their aggressive policy of generating or acquiring more asset from the debt lead to have highest leverage ratio (2.57) followed by Shopper's Stop (1.86) and Provogue (1.40). It tells that big, medium and small company are generating or acquiring Rs 2.57 , Rs 1.80 and Rs. 1.40 respectively for every Rs.1in the net worth. This further affects their RONW.

Pantaloon with high leverage ratio and low ROA is the leader company in providing maximum value to their share holders. It also tells about Pantaloon's management effectiveness regarding the use of resources by the owners for maximizing shareholders and owner's welfare. But Provogue in spite of having high ROA is not able to achieve high RONW due to its low leverage ratio. Shopper's Stop RONW is the second highest among the sample due to its decent ROA and consistent policy of generating or acquiring asset through debt for viable use. Thus for the large, medium and small companies, RONW is 10.99%, 8.88% and 7.46%, respectively. Further figure 2, would provide further details of financial indicators performance for 2007.

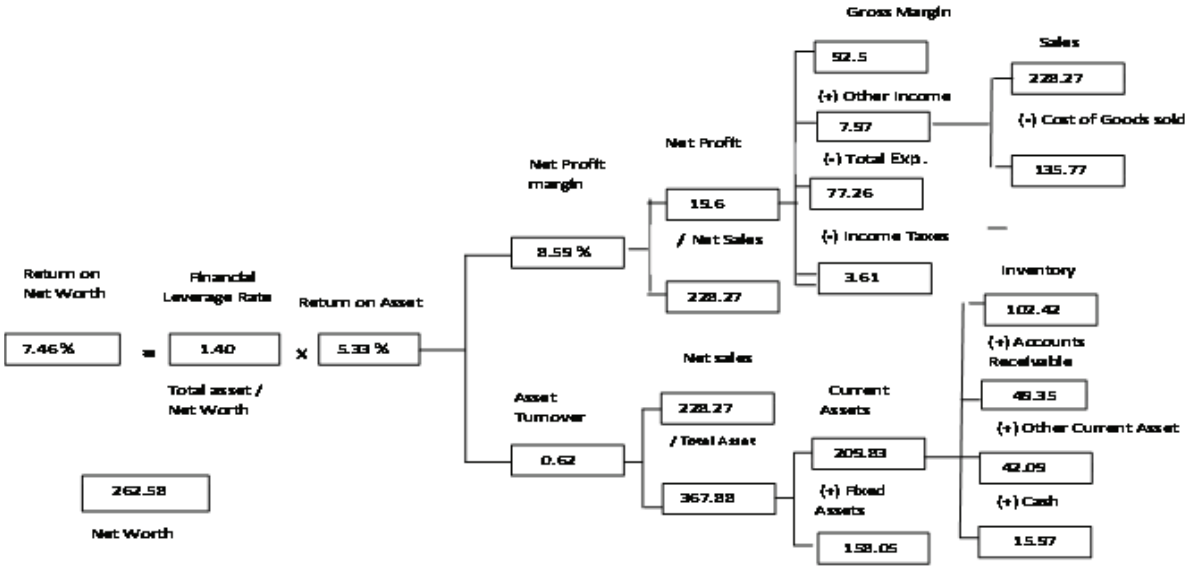


Figure No – 3: Provogue (India) Ltd. (Rs. Cr)

4.2 Comparing the financial ratio used for establishing SPM for three years from 2005 to 2007(Table No.1)

Financial ratio is used as an index and yardstick for evaluating the financial position and performance of the firm. It helps in analysis for making qualitative judgments about the company's financial position and performance. Financial ratios which have been used for establishing SPM are net profit margin (%), asset turnover ratio, return on asset (ROA), financial leverage and return on net worth (RONW). The description of retail financial indicators has been presented for Pantaloon, shoppers' stop and Provogue separately.

4.2.1 Pantaloon Retail Ltd.

Net Profit margin (%) has been decreased by 3.63% in 2006 along with low generation of sales from the asset as a result of which asset turnover ratio is lowered by 18.02%. as a combined effect of these indicators is seen on ROA and financial leverage which also is further decreased by 21.00%. and 11.43% respectively. Since RONW is dependent on ROA and financial leverage, it is also decreased by 30.02%. All these changes indicates that company's cost of goods sold and other operating expenses has been increased rapidly in proportion to sales .Pantaloon has idle or

improperly used assets which increases the need for costly financing and expenses for maintenance and up keeping .

In 2007, net profit margin shows a positive sign with an increment by 8.49% .But Pantaloon is not able to sustain the effect of this change on other factors also. Asset turnover ratio is further lowered by 13.93 %. Increment in net profit margin is over ridden by lowering asset turnover ratio as a result of which ROA is further lowered by 6.62% along with financial leverage by 3.38%, thus lowering RONW by 9.78%. As company's RONW is falling that indicates that company is borrowing more money for the expansion of their retail operations followed by the under utilization of it in generating asset. Company should improve RONW by getting rid of excessive and idle asset. Thus investing that fund for other purposes like implementation of cost reduction techniques, expansion, paying advances to suppliers to avail maximum purchase benefit etc.

4.2.2 Shopper's Stop Ltd.

In 2006, net profit margin increased by 9.65% but asset turnover ratio showed a downward trend lowering it by 22.1% .As a result of which ROA is declined by 14.59%. Financial leverage is also declined by 41.45% indicating that company is not generating enough amounts of assets from the money borrowed by them which in term reduces their RONW by 50%.

In 2007, net profit margin declined by 27.52% indicating an increase in COGS and other operating expenses. On the other hand, Shopper's Stop management gave importance to management of their asset which resulted in a marginal increment of asset turnover ratio by 0.52% .But this marginal increment is sufficient enough to increase ROA. Resulting in a decline of 27.12% in ROA indicating that profit generated from the investment made in asset is reduced to 4.78 from 6.56. Financial leverage increased by 21.27 % pointing towards proper management of the investment made in the asset .It is not enough to increase RONW which registered a negative growth of 11.62%. Company should improve RONW by improving upon their operating expenses such as selling, distribution etc which acts as a barrier in facilitating net profit growth and avoid holding unnecessary cash balance.

4.2.3 Provogue India Ltd.

A tremendous increment .i.e.22.22 % in net profit margin in 2006 marked Provogue followed by a tremendous decline in asset turnover ratio by 33.81% indicating about ineffectiveness of the management in handling asset. Company might be facing problems like high stock carrying, slow debt collection, high capital blocked in fixed asset etc. which acts a dominating force further affecting ROA and financial leverage by 18.85% and 18.06% respectively effecting RONW which registered a negative growth by 33.31%.

Year 2007 embarks net profit margin with an increment of 7.24% indicating company's management effectiveness in curbing upon COGS and other operating expenses But Company's ineffectiveness regarding managing their asset put increment of net profit margin in vain which lowers asset turnover ratio by 20.31% which further affected ROA and financial leverage ratio. Both registered a negative growth of 14.54% and 20.42% respectively which affected RONW by 31.99% adversely. Company should improve RONW by effectively managing their assets by avoiding high stocks, slow debt collection, not availing unlimited credit, blocking capital in fixed and idle or unused asset

4.3 To analyze the deviation of COGS, total expenses and current assets from its peer average for the year 2005(bench mark-the best performance so far) and its peer average for the year 2007

In every business, the concept of increasing return on asset (ROA) can be achieved by improving upon two factors .i.e. profit maximization (increasing sales and sustaining expenses) and reduction in unutilized asset. In year 2005, ROA performance of Pantaloon, Shopper's Stop and Provogue was a bench mark performance so far. Peer group average performance for year 2005 is far than year 2007. There is drastic increase in expenses, followed by other income sources and assets embarking up rising top line figures from 2005 to 2007. Industry average(peer average) for items/sales has significantly increased by .i.e. COGS (3.35%) , other income (157.5%) , expenses (5.55%) , account receivable (37.5%) , inventory (17.50%) , cash (346.50) , other current asset (19.35%) whereas sales (155%) . In spite of increase in sales, companies are not able to reach or cross the bench mark level performance. This deviation plays a major role in determining the strategies for enhancing the performance of various financial indicators to achieve the desired level of performance. Each sample has been examined separately. Here each item of individual sample is expressed as percentage of sales and then compared for year 2005 & 2007. To study deviation of items .i.e. COGS, expenses and current asset between bench mark year (2005) and current year (2007) difference between company's individual item/sales(%)of year 2005 v/s 2007 and peer average of 2005 v/s 2007

4.3.1 Pantaloon Retail Ltd.

Pantaloon sales have increased by 182.1% but other expenses and current asset rises unproportional. From year 2005 to 2007 changes in items/sales(%) are as follows :- COGS by 5.62% ; other income by 166.30% ; expenses by 5.56% ; inventory by 7.86% ; account receivable by 77.69% ; cash by 176.30% ;and other current asset by 126.60%. At the same time when these items are compared among the peer average of year 2005 and 2007 certain difference in items has been noticed. They are as follows: COGS by 2.48%; other income by 2.67% ; expenses by -0.65% ; account receivable by 40.36% ; inventory by -7.53% ; cash by 44.42% ;and other current by -3.07%. Here increment in any item is referred to an increase in that item in respect to share of difference between actual item/sales (%) v/s peer average value of current year and actual item/sales (%) v/s peer average value of previous year

4.3.2 Shopper's Stop Ltd.

Shopper's Stop sales have increased by 73.18%. From year 2005 to 2007 changes in items/sales(%) are as follows :- COGS by -4% ; other income by 89.44% ; expenses by 9.94% ; inventory by 16.08% ; account receivable by 62.26% ; cash by 5,775% ;and other current asset by 8.72%. At the same time when these items are compared among the peer average of year 2005 and 2007 certain difference in items has been noticed. They are as follows: COGS by -7.5% ; other income by -69.86% ; expenses by 4.27% ; account receivable by 62.18% ; inventory by -2.32% ; cash by 702.21% ;and other current by 14.73%.

4.3.3 Provogue (India) Ltd.

Provogue sales have increased by 107.12%. From year 2005 to 2007 changes in items/sales(%) are as follows :- COGS by -4.16% ; other income by -32.84% ; expenses by -1.08% ; inventory by 35% ; account receivable by 19.44% ; cash by 278.37% ;and other current asset by 6.40%. At the same time when these items are compared among the peer average of year 2005 and 2007 certain difference in items has been noticed. They are as follows: COGS by -8.5% ; other income by -190.34% ; expenses by 5.66% ; account receivable by -1.8% ; inventory by 8.94% ; cash by 14.84%; and other current by -41.91%.

5. Recommendation to increase ROA 10%

5.1 Pantaloon Retail India Ltd.

Basic strategies that Pantaloon's management can apply are increase sales, increase other income sources, reduce cost of goods sold, reduce expenses, reduce inventory, reduce account receivable and/or reduce cash and other current asset. It is assumed that they desire to increase ROA by 10% keeping parallel to targeted growth of India's GDP. SPM model has been applied independently to various factors like sales; COGS, etc. study the various changes independently. To achieve a 10% increase in the ROA, Pantaloon management would have to increase sales by 12.88%.To achieve this sale , company would have to corresponding increase its cost of goods sold (COGS) ; expenses ; inventory ; other income sources ; account receivable ; cash and other current asset. As a result of which net profit remains the same time asset turn over ratio increase helping ROA to increase by 10%.

Next strategy is to increase other income sources, assuming all remain same .Other income sources have to be increase by Rs 11.95 cr (4.02%). As company's other income sources are above of peer average. It is not possible to increase it. It might divert the attention of the firm from its regular business course. It is only a short term gain followed by long term diversion of business.

Another strategy is to reduce cost of goods sold (COGS) and expenses, assuming all remains the same. Then COGS is to be reduced by Rs.11.94 cr (0.54%) to achieve an increase in ROA by 10%. Likewise, if an expense is reduced by Rs.11.95 cr (1.27%) same result can be achieved. Reducing COGS and expenses simultaneously by Rs 11.95 cr would be another better way to increase ROA by 10% .Reduction in COGS and expenses would increase the profit margin resulting in an increase in ROA.

Reducing current asset .i.e. inventory, cash, account receivable and other current asset is the last set of strategy to be applied to achieve the desire ROA .Assuming all remains same. Each asset is to examined separately in order to achieve an desired ROA. Then inventory has to be reduced by Rs 254.20 cr (28.69%) which might result in a loss in sale due to stock outs. Another ways to reduce cash by Rs 71.30 cr (43.75%) .As Company's cash balance is well below the peer average indicates that company is follows a policy of holding less cash. Then it is advisable not to decrease cash balance. It might increase the risk of crisis of short term capital for meeting day-to-obligations. Reducing account receivable by Rs.254.31cr. (389.95%) indicating that company has to follow a strict policy regarding giving credit to its customers. This appears to be an unrealistic scenario as Pantaloon's account receivable is already below than peer average. It indicates that max. sale of the company is in cash. In the last desired ROA can be achieved by reducing other current asset by Rs254.11 cr (40%). This change is substantial which may result in loss in purchasing goods at a low

and attractive price from its suppliers by making advance payment along with attracting world famous brands to its counter who generally ask for some type of special deposits before providing goods.

It appears best and suitable policy for Pantaloon to increase ROA by 10% lies in the combination of increasing sales and decrease its cost of goods sold expenses. As discussed above decrease in inventory, cash, account receivable and other current asset is affecting firm's performance adversely in spite of increasing ROA. Company's success lies in increasing sales and decreasing expenses. So, the model suggests that a modest increase in sales by 1.5% followed by a decrease in COGS by 0.5% is sufficient enough to increases ROA by 10%. As these changes are easily to achieve without any special effort.

5.2 Shopper's Stop Ltd.

Using SPM, financial data are analysed. These data forms the base on the basis of which financial indicators are identified and calculated. Among the samples Shopper's Stop is found to the best in the area of managing asset effectively .In order to achieve an increase in ROA by 10% , Shopper's Stop management can apply following strategies .i.e. increase sales & other income sources , decrease expenses .i.e. COGS and expenses and / or finally reduce asset .i.e. inventory, account receivable , cash and other current assets. Other current assets include loans and advances also.

If a sale of Shopper's Stop is increased by Rs.71.76 cr (8.46%), then ROA can be increased by 10%. To achieve this sale company correspondingly has to increase its other income, COGS, expenses, account receivable, inventory, cash and other current asset. As a result of which profit margin would remain almost the same while asset turn over ratio increases. Thus acts as a catalyst in achieving desired ROA. Increasing sales in a short span of time appeared to be a little tough exercise. It requires a handsome amt. of effort to push sales up equivalent to India's GDP growth rate (about 10%). Another strategy is to increase other income sources, keeping all things constant. Then it requires an increment of Rs.2.61 cr (10.12%) which is below the peer average. But increasing this might result in the deviation from the vision and mission of the company.

A successful company is one which is able to curb upon the expenses. It is considered to be one of the foremost strategies. Assuming all things remain the same, Shopper's Stop management can cut COGS by Rs2.6 cr (0.46%) to achieve the desired ROA. As, company's COGS is far below the peer average giving an indication that the further reduction in COGS is not possible. Another method is to reduce expenses by Rs.2.61 cr (1.03%). It would be more realistic scenario when combination of both COGS and expenses is applied simultaneously to reduce cost by Rs 2.6cr to achieve desired ROA.

Another similar approach is to decrease current asset .i.e. inventory, account receivable, cash and other current asset. Assuming all things remain the same, inventory is to be reduced by Rs.49.53cr (42.92%). This reduction might result in loss of sale due to stock out. Shopper's Stop inventory levels are far below peer average of inventory to sales .So, it is not favourable for the company to reduce stock. Another way is to reduce account receivable level by Rs49.56 cr (680.77%) but it is not suggested to do so .Already Company's account receivable level is much better than industry peer average. So, further reduction is not possible. It indicates that company is following a controllable credit policy. Cash plays an important role in carrying out day-to-day business transaction. To achieve the desired level of ROA, cash level can be reduced by Rs49.51cr (49.77%). Company's cash to sales ratio is greater than industry peer average indicating company's

ineffectiveness in the management of cash. Suddenly reducing cash to such extent might result in cash shortage. It can be reduce to a certain extent only. The final approach is to reduce other current asset by Rs. 49.41 (39.81%). As Shopper's stop other current asset to sales ratio is much better than industry peer average then further reduction can effect the growth of the company.

Shoppers Stop management should look after those are area which is under performing relative to the peer average. So combination of certain factors would help in achieving the increment in ROA with an ease. It is recommended that to increase sales by 2% along with reduction in expenses by 0.5% and cash by 11%.

5.3 Provogue India Ltd.

Among the samples, Provogue India Ltd. is first in terms of ROA but last in terms of RONW. So, to increase its ROA by 10% company's management can increase the sales by Rs153.71 cr (67.34%). To get this sale company correspondingly has to increase its other income, COGS, expenses, account receivable, inventory, cash and other current asset. Nevertheless to get this much sale in short span is not easy to achieve.

Another strategy that Provogue management can apply is to increase other income sources by Rs 1.96 (24.59%) by keeping all things remain same. It is not advisable to increase other income source because it is an uncertain source of income. It is only for a short term.

On of the most secret weapon which is often generally used by management to increase ROA is to reduce expenses and increase profit margin. Another approach is to reduce COGS by Rs 1.99 cr (1.44%) keeping all things same. Further reduction in COGS is further not possible to achieve. Since Provogue's COGS is much lower than industry peer average indicating management effectiveness in purchasing / manufacturing goods at a lower price but selling than at a high rate. Similarly reducing expenses by Rs2.56 cr (3.29%) keeping all thing remain the same. As operating expenses is more than industry peer average which acts as a slow poison in eating away the advantage of low COGS. In a realistic situation it is possible to curb upon excessive operating expenses to neutralise the effect of slow poison and acts as a lifesaver.

Another approach is to reduce current asset. Each component of current asset is examined keeping all thing remain same. Reducing inventory by Rs 33.41cr (32.62%) to achieve the desired increment in ROA .i.e. 10%. Company is having excess of stock with them. This is supported by the fact that its inventory peer average is much more than industry peer average. So, it is advisable to decrease stock which would result in increase in asset turn over ratio. On other hand account receivable can be reduced by Rs.33.41cr (74.38%). As company's account receivable to sales is much above the industry level indicating towards company's lenient credit policy. Company is adopting a policy providing credit to generate sales. It is not possible to reduce account receivable to such a extent since it can result in loss of sale due to tight credit policy. But an optimum level of reduction is possible. Another approach is to reduce cash by Rs 33.42 cr (209.08%) to achieve the increase in ROA by 10%. It is not possible to decrease cash to such an extent as company cash level is almost equivalent to industry peer average. Reducing cash can adversely affect the day-to-day working of the company. Final approach is to reduction of other current asset by Rs.33.42 (79.39%) to increase ROA by 10%. It is not a realistic situation to reduce other current asset as it is already below industry norms.

As Provogue management should look toward reducing unused and unutilised asset. So, a combination of change can be applied to modestly increase sales by 2% tripled with decrease in expenses by 2% ; inventory also by 2% and account receivable by 2% to increase ROA by 10%. It

is further advised to Provogue management that further increment in ROA can be achieved by working upon the under performance area in comparison to industry peer average.

References

[1.] N.VISWANADHAM and POORNIMA Luthra (2005), Models for measuring and predicting shareholder value: A study of Third Party Software Service Providers, Sadhana, Vol.30, Part 2 & 3, April/June 2005, Pg. No. 475- 498.

[2.] EVANS, R.JOEL (2005), Are the largest public retailers top financial performers? A longitudinal analysis, International Journal of Retail & Distribution Management, Vol.33 Issue 11, Pg. No. 842-857.

[3.] Andrew STAPLETON, HANNA, Joe B, STEVE YAGLA, Jay Johnson, Dan Markussen (2002), International Journal of Logistics Management, Vol.13 Issue 1, Pg. No. 89-107.

[4.] LAMBERT, Douglas M, RENAN, Burduroglo (2000), International Journal of Logistics Management, Vol.11 Issue 1, Pg. No. 1- 16.

[5.] KOENIG , F. Harold (1994) , Marketing Education Review , Vol.4 Issue 1 , Pg. No. 36-44

[6.] MACKAY J.R. (1992), Does your financial road map lead to profitability? , Textile Rental, Vol. 75 Issue 7, Pg. 22.

[7.] BHABATOSH Banerjee, *Financial Policy & Management Accounting,* 7th Edition, Pg.No. 330 – 360.

[8.] Leopold A. BERNSTEIN, *Financial Statement Analysis Theory, Application & Interpretation,* Pg.No 562,563.

[9.] Patrick M.DUNNE, Robert F.CUSCH, David A.GRIFFITH, *Retailing,* Fourth Edition, Pg.No 37 – 67.

[10.] TEXSUMMIT 2007, Background Paper, Ministry of Textiles, Govt. of India .p.1.

Table No – 2. Intra Company Ratio Comparison

Pantaloon Retail Ltd.						
Year	Sales (Cr.)	Profit Margin (%)	Asset Turnover	ROA(%)	Financial Leverage	RONW (%)
2007 (Latest Performance)	3031.44	3.96	1.08	4.27	2.57	10.99
2006	1758.5	3.65	1.25	4.58	2.66	12.18
2005 (Max.Performance)	1018.25	3.79	1.53	5.79	3.00	17.40

Shoppers' Stop Ltd.						
Year	Sales (Cr.)	Profit Margin (%)	Asset Turnover	ROA(%)	Financial Leverage	RONW (%)
2007 (Latest Performance)	846.31	3.09	1.55	4.78	1.86	8.88
2006	634.93	4.27	1.54	6.56	1.53	10.04
2005 (Max.Performance)	488.69	3.89	1.97	7.69	2.61	20.08

Provogue India Ltd.						
Year	Sales (Cr.)	Profit Margin (%)	Asset Turnover	ROA(%)	Financial Leverage	RONW (%)
2007 (Latest Performance)	228.27	8.59	0.62	5.33	1.40	7.46
2006	149.25	8.01	0.78	6.23	1.76	10.98
2005 (Max.Performance)	110.21	6.55	1.17	7.68	2.15	16.51

Table No-1						
Comparison of change in various financial indicators of different companies						
Year	Company	Profit Margin (%)	Asset Turnover	ROA (%)	Financial Leverage	RONW
2006-07	Pantaloon Retail Ltd.	8.49	-13.93	-6.62	-3.38	-9.78
	Shopper's Stop Ltd.	-27.52	0.52	-27.12	21.27	-11.62
	Provogue India Ltd.	7.24	-20.31	-14.54	-20.42	-31.99
2005-06	Pantaloon Retail Ltd.	-3.63	-18.02	-21.00	-11.43	-30.02
	Shopper's Stop Ltd.	9.65	-22.11	-14.59	-41.45	-50.50
	Provogue India Ltd.	22.22	-33.61	-18.85	-18.06	-33.55

Table No.3

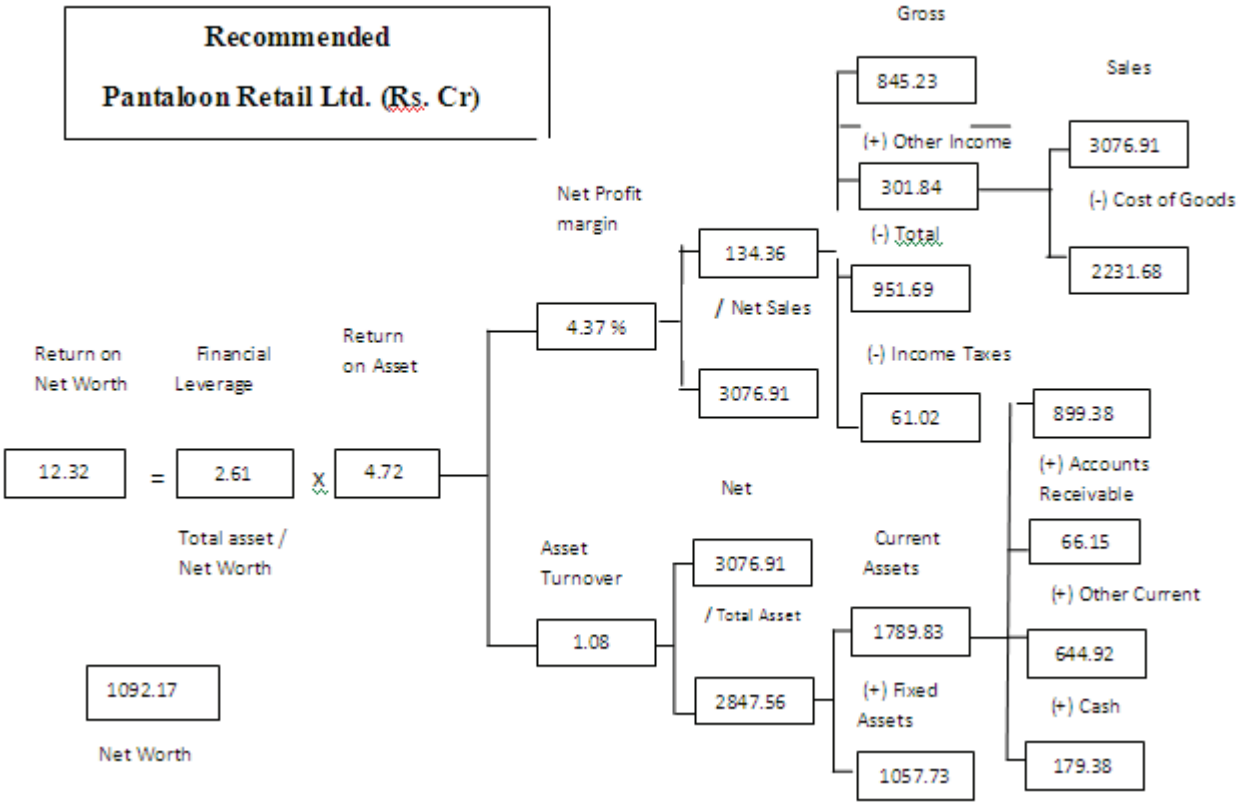
Target Firm (07) Comparison with Peer Group (07) – Items as a % of sales

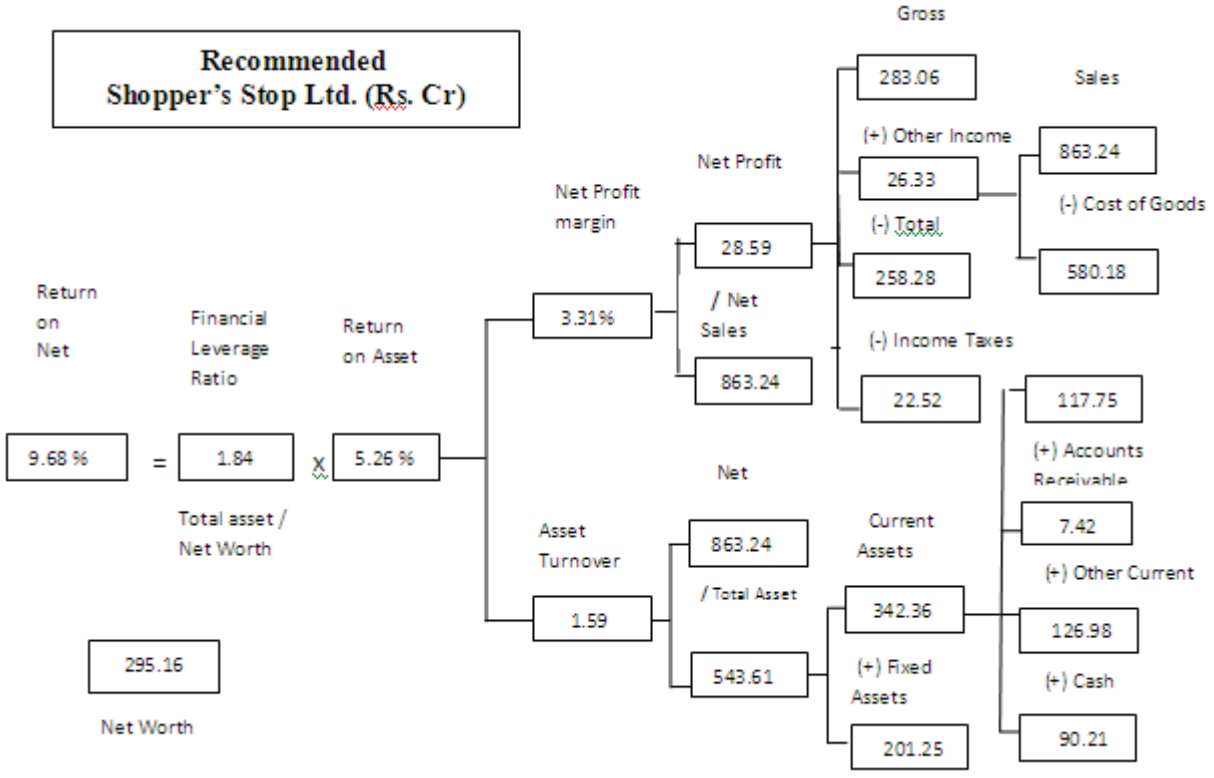
		COGS	Other Income	Expenses	Account Receivable	Inventory	Cash	Other Current Asset
Peer Average (07)		**70.97%**	**8.06%**	**30.93%**	**2.97%**	**26.88%**	**6.79%**	**19.53%**
Pantaloon Retail (India) Ltd.	Actual(07)	72.89%	9.81%	30.95%	2.15%	29.23%	5.83%	20.96%
	v/s							
	Peer (07)	1.92%	1.75%	0.02%	-0.82%	2.35%	-0.96%	1.43%
Shopper's Stop Ltd.	Actual(07)	67.21%	3.05%	30.08%	0.86%	13.64%	11.75%	14.71%
	v/s							
	Peer (07)	-3.76%	-5.01%	-0.85%	-2.11%	-13.24%	4.96%	-4.82%
Provogue (India) Ltd.	Actual(07)	59.48%	3.13%	33.85%	21.62%	44.87%	7.00%	18.44%
	v/s							
	Peer (07)	-11.49%	-4.93%	2.95%	18.65%	17.99%	0.21%	-1.09%

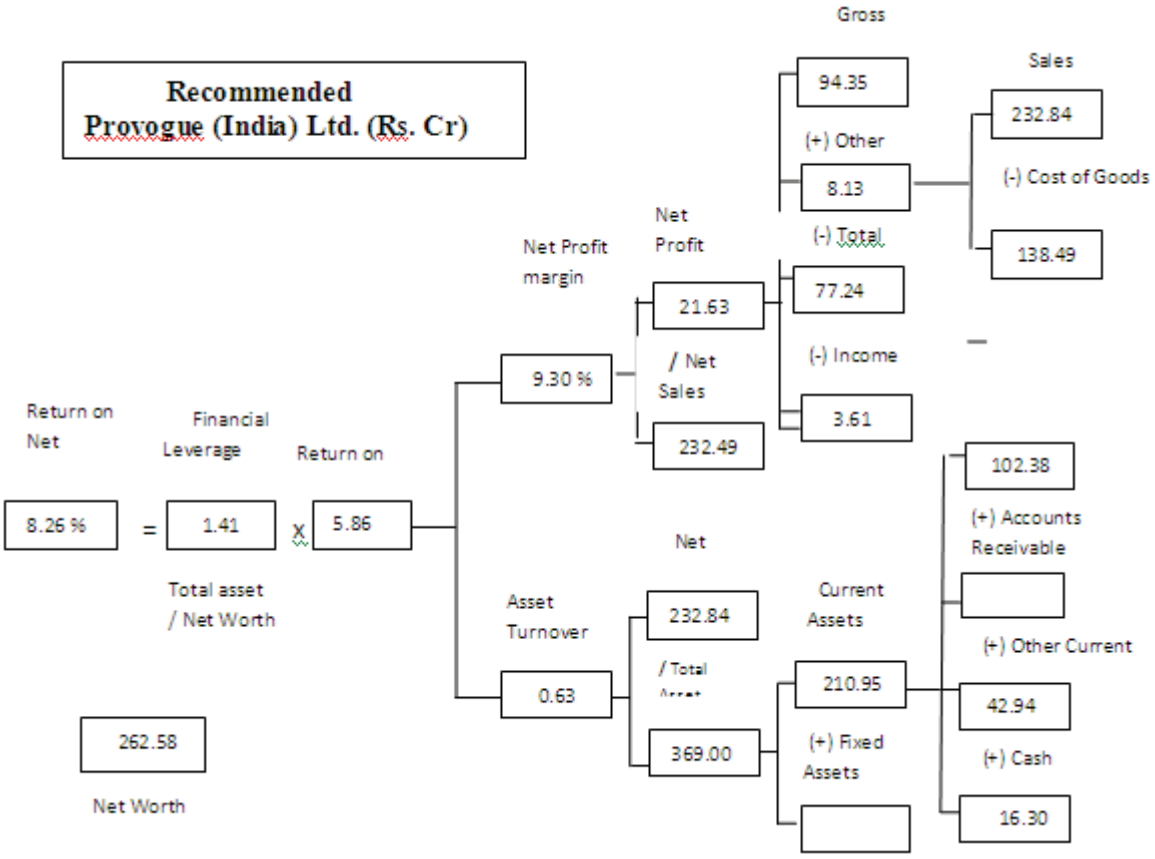
Table No. 4

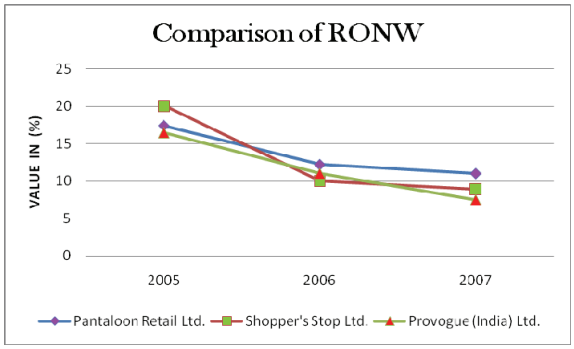
Target Firm (05) Comparison with Peer Group (05) – Items as a % of sales

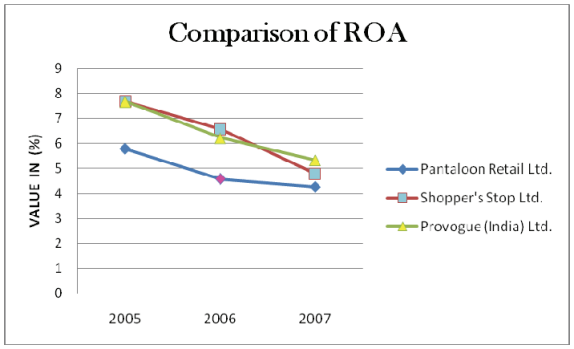
		COGS	Other Income	Expenses	Account Receivable	Inventory	Cash	Other Current Asset
Peer Average (05)		**68.68%**	**3.13%**	**29.30%**	**2.16%**	**22.88%**	**1.52%**	**11.09%**
Pantaloon Retail (India) Ltd.	Actual(05)	68.78%	3.69%	29.70%	1.21%	27.10%	2.11%	9.25%
	v/s							
	Peer (05)	0.10%	0.56%	0.40%	-0.95%	4.22%	0.59%	1.84%
Shopper's Stop Ltd.	Actual(05)	70.01%	1.61%	27.36%	0.53%	11.75%	0.20%	13.53%
	v/s							
	Peer (05)	1.33%	-1.52%	-1.94%	-1.63%	-11.13	-1.32%	2.44%
Provogue (India) Ltd.	Actual(05)	61.95%	4.66%	34.22%	18.10%	33.23%	1.85%	17.33%
	v/s							
	Peer (05)	-6.73%	1.53%	4.92%	15.94%	10.35%	0.33%	6.24%

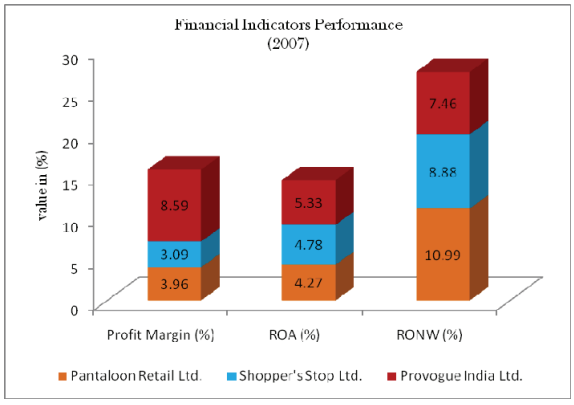

Recommended
Pantaloon Retail Ltd. (Rs. Cr)
Return on Net Worth
Financial Leverage
Return on Asset
12.32
=
2.61
X
4.72
Total asset / Net Worth
1092.17
Net Worth
Net Profit margin
4.37 %
Asset Turnover
1.08
134.36
/ Net Sales
3076.91
Net
3076.91
/ Total Asset
2847.56
Gross
845.23
(+) Other Income
301.84
(-) Total
951.69
(-) Income Taxes
61.02
Sales
3076.91
(-) Cost of Goods
2231.68
Current Assets
1789.83
(+) Fixed Assets
1057.73
899.38
(+) Accounts Receivable
66.15
(+) Other Current
644.92
(+) Cash
179.38

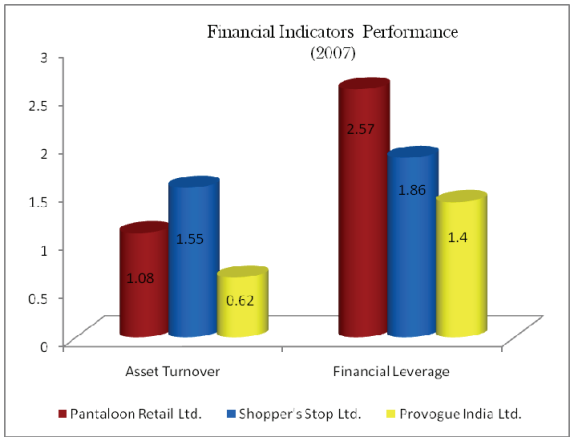

Recommended
Shopper's Stop Ltd. (Rs. Cr)
Return on Net
Financial Leverage Ratio
Return on Asset
9.68 %
=
1.84
X
5.26 %
Total asset / Net Worth
295.16
Net Worth
Net Profit margin
3.31%
Asset Turnover
1.59
Net Profit
28.59
/ Net Sales
863.24
Net
863.24
/ Total Asset
543.61
Gross
283.06
(+) Other Income
26.33
(-) Total
258.28
(-) Income Taxes
22.52
Sales
863.24
(-) Cost of Goods
580.18
Current Assets
342.36
(+) Fixed Assets
201.25
117.75
(+) Accounts Receivable
7.42
(+) Other Current
126.98
(+) Cash
90.21

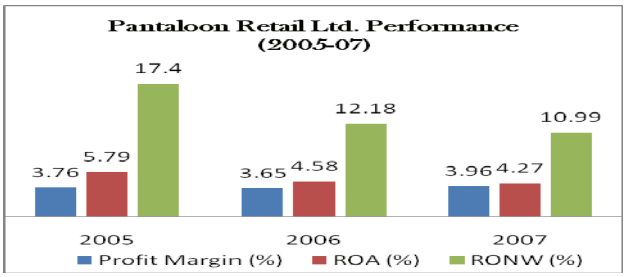

Recommended
Provogue (India) Ltd. (Rs. Cr)
Return on Net
Financial Leverage
Return on
8.26 %
=
1.41
X
5.86
Total asset / Net Worth
Net Profit margin
9.30 %
Asset Turnover
0.63
Net Profit
21.63
/ Net Sales
232.49
Net
232.84
/ Total
369.00
Gross
94.35
(+) Other
8.13
(-) Total
77.24
(-) Income
3.61
Sales
232.84
(-) Cost of Goods
138.49
Current Assets
210.95
(+) Fixed Assets
102.38
(+) Accounts Receivable
(+) Other Current
42.94
(+) Cash
16.30
262.58
Net Worth

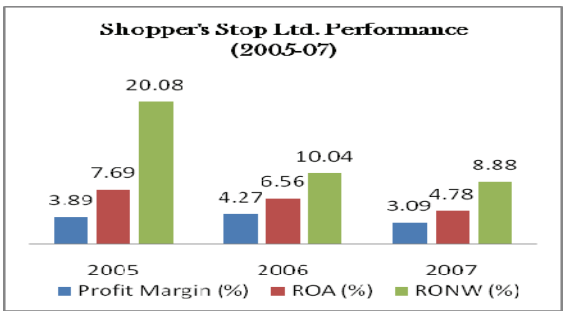

Comparison of RONW
VALUE IN (%)
25
20
15
10
5
0
2005
2006
2007
Pantaloon Retail Ltd.
Shopper's Stop Ltd.
Provogue (India) Ltd.

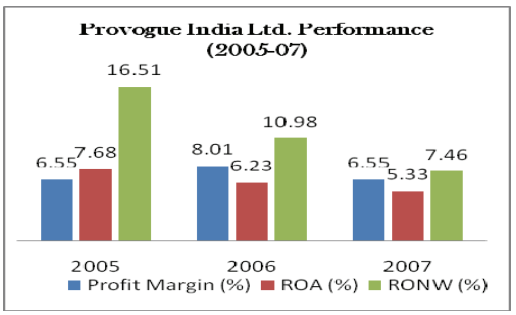

Comparison of ROA
VALUE IN (%)
9
8
7
6
5
4
3
2
1
0
2005
2006
2007
Pantaloon Retail Ltd.
Shopper's Stop Ltd.
Provogue (India) Ltd.


Financial Indicators Performance
(2007)
value in (%)
30
25
20
15
10
5
0
8.59
3.09
3.96
5.33
4.78
4.27
7.46
8.88
10.99
Profit Margin (%)
ROA (%)
RONW (%)
Pantaloon Retail Ltd.
Shopper's Stop Ltd.
Provogue India Ltd.


Financial Indicators Performance
(2007)
3
2.5
2
1.5
1
0.5
0
1.08
1.55
0.62
2.57
1.86
1.4
Asset Turnover
Financial Leverage
Pantaloon Retail Ltd.
Shopper's Stop Ltd.
Provogue India Ltd.







Current address

A. Selvarasu,
Professor of Marketing in Sardar Vallabhbhai Patel Institute of Textile Management
(Autonomous Institute, Government of India, Ministry of Textiles), Coimbatore, India.
e-mail: jeemanse@rediffmail.com

José António Candeias Bonito Filipe, Professor Auxiliar
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel.+351 217 903 000
e-mail: jose.filipe@iscte.pt

Manuel Alberto Martins Ferreira, Professor Catedrático
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel.+351 217 903 000
e-mail: manuel.ferreira@iscte.pt

Maria Isabel Craveiro Pedro, Professor Auxiliar
IST - Instituto Superior Técnico
CEGIST – Centro de Estudos Gestão
Av. Rovisco Pais 1100-099 Lisboa (Lisbon, Portugal), Tel.+351 21 423 35 07
e-mail: ipedro@ist.utl.pt

Journal of Applied Mathematics

volume II (2009), number III

THE RISK OF A SMALL CURRENCY PORTFOLIO
—
BACKTESTING RESULTS BY COPULA APPROACH

TICHÝ Tomáš, (CZ)

Abstract. For financial institutions, foreign exchange (FX) rates commonly constitute the most important part of the market risk. In order to assess the risk of opened position various models can be used. However, since real FX returns exhibit higher than normal kurtosis and since the very far tails of the distribution should also be measured, the Monte Carlo simulation of multidimensional Lévy processes seems to be the most efficient approach. In this paper we apply two basic copula functions to connect multidimensional Lévy models and provide a model for portfolio probability distribution. Next, VaR (and AVaR) is calculated and the backtesting procedure is carried out.

Key words and phrases. FX rate portfolio, multidimensional Lévy models, VaR, AVaR, backtesting.

Mathematics Subject Classification. 60G35, 60G70.

1 Introduction

Standard approaches to model the evolution of financial quantities (such as stock prices, interest rates, foreign exchange rates) are based on the *Gaussian distribution* (through the application of a Wiener process or a Brownian motion) with possible combination with the Poisson distribution (a pure jump process). More advanced models belong to the family of *Lévy processes*, processes with infinite activity of jumps. Up to now, there have been introduced various types of such models, with some of them defined as subordinated (geometric) Brownian motions. Hence, the standard clock time is replaced by a suitable stochastic process, obviously a nonegative one, to model the arrival of a new information.

From the riskmanagement perspective, it is important to measure the risk of overall position, ie how the portfolio value can change during the passage of time. Generally, the value of any

portfolio is sensitive to the evolution of several distinct stochastic factors. As an implication, when the portfolio risk is estimated, the dependency among particular sources of randomness must be taken into account.

The intention of this paper is to make a further contribution to the theory and practice of portfolio modeling by multidimensional Lévy models (VG and NIG models) via elliptical copula functions (Gaussian and Student). For more details on other alternatives and previous research with the same date, see [10] and [17].

We proceed as follows. In Section 2, we start with the definition of the Lévy models family, including the tools for multidimensional problems. In Section 3 we describe the data set of six distinct FX rate market quotes taken over last 8 years, each with respect to CZK (Česká koruna) and finally, in Section 4 the results are obtained. More particularly, GBM, VG, and NIG models are applied to estimate a one day VaR and AVaR of an equally weighted portfolio by Monte Carlo simulation approach assuming either Gaussian or Student copula approach. The parameters of the models are regularly estimated on the basis of 1 000 preceding business days for each day over the last four years of the time series available.

2 Lévy models

The processes, which belong to the broadly defined Lévy-type models family,[1] can be characterized by independent and stationary increments. Another typical feature is a so called stochastic continuity – the probability, that a jump will occur within a particular time interval τ is zero.

2.1 Formal definition

The key step within the definition of advanced Lévy models is to formulate a characteristic function ϕ. Its use allow us to avoid several problems connected with the application of probability distribution function of a random variable X, $F_X(x)$. The relationship between the characteristic function of random variable X, $\phi_X(u)$, and its distribution function, $F_X(x)$, is given by Fourier-Stieltjes transformation:

$$\phi_X(u) = \mathbb{E}[\exp(\imath u X)] = \int_{-\infty}^{\infty} \exp(\imath u x) dF_X(x). \tag{1}$$

For a characteristic function, it holds in general, that $\phi(0) = 1$ and $|\phi(u)| \leq 1$ for all $u \in \mathbb{R}$. It also holds, that a characteristic function always exists, is continuous and determines the distribution function of a given probability distribution uniquely.

Suppose a probability distribution, which is infinitely divisible. Then, a Lévy process is a stochastic process $X(t)$ with zero origin and independent and stationary increments defined for all such infinitely divisible distribution. Moreover, increments of such processes over time interval $\tau \geq 0$, i.e. $X_{t+\tau} - X_t$, has characteristic function $(\phi(u))^\tau$.

[1] For more on Lévy models see [4] or [1].

Cumulant of the characteristic function $\Phi(u) = \ln \phi(u)$ is denoted as a characteristic exponent and fulfills the Lévy-Khintchin formula:

$$\Phi(u) = \imath\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} \left(\exp(\imath u x) - 1 - \imath u x \mathbb{I}_{|x|<1}\right) \nu(dx). \tag{2}$$

Here $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$ and ν is a measure on $\mathbb{R}\backslash\{0\}$ with

$$\int_{-\infty}^{\infty} \inf[1, x^2]\nu(dx) = \int_{-\infty}^{\infty} (1 \wedge x^2)\nu(dx) < \infty. \tag{3}$$

For a given infinitely divisible distribution, we can define a so called triple of Lévy characteristics,

$$\{\gamma, \sigma^2, \nu(dx)\}.$$

The former two define the drift of the process (deterministic part) and its diffusion. The latter is a Lévy measure. If it can be formulated as $\nu(dx) = u(x)dx$, it is a Lévy density. It is similarly to the probability density, with the exception that it need not be integrable and zero at origin.

2.2 Subordinated exponential Lévy models

The admissible prices of financial assets are usually restricted only to positive values, so that *exponential Lévy models*, i.e. models with a Lévy process $X(t)$ in the exponential should be preferred. It gives us the following formula to describe the dynamic of an asset price $\mathcal{S}(t)$:

$$\mathcal{S}(t) = \mathcal{S}e^{\mu t + X(t)}. \tag{4}$$

Here, in the exponential part of the model the Lévy process $X(t)$ is accompanied by a deterministic drift term, μ.

Many Lévy models commonly applied in Finance are formulated as a (geometric) Brownian motion driven by a particular intrinsic process (*subordinator/subordinated process*). From an economic point of view, such processes can be understood as a GBM within a (random) business time (it depends on economic activity, arrival of new information, etc.).

Denoting $\mathcal{Z}(t; \sigma, \mu)$ as a Wiener process in dependency on time t with parameters $\mu = 1$ and $\sigma = \sqrt{t}$, i.e. $\mathcal{Z}_t = \varepsilon\sqrt{t}$, $\varepsilon \in \mathcal{N}(0; 1)$, we can define Brownian motion $X(t; \theta, \vartheta)$ with drift θ and volatility ϑ driven by another Lévy process $\ell(t)$ with a unit mean and a variance specified by ν simply when we replace t by $\ell(t)$. Thus

$$X_t = \theta\ell(t) + \vartheta\mathcal{Z}(\ell_t), \tag{5}$$

which can be rewritten as:

$$X_t = \theta\ell(t) + \vartheta\varepsilon\sqrt{\ell(t)}. \tag{6}$$

This relation can be interpreted in such a way that the increment dX within an infinitesimal time interval dt is of normal distribution with mean $\theta\ell(dt)$ and variance $\vartheta^2\ell(dt)$. The mean of the driving process $\ell(t)$ should be dt and its variance will determine the *fat tails*. Similarly, the mean controls the asymmetry.

Very useful subordinators are *a gama proces* leading to the *Variance gamma model* (the variance is not given by standard time but by the so called gamma-time, hence the Variance gamma model) and *an inverse Gaussian process* leading to *NIG* model (*Normal inverse Gaussian model*).[2]

Concerning the parameter estimation, we can get them e.g. by maximization of the likelihood function (on the basis of a Lévy process density and the set of real data set) or solving the equations for particular moments (on the basis of the characteristic function and the empirically estimated moments of the distribution).

2.3 Multidimesional Lévy models

The approaches to model the dependency structure of random terms differs due to the probability distribution we consider. Since the standard approach to market risk modeling is still based on the application of a (geometric) Brownian motion, ie the higher moments of the underlying distribution are ignored, Cholesky decomposition is sufficient. However, the subordinated Lévy models introduced in the preceding section are defined by means of two distinct distributions. Furthermore, the subordinator is not Gaussian. This fact obviously does not allow us to apply the Cholesky decomposition if we are not equipped by other more advance tools. Bellow, we provide the theoretical analysis of multidimensional subordinated Lévy models.

Suppose that the evolution of a financial quantity can be described well only by the Lévy model (5). Consider two assets. Since each process consists of two random terms, the subordinator ℓ_i and the Wiener process ε_i, the covariance formula for two possibly dependent subordinated processes of the Lévy type, $\mathcal{X}_1$ and $\mathcal{X}_2$, is the following:

$$\mathtt{cov}[\mathcal{X}_1, \mathcal{X}_2] = \theta_1\theta_2\mathtt{cov}[\ell_1, \ell_2] + \vartheta_1\vartheta_2\mathbb{E}[\sqrt{\ell_1\ell_2}]\mathbb{E}[\varepsilon_1\varepsilon_2]. \tag{7}$$

It is useful to derive also its correlation counterpart:

$$\begin{aligned} \mathtt{cor}[\mathcal{X}_1, \mathcal{X}_2] &= \frac{\mathtt{cov}[\mathcal{X}_1, \mathcal{X}_2]}{\mathtt{var}[\mathcal{X}_1]\mathtt{var}[\mathcal{X}_2]} \\ &= \frac{\theta_1\theta_2\mathtt{cov}[\ell_1, \ell_2] + \vartheta_1\vartheta_2\mathbb{E}[\sqrt{\ell_1\ell_2}]\mathbb{E}[\varepsilon_1\varepsilon_2]}{\sqrt{\vartheta_1^2 + \nu_1\theta_1^2}\sqrt{\vartheta_2^2 + \nu_2\theta_2^2}}. \end{aligned} \tag{8}$$

Looking at either (7) or (8), we can see that the dependency between both processes can arise either through the dependency of subordinators ℓ or through the dependency of Wiener processes – standard normal variables ε (recall that ℓ and ε should be mutually independent).

Another possibility is to take the advantage of suitable copula functions.

[2]For more details on Variance gamma model see e.g. Madan and Seneta [12] (for symmetric case) and Madan and Milne [11] and Madan *et al.* [13] (for asymmetric case). Similarly, Normal Inverse Gaussian (NIG) model is due to Barndorff-Nielsen [2] and [3] and its generalisation the Hyperbolic model was introduced in Eberlein and Keller [9].

2.4 Dependency via copula functions

A useful tool of dependency modeling are the copula functions,[3] i.e. the projection of the dependency among particular distribution functions into $[0,1]$,

$$\mathcal{C} : [0,1]^n \to [0,1] \text{ na } \mathbb{R}^n,\ n \in \{2,3,\ldots\}. \tag{9}$$

In fact, it can be regarded as a multidimensional distribution function with marginals in the form of standardized uniform distribution.

For simplicity assume once again two potentially dependent random variables with marginal distribution functions F_X, F_Y and joint distribution function $F_{X,Y}$. Then, following the Sklar's theorem:

$$F_{X,Y}(x,y) = \mathcal{C}(F_X(x), F_Y(y)). \tag{10}$$

If both F_X, F_Y are continuous a copula function $\mathcal{C}$ is unique. Sklar's theorem implies also an inverse relation,[4]

$$\mathcal{C}(u,v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)). \tag{11}$$

Formulation (10) above should be understood such that the joint distribution function gives us two distinct information: (i) marginal distribution of random variables, (ii) dependency function of distributions. Hence, while the former is given by $F_X(x)$ and $F_Y(y)$, a copula function specifies the dependency, nothing less, nothing more. That is, only when we put both information together, we have sufficient knowledge about the pair of random variables X, Y.

For the modeling purposes, we can combine e.g. VG and NIG marginals by means of Gaussian or Student copula function. Since with the latter we can stress the tails, it might be more suitable for financial returns modeling.

3 Description of FX rate data

The data set we consider comprises of daily effective FX rates for EUR, GBP, HUF, PLN, SKK, and USD with respect to CZK as published by the Czech National Bank, ie generally the market quotes at 2 p.m. We monitor the market data starting on January 1, 2000. The last quotes were taken on December 31, 2007. It follows that we dispose of a time series of 2014 observations for log-returns of six distinct FX rates. For each FX rate basic descriptive statistics – mean, standard deviation, variance, skewness and kurtosis – of daily log-returns (per annum, if applicable) were evaluated, see Table 1.

We can see that the mean returns p.a. (the drift over the whole length) varies substantially between -1% (SKK) and -9% (USD). The standard deviation of two FX rates is around 5% (SKK, EUR), another two are close to 8% (GBP, HUF) and the last two are slightly above 10% (PLN, USD). Except the SKK rate, the skewness is significantly negative, the highest is for HUF (-0.8). By contrast, the highest kurtosis can be observed for the PLN rate (12), while the

[3] Basic reference for the theory of copula functions is [14], while [15] and [7] target mainly on the application issues in finance.

[4] To simplify the notation, $F_X^{-1}(u)$ is the inverse to the distribution function, if it exists, or at least its generalization $F_X^{-1}(u) = \inf\{x : F(x) \geq u\}$.

Table 1: Descriptive statistics of daily log-returns (p.a.)

Parameter		EUR	GBP	HUF	PLN	SKK	USD
mean	μ	-0.038	-0.058	-0.038	-0.019	-0.009	-0.086
variance	σ^2	0.0028	0.0067	0.0062	0.010	0.0034	0.0112
standard deviation	σ	0.053	0.082	0.079	0.100	0.058	0.106
skewness	κ_3	-0.297	-0.411	-0.790	-0.533	0.0646	-0.122
kurtosis	κ_4	7.432	5.320	9.748	12.209	7.618	4.047

USD is not very far from the Gaussian. When testing if the distribution can be regarded to be the Gaussian, several tests of Jarque-Bera type can be used. Here, the hypothesis of normality must be strongly rejected for all FX rates, including the USD rate data.

In order to simplify further calculations, the portfolio construction and to concentrate our attention on the non-linearity and non-normality, we can normalize the vectors of returns to get standardized time series with zero mean and unit variance. Bellow, we report the correlation matrix.

$$\mathbf{R} = \begin{pmatrix} 1. & 0.54 & 0.45 & 0.24 & 0.61 & 0.40 \\ 0.54 & 1. & 0.25 & 0.35 & 0.37 & 0.65 \\ 0.45 & 0.25 & 1. & 0.46 & 0.50 & 0.17 \\ 0.24 & 0.35 & 0.46 & 1. & 0.35 & 0.38 \\ 0.61 & 0.37 & 0.50 & 0.35 & 1. & 0.22 \\ 0.40 & 0.65 & 0.17 & 0.38 & 0.22 & 1 \end{pmatrix}.$$

4 Portfolio modeling

Consider a portfolio with equal sensitivity to each variable (FX rate)

$$\Pi_B = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6).$$

The true loss on a one-day window of the portfolio, ie. return with minus sign, incurred during last four years is depicted in (Figure 1).

The modeling by Lévy processes allows us to fit also the higher moments of the distribution. However, it simultaneously make the estimation of the parameters more data demanding (kurtosis is commonly less stable than variance). We therefore use 4 preceding years to fit the model, in such a way, the instability of the kurtosis and skewness fitting should be prevented, and estimate the risk (by MC simulation, assuming 500 000 trials) for the next day considering VG model, NIG model and GBM, each of them either coupled together by Gaussian or Student copula function in order to obtain a multidimensional distribution.

Each model is considered for three distinct significance levels, $\alpha = \{0.05, 0.01, 0.001\}$. Hence, the first level corresponds to the Basel II requirement, while the others are related to the Economic capital calculation for banks with good or very good ratings. Next, the estimated VaR is compared to the true loss incurred on a given day. If the VaR is exceeded, we get an

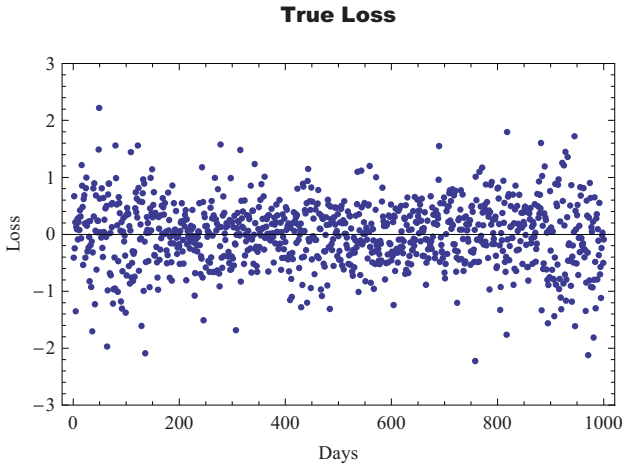


Figure 1: True loss over the last four years

Table 2: Backtesting results, Gaussian copula

	Assumption	VG	NIG	GBM
VaR (0.05)	50	31	31	31
VaR (0.01)	10	7	7	9
VaR (0.001)	1	1	1	3

exception day. This procedure is repeated over 4 years. The backtesting results are depicted in Table 2.

Since the last confidence level is very high, the data we have should provide us with one exception only.[5] Another interesting results are apparent from Figure 2. While the left part depicts VaR estimation, on the right we can observe the magnitude of AVaR. Particular curves depicts the evolution in time for $\alpha = 0.05$ (lower two) and $\alpha = 0.001$. Since the results for VG and NIG are almost interchangeable, we produce only one of them (VG, gray curves) accompanied by the GBM (black curves).

It is interesting to see that although the data are normalized (variance is still one and the average return is zero), the VaR and AVaR slightly decline for each combination of the model and probability. The reason might be the dependency among particular assets – since the linear correlation falls slightly, there was a bigger effect of diversification. Moreover, we observe that the VG and NIG models provide us with not very stable results. This might be give by (a) the changes in the skew and kurtosis for various intervals; and (b) the complexity of the model, ie many trials are needed to get a smooth results.

Next, we proceed to the Student copula, ie the dependency in both tails is emphasized and so is the probability of the extremal scenarios. In theory, distinct numbers of degrees of freedom

[5]The confidence of the results can be tested e.g. by Kupiec's statistic LR: $LR = -2\ln[(1-\alpha)^{N-M}\alpha^M] + 2\ln[(1-M/N)^{N-M}(M/N)^M]$.

might be used for particular models, however, for simplicity, we assume it to be 7.

The backtesting results (Table 3) indicates no clear difference among the Gaussian copula and Student copula, when the significance is 5%. The overestimation is the same. Surprisingly, the error (more particulary, the overestimation of the loss magnitude) is increased for 1% significance. Moreover, considering $\alpha = 0.1\%$, the number of exception is identical with the assumption, even for the GBM model, which is quite surprising.

Table 3: Backtesting results, Student copula

	Assumption	VG	NIG	GBM
VaR (0.05)	50	32	31	32
VaR (0.01)	10	5	5	7
VaR (0.001)	1	1	1	1

By inspection of Figure 3 it is apparent that the higher significance of the correlation in tails results into apparent increase of the estimated risk, both by VaR and $AVaR$, especially for both Lévy models.

5 Conclusions

The assessment of the market risk of the portfolio is a crucial issue of the risk management unit of financial institutions. Since real market returns commonly departure from the Gaussianity, the models fitting well the fatter tails (mainly left) are desired. A large group of models targeting on higher moments of the distribution (skewness and kurtosis) are so called subordinated GBMs from the Lévy models family (VG, NIG). However, the dependency modeling of such models is further complicated.

In this paper, we focused on a FX rate portfolio. First, we normalized the series of six distinct FX log-returns to get zero mean and unit variance. In this way, the data on equally weighted portfolio differ only due to the skewness and kurtosis. Next, we modeled the FX returns independently with a standard model (GBM) and two subordinated alternatives (VG and NIG). Each model of the marginal distribution of FX returns was evaluated with two elliptical copula functions, either Gaussian or Student.

It was documented that although the Gaussian copula is symmetric with no emphasis on the tails, the fat tailed marginals (VG, NIG) provide us with considerable improvements against the standard market model – both, the skewness and kurtosis are preserved. Clearly, the Student copula allows us to fit the heaviness of the tails leaving the skewness to be controlled only slightly by the marginals. For the series we considered, VG model needs more degrees of freedom than NIG model.

All the results were obtained for the overall series of data, it might be therefor interesting to compare particular models of the marginal distribution jointly with various copula functions within the predictive task.

The preliminary backtesting results indicate important improvements of copula approach with Lévy marginals for risk assessment. However, the data series should be studied more carefully before final conclusion could be made (subseries, clusters, various weighting). Moreover,

since not only the probability of a loss is important, also its magnitude should be measured. In such a case, however, the fitting is more problematic. Finally, the results can be used to obtain risk adjusted performance measures, such as RORAC or RAROC.

Acknowledgement

The paper is based on research activities sponsored through GAČR (Czech Science Foundation – Grantová Agentura České Republiky) project No. 402/08/1237. The support is greatly acknowledged.

References

[1] APPLEBAUM, D. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2004.

[2] BARNDORFF-NIELSEN, O.E. Normal inverse Gausian distributions and the modeling of stock returns, *Research report* No. 300, Department of Theoretical Statistics, Aarhus University, 1995.

[3] BARNDORFF-NIELSEN, O.E. Processes of Normal Inverse Gaussian Type, *Finance and Stochastics* **2**, 41–68, 1998.

[4] BERTOIN, J. *Lévy Processes*. Cambridge University Press, 1998.

[5] CARR, P., GEMAN, H., MADAN, D.B., YOR, M. The fine structure of asset returns: An empirical investigation, *Journal of Business* **75**, 305–332, 2002.

[6] CARR, P., GEMAN, H., MADAN, D.B., YOR, M. Stochastic Volatility for Lévy Processes, *Mathematical Finance* **13**, 345–382, 2003.

[7] CHERUBINI, G., LUCIANO, E., VECCHIATO, W. *Copula Methods in Finance*. Wiley, 2004.

[8] CONT, R., TANKOV, P. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC press, 2004.

[9] EBERLEIN, E., KELLER, U. Hyperbolic Distributions in Finance. *Bernoulli* **1**, 281–299, 1995.

[10] GURNÝ, P., TICHÝ, T. Dependency structure models of financial asset returns. In: *Financial management of firms and financial institutions*, Ostrava, 2007.

[11] MADAN, D.B., MILNE, F. Option pricing with VG martingale components, *Mathematical Finance* **1**, 39–56, 1991.

[12] MADAN, D.B., SENETA, E. The VG model for Share Market Returns, *Journal of Business* **63** (4), 511–524, 1990.

[13] MADAN, D.B., CARR, P., CHANG, E.C. The variance gamma process and option pricing, *European Finance Review* **2**, 79–105,1998.

[14] NELSEN, R.B. *An Introduction to Copulas*. 2nd ed. Springer, 2006.

[15] RANK, J. *Copulas. From theory to application in finance.* Risk books, 2007.

[16] ROCKAFELLER, R.T., URAYSEV, S. Conditional Value at Risk for general loss distributions. *Journal of Banking and Finance* **26**, 1443 – 1471, 2002.

[17] TICHÝ, T. Dependency models for a small FX rate sensitive portfolio. In: Stavrek, D. and Polouek, S. (ed) *Consequences of the European Monetary Intergration on Financial Markets.* Newcastle: Cambridge Scholars Publishing, 2008.

Current address

Tomáš Tichý, Ing. Ph.D.
Department of Finance, Faculty of Economics, VŠ-Technical university Ostrava,
Sokolská 33, 701 21, Czech Republic, +420 59 699 2336,
e-mail: tomas.tichy@vsb.cz

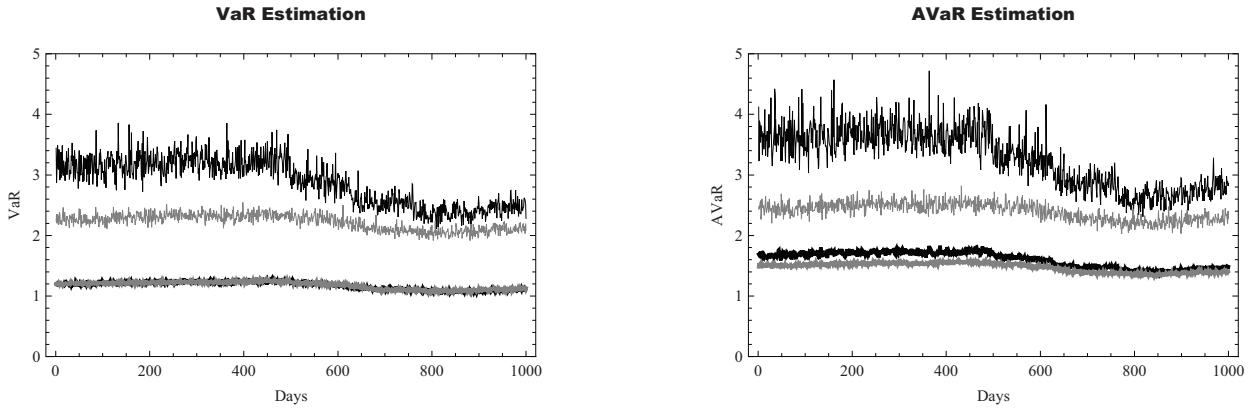


Figure 2: VaR and AVaR estimation, Gaussian copula for VG (black) and GBM (grey) models assuming significations of 0.05 (bottom) and 0.001 (top).

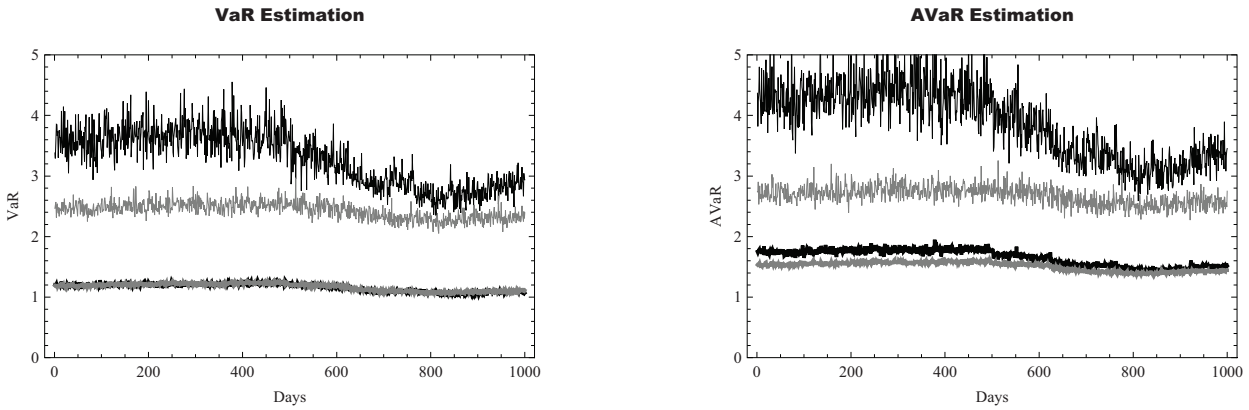


Figure 3: VaR and AVaR estimation, Student copula for VG (black) and GBM (grey) models assuming significations of 0.05 (bottom) and 0.001 (top).

Journal of Applied Mathematics

volume II (2009), number III

ANALYSIS OF CZECH FINANCIAL TIME SERIES

TREŠL Jiří, (CZ)

Abstract. At present, two main approaches to the volatility modelling exist: the GARCH and stochastic volatility models. In this paper, the GARCH models have been applied both to selected Czech capital market and exchange rates time series. The goal of this investigation consists in possible differences in behaviour with respect to time scale (days, weeks, months). The second level is given by differences between stock prices and exchange rates. Finally, besides univariate models, multivariate GARCH were also applied to discover possible dynamic relations among different time series.

Key words. Stock returns, exchange rates, volatility, GARCH models, dynamic correlation

Mathematics Subject Classification: Primary 62P20; Secondary 91B84.

1 Elementary descriptive statistics

The main aim of this study is to compare the behaviour of stock returns and exchange rate returns at Czech financial markets. Input data are given as daily values of corresponding time series values x_t during the period 2001-2007, i.e.1757 daily values of stock prices. We have selected four most liquid stocks: ***CEZ*** (energetics), ***KB*** (finance), ***TEL*** (telecommunication) and ***UP*** (petrochemicals). As for exchange rates, there are 1302 daily values observed within 2003-2007, and, namely, ***CHF***, ***EUR***, ***GBP*** and ***USD***. The subject of our analysis were daily logarithmic returns expressed as percentage and computed as

$$y_t = 100\left(\ln x_t - \ln x_{t-1}\right) \tag{1.1}$$

Similarly, 5-day (weekly) and 21-day (monthly) returns were computed too using non-overlapping intervals. Thus, lengths of corresponding weekly time series are 364 (stocks) and 273 (exchange rates). Monthly time series have lengths 86 (stocks) and 65 (exchange rates).

First, some elementary summary statistics related to daily returns were computed. The results obtained are compiled in the folowing tables.

Table 1: Summary statistics for stock returns

	Average	Median	StDev	LowerQ	UpperQ	InterQ	Skewness	Kurtosis
CEZ_01	0.15	0.17	1.95	-0.81	1.20	2.01	-1.11	13.33
CEZ_05	0.75	1.03	3.53	-1.06	2.83	3.89	-0.45	1.89
CEZ_21	2.95	3.87	8.53	-3.19	8.19	11.38	-0.21	0.56
KB_01	0.09	0.14	1.83	-0.94	1.08	2.02	-0.08	2.38
KB_05	0.43	0.71	3.52	-1.63	2.65	4.28	-0.33	0.72
KB_21	1.58	3.00	8.72	-3.57	7.44	11.01	-0.45	-0.01
TEL_01	0.00	0.02	2.07	-0.82	0.87	1.69	-0.45	5.78
TEL_05	0.02	0.32	3.90	-1.46	1.90	3.36	-0.62	3.09
TEL_21	0.08	0.98	9.74	-4.02	4.58	8.60	-0.33	2.96
UP_01	0.09	0.12	2.29	-0.75	1.03	1.78	-0.61	10.54
UP_05	0.48	0.55	4.63	-1.47	2.68	4.15	-1.39	12.28
UP_21	1.79	1.95	11.54	-3.24	7.48	10.72	-0.07	1.93

Table 2: Summary statistics for exchange rate returns

	Average	Median	StDev	LowerQ	UpperQ	InterQ	Skewness	Kurtosis
CHF_01	-0.02	-0.02	0.44	-0.26	0.20	0.46	0.15	13.12
CHF_05	-0.11	-0.10	0.64	-0.62	0.37	0.99	-0.07	-0.36
CHF_21	-0.47	-0.52	1.04	-1.34	0.31	1.65	-0.19	-0.43
EUR_01	-0.01	-0.01	0.33	-0.20	0.18	0.38	0.10	2.69
EUR_05	-0.07	-0.06	0.55	-0.42	0.29	0.71	-0.24	0.28
EUR_21	-0.30	-0.44	0.99	-1.00	0.52	1.52	0.03	-0.86
GBP_01	-0.02	0.00	0.50	-0.33	0.29	0.62	0.04	1.18
GBP_05	-0.12	-0.08	0.84	-0.64	0.44	1.08	-0.15	0.22
GBP_21	-0.51	-0.56	1.68	-1.68	0.63	2.31	0.08	0.02
USD_01	-0.04	-0.02	0.66	-0.41	0.35	0.76	0.04	1.33
USD_05	-0.20	-0.24	1.21	-0.98	0.52	1.50	0.26	0.19
USD_21	-0.83	-0.66	2.46	-2.36	1.07	3.43	-0.35	-0.22

2 Univariate modelling

The next step is the possibility of modelling of return time series. First, the behaviour of daily returns was investigated. In general, ACF values statistically significant at 5% level occured up to relatively high orders. Therefore, combined AR-GARCH models suitable for the modelling in the presence of heteroscedasticity were employed . The governing equations are [4], [8]

$$\begin{aligned} y_t &= \varphi_1 y_{t-1} + \ldots + \varphi_m y_{t-m} + \varepsilon_t & \varepsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 & e_t &\approx N(0,1) \end{aligned} \tag{2.1}$$

where σ_t denotes conditional standard deviation and e_t is normal white noise. First, we employed the simplest GARCH(1,1) model, which was quite efficient in most cases.

Second, EGARCH(1,1) model was employed to model possible asymmetric reaction with respect to positive and negative shocks e_t. General form of conditional variance in EGARCH models can be written as [7]

$$\log\left(\sigma_t^2\right) = \omega + \sum_{j=1}^{p} \beta_j \log\left(\sigma_{t-j}^2\right) + \sum_{i=1}^{q} \left(\alpha_i \left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| + \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right) \tag{2.2}$$

Clearly, if $\gamma_1 = 0$, then both positive and negative shocks exert the same influence on volatility. On the other hand, for $-1 < \gamma_1 < 0$, positive shocks influence volatility less than negative ones. Indeed, this second case actually occurs, as can be seen from the Table 5. In all cases, EGARCH (1,1) model led to slightly better results in comparison with GARCH (1,1).

Table 3: Estimated parameters of AR-EGARCH models for daily stock returns (SL=0.05)

	φ (1)	φ (2)	φ (4)	φ (05)	φ (7)	φ (8)	ω	α	β	γ
CEZ_01	0.054	-	-	-	-	0.054	-	0.241	0.892	-0.085
KB_01	0.072	-	0.058	-	-0.047	-	-0.075	0.216	0.924	-0.079
TEL_01	0.048	-	-	-	-	-	-0.174	0.267	0.996	-
UP_01	0.052	0.037	-	-	-	-	-0.170	0.420	0.940	-0.090

Obviously, there is strong direct dependence of conditional standard deviation on its previous values, manifested itself by large values of β parameter. Second, β values vary only slightly among individual stocks returns. On the other hand, GARCH (1,1) models proved to be sufficient for exchange rate daily returns and the results are presented in the Table 4.

Table 4: Estimated parameters of AR-GARCH models for daily exchange rate returns (SL=0.05)

	φ (1)	φ (3)	φ (6)	ω	α	β
CHF_01	-0.056	-	-	-	0.036	0.946
EUR_01	-0.072	-	-0.055	0.002	0.053	0.929
GBP_01	-	-0.052	-	-	0.023	0.971
USD_01	-	-	-	-	0.029	0.967

As for weekly stock returns, the matter is more stratified. There is no need for GARCH model in the case of *KB* and EGARCH is more efficient only for *TEL*. Again, this situation is summarized in the Table 5.

Table 5: Estimated parameters of AR-EGARCH models for weekly stock returns (SL=0.05)

	φ (1)	φ (2)	φ (3)	φ (4)	φ (5)	ω	α	β	γ
GARCH(1,1)									
CEZ_05	0.307	-	-	-	-	-	-	0.864	-
KB_05	0.190	-	-	-	-	-	-	-	-
UP_05	0.235	-	0.097	-0.113	0.067	-	0.159	0.742	-
EGARCH(1,1)									
TEL_05	0.261	-0.148	-	-	-	0.085	-	0.985	-0.139

On the contrary, the modelling weekly exchange rate returns does not demand GARCH at all and simple AR models are quite sufficient.

Table 6: Estimated parameters of AR models for weekly exchange rate returns (SL=0.05)

	CHF_05	*EUR_05*	*GBP_05*	*USD_05*
AR (1)	0.140	0.209	0.201	0.291
AR (2)	-0.121	-0.161	-	-

As for monthly stock returns, classical ARMA models are fully satisfactory.

Table 7: Estimated parameters of ARMA models for monthly stock returns (SL=0.05)

	CEZ_21	*KB_21*	*TEL_21*	*UP_21*
AR (7)	-	-0.633	-	-
AR(10)	-0.403	-	-	-
MA(7)	-	0.891	0.268	0.271
MA(10)	0.878	-	-	-

The same is true for 21-day exchange rate returns. Moreover, only MA models are needed.

Table 8: Estimated parameters of MA models for monthly exchange rate returns (SL=0.05)

	CHF_05	*EUR_05*	*GBP_05*	*USD_05*
MA (1)	-	0.290	0.244	0.283
MA (3)	-	-	-	-0.345
MA (5)	0.427	-	-	-
MA (7)	-	-	-	-0.763
MA (9)	-0.507	-	-0.512	-
MA (10)	-	-	-0.531	-

3 Multivariate modelling

First of all, the values of cross-correlation function for zero time lag were computed. Clearly, it is the case of synchronous correlation between all time series of returns under consideration.

Table 9: Estimated values of sample cross-correlation function for stock returns (SL=0.05)

	CEZ KB	CEZ TEL	CEZ UP	KB TEL	KB UP	TEL UP
daily	0.465	0.362	0.417	0.408	0.349	0.276
weekly	0.397	0.294	0.438	0.368	0.256	0.236
monthly	0.465	0.302	0.558	0.288	-	0.236

Table 10: Estimated values of sample cross-correlation function for exchange rate returns (SL=0.05)

	CHF EUR	CHF GBP	CHF USD	EUR GBP	EUR USD	GBP USD
daily	0.689	0.456	0.324	0.638	0.531	0.634
weekly	0.748	0.442	0.425	0.572	0.593	0.615
monthly	0.739	0.492	0.516	0.660	0.629	0.633

To investigate dynamical dependence among individual returns, Granger causality test was applied in two-dimensional system of jointly stationary time series [1]. We say, variable *x* Granger cause variable *y,* if delayed values of *x* variable improve prediction of *y*, despite that delayed values of *y* are introduced as explanatory variables. The model assumed is bivariate VAR(p) in the form

$$\begin{aligned} x_t &= c_1 + \sum_{i=1}^{p} \alpha_{1i} x_{t-i} + \sum_{i=1}^{p} \beta_{1i} y_{t-i} + u_{1t} \\ y_t &= c_2 + \sum_{i=1}^{p} \alpha_{2i} x_{t-i} + \sum_{i=1}^{p} \beta_{2i} y_{t-i} + u_{2t} \end{aligned} \tag{3.1}$$

Then the test of Granger causality in direction $x \to y$ can be understood as *F*-test of parameters $\alpha_{21}, \alpha_{22}, ..., \alpha_{2p}$ in the regression model (4), whereas the test of Granger causality in direction $y \to x$ is related to parameters $\beta_{11}, \beta_{12}, ..., \beta_{1p}$. The results obtained are summarized in the Table 11.

Table 11: Results of testing Granger causality

1-DAY	CEZ → UP	TEL → KB	-	-	-
P-value	0.021	0.047	-	-	-
5-DAY	CEZ → UP	KB → CEZ	UP → KB	-	-
P-value	0.045	0.048	0.008	-	-
21-DAY	-	-	-	-	-
P-value	-	-	-	-	-
1-DAY	USD → EUR	USD → CHF	EUR → CHF		
P-value	0.016	0.014	0.001		
5-DAY	USD → EUR				
P-value	0.029				
21-DAY	USD → EUR	USD → GBP	GBP → USD	GBP → CHF	GBP → EUR
P-value	0.049	0.006	0.005	0.004	0.002

Further, it is of interest to generalize univariate GARCH(1,1) employed former to multivariate case. This approach allows us to study time-varying behaviour of conditional covariances, which is an important problem in portfolio theory. A general multivariate GARCH model related to *k*-dimensional random process ε_t can be written in the form [6]

$$\varepsilon_t = \mathbf{e}_t \sqrt{\mathbf{H}_t} \qquad \varepsilon_t = \left(\varepsilon_{1t},...,\varepsilon_{kt}\right)^T \tag{3.2}$$

where $\mathbf{e}_t$ is a *k*-dimensional iid process with zero mean and covariance matrix equal to identity one. As a generalization of univariate case, $\mathbf{H}_t$ denotes time-varying conditional covariance matrix that needs to be specified. A general representation for the multivariate analogue of the GARCH(1,1) is so-called VEC model [5]

$$vech\left(\mathbf{H}_t\right) = \Omega^* + \mathbf{A}^*\ vech\left(\varepsilon_{t-1}\varepsilon_{t-1}{}^T\right) + \mathbf{B}^*\ vech\left(\mathbf{H}_{t-1}\right) \tag{3.3}$$

where *vech* operator stacks the lower portion of a matrix in a vector. For example, in the simplest bivariate case, this expression takes the form

$$\begin{bmatrix} h_{11t} \\ h_{12t} \\ h_{22t} \end{bmatrix} = \begin{bmatrix} \omega_{11}^* \\ \omega_{12}^* \\ \omega_{22}^* \end{bmatrix} + \begin{bmatrix} \alpha_{11}^* & \alpha_{12}^* & \alpha_{13}^* \\ \alpha_{21}^* & \alpha_{22}^* & \alpha_{23}^* \\ \alpha_{31}^* & \alpha_{32}^* & \alpha_{33}^* \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11}^* & \beta_{12}^* & \beta_{13}^* \\ \beta_{21}^* & \beta_{22}^* & \beta_{23}^* \\ \beta_{31}^* & \beta_{32}^* & \beta_{33}^* \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix} \tag{3.4}$$

containing 21 parameters to be estimated, and, generally, (k(k+1)/2)*(1+2(k(k+1)/2)). Thus, to overcome this shortcoming, diagonal VEC model was constructed with elements [3]

$$h_{ijt} = \omega_{ij} + \alpha_{ij}\ \varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}\ h_{ij,t-1} \qquad i,j = 1,2,...,k \tag{3.5}$$

containing generally 3(k(k+1)/2) parameters. Again, written explicite for bivariate case k=2

$$\begin{aligned} h_{11t} &= \omega_{11} + \alpha_{11}\ \varepsilon_{1,t-1}^2 + \beta_{11}\ h_{11,t-1} \\ h_{22t} &= \omega_{22} + \alpha_{22}\ \varepsilon_{2,t-1}^2 + \beta_{22}\ h_{22,t-1} \\ h_{12t} &= \omega_{12} + \alpha_{12}\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} + \beta_{12}\ h_{12,t-1} \end{aligned} \tag{3.6}$$

and there are 9 parameters to be estimated.

Bollerslev developed an alternative approach by assuming time-invariant conditional correlations ρ_{ij} between the elements of ε_t (CCC model). This model can be written as [2]

$$\begin{aligned} h_{iit} &= \omega_{ii} + \alpha_{ii}\ \varepsilon_{i,t-1}^2 + \beta_{ii}\ h_{ii,t-1} \qquad & i = 1,2,...,k \\ h_{ijt} &= \rho_{ij}\sqrt{h_{iit}}\sqrt{h_{jjt}} \qquad & i \neq j \end{aligned} \tag{3.7}$$

The results of computation are summarized in the tables below.

Table 12: Estimated parameters of diagonal VEC-GARCH model for daily stock returns. Parameters ρ_{ij} were computed using CCC model

	φ (1)	ω	α	β	ρ
CEZ	0.055	0.241	0.094	0.851	
KB	0.078	0.145	0.076	0.893	
TEL	0.052	0.019	0.077	0.919	
UP	0.051	0.194	0.131	0.830	
CEZ_KB		0.057	0.044	0.914	0.432
CEZ_TEL		0.032	0.053	0.912	0.401
CEZ_UP		0.071	0.067	0.876	0.389
KB_TEL		0.016	0.041	0.937	0.372
KB_UP		0.051	0.046	0.907	0.352
TEL_UP		0.018	0.052	0.915	0.300

Table 13: Estimated parameters of diagonal VEC-GARCH model for daily exchange rate returns. Parameters ρ_{ij} were computed using CCC model

	φ (1)	ω	α	β	ρ
CHF	-0.078	0.005	0.060	0.904	
EUR	-0.053	0.002	0.044	0.932	
GBP	-	0.002	0.025	0.966	
USD	-	0.003	0.021	0.971	
CHF_EUR		0.002	0.054	0.919	0.758
CHF_GBP		0.002	0.030	0.935	0.475
CHF_USD		0.002	0.038	0.912	0.323
EUR_GBP		0.002	0.029	0.944	0.615
EUR_USD		0.002	0.030	0.939	0.519
GBP_USD		0.001	0.020	0.970	0.616

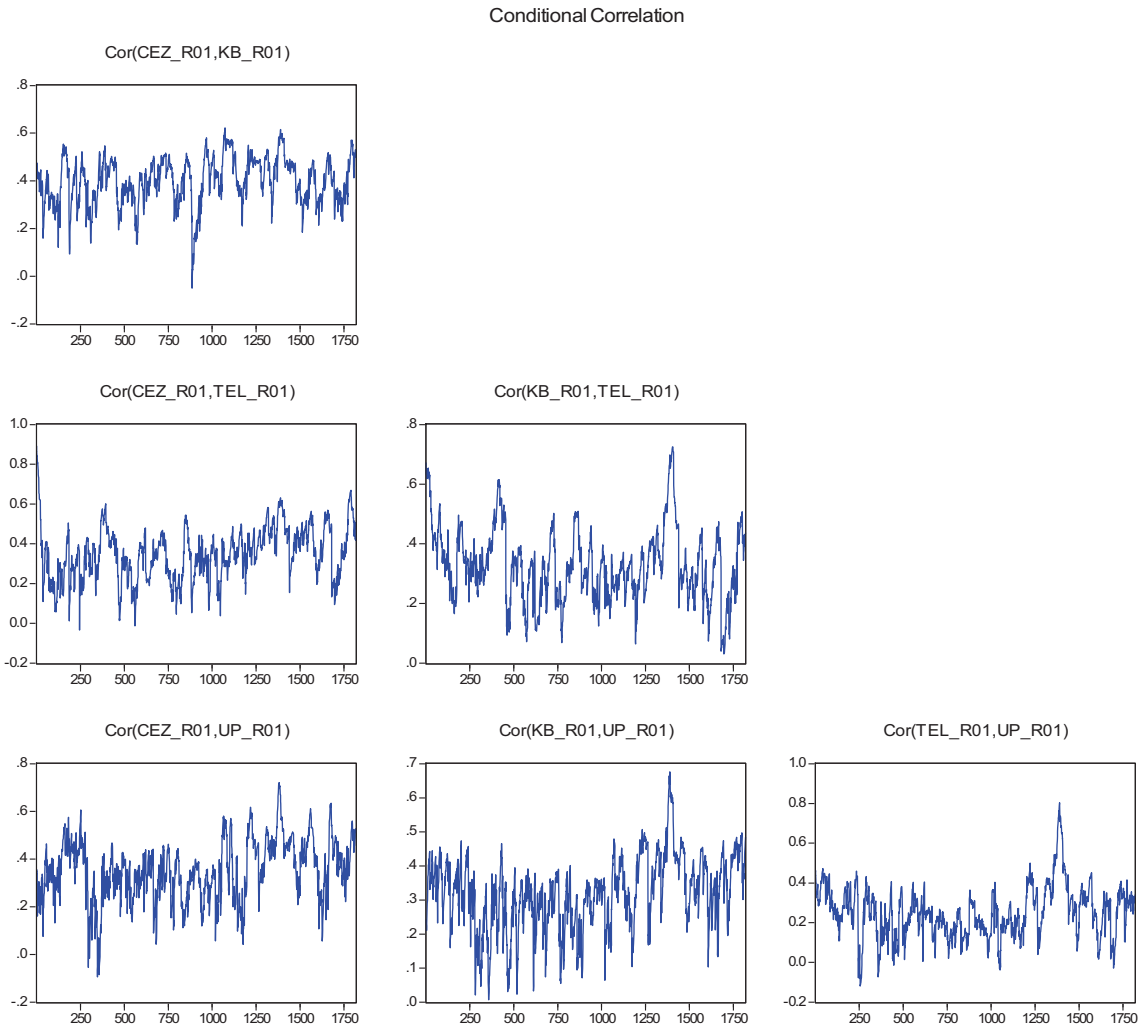


Figure 1: Time-varying conditional correlation between daily stock returns

Table 14: Estimated parameters of diagonal VEC-GARCH model for weekly stock returns. Parameters ρ_{ij} were computed using CCC model

	φ (1)	ω	α	β	ρ
CEZ	0.229	-	0.033	0.919	
TEL	0.227	-	0.047	0.941	
UP	0.210	0.772	0.064	0.889	
CEZ_TEL		-	0.073	0.625	0.366
CEZ_UP		-	0.015	0.969	0.393
TEL_UP		-	0.068	0.625	0.309

Again, like in univariate case, there was no need for multivariate GARCH model in the case of weekly exchange rate returns.

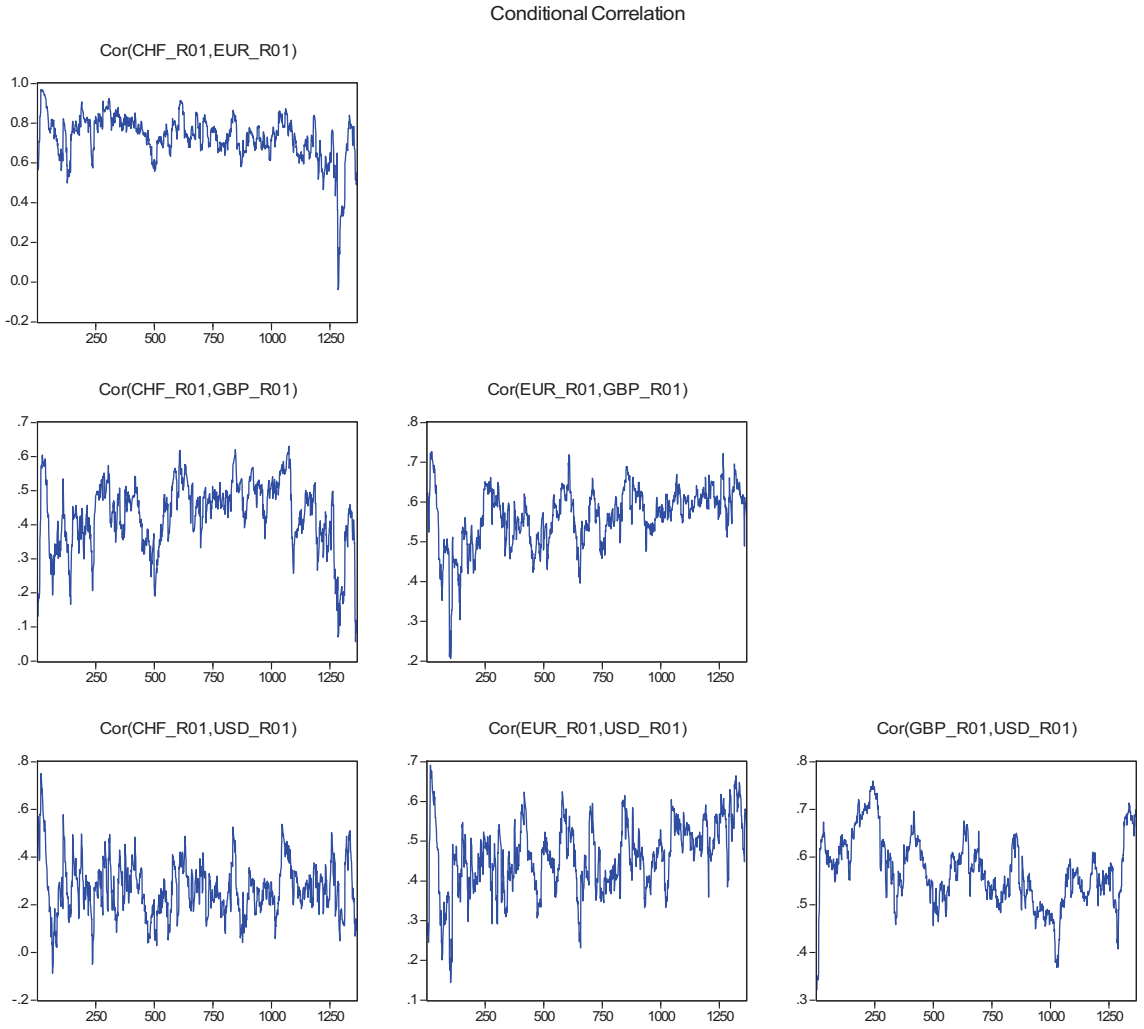


Figure 2: Time-varying conditional correlation between daily exchange rate returns

Conclusion

As for descriptive statistics, there is clear tendency to positive kurtosis and negative skewness in the case of daily and weekly stock returns. One the other hand, these findings are not repeated in exchange rate returns. Thus, GARCH models proved to be unavoidable for the modelling of daily and mostly also weekly stock returns and daily exchange rate returns. Second, asymmetric EGARCH(1,1) model capable to capture non-symmetry in reaction to positive and negative shocks was needed in the case of daily stock returns. It was manifested that positive shocks influenced volatility less than negative ones. On contrary, the modelling of weekly and monthly exchange rate returns demands only ARMA models. The same is true also for monthly stock returns.

Further, the values of cross-correlation function were always positive, signalizing some measure of coherent movement among time series. The measure of cross-correlation is higher in the case of exchange rate returns. Granger causality test revealed some directions of possible influence, exerted

by lagged values of explanatory time series (for example the influence of lagged CEZ returns on UP ones and lagged USD returns on EUR ones). Finally, the use of multivariate GARCH(1,1) model led to the possibility of modelling dynamical time-varying correlations among individual daily returns. In practice, this can help in the problems connected with portfolio theory.

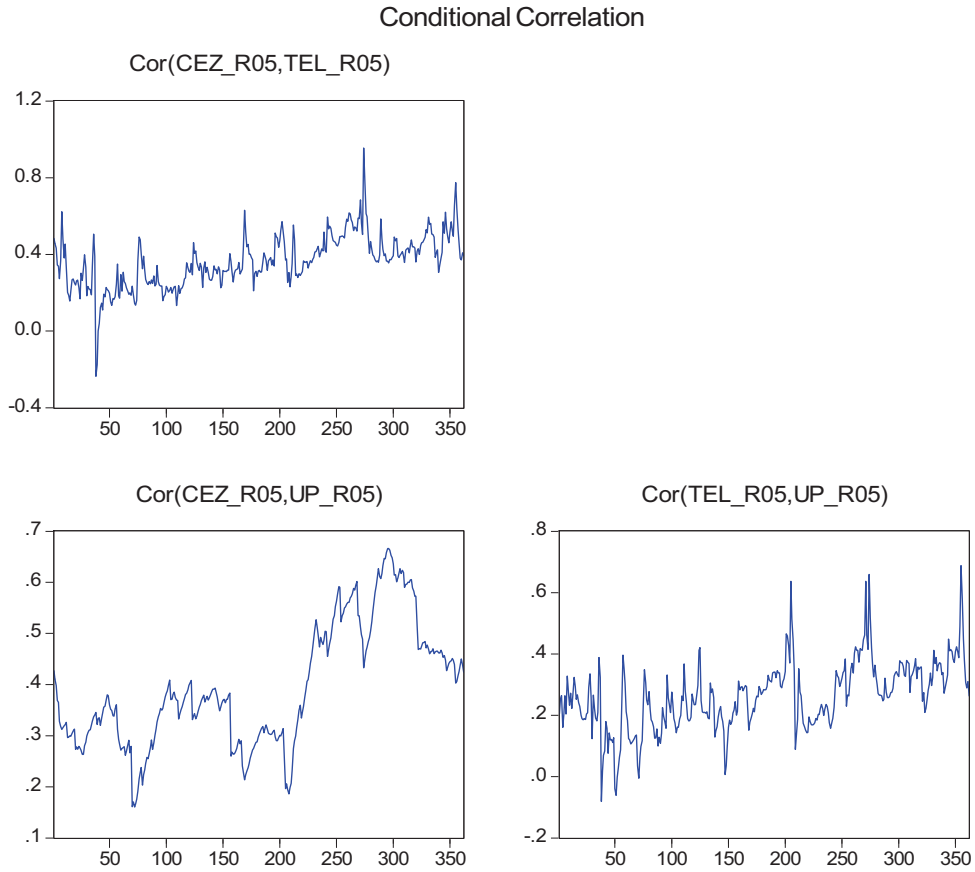


Figure 3: Time-varying conditional correlation between weekly stock returns

References

[1.] ALEXANDER, C.: *Market Models.* Wiley, New York 2001.

[2.] BOLLERSLEV, T.: *Modelling the Coherence in Short-Run Nominal Exchange Rates: Multivariate Generalized ARCH Approach.* Review of Economics and Statistics 72(1990), pp.498-505.

[3.] BOLLERSLEV, T. et al.: *Capital Asset Pricing Model with Time-Varying Covariances.* Journal of Political Economy 96(1988), pp.116-131.

[4.] ENGLE, R.: *ARCH (Selected Readings).* Oxford Univ.Press, Oxford 1995.

[5.] ENGLE, R. – KRONER, K.: *Multivariate Simultaneous Generalized ARCH.* Econometric Theory 11(1995), pp.122-150.

[6.] FRANSES, P. – van DIJK , D.: *Non-Linear Time Series Models in Empirical Finance.* Cambridge Univ.Press, Cambridge 2006.

[7.] NELSON, D.: *Conditional Heteroskedasticity in Asset Returns: a New Approach.* Econometrica 59(1991), pp.347-370.
[8.] TSAY, R.: *Analysis of Financial Time Series.* Wiley, New York 2002.

Current address

Trešl Jiří, doc.Ing.CSc.
University of Economics, Prague
Department of Statistics and Probability
W.Churchill Sq.4, 130 67 Prague 3, Czech Republic
e-mail: tresl@vse.cz

Journal of Applied Mathematics

volume II (2009), number III

VECTORIZED TABLE DRIVEN ALGORITHMS FOR DOUBLE PRECISION ELEMENTARY FUNCTIONS USING TAYLOR EXPANSIONS

BARRERA Tony, (SE), SPåNGBERG Daniel, (SE), HAST Anders (SE), BENGTSSON Ewert (SE)

Abstract. This paper presents fast implementations of the inverse square root and arcsine, both in double precision. In single precision it is often possible to use a small table and one ordinary Newton-Raphson iteration to compute elementary functions such as the square root. In double precision a substantially larger table is necessary to obtain the desired precision, or, if a smaller table is used, the additional Newton-Raphson iterations required to obtain the precision often requires the evaluation of other expensive elementary functions. Furthermore, large tables use a lot of the cash memory that should have been used for the application code.

Obtaining the desired precision using a small table can instead be realised by using a higher order method than the second order Newton-Raphson method. A generalization of Newton's method to higher order is Householder's method, which unfortunately often results in very complicated expressions requiring many multiplications, additions, and even divisions.

We show how a high-order method can be used, which only requires a few extra additions and multiplications for each degree of higher order. The method starts from the Taylor expansion of the difference of the value of the elementary function and a starting guess value for each iteration. If the Taylor series is truncated after the second term, ordinary Newton iterations are obtained. In several cases it is possible to algebraically simplify the difference between the true value and the starting guess value. In those cases we show that it is advantageous to use the Taylor series to higher order to obtain the fast convergent method. Moreover, we will show how the coefficients of a Chebyshev polynomial can be fitted to give as little error as possible for the functions close to zero and in the same time reduce the terms in the Taylor expansion.

In the paper we benchmark two example implementations of the method on the x86_64 architecture. The first is the inverse square root, where the actual table (to 12 bit precision) is provided by the processor hardware. The inverse square root is important in many application programs, including computer graphics, and explicit particle simulation codes, for instance the Monte Carlo and Molecular Dynamics methods of statistical mechanics.

The other example is the arcsine function, which has a slow converging Taylor expansion and where no tables are provided by the hardware.
The vectorized versions of the implementations of the inverse square root are 3.5 times faster than compiled code on the Athlon64 and about 5 times faster on the Core 2. The scalar version of the arcsine function is, depending on order and table size, between 2 and 3 times faster than the compiled code, and the vectorized version is between 3 and 4 times faster on the Athlon64, while it is between 4 and 5 times faster than the compiled version on the Core 2.

Key words and phrases. Vectorized algorithms, Taylor expansion, Double precision

Mathematics Subject Classification. Primary 60A05, 08A72; Secondary 28E10.

1 Introduction

For many years it has been said that using table driven algorithms to compute elementary functions on modern processors is inefficient due to the replacement of application program data cache lines by table entries of the elementary function. Still many attempts to implement efficient table algorithms has been done such as [1, 2, 3]

In single precision it is often possible to use a small table and one ordinary Newton-Raphson iteration. In double precision a substantially larger table is necessary to obtain the desired precision, or, if a smaller table is used, the additional Newton-Raphson iterations required to obtain the precision often requires the evaluation of other expensive elementary functions.

Obtaining the desired precision using a small table can instead be realised by using a higher order method than the second order Newton-Raphson method. A generalization of Newton's method to higher order is Householder's method, which unfortunately often results in very complicated expressions requiring many multiplications, additions, and even divisions.

We have developed a new high-order method which only requires a few extra additions and multiplications for each degree of higher order. The method starts from the Taylor expansion of the difference of the value of the elementary function and a starting guess value for each iteration. If the Taylor series is truncated after the second term, ordinary Newton iterations are obtained. In several cases it is possible to algebraically simplify the difference between the true value and the starting guess value. In those cases we show that it is advantageous to use the Taylor series to higher order to obtain the fast convergent method.

In practice, when a fixed precision is required (such as single or double precison) it is more efficient (fewer table entries or fewer terms in the expansion) to replace the Taylor expansion by a Chebyshev polynomial. The drawback of using the Chebyshev polynomial is that the magnitude of the obtained absolute errors are of the same size over the whole function range, so for values of the elementary function close to zero, the relative errors become large, and therefore do not give proper double precision results. For this reason we have fitted polynomial coefficients to obtain the desired relative errors over the whole function range.

For a fixed precision, such as our selected double precision, the size of the table varies, depending on the order of the method. More terms in the polynomial are required for smaller tables, so depending on the cache-use of the application program, there exist an optimal combination.

We benchmark two example implementations of the method on the x86_64 architecture.

The first is the inverse square root, where the actual table (to 12 bit precision) is provided by the processor hardware. The inverse square root is important in many application programs, including computer graphics, and explicit particle simulation codes, for instance the Monte Carlo and Molecular Dynamics methods of statistical mechanics. The other example is the arcsine function, where no tables are provided by the hardware, so we have developed four different versions, depending on order and table size. The arcsine function is generally hard to compute since many coefficients of the Taylor series is needed for double precision [4]. Nonetheless this paper presents a fast algorithm for double precision arcsine.

2 Theory

The general Taylor series of a function, $f(p)$, expanded around the point p_k is given by

$$f(p) = f(p_k) + \frac{f^{(1)}(p_k)}{1!}(p - p_k) + \frac{f^{(2)}(p_k)}{2!}(p - p_k)^2 + \cdots + \frac{f^{(m)}(p_k)}{m!}(p - p_k)^m \tag{1}$$

Truncating the Taylor series after the linear term, rearranging and dividing with $f^{(1)}(p_k)$, returns

$$p = p_k - \frac{f(p) - f(p_k)}{f^{(1)}(p_k)}, \tag{2}$$

which is identical to Newton's method, provided that the function, $f(p)$, has a zero at p. However, here, we keep the higher terms in the Taylor series, to provide for higher convergence. In general, the residual error can be written as

$$\epsilon = p - p_k, \tag{3}$$

although it is advantageous to have more elaborate error functions. For each general function we wish to find a function in the residual error:

$$R = R(p, p_k, x) \tag{4}$$

Below we show appropriate such functions for the inverse square root and the arcsine functions.

2.1 Inverse square root

The inverse square root is used in many applications in computer graphics and molecular dynamics and it can also be computed in many ways, using the Newton-Raphson method [5, 6] or by computing its Taylor expansion [7].

The inverse square root, $p = 1/\sqrt{x}$, does not have a Taylor series. Therefore we use the well known variable substitution $x = 1 - z$ to obtain the Taylor expansion around $z = 0$, [7]:

$$\frac{1}{\sqrt{1-z}} = 1 + \frac{1}{2}z + \frac{3}{8}z^2 + \frac{5}{16}z^3 \ldots \tag{5}$$

Setting the residual error function to

$$R = 1 - xp_k^2 = 1 - \frac{p_k^2}{p^2}, \tag{6}$$

which is small when p_k is a close starting guess to p, and noting that

$$\frac{1}{\sqrt{x}} = \frac{p_k}{\sqrt{xp_k^2}} = \frac{p_k}{\sqrt{1-R}} \tag{7}$$

allows the calculation of p:

$$p = p_k\left(1 + \frac{1}{2}R + \frac{3}{8}R^2 + \frac{5}{16}R^3 \dots\right) \tag{8}$$

Truncating this series gives a new, better, estimate of p, p_{k+1}, and the series can be iterated to arbitraty precision. If the series is truncated already after the linear term, iterations identical to the Newton method is obtained.

2.2 Arcsine

The acsine can generally be computed by the CORDIC algorithm [8], however here we will show how we can compute it efficiently using the proposed technique. The arcsine, $p = \arcsin(x)$, can be Taylor expanded around $x = 0$:

$$\arcsin(x) = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7 + \frac{35}{1152}x^9 + \frac{63}{2816}x^{11} \dots \tag{9}$$

Setting the residual error to

$$p - p_k = \arcsin(R), \tag{10}$$

returns the residual error function, R:

$$R = \sin(p_k - \arcsin(x)), \tag{11}$$

and using the trigonometric addition formulae and the relation $\cos(\arcsin(x)) = \sqrt{1-x^2}$, yields

$$R = \sqrt{1-x^2}\sin(p_k) - x\cos(p_k), \tag{12}$$

or alternatively

$$R = x\sqrt{1-x_k^2} - x_k\sqrt{1-x^2}, \tag{13}$$

where $p_k = \arcsin(x_k)$. Thus, when x_k is close to x, R is small. The final value of p can be found by iterating

$$p = p_k - R - \frac{1}{6}R^3 - \frac{3}{40}R^5 - \frac{5}{112}R^7 - \frac{35}{1152}R^9 - \frac{63}{2816}R^{11} \dots \tag{14}$$

3 Method

3.1 Implementation

For practical reasons it may be inefficient to have a short polynomial expansion and iterating several times. For instance, in the case of the arcsin function, the values of $\cos(p_k)$ and $\sin(p_k)$ are needed every iteration, and they are about as expensive to evaluate as the arcsin function itself. However, by providing a table of initial guesses for a number of $x_k \approx x$, i.e. $p_k = \arcsin(x_k)$, as well as $\cos(p_k)$ and $\sin(p_k)$ and providing a polynomial expansion of sufficient order it is possible to give the desired precision in a single iteration.

We have implemented the inverse square root, square root and arcsine in x86_64 assembler using gcc extensions and inline assembly to allow for easy calling of the routines from C. The implementation of the routines allow the evaluation of a single (scalar), 2, 4, 8, and a large number of values at a time (vector). To minimize the amount of code having to be written, we have written a number of macros which can be combined to form the various routines. The use of these macros generates code where the subsequent instructions are dependent on each other. For this reason a simple scheduler was written which is capable of scheduling the instructions from various macro blocks together. The goal of the scheduler is to place loads as early as possible and to place dependent instructions as far as possible from each other.

The program code, including the macro processor and scheduler can be downloaded from http://www.uppmax.uu.se/Members/daniels/fastmath/fastmath

3.2 Benchmarks

The benchmarks were run by generating a vector containing random numbers in an appropriate range. The implemented functions were then called to evaluate all the values in the vector. The same vector was evaluated $10^8/n$ times, where n is the length of the vector. This makes the test for each vector length last on the order of one second. The vector length was varied from 10 elements to about 10^6 elements, and the timing for the routines were estimated where the performance drops off due to the vector no longer fitting in the cache (apart from the arcsin implementation using the very large tables, where the values were taken for vector lengths corresponding to about 3 MB in size, as noted in the text). The tests were rerun and checked for consistency.

4 Results and Discussion

4.1 The inverse square root function

4.1.1 Implementation

In the case of the inverse square root function, the x86_64 instruction set contains an instruction `rsqrtps`, which is capable of at the same time give four 12 bit accurate estimates to the inverse

square root from single precision inputs [9, 10, 11].

While it is possible to use a fifth order method to give at least double precision accuracy, if the initial 12 bit estimates are converted to double precision directly, a higher performance version is possible, by first doing an ordinary Newton iteration in single precision, since it is possible to do single precision math at twice the speed of double precision math [9, 10, 11], and then convert the now 23 bit estimates to double precision and do a third order iteration to obtain the full double precision. Our implementation is shown in the appendix A.1.1.

4.1.2 Performance

Table 1 shows the performance of the inverse square root implementation. The performance of our implementation (Newton+New) is higher in all cases (scalar to vector) than the compilers library code, with the vector version beeing 3.5 times faster on the Athlon 64 and 4.7 times faster on the Core 2.

The reason for the increased performance of the routines computing two values or more at the same time over the scalar versions comes from two sources. First, the throughput of multiply and addition instructions on four single precision values at a time is twice that of double precision operations for the AMD64 processor [9]. On the Intel Core2 and AMD Family 10h processors the throughput of double precision operations on two values at the same time is two times as high as scalar operations, while the throughput of single precision operations on four values at the same time is four times that of scalar operations [10, 11]. Second, when more values are processed at a time, unrolling the loops and scheduling places dependencies further apart, allowing for improved software pipelining.

4.2 The square root function

The square root of a value can be estimated quickly if the inverse square root of the value has been computed:

$$\sqrt{x} = \frac{x}{\sqrt{x}} \tag{15}$$

However, if the square root of zero is being computed, the above formula does obviously not work. If evaluated as it stands it returns a NaN (not a number) for $\sqrt{0}$. However, the x86_64

Table 1: The performance of the inverse square root implementation on the AMD Athlon 64 and the Intel Core2 in cycles per determined value.

Implementation	scalar	2 at a time	4 at a time	8 at a time	vector
		AMD Athlon 64			
Compiler (gcc 4.1.2)	43.6				
Newton+New	36.1	24.6	18.0	13.0	12.2
		Intel Core 2			
Compiler (gcc 4.1.2)	39.8				
Newton+New	26.2	31.5	17.3	10.2	8.4

instruction set contains an instruction, `cmpordpd`, which returns true (all bits 1) if the two numbers compared are ordered (normal numbers) and zero (all bits 0) if the two numbers are NaNs. Since the double precision representation of zero corresponds to all bits zero, the value returned from the formula can therefore be masked using the return value of the `cmpordpd` instruction. Our implementation of the square root function is shown as part of the appendix A.1.2.

4.3 The arcsine function

4.3.1 Implementation

No hardware assisted initial estimate tables are available for the arcsine function in the x86_64 instruction set. In this case we therefore derive our own tables. The goal is to quickly find the value of R, the residual error function, which is later to be fed into a polynomial expansion of $\arcsin(R)$. The residual error function is:

$$R = \sqrt{1 - x^2}\sin(p_k) - x\cos(p_k). \tag{16}$$

We compute the value $\sqrt{1 - x^2}$ using the routines described in section 4.1. We also look up the values p_k, $\cos(p_k)$, and $\sin(p_k)$ in a table. The value p_k corresponds to $\arcsin(x_k)$, where x_k is a value close to the original x:

$$x_k = \frac{\mathrm{ent}(C(x+1))}{C}, \tag{17}$$

where C depends on the size of the table. Our implementation is shown in the appendix A.1.2

The Taylor expansion of the arcsine function to order 11 is

$$x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \frac{5}{112}x^7 + \frac{35}{1152}x^9 + \frac{63}{2816}x^{11} \tag{18}$$

The Chebyshev polynomial approximation to $\arcsin(x)$ in the range $x = -0.0125$ to 0.0125 with the largest absolute error slightly larger than $3 \cdot 10^{-17}$ is

$$1.00000000000001863010738x + .16666666571281759107583 4x^3 + .07501220871455619338687 22x^5 \tag{19}$$

The drawback of using a Chebyshev polynomial is that the magnitude of the obtained absolute errors are of the same size over the whole function range, so for values of $\arcsin(x)$ close to zero, the relative errors become large. For this reason we have fitted polynomial coefficients to obtain the desired relative errors over the whole function range, combining the best of the Taylor and Chebyshev polynomials.

For the arcsin function four different polynomial expansions were derived using a weighted

least squares method. The expansions are

$$\begin{aligned} a5(x) &= x + .16666666612865512862330 5x^3 + .07501024024355892521474 05x^5 \quad (20) \\ a7(x) &= x + .16666666669790574372742 7x^3 + .07499989765816825642887 84x^5 \\ &+ .04474302778051425403870 78x^7 \quad (21) \\ a9(x) &= x + .16666666666002119867929 9x^3 + .07500000749135134659344 39x^5 \\ &+ .04464006524781889882187 34x^7 + .03080481374174533170151 42x^9 \quad (22) \\ a11(x) &= x + .16666666666890968765731 5x^3 + .07499999868224255370696 29x^5 \\ &+ .04464312863363273795087 46x^7 + .03035663271733917784216 22x^9 \\ &+ .02346213199762025291392 91x^{11} \quad (23) \end{aligned}$$

The range with a largest relative error (see later) of about $6 \cdot 10^{-15}$ for the different polynomials is different. For the $a5(x)$ polynomial the range is ± 0.00125, for the $a7(x)$ polynomial it range is ± 0.04, for the $a9(x)$ polynomial ± 0.085, and for the $a11(x)$ polynomial it is ± 0.14.

Figure 1 gives the absolute error for the $a5(x)$ polynomial in the range $x = -0.0125$ to 0.0125, in comparison with the Chebyshev polynomial (eqn. 19) and the Taylor expansion to fifth order. The smallest absolute error is obtained with the Chebyshev polynomial, while the error of the Taylor expansion becomes very large at the larger range. Close to $x = 0$ the slope of the absolute error of the Taylor and $a5(x)$ polynomials is zero, while it is largely positive for the Chebyshev polynomial due to the first term ($1.0000000000000186301073 8x$). Figure 2 shows the relative error for x close to zero for the three polynomials. The relative error for the Chebyshev polynomial is $1.863 \cdot 10^{-14}$. The largest relative error for the $a5(x)$ polynomial is $5.7475 \cdot 10^{-15}$ at $x = \pm 0.01087$.

The largest possible range for each polynomial must not be smaller than the largest error

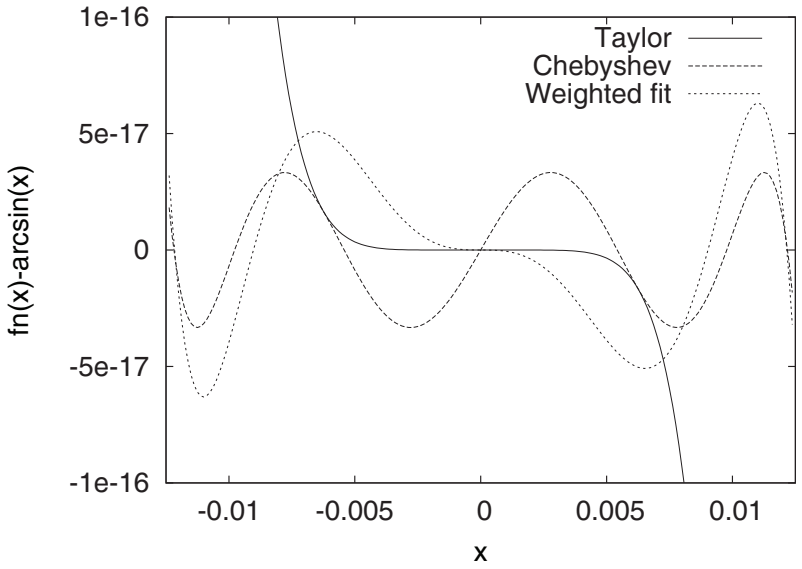


Figure 1: Absolute error in the polynomial expansions to fifth order to the arcsin(x) function in the range -0.0125 to 0.0125.

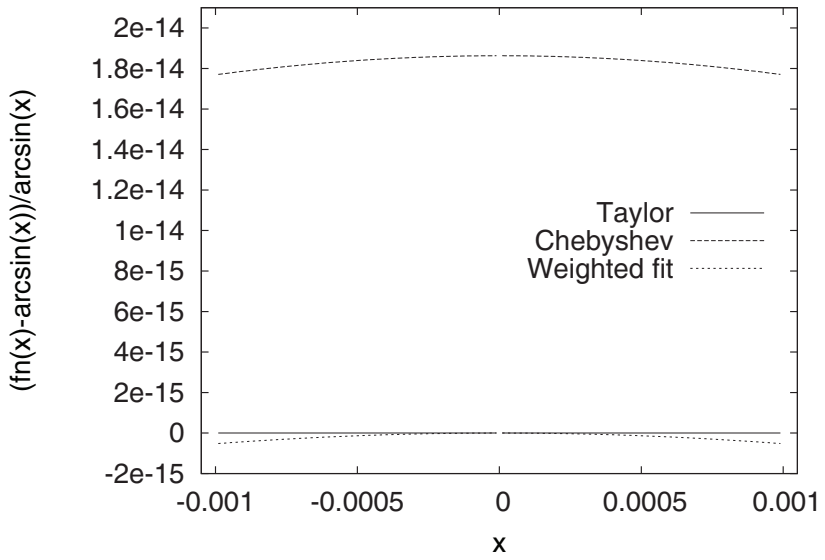


Figure 2: Relative error in the polynomial expansions to fifth order to the arcsin(x) function in the range -0.001 to 0.001.

from each starting guess. Table 2 shows the number of starting guesses required and the obtained largest error in each table in combination with the different polynomials. Each part of the table contains three values: The starting guess ($p_k = \arcsin(x_k)$) and the sine and cosine of the starting guess ($\sin(p_k)$, $\cos(p_k)$).

Figure 3 shows the resulting error of the arcsin implementation over the whole range. The relative difference to the compiler library code is mostly in the last bit.

4.3.2 Performance

Table 3 shows the performance of the arcsine implementation. The performance is higher in all cases than the compilers library code. In the case of the scalar implementation, the speedup is 2-3 on the Athlon 64 and Core 2. The vector implementation is up to 3.6 times faster on the Athlon 64 and 5.3 times faster on the Core 2. It should be noted that the shortest polynomials ($a5(x)$ and $a7(x)$) need fairly large tables, which will have negative effects on application code, since the values in the table will replace application data in the cache. However, the small

Table 2: The size of the tables of starting guesses and the two other values required for the evaluation of the arcsin function using the different polynomials.

Polynomial	Largest starting guess error	Number of starting guesses	Size of table (kB)
$a5(x)$	0.001238	6528	153
$a7(x)$	0.03958	640	15.0
$a9(x)$	0.08496	140	3.28
$a11(x)$	0.1264	64	1.50

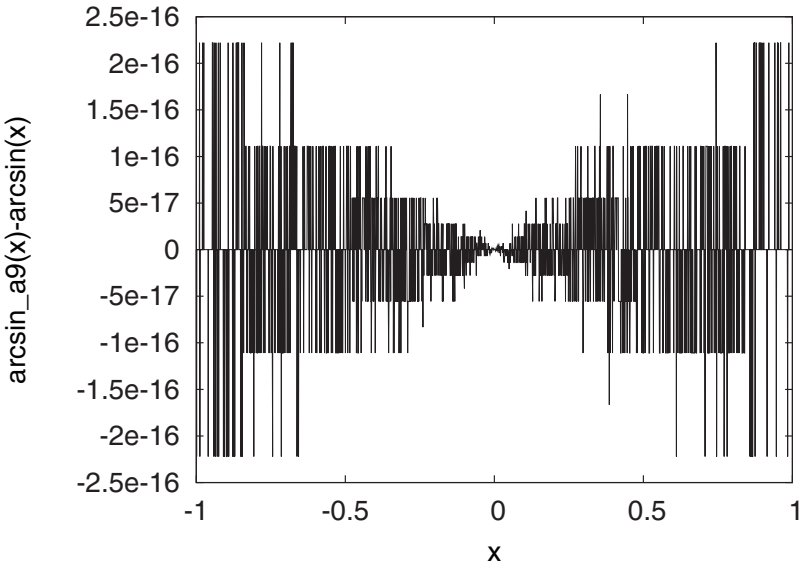


Figure 3: Resulting absolute difference between the implementation in this paper and the compilers library code over the whole range from -1 to 1 using the $a9(x)$ polynomial.

tables required by the $a9(x)$ and $a11(x)$ polynomials, 3.28 kB and 1.50 kB, respectively, should usually not cause any problems. The speedup using the $a11(x)$ polynomial is 3.2 on the Athlon 64 and 4.5 on the Core 2.

5 Conclusions

We have shown that it is possible to use the Taylor expansion in order to implement very efficiently the computation of such functions as the inverse square root and the arcsine functions. The proposed method is especially efficient when high precision is necessary. The vectorized versions of the implementations of the inverse square root are 3.5 times faster than compiled code on the Athlon64 and about 5 times faster on the Core 2. The scalar version of the arcsine function is, depending on order and table size, between 2 and 3 times faster than the compiled code, and the vectorized version is between 3 and 4 times faster on the Athlon64, while it is between 4 and 5 times faster than the compiled version on the Core 2.

Acknowledgement

The authors wish to thank UPPMAX for using their computing resources.

References

[1] Gal's accurate tables method revisited Stehle, D. Zimmermann, P. LORIA, Villers les Nancy, France; ARITH-17 2005. 17th IEEE Symposium on Computer Arithmetic, 2005

Table 3: The performance of the arcsin implementation on the AMD Athlon 64 and the Intel Core2 in cycles per determined value.

Polynomial	scalar	2 at a time	4 at a time	8 at a time	vector[a]
		AMD Athlon 64			
Compiler (gcc 4.1.2)	140				
$a5(x)$	50.8	53.2	50.4	57.9	57.0
$a7(x)$	48.0	46.2	44.0	42.2	38.3
$a9(x)$	53.2	50.0	47.7	44.7	40.5
$a11(x)$	60.1	50.8	49.7	47.7	43.3
		Intel Core 2			
Compiler (gcc 4.1.2)	124				
$a5(x)$	36.8	25.5	31.8[b]	29.4	27.4
$a7(x)$	41.0	23.1	34.2[b]	25.3	23.3
$a9(x)$	49.5	26.2	31.0[b]	26.0	26.1
$a11(x)$	58.3	31.1	31.9[b]	28.1	27.3

[a] Optimal vector length differs depending on CPU and which polynomial/size of table but is usually obtained at about 200-300 elements. For the routines with the large tables, the performance for larger input vector lengths deteriorates. The values in the table are given for working set sizes of about 3 MB.

[b] The performance of the routines processing 4 values at a time is not optimal on the Core 2. In the routines for 2 values at a time the `sqrtpd` instruction is used. In the routines for 4, 8, and vector the inverse square root times the value itself is used (see Sec. 4.2).

pp. 257- 264

[2] Table-lookup algorithms for elementary functions and their erroranalysis Tang, P.T.P. Proceedings, 10th IEEE Symposium on Computer Arithmetic, 1991 pp. 232-236.

[3] Table-driven implementation of the logarithm function in IEEE floating-point arithmetic Ping Tak, Peter Tang, ACM Transactions on Mathematical Software (TOMS), Volume 16 , Issue 4, 1990 pp. 378 - 400.

[4] Fast Trigonometric Functions Using INTEL's SSE2 Instructions. L. Nyland, M. Snyder. Intel Tech. Rep., available online at: http://www.weblearn.hs-bremen.de/risse/RST/docs/Intel/03-041.pdf.

[5] Reciprocation, Square Root, Inverse Square Root, and Some Elementary Functions Using Small Multipliers Milos D. Ercegovac , Toms Lang, Jean-Michel Muller , Arnaud Tisserand IEEE Transactions on Computers, Volume 49 , Issue 7 (July 2000) , Special issue on computer arithmetic pp. 628 - 637

[6] Computing the Inverse Square Root Ken Turkowski Book chapter: Graphics GemsV, pp. 16-21.

[7] Floating-point division and square root using a Taylor-series expansion algorithm Taek-Jun Kwon; Sondeen, J.; Draper, J. MWSCAS 2007. 50th Midwest Symposium on Circuits and Systems, Volume , Issue , 5-8 Aug. 2007, pp. 305 - 308.

[8] CORDIC-based computation of arccos and arcsin Tomas Lang, Elisardo Antelo Proceedings of the IEEE International Conference on Application-Specific Systems, Architectures and Processors, 1997, pp 132-143
[9] Software Optimization Guide for AMD64 Processors, rev 3.06, September 2005, AMD.
[10] Intel 64 and IA-32 Architectures Optimization Reference Manual, November 2006, Intel.
[11] Software Optimization Guide for AMD Family 10h Processors, rev 3.00, April 2007, AMD.

Current address

Tony Barrera
Uppsala, Sweden

Daniel Spångberg
Materials Chemistry, The Ångström Laboratory, Uppsala University, Box 538, S-751 21, Uppsala, Sweden, +46 18 471 3771, daniels@mkem.uu.se

Anders Hast
Creative Media Lab, University of Gävle, Gävle, Sweden, +46 26 648720, aht@hig.se

Ewert Bengtsson
Centrum för bildanalys, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden, +46 18 471 3467, Ewert.Bengtsson@cb.uu.se

A Appendix

A.1 Example implementations

Here the implementation of the macros / functions for the inverse square root and arcsin function for four values at a time using the shortest polynomial $a5(x)$ are shown.

Note that the macros for doing the inverse square roots and square roots are used by the arcsin implementation.

A.1.1 Inverse square root

The routine for obtaining four inverse square roots at a time is named `fast_math_quad_isqrt`:

```
static double __attribute__ ((aligned (16))) dual_one[2]={1.,1.};
static double __attribute__ ((aligned (16))) dual_c1[2]={0.5,0.5};
static double __attribute__ ((aligned (16))) dual_c2[2]={0.375,0.375};
static float __attribute__ ((aligned (16))) quad_onehalff[4]={1.5f,1.5f,1.5f,1.5f};
static float __attribute__ ((aligned (16))) quad_halff[4]={0.5f,0.5f,0.5f,0.5f};

void fast_math_quad_isqrt(double *x, double *y)
{
  register double *x0 __asm__("rdi")=x;
  register double *y0 __asm__("rsi")=y;
  __asm__ __volatile__ ( "\n\t"  /* */ \
     SCHEDULE
     "movlpd (%%rdi),%%xmm0\n\t"  /* */ \
     "movhpd 8(%%rdi),%%xmm0\n\t"  /* */ \
     "movlpd 16(%%rdi),%%xmm1\n\t"  /* */ \
     "movhpd 24(%%rdi),%%xmm1\n\t"  /* */ \
     MACRO EXPAND quad_isqrt(%%xmm0,%%xmm1,%2,%3,%4,%5,%6,%%xmm2,%%xmm3,%%xmm4,%%xmm5, \
                             %%xmm6,%%xmm7,%%xmm8,%%xmm9)
     "movlpd %%xmm0,(%%rsi)\n\t"  /* */ \
     "movhpd %%xmm0,8(%%rsi)\n\t"  /* */ \
     "movlpd %%xmm1,16(%%rsi)\n\t"  /* */ \
     "movhpd %%xmm1,24(%%rsi)\n\t"  /* */ \
     END SCHEDULE
     :
     :
     "r" (y0),
     "r" (x0),
     "m" (*quad_onehalff),
     "m" (*quad_halff),
     "m" (*dual_one),
     "m" (*dual_c1),
     "m" (*dual_c2)
     : "memory","xmm0","xmm1","xmm2","xmm3","xmm4","xmm5","xmm6","xmm7","xmm8","xmm9");
}
```

The macro `quad_isqrt`:

```
MACRO DEFINE quad_isqrt(xinout0,xinout1,xval1.5f,xval0.5f,xval1.0,xval0.5,xval0.375,xr0,xr1,xr2,xr3,xr4,xr5,xr6,xr7)
"movapd xinout0,xr5\n\t"  /* lo copy input values */ \
"movapd xinout1,xr6\n\t"  /* hi copy input values */ \
"cvtpd2ps xinout0,xr0\n\t"  /* lo input values in single precision */ \
"cvtpd2ps xinout1,xr7\n\t"  /* hi input values in single precision */ \
"movlhps xr7,xr0\n\t"  /* all input values in single precision */ \
```

```
MACRO EXPAND quad_isqrt_single(xr0,xval0.5f,xval1.5f,xr1,xr2)
"cvtps2pd xr2,xr0\n\t"  /* lo g=better precision inverse square root */ \
"movhlps xr2,xr2\n\t"  /* hi value in single precision */ \
"cvtps2pd xr2,xr7\n\t"  /* hi g=better precision inverse square root */ \
MACRO EXPAND dual_isqrt_double(xr5,xr0,xval1.0,xval0.375,xval0.5,xr3,xr4,xinout0)
MACRO EXPAND dual_isqrt_double(xr6,xr7,xval1.0,xval0.375,xval0.5,xr3,xr4,xinout1)
END MACRO
```

The macro `quad_isqrt_single`:

```
MACRO DEFINE quad_isqrt_single(xinput,xval0.5f,xval1.5f,xr0,xoutput)
"rsqrtps xinput,xr0\n\t"  /* tmp=inverse square root 12 bit precision */ \
"mulps xval0.5f,xinput\n\t"  /* 0.5*x */ \
"mulps xr0,xinput\n\t"  /* 0.5*x*tmp */ \
"mulps xr0,xinput\n\t"  /* 0.5*x*tmp*tmp */ \
"movaps xval1.5f,xoutput\n\t"  /* 1.5 */ \
"subps xinput,xoutput\n\t"  /* 1.5-0.5*x*tmp*tmp */ \
"mulps xr0,xoutput\n\t"  /* tmp*(1.5-0.5*x*tmp*tmp) better precision inverse square root */ \
END MACRO
```

The macro `dual_isqrt_double`:

```
MACRO DEFINE dual_isqrt_double(xinputx,xinputguess,xval1.0,xval0.375,xval0.5,xr0,xr1,xoutput)
"mulpd xinputguess,xinputx\n\t"  /* g*x */ \
"mulpd xinputguess,xinputx\n\t"  /* g*g*x */ \
"movapd xval1.0,xr0\n\t"  /* 1.0 */ \
"movapd xr0,xoutput\n\t"  /* 1.0 */ \
"subpd xinputx,xr0\n\t"  /* R=1.0-g*g*x */ \
"movapd xr0,xr1\n\t"  /* R */ \
"mulpd xr0,xr0\n\t"  /* R2=R*R */ \
"mulpd xval0.375,xr0\n\t"  /* c2*R2 */ \
"mulpd xval0.5,xr1\n\t"  /* c*R */ \
"addpd xr0,xoutput\n\t"  /* 1.+c2*R2 */ \
"addpd xr1,xoutput\n\t"  /* 1.+c*R+c2*R2 */ \
"mulpd xinputguess,xoutput\n\t"  /* result g*(1.+c*R+c2*R2) */ \
END MACRO
```

A.1.2 Arcsine

The table that contains all the 6528 required starting guesses and other values is named `fast_math_arcsin_table6528` and is not shown here. The routine that can be called from a C program is named `fast_math_quad_arcsin_4`:

```
#define AS_4_C1 .166666666128655128623305
#define AS_4_C2 .075010240243558925214740 5
static double __attribute__ ((aligned (16))) dual_arcsin_4_c1[2]={AS_4_C1,AS_4_C1};
static double __attribute__ ((aligned (16))) dual_arcsin_4_c2[2]={AS_4_C2/AS_4_C1,AS_4_C2/AS_4_C1};
static double __attribute__ ((aligned (16))) dual_arcsin_c_table_6528[2]={(double)(6528-1)/2.,(double)(6528-1)/2.};
void fast_math_quad_arcsin_4(double *x, double *y)
{
  register double *x0 __asm__("rdi")=x;
  register double *y0 __asm__("rsi")=y;
  __asm__ __volatile__ (
     "\n\t"  /* */ \
     SCHEDULE
     "movupd (%%rdi),%%xmm0\n\t"  /* */ \
     "movupd 16(%%rdi),%%xmm1\n\t"  /* */ \
     "lea %10,%%rax\n\t"  /* table base */ \
     MACRO EXPAND quad_arcsin4(%%xmm0,%%xmm1,%2,%3,%4,%5,%6,%7,%8,%9,%%rax,6528,%%r8,%%r9,%%r10,%%xmm2,%%xmm3,%%xmm4, \
                               %%xmm5,%%xmm6,%%xmm7,%%xmm8,%%xmm9,%%xmm10,%%xmm11,%%xmm12,%%xmm13,%%xmm14,%%xmm15)
     "movupd %%xmm0,(%%rsi)\n\t"  /* */ \
     "movupd %%xmm1,16(%%rsi)\n\t"  /* */ \
     END SCHEDULE
```

```
     :
     :
     "r" (y0),
     "r" (x0),
     "m" (*quad_onehalff),
     "m" (*quad_halff),
     "m" (*dual_one),
     "m" (*dual_c1),
     "m" (*dual_c2),
     "m" (*dual_arcsin_c_table_6528),
     "m" (*dual_arcsin_4_c1),
     "m" (*dual_arcsin_4_c2),
     "m" (*fast_math_arcsin_table6528)
     :
     "memory","rax","r8","r9","r10","xmm1","xmm2","xmm3","xmm4","xmm5","xmm6", \
     "xmm7","xmm8","xmm9","xmm10","xmm11","xmm12","xmm13","xmm14","xmm15");
}
```

The macro quad_arcsin4:

```
MACRO DEFINE quad_arcsin4(xinout0,xinout1,xval1.5f,xval0.5f,xval1.0,xval0.5,xval0.375,xval_arcsin_c,xval_arcsin_c1, \
             xval_arcsin_c2,rbasetable,tablesize,r0,r1,r2,xr0,xr1,xr2,xr3,xr4,xr5,xr6,xr7,xr8,xr9,xr10,xr11,xr12,xr13)
"mov $0xFFFFFFFF,r2\n\t" /* Address offset mask */ \
MACRO EXPAND dual_arcsin_r2(xinout0,xval1.0,xr0,xr2)
MACRO EXPAND dual_arcsin_r2(xinout1,xval1.0,xr1,xr3)
MACRO EXPAND quad_sqrt_less_reg(xr2,xr3,xval1.5f,xval0.5f,xval1.0,xval0.5,xval0.375,xr4,xr5,xr6,xr7,xr8,xr9)
MACRO EXPAND dual_arcsin4_part(xinout0,xr2,xval1.0,xval_arcsin_c,xval_arcsin_c1,xval_arcsin_c2, \
                               rbasetable,tablesize,r0,r1,r2,xr10,xr11,xr12,xr13)
MACRO EXPAND dual_arcsin4_part(xinout1,xr3,xval1.0,xval_arcsin_c,xval_arcsin_c1,xval_arcsin_c2, \
                               rbasetable,tablesize,r0,r1,r2,xr4,xr5,xr7,xr0)
END MACRO
```

The macro dual_arcsin_r2:

```
MACRO DEFINE dual_arcsin_r2(xinput,xval1.0,xr0,xoutput)
"movapd xinput,xr0\n\t" /* x=input value */ \
"movapd xval1.0,xoutput\n\t" /* 1.0 */ \
"addpd xval1.0,xr0\n\t" /* 1.0+x */ \
"subpd xinput,xoutput\n\t" /* 1.0-x */ \
"mulpd xr0,xoutput\n\t" /* r**2=1-x**2=(1.0-x)*(1.0+x) */ \
END MACRO
```

The macro dual_arcsin4_part:

```
MACRO DEFINE dual_arcsin4_part(xinout,xrval,xval1.0,xval_arcsin_c,xval_arcsin_c1,xval_arcsin_c2,rbasetable, \
                               tablesize,r0,r1,r2,xr0,xr1,xr2,xr3)
MACRO EQUATE sin_offset tablesize*8
MACRO EQUATE cos_offset tablesize*16
"movapd xinout,xr0\n\t" /*  x=input value */ \
"addpd xval1.0,xinout\n\t" /* x+1. */ \
"mulpd xval_arcsin_c,xinout\n\t" /* c*(x+1.) */ \
"cvttpd2dq xinout,xinout\n\t" /* index=(int)(c*(x+1.)) */ \
"movd xinout,r0\n\t" /* index */ \
"mov r0,r1\n\t" /* index */ \
"and r2,r0\n\t" /* low index */ \
"shr $32,r1\n\t" /* high index */ \
"movlpd (rbasetable,r0,8),xinout\n\t" /* low pk */ \
"movhpd (rbasetable,r1,8),xinout\n\t" /* high pk */ \
"movlpd sin_offset(rbasetable,r0,8),xr1\n\t" /* low sin(pk)   */ \
"movhpd sin_offset(rbasetable,r1,8),xr1\n\t" /* high sin(pk)  */ \
"movlpd cos_offset(rbasetable,r0,8),xr2\n\t" /* low cos(pk) */ \
"movhpd cos_offset(rbasetable,r1,8),xr2\n\t" /* high cos(pk) */ \
"mulpd xrval,xr1\n\t" /* r*sin(pk) */ \
"mulpd xr0,xr2\n\t" /* x*cos(pk) */ \
"subpd xr2,xr1\n\t" /* R=r*sin(pk)-x*cos(pk) */ \
"movapd xr1,xr2\n\t" /* R */ \
"movapd xr1,xr3\n\t" /* R */ \
```

```
"mulpd xr1,xr1\n\t" /* R2=R*R */ \
"mulpd xr1,xr2\n\t" /* R3=R2*R */ \
"mulpd xval_arcsin_c1,xr2\n\t" /* c1*R3 */ \
"mulpd xval_arcsin_c2,xr1\n\t" /* c2*R2 */ \
"addpd xval1.0,xr1\n\t" /* 1.+(c2*R2) */ \
"mulpd xr2,xr1\n\t" /* c1*R3*(1.+(c2*R2)) */ \
"addpd xr3,xr1\n\t" /* R+c1*R3*(1.+(c2*R2)) */ \
"subpd xr1,xinout\n\t" /* arcsin(x)=pk-(R+c1*R3*(1.+(c2*R2))) */ \
END MACRO
```

The macro quad_sqrt_less_reg:

```
MACRO DEFINE quad_sqrt_less_reg(xinout0,xinout1,xval1.5f,xval0.5f,xval1.0,xval0.5,xval0.375,xr0,xr1,xr2,xr3,xr4,xr5)
"movapd xinout0,xr4\n\t"  /* Store input value */ \
"movapd xinout1,xr5\n\t"  /* Store input value */ \
MACRO EXPAND quad_isqrt_less_reg(xinout0,xinout1,xval1.5f,xval0.5f,xval1.0,xval0.5,xval0.375,xr0,xr1,xr2,xr3)
"mulpd xr4,xinout0\n\t"  /* Resulting square root: x*(1./sqrt(x)) */ \
"movapd xinout0,xr0\n\t"  /* Store resulting value */ \
"cmpordpd xr0,xr0\n\t"  /* ordered compare. true if both numbers are ordinary numbers */ \
"andpd xr0,xinout0\n\t"  /* Mask out bad nan values to zero */ \
"mulpd xr5,xinout1\n\t"  /* Resulting square root: x*(1./sqrt(x)) */ \
"movapd xinout1,xr1\n\t"  /* Store resulting value */ \
"cmpordpd xr1,xr1\n\t"  /* ordered compare. true if both numbers are ordinary numbers */ \
"andpd xr1,xinout1\n\t"  /* Mask out bad nan values to zero */ \
END MACRO
```

The macro quad_isqrt_less_reg:

```
MACRO DEFINE quad_isqrt_less_reg(xinout0,xinout1,xval1.5f,xval0.5f,xval1.0,xval0.5,xval0.375,xr0,xr1,xr2,xr3)
"cvtpd2ps xinout0,xr0\n\t"  /* lo input values in single precision */ \
"cvtpd2ps xinout1,xr3\n\t"  /* hi input values in single precision */ \
"movlhps xr3,xr0\n\t"  /* all input values in single precision */ \
MACRO EXPAND quad_isqrt_single(xr0,xval0.5f,xval1.5f,xr1,xr2)
"cvtps2pd xr2,xr0\n\t"  /* lo g=better precision inverse square root */ \
"movhlps xr2,xr2\n\t"  /* hi value in single precision */ \
"cvtps2pd xr2,xr3\n\t"  /* hi g=better precision inverse square root */ \
MACRO EXPAND dual_isqrt_double_less_reg(xinout0,xr0,xval1.0,xval0.375,xval0.5,xr1,xr2)
MACRO EXPAND dual_isqrt_double_less_reg(xinout1,xr3,xval1.0,xval0.375,xval0.5,xr1,xr2)
END MACRO
```

The macro dual_isqrt_double_less_reg:

```
MACRO DEFINE dual_isqrt_double_less_reg(xinout,xinputguess,xval1.0,xval0.375,xval0.5,xr0,xr1)
"mulpd xinputguess,xinout\n\t"  /* g*x */ \
"mulpd xinputguess,xinout\n\t"  /* g*g*x */ \
"movapd xval1.0,xr0\n\t"  /* 1.0 */ \
"subpd xinout,xr0\n\t"  /* R=1.0-g*g*x */ \
"movapd xval1.0,xinout\n\t"  /* 1.0 */ \
"movapd xr0,xr1\n\t"  /* R */ \
"mulpd xr0,xr0\n\t"  /* R2=R*R */ \
"mulpd xval0.375,xr0\n\t"  /* c2*R2 */ \
"mulpd xval0.5,xr1\n\t"  /* c*R */ \
"addpd xr0,xinout\n\t"  /* 1.+c2*R2 */ \
"addpd xr1,xinout\n\t"  /* 1.+c*R+c2*R2 */ \
"mulpd xinputguess,xinout\n\t"  /* result g*(1.+c*R+c2*R2) */ \
END MACRO
```

Journal of Applied Mathematics

volume II (2009), number III

MATHEMATICAL MODELLING OF TWO-STAGE CONVERTER USING COMPLEX CONJUGATED MAGNITUDES AND ORTHOGONAL PARK/CLARKE TRANSFORMATION METHODS

DOBRUCKÝ Branislav, (SR), ŠUL Róbert, (SR), BEŇOVÁ Mariana, (SR)

Abstract. The paper deals with mathematical modelling of two-stage frequency converter. Two special methods of investigation are used here. The first one, method of complex conjugated amplitude, is used for steady-state investigation. The second one, orthogonal Park/Clarke transformation is suitable for investigation of three-phase electric circuits. The combination of both methods is very useful for analysis of three-phase electric motors in steady-state condition.

Key words. Orthogonal transformation, Fourier series, Computer simulation

Mathematics Subject Classification: Primary A2B05, Secondary 68M01, 37M10.

1 Introduction – Methods Used for Modelling and Calculation

It is well known in electrical engineering, that, since Φ–function method [1] is generally used for transient state solution, the method of complex amplitude has been introduced by Takeuchi [2] for analysis of converter circuit supplied electric machines in steady-state. The principle is based on substitution of trigonometric function by exponential one with complex argument. After determination of investigated variable in complex form, the variable can be that transformed back into time domain. Regarding to non-harmonic time waveforms of converter quantities the Fourier analysis is used for variables as the first step. Similar approach, but not the same, has been used by Kneppo in the seventies of the 20th century [3]. This way of investigation in electrical engineering has been called symbolic calculus.

Method of orthogonal transformation for electrical quantities was introduced by Park [4] for three-phase electric machines. The method makes possible to transform symmetrical 3-phase system into equivalent two-phase orthogonal system. This transformation decreases number of differential equations (from 3 to 2), and removes variable coefficients in the equations. Besides, trajectories of the quantities in complex Gauss plane denote themselves by six-side symmetry, thus the steady-state quantities can be calculated in only one sixth of time period. Clarke's multiplicative

transformation constant (equal 2/3) provides the invariances of voltage and current quantities in the both coordinating systems.

2 Using of Complex Conjugated Amplitude Methods for Electrical Circuit Fed by Single-Phase Inverter

For rectangular form of electric voltage (current), the sum of its odd harmonic components can be written as:

$$u(t) = \frac{4U_0}{\pi} \sum_{\nu=0}^{\infty} \frac{\sin(2\nu+1)\omega t}{2\nu+1}, \tag{2.1}$$

for non-negative integer ν (from interval $\langle 0; \infty)$), constant supply voltage of inverter U_0 and constant angular frequency ($= 2\pi f_0$)

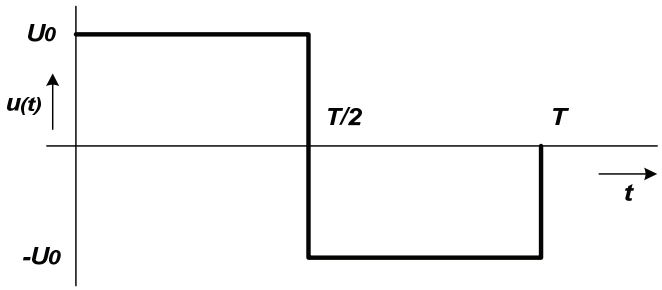


Fig. 1 Rectangular time-waveform of single phase inverter voltage with sine harmonic components

Let us apply such a voltage to passive *R-L* circuit whose complex impedance is:

$$\underline{Z}_{2\nu+1} = R + j(2\nu+1)\omega L = \left|Z_{2\nu+1}\right| e^{j\varphi_{2\nu+1}} \tag{2.2}$$

where: $\underline{Z}_{2\nu+1}$ complex impedance,
R resistance of the circuit,
j complex unit $\left(=\sqrt{-1}\right)$,
L inductance of the circuit,
$\left|\underline{Z}_{2\nu+1}\right|$ module (magnitude) of complex impedance $\left(=\sqrt{R^2 + (2\nu+1)^2 \omega^2 L^2}\right)$,
$\varphi_{2\square+1}$ argument (phase angle) of $\underline{Z}$ $\left(= \arctan \frac{(2\nu+1)\omega L}{R}\right)$.

Using Euler relations the non-harmonic voltage can be expressed as:

$$u(t) = \frac{2U_0}{\pi} \sum_{\nu=0}^{\infty} \frac{e^{j(2\nu+1)\omega t} - e^{-j(2\nu+1)\omega t}}{j(2\nu+1)} \tag{2.3}$$

Then corresponding complex current is:

$$i(t)=\frac{2U_0}{\pi}\sum_{\nu=0}^{\infty}\frac{e^{j(2\nu+1)\omega t}-e^{-j(2\nu+1)\omega t}}{j(2\nu+1)\left(R+j(2\nu+1)\omega L\right)}, \tag{2.4}$$

that can be written in complex conjugated magnitude form:

$$i(t)=\frac{1}{\sqrt{2}}\sum_{\nu=0}^{\infty}I_{2\nu+1}e^{j(2\nu+1)\omega t}+I_{2\nu+1}^{*}e^{-j(2\nu+1)\omega t}\ , \tag{2.5}$$

where complex magnitude of current will be:

$$I_{2\nu+1}=\frac{2\sqrt{2}}{\pi}\cdot\frac{U_0}{j(2\nu+1)\left[R+j(2\nu+1)\omega L\right]}=\frac{2\sqrt{2}}{\pi}\frac{U_0e^{-j\varphi_{2\nu+1}}}{\sqrt{R^2+(2\nu+1)^2\omega^2L^2}}, \tag{2.6}$$

and complex conjugate current magnitude:

$$I_{2\nu+1}^{*}=\frac{2\sqrt{2}}{\pi}\frac{U_0e^{j\varphi_{2\nu+1}}}{\sqrt{R^2+(2\nu+1)^2\omega^2L^2}}. \tag{2.7}$$

Finally, by adapting of (2.5), for current form in time domain one will obtain:

$$i(t)=\frac{4U_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin\left[(2\nu+1)(\omega t-\varphi_{2\nu+1})\right]}{(2\nu+1)\sqrt{R^2+(2\nu+1)^2\omega^2L^2}}. \tag{2.8}$$

Note: Eq. (8) is approximated numerical solution of ordinary differential equation:

$$\frac{di(t)}{dt}=-\frac{R}{L}\cdot i(t)+\frac{1}{L}\cdot u(t) \tag{2.9}$$

The relation for resulting current wave-form can be obtained also in compact closed form using classical analytical solution, Laplace transform or z-transform [7]:

$$i(t)=\frac{U_0}{R}\cdot\left(1-e^{-\frac{t}{\tau}}\right)+I_0 \tag{2.9a}$$

and $$i(t)=\frac{U_0}{R}\cdot\frac{1-e^{-\frac{T}{2\tau}}}{1+e^{-\frac{T}{2\tau}}}=\frac{U_0}{R}\cdot\tanh\left(\frac{T}{4\tau}\right) \tag{2.9b}$$

where τ is time constant of the circuit: $\tau = \frac{L}{R}$.

Anyway, the solution (2.8) makes possible to analyse more exactly each harmonic component comprised in total waveform.

$$i_{2\upsilon+1}(t) = \frac{4U_0}{\pi} \cdot \frac{1}{2\nu+1} \cdot \frac{1}{\left|\underline{Z}_{2\upsilon+1}\right|} \cdot \sin\left[(2\nu+1)(\omega t - \varphi_{2\nu+1})\right] \tag{2.10}$$

The waveforms whose Fourier series analysis leads to cosine functions of the harmonics can be expressed by the similar way, Fig. 2.

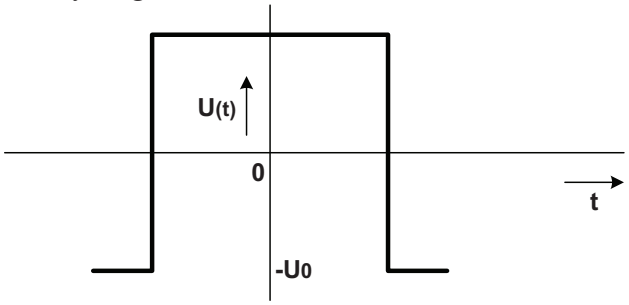


Fig. 2 Rectangular time-waveform of single phase inverter voltage with cosine harmonic components

$$u(t) = (-1)^{\nu} \cdot \frac{4U_0}{\pi} \sum_{\nu=0}^{\infty} \frac{\cos\left[(2\nu+1)(\omega t - \varphi_{2\nu+1})\right]}{(2\nu+1)} \tag{2.11}$$

3 Electrical Model of 2-Stage Frequency Converter

Scheme of the 2-stage converter is shown in Fig. 3. It comprises two semiconductor type of converters [6]:

- single-phase voltage inverter as the first stage,
- three-phase matrix converter or cykloconverter as the second stage.

The first stage operates with constant voltage U_0 and fixed frequency f_0. The second one supplies passive *R-L* or active load (electric motor) with variable output frequency and which is muchlesser then frequency of AC interlink between stages.

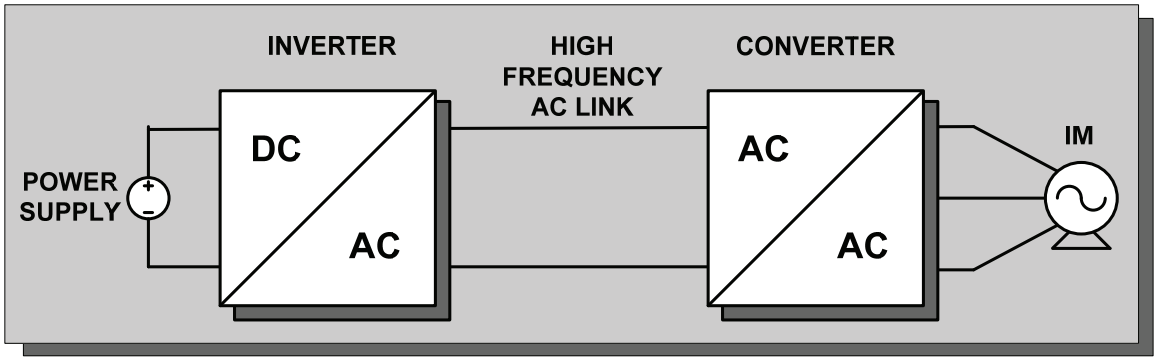


Fig. 3 Overall schematic diagram of 2-stage 3phase DC/AC/AC converter

Control of such system is described in greater detail in [5]. It is possible to use control methods based on fuzzy logic described in [6].

Considering rectangular form of the phase-current length of $2\pi/3$ radians with I_0 equal U_0/R, the scheme can be reconfigurated to the scheme of three-phase current inverter with *R-L* load [2], Fig. 4, whereas commutating capacitors could be omitted because of switches of inverter are switch-off capability.

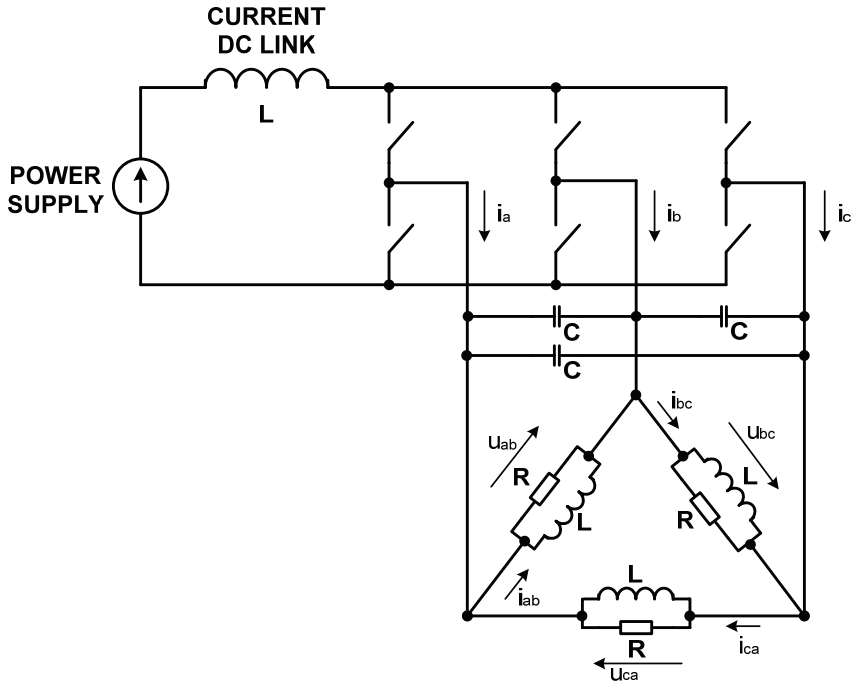


Fig. 4 Transfigurated scheme of three-phase current inverter with *R-L* load in delta connection

4 Mathematical Modeling using Complex Magnitudes Method and Park/Clarke Transformation

The output phase current of three-phase current inverter (Fig.5) can be expressed by Fourier series:

$$i(t) = \sum_{n=0}^{\infty} \left[a_n \cos(n\omega t) + b_n \sin(n\omega t) \right], \tag{4.1}$$

where coefficients a_n and b_n will be equal zero for even n. The coefficients for odd n will then be:

$$a_{2\nu+1} = \frac{2I_0}{\pi} \sin \frac{\frac{2\pi}{3}(2\nu+1)}{2\nu+1} \quad \text{and} \quad b_{2\nu+1} = \frac{2I_0}{\pi} \left(1 - \cos \frac{\frac{2\pi}{3}(2\nu+1)}{2\nu+1} \right). \tag{4.2}$$

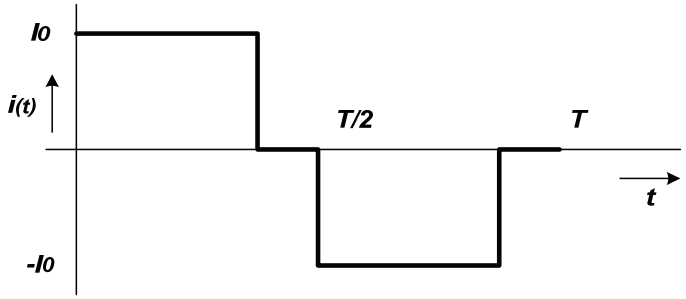


Fig. 5 Time-waveform of output phase current of the inverter

Using equations (4.1)-(4.2) and complex magnitudes method, the current will be:

$$i_a(t)=\frac{2I_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin\left[\frac{2\pi}{3}(2\nu+1)\right]\cdot\cos(2\nu+1)\omega t}{2\nu+1}+\left[\frac{1-\cos\left[\frac{2\pi}{3}(2\nu+1)\right]\cdot\sin[(2\nu+1)\omega t]}{2\nu+1}\right]=$$

$$=\frac{4I_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin\left(\frac{\pi}{3}(2\nu+1)\right)\cdot\cos\left[(2\nu+1)\left(\omega t-\frac{\pi}{3}\right)\right]}{2\nu+1} \tag{4.3}$$

4.1 Implementation of orthogonal transformation

Based on definition of complex-time vector by Park ([4] and others):

$$\underline{i}(t)=\frac{2}{3}\left[i_a(t)+\underline{a}\cdot i_b(t)+\underline{a}^2\cdot i_c(t)\right]=i_\alpha(t)+j\cdot i_\beta(t), \tag{4.4}$$

the real- and imaginary parts of the vector can be obtained:

$$\underline{i}_\alpha(t)=\frac{1}{3}\left[2i_a(t)-i_b(t)-i_c(t)\right] \tag{4.5a}$$

$$\underline{i}_\beta(t)=\frac{\sqrt{3}}{3}\left[i_b(t)-i_c(t)\right] \tag{4.5b}$$

Considering sum of phase currents to be zero:

$$i_a(t)+i_b(t)+i_c(t)=0, \tag{4.6}$$

real- and imaginary parts will be:

$$i_\alpha(t)=i_a(t), \qquad i_\beta(t)=\frac{i_b(t)-i_c(t)}{\sqrt{3}}. \tag{4.7a,b}$$

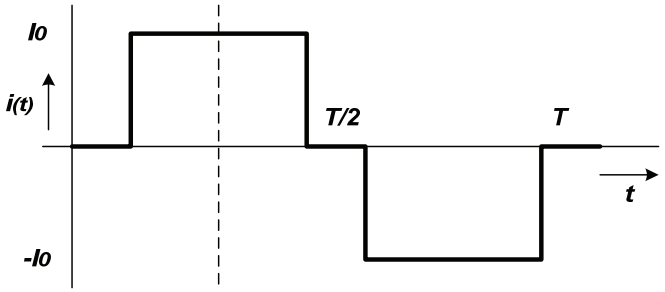


Fig. 6 Phase position of the *a*-phase current for transformation

Current of *a*-phase of the inverter then will be more simply than that of (4.3):

$$i_\alpha(t)=i_a(t)=\frac{4I_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin\left(\frac{\pi}{3}(2\nu+1)\right)\cdot\sin[(2\nu+1)\omega t]}{2\nu+1} \tag{4.8}$$

Current of *b*-phase lags the current of *a*-phase, thus:

$$i_b(t)=\frac{4I_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin\left(\frac{\pi}{3}(2\nu+1)\right)\cdot\sin\left[(2\nu+1)\left(\omega t-\frac{2\pi}{3}\right)\right]}{2\nu+1} \tag{4.9}$$

4.2 Determination of line-to-line voltage u_{ab} , u_{bc} ,u_{ca} of the inverter in steady-state:

Let the investigated quantities of the inverter, that means voltage on capacitors, i.e. u_{ab} , u_{bc} , u_{ca} be state-variables. To derive their time-waveforms it is necessary to know line-to-line currents, at first (Fig. 7).

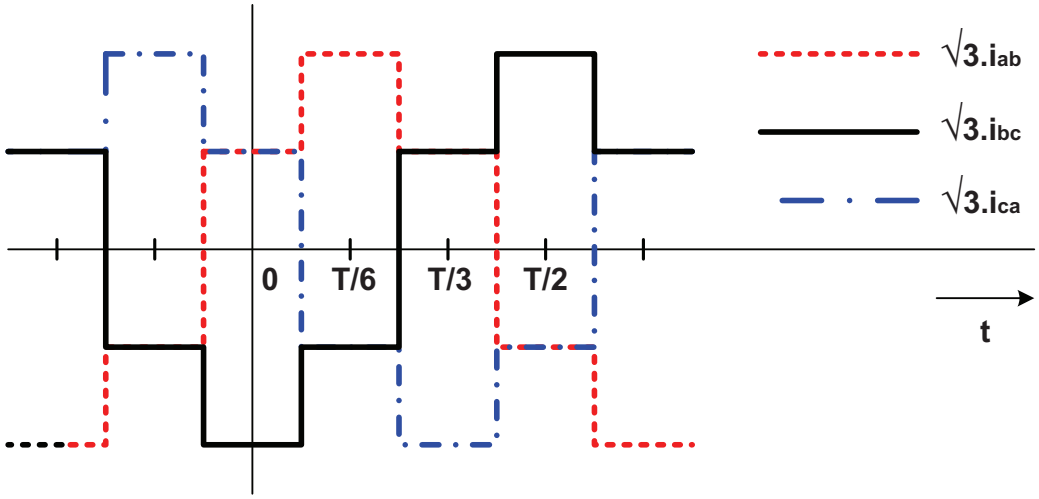


Fig. 7 Time-waveforms of line-to-line currents considering capacitors

Based on (4.5b), (4.7b) the differences of the phase currents are, indeed, the β-componets of Park complex time-vector, multiplied by constant √3.

Then the difference of phase-currents $i_b - i_c$, i.e. i_β x √3:

$$\sqrt{3}i_\beta(t) = i_b(t) - i_c(t) = -\frac{8I_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2\left(\frac{\pi}{3}(2\nu+1)\right)\cdot\cos\left[(2\nu+1)(\omega t)\right]}{2\nu+1} \tag{4.10}$$

and the difference of phase-currents $i_a(t) - i_b(t)$:

$$i_a(t) - i_b(t) = \frac{8I_0}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2\left(\frac{\pi}{3}(2\nu+1)\right)\cdot\sin\left[(2\nu+1)\left(\omega t+\frac{\pi}{6}\right)\right]}{2\nu+1} \tag{4.11}$$

Note: it is to be aware of the fact that Eqs. (4.8) and (4.10) create orthogonal series for i_α and i_β [7], [8]-[10], [11], which can be processed by orthogonal Fourier series rules.

Since

$$i_a(t) - i_b(t) = \frac{1}{R + j(2\nu+1)\omega L + j(2\nu+1)\omega C}\cdot\left[\left(u_{ab}(t) - u_{ca}(t)\right) - \left(u_{bc}(t) - u_{ab}(t)\right)\right] \tag{4.12}$$

and

$$u_{ab}(t) + u_{bc}(t) + u_{ca}(t) = 0 \tag{4.12a}$$

then

$$i_a(t) - i_b(t) = 3\cdot\frac{1}{R + j(2\nu+1)\omega L + j(2\nu+1)\omega C}\cdot u_{ab}(t). \tag{4.12b}$$

Line-to-line voltage u_{ab} then will be:

$$u_{ab}(t) = \frac{R + j(2\nu+1)\omega L}{1 - (2\nu+1)2\omega LC + j(2\nu+1)\omega RC}\cdot\frac{i_a(t) - i_b(t)}{3}. \tag{4.13}$$

After substituting $i_a(t) - i_b(t)$ by (23b) the line-to-line voltage can be expressed as:

$$\underline{u}_{ab}(t) = \frac{\left|\underline{Z}_{2\upsilon+1}\right|e^{j\varphi_{2\upsilon+1}}}{\left|\underline{Z}^{+C}_{2\upsilon+1}\right|e^{j\varphi^{+C}_{2\upsilon+1}}}\cdot\frac{8I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2\left[(2\nu+1)\frac{\pi}{3}\right]}{(2\nu+1)}\cdot\sin\left[(2\upsilon+1)\left(\omega t - \frac{\pi}{6} - \varphi_{2\upsilon+1} + \varphi^{+C}_{2\upsilon+1}\right)\right] \tag{4.14a,b}$$

$$u_{ab}(t) = -\frac{4I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2(2\nu+1)\frac{\pi}{3}}{(2\nu+1)}\cdot\frac{R + j(2\nu+1)\omega L}{1 - (2\nu+1)^2\omega LC + j(2\nu+1)\omega RC}\cdot\left(e^{j(2\nu+1)\omega\left(t-\frac{2\pi}{3}\right)} - e^{-j(2\nu+1)\omega\left(t-\frac{2\pi}{3}\right)}\right)$$

This can also be written in complex conjugated magnitude form:

$$u(t)=\frac{1}{\sqrt{2}}\sum_{\nu=0}^{\infty}\left(U_{ab_{2\nu+1}}e^{j(2\nu+1)\omega t}+U^{*}_{ab_{2\nu+1}}e^{-j(2\nu+1)\omega t}\right) \tag{4.15}$$

where complex magnitude of line-to-line voltage will be:

$$U_{2\nu+1}(t)=-\frac{4\sqrt{2}I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2(2\nu+1)\frac{\pi}{3}}{(2\nu+1)}\cdot\frac{R+j(2\nu+1)\omega L}{1-(2\nu+1)^2\omega^2LC+j(2\nu+1)\omega RC}e^{-j(2\nu+1)\omega\frac{2\pi}{3}} \tag{4.16a}$$

$$U^{*}_{2\nu+1}(t)=\frac{4\sqrt{2}I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2(2\nu+1)\frac{\pi}{3}}{(2\nu+1)}\cdot\frac{R+j(2\nu+1)\omega L}{1-(2\nu+1)^2\omega^2LC+j(2\nu+1)\omega RC}e^{j(2\nu+1)\omega\frac{2\pi}{3}} \tag{4.16b}$$

After completing of (4.15):

$$u_{ab}(t)=\frac{8I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2\left[(2\nu+1)\frac{\pi}{3}\right]}{(2\nu+1)}\cdot\frac{|\underline{Z}_{2\upsilon+1}|}{|\underline{Z}^{+C}_{2\upsilon+1}|}\cdot\sin\left[(2\upsilon+1)\left(\omega t-\frac{\pi}{6}\right)+\varphi_{2\upsilon+1}-\varphi^{+C}_{2\upsilon+1}\right] \tag{4.17}$$

where

$$|Z_{2\nu+1}|=\sqrt{\left(R^2+(2\nu+1)^2\omega^2L^2\right)};\ \varphi_{2\nu+1}=\arctan\left(\frac{(2\nu+1)\omega L}{R}\right) \tag{4.18}$$

and

$$|\underline{Z}^{+C}_{2\upsilon+1}|=\sqrt{\left(1-(2\nu+1)^2\omega 2LC\right)^2+\left((2\nu+1)^2\omega RC\right)^2};\ e^{\varphi^{+C}_{2\upsilon+1}}=\arctan\left(\frac{(2\nu+1)\omega RC}{1-(2\nu+1)^2\omega^2LC}\right) \tag{4.19}$$

Time relations for other line-to-line voltages u_{bc} and u_{ca} can be obtained by the similar way:

$$u_{bc}(t)=\frac{8I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2\left[(2\nu+1)\frac{\pi}{3}\right]}{(2\nu+1)}\cdot\frac{|\underline{Z}_{2\upsilon+1}|}{|\underline{Z}^{+C}_{2\upsilon+1}|}\cdot\sin\left[(2\upsilon+1)\left(\omega t-\frac{\pi}{6}\right)+\varphi_{2\upsilon+1}-\varphi^{+C}_{2\upsilon+1}\right], \tag{4.20a}$$

and finally:

$$u_{ca}(t)=\frac{8I_0}{3\pi}\sum_{\nu=0}^{\infty}\frac{\sin^2\left[(2\nu+1)\frac{\pi}{3}\right]}{(2\nu+1)}\cdot\frac{|\underline{Z}_{2\upsilon+1}|}{|\underline{Z}^{+C}_{2\upsilon+1}|}\cdot\sin\left[(2\upsilon+1)\left(\omega t-\frac{\pi}{6}\right)+\varphi_{2\upsilon+1}-\varphi^{+C}_{2\upsilon+1}\right]. \tag{4.20b}$$

5 Simulation Experiment Results

The following time functions and time-waveforms have been programmed in MatLab programming environment:

– current of single-phase inverter, Eq. (2.8),
– time-waveforms of phase currents $i_a(t)$, $i_b(t)$, $i_c(t)$, Eq. (4.8), (4.10),
– time-waveforms of orthogonal series i_α and i_β of Eqs. (4.3) and (4.7b),
– time-waveforms of phase currents $i_a(t)$ and voltage u_{ab} ,
– time-waveforms of state-space quantities, i.e. line-to-line terminal voltages u_{ab} , u_{bc} , u_{ca}, by the Eqs. (4.17) and (4.20a,b).

Parameters of the circuit: $R = 1$ Ohm, $L = 5$ mH, $C = 50$ μF, $U_0 = 100$ V, $I_0 = 10$ A.

Parameters of the simulation: time increment $\Delta T = 1$ μs, number of considered odd harmonic componens from 1 up to 999.

Version of MatLab programming environment: R2007b

Corresponding results of simulation experiments are given in Fig. 8a,b, Fig. 9a,b,c, Fig. 10a,b, Fig. 11a,b and Fig. 12a,b,c.

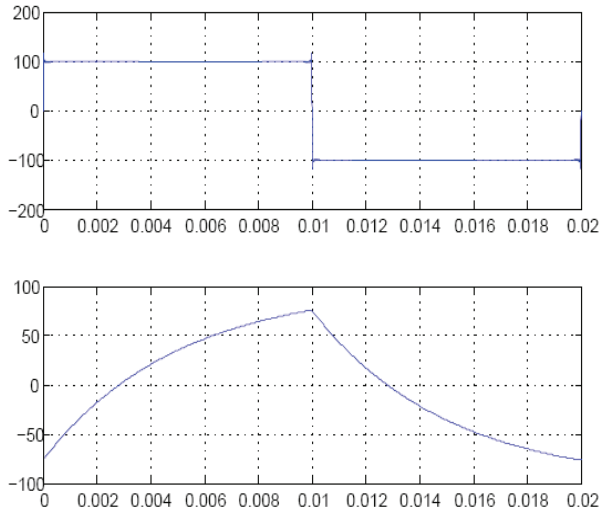


Fig. 8a,b Time-waveform of supply voltage (up), current response (bottom)

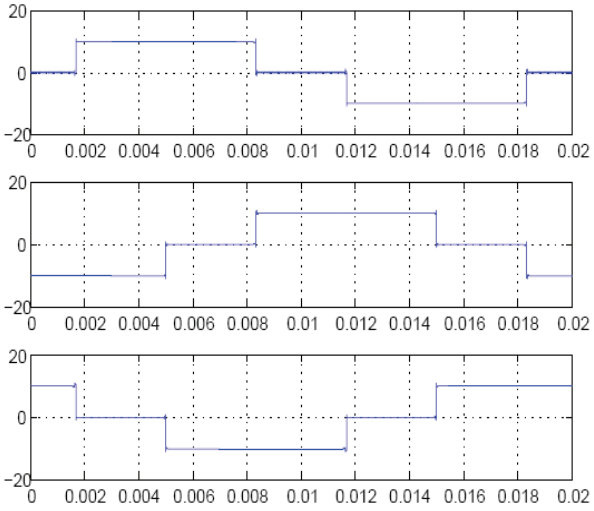


Fig. 9a,b,c Time-waveforms of phase currents $i_a(t)$, $i_b(t)$, $i_c(t)$ from up to bottom

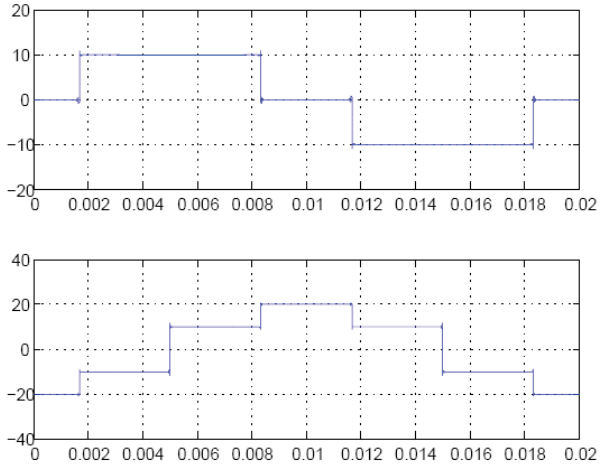


Fig. 10a,b Time-waveforms of $i_\alpha(t)$ and $i_\beta(t)$

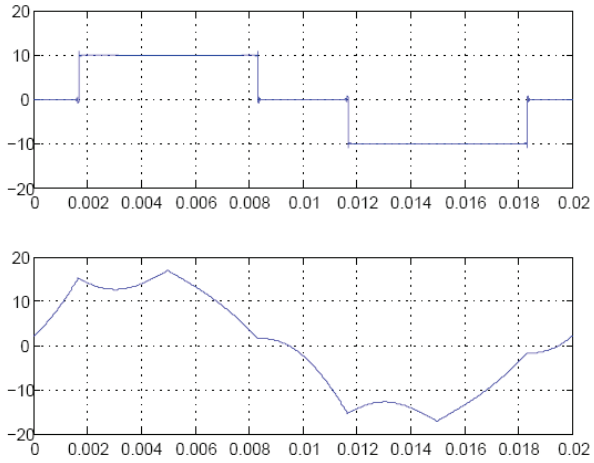


Fig. 11a,b Time-waveforms of $i_\alpha(t)$ and $u_{ab}(t)$

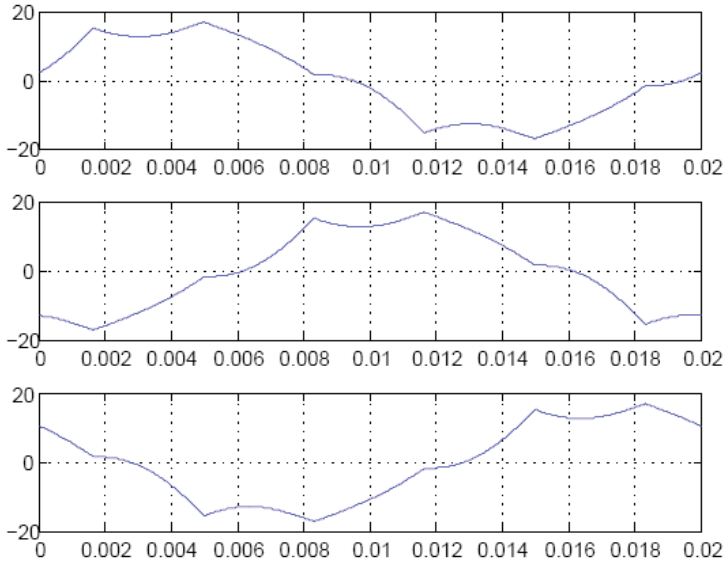


Fig. 12 Time-waveforms of line-to-line voltage u_{ab} (a), line-to-line voltage u_{bc} (b), line-to-line voltage u_{ca} (c)

6 Conclusions

The simulation experiments have shown very good coincidences of theoretical and simulated results. The time-waveforms of line-to-line voltages u_{ab} , u_{bc} , u_{ca} are very sensitive to choice of the values of inductors and capacitors of the inverter. Similarly, ripple of input current of the inverter depends on value of input smoothing inductor.

The plans for the near future are: investigation of the time waveforms of state-space quantities under operation with electrical induction motor [5], and also dynamical behaviour of the system of two-stage frequency converter connected to wheel motor.

Acknowledgement

The authors would like to thank to the Slovak Grant Agency of the Ministry of Education of the Slovak Republic for financial support of the Vega projects No. 1/3086/06 “New Method of Simulation, Modelling and Control of Mechatronic Systems“, and No. 1/0470/09 “Research of Topology and Control of Power Electronic System with Single-Phase HF Input and Two-Phase Orthogonal Output for SM/IM Electrical Motor.

References

[1] T.J. TAKEUCHI: *Method of Φ-Function to Analyse SCR Rectifier Circuit*. In: Journal of IEEJ (JP), Vol. 83-10, No. 901, 1963.

[2] T.J. TAKEUCHI: *Theory of SCR Circuit and Application to Motor Control*. Electrical Engineering College Press, Tokyo (JP), 1968 - Russian Edition, Energia Publisher, Leningrad-St. Petersburg (RF), 1973, 248 pp., T 051(01)-73 130-73.

[3] Ľ. KNEPPO: *Transient Phenomena in Electric Networks.* Textbook, Edition Center of STU University, Bratislava (SK), 1973.
[4] R.H. PARK: *Two Reaction Theory of Synchronous Machine. Generalized Method of Analysis* – Part I. IEEE Transactions, Vol. 48, July 1929, pp. 716-730
[5] I. GONTHIER, F. BERNOT, S.D. BOCUS, S. ELBAROUDI, A. BERTHON: *High-Efficiency Soft-Commutated DC-AC-AC Converter for Electric Vehicles*. In: Journal ElectroMotion (5), Mediamira Science Publisher (RO), 1998, No. 2, pp. 54-64, ISSN 1223-057X.
[6] B. DOBRUCKÝ, P. KUCER, M. ALEXÍK: *Fuzzy Control of AC Traction Drive*. In: Journal ElectroMotion (5), Mediamira Science Publisher (RO), 1998, No. 2, pp. 65-74, ISSN 1223-057X.
[7] M. MARČOKOVÁ: *Equiconvergence of Two Fourier Series*. Journal of Approximation Theory, Vol. 80, No.2 (1995), 151 – 163.
[8] M. MARČOKOVÁ, V. Stanček: *On Polynomials Orthogonal in a Finite (Symmetric) Interval*. Acta mathematica 9 (2006), pp. 209 – 214.
[9] B. DOBRUCKÝ, M. MARČOKOVÁ, R. ŠUL: *Prediction of Periodical Variable Structure System Behaviour using Minimum Data Acquisition Time*. In: Proc. of the 26th IASTED MIC'07 – Int'l Conf. on Modeling, Identification and Control, Innsbruck (AT), 12-14 Feb. 2007, ISBN 978-0-88986-633-1.
[10] B. DOBRUCKÝ, M. MARČOKOVÁ, M. KABAŠTA: *Using Orthogonal Transform for Solution of Matrix Con-verter Power Circumstances in Mathematica Environment*. In: Proc. of ApliMat 08 Int'l Conf. on Applied Mathematics, Bratislava (SK), Feb. 2008, pp. 54, ISBN 978-80-893-13-02-0
[11] V. GULDAN, S. HYČKOVÁ: *Orthogonal Polynomials and Sums of Some Infinite Series* (in Slovak). Proceedings of 5th didactics conference in Žilina – DidZA, June 12, 2008, 5 p. on CD (2008).
[12] B. DOBRUCKÝ, M. BEŇOVÁ, M. POKORNÝ: *Using Virtual Two Phase Theory for Instantaneous Single-Phase Power System Demonstration*. In: Przeglad Elektrotechniczny (PL), Vol. 85 (2009), No. 1, pp. 174-178, ISSN 0033-2097.

Current Address

Branislav Dobrucký, prof. Ing. PhD.
University of Žilina, Faculty of Electrical Engineering,
Univerzitná 1, 01 026 Zilina, Slovakia,
e-mail: dobrucky@fel.uniza.sk

Róbert Šul, Ing.
University of Žilina, Faculty of Electrical Engineering,
Univerzitná 1, 01 026 Zilina, Slovakia,
e-mail: robert.sul@fel.uniza.sk

Mariana Beňová, Ing. PhD.
University of Žilina, Faculty of Electrical Engineering,
Univerzitná 1, 01 026 Zilina, Slovakia
e-mail: benova@fel.uniza.sk

Journal of Applied Mathematics

volume II (2009), number III

COMPARISON OF TWO METHODS FOR SOLVING NONLINEAR PARABOLIC MODEL IN POROUS MEDIA

FTOREK Branislav, (SK), TOMAŠOVIČ, Peter, (SK)
DOROCIAKOVÁ Božena, (SK)

Abstract. We consider characteristics-based method for solving nonlinear convection diffusion equations. These equations arise in the transport of contaminant in porous media. We present two kinds of combinations to the derivatives with respect to time in the convection term. Numerical results are given.

Key words and phrases. nonlinear parabolic equation, method of characteristics.

Mathematics Subject Classification. Primary 65N30.

1 Introduction

In this paper we compare two schemes to the following model:

$$\partial_t C(x,t) + \partial_t \psi(C(x,t)) + q \cdot \partial_x C(x,t) + D\partial_{xx} C(x,t) = 0,\ t \in (0,T), \tag{1}$$

$$C(-\infty,t) = C(\infty,t) = 0, \tag{2}$$

which arises in the contaminant transport in porous media, C is the concentrative of contaminant, $\psi(C)$ is a nonlinear (degenerate) function and q, D (constants) represent convection and diffusion coefficients, where q is much more than D.

Firstly, we consider the scheme that has been developed by Kačur and Mahmood in [4] (see also [5, 7, 3]). At the time $t = t_i$ we find C_i from solving the following linear system:

$$(C_{i,l} - C_{i-1} \circ \varphi^i, v) + (\lambda_{i,l}(C_{i,l} - C_{i-1}), v) + \tau(D\partial_x C_{i,l}, \partial_x v) = 0,$$

$$\lambda_{i,l} = \frac{\psi_n(C_{i,l}) - \psi_n(C_{i-1})}{C_{i,l} - C_{i-1}}, \qquad \lambda_{i,0} = \psi'(t_i, C_{i-1}), \tag{3}$$

for $l = 1, ...$ If $|\lambda_{i,l} - \lambda_{i,l-1}|_\infty \leq c\tau$ ($c\tau$ is given) we put $l = l_i$, where l_i represents the number of iterations in time t_i, φ^i is a characteristic position and τ is a time step size. Then $\lambda_i = \lambda_{i,l_i}$ and $C_i = C_{i,l_i}$.The inner product $(C_{i-1} \circ \varphi^i, v)$ represents the approximation of the transport part

$$\partial_t C + q \cdot \partial_x C,$$

which can be computed using e.g. Bermejo approach [1].

In the degenerate case we replace ψ by ψ_n in the approximation scheme (3) where

$$\psi_n(s) = \max\{\tau, \min\{\psi(s), \tau^{-1}\}\}, \tag{4}$$

is a regularization of ψ.

Knabner et.al. [6] have developed the following formulation of the equation (1):

$$\partial_t \psi(C) - \alpha(x)\partial_t C + (1 + \alpha(x)) \left[\partial_t C + \frac{q}{1 + \alpha(x)} \partial_x C\right] = 0. \tag{5}$$

The scheme which they have proposed is called FIS (full implicit scheme):

$$C_j^i + \psi(C_j^i) = (1 + \alpha_j^i)C^{i-1} \circ \varphi_j^i - \alpha_j C_j^{i-1} + \psi(C_j^{i-1}), \tag{6}$$

where Galerkin characteristics discretization have been used [1](the indices i, j denote the time and space discretizations), where

$$\varphi_j^i = x_j - q\tau b_{j-1/2}, \quad b(x) = \frac{1}{1 + \alpha(x)}, \tag{7}$$

$$\varphi_j^i = x_{j-m} + b_{j-m-1/2} \left[\sum_{k=1}^{m} \left(\frac{h}{b_{j+1/2-k}}\right) - q\tau\right], \tag{8}$$

where

$$b_{j-1/2} = \frac{1}{2}(b(x_j) + b(x_{j-1})), \quad h = x_j - x_{j-1},$$

$m \geq 0$ is the first integer for which

$$\sum_{k=1}^{m+1} \left(\frac{h}{b_{j+1/2-k}}\right) > q\tau.$$

They have suggested for choosing α_j^i to be

$$\alpha_j^i = \begin{cases} \frac{\psi(C^{i-1}(\varphi_j^i)) - \psi(C_j^{i-1})}{C^{i-1}(\varphi_j^i) - C_j^{i-1}}, & \text{if} \quad C_j^{i-1} \neq C_j^{i-1}(\varphi_j^i) \\ \psi'(C_j^{i-1}), & otherwise \end{cases} \tag{9}$$

When $\alpha = 0$ we get our scheme (3) which we call FVS scheme (field velocity scheme). In the degenerate case we regularize the function ψ as in (4) or substituting it by ψ_ε which differs only in $[0, \varepsilon]$, e.g. defined by $\psi_\varepsilon(s) = \psi(\varepsilon)/\varepsilon s$. Solution of nonlinear system has been avoided by the relaxation procedure using α as in (9). Then we need to interpolate the function $\psi(C^{i-1}(\varphi_j^i))$ in (9). This can be done by using the linear basis function or cubic spline interpolation.

2 Numerical implementation

The question is: which scheme we prefer? To answer this question we have done a rigorous practical analysis for these two schemes carefully for convection dominant problem (Similar comparison can be done for hyperbolic problem). We have used

$$\psi(C) = (C)^{0.5}, \quad q = 1, \quad D = 0.01.$$

We restrict our attention to the following points. These results were tested for different initial conditions and we have chosen an example for each point:

Firstly, it seems that the scheme (6) is not stable for hyperbolic (convection-dominant) problem, because accurate characteristic position needs to solve (6) iteratively and this leads the scheme (6) to be very expensive and also makes the scheme sensitive to the refinement in the space. For the convection- dominant problem the error of the nonaccurate characteristic position will exist in three places: in the convective part, in the degenerate part and in the diffusion part. While this situation does not exist for the scheme (3) because we have generally accurate characteristic position (we do not need to find it iteratively). In addition we add the justification by Douglas et al. [2] (see also [8]) for nonconstant coefficients problems, where the mass balance failure for Galerkin characteristics approach (Mass balance can be obtained by integral (1) over $(0, T)$ and the space) was used. Consequently, one can see that the scheme (3) has better reservation for the profile much more than the scheme (6) when we use smaller step size $h = \Delta x$ and the mass balance problem still arises. We have used the following initial condition

$$C(x, 0) = \begin{cases} 1, & x \in (0.25, 0.5) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

(see Figures (1, 2)).

Secondly, it seems that the mass balance problem for the scheme (3) is much more than in the scheme (5). Here the initial condition is

$$C(x, 0) = \begin{cases} 1, & x \in (0.25, 0.3) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

see Figure (3).

Acknowledgement

The first two authors gratefully acknowledge the Scientific Grant Agency (VEGA) of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences for supporting this work under the Grants No. 2/6169/26 and No. 1/0867/08.

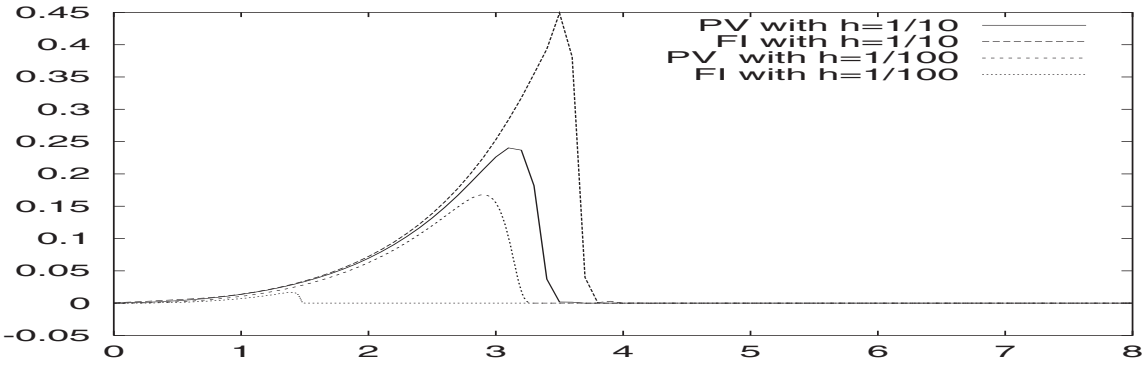


Figure 1: FI and PV schemes at t=3 hours for different space discretizations and $\Delta t = 0.05$

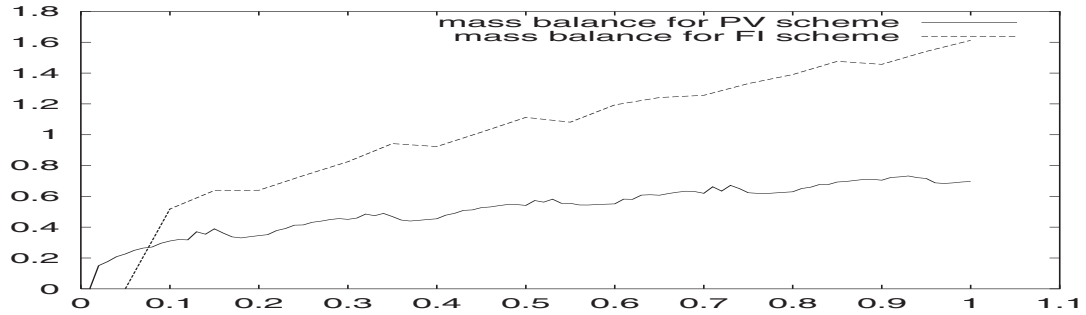


Figure 2: Mass balance problem for FI and PV schemes at t=1 hour, $\Delta x = 1/20$ and $\Delta t = 0.1$

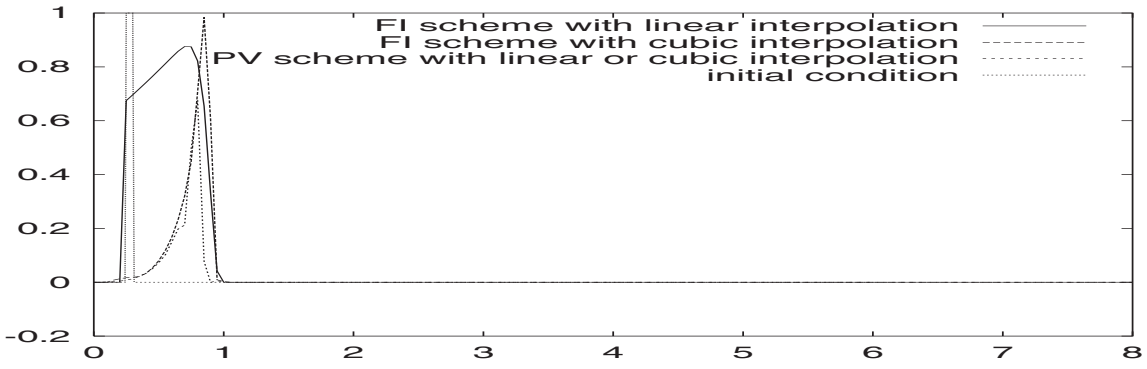


Figure 3: Mass balance profiles for FI and PV scheme

The third author was supported by the Scientific Grant Agency (VEGA) of the Ministry of Education of Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) 1/0843/08 (21).

References

[1] BERMEJO, R.:*A Galerkin-Characteristics algorithm for transport-diffusion equation*. SIAM J. Numer. Annal. 32, pp. 425-455, 1995.

[2] DOUGLAS, J., HUANG, C.-S., F. Pereira, F.: *The modified method of characteristics with adjusted advection*. Numer. Math. 83, pp. 353-369, 1999.

[3] JÄGER, W., KAČUR, J.: *Solution of porous medium systems by linear approximation scheme*. Num. Math. 60, pp. 407-427, 1991.

[4] KAČUR, J., MAHMOOD, M.:*Solution of Solute transport in unsaturated porous media by the method of characteristics*. Numer. Methods Partial Differential Eq. 19, pp. 732-761, 2003.

[5] KAČUR, J., MAHMOOD, M.:*Solution of convection-diffusion problems with memory terms by the method of characteristics*. Inter. J. Numerical Analysis and Modelling, Vol. 6, No. 1, 2009.

[6] KNABNER, P., BARRETT, J., KAPPMEIER, H.: *Lagrange-Galerkin approximation for advection-dominant nonlinear contaminant transport in porous media*. Computational methods in water resources X, Vol 1, A. Norwell, Ma, Kluwer Academic Publishers, pp. 299-307, 1994.

[7] MAHMOOD, M.:*The regularized Galerkin characteristics algorithm for contaminant transport with equilibrium and non-equilibrium adsorption*. PhD. thesis, Comenius University, March 2002.

[8] MAHMOOD, M.: *Analysis of Galerkin-characteristics algorithm for variably saturated flow*. Inter. J. Comp. Math., Volume 85 Issue 3, 509, 2008.

Current address

Branislav Ftorek, Mgr., PhD.
Department of Appl. Mathematics, Faculty of Mechanical Engineering
University of Žilina, J. M. Hurbana 15, 010 26 Žilina, Slovakia
branislav.ftorek@fstroj.uniza.sk

Peter Tomašovič, Ing.
Department of Appl. Mathematics, Faculty of Mechanical Engineering
University of Žilina, J. M. Hurbana 15, 010 26 Žilina, Slovakia
peter.tomasovic@fstroj.uniza.sk

Božena Dorociaková, RNDr., PhD.
Department of Appl. Mathematics, Faculty of Mechanical Engineering
University of Žilina, J. M. Hurbana 15, 010 26 Žilina, Slovakia
bozena.dorociakova@fstroj.uniza.sk

Journal of Applied Mathematics

volume II (2009), number III

SOLUTION OF TORSION OF PRISMATIC BAR WITH TRIANGULAR CROSS SECTION AREA USING PROGRAM *MATHEMATICA*

JANČO Roland, (SK), KOVÁČOVÁ Monika, (SK)

Abstract. In real design of bar and beam which is load by torque we need properties of cross section area. No all time you have circular cross section area in real problems. For solution of non-circular cross section area we used Saint-Venant's principle. In this paper is short introduction how to used Saint-Venant's principle to solution of triangular cross section area. Theoretical solutions for triangular cross section are compared by numerical solution solved in program Mathematica with package Structural Mechanics.

Key words and phrases. Torsion, Saint-Venant's principle, triangular cross section.

Mathematics Subject Classification. Primary 74A10, 74B05, 74G50 ; Secondary 74K10.

1 Introduction

Because many engineering structures, such as beams, shafts, and airplane wings, are subjected to torsional forces, the torsional problem has been of practical importance in structural analysis for a long time. Saint-Venant (1885) was the first to provide the correct solution to the problem of torsion of bars subjected to moment couples at the ends. He made certain assumptions about the deformation of the twisted bar, and then showed that his solutions satisfied the equations of equilibrium and the boundary conditions. From the uniqueness of solutions of the elasticity equations, it follows that the assumed forms for the displacements are the exact solutions to the torsional problem. The Saint-Venant principe is adopted in Structural Mechanics packages.

In this paper is contains of theoretical background of solution triangular cross-section area properties for torsional problems using Saint-Venant principle and comparison of theoretical solution with solution from Structural Mechanics package in Mathematica.

2 Theoretical background

If bar is loaded by equal and opposite torques T on its ends, we anticipate that the relative rigid-body displacement of initially plane section will consist of rotation, leading to a twist per unit length ϑ. These sections may also deform out of plane, but this deformation must by same for all values of z. These kinematic considerations lead to the candidate displacement field

$$u_x = -\vartheta\, z\, y; \qquad u_y = \vartheta\, z\, x; \qquad u_z = \vartheta\, f(x,\, y)\,, \tag{1}$$

where f is an unknown function of x,y describing the out-of-plane deformation.

Substituting these kinematic consideration into the strain-displacement relations $e_{ij} = \frac{1}{2}\left(\frac{\partial\, u_i}{\partial\, x_j} + \frac{\partial\, u_j}{\partial\, x_i}\right)$ yields

$$e_{xy} = 0; \qquad e_{zx} = \frac{\vartheta}{2}\left(\frac{\partial\, f}{\partial\, x} - y\right); \qquad e_{zy} = \left(\frac{\partial\, f}{\partial\, y} + x\right) \tag{2}$$

and it follow from Hooke's law in form $\sigma_{ij} = \lambda\, e_{mm}\, \delta_{ij} + 2\,\mu\, e_{ij}$ [1] that

$$\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = 0 \tag{3}$$

and

$$\sigma_{xy} = 0; \qquad \sigma_{zx} = \mu\vartheta\left(\frac{\partial\, f}{\partial\, x} - y\right); \qquad \sigma_{zy} = \mu\vartheta\left(\frac{\partial\, f}{\partial\, y} + x\right). \tag{4}$$

There are no body forces, so substitution into the equilibrium equations $\frac{\partial\, \sigma_{ij}}{\partial\, x_j} + P_i = 0$ from [1] yields

$$\nabla^2 f = 0. \tag{5}$$

The torsion problem is therefore reduced to the determination of harmonic function f such that the stresses (4) satisfy the traction-free condition on the curved surfaces of the bar. The twist per unit length ϑ can be determined by evaluating the torque on the cross-section Ω

$$T = \int\int_{\Omega} (x\, \sigma_{zy} - y\, \sigma_{zx}) \mathrm{d}x\, \mathrm{d}y\,. \tag{6}$$

2.1 The bar of triangular cross-section area

For solution of the rectangular bar we used Prandtl's stress function defined by

$$\boldsymbol{\tau} \equiv \mathbf{i}\, \sigma_{zx} + \mathbf{j}\sigma_{zy} = \mathrm{curl}\, \mathbf{k}\phi \tag{7}$$

or

$$\sigma_{zx} = \frac{\partial\, \phi}{\partial\, y}; \qquad \sigma_{zy} = -\frac{\partial\, \phi}{\partial\, x}\,. \tag{8}$$

With this representation, the traction-free boundary condition can be written

$$\boldsymbol{\tau}.\mathbf{n} = \sigma_{zn} = \frac{\partial \phi}{\partial t} = 0 \tag{9}$$

where $\mathbf{n}$ is the local normal to the boundary of Ω and n, t are a corresponding set of local orthogonal coordinates respectively normal and tangential to the boundary, Thus ϕ must be constant around the boundary and for simply-connected bodies this constant can be taken as zero without loss of generality giving the simple condition

$$\phi = 0 \tag{10}$$

on the boundary and from this boundary condition we obtain

$$\nabla^2 \phi = -2\,\mu\,\vartheta\,. \tag{11}$$

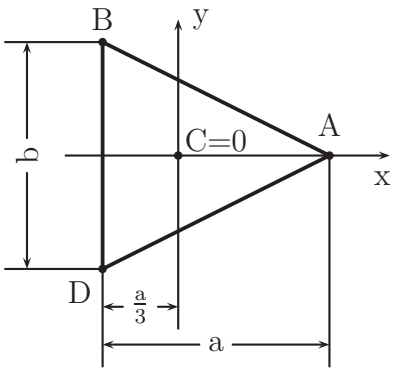


Figure 1: Definition of cross section area

This results can be used to obtain an approximate solution for torsion of the triangular bar $-\frac{a}{3} < x < \frac{2a}{3}$, $-\frac{b}{2} < y < \frac{b}{2}$, where $b = \frac{2a}{\sqrt{3}}$.

The stress function (10) satisfied the governing equation (11) and the boundary condition on the AB lines

$$\phi = 0; \qquad y = -\frac{\mathrm{b}}{2\mathrm{a}}x + \frac{\mathrm{b}}{3}\,, \tag{12}$$

on the line DA

$$\phi = 0; \qquad x = \frac{\mathrm{b}}{2\mathrm{a}}x - \frac{\mathrm{b}}{3}\,, \tag{13}$$

but it does not satisfy the corresponding boundary condition on the line BD

$$\phi = 0; \qquad x = -\frac{\mathrm{a}}{3}\,. \tag{14}$$

Stress function, which satisfied the both boundary conditions (12), (13) and (14) is

$$\phi = C\left[\left(x + \frac{\mathrm{a}}{3}\right)\left(x + \sqrt{3} - \frac{2}{3}\mathrm{a}\right)\left(x - \sqrt{3} - \frac{2}{3}\mathrm{a}\right)\right], \tag{15}$$

where C is a arbitrary constant. When we put stress function (15) to (10), we get

$$\nabla^2 \phi \;=\; 2\,C \;=\; -2\,\mu\,\vartheta \tag{16}$$

Integrating eqn. (6) by parts and using the fact that $\phi = 0$ on the boundary of Ω, we obtain the simple expression

$$T \;=\; 2 \int\int_{\Omega} \phi \mathrm{d}x\,\mathrm{d}y\,. \tag{17}$$

The transmitted torque is obtained from equations (17, 15) as

$$T \;=\; 2 \int_{-\frac{a}{3}}^{\frac{2a}{3}} \int_{\frac{b}{2a}x-\frac{b}{3}}^{-\frac{b}{2a}x+\frac{b}{3}} C \left[\frac{1}{2}\left(x^2 \,+\, y^2\right) \,-\, \frac{1}{2a}\left(x^3 - 3xy^2\right) \,-\, \frac{2}{27}a^2\right] \mathrm{d}x\,\mathrm{d}y\,. \tag{18}$$

After integration of eqn. 18, we get

$$T \;=\; -\frac{a^4}{15\,\sqrt{3}} C\,. \tag{19}$$

From eqn. (16) and eqn. (19), we have

$$C \;=\; -\frac{15\,\sqrt{3}}{a^4} T \qquad\qquad \text{and} \qquad\qquad \vartheta \;=\; -\frac{C}{\mu} \;=\; \frac{15\,\sqrt{3}}{a^4 \mu} T\,. \tag{20}$$

The torsional rigidity of the section K is generally defined $K \;=\; \mu\, J_K$ such that

$$T \;=\; \mu\,\vartheta\, J_K \tag{21}$$

The torsional rigidity of triangular section K using eqn. (20) and eqn. (21) is

$$K \;=\; \mu\, J_K \;=\; \mu\, \frac{a^4}{15\,\sqrt{3}} \;=\; \mu\, \frac{3\sqrt{3}\,b^4}{240} \tag{22}$$

The maximum shear stress occurs at the point $\left(-\frac{a}{3},\, 0\right)$ and is

$$\tau_{max} \;=\; \left|\frac{\partial\,\phi}{\partial\,x}\left(-\frac{a}{3},\, 0\right)\right| \;=\; \left|-\frac{15\sqrt{3}T}{a^4}\left[x \,-\, \frac{3x^2 - 3y^2}{2a}\right]\right| \;=\; \frac{15\sqrt{3}T}{2a^3} \;=\; \frac{20T}{b^3}\,. \tag{23}$$

2.2 Equation for solution the equilateral triangle bar in Mechanics of Materials

In mechanics of material [3] we used for solution of maximum shearing stress the equation

$$\tau_{max} = \frac{20\,T}{b^3} \tag{24}$$

and angle of twist is defined by

$$\varphi \;=\; \frac{46\,T\,\ell}{b^4\,G}\,, \tag{25}$$

where ℓ is the length of bar, G is the shear modulus of elasticity for the material.

3 Solution of torsion in program *MATHEMATICA*

For solution of torsional problems in program *MATHEMATICA* we were used the package "Structural Mechanics", which consist of solution the following types of cross sections:

- circular cross sections
- elliptical cross sections
- rectangular cross sections
- equilateral-triangular cross sections
- sectorial-type cross sections
- semicircular cross sections

3.1 Triangular Cross Section - Graphically

First, generate the graphical representation of the equilateral triangle with height a using the coordinates of the vertices given in points. Here plot the shape of the cross section for $a = 1$ in the Fig. 2. In the Fig. 3 is displayed of cross section of area.

```
In[1]:= << StructuralMechanics`

In[2]:= $DefaultFont = {"Courier", 7};

In[3]:= points = {{-a / 3, -a / Sqrt[3]},
            {-a / 3, a / Sqrt[3]}, {2 a / 3, 0}, {-a / 3, -a / Sqrt[3]}} /. a → 1;
```

Figure 2: Definition of cross section area

Visualization of equilateral-triangular beam twisted by $\theta = \pi/(4\,\ell)$ is in the Fig. 4, where the beam length is $\ell_z = 7$. The end of the beam at $z = 7$ is rotated by $\pi/4$ with respect to the root cross section at $z = 0$. In the Fig.5 is shown how the section at z = 1 is rotated after the force is applied.

3.2 Torsional Analysis Functions for Rectangular Cross Sections

In package `Structural Mechanics` you can calculate the analytical equation for rectangular cross section which is including. Definition of torsional rigidity in Mechanics of Materials [3] is $G\,J_k$ for triangular cross section $G\,b^4/\,46 = a^4\,G/(15\sqrt{3})$, which is defined by equation 25. Result from program *Mathematica* is in Fig. 6 which is same from theoretical solution, see eqn. 22.

To avoid replacing x, y, and z in the stress components, use the following replacements for only the right-hand side of the replacement rules. Result is shown in Fig. 7.

```
In[5]:= ss = CrossSectionPlot[points, PlotRange → {{-1, 1}, {-1, 1}},
            Frame → True, DefaultFont → {"Courier", 12}, Axes → None];
```

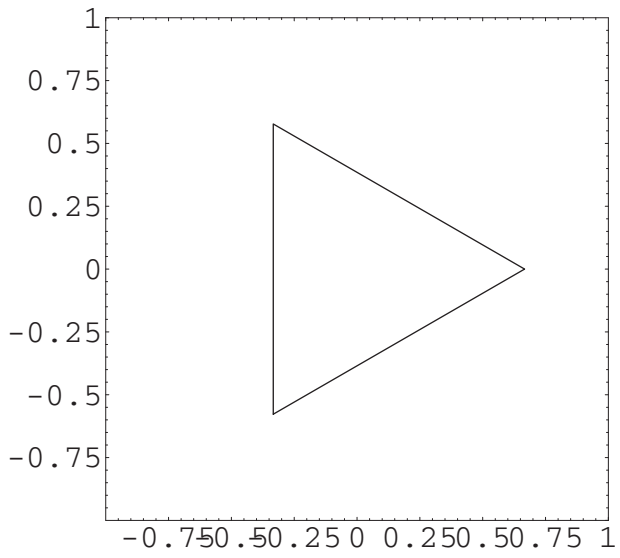


Figure 3: Quilateral triangle cross section area

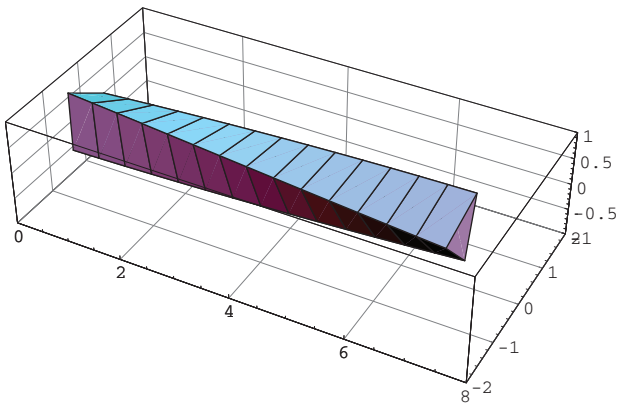


Figure 4: Deformation of triangular bar

Next extract the traction components T_{xz}, T_{yz} from the stress tensor str, see Fig. 8.

In the Fig. 9 is the stress component σ_{yz} a varies along the x axis from $-\frac{1}{3}$ to $\frac{2}{3}$

Similarly, you can obtain a plot showing (Fig. 10) how the value of σ_{yz} a changes along the y axis.

```
In[23]:= RotatedCrossSection[points /. a → 1, 1, π / 28, DefaultFont → {"Courier", 12}];
```

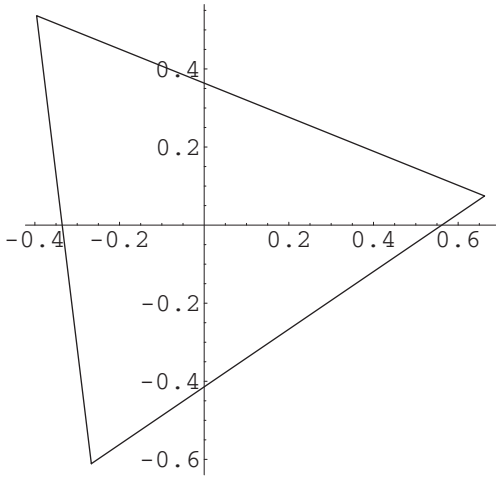


Figure 5: Twisting of cross section area

```
In[7]:= Clear[a, G]
        TorsionalRigidity[EquilateralTriangle, {a}, G]

Out[8]= a^4 G / (15 √3)
```

Figure 6: Torsional rigidity K

```
In[9]:= str = TorsionalStresses[EquilateralTriangle, {a}, G, θ, {x, y}]

Out[9]= {0, 0, 0, 0, -G (y + 3 x y / a) θ, G (x - (3 x^2 - 3 y^2) / (2 a)) θ}
```

Figure 7: Stress vector

```
In[18]:= {Txz, Tyz} = {σxz, σyz} /. str

Out[18]= {-y - 3 x y, x + 1/2 (-3 x^2 + 3 y^2)}
```

Figure 8: Selected component from stress vector

4 Conclusion

In this paper is presented the theoretical solution of torsion properties for triangular cross section area of bar loading by torque. This properties was derived using Saint-Venant's principle. Result is presented by equation (22) and (23). In program *MATHEMATICA* was implemented Saint-Venant's principle in package `Structural Mechanics`. This package include two way solution of non-circular cross section area, first way is graphically and second way is solution of analytically. Both this way is described in this paper. When we compare theoretical solution derived in this paper by numerical solution using package `Structural Mechanics`, the results are identically.

```
In[20]:= Plot[Tyz /. y -> 0, {x, -1 / 3, 2 / 3},
           PlotRange -> All,
          Frame -> True, DefaultFont → {"Courier", 12},
           AxesLabel -> {"x", ""}];
```

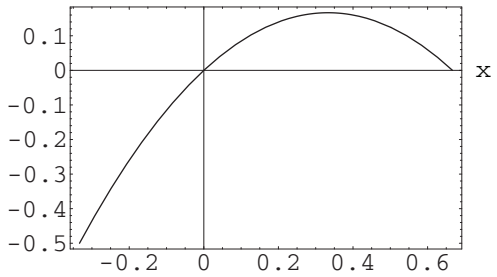


Figure 9: The stress component σ_{yz} a varies along the x axis

```
In[21]:= Plot[Tyz /. x -> -1 / 3, {y, -1 / 2, 1 / 2},
           PlotRange -> {All, {.0, -.6}},
           Frame -> True , DefaultFont → {"Courier", 12},
           AxesLabel -> {"y", " "}];
```

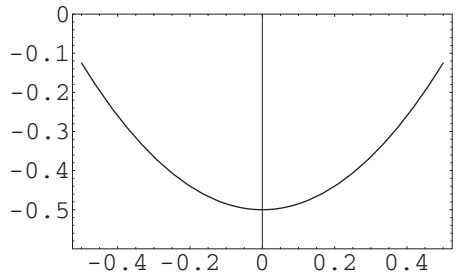


Figure 10: The stress component σ_{yz} a varies along the y axis

Acknowledgement

The paper was supported by grant from Grant Agency of VEGA no. 1/4106/07.

References

[1] HOLZAPFEL, G.A.: *Nonlinear Solid Mechanics: A Continuum Approach for Engineering.* Wiley,2004.

[2] BARBER, J.R.: *Elasticity*, Second Edition, Series: Solid Mechanics and Its Applications, Vol. 107 Springer Verlag, Berlin, 2003.

[3] HIBBELER, R.C.: *Mechanics of Materials*, SI Second Edition, Prentice Hall, Singapore, 2005.

Current address

MSc. Roland Jančo, PhD. ING-PAED IGIP
Institute of Applied Mechanics and Mechatronics, Section of Strength of Material, Faculty of Mechanical Engineering, Slovak University of Technology Bratislava, Nám. slobody 17, 812 31 Bratislava, Slovak Republic, e-mail: roland.janco@stuba.sk .

Mgr. Monika Kováčová, PhD.
Institute of natural sciences, humanities and social sciences, Faculty of Mechanical Engineering, Slovak University of Technology Bratislava, Nám. slobody 17, 812 31 Bratislava, Slovensk Republika, e-mail: monika.kovacova@stuba.sk .

Journal of Applied Mathematics

volume II (2009), number III

RELIABILITY LIKELIHOOD RATIO CONFIDENCE BOUNDS

KOVÁČOVÁ Monika, (SK)

Abstract. One of the most confusing concepts to a novice reliability engineer is estimating the precision of an estimate. This is an important concept in the field of reliability engineering, leading to the use of confidence intervals (or bounds). In this paper, we will try to briefly present the concept in relatively simple terms but based on solid common sense.[5]
We will present one method for calculating confidence bounds thorough the likelihood ratio bounds (LRB) method. Conceptually, this method is a great deal simpler than that of the Fisher matrix; although that does not mean that the results are of any less value. In fact, the LRB method is often preferred over the FM method in situations where there are smaller sample sizes.
We will show in this paper how to use computing software *MATHEMATICA* for calculating confidence bounds.

Key words and phrases. confidence bounds, computing software *MATHEMATICA*, reliability

Mathematics Subject Classification. 60K10, 62F25

1. Introduction

We will display some basic ideas of parametric maximum likelihood methods and their usability in reliability engineering. First we will show how to construct the likelihood (probability of the data) function and the basic ideas behind using this function to estimate a parameter will be explain. Due to usability these methods for the engineers we will describe the progress in computation in programming system *MATHEMATICA*. We will compute one of three possible different situations – methods for computing confidence intervals for parameters and functions of parameters.

As is known, parametric distribution, e.g. widely known Weibull distribution, when used appropriately, can provide a simple, versatile, visually appealing failure-time model. Maximum likelihood is perhaps the most versatile method for fitting statistical models to data. The appeal of maximum likelihood estimation stems from the fact that it can be applied to a wide variety of statistical models and kinds of data (e.g. continuous, discrete, categorical, censored, truncated) where other popular methods, like least squares, are not, in general, satisfactory.
In typical applications, the goal is to use a parametric statistical model to describe a set of data or a process or population that generated a set of data. Modern computing software had expanded the

feasible areas of application for maximum likelihood methods and allowed engineers to compute that estimations in hand without specific software need.

Statistical theory shows that, under standard regularity conditions, maximum likelihood estimators are "optimal" in large samples. More specifically, this means that maximum likelihood are consistent and asymptotically (as the sample size increased) efficient. That is, among consistent competitors to maximum likelihood estimators, none has a smaller asymptotic variance.

Statistical modeling, in practice, is an iterative procedure of fitting proposed models in search of a model that provides an adequate description of the population or process of interest, without being unnecessarily complicated. Application of maximum likelihood methods generally starts with a set of data and a tentative statistical model for the data. The tentative model is often suggested by the initial graphical analysis, physical theory, previous experience with similar data, or other expert knowledge.

2. Parametric Likelihood in General

Using a parametric model of the observed data the likelihood function can be viewed as the probability of the observed data, written as a function of the model's parameters. For a parametric model, the number of parameters is usually small relative to the nonparametric models. Most popular are Exponential and Weibull distributions in reliability engineering.

The exponential distribution has only one parameter and its hazard function is constant. Hazard function of this distribution sometimes mean failure rate. The Weibull distribution has a hazard rate function that is not constant over time. It is used to model for both increasing and decreasing failure rates. Weibull is a distribution that can be used to model a wide range of phenomena. It is used to model the infant mortality or the wear-out period in the bathtub curve. Weibull analysis is extensively used to study mechanical, chemical, electrical, electronic, material, and human failures. The primary advantages of Weibull analysis are its ability to

- Provide moderately accurate failure analysis and failure forecasts with extremely small data samples, making solution possible at the earliest indications of a problem.
- Provide simple and useful graphical plots for individual failure modes that can be easily interpreted and understood, even when data inadequacies exist.
- Represent a broad range of distribution shapes so that the distribution with the best fit can be selected.
- Provide physics of failure clues based on the slope of the Weibull probability plot.

Although the use of normal or lognormal distribution generally requires at least 20 failures or knowledge from prior existence, Weibull analysis works extremely well when there are as few as 3-5 failures, which is critical when the result of a failure involves safety or extreme costs. Parameters for the Weibull distribution for most components can be found in many references [4].

For a set of n independent observations, the likelihood function can be written as the following joint probability

$$L(\theta) = L(\theta, data) = c\prod_{i=1}^{n} L_i(\theta, data_i) \tag{1}$$

The quantity c in term (1) is a constant term that does not depend on the data or on θ (in general θ be a vector). For computation purposes, let $c = 1$. In reliability engineering, if a failure time is

known to have occurred between times t_{i-1} and t_i, the probability of this event is contribution terms $L_i(\theta, data_i)$

$$L_i(\theta, data_i) = L_i(\theta) = \int_{t_{i-1}}^{t_i} f(t,\theta)dt = F(t_i,\theta) - F(t_{i-1},\theta) \quad (2)$$

For a given set of data, $L(\theta)$ can be viewed as a function of θ. The dependence of $L(\theta)$ on the data will be understood and is usually suppressed in notation. The values of θ for which $L(\theta)$ is relatively large are more plausible than values of θ for which the probability of the data is relatively small. There may or may not be a unique value of θ that maximizes $L(\theta)$. Regions in the space of θ with relatively large $L(\theta)$ can be used to define confidence regions for θ. One can also use maximum likelihood to estimate functions of θ. We will show how to make these concepts operational for the two parameter Weibull dinstribution, using simple examples for illustration.

2.1. Likelihood Function and its Maximum

Given a sample of n independent observations, denoted generically by $data_i$, $i = 1,...,n$ and a specific model, the total likelihood $L(\theta)$ for the sample is given by equation (1). For some purposes, it is convenient to use the log lokelihood $\tilde{L}_i(\theta) = \log\left[L(\theta)\right]$. For all practical problems $\tilde{L}(\theta)$ will be representable in computer memory without special scaling [which is not so for $L(\theta)$] because of possible extreme exponent values, and some theory for maximum likelihood is developed more naturally in terms of sums like

$$\tilde{L}_i(\theta) = \log\left[L(\theta)\right] = \sum_{i=1}^{n} \tilde{L}_i(\theta),$$

rather than in terms of the product in equation (1). Note that the maximum of $\tilde{L}(\theta)$, if one exists, occurs at the same value of θ as the maximum of $L(\theta)$.

2.2. Likelihood Confidence Intervals for θ

The likelihood function provides a versatile method for assessing the information that the data contains on parameter, or functions of parameters. Specifically, the likelihood function provides a generally useful method for finding approximate confidence intervals for parameters and functions of parameters.
An approximate $100(1-\alpha)\%$ likelihood-based confidence interval for θ is the set of all values of θ such that

$$-2\log\left[R(\theta)\right] \le \chi^2_{(1-\alpha,1)}$$

or, equivalently, the set defined by

$$R(\theta) \ge \exp\left[-\chi^2_{(1-\alpha,1)}/2\right].$$

The theoretical justification for this interval would be explained as follows.

Assume that we want to estimate θ_1, from the partition $\theta=(\theta_1,\theta_2)$. Let r_1 denote the length of θ_1 (θ_1 means in general vector). The profile likelihood for θ_1 is

$$R(\theta_1)=\max\left[\frac{L(\theta_1,\theta_2)}{L(\hat{\theta})}\right]. \tag{3}$$

When the length of θ_2 is 0 (as in one-parameter exponential distribution) is a relative likelihood for $\theta=\theta_1$. Otherwise we have a "maximized relative likelihood" for θ_1. In either case, $R(\theta_1)$ is commonly known as a "profile likelihood" because it provides a view of the profile of $L(\theta)$ as viewed along a line that is perpendicular to the axes of θ_1.

- When θ_1 is of length 1, $R(\theta_1)$ is a curve projected onto plane.
- When θ_1 is of length 2 or more, $R(\theta_1)$ is a surface projected onto a three-dimensional hyper-plane.

In either case the projection is in a direction perpendicular to the coordinate axes for θ_1. When θ_1 is of length 1 or 2, it is useful to display $R(\theta_1)$ graphically.

Asymptotically, $LLR_n(\theta_1)=-2\log\left[R(\theta_1)\right]$ when evaluated at the true θ_1, has a chi-square distribution with r_1 degrees of freedom. To do a likelihood-ration significance test, we would reject the null hypothesis that $\theta=\theta_0$, at the α level of significance, if

$$LLR_n(\theta_1)=-2\log\left[R(\theta)\right]>\chi^2_{(1-\alpha,r_1)}.$$

So one-sided approximate $100(1-\alpha)\%$ confidence bound can be obtained by drawing the horizontal line at $\exp\left[-\chi^2_{(1-2\alpha,1)}/2\right]$ and using the appropriate endpoint of the resulting two-sided confidence interval.

2.3. Relationship between Confidence Intervals and Significance Tests

Significance testing (or sometimes called hypothesis testing) is a statistical technique widely used in many areas of science. The basic idea is to assess the reasonableness of a claim or hypothesis about a model or parameter value, relative to observed data. One can test a hypothesis by first constructing a $100(1-\alpha)\%$ confidence interval for the quantity of interest and then checking to see if the interval encloses the hypothesized value or not. If not, then the hypothesis is rejected "at the α level of significance". If the interval encloses the hypothesized value, then the appropriate conclusion is that the data are consistent with the hypothesis (it is important, however, to note that failing to reject a hypothesis is not the same as saying that the hypothesis is true).

Most practitioners find confidence intervals much more informative than yes/no result of an significance test. See e.g. Hahn and Meeker (1991) and other references for further discussion of this subject.

To be more formal, a likelihood ratio test for a single parameter model can be done by comparing the maximum of the likelihood under the "null hypothesis" to the maximum of the likelihood over all possible values for the parameter. A likelihood much smaller under the null hypothesis provides evidence to refute the hypothesis.

Specifically, for the Exponential distribution, the single-point null hypothesis $\theta=\theta_0$ should be rejected if

$$-2\log\left[\frac{L(\theta_0)}{L\left(\hat{\theta}\right)}\right] > \chi^2_{(1-\alpha,1)}, \tag{4}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ. Rejection implies that the data are not consistent with the null hypothesis. Using the previous definition from the section 2.2, it is easy to see that the a likelihood-based confidence interval is the set of all values of θ that would not be rejected under the likelihood ratio test defined in (4).

3. Likelihood Ratio Confidence Bounds

In the next section we will present a method for calculating confidence bounds via the likelihood ration bounds (LRB) method. To calculate confidence bounds for distributions parameters is a one of key questions in modeling MTBF. Conceptually, as was seen in the previous section, **this method is a great deal simpler than that of the Fisher matrix, although that does not mean that the results are of any less value. In fact, the LRB method is often preferred over the Fisher matrix method in situations where there are smaller sample sizes. Note, even through LRB bounds are conceptually simpler, they are computationally intensive and require a much longer time to plot than Fisher matrix bounds.**
It is need to note that especially for time failure data sets modeled by Weibull distribution function is the estimation for likelihood confidence bound has a great deal.

Likelihood ratio confidence bounds are based on the equation [2, 6, 7]:

$$-2.\log\left(\frac{L(\theta)}{L(\hat{\theta})}\right) \geq \chi^2_{\alpha,k} \tag{5}$$

where:

- $L(\theta)$ is the likelihood function for the unknown parameter vector θ
- $L(\hat{\theta})$ is the likelihood function for the unknown parameter vector $\hat{\theta}$
- $\chi^2_{\alpha,k}$ is the chi-squared statistic with probability α and k degrees of freedom, where k is the number of quantities jointly estimated.

Let x be a continuous random variable with pdf: $f(x,\theta_1,\theta_2,...,\theta_k)$, where $\theta_1,\theta_2,...,\theta_k$ are k unknown constant parameters that need to be estimated, one can conduct an experiment and obtain R independent observation $x_1,x_2,...,x_R$ which correspond in the case of life data analysis to failure times. The likelihood function is given by:

$$L(x_1,x_2,...,x_R\,|\theta_1,\theta_2,...\theta_k) = L = \prod_{i=1}^{R} f(x_i,\theta_1,\theta_2,...\theta_k) \qquad i = 1,2,...R$$

The maximum likelihood estimators (MLE) of $\theta_1,\theta_2,...,\theta_k$ are obtained by maximizing L. These are representing by the $L\left(\hat{\theta}\right)$ term in the denominator of the ratio in equation (5). Since the values of the data points are knows, and the values of the parameter estimates $\hat{\theta}$ have been calculated using MLE methods, the only unknown term in eq. (5) is the $L(\theta)$ term in the numerator of the ratio. It remains to find the values of the unknown parameter vector θ that satisfy eq. (5).

For distribution that have two parameters (such as two-parameter Weibull distribution.), the values of the parameters that satisfy this equation will change based on the desired confidence level δ , but at a given value δ there is only a certain region of values for θ_1, θ_2 for which eq. (5) holds true. **That region can be represented graphically as a contour plot.** The region of the contour plot essentially represents a cross-section of the likelihood function surface that satisfies the condition of eq. (5). If δ is the confidence level, then $a = \delta$ for two-sided bounds and $\alpha = 2\delta - 1$ for one-sided.

The bounds on the parameters are calculated by finding the extreme values of the contour plot on each axis for a given confidence level. Since each axis represents the possible values of a given parameter, the boundaries of the contour plot represent the extreme values of the parameter that satisfy

$$-2.\log\left(\frac{L(\theta_1,\theta_2)}{L(\hat{\theta}_1,\hat{\theta}_2)}\right) \geq \chi^2_{\alpha,k}$$

For two-parameter distributions, the contour plot will be a two dimensional plot.

This equation can be rewritten as

$$L(\theta_1,\theta_2) = L(\hat{\theta}_1,\hat{\theta}_2).e^{\frac{-\chi^2_{\alpha,1}}{2}} \tag{6}$$

The task now becomes to find the values of parameters θ_1 and θ_2 so that the equality in eq. (5) is satisfied. Unfortunatelly, there is no closed-form solution, thus these values must be arrived at numerically. One method of doing this is to hold one parameter constant and iterate on the other until an acceptable solution is reached. This can prove to be rather tricky, since thre will be two solutions for one parameter if the other is held constant. **In situations such as these, it is best to being the iterative calculation with values close to those of the MLE values**, so as to ensure that one is not attempting to perform calculations outside of the region of the contour plot where no solution exists.

Let us present now to computing the confidence bounds for our pseudo-random numbers generating via pseudorandom generator.

```
In[1]:= x = RandomReal[WeibullDistribution[1.5, 30], 5]

Out[1]= {25.3153, 6.02835, 13.9177, 44.6242, 41.471}
```

Five units were put on a reliability test and experienced failures at previous output. Assuming a Weibull distribution, the MLE parameter estimates are calculated (or taking from our previous example) to be $\hat{\beta} = 1.5$ and $\hat{\eta} = 30$. We want to calculate the 90 % two-sided confidence bounds on these parameters using the likelihood ratio method described in the previous section.

The first step is to calculate the likelihood function for the parameter estimates.

$$L(\hat{\beta},\hat{\eta}) = \prod_{i=1}^{N} f(x_i,\hat{\beta},\hat{\eta}) = \prod_{i=1}^{5} \frac{\hat{\beta}}{\hat{\eta}}.\left(\frac{x_i}{\hat{\eta}}\right)^{\hat{\beta}-1}.e^{-\left(\frac{x_i}{\hat{\eta}}\right)^{\hat{\beta}}}$$

$$L(\hat{\beta},\hat{\eta}) = \prod_{i=1}^{5} \frac{1.5}{30}.\left(\frac{x_i}{30}\right)^{1.5-1}.e^{-\left(\frac{x_i}{30}\right)^{1.5}} = 5.39525\times10^{-9}$$

where x_i are the original time-to-failure data points.

```
In[2]:= f[β_, η_, T_] = PDF[WeibullDistribution[β, η], T]
```

$$\text{Out[2]=}\ \beta T^{\beta-1}\eta^{-\beta}e^{-\left(\frac{T}{\eta}\right)^{\beta}}$$

```
In[3]:= ls = Apply[Times, f[1.5, 30, x]]
```

$$\text{Out[3]=}\ 5.39525\times10^{-9}$$

```
In[4]:= ps = ls * Exp[-(InverseCDF[ChiDistribution[1], 0.9])^2 / 2]
```

$$\text{Out[4]=}\ 1.39479\times10^{-9}$$

We can now rearrange eq. (2) to the form

$$L(\beta,\eta)-L(\hat{\beta},\hat{\eta}).e^{\frac{-\chi^2_{\alpha,1}}{2}}=0$$

Since our specified confidence level, δ, is 90%, we can calculate the value of the chi-squared statistics $\chi^2_{0.9,1}=2.705543$. We then substitute this information into the equation

$$L(\beta,\eta)-5.39525\times10^{-9}.e^{\frac{-2.705543}{2}}=L(\beta,\eta)-1.39479\times10^{-9}=0$$

$$L(\beta,\eta)-1.39479\times10^{-9}=0$$

or in *MATHEMATICA*® notation

$$ff(\beta,\eta)-1.39479\times10^{-9}=0\ \text{(see the next picture)}$$

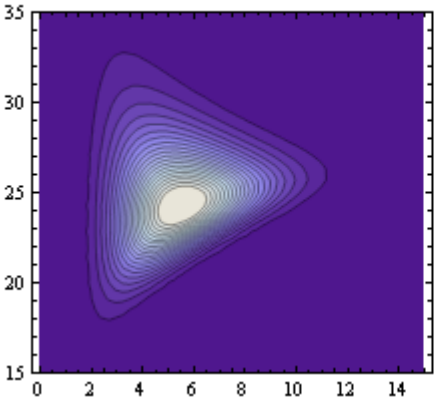
```
In[5]:= ff[β_, η_] = Apply[Times, f[β, η, x]] - ps
```

$$\text{Out[5]=}\ 5.16885\times10^{6\beta-1}\beta^5\eta^{-5\beta}$$
$$\exp\left(-15.9203^{\beta}\left(\frac{1}{\eta}\right)^{\beta}-20.7877^{\beta}\left(\frac{1}{\eta}\right)^{\beta}-22.7351^{\beta}\left(\frac{1}{\eta}\right)^{\beta}-23.1297^{\beta}\left(\frac{1}{\eta}\right)^{\beta}-29.7009^{\beta}\left(\frac{1}{\eta}\right)^{\beta}\right)-1.39479\times10^{-9}$$

The values of the parameters that satisfy this equation will change based on the desired confidence level δ. The region can be represented graphically as a Contour Plot *MATHEMATICA*® function.

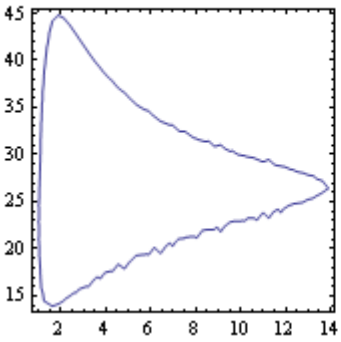
```
In[6]:= ContourPlot[ff[β, η], {β, 0, 15}, {η, 15, 35}, Contours → 20, PlotRange → All]
```

Out[6]=



```
In[7]:= ContourPlot[ff[β, η] == 0, {β, 0, 15}, {η, 5, 55}, Contours → 40, PlotRange → All]
```

Out[7]=



The solution is an iterative process, that reguires setting the value of η and finding the appopriate values of β.

```
In[8]:= zac = 14;
        kon = 44;
        body =
          Table[{η1, FindRoot[ff[β, η] /. η → η1, {β, 1}][[1, 2]],
            FindRoot[ff[β, η] /. η → η1, {β, 4}][[1, 2]]},
           {η1, zac, kon}];
        body // TableForm
```

```
In[12]:= TableForm[body[[1 ;; 15]],
           TableHeadings → {None, {"  η  ", "  min β", "  max β"}}]
```

Out[12]//TableForm=

η	min β	max β
14	1.41302	1.41302
15	1.20252	1.20252
16	1.13245	2.92151
17	1.09345	3.54451
18	1.06918	4.25105
19	1.05361	5.06153
20	1.04381	5.99341
21	1.03817	23.2302
22	1.03569	1.03569
23	1.03574	1.03574
24	1.0379	1.0379
25	1.04188	1.04188
26	1.04748	1.04748
27	1.05457	1.05457
28	1.06309	1.06309

```
In[13]:= TableForm[body[[16 ;; 30]],
           TableHeadings → {None, {"  η  ", "  min β", "  max β"}}]
```

Out[13]//TableForm=

η	min β	max β
29	1.073	11.3061
30	1.08429	−28.5507
31	1.09701	8.68849
32	1.11124	7.67967
33	1.12707	6.83797
34	1.14468	6.12968
35	1.16428	5.52679
36	1.18616	5.00731
37	1.21073	4.55415
38	1.23856	4.15388
39	1.27045	3.79569
40	1.30761	3.47053
41	1.35201	3.17026
42	1.40711	1.40711
43	1.48018	2.60774

As we can determined from the table, the lowest calculated value for β is 1.03569 while the highest is 23.2302 . These represent the two-sided 90 % confidence limits on this parameter. Since solution for the equation do not exist for values of η bellow 14 or above 44, these can be considere the 90 % confidence limits for this parameter.

Acknowledgement

The paper was supported by grant from Grant Agency of VEGA no. 1/4106/07.
General discussion about using the likelihood ratio method to calculate confidence bounds has been quoted based on articles written by ReliaSoft Corporation and published on www.weibull.com reliability engineering resource website.

References

[1.] KUMAR D. U.: Reliability, Maintenance and Logistic Support, Kluwer Academic Publishers, ISBN: 0-412-84240-8
[2.] MEEKER W. Q., ESCOBAR L. A.: Statistical Methods for Reliability Data, John Wiley & Sons, Inc., 1998, ISBN: 978 – 0 471 – 14328 - 4
[3.] WASSERMAN G.: Reliability, Verification, Testing and Analysis in Engineering Design, Marcel Dekker Inc., NY, 2003.
[4.] http://www.barringer1.com/wdbase.htm
[5.] http://www.weibull.com/hotwire/issue34/relbasics34.htm
[6.] http://www.weibull.com/hotwire/issue42/relbasic42.htm
[7.] http://www.weibull.com/hotwire/issue18/relbasics18.htm

Current address

Monika Kováčová, Mgr. PhD.
Institute of natural sciences, humanities and social sciences,
Faculty of Mechanical Engineering,
Slovak University of Technology Bratislava, Nám. slobody 17, 812 31 Bratislava,
Slovak Republic,
e-mail: monika.kovacova@stuba.sk .

Journal of Applied Mathematics

volume II (2009), number III

THE BEST LEAST ABSOLUTE DEVIATION LINEAR REGRESSION: PROPERTIES AND TWO EFFICIENT METHODS

Kuzmanović I., (HR), Sabo K., (HR), Scitovski R., (HR), Vazler I., (HR)

Abstract. For the given set of data, among which outliers (wild points) are expected, the problem of determining the best Least Absolute Deviations (LAD) linear regression is considered. Particularly, the problem of determining the best weighted LAD-line and the best LAD-plane is considered and efficient algorithms for solving these problems are given. Algorithms are illustrated by several examples as well as compared with other methods known in literature. The proposed methods proved to be sufficiently efficient for being considered as giving a solution in real time. Therefore, they are suitable for various applications, as e.g. in robotics.

Key words and phrases. Torsion, Saint-Venant's principle, triangular cross section.

Mathematics Subject Classification. 65D10, 65C20, 62J05, 90C27, 90C56, 90B85, 34K29

1 Introduction

The problem of determining parameters of the hyperplane, in order to have its graph passing as close as possible (in some sense) to the given points, is an old problem which has been solved in various ways. Most frequently it is assumed that the errors can occur only in measured values of the independent variable. In this case, if we use the l_2 norm, it is the *Ordinary Least Squares* (OLS) problem. In many technical and other applications (where the so-called "outliers" can be expected) using the l_1 norm is much more interesting. In literature this approach is better known as the *Least Absolute Deviations* (LAD) problem (see e. g. [3], [13], [16]). For example, while calculating movements of robots, based on the data obtained from a stereographic camera, it is important to estimate the position of a hyperplane efficiently and in real time (see e. g. [5]). At the same time, the so-called outliers among the data should not affect the obtained

results. It is known that this sort of problems can be solved by applying the LAD approach (see e. g. [4], [12]).

If one assumes that the errors can occur in measured values of both (dependent and independent) variables, and then in general the l_p norm is used, it is called the *Total-Least-p* (TLp) problem (see e. g. [1]).

In this paper we give an overview of basic properties and facts related to estimation of parameters of the best LAD-line and LAD-plane and propose an efficient method for searching optimal parameters. Our methods have been compared to numerous other methods known from literature, and have shown to be much more efficient.

At the beginning let us give an important lemma whose proof can be seen in [14], and which is used in various situations throughout the whole text.

Lemma 1.1 *Let (ω_i, y_i), $i \in I$, $I = \{1, \ldots, m\}$, $m \geq 2$, be the data, where $y_1 \leq y_2 \leq \ldots \leq y_m$ are real numbers, and $\omega_i > 0$ corresponding data weights. Denote*

$$\nu_0 = \max\{\nu \in I : \sum_{i=1}^{\nu} \omega_i - \frac{1}{2}\sum_{i=1}^{m} \omega_i \leq 0\}.$$

Furthermore, let $F : \mathbb{R} \to \mathbb{R}$ be a function defined by the formula

$$F(\alpha) = \sum_{i=1}^{m} \omega_i |y_i - \alpha|. \tag{1}$$

(i) If $\sum_{i=1}^{\nu_0} \omega_i < \frac{1}{2}\sum_{i=1}^{m} \omega_i$, then the minimum of the function F is attained at the point $\alpha^\star = y_{\nu_0+1}$.

(ii) If $\sum_{i=1}^{\nu_0} \omega_i = \frac{1}{2}\sum_{i=1}^{m} \omega_i$, then the minimum of the function F is attained at every point $\alpha^\star$ from the segment $[y_{\nu_0}, y_{\nu_0+1}]$.

Corollary 1.2 *Let $y_1 \leq y_2 \leq \ldots \leq y_m$, $m > 1$ be the data with weights $\omega_1 = \cdots = \omega_m = 1$. Then*

(i) if m is odd ($m = 2k + 1$), then the minimum of the function F is attained at the point $\alpha^\star = y_{k+1}$;

(ii) if m is even ($m = 2k$), the minimum of the function F is attained at every point $\alpha^\star$ from the segment $[y_k, y_{k+1}]$.

Remark 1.3 *If the minimum of the function F defined by (1) is attained in real number α^*, then*

$$F(\alpha) = \sum_{i=1}^{m} \omega_i |y_i - \alpha| \geq \sum_{i=1}^{m} \omega_i |y_i - \alpha^*|,$$

where the equality holds if and only if $\alpha = \alpha^*$.

2 The best weighted LAD-line

Let $I = \{1, \ldots, m\}$, $m \geq 2$ be a set of indices, $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 : i \in I\}$ a set of points in the plane, and $\omega_i > 0$ corresponding data weights. The best weighted LAD-line should be determined, i.e. we should determine optimal parameters $a^*, b^* \in \mathbb{R}$ of the function $f(x; a, b) = ax + b$ such that

$$G(a^*, b^*) = \min_{(a,b)\in\mathbb{R}^2} G(a, b), \qquad G(a, b) = \sum_{i=1}^{m} \omega_i |y_i - ax_i - b|. \tag{2}$$

Proof of the following existence problem is simple, and it can be proved by means of the principle applied in [8], [9].

Theorem 2.1 *Let $I = \{1, \ldots, m\}$, $m \geq 2$ be a set of indices, $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 : i \in I\}$ a set of points in the plane, and $\omega_i > 0$ corresponding data weights. Then there exists the best weighted LAD-line, i.e. the problem (2) has a solution. Especially, if $x_1 = \cdots = x_m =: \xi$, then there exist infinitely many best LAD-lines of the form $y = a^\star(x - \xi) + \mu$, whereby $a^\star$ is an arbitrary real number, and μ a weighted median of the data $y_1, \ldots, y_m$.*

The following lemma shows that for a linear function whose graph passes through some point $T_0(x_0, y_0) \in \mathbb{R}^2$ there is the best weighted LAD-line whose graph passes also through some point $T_i(x_i, y_i) \in \Lambda$, for which $x_i \neq x_0$. Especially, point T_0 can be one of the points of the set Λ. In that case, the lemma claims that there exists the best weighted LAD-line whose graph passes through one more point $T_i(x_i, y_i) \in \Lambda$, for which $x_i \neq x_0$.

Lemma 2.2 *Let $I = \{1, \ldots, m\}$, $m \geq 2$ be a set of indices and*

(i) $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 : i \in I\}$ *a set of points in the plane, such that* $\min\limits_{i\in I} x_i < \max\limits_{i\in I} x_i$.

(ii) $\omega_i > 0$, $i \in I$ *corresponding data weights,*

(iii) $T_0(x_0, y_0) \in \mathbb{R}^2$,

(iv) $\overline{f}(x; a) = a(x - x_0) + y_0$, *a linear function whose graph passes through the point* $T_0(x_0, y_0) \in \mathbb{R}^2$.

Then there exists $a^* \in \mathbb{R}$ such that

$$\overline{G}(a^*) = \min_{a\in\mathbb{R}} \overline{G}(a), \qquad \overline{G}(a) = \sum_{i=1}^{m} \omega_i |y_i - \overline{f}(x_i; a)| = \sum_{i=1}^{m} \omega_i |y_i - y_0 - a(x_i - x_0)|, \tag{3}$$

and the graph of the linear function $x \mapsto \overline{f}(x; a^*)$ passes through at least one more point $T_\nu(x_\nu, y_\nu) \in \Lambda$, whereby $x_\nu \neq x_0$.

Proof. Denote $I_0 = \{i \in I : x_i = x_0\}$. It can be seen that $I \setminus I_0 \neq \emptyset$, i.e. that there exists $i_0 \in I$, such that $x_{i_0} \neq x_0$. Otherwise it would be $x_i = x_0$, $\forall i \in I$, which contradicts assumption (i). Functional (3) can be written as

$$\overline{G}(a) = \sum_{i \in I_0} \omega_i |y_i - y_0| + \sum_{i \in I \setminus I_0} \omega_i |x_i - x_0| \left| \frac{y_i - y_0}{x_i - x_0} - a \right|. \tag{4}$$

According to *Lemma 1.1*, there exists $\nu \in I \setminus I_0$ such that

$$\overline{G}(a^*) = \min_{a \in \mathbb{R}} \overline{G}(a) = \sum_{i \in I_0} \omega_i |y_i - y_0| + \sum_{i \in I \setminus I_0} \omega_i |x_i - x_0| \left| \frac{y_i - y_0}{x_i - x_0} - \frac{y_\nu - y_0}{x_\nu - x_0} \right|,$$

from where there follows $a^* = \frac{y_\nu - y_0}{x_\nu - x_0}$.

It can be simply seen that the graph of function $x \mapsto \overline{f}(x; a^*) = \frac{y_\nu - y_0}{x_\nu - x_0}(x - x_0) + y_0$ passes through at least one point $T_\nu(x_\nu, y_\nu) \in \Lambda$, for which $x_\nu \neq x_0$. □

Theorem 2.3 *Let $I = \{1, \ldots, m\}$, $m \geq 2$ be a set of indices, $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 : i \in I\}$ a set of points in the plane, such that $\min_{i \in I} x_i < \max_{i \in I} x_i$, and $\omega_i > 0$ corresponding data weights. Then there exists the best weighted LAD-line which passes through at least two different points from Λ.*

Proof. According to *Theorem 2.1*, there exists the best weighted LAD-line $f(x, a^*, b^*) = a^* x + b^*$ with optimal parameters a^*, b^*.

According to*Lemma 1.1*, there exists $\mu \in I$, such that the minimum of the functional

$$G(a^*, b) = \sum_{i=1}^{m} \omega_i |y_i - a^* x_i - b|,$$

is attained for $b^+ = y_\mu - a^* x_\mu$, i.e.

$$G(a^*, b^*) = \min_{b \in \mathbb{R}} G(a^*, b) = G(a^*, b^+).$$

It means that also at the point (a^*, b^+) functional G attains its global minimum, and since $y_\mu = a^* x_\mu + b^+$, it means that there exists the best weighted LAD-line passing through the point $T_\mu(x_\mu, y_\mu)$, given by

$$\overline{f}(x; a^*) := f(x; a^*, y_\mu - a^* x_\mu) = a^*(x - x_\mu) + y_\mu.$$

According to *Lemma 2.2*, then there exists at least one more point from the set $\Lambda \setminus \{T_\mu\}$, through which there passes the best weighted LAD-line. □

Remark 2.4 *Notice that* Theorem 2.3 *is proved by using* Lemma 2.2*, whereas* Lemma 2.2 *is proved by applying* Lemma 1.1*. Thus, generally the best weighted LAD-line does not have to be unique. See also e. g. [11], [16].*

2.1 Methods for searching the best LAD-line

For solving this problem we can use general minimization methods without derivatives, e.g.: *Differential Evolution, Nelder Mead, Random Search, Simulated Annealing*, etc. (see e. g. [10]). There is also a certain number of methods which are based on Linear Programming (see e. g. [2], [13]) or they can be different special cases of the Gauss-Newton method (see e. g. [6], [13], [16]). Moreover, there are some methods specialized for solving this problem (see e. g. [12], [18]).

2.1.1 Two Points Method

For the purpose of developing the algorithm for searching the best LAD-line, first note that *Theorem 2.3* refers to the fact that the best LAD-line should be searched for among those lines which pass through at least two different points of the set Λ. The following algorithm is based on that fact. To get as close as possible from the beginning to the best LAD-line, as the initial point we can choose the centroid of the data $T_p(x_p, y_p)$, where

$$x_p = \frac{1}{\omega}\sum_{i=1}^{m}\omega_i x_i, \quad y_p = \frac{1}{\omega}\sum_{i=1}^{m}\omega_i y_i, \qquad \omega = \sum_{i=1}^{m}\omega_i, \tag{5}$$

which are quickly calculated, and probably pass closely to the best LAD-line. After that, in accordance with *Lemma 2.2*, the following algorithm is developed.

Algorithm [Two Points]

Step 1: Give $m \geq 2$, input points $T_i(x_i, y_i)$, $i \in I$, $I = \{1, \ldots, m\}$, and corresponding data weights $\omega_i > 0$; According to (5), determine the point $T_p(x_p, y_p)$ and define the set $I_0 = \{i \in I : x_i = x_p\}$;

Step 2: In accordance with *Lemma 2.2*,

- solve the LAD problem for the function $\overline{f}(x; a) = a(x - x_p) + y_p$ by minimizing functional (4), denote the solution by a_1 and determine a new point $T_{i_1}(x_{i_1}, y_{i_1}) \in \Lambda$ for which $x_{i_1} \neq x_p$;
- Put $b_1 = -a_1 x_p - y_p$ and calculate $G_1 = \sum_{i=1}^{m} \omega_i |y_i - a_1 x_i - b_1|$;

Step 3: Define the set $I_0 = \{i \in I : x_i = x_{i_1}\}$. In accordance with *Lemma 2.2*,

- solve the LAD problem for the function $\overline{f}(x; a) = a(x - x_{i_1}) + y_{i_1}$ by minimizing functional (4), denote the solution by a_2 and determine a new point $T_{i_2}(x_{i_2}, y_{i_2}) \in \Lambda$ for which $x_{i_2} \neq x_{i_1}$;
- Put $b_2 = -a_2 x_{i_1} - y_{i_1}$ and calculate $G_2 = \sum_{i=1}^{m} \omega_i |y_i - a_2 x_i - b_2|$,

Step 4: If $G_2 > G_1$, put $\{i_1 = i_2,\ G_1 = G_2\}$ and go to *Step 3*; If not, STOP.

It can be shown (see [14]) that Algorithm [Two Points] in finitely many steps leads to the best LAD-line. From the Algorithm and *Lemma 2.2* it is clear that the number of steps is less than the number of given points $T_1, \ldots, T_m$. Practically, considering a favourable choice of the initial point according to (5), the number of steps will be considerably smaller from m. In this way maximum efficiency of the algorithm is ensured, which is implementable in real time.

Example 2.5 *We are given the set of points* $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, 8\}$, *where* $\omega_1 = \ldots = \omega_8 = 1$, *and*

x_i	1	2	3	4	5	6	7	8
y_i	7	14	10	17	15	21	26	23

Algorithm [Two Points] is initialized so that first the centroid of the data has been calculated $T_p(5, 17) \notin \Lambda$. After that, the algorithm chooses the first point $T_1(1, 7) \in \Lambda$ and determines the linear function $f_1(x) = 2.5\,x + 4.5$, whose graph passes through those two points (see the left illustration in *Figure 1*). The sum of absolute deviations is $G_1 = 18$.

After that, the algorithm fixes the point $T_1(1, 7)$ and chooses a new point $T_6(6, 21)$ and a new linear function $f_2(x) = 2.8\,x + 4.2$, whose graph passes through those two points (see the right illustration in *Figure 1*). The sum of absolute deviations is now $G_2 = 17.4$, which also shows a global minimum for this problem.

Naturally, the same result is obtained by applying the module `NMinimize` using the *Nelder-Mead* method from the software tool *Mathematica* (see also [10], [19]).

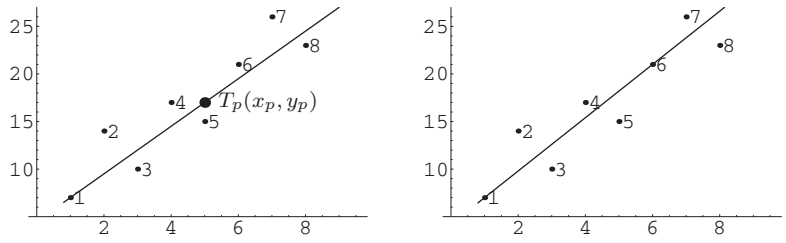


Figure 1: Illustration of the Algorithm[Two Points]

Example 2.6 *We are given the set of points* $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, m\}$, *where*

$$x_i = \tfrac{10i}{m}, \quad \omega_i = 1, \qquad i = 1, \ldots, m,$$
$$y_i = 3x_i + 2 + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2).$$

Efficiency of algorithms Algorithm [Two Points] will be compared with some special methods for searching the best LAD-line: Wesolowsky(1981) and Li and Arce (2004), but also with general minimization algorithms, which are also included in software tool *Mathematica*: `DifferentialEvolution, NelderMead, RandomSearch, and SimulatedAnnealing` (see also [10], [19]).

m	10	100	500	1 000	5 000	10 000
Algorithm [Two Points]	2(0)	3(0.03)	3(0.42)	2(1)	2(24.5)	3(75)
Algorithm (Li-Arce, [12])	2(0)	2(0.04)	3(1.03)	4(4.53)	1(105.8)	1(529)
Algorithm (Wesolowsky, [18])	3(0)	3(0.04)	3(1.09)	3(4.34)	3(127.8)	3(635)
Differential Evolution	(0.3)	(0.44)	(1.86)	(6.95)	(194)	(825)
Nelder-Mead	(0.14)	(0.16)	(1.36)	(6.42)	(186)	(775)
Random Search	(0.25)	(0.47)	(3.2)	(14.75)	(322)	(1168)
Simulated Annealing	(0.25)	(0.17)	(1.66)	(6.53)	(209)	(852)

Table 1: Comparison of algorithms for solving the LAD problem

For $m = 10, 100, 500, 1\,000, 5\,000, 10\,000$ by means of every method we will measure time in seconds (numbers in brackets given in *Table 1*), and with Algorithm [Two Points], Algorithm[Wesolowsky](1981) and Algorithm[Li and Arce] (2004) also the number of iterations.

As it can be seen, Algorithm [Two Points] comes to a global minimum in only few steps (most often 2–3 steps). Required time with Algorithm [Two Points] is a bit longer, which is probably the result of a direct application of *Mathematica*-module `Sort`. Considering development and low complexity of this algorithm, it can be expected that with careful programming this time would be significantly shorter. From *Table 1* it can be seen that there is significant superiority of both proposed algorithms in relation to other compared methods, which can be further improved by careful programming.

3 The best LAD-plane

Let $I = \{1, \ldots, m\}$, $m \geq 3$ be a set of indices, $\Lambda = \{T_i(x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ a set of points in space. The best LAD-plane should be determined, i.e. we should determine optimal parameters $a^*, b^*, c^* \in \mathbb{R}$ of the function $f(x, y; a, b, c) = ax + by + c$ such that

$$G(a^*, b^*, c^*) = \min_{(a,b,c)\in\mathbb{R}^3} G(a, b, c), \qquad G(a, b, c) = \sum_{i=1}^{m} |z_i - ax_i - by_i - c|. \tag{6}$$

All assertions that follow might be easily proved also in the case if data have corresponding weights $\omega_i > 0$. Analogously to *Theorem 2.1* there holds the following theorem.

Theorem 3.1 (Theorem on the existence of the best LAD-plane) *Let $I = \{1, \ldots, m\}$, $m \geq 3$ be a set of indices, and $\Lambda = \{T_i(x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ a set of points in space. Then there exists the best LAD-plane, i.e. problem (6) has a solution. Especially, if the data (x_i, y_i), $i = 1, \ldots, m$ lie on some line $\alpha x + \beta y + \gamma = 0$, $\alpha^2 + \beta^2 \neq 0$, then there exist infinitely many LAD-planes.*

Lemma 3.2 *Let $I = \{1, \ldots, m\}$, $m \geq 3$ be a set of indices and*

(i) $\Lambda = \{T_i(x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ a set of points in space such that $\min\limits_{i\in I} x_i < \max\limits_{i\in I} x_i$, $\min\limits_{i\in I} y_i < \max\limits_{i\in I} y_i$, whereby the data (x_i, y_i), $i \in I$ do not lie on a line.

(ii) $T_0(x_0, y_0, z_0) \in \mathbb{R}^3$,

(iii) $\overline{f}(x, y; a, b) = a(x - x_0) + b(y - y_0) + z_0$, $(a, b) \in \mathbb{R}^2$ *a linear function whose graph passes through the point* $T_0(x_0, y_0, z_0) \in \mathbb{R}^3$.

Then there exists $(a^*, b^*) \in \mathbb{R}^2$ *such that* $\overline{G}(a^*, b^*) = \min\limits_{(a,b)\in\mathbb{R}^2} \overline{G}(a, b)$, *where*

$$\overline{G}(a, b) = \sum_{i=1}^{m} |z_i - \overline{f}(x_i, y_i; a, b)| = \sum_{i=1}^{m} |z_i - z_0 - a(x_i - x_0) - b(y_i - y_0)|, \tag{7}$$

and the graph of the linear function $x \mapsto \overline{f}(x, y; a^*, b^*)$ *passes through at least two different points* $T_\nu(x_\nu, y_\nu, z_\nu), T_\mu(x_\mu, y_\mu, z_\mu) \in \Lambda$, *whereby* $(x_\mu, y_\mu) \neq (x_0, y_0)$, *and* $(x_\nu, y_\nu) \neq (x_0, y_0)$.

PROOF. Let $I_0 = \{i \in I : y_i = y_0\}$ and $I_1 = I \setminus I_0$. Notice that due to condition (i) there holds $I_1 \neq \emptyset$. It can be written

$$\overline{G}(a, b) = \sum_{i\in I_1} |y_i - y_0| \left| \frac{z_i - z_0}{y_i - y_0} - a\frac{x_i - x_0}{y_i - y_0} - b \right| + \sum_{i\in I_0} |(z_i - z_0) - a(x_i - x_0)|$$

According to *Lemma 1.1*, there exists $\mu \in I_1$, such that

$$\begin{aligned}
&\overline{G}(a, b) \geq \textstyle\sum_{i\in I_1} |y_i - y_0| \left| \frac{z_i - z_0}{y_i - y_0} - a\frac{x_i - x_0}{y_i - y_0} - \left(\frac{z_\mu - z_0}{y_\mu - y_0} - a\frac{x_\mu - x_0}{y_\mu - y_0} \right) \right| + \textstyle\sum_{i\in I_0} |(z_i - z_0) - a(x_i - x_0)| \\
&= \sum_{i\in I_1\setminus\{\mu\}} \left| (z_i - z_0) - (y_i - y_0)\frac{z_\mu - z_0}{y_\mu - y_0} - a\left(x_i - x_0 - (y_i - y_0)\frac{x_\mu - x_0}{y_\mu - y_0} \right) \right| + \sum_{i\in I_0} |(z_i - z_0) - a(x_i - x_0)| \\
&= \overline{G}\left(a, \tfrac{z_\mu - z_0}{y_\mu - y_0} - a\tfrac{x_\mu - x_0}{y_\mu - y_0} \right).
\end{aligned}$$

Obviously

$$\min_{(a,b)\in\mathbb{R}^2} \overline{G}(a, b) = \min_{a\in\mathbb{R}} \overline{G}\left(a, \frac{z_\mu - z_0}{y_\mu - y_0} - a\frac{x_\mu - x_0}{y_\mu - y_0} \right).$$

Let $J_1 = \{i \in I_1 \setminus \{\mu\} : x_i - x_0 - (y_i - y_0)\frac{x_\mu - x_0}{y_\mu - y_0} = 0\}$ and $J_0 = \{i \in I_0 : x_i = x_0\}$. Notice that due to condition (i) at least one of the sets $I_1 \setminus J_1$ or $I_0 \setminus J_0$ is non-empty. It can be written

$$\begin{aligned}
&\overline{G}\left(a, \tfrac{z_\mu - z_0}{y_\mu - y_0} - a\tfrac{x_\mu - x_0}{y_\mu - y_0} \right) = \\
&= \sum_{i\in I_1\setminus J_1} \left| (z_i - z_0) - (y_i - y_0)\tfrac{z_\mu - z_0}{y_\mu - y_0} - a\left(x_i - x_0 - (y_i - y_0)\tfrac{x_\mu - x_0}{y_\mu - y_0} \right) \right| + \sum_{i\in J_1} \left| (z_i - z_0) - (y_i - y_0)\tfrac{z_\mu - z_0}{y_\mu - y_0} \right| \\
&+ \sum_{i\in I_0\setminus J_0} |(z_i - z_0) - a(x_i - x_0)| + \sum_{i\in J_0} |(z_i - z_0)|.
\end{aligned}$$

According to *Lemma 1.1*, there exists $\nu' \in I_1 \setminus J_1$, i.e. there exists $\nu'' \in I_0 \setminus J_0$, such that

$$\begin{aligned}
&\textstyle\sum_{i\in I_1\setminus J_1} \left| (z_i - z_0) - (y_i - y_0)\frac{z_\mu - z_0}{y_\mu - y_0} - a\left(x_i - x_0 - (y_i - y_0)\frac{x_\mu - x_0}{y_\mu - y_0} \right) \right| \\
&\qquad \geq \textstyle\sum_{i\in I_1\setminus J_1} \left| (z_i - z_0) - (y_i - y_0)\frac{z_\mu - z_0}{y_\mu - y_0} - a'_\nu\left(x_i - x_0 - (y_i - y_0)\frac{x_\mu - x_0}{y_\mu - y_0} \right) \right|,
\end{aligned}$$

i.e.

$$\sum_{i\in I_0\setminus J_0} |(z_i - z_0) - a(x_i - x_0)| \geq \sum_{i\in I_0\setminus J_0} |(z_i - z_0) - a_{\nu''}(x_i - x_0)|,$$

where

$$a_{\nu'} = \frac{(z_{\nu'} - z_0) - (y_{\nu'} - y_0)\frac{z_\mu - z_0}{y_\mu - y_0}}{x_{\nu'} - x_0 - (y_{\nu'} - y_0)\frac{x_\mu - x_0}{y_\mu - y_0}}, \qquad a_{\nu''} = \frac{z_{\nu''} - z_0}{x_{\nu''} - x_0}.$$

Notice that $\nu', \nu'' \neq \mu$ and that there holds

$$\begin{aligned}\min_{(a,b)\in\mathbb{R}^2} \overline{G}(a,b) &= \min_{a\in\mathbb{R}} \overline{G}\left(a, \frac{z_\mu - z_0}{y_\mu - y_0} - a\frac{x_\mu - x_0}{y_\mu - y_0}\right) \\ &= \min\left\{\overline{G}\left(a_{\nu'}, \frac{z_\mu - z_0}{y_\mu - y_0} - a_{\nu'}\frac{x_\mu - x_0}{y_\mu - y_0}\right), \overline{G}\left(a_{\nu''}, \frac{z_\mu - z_0}{y_\mu - y_0} - a_{\nu''}\frac{x_\mu - x_0}{y_\mu - y_0}\right)\right\},\end{aligned}$$

It can be simply seen that the plane

$$z = a_{\nu'}(x - x_0) + \left(\frac{z_\mu - z_0}{y_\mu - y_0} - a_{\nu'}\frac{x_\mu - x_0}{y_\mu - y_0}\right)(y - y_0) + z_0$$

passes through mutually different points (x_0, y_0, z_0), (x_μ, y_μ, z_μ), $(x_{\nu'}, y_{\nu'}, z_{\nu'})$, whereby $(x_\mu, y_\mu) \neq (x_0, y_0)$ and $(x_{\nu'}, y_{\nu'}) \neq (x_0, y_0)$, whereas the plane

$$z = a_{\nu''}(x - x_0) + \left(\frac{z_\mu - z_0}{y_\mu - y_0} - a_{\nu''}\frac{x_\mu - x_0}{y_\mu - y_0}\right)(y - y_0) + z_0$$

passes through mutually different points (x_0, y_0, z_0), (x_μ, y_μ, z_μ), $(x_{\nu''}, y_{\nu''}, z_{\nu''})$, whereby $(x_\mu, y_\mu) \neq (x_0, y_0)$ and $(x_{\nu''}, y_{\nu''}) \neq (x_0, y_0)$. □

The following theorem shows that under natural conditions on the data there exists the best LAD-plane. This result will be used for developing an efficient algorithm for searching the best LAD-plane.

Theorem 3.3 *Let $I = \{1, \ldots, m\}$, $m \geq 3$ be a set of indices and $\Lambda = \{T_i(x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ a set of points in space, such that $\min_{i\in I} x_i < \max_{i\in I} x_i$, $\min_{i\in I} y_i < \max_{i\in I} y_i$, whereby the data (x_i, y_i), $i \in I$ do not lie on a line.*

Then there exists the best LAD-plane passing through at least three different points from Λ.

Proof. According to *Theorem 3.1*, there exists the best LAD-plane $f(x, y; a^*, b^*, c^*) = a^*x + b^*y + c^*$ with optimal parameters a^*, b^*, c^*.

According to *Lemma 1.1*, there exists $\mu \in I$, such that the minimum of the functional

$$G(a^*, b^*, c) = \sum_{i=1}^{m} |z_i - a^*x_i - b^*y_i - c|,$$

is attained for $c^+ = z_\mu - a^*x_\mu - b^*y_\mu$, i.e.

$$G(a^*, b^*, c^*) = \min_{c\in\mathbb{R}} G(a^*, b^*, c) = G(a^*, b^*, c^+).$$

It means that also at point (a^*, b^*, c^+) functional G attains its global minimum, and since $z_\mu = a^* x_\mu + b^* y_\mu + c^*$, it means that there exists the best LAD-plane passing through the point $T_\mu(x_\mu, y_\mu, z_\mu)$ given by

$$\overline{f}(x, y; a^*, b^*) := f(x, y; a^*, b^*, z_\mu - a^* x_\mu - b^* y_\mu) = a^*(x - x_\mu) + b^*(y - y_\mu) + z_\mu.$$

According to *Lemma 3.2*, then there exist at least two more different points from the set $\Lambda \backslash \{T_\mu\}$, through which there passes the best LAD-plane. □

Remark 3.4 *Notice that* Theorem 3.3 *is proved by using* Lemma 3.2*, whereas* Lemma 3.2 *is proved by applying* Lemma 1.1*. Thus, generally the best weighted LAD-line does not have to be unique. See also e. g. [3], [16].*

3.1 Methods for searching the best LAD-plane

As with the LAD-line, this problem might be solved by applying general minimization methods without using derivatives (see e. g. [10]), methods based on Linear Programming (see e. g. [2], [16]) or various specializations of the Gauss-Newton method (see e. g. [6], [13], [16]).

3.1.1 Three Points Method

For the purpose of developing an algorithm for searching the best LAD-plane, notice first that *Theorem 3.3* refers to the fact that the best LAD-line should be searched for among those planes which pass through at least three different points of the set Λ. The following algorithm is based on that fact. To get as close as possible from the beginning to the best LAD-line, as the initial point we can choose the centroid of the data $T_p(x_p, y_p, z_p)$, where

$$x_p = \frac{1}{m}\sum_{i=1}^{m} x_i, \quad y_p = \frac{1}{m}\sum_{i=1}^{m} y_i \quad z_p = \frac{1}{m}\sum_{i=1}^{m} z_i \tag{8}$$

Let us first choose initial approximation (a_0, b_0), e.g. by solving the OLS problem

$$\sum_{i \in I_0} (y_i - y_p)^2 \left(\frac{z_i - z_p}{y_i - y_p} - a\frac{x_i - x_p}{y_i - y_p} - b \right)^2, \qquad I_0 = \{i \in I : y_i \neq y_p\}, \tag{9}$$

and define the plane passing through the centroid $\overline{f}(x, y) = a_0(x - x_p) + b_0(y - y_p) + z_p$. Define also

$$\overline{G}(a, b) = \sum_{i=1}^{m} |z_i - z_p - a(x_i - x_p) - b(y_i - y_p)|. \tag{10}$$

For $c_0 = z_p - a_0 x_p - b_0 y_p$ denote $G_0 = G(a_0, b_0, c_0)$.

In initialization of the algorithm the following two mutually different points will be searched for in the set Λ, such that each of them decreases the value of the minimizing function.

Notice first that if there exists such $i_0 \in I$ that $T_p = T_{i_0}$, then the term of the sum (10) for $i = i_0$ vanishes. Denote $I' = I \setminus \{i_0\}$ and $I_0 = \{i \in I' : y_i = y_p\}$. Now it can be written

$$\overline{G}(a_0, b_0) = 0 + \sum_{i \in I_0} |z_i - z_p - a_0(x_i - x_p)| + \sum_{i \in I' \setminus I_0} |y_i - y_p| \left| \frac{z_i - z_p}{y_i - y_p} - a_0 \frac{x_i - x_p}{y_i - y_p} - b_0 \right|. \quad (11)$$

By solving the weighted median problem for the second sum in (11) we obtain a new point $T_{i_1}(x_{i_1}, y_{i_1}, z_{i_1})$, $i_1 \in I' \setminus I_0$ from the set Λ and a new value of the parameter $b_1 = \frac{z_{i_1} - z_p}{y_{i_1} - y_p} - a_0 \frac{x_{i_1} - x_p}{y_{i_1} - y_p}$. Notice that for $c_1 = z_p - a_0 x_p - b_1 y_p$ there holds $G_1 = G(a_0, b_1, c_1) < G_0$. Therefore, it holds

$$\overline{G}(a_0, b_0) \geq \sum_{i \in I} \left| z_i - z_p - a_0(x_i - x_p) - \left(\frac{z_{i_1} - z_p}{y_{i_1} - y_p} - a_0 \frac{x_{i_1} - x_p}{y_{i_1} - y_p} \right) (y_i - y_p) \right| =: \overline{G}(a_0). \quad (12)$$

Notice that this sum cancels out for $i \in \{i_0, i_1\}$. By means of notations $I' = I \setminus \{i_0, i_1\}$ and $I_0 = \{i \in I' : (x_i - x_p) - \frac{x_{i_1} - x_p}{y_{i_1} - y_p}(y_i - y_p) = 0\}$ functional (12) can be written as

$$\begin{aligned} \overline{G}(a_0) &= \sum_{i \in I_0} \left| z_i - z_p - \tfrac{z_{i_1} - z_p}{y_{i_1} - y_p}(y_i - y_p) \right| \\ &+ \sum_{i \in I' \setminus I_0} |x_i - x_p - \tfrac{x_{i_1} - x_p}{y_{i_1} - y_p}(y_i - y_p)| \left| \frac{z_i - z_p - \frac{z_{i_1} - z_p}{y_{i_1} - y_p}(y_i - y_p)}{x_i - x_p - \frac{x_{i_1} - x_p}{y_{i_1} - y_p}(y_i - y_p)} - a_0 \right|. \end{aligned} \quad (13)$$

By solving the weighted median problem for the second sum in (13) we obtain a new point $T_{i_2}(x_{i_2}, y_{i_2}, z_{i_2})$, $i_2 \in I' \setminus I_0$ from the set Λ and a new value of the parameter $a_1 = \frac{z_{i_2} - z_p - \frac{z_{i_1} - z_p}{y_{i_1} - y_p}(y_{i_2} - y_p)}{x_{i_2} - x_p - \frac{x_{i_1} - x_p}{y_{i_1} - y_p}(y_{i_2} - y_p)}$.
Notice that for $c_1 = z_p - a_1 x_p - b_1 y_p$ there holds $G_2 = G(a_1, b_1, c_1) < G_1 < G_0$.

Now we can move on to algorithm development.

Algorithm [Three Points]

Step 1: Input points $T_i(x_i, y_i, z_i) \in \Lambda$, $i \in I = \{1, \ldots, m\}$ and check conditions from *Theorem 3.3*

Step 2: Determine point T_p according to (8) and initial approximation (a_0, b_0) by solving the OLS problem (9); Put $c_0 = z_p - a_0 x_p - b_0 y_p$ and calculate $G_0 = G(a_0, b_0, c_0)$.

Step 3: Put $I' = I \setminus \{i_0\}$ and $I_0 = \{i \in I' : y_i = y_p\}$ and by solving the weighted median problem for the second sum from (11) determine a new point $T_{i_1}(x_{i_1}, y_{i_1}, z_{i_1})$, $i_1 \in I' \setminus I_0$ from the set Λ and a new value of the parameter $b_1 = \frac{z_{i_1} - z_p}{y_{i_1} - y_p} - a_0 \frac{x_{i_1} - x_p}{y_{i_1} - y_p}$; Put $c_1 = z_p - a_0 x_p - b_1 y_p$ and calculate $G_1 = G(a_0, b_1, c_1)$.

Step 4: Put $I' = I \setminus \{i_0, i_1\}$ and $I_0 = \{i \in I' : (x_i - x_p) - \frac{x_{i_1} - x_p}{y_{i_1} - y_p}(y_i - y_p) = 0\}$ and by solving the weighted median problem for the second sum from (13) determine a new point $T_{i_2}(x_{i_2}, y_{i_2}, z_{i_2})$, $i_2 \in I' \setminus I_0$ from the set Λ and a new value of the parameter $a_1 = \frac{z_{i_2} - z_p - \frac{z_{i_1} - z_p}{y_{i_1} - y_p}(y_{i_2} - y_p)}{x_{i_2} - x_p - \frac{x_{i_1} - x_p}{y_{i_1} - y_p}(y_{i_2} - y_p)}$; Put $c_1 = z_p - a_1 x_p - b_1 y_p$ and calculate $G_2 = G(a_1, b_1, c_1)$.

Step 5: Knowing points T_{i_1} and T_{i_2}, in accordance with *Lemma 3.2* define the set

$$I_1 = \{i \in I \setminus \{i_2\} : y_i \neq y_{i_1} \quad \& \quad x_i - x_{i_1} - (y_i - y_{i_1})\frac{x_{i_2} - x_{i_1}}{y_{i_2} - y_{i_1}} \neq 0\},$$

in accordance with *Lemma 1.1* solve the weighted median problem

$$\sum_{i \in I_1} \left| (z_i - z_{i_1}) - (y_i - y_{i_1})\frac{z_{i_2} - z_{i_1}}{y_{i_2} - y_{i_1}} - a\left(x_i - x_{i_1} - (y_i - y_{i_1})\frac{x_{i_2} - x_{i_1}}{y_{i_2} - y_{i_1}}\right)\right| \to \min_a,$$

and by $\nu_1 \in I_1$ denote the index for which the minimum a_3 is attained; Put $b_3 := \frac{z_{i_2} - z_{i_1}}{y_{i_2} - y_{i_1}} - a_3\frac{x_{i_2} - x_{i_1}}{y_{i_2} - y_{i_1}}$ and $G_3 = G(a_3, b_3, z_{i_1} - a_3 x_{i_1} - b_3 y_{i_1})$;

Step 6: In accordance with *Lemma 3.2*, define the set $I_2 = \{i \in I : y_i = y_{i_1} \quad \& \quad x_i \neq x_{i_1}\}$;
If $I_2 \neq \emptyset$, in accordance with *Lemma 1.1*, solve the weighted median problem

$$\sum_{i \in I_2} |(z_i - z_{i_1}) - a(x_i - x_{i_1})| \to \min_a,$$

and by $\nu_2 \in I_2$ denote the index for which the minimum a_4 is attained; Put $b_4 := \frac{z_{i_2} - z_{i_1}}{y_{i_2} - y_{i_1}} - a_4\frac{x_{i_2} - x_{i_1}}{y_{i_2} - y_{i_1}}$ and $G_4 = G(a_4, b_4, z_{i_1} - a_4 x_{i_1} - b_4 y_{i_1})$;
else $G_4 = G_3$.

Step 7: If $\big[G_4 < G_3$, $G_3 = G_4$; $a_3 = a_4$; $b_3 = b_4\big]$;
If $\big[G_3 < G_2$, put $G_2 = G_3$; $\quad T_{i_1} = T_{i_2}$;

If $\big[G_3 < G_4$, put $T_{i_2} = T(x_{\nu_1}, y_{\nu_1}, z_{\nu_1})$,
else put $T_{i_2} = T(x_{\nu_2}, y_{\nu_2}, z_{\nu_2})\big]$;
and go to *Step 5*;

else STOP$\big]$

Solution: parameters: a_3, b_3, $c_3 = z_{i_1} - a_3 x_{i_1} - b_3 y_{i_1}$.

Example 3.5 *Let a set of points* $\Lambda = \{T_i(x_i, y_i, z_i) : i = 1, \ldots, 20\}$ *be given, where numbers* (x_i, y_i, z_i) *are given in the table below.*

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	8.8	9.9	2.8	9.5	9.1	4.1	3.7	6.5	8.	4.5	9.5	6	6.	4.7	3.	1.1	2.5	2	8.1	8.8
y_i	6.5	5.6	6.	3.6	9.3	3.9	9.6	7	9.2	9.2	7.5	8	4.4	3.4	7.1	10.	0.9	4	3.1	0.8
z_i	41	38	23	26	48	18	36	36	44	33	43	31	33	25	25	33	3	13	30	16

Algorithm [Three Points] is intialized such that first the centroid $T_p(5.93, 5.955, 29.75) \notin \Lambda$ of the data is calculated. After that, the algorithm chooses two new points $T_2(9.9, 5.6, 38)$, $T_5(9.1, 9.3, 48) \in \Lambda$ and determines a linear function $f_0(x, y) = -3.41774 + 2.3655x + 3.21416y$, whose graph passes through these three points (see Initial approximation in *Fig. 1*). The sum of absolute deviations is $G_0 = 50.3662$. After that the algorithm fixes points $T_2(9.9, 5.6, 38)$, $T_5(9.1, 9.3, 48) \in \Lambda$, chooses a new point $T_6(4.1, 3.9, 18)$ and determines a new linear function

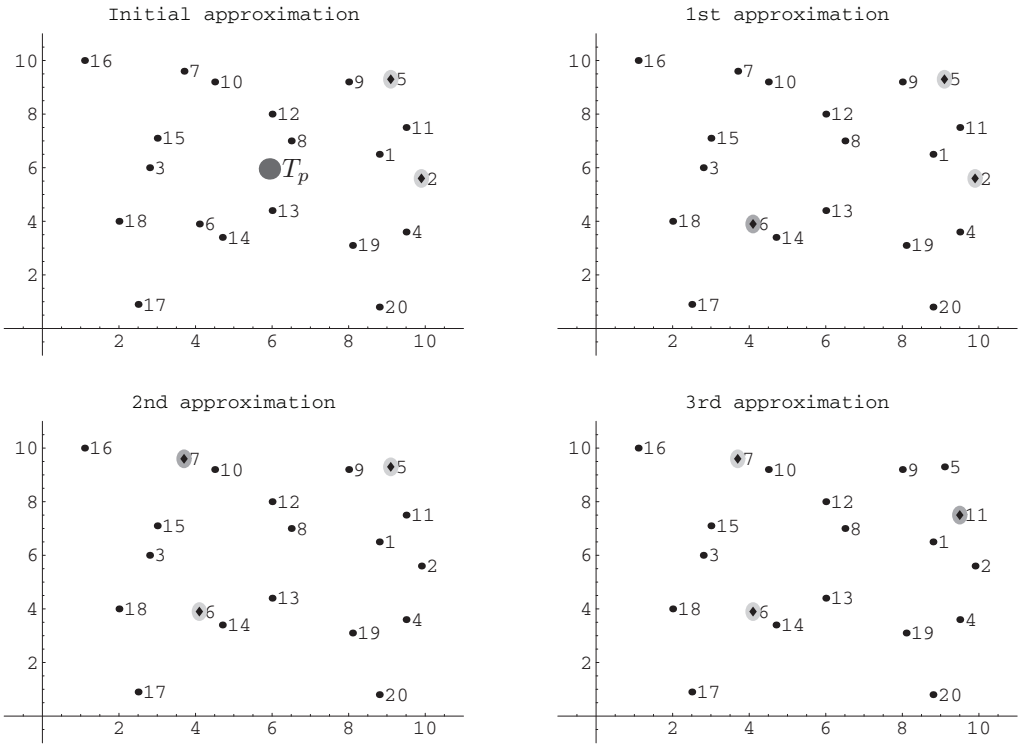


Figure 2: Illustration of Algorithm[Three Points]

$f_1(x, y) = -4.88782 + 2.49781x + 3.24277y$, whose graph passes through these three points (see 1st approximation in *Fig. 2*). The sum of absolute deviations is now $G_1 = 49.3085$.

In the next step the algorithm fixes points $T_5(9.1, 9.3, 48)$, $T_6(4.1, 3.9, 18) \in \Lambda$, chooses a new point $T_7(3.7, 9.6, 36)$ and determines a new linear function $f_2(x, y) = -4.84344 + 2.40705x + 3.32681y$, whose graph passes through these three points (vidi 2nd approximation in *Fig. 2*). The sum of absolute deviations is now $G_2 = 48.7143$.

In the next step the algorithm fixes points $T_6(4.1, 3.9, 18)$, $T_7(3.7, 9.6, 36) \in \Lambda$, chooses a new point $T_{11}(9.5, 7.5, 43)$ and determines a new linear function $f_3(x, y) = -4.86313 + 2.41155x + 3.32713y$, whose graph passes through these three points (see 3rd approximation in *Fig. 2*). The sum of absolute deviations is now $G_3 = 48.6958$.

In the next step the algorithm fixes points $T_7(3.7, 9.6, 36)$, $T_{11}(9.5, 7.5, 43) \in \Lambda$, chooses a new point $T_6(4.1, 3.9, 18)$. Since this combination of points has repeated, the process is completed.

Naturally, the same result is obtained by applying the module `NMinimize` using the *Nelder-Mead* method from the software system *Mathematica* (see also [10], [19]).

References

[1] A. ATIEG, G. A. WATSON, *Use of lp norms in fitting curves and surfaces to data* ANZIAM J. (Journal of Australian Mathematical Society, Series B – Applied Mathematics), **45**(2004)

C187-C200

[2] A. BARRODALE, F. D. K. ROBERTS, *An improved algorithm for discrete l_1 linear approximation*, SIAM J. Numer. Anal. **10**(1973) 839–848

[3] P. BLOOMFIELD, W. STEIGER, *Least Absolute Deviations: Theory, Applications, and Algorithms*, Birkhauser, Boston, 1983.

[4] J. BRIMBERG, H. JUEL, A. SCHÖBEL, *Properties of three-dimensional median line location models*, Annals of Operations Research **122**(2003) 71–85

[5] R. CUPEC, *Scene Reconstruction and Free Space Representation for Biped Walking Robots*, VDI Verlag, Düsseldorf, 2005.

[6] E. Z. DEMIDENKO, *Optimization and Regression*, Nauka, Moscow, 1989, in Russian

[7] Y. DODGE (ED.) *Statistical Data Analysis Based on the L_1-norm and Related Methods*, Proceedings of The Third International Conference on Statistical Data Analysis Based on the L_1-norm and Related Methods, Elsevier, Neuchâtel, 1997.

[8] K. P. HADELER, D. JUKIĆ, K. SABO, *Least squares problems for Michaelis Menten kinetics*, Mathematical Methods in the Applied Sciences, **30**(2007) 1231-1241

[9] D. JUKIĆ, K. SABO, R. SCITOVSKI, *Total least squares fitting Michaelis-Menten enzyme kinetic model function*, Journal of Computational and Applied Mathematics, **201**(2007) 230 - 246

[10] C. T. KELLEY, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.

[11] N. M. KORNEENKO, H. MARTINI, *Hyperplane approximation and related topics*, in: *New Trends in Discrete and Computational Geometry*, (J. Pach, Ed.), Springer-Verlag, Berlin, 1993.

[12] Y. LI, *A maximum likelihood approach to least absolute deviation regression*, EURASIP Journal on Applied Signal Processing, **12**(2004), 1762-1769

[13] M. R. OSBORNE, *Finite Algorithms in Optimization and Data Analysis*, Wilwy, Chichester, 1985.

[14] K. SABO, R. SCITOVSKI, *The best least absolute deviation line: properties and two efficient methods*, manuscript

[15] H. SPÄTH, G. A. WATSON, *On orthogonal linear l_1 approximation*, Numerische Mathematik **51**(1987) 531–543

[16] G. A. WATSON, *Approximation Theory and Numerical Methods*, John Wiley & Sons, Chichester, 1980

[17] G. A. WATSON, *On the Gauss-Newton method for l_1 orthogonal distance regression*, IMA J. Num. Anal. **22**(2002) 345-357

[18] G. O. WESOLOWSKY, *A new descent algorithm for the least absolute value regression problem*, Communications in Statistics, Simulation and Computation, **B10**(1981) 479-491

[19] S. WOLFRAM, *The Mathematica Book*, Wolfram Media, Champaign, 2007

Current address

I. Kuzmanović, K. Sabo, R. Scitovski, I. Vazler
Department of Mathematics, University of Osijek, Gajev trg 6, HR – 31 000 Osijek, Croatia
e-mail: ivana.kuzmanovic@mathos.hr, kristian.sabo@mathos.hr, rudolf.scitovski@mathos.hr, ivan.vazler@mathos.hr

Journal of Applied Mathematics

volume II (2009), number III

SOME PROBLEMS OF MICROHARDNESS OF METALS

NAVRÁTIL Vladislav, (CZ), NOVOTNÁ Jiřina, (CZ)

Abstract. The purpose of our contribution was to study the load dependence of the microhardness. This well known phenomena is called the Indentation Size Effect (ISE) and was investigated for two sets of specimen: pure copper and zinc. Variation of the microhardness with depth of indent (or applied load) was compared with various existing models. For the materials investigated the ISE is not artefact and can be explained by the hypothesis of presence of geometrically necessary dislocations, i.e. dislocations produced to accomodate the gradient in strain surrounding the indent.

MATHEMATICAL MODELS OF HARDNESS

Abstrakt. .Cílem našeho příspěvku je studium závislosti mikrotvrdosti na hloubce vtisku (nebo na aplikovaném zatížení). Tato závislost je zpravidla klesající a je známa jako Indentation Size Effect (ISE). Studovali jsme ji pro čistou měď a zinek a výsledky porovnali se známými modely. Naměřené výsledky ukazují, že pro námi vyšetřované materiály (Cu a Zn) není ISE artefaktem, ale je vlastností materiálu. Kvalitativně lze tento jev objasnit pomocí předpokladu o existenci tzv. „geometricky nutných dislokací“, tj. dislokací, vznikajících v okolí vtisku..

1 Introduction

According to definition *Hardness of material is ability of a body to resist permanent deformation.* Hardness measurements usually fall into three main cathegories: scratch hardness, indentation hardness and dynamic hardness. The methods most widely used in determining the hardness of metals are static indentation methods. In the Brinell test [1] the indenter consist of a hard steel ball made of tungsten carbide or even of diamond. Another type of indenter which has received wide use is the conical or pyramidal indenter as used in the Ludvik [2] and Vickers [3] hardness tests respectively. These indenters are now usually made of diamond. There are other types of indenters (Knoop, Berkovich) and other various methods (Rockwell) very well described in [3].

The diamond pyramidal indenter was first introduced in hardness measurements by Smith and Sandland [4] and was later developed by Messrs. Vickers – Armstrong , Ltd. The indenter is in the form of a square pyramid and the opposite faces make an angle of 136^0 with one another. The

choice of this angle is based on an analogy with the Brinell test. Vickers hardness HV is defined as the load divided by the surface area of the indentation:

$$HV = 0{,}927\,P \tag{1}$$

(P is yield pressure).

In making Vickers hardness measurements the lengths d of the diagonals of the indentation are measured. If the mean value of the diagonal is d, the yield pressure is $\frac{2P}{d^2}$ (P is the load).

Hence the Vickers hardness is

$$HV = 0{,}927\frac{2P}{d^2} \tag{2}$$

The loads usually rang from 10 to 1200 N (diagonal length is less than 1 mm). Smaller loads may also be used when a micro-hardness or nano-hardness is needed.

2. Indentation Size Effect

According to experimental results are mechanical properties, such as flow stress or hardness size dependent. This dependency is clearly observed when the size becomes significantly small, practically less than a micron (i.e. microhardness or nanohardness). This is in contrast to classical continuum mechanics plasticity theory where it is usually assumed that the mechanical properties of a material do not depend on scale.

In reference to indentation testing, the phenomenon whereby the indentation hardness increases with decreasing size is called the Indentation Size Effect (ISE).

Earlier researchers had offered different explanations for the ISE. All of these suggestions have some merit and are likely to have contributed to the ISE. There are some of the explanations of the ISE:

- abroaded surface layers and oxides,
- chemical contamination,
- inadequate measurement capability of small areas of indents,
- elastic recovery of indents,
- indenter-specimen friction.

However it has become clear that the inherent size dependence in many cases is not an artifact caused by oxide and friction. In the last two decades a number of authors [5,6,7] have argued that the ISE resulted from an increase in dislocation density at decreased loads caused by the presence of geometrically necessary dislocations, that is, dislocations produced to accommodate the gradient in strain surrounding the indent.

The simplest relationship describing ISE was proposed by Meyer [3]:

$$P = A.d^n \quad i.e.\ HV = C.d^{n-2} \tag{3}$$

Where A, n and C are constants.

According to energy-balance considerations and "Proportional Specimen Resistance" (PSR) model [8] the relationship between the applied load and the resulting indentation size is

$$P = a_1.d + a_2 d^2 \quad i.e.\ HV = b_2 + b_1 d^{-1} \tag{4}$$

Fröhlich et al. [9] related the $a_1 d$ term in (9) to the energy consumed in creating new surface (indentation facets and microcracking) while Li and Bradt [10] related this term to frictional and elastic contribution in their PSR model. The $a_2 d^2$ term was thought to be the work of permanent deformation [9] or the volume energy of deformation [10].
According to Gong et. al. [11] the modified equation should considered:

$$P = a_0 + a_1 d + a_2 d^2, \quad i.e.\ HV = b_0 d^{-2} + b_1 d^{-1} + b_2 \tag{5}$$

Where a_0, a_1, a_2, b_0, b_1, b_2 are constants

In present work the new form of hardness-indentation depth h ($h = \frac{1}{7} d$) is proposed:

$$y = y_0 + a.e^{-b.h} \tag{6}$$

Where $y_0 = HV_0$ is Vickers hardness extrapolated for great loads, $y = HV$ is Vickers hardness for indentation depth h and a and b are constants.

3. Experimental procedure

The samples studied were 99,99 % pure copper and zinc polycrystals. Microhardness measurements were made with Zeiss Neophot microscope and Hannemann microhardness tester. The diagonals of the indentation prints were measured by optical way. In order to minimize error, the hardness measurements were repeated to ten times.

4. Results and Conclusion

The results of our experiments are shown in Fig.1 and 2. To conclude, the ISE is not any artifact (unlike the results of Iost and Bigot [8]). From the Fig.1 we can see that for zinc (h.c.p. structure) the ISE exist unlike copper (f.c.c. structure – Fig.2), where ISE was not observed.

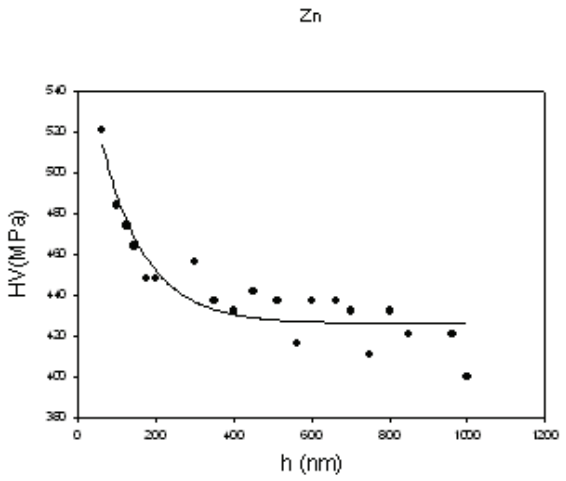


Fig.1. Hardness-indentation depth dependence for Zn

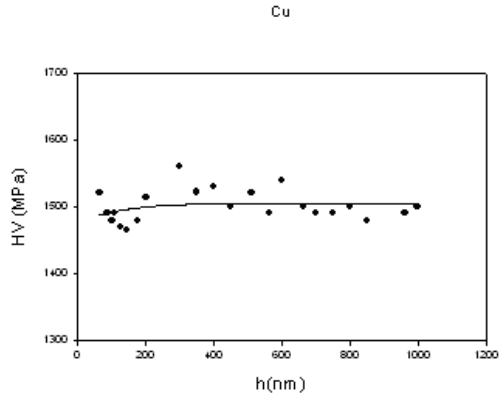


Fig.2. Hardness-indentation depth dependence for Cu

Our results confirm the theory of ISE based on the presence of geometrically necessary dislocations, i.e. dislocations produced to accommodate the gradient in strain surrounding the indent. At the beginning of indentation process the dislocations in f.c.c. metals have best conditions for moving in various slip planes (cross slip), than dislocations in h.c.p. metals. From the point of view of this qualitatively explanation the hardness of f.c.c. metals is independent on depth of indent unlike of h.c.p. metals, where the dislocations have not such conditions (so much alternative slip planes) and the ISE is observed.

References

[1.] BRINELL, J.A.: *J.Iron & Steel Inst*. **59**, 243
[2.] LUDWIK, P.: *Elemente der Tech. Mech*. Berlin 1909
[3.] TABOR, D.: *Microindentation Techniques in Materials Science and Engineering.*ASTM Speciál Publ. 889, pp. 129-159 (1985).
[4.] SMITH, R., SANDLAND, G.: *Proc. Inst. Mech: Engrs*. **1**, 623, (1922)
[5.] NIX, W.D., GAO, H.J.: *Mech. Phys. Solids* **46**, 411, (1998)
[6.] MA, Q., CLARKE, D.R.J.: *J. Mater. Res*. **10**, 853, (1995)
[7.] GURTIN, M.E.: *J.Mech. Phys. Solids*, **50**, 5, (2002)
[8.] IOST, A, BIGOT, R.: *J. Mater. Sci*. 31, 3573, (1996)
[9.] FRŐHLICH, P., GRAU, P., GELLMANN.: *Phys. Stat. Sol*. **42**, 79, (1977)
[10.] LI, H., BRADT, R.C.: *J. Mater. Sci*. **28**, 917, (1993)
[11.] GONG, J., WU, J., GUAN, Z.: *J. Mater. Sci. Lett*. **17**, 473, (1998)

Current Address

PhDr. Jiřina Novotná, PhD.,
katedra matematiky, Pedagogická fakulta Masarykovy University, Poříčí 31, 603 00 Brno.
e-mail: novotna@ped.muni.cz

Prof.RNDr. Vladislav Navrátil, CSc.,
katedra fyziky, Pedagogická fakulta Masarykovy University, Poříčí 7, 603 00 Brno.
e-mail:navratil@ped.muni.cz